Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

Multi-innovation gradient identification for input nonlinear state space systems ⋆

Xuehai Wang [∗,∗∗], **Yanjun Liu** [∗], **Feng Ding** [∗]

[∗] *Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, People's Republic of China (xuehaiwang735@163.com; yjl@jiangnan.edu.cn; fding@jiangnan.edu.cn);*
[∗∗] *School of Mathematics and Physics, Henan University of Urban Construction, Pingdingshan 467036, People's Republic of China (xuehaiwang735@163.com).*

Abstract: This paper presents a multi-innovation stochastic gradient algorithm for an input nonlinear state space system by deriving the identification model of the system and by decomposing the model into two sub-models. The basic idea is to design a state observer to estimate the unmeasurable states and to estimate interactively the unknown parameters of two subsystems by using the hierarchical identification principle. The simulation results show that the proposed algorithm is efficient.

Keywords: Recursive identification, Gradient search, Multi-innovation identification, State space model

1. INTRODUCTION

Nonlinearities are often encountered in fault detection and system identification (Jiang & Chowdhury, 2005; Zhang, et al., 2014; Wang & Ding, 2014). The nonlinear system modeling has been widely used in the industrial processes, such as the microbial batch fermentation and pH process (Chen, et al., 2014; Liu, 2013). Much effort was devoted to the identification of the block-oriented nonlinear systems, which include the input nonlinear systems, i.e., a static nonlinear block followed by a linear dynamic one, and the output nonlinear systems as the special cases (Chai, Loxton, Teo, & Yang, 2013; Enqvist & Ljung, 2005; Vörös, 2007). For example, Yu, Zhang, and Xie (2014) studied the parameter estimation problem for Hammerstein systems using a blind identification based deterministic estimation algorithm. Li and Wen (2013) transformed the known basis functions into the orthogonal basis functions and proposed a normalized iterative algorithm for Hammerstein systems by applying a fixed-point iteration technique.

Some algorithms have been reported for the nonlinear systems, such as the over-parametrization methods (Bai, 1998; Liu & Bai, 2007), the least squares methods (Krstic, 2009; Vörös, 2005), the iterative methods (Bai & Li, 2004; Vörös, 1999; Wang & Tang, 2014) and the maximum-likelihood methods (Hagenblad, Ljung, & Wills, 2008; Li, 2013). The over-parametrization methods transform the output of the original nonlinear system into a linear function on the parameter space so that any linear estimation algorithms can be applied (Ding & Chen, 2005). However, the resulting cross-product increases the dimensions of the identification parameters and leads to a heavy computational burden. The hierarchical identification algorithm is to decompose a system into two subsystems with small dimensions, and to identify the parameters of each subsystem respectively, thus the computational efficiency can be improved (Ding, 2013). Raja and Chaudhary (2014) presented a two-stage fractional least mean squares adaptive algorithm for controlled autoregressive moving average systems. Ding and Duan (2013) derived a two-stage recursive least squares algorithm and a two-stage gradient algorithm for Box-Jenkins models.

State space models can describe the dynamic of the systems and play a key role in the controller design (Dhawan & Kar, 2011; Wu, Wang, & Li, 2012), signal filtering (Levy & Nikoukhah, 2013) and system analysis (Hmida, Khémiri, Ragot, & Gossa, 2012; Mulders, Schoukens, & Vanbeylen, 2013; Pence, Fathy, & Stein, 2011). The identification of the state space models may involve not only the estimation of the unknown model parameters, but also the unmeasurable system states. The combined state and parameter estimation based recursive identification is important since it can be implemented with advanced control strategies or monitoring process behavior to improve the online control performance of the key quality variables. The maximum likelihood method or the expectation-maximization method have been reported for estimating the state and parameter simultaneously (Mader, Linke, et al., 2014). Tulsyan, Huang, Gopaluni, and Forbes (2013) presented a Bayesian approach for nonlinear state space models with an adaptive sequential-importance-resampling filter. Schön, Wills, and Ninness (2011) proposed an expectation-maximization algorithm for nonlinear dynamic systems in state-space form using a particle smoother.

⋆ This work was supported by the National Natural Science Foundation of China (No. 61304138) and the PAPD of Jiangsu Higher Education Institutions.

The importance of the canonical forms of the state space models is widely admitted. Mercère and Bako (2011) presented the canonical form of the multivariable systems and discussed its application in the subspace identification. Based on the observability canonical form, Ding (2014) derived a combined state and least squares algorithm for state space systems by means of the Kalman filter. Recently, for the nonlinear state space systems with one-time delay, Gu and Ding (2014) investigated the gradient based and the least squares based iterative algorithms. Wang, Ding, and Liu (2014) studied the least squares based parameter and state estimation algorithm for the input nonlinear state space systems. This paper develops a highly accurate recursive identification algorithm for input nonlinear state space systems using the hierarchical identification and the multi-innovation identification theory. The multi-innovation identification algorithm is introduced to improve the computational efficiency of the stochastic gradient methods by utilizing current innovations and past innovations repeatedly.

The rest of this paper is organized as follows. Section 2 derives the identification model for input nonlinear state space systems. Section 3 develops a multi-innovation stochastic gradient (MISG) algorithm to estimate the system parameters and states jointly. Section 4 provides an illustrative example to show the effectiveness of the proposed algorithms. Finally, concluding remarks are given in Section 5.

2. PROBLEM DESCRIPTION AND SYSTEM MODEL

Consider the following input nonlinear state space systems,

$$\begin{aligned}\boldsymbol{x}(t+1) &= \boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{b}\bar{u}(t), & (1)\\ \bar{u}(t) &= g(u(t)), & (2)\\ y(t) &= \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}(t)+w(t), & (3)\\ w(t) &= H(z)v(t), & (4)\end{aligned}$$

where $\boldsymbol{x}(t) := [x_1(t), x_2(t), \cdots, x_n(t)]^{\mathrm{T}} \in \mathbb{R}^n$ is the state vector, $\bar{u}(t) \in \mathbb{R}$ is the output of the nonlinear part and unmeasurable, $u(t) \in \mathbb{R}$ is the system input, $y(t) \in \mathbb{R}$ is the system output, $v(t) \in \mathbb{R}$ is a random noise with zero mean, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{c} \in \mathbb{R}^n$ are the parameter matrix and vectors of the system,

$$\boldsymbol{A} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ a_1 & a_2 & a_3 & \cdots & a_n \end{bmatrix} \in \mathbb{R}^{n\times n},$$

$$\boldsymbol{b} := [b_1, b_2, \cdots, b_n]^{\mathrm{T}} \in \mathbb{R}^n,$$

$$\boldsymbol{c} := [1, 0, \cdots, 0]^{\mathrm{T}} \in \mathbb{R}^n.$$

The disturbance $w(t) \in \mathbb{R}$ is a process noise, $H(z)$ is a finite impulse response model:

$$H(z) = 1 + h_1 z^{-1} + h_2 z^{-2} + \cdots + h_{n_h} z^{-n_h}.$$

The nonlinear output $\bar{u}(t)$ is a linear combination of a known basis with coefficients $\boldsymbol{\gamma} := (\gamma_1, \gamma_2, \cdots, \gamma_m)^{\mathrm{T}} \in \mathbb{R}^m$, and can be expressed as

$$\bar{u}(t) = g(u(t)) = \sum_{j=1}^{m} \gamma_j g_j(u(t)). \qquad (5)$$

From (1), we have

$$\begin{aligned}x_1(t) &= a_1x_1(t-n)+a_2x_2(t-n)+\cdots+a_nx_n(t-n)\\ &\quad +b_1\bar{u}(t-1)+b_2\bar{u}(t-2)+\cdots+b_n\bar{u}(t-n)\\ &= \boldsymbol{x}^{\mathrm{T}}(t-n)\boldsymbol{a}+\sum_{i=1}^{n} b_i \sum_{j=1}^{m}\gamma_j g_j(u(t-i))\\ &= \boldsymbol{x}^{\mathrm{T}}(t-n)\boldsymbol{a}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{G}(t)\boldsymbol{\gamma}, \qquad (6)\end{aligned}$$

where

$$\boldsymbol{G}(t) := \begin{bmatrix} g_1(u(t-1)) & g_2(u(t-1)) & \cdots & g_m(u(t-1)) \\ g_1(u(t-2)) & g_2(u(t-2)) & \cdots & g_m(u(t-2)) \\ \vdots & \vdots & & \vdots \\ g_1(u(t-n)) & g_2(u(t-n)) & \cdots & g_m(u(t-n)) \end{bmatrix} \in \mathbb{R}^{n\times m}. \qquad (7)$$

From (4), we have

$$\begin{aligned}w(t) &= \sum_{i=1}^{n_h} h_i v(t-i) + v(t)\\ &= \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{h}+v(t), \qquad (8)\end{aligned}$$

where

$$\boldsymbol{\varphi}(t) := [v(t-1), v(t-2), \cdots, v(t-n_h)]^{\mathrm{T}} \in \mathbb{R}^{n_h},$$

$$\boldsymbol{h} := [h_1, h_2, \cdots, h_{n_h}]^{\mathrm{T}} \in \mathbb{R}^{n_h}.$$

Inserting (6) and (8) into (3) yields

$$\begin{aligned}y(t) &= x_1(t)+w(t) & (9)\\ &= \boldsymbol{x}^{\mathrm{T}}(t-n)\boldsymbol{a}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{G}(t)\boldsymbol{\gamma}+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{h}+v(t). & (10)\end{aligned}$$

Define the intermediate variable

$$y_1(t) := y(t)-\boldsymbol{x}^{\mathrm{T}}(t-n)\boldsymbol{a}-\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{h}$$

and the information vectors

$$\boldsymbol{\eta}(\boldsymbol{\gamma},t) := \begin{bmatrix} \boldsymbol{x}(t-n) \\ \boldsymbol{G}(t)\boldsymbol{\gamma} \\ \boldsymbol{\varphi}(t) \end{bmatrix} \in \mathbb{R}^{2n+n_h},$$

$$\boldsymbol{\xi}(\boldsymbol{b},t) := \boldsymbol{G}^{\mathrm{T}}(t)\boldsymbol{b} \in \mathbb{R}^m.$$

The model in (10) can be decomposed into the following two identification sub-models,

$$\begin{aligned}S_1: y(t) &= [\boldsymbol{x}^{\mathrm{T}}(t-n), \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t)]\boldsymbol{\vartheta}+v(t)\\ &= \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\gamma},t)\boldsymbol{\vartheta}+v(t), & (11)\\ S_2: y_1(t) &= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{G}(t)\boldsymbol{\gamma}+v(t)\\ &= \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{b},t)\boldsymbol{\gamma}+v(t). & (12)\end{aligned}$$

The parameter vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{\gamma}$ and $\boldsymbol{h}$ to be estimated are included in sub-models (11) and (12).

3. THE MULTI-INNOVATION STOCHASTIC GRADIENT ALGORITHM

Consider p data from $j = t-p+1$ to $j = t$. Define the cost functions

$$J_1(\boldsymbol{\vartheta}) := \sum_{j=0}^{p-1}[y(t-j) - \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\gamma}, t-j)\boldsymbol{\vartheta}]^2,$$

$$\begin{aligned}J_2(\boldsymbol{\gamma}) &:= \sum_{j=0}^{p-1}[y_1(t-j) - \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{b}, t-j)\boldsymbol{\gamma}]^2 \\ &= \sum_{j=0}^{p-1}[y(t-j) - \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{b}, t-j)\boldsymbol{\gamma} \\ &\quad -\boldsymbol{x}^{\mathrm{T}}(t-n-j)\hat{\boldsymbol{a}}(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t-j)\hat{\boldsymbol{h}}(t)]^2.\end{aligned}$$

Note that $\boldsymbol{\gamma}$ in $J_1(\boldsymbol{\vartheta})$ and $\boldsymbol{b}$ in $J_2(\boldsymbol{\gamma})$ are unknown and can be replaced by their corresponding estimates $\hat{\boldsymbol{\gamma}}(t-1)$ and $\hat{\boldsymbol{b}}(t)$. Using the gradient search and minimizing $J_1(\boldsymbol{\vartheta})$ and $J_2(\boldsymbol{\gamma})$ give

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t-1) + \frac{1}{r_1(t)}\sum_{j=0}^{p-1}\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\gamma}}(t-1), t-j) \\ &\quad \times[y(t-j) - \hat{\boldsymbol{\eta}}^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}(t-1), t-j)\hat{\boldsymbol{\vartheta}}(t-1)],\end{aligned} \tag{13}$$

$$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\gamma}}(t-1), t)\|^2, \tag{14}$$

$$\begin{aligned}\hat{\boldsymbol{\gamma}}(t) &= \hat{\boldsymbol{\gamma}}(t-1) + \frac{1}{r_2(t)}\sum_{j=0}^{p-1}\boldsymbol{\xi}(\hat{\boldsymbol{b}}(t), t-j) \\ &\quad \times[y(t-j) - \boldsymbol{\xi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(t), t-j)\hat{\boldsymbol{\gamma}}(t-1) \\ &\quad -\hat{\boldsymbol{x}}^{\mathrm{T}}(t-n-j)\hat{\boldsymbol{a}}(t) - \hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-j)\hat{\boldsymbol{h}}(t)],\end{aligned} \tag{15}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\xi}(\hat{\boldsymbol{b}}(t), t)\|^2, \tag{16}$$

where p represents the innovation length. From (8) and (3), we have

$$\begin{aligned}v(t) &= w(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{h} \\ &= y(t) - x_1(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{h}.\end{aligned}$$

Let $\hat{\boldsymbol{\varphi}}(t)$ be the estimate of $\boldsymbol{\varphi}(t)$,

$$\hat{\boldsymbol{\varphi}}(t) := [\hat{v}(t-1), \hat{v}(t-2), \cdots, \hat{v}(t-n_h)]^{\mathrm{T}} \in \mathbb{R}^{n_h}. \tag{17}$$

Define

$$\hat{\boldsymbol{x}}(t) := [\hat{x}_1(t), \hat{x}_2(t), \cdots, \hat{x}_n(t)]^{\mathrm{T}} \in \mathbb{R}^n. \tag{18}$$

Replacing $x_1(t)$, $\boldsymbol{\varphi}(t)$ and $\boldsymbol{h}$ with $\hat{x}_1(t)$, $\hat{\boldsymbol{\varphi}}(t)$ and $\hat{\boldsymbol{h}}(t)$, the estimate of $v(t)$ can be computed through

$$\hat{v}(t) = y(t) - \hat{x}_1(t) - \hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{h}}(t). \tag{19}$$

Define the stacked output vector $\boldsymbol{Y}(p,t)$ and the stacked information matrices $\boldsymbol{\Gamma}(p,t)$, $\boldsymbol{\Xi}(p,t)$, $\boldsymbol{X}(p,t)$, $\boldsymbol{\Psi}(p,t)$ and $\boldsymbol{\Phi}(p,t)$ as

$$\boldsymbol{Y}(p,t) := [y(t), y(t-1), \cdots, y(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^p, \tag{20}$$

$$\begin{aligned}\boldsymbol{\Gamma}(p,t) &:= [\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\gamma}}(t-1), t), \hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\gamma}}(t-1), t-1), \cdots, \\ &\quad \hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\gamma}}(t-1), t-p+1)] \in \mathbb{R}^{(2n+n_h)\times p},\end{aligned} \tag{21}$$

$$\begin{aligned}\boldsymbol{\Xi}(p,t) &:= [\boldsymbol{\xi}(\hat{\boldsymbol{b}}(t), t), \boldsymbol{\xi}(\hat{\boldsymbol{b}}(t), t-1), \cdots, \\ &\quad \boldsymbol{\xi}(\hat{\boldsymbol{b}}(t), t-p+1)] \in \mathbb{R}^{m\times p},\end{aligned} \tag{22}$$

$$\begin{aligned}\boldsymbol{X}(p,t) &:= [\hat{\boldsymbol{x}}(t-n), \hat{\boldsymbol{x}}(t-n-1), \cdots, \hat{\boldsymbol{x}}(t-n-p+1)] \\ &\quad \in \mathbb{R}^{n\times p},\end{aligned} \tag{23}$$

$$\boldsymbol{\Phi}(p,t) := [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \cdots, \hat{\boldsymbol{\varphi}}(t-p+1)] \in \mathbb{R}^{n_h\times p}. \tag{24}$$

Introduce two innovation vectors

$$\begin{aligned}\boldsymbol{E}_1(p,t) &:= \boldsymbol{Y}(p,t) - \boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1) \in \mathbb{R}^p, \\ \boldsymbol{E}_2(p,t) &:= \boldsymbol{Y}(p,t) - \boldsymbol{\Xi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\gamma}}(t-1) \\ &\quad -\boldsymbol{X}^{\mathrm{T}}(p,t)\hat{\boldsymbol{a}}(t) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{h}}(t) \in \mathbb{R}^p.\end{aligned}$$

Then Equations (13) and (15) can equivalently be rewritten as

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\Gamma}(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t), \\ \hat{\boldsymbol{\gamma}}(t) &= \hat{\boldsymbol{\gamma}}(t-1) + \frac{\boldsymbol{\Xi}(p,t)}{r_2(t)}\boldsymbol{E}_2(p,t).\end{aligned}$$

The unknown state $\boldsymbol{x}(t)$ in the information vector $\boldsymbol{\eta}(\hat{\boldsymbol{\gamma}}(t-1), t)$ can be replaced with the state observer $\hat{\boldsymbol{x}}(t)$ in the preceding section, and the unknown $v(t-i)$ are replaced with its estimate $\hat{v}(t-i)$. Thus we can summarize the multi-innovation gradient identification algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\Gamma}(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t), \tag{25}$$

$$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\gamma}}(t-1), t)\|^2, \tag{26}$$

$$\hat{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(t-1) + \frac{\boldsymbol{\Xi}(p,t)}{r_2(t)}\boldsymbol{E}_2(p,t), \tag{27}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\xi}(\hat{\boldsymbol{b}}(t), t)\|^2, \tag{28}$$

$$\boldsymbol{E}_1(p,t) = \boldsymbol{Y}(p,t) - \boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{29}$$

$$\begin{aligned}\boldsymbol{E}_2(p,t) &= \boldsymbol{Y}(p,t) - \boldsymbol{\Xi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\gamma}}(t-1) \\ &\quad -\boldsymbol{X}^{\mathrm{T}}(p,t)\hat{\boldsymbol{a}}(t) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{h}}(t),\end{aligned} \tag{30}$$

$$\boldsymbol{\xi}(\hat{\boldsymbol{b}}(t), t) = \boldsymbol{G}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t), \tag{31}$$

$$\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\gamma}}(t-1), t) = \begin{bmatrix}\hat{\boldsymbol{x}}(t-n) \\ \boldsymbol{G}(t)\hat{\boldsymbol{\gamma}}(t-1) \\ \hat{\boldsymbol{\varphi}}(t)\end{bmatrix}, \hat{\boldsymbol{\vartheta}}(t) = \begin{bmatrix}\hat{\boldsymbol{a}}(t) \\ \hat{\boldsymbol{b}}(t) \\ \hat{\boldsymbol{h}}(t)\end{bmatrix}, \tag{32}$$

$$\hat{\boldsymbol{\gamma}}(t) = [\hat{\gamma}_1(t), \hat{\gamma}_2(t), \cdots, \hat{\gamma}_m(t)]^{\mathrm{T}}. \tag{33}$$

Based on the Kalman filtering theory, the following state observer algorithm can be used for estimating the state vector $\boldsymbol{x}(t)$,

$$\hat{\boldsymbol{x}}(t+1) = \hat{\boldsymbol{A}}(t)\hat{\boldsymbol{x}}(t) + \hat{\boldsymbol{b}}(t)\hat{u}(t), \tag{34}$$

$$\hat{\boldsymbol{A}}(t) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \hat{a}_1(t) & \hat{a}_2(t) & \hat{a}_3(t) & \cdots & \hat{a}_n(t)\end{bmatrix}, \tag{35}$$

$$\hat{\boldsymbol{b}}(t) = [\hat{b}_1(t), \hat{b}_2(t), \cdots, \hat{b}_n(t)]^{\mathrm{T}}, \tag{36}$$

$$\hat{u}(t) = \sum_{i=1}^{m}\hat{\gamma}_i(t)g_i(u(t)). \tag{37}$$

Equations (17)–(37) form the multi-innovation stochastic gradient algorithm for Hammerstein state space systems.

4. EXAMPLE

Consider the following Hammerstein state space system,

$$\begin{aligned}
\boldsymbol{x}(t+1) &= \begin{bmatrix} 0 & 1 \\ 0.84 & 0.13 \end{bmatrix} \boldsymbol{x}(t) + \begin{bmatrix} 0.71 \\ 0.7042 \end{bmatrix} \bar{u}(t), \\
y(t) &= [1,0]\boldsymbol{x}(t) + H(z)v(t), \\
\bar{u}(t) &= 0.05u(t) + 0.09u^2(t), \\
H(z) &= 1 + 0.43z^{-1}.
\end{aligned}$$

The parameter vector to be estimated is

$$\begin{aligned}
\boldsymbol{\Theta} &:= [a_1, a_2, b_1, b_2, h_1, \gamma_1, \gamma_2]^{\mathrm{T}} \\
&= [0.84, 0.13, 0.71, 0.7042, 0.430, 0.05, 0.09]^{\mathrm{T}},
\end{aligned}$$

where $\|\boldsymbol{b}\|^2 = b_1^2 + b_2^2 = 1$. In simulation, the input $\{u(t)\}$ is taken as an uncorrelated stochastic signal sequence with zero mean and unit variance, and $\{v(t)\}$ as a white noise sequence with zero mean and variance $\sigma^2 = 0.30^2$. Take the data length $L = 3000$ and use the MISG algorithm to estimate the parameters of this system. The parameter estimates with different innovation lengths $p = 1$, $p = 3$ and $p = 8$ are shown in Tables 1–3, and the estimation errors $\delta := \|\hat{\boldsymbol{\Theta}}(t) - \boldsymbol{\Theta}\|/\|\boldsymbol{\Theta}\|$ versus t are shown in Figure 1.

From Tables 1–3 and Figure 1, we can draw the following conclusions.

- The estimation errors given by the stochastic gradient algorithm and the MISG algorithm become smaller with increasing t – see Tables 1–3.
- For the MISG algorithm, a larger innovation length p leads to a faster convergence rate – see Figure 1.

5. CONCLUSIONS

A multi-innovation stochastic gradient identification algorithm is presented for input nonlinear state space systems. The simulation test indicates that increasing the innovation length can improve the parameter estimation accuracy. The proposed algorithm can be extended to other nonlinear systems (Yu, Mao, Jia, & Yuan, 2014) and time-delay systems (Li & Shi, 2012).

REFERENCES

Bai E.W. (1998). An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*, 34(3), 333-338.

Bai E.W., and Li D. (2004). Convergence of the iterative Hammerstein system identification algorithm. *IEEE Transactions on Automatic Control*, 49(11), 1929-1940.

Chai Q.Q., Loxton R., Teo K.L., and Yang C.H. (2013). Time-delay estimation for nonlinear systems with piecewise-constant input. *Applied Mathematics and Computation*, 219(17), 9543-9560.

Chen X., Xue A., Peng D.L., and Guo Y.F. (2014). Modeling of pH neutralization process using fuzzy recurrent neural network and DNA based NSGA-II. *Journal of the Franklin Institute – Engineering and Applied Mathematics*, 351(7), 3847-3864.

Dhawan A., and Kar H. (2011). An improved LMI-based criterion for the design of optimal guaranteed cost controller for 2-D discrete uncertain systems. *Signal Processing*, 91(4), 1032-1035.

Ding F. (2013). Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling. *Applied Mathematical Modelling*, 37(4), 1694-1704.

Ding F. (2014). Combined state and least squares parameter estimation algorithms for dynamic systems. *Applied Mathematical Modelling*, 38(1), 403-412.

Ding F., and Chen T. (2005). Identification of Hammerstein nonlinear ARMAX systems. *Automatica*, 41(9), 1479-1489.

Ding F., and Duan H.H. (2013). Two-stage parameter estimation algorithms for Box-Jenkins systems. *IET Signal Processing*, 7(8), 646-654.

Enqvist M., and Ljung L. (2005). Linear approximations of nonlinear FIR systems for separable input processes. *Automatica*, 41(3), 459-473.

Gu Y., and Ding F. (2014). Parameter estimation for an input nonlinear state space system with time delay, *Journal of the Franklin Institute – Engineering and Applied Mathematics*, 351(12), 5326-5339.

Hagenblad A., Ljung L., and Wills A. (2008). Maximum likelihood identification of Wiener models. *Automatica*, 44(11), 2697-2705.

Hmida F.B., Khémiri K., Ragot J., and Gossa M. (2012). Three-stage Kalman filter for state and fault estimation of linear stochastic systems with unknown inputs. *Journal of the Franklin Institute – Engineering and Applied Mathematics*, 349(7), 2369-2388.

Jiang B., and Chowdhury F.N. (2005). Parameter fault detection and estimation of a class of nonlinear systems using observers. *Journal of the Franklin Institute – Engineering and Applied Mathematics*, 342(7), 725–736.

Krstic M. (2009). On using least-squares updates without regressor filtering in identification and adaptive control of nonlinear systems. *Automatica*, 45(3), 731-735.

Levy B.C., and Nikoukhah R. (2013). Robust State Space Filtering Under Incremental Model Perturbations Subject to a Relative Entropy Tolerance. *IEEE Transactions on Automatic Control*, 58(3), 682-695.

Li J.H. (2013). Parameter estimation for Hammerstein CARARMA systems based on the Newton iteration. *Applied Mathematics Letters*, 26(1), 91-96.

Li H., and Shi Y. (2012). Robust H-infty filtering for nonlinear stochastic systems with uncertainties and random delays modeled by Markov chains, *Automatica*, 48(1), 159-166.

Li G.Q., and Wen C.Y. (2013). Convergence of fixed-point iteration for the identification of Hammerstein and Wiener systems. *International Journal of Robust and Nonlinear Control*, 23(13), 1510-1523.

Liu C.Y. (2013) Modelling and parameter identification for a nonlinear time-delay system in microbial batch fermentation. *Applied Mathematical Modelling*, 37(10-11), 6899-6908.

Liu Y., and Bai E.W. (2007). Iterative identification of Hammerstein systems. *Automatica*, 43(2), 346-354.

Mercère G., and Bako L. (2011). Parameterization and identification of multivariable state-space systems: A canonical approach. *Automatica*, 47(8), 1547-1555.

Mader W., Linke Y., Mader M., Sommerlade L., Timmer J., and Schelter B. (2014). A numerically efficient implementation of the expectation maximization algorithm for state space models. *Applied Mathematics and Computation*, 241, 222-232.

Table 1. The MISG estimates and errors with $p = 1$

t	100	200	500	1000	2000	3000	True
a_1	0.51359	0.53236	0.55166	0.56302	0.56413	0.57026	0.84000
a_2	0.29217	0.30890	0.32556	0.33592	0.33676	0.34249	0.13000
b_1	0.64088	0.64163	0.64053	0.64003	0.63957	0.63947	0.71000
b_2	0.76764	0.76701	0.76794	0.76835	0.76873	0.76882	0.70420
h_1	1.04698	1.02740	0.93650	0.93191	0.92790	0.92554	0.43000
γ_1	0.01925	0.05292	0.03502	0.03106	0.03033	0.02763	0.05000
γ_2	0.45292	0.41385	0.37070	0.33491	0.29097	0.27826	0.09000
$\delta(\%)$	58.43671	55.79285	49.28239	47.91747	46.60821	46.15376	

Table 2. The MISG estimates and errors with $p = 3$

t	100	200	500	1000	2000	3000	True
a_1	0.66639	0.68320	0.71943	0.72460	0.72949	0.73495	0.84000
a_2	0.17307	0.18519	0.21861	0.22351	0.22822	0.23356	0.13000
b_1	0.66106	0.65973	0.66033	0.66049	0.66048	0.66050	0.71000
b_2	0.75033	0.75150	0.75098	0.75083	0.75084	0.75083	0.70420
h_1	0.74457	0.74056	0.73754	0.73651	0.73403	0.73225	0.43000
γ_1	0.12732	0.11880	0.05048	0.02801	0.01935	0.01763	0.05000
γ_2	0.45465	0.34179	0.24566	0.16457	0.11551	0.09950	0.09000
$\delta(\%)$	37.82532	32.04669	27.57636	25.69972	25.06640	24.89608	

Table 3. The MISG estimates and errors with $p = 8$

t	100	200	500	1000	2000	3000	True
a_1	0.76889	0.78661	0.83522	0.84053	0.84585	0.84883	0.84000
a_2	0.07563	0.08015	0.12405	0.12874	0.13381	0.13672	0.13000
b_1	0.71084	0.70931	0.70940	0.70937	0.70935	0.70934	0.71000
b_2	0.70336	0.70490	0.70480	0.70484	0.70486	0.70486	0.70420
h_1	0.32987	0.36079	0.39332	0.40920	0.42416	0.43059	0.43000
γ_1	0.17129	0.08632	0.02044	0.03428	0.02372	0.03227	0.05000
γ_2	0.48730	0.29787	0.14186	0.07007	0.07667	0.07139	0.09000
$\delta(\%)$	31.52362	16.88089	5.08940	2.37246	2.22821	2.02323	

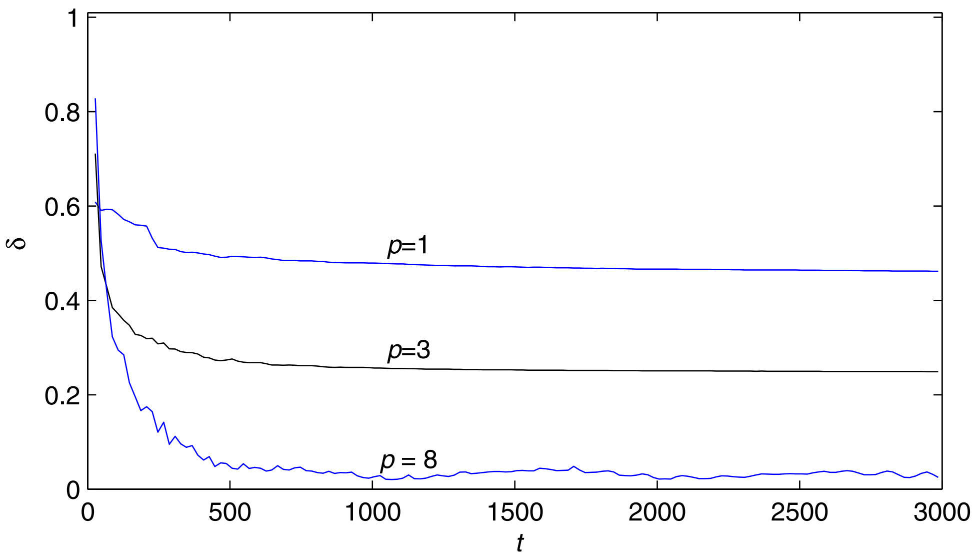

Fig. 1. Parameter estimation errors δ versus t with different innovation length p

Mulders A.V., Schoukens J., and Vanbeylen L. (2013). Identification of systems with localised nonlinearity: From state-space to block-structured models. *Automatica*, 49(5), 1392-1396.

Pence B.L., Fathy H.K., and Stein J.L. (2011). Recursive maximum likelihood parameter estimation for state space systems using poly- nomial chaos theory. *Automatica*, 47(11), 2420-2424.

Raja M.A.Z., and Chaudhary N.I. (2014). Two-stage fractional least mean square identification algorithm for parameter estimation of CARMA systems. *Signal Processing*, 107, 327-339.

Schön T.B., Wills A., and Ninness B. (2011). System identification of nonlinear state-space models. *Automatica*, 47(1), 39-49.

Tulsyan A., Huang B., Gopaluni R.B., and Forbes J.F. (2014). On simultaneous on-line state and parameter estimation in non-linear state-space models. *Journal of Process Control*, 23(4), 516-526.

Vörös J. (1999). Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities. *IEEE Transactions on Automatic Control*, 44(11), 2145-2149.

Vörös J. (2005). Identification of Hammerstein systems with time-varying piecewise-linear characteristics. *IEEE Transactions on Circuits and Systems– II: Express Brief*, 52 (12), 865-869.

Vörös J. (2007). Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities. *Systems & Control Letters*, 56(2), 99-105.

Wang X.H., and Ding F. (2014). Performance analysis of the recursive parameter estimation algorithms for multivariable Box-Jenkins systems, *Journal of the Franklin Institute – Engineering and Applied Mathematics*, 351(10), 4749-4764.

Wang D.Q., Ding F., and Liu X.M. (2014). Least squares algorithm for an input nonlinear system with a dynamic subspace state space model, *Nonlinear Dynamics*, 75(1-2), 49-61.

Wang C., and Tang T. (2014). Several gradient-based iterative estimation algorithms for a class of nonlinear systems using the filtering technique. *Nonlinear Dynamics*, 77(3), 769-780.

Wu H.N., Wang J.W., and Li H.X. (2012). Design of distributed H-infinity fuzzy controllers with constraint for nonlinear hyperbolic PDE systems. *Automatica*, 48(10), 2535-2543.

Yu F., Mao Z.Z., Jia M.X., and Yuan P. (2014). Recursive parameter identification of Hammerstein-Wiener systems with measurement noise, *Signal Processing*, 105, 137-147.

Yu C.P, Zhang C.S., and Xie L.H. (2014). A new deterministic identification approach to Hammerstein systems. *IEEE Transactions on Signal Processing*, 64(1), 131-140.

Zhang Y.Q., Shi P., Nguang S.K., Karimi H.R., and Agarwal R.K. (2014). Robust finite-time fuzzy H-infinity control for uncertain time-delay systems with stochastic jumps. *Journal of the Franklin Institute – Engineering and Applied Mathematics*, 351(8), 4211–4229.

