International Journal of Control, Automation and Systems 18(11) (2020) 2879-2889
http://dx.doi.org/10.1007/s12555-019-0799-5
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Maximum Likelihood Iterative Algorithm for Hammerstein Systems with Hard Nonlinearities

Yan Pu, Yongqing Yang*, and Jing Chen

Abstract: In this paper, we consider several iterative algorithms for Hammerstein systems with hard nonlinearities. The Hammerstein system is first simplified as a polynomial identification model through the key term separation technique, and then the parameters are estimated by using the maximum likelihood (ML) based gradient-based iterative algorithm. Furthermore, an ML least squares auxiliary variable algorithm and an ML bias compensation gradient-based iterative algorithm are developed to identify the saturation system with colored noise. Simulation results are included to illustrate the effectiveness of the proposed algorithms.

Keywords: Gradient search, Hammerstein system, key term separation, maximum likelihood, saturation nonlinearity.

1. INTRODUCTION

Nonlinear system modeling widely exists in system identification and has attracted a lot of attention in recent years [1, 2], inlcuding bilinear systems [3, 4]. The well-known nonlinear systems include the Hammerstein model which is composed of a static nonlinear part followed by a linear time invariant (LTI) part, and the Wiener model is composed of a LTI part followed by a static nonlinear part [5, 6]. They are already widely used in the field of applications ranging from system modeling to chemical processes. Nonlinearities can be roughly divided into two types: the hard nonlinearity and the polynomial nonlinearity. Compared with the polynomial nonlinearity, the hard nonlinearity has various kinds of structures, which leads to the difficulties to choose a general model structure to represent data from the hard nonlinear system [7, 8]. The saturation nonlinearity, one kind of hard nonlinearity, is often encountered in engineer practice, and there exist lots of controller design methods for systems with saturation nonlinearity [9, 10]. Notice that a robust controller always has the assumption that the parameters of the nonlinear systems should be known in advance. However, there are a few literatures on parameter identification of hard nonlinear systems [11]. The focus of this paper is to develop some identification algorithms for such nonlinear systems with different kinds of noises.

The existing estimation algorithms for nonlinear systems include the recursive algorithms [12], the iterative algorithms [13] and the multi-innovation identification methods [14]. Among these algorithms, the ML algorithm has many optimal properties such as sufficiency, efficiency and consistency [15], which makes the ML algorithm be used widely in nonlinear system identification [16]. The idea of the ML algorithm is that a likelihood function can be constructed based on the input-output data and parameters, and then the estimators can be obtained by maximizing the likelihood functions. For instance, Schön and Wills provided a maximum likelihood method for nonlinear state-space systems [17], Vanbeylen proposed blind maximum likelihood methods for Hammerstein systems and Wiener systems [18, 19].

In this paper, some maximum likelihood based iterative algorithms are proposed to identify the Hammerstein saturation systems with white and colored noises. First, the complex Hammerstein saturation system is simplified by utilizing the key term separation technique. Then a maximum likelihood gradient-based iterative (ML-GI) algorithm is presented to identify the Hammerstein system with white noise. Furthermore, a maximum likelihood based least squares auxiliary variable algorithm (ML-LSAV) algorithm and a maximum likelihood bias compensation gradient-based iterative (ML-BCGI) algorithm are developed for this nonlinear system with colored noises. Compared with the ML-LSAV algorithm, the ML-BCGI algorithm can get the unbiased parameters of the Hammerstein saturation system with colored noise.

Briefly, the rest of this paper is organized as follows:

Manuscript received September 23, 2019; revised January 15, 2020; accepted February 12, 2020. Recommended by Associate Editor Yang Tang under the direction of Editor Hamid Reza Karimi. This work is supported by the National Natural Science Foundation of China (No. 61973137) and the Joint Funds of the National Natural Science Foundation of China (No. U1734210).

Yan Pu, Yongqing Yang, and Jing Chen are with School of Science, School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, P. R China (e-mails: puyan@jiangnan.edu.cn, yongqingyang@163.com, chenjing1981929@126.com).
* Corresponding author.

Section 2 introduces a Hammerstein system with saturation nonlinearity. The ML gradient-based iterative algorithm is developed in Section 3. Section 4 discusses two different algorithms for the saturation system with colored noise, and the numerical examples are illustrated in Section 5. Finally, conclusions are drawn to summarize this paper in Section 6.

2. PROBLEM FORMULATION

The Hammerstein system which has saturation nonlinearity is shown in Fig. 1, and is written by

$$C(\zeta)y(\tau)=D(\zeta)\bar{x}(\tau)+v(\tau),$$
$$\bar{x}(\tau)=\eta(u(\tau)),$$

in which $\{u(\tau)\}$ and $\{\bar{x}(\tau)\}$ are the input and output sequence of the nonlinear part, and $\{u(\tau)\}$ is taken as a persistent excitation signal sequence with zero mean and unit variance, the output sequence of the system is $\{y(\tau)\}$, and $\{v(\tau)\}$ is a white noise sequence with zero mean, $C(\zeta)$ and $D(\zeta)$ are polynomials in the shift operator ζ^{-1} $[\zeta^{-1}y(\tau)=y(\tau-1)]$ and

$$C(\zeta)=1+c_1\zeta^{-1}+c_2\zeta^{-2}+\cdots+c_{n_c}\zeta^{-n_c}, \tag{1}$$
$$D(\zeta)=1+d_1\zeta^{-1}+d_2\zeta^{-2}+\cdots+d_{n_d}\zeta^{-n_d}. \tag{2}$$

The nonlinear part $\eta(u(\tau))$ is shown in Fig. 2, and $\bar{x}(\tau)$ can be expressed as

$$\bar{x}(\tau)=\eta(u(\tau))=\begin{cases} h_1, & u(\tau)\geqslant r,\\ s(u(\tau)), & -r<u(\tau)<r,\\ h_2, & u(\tau)\leqslant -r,\end{cases} \tag{3}$$

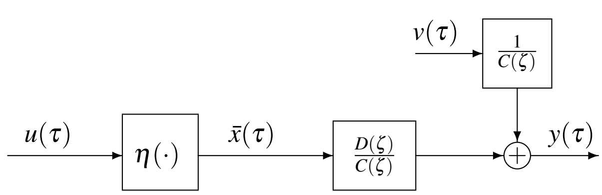


Fig. 1. The Hammerstein system.

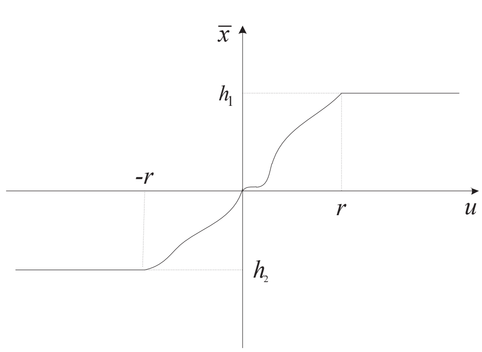


Fig. 2. The saturation Hammerstein system.

where $s(u(\tau))=\gamma_1u(\tau)+\gamma_2u^2(\tau)+\cdots+\gamma_nu^n(\tau)$, $h_1=\gamma_1r+\gamma_2r^2+\cdots+\gamma_nr^n$, $h_2=\gamma_1(-r)+\gamma_2(-r)^2+\cdots+\gamma_n(-r)^n$, γ_1, γ_2, $\cdots$, γ_n are the coefficients of the polynomial, $-r$, r $(r>0)$ are the saturation points.

A switching function $\xi(\cdot)$ is employed as follows:

$$\xi[u(\tau)]=\begin{cases} 1, & \text{if } u(\tau)\leqslant 0,\\ 0, & \text{if } u(\tau)>0,\end{cases} \tag{4}$$

and then $\bar{x}(\tau)$ can be expressed as

$$\begin{aligned}
\bar{x}(\tau)=&h_1\xi[r+u(\tau)]+h_2\xi[r-u(\tau)]\\
&+s[u(\tau)]\xi[-r-u(\tau)]\xi[-r+u(\tau)]\\
=&[\gamma_1r+\gamma_2r^2+\cdots+\gamma_nr^n]\xi[r-u(\tau)]\\
&+[\gamma_1(-r)+\gamma_2(-r)^2+\cdots+\gamma_n(-r)^n]\\
&\times\xi[u(\tau)+r]+[\gamma_1u(\tau)+\gamma_2u^2(\tau)+\cdots\\
&+\gamma_nu^n(\tau)]\xi[-u(\tau)-r]\xi[-r+u(\tau)]\\
=&\gamma_1r\left(\xi[-u(\tau)+r]+(-1)\xi[r+u(\tau)]\right)+\gamma_2r^2\\
&\times\left(\xi[-u(\tau)+r]+(-1)^2\xi[u(\tau)+r]\right)+\cdots\\
&+\gamma_nr^n\left(\xi[r-u(\tau)]+(-1)^n\xi[r+u(\tau)]\right)\\
&+[\gamma_1u(\tau)+\gamma_2u^2(\tau)+\cdots+\gamma_nu^n(\tau)]\\
&\times\xi[-u(\tau)-r]\xi[u(\tau)-r]\\
=&\sum_{j=1}^{n}\gamma_jr^j\xi[r-u(\tau)]+\sum_{j=1}^{n}\gamma_j(-r)^j\xi[u(\tau)+r]\\
&+\sum_{j=1}^{n}\gamma_ju^j(t)\xi[-u(\tau)-r]\xi[u(\tau)-r]\\
=&\sum_{j=1}^{n}\gamma_jr^jf_j(\tau)+\sum_{j=1}^{n}\gamma_ju^j(\tau)g(\tau),
\end{aligned} \tag{5}$$

where $f_j(\tau)=\xi[r-u(\tau)]+(-1)^j\xi[r+u(\tau)]$, $(j=1,2,\cdots,n)$ and $g(\tau)=\xi[-u(\tau)-r]\xi[u(\tau)-r]$. The equivalent form of the Hammerstein system can be written by

$$\begin{aligned}y(\tau)=&[1-C(\zeta)]y(\tau)+\bar{x}(\tau)\\&+[D(\zeta)-1]\bar{x}(\tau)+v(\tau).\end{aligned} \tag{6}$$

Substituting $\bar{x}(\tau)$ in (5) into (6) gives an analytic model

$$\begin{aligned}
y(\tau)=&-\sum_{i=1}^{n_c}c_iy(\tau-i)+\bar{x}(\tau)+\sum_{i=1}^{n_d}d_i\bar{x}(\tau-i)+v(\tau)\\
=&-\sum_{i=1}^{n_c}c_iy(\tau-i)+\sum_{j=1}^{n}\gamma_jr^jf_j(\tau)\\
&+\sum_{j=1}^{n}\gamma_ju^j(\tau)g(\tau)+\sum_{i=1}^{n_d}d_i\bar{x}(\tau-i)+v(\tau).
\end{aligned} \tag{7}$$

Let θ be the parameter vector and $\varphi(\tau)$ be the information vector

$$\begin{aligned}\theta=&[c_1,c_2,\cdots,c_{n_c},\gamma_1r,\gamma_2r^2,\cdots,\gamma_nr^n,\\&\gamma_1,\gamma_2,\cdots,\gamma_n,d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbb{R}^{n_0},\end{aligned}$$

$$\begin{aligned} n_0 &:= n_c + n_d + 2n, \\ \varphi(\tau) &= [-y(\tau-1), -y(\tau-2), \cdots, -y(\tau-n_c), \\ &\quad f_1(\tau), f_2(\tau), \cdots, f_n(\tau), u(\tau)g(\tau), \\ &\quad u^2(\tau)g(\tau), \cdots, u^n(\tau)g(\tau), \bar{x}(\tau-1), \cdots, \\ &\quad \bar{x}(\tau-n_d)]^{\mathrm{T}} \in \mathbb{R}^{n_0}. \end{aligned}$$

Then the Hammerstein system with saturation nonlinearity can be expressed as a simple form

$$y(\tau) = \varphi(\tau)^{\mathrm{T}}\theta + v(\tau). \tag{8}$$

3. THE ML ITERATIVE ESTIMATION ALGORITHM

Given the data sets $y_N := \{y(1), y(2), \cdots, y(N)\}$ and $u_N := \{u(1), u(2, \cdots, u(N))\}$, and $v(\tau)$ is uncorrelated with $u(i)$, $y(i)$ for $i < \tau$, the likelihood function $L(y_N|u_{N-1}, \theta)$ is equal to the joint conditional probability density function $P(y_N|u_{N-1}, \theta)$, which can be expressed as

$$\begin{aligned} L(y_N|u_{N-1},\theta) &= P(y_N|u_{N-1},\theta) \\ &= \prod_{\tau=1}^{N} p\Big(-\sum_{i=1}^{n_c} c_i y(\tau-i) + \bar{x}(\tau) + \sum_{i=1}^{n_d} d_i \bar{x}(\tau-i) + v(\tau) \\ &\quad \mid y(1), y(2), \cdots, y(\tau-1), u(1), \cdots, u(\tau-1), \theta\Big), \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\left(-\frac{1}{2\sigma^2}\sum_{\tau=1}^{N} v^2(\tau)\right) + l, \end{aligned} \tag{9}$$

where l is a constant. The ML estimate is obtained by maximizing the likelihood function, i.e.,

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(y_N|u_{N-1}, \theta).$$

Define the logarithm of $L(y_N|u_{N-1})$ as

$$\begin{aligned} l(y_N|u_N-1,\theta) &:= \ln \mathbf{L}(\mathbf{y_N}|\mathbf{u_N}-\mathbf{1},\theta) \\ &= \ln l - \frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{\tau=1}^{N} v^2(\tau). \end{aligned} \tag{10}$$

Letting the derivative of the log-likelihood function equal to zero gives

$$\left.\frac{\partial l(y_L|u_{N-1},\theta)}{\partial\sigma^2}\right|_{\hat{\sigma}^2} = 0,$$

whose solution is given by

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{\tau=1}^{N} v^2(\tau). \tag{11}$$

Substituting (11) into (10) gives

$$l(y_N|u_{N-1},\theta) = k_1 - \frac{N}{2}\ln\frac{1}{N}\sum_{\tau=1}^{N} v^2(\tau), \tag{12}$$

where $k_1 = \ln l - \frac{N}{2}\ln 2\pi - \frac{N}{2}$. The maximum value of (12) can be achieved by minimizing the following cost function

$$J(\hat{\theta}) = \frac{1}{2}\sum_{\tau=1}^{N} v^2(\tau)\Big|_{\hat{\theta}}, \tag{13}$$

where $v(\tau)$ is given by

$$v(\tau) = C(\zeta)y(\tau) - D(\zeta)\bar{x}(\tau) = y(\tau) - \varphi^{\mathrm{T}}(\tau)\theta.$$

The vector form of (13) can be written as

$$J(\hat{\theta}) = \frac{1}{2}(Y_N - H_N^{\mathrm{T}}\theta)^{\mathrm{T}}(Y_N - H_N^{\mathrm{T}}\theta)\Big|_{\hat{\theta}_{ML}},$$

in which $H_N := [\varphi(1), \varphi(2), \cdots, \varphi(N)] \in \mathbb{R}^{n_0\times N}$ and $Y_N := [y(1), y(2), \cdots, y(N)]^{\mathrm{T}} \in \mathbb{R}^N$.

The parameter estimate $\hat{\theta}$ by using the ML gradient-based iterative (ML-GI) algorithm is as follows:

$$\begin{aligned} \hat{\theta}_k &= \hat{\theta}_{k-1} - \frac{\mu_k}{2}\mathrm{grad}[J(\hat{\theta}_{k-1})] \\ &= \hat{\theta}_{k-1} + \mu_k \hat{H}_N^{k-1}[Y_N - (\hat{H}_N^{k-1})^{\mathrm{T}}\hat{\theta}_{k-1}] \\ &= [I - \mu_k \hat{H}_N^{k-1}(\hat{H}_N^{k-1})^{\mathrm{T}}]\hat{\theta}_{k-1} + \mu_k \hat{H}_N^{k-1} Y_N, \end{aligned} \tag{14}$$

where μ_k is the convergence factor. In fact, in order to ensure the convergence of the parameter $\hat{\theta}_k$, all the eigenvalues of the matrix $[I - \mu_k \hat{H}_N^{k-1}(\hat{H}_N^{k-1})^{\mathrm{T}}]$ should be in the unit circle, which means the value of μ_k satisfies $0 < \mu_k < 2\{\lambda_{\max}[\hat{H}_N^{k-1}(\hat{H}_N^{k-1})^T]\}^{-1}$.

Since the unknown parameters $\gamma_1, \gamma_2, \cdots, \gamma_n$, r and inner variables $\bar{x}(\tau - i)$ are contained in the information vector $\varphi(\tau)$, the parameter estimate θ cannot be directly obtained by the gradient-based iterative algorithm in (14). The solution for this problem is to replace the unknown variables with their estimates by using the iterative estimation technique. The unknown variables are replaced with their estimates. Define

$$\begin{aligned} \hat{\varphi}^{k-1}(\tau) &:= [-y(\tau-1), -y(\tau-2), \cdots, -y(\tau-n_c), \\ &\quad \hat{f}_1^{k-1}(\tau), \hat{f}_2^{k-1}(\tau), \cdots, \hat{f}_n^{k-1}(\tau), \\ &\quad u(\tau)\hat{g}^{k-1}(\tau), u^2(\tau)\hat{g}^{k-1}(\tau), \cdots, \\ &\quad u^n(\tau)\hat{g}^{k-1}(\tau), \hat{\bar{x}}_{k-1}(\tau-1), \hat{\bar{x}}_{k-1}(\tau-2), \\ &\quad \cdots, \hat{\bar{x}}_{k-1}(\tau-n_d)]^{\mathrm{T}} \in \mathbb{R}^{n_0}, \\ &\quad n_0 := n_c + n_d + 2n, \\ \hat{H}_N^{k-1} &:= [\hat{\varphi}^{k-1}(1), \hat{\varphi}^{k-1}(2), \cdots, \hat{\varphi}^{k-1}(N)]. \end{aligned}$$

Replacing the unknown parameters $\gamma_1, \gamma_2, \cdots, \gamma_n$ and r in (5) with their estimates $\hat{\gamma}_1^{k-1}, \hat{\gamma}_2^{k-1}, \cdots, \hat{\gamma}_n^{k-1}$ and $\hat{r}_{k-1}$, then the estimator $\hat{\bar{x}}_{k-1}(\tau)$ can be computed by

$$\hat{\bar{x}}_{k-1}(\tau) = \sum_{j=1}^{n}\hat{\gamma}_j^{k-1}\hat{r}_{k-1}^j \hat{f}_j^{k-1}(\tau) + \sum_{j=1}^{n}\hat{\gamma}_j^{k-1}u^j(\tau)\hat{g}^{k-1}(\tau).$$

Thus the ML-GI identification algorithm can be summarized as follows:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \mu_k \hat{H}_N^{k-1}[Y_N - (\hat{H}_N^{k-1})^{\mathrm{T}}\hat{\theta}_{k-1}], \tag{15}$$

$$Y_N = [y(1), y(2), \cdots, y(N)]^{\mathrm{T}}, \tag{16}$$

$$\hat{H}_N^{k-1} = [\hat{\varphi}^{k-1}(1), \hat{\varphi}^{k-1}(2), \cdots, \hat{\varphi}^{k-1}(N)], \tag{17}$$

$$\begin{aligned}\hat{\varphi}^{k-1}(\tau) = [&-y(\tau-1), -y(\tau-2), \cdots, -y(\tau-n_c),\\ &\hat{f}_1^{k-1}(\tau), \hat{f}_2^{k-1}(\tau), \cdots, \hat{f}_n^{k-1}(\tau),\\ &u(\tau)\hat{g}^{k-1}(\tau), u^2(\tau)\hat{g}^{k-1}(\tau), \cdots,\\ &u^n(\tau)\hat{g}^{k-1}(\tau), \hat{\bar{x}}_{k-1}(\tau-1),\\ &\hat{\bar{x}}_{k-1}(\tau-2), \cdots, \hat{\bar{x}}_{k-1}(\tau-n_d)]^{\mathrm{T}},\end{aligned} \tag{18}$$

$$\begin{aligned}\hat{\bar{x}}_{k-1}(\tau) = &\sum_{j=1}^{n} \hat{\gamma}_j^{k-1}\hat{r}_{k-1}^j \hat{f}_j^{k-1}(\tau)\\ &+ \sum_{j=1}^{n} \hat{\gamma}_j^{k-1} u^j(\tau)\hat{g}^{k-1}(\tau),\end{aligned} \tag{19}$$

$$\mu_k < 2\{\lambda_{\max}[\hat{H}_N^{k-1}(\hat{H}_N^{k-1})^T]\}^{-1}. \tag{20}$$

Remark 1: The Hammerstein saturation system in this section contains Gaussian white noise. However, we may encounter colored noise models in some practical problems. Then the identification of Hammerstein saturation systems with colored noise will be discussed in the next section.

The flowchart of calculating the estimate $\hat{\theta}_k$ in the ML-GI algorithm for the Hammerstein saturation nonlinear system is presented in Fig. 3.

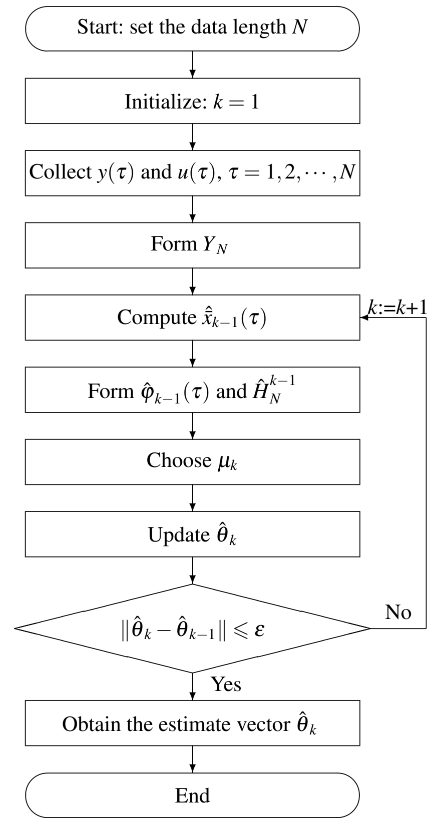


Fig. 3. The flowchart of the ML-GI algorithm for $\hat{\theta}_k$.

4. HAMMERSTEIN SATURATION SYSTEM WITH COLORED NOISE

In this section, some identification methods for the Hammerstein saturation system with colored noise are proposed. Consider the system

$$y(\tau) = \frac{D(\zeta)}{C(\zeta)}\bar{x}(\tau) + \frac{G(\zeta)}{C(\zeta)}v(\tau), \tag{21}$$

where $G(\zeta) := 1 + g_1\zeta^{-1} + g_2\zeta^{-2} + \cdots + g_{n_g}\zeta^{-n_g}$, the definition of $C(\zeta), D(\zeta)$ and the saturation nonlinearity part are mentioned in Section 2.

By utilizing the key term separation technique, the nonlinear system can be transformed as

$$\begin{aligned}y(\tau) = &-\sum_{i=1}^{n_c} c_i y(\tau-i) + \sum_{j=1}^{n} \gamma_j r^j f_j(\tau) + \sum_{i=1}^{n_d} d_i \bar{x}(\tau-i)\\ &+ \sum_{j=1}^{n} \gamma_j u^j(\tau) g(\tau) + \sum_{i=1}^{n_g} g_i v(\tau-i) + v(\tau).\end{aligned} \tag{22}$$

Let ϑ and $\psi(\tau)$ be the parameter vector and information vector, respectively,

$$\begin{aligned}\vartheta := [&c_1, c_2, \cdots, c_{n_c}, \gamma_1 r, \gamma_2 r^2, \cdots, \gamma_n r^n, \gamma_1, \gamma_2, \cdots,\\ &\gamma_n, d_1, d_2, \cdots, d_{n_d}, g_1, g_2, \cdots, g_{n_g}]^{\mathrm{T}},\end{aligned} \tag{23}$$

$$n_1 := n_c + n_d + n_g + 2n,$$

$$\begin{aligned}\psi(\tau) := [&-y(\tau-1), -y(\tau-2), \cdots, -y(\tau-n_c),\\ &f_1(\tau), f_2(\tau), \cdots, f_n(\tau), u(\tau)g(\tau),\\ &u^2(\tau)g(\tau), \cdots, u^n(\tau)g(\tau), \bar{x}(\tau-1),\\ &\bar{x}(\tau-2), \cdots, \bar{x}(\tau-n_d), v(\tau-1),\\ &v(\tau-2), \cdots, v(\tau-n_g)]^{\mathrm{T}} \in \mathbb{R}^{n_1}.\end{aligned} \tag{24}$$

Then, the system can be written as

$$y(\tau) = \psi^{\mathrm{T}}(\tau)\vartheta + v(\tau). \tag{25}$$

4.1. The maximum likelihood least squares auxiliary variable algorithm

Case 1: Define the stacked information matrix $\Psi(N)$(N is the data length), the stacked noise vector $V(N)$ and the stacked output vector $Y(N)$ as

$$Y(N) := [y(1), y(2), \cdots, y(N)]^{\mathrm{T}} \in \mathbb{R}^N,$$

$$\Psi(N) := [\psi(1), \psi(2), \cdots, \psi(N)] \in \mathbb{R}^{n_1 \times N},$$

$$V(N) := [v(1), v(2), \cdots, v(N)]^{\mathrm{T}} \in \mathbb{R}^N.$$

The vector form of the model (25) can be written by

$$Y(N) = \Psi^{\mathrm{T}}(N)\vartheta + V(N). \tag{26}$$

The cost function is

$$J(\hat{\vartheta})=\frac{1}{2}[Y(N)-\Psi^{\mathrm{T}}(N)\vartheta]^{\mathrm{T}}[Y(N)-\Psi^{\mathrm{T}}(N)\vartheta].$$

Letting the derivative of the cost function equal to zero gives

$$\left.\frac{\partial J(\vartheta)}{\partial\vartheta}\right|_{\hat{\vartheta}}=0.$$

Then, the parameter estimate is

$$\hat{\vartheta}=[\Psi(N)\Psi^{\mathrm{T}}(N)]^{-1}\Psi(N)Y(N). \tag{27}$$

However, the information vector $\psi(\tau)$ contains the noise $v(\tau-i)$, which means that $\psi(\tau)$ is relevant to $v(\tau)$. So that $E\{[\Psi(N)\Psi^{\mathrm{T}}(N)]^{-1}\Psi(N)V(N)\}\neq 0$. Then

$$\begin{aligned}\mathrm{E}(\hat{\vartheta})&=\mathrm{E}\{[\Psi(N)\Psi^{\mathrm{T}}(N)]^{-1}\Psi(N)Y(N)\}\\&=\mathrm{E}\{[\Psi(N)\Psi^{\mathrm{T}}(N)]^{-1}\Psi(N)[\Psi(N)^{\mathrm{T}}\vartheta+V(N)]\}\\&=\vartheta+\mathrm{E}\{[\Psi(N)\Psi^{\mathrm{T}}(N)]^{-1}\Psi(N)V(N)\}\\&\neq\vartheta.\end{aligned} \tag{28}$$

Thus the parameter estimate of ϑ is biased. To overcome this problem, let $w(\tau)=D(z)v(\tau)$, and define the information vector $\Psi_w(t)$ and the vector ϑ_w as

$$\begin{aligned}\vartheta_w&:=[c_1,c_2,\cdots,c_{n_c},\gamma_1 r,\gamma_2 r^2,\cdots,\gamma_n r^n,\\&\qquad\gamma_1,\gamma_2,\cdots,\gamma_n,d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbb{R}^{n_0},\\\psi_w(\tau)&:=[-y(\tau-1),-y(\tau-2),\cdots,-y(\tau-n_c),\\&\qquad f_1(\tau),f_2(\tau),\cdots,f_n(\tau),u(\tau)g(\tau),\\&\qquad u^2(\tau)g(\tau),\cdots,u^n(\tau)g(\tau),\bar{x}(\tau-1),\\&\qquad\bar{x}(\tau-2),\cdots,\bar{x}(\tau-n_d)]^{\mathrm{T}}\in\mathbb{R}^{n_0},\\Y(N)&:=[y(1),y(2),\cdots,y(N)]^{\mathrm{T}}\in\mathbb{R}^{N},\\\Psi_w(N)&:=[\psi_w(1),\psi_w(2),\cdots,\psi_w(N)]\in\mathbb{R}^{n_0\times N},\\W(N)&:=[w(1),w(2),\cdots,w(N)]^{\mathrm{T}}\in\mathbb{R}^{N},\end{aligned}$$

where $w(\tau)=v(\tau)+\sum_{i=1}^{n_g}g_i v(\tau-i)$.

Then, the system (22) can be written as

$$\begin{aligned}y(\tau)&=-\sum_{i=1}^{n_c}c_i y(\tau-i)+\bar{x}(\tau)+\sum_{i=1}^{n_d}d_i\bar{x}(\tau-i)+w(\tau)\\&=\psi_w^{\mathrm{T}}(\tau)\vartheta_w+w(\tau),\end{aligned} \tag{29}$$

where

$$\begin{aligned}\mathrm{E}[w(\tau)]&=\mathrm{E}[v(\tau)+\sum_{i=1}^{n_g}g_i v(\tau-i)]\\&=\mathrm{E}[v(\tau)]+\sum_{i=1}^{n_g}g_i\mathrm{E}[v(\tau-i)]=0,\end{aligned} \tag{30}$$

$$\begin{aligned}w(\tau)&=v(\tau)+g_1v(\tau-1)+g_2v(\tau-2)+\cdots\\&\quad+g_{n_g}v(\tau-n_g),\\w(\tau-1)&=v(\tau-1)+g_1v(\tau-2)+g_2v(\tau-3)+\cdots\\&\quad+g_{n_g}v(\tau-1-n_g),\\&\mathrm{cov}[w(\tau),w(\tau-1)]\neq 0,\end{aligned} \tag{31}$$

which means $w(\tau)$ is a colored noise.

The vector form of the model (29) can be expressed as

$$Y(N)=\Psi_w^{\mathrm{T}}(N)\vartheta_w+W(N),$$

where $\mathrm{E}\{[\Psi_w(N)\Psi_w^{\mathrm{T}}(N)]^{-1}\Psi_w(N)W(N)\}\neq 0$. Then we can get $\mathrm{E}(\hat{\vartheta}_w)\neq\vartheta_w$, that is to say, the estimate ϑ_w of ϑ is still biased.

Case 2: The key to solve the problem in Case 1 is to construct the auxiliary variable information vector which is irrelevant with the noise vector. The maximum likelihood based least squares auxiliary variable (ML-LSAV) algorithm is proposed in this section.

Define the auxiliary variable information vector $\psi_w^*(\tau)$ to replace $\psi_w(\tau)$ in (29) as

$$\begin{aligned}\psi_w^*(\tau):=&[u(\tau-1),u(\tau-2),\cdots,u(\tau-n_c),\\&f_1(\tau),f_2(\tau),\cdots,f_n(\tau),u(\tau)g(\tau),\\&u^2(\tau)g(\tau),\cdots,u^n(\tau)g(\tau),\\&\bar{x}(\tau-1),\bar{x}(\tau-2),\cdots,\bar{x}(\tau-n_d)]^{\mathrm{T}}\in\mathbb{R}^{n_0}.\end{aligned}$$

Remark 2: The auxiliary variable information vector $\psi_w^*(\tau)$ is constructed by using the irrelevant variables $u(\tau-i)$ to replace the output $y(\tau-i)$ in the w-information vecter $\psi_w(\tau)$.

Define the new stacked information matrix $\Psi_w^*(N)$ as

$$\Psi_w^*(N):=[\psi_w^*(1),\psi_w^*(2),\cdots,\psi_w^*(N)]\in\mathbb{R}^{n_0\times N},$$

where $\vartheta_w^*:=[c_1,\ c_2,\ \cdots,\ c_{n_c},\ \gamma_1 r,\ \gamma_2 r^2,\ \cdots,\ \gamma_n r^n,\ \gamma_1,\ \gamma_2,\ \cdots,\ \gamma_n,\ d_1,\ d_2,\ \cdots,\ d_{n_d}]^{\mathrm{T}}\in\mathbb{R}^{n_0}$. The vector form of the model can be written as $Y(N)=\Psi_w^{*\mathrm{T}}(N)\vartheta_w^*+W(N)$. Then we get the estimate $\hat{\vartheta}_w^*=[\Psi_w^*(N)\Psi_w^{\mathrm{T}}(N)]^{-1}\Psi_w^*(N)Y(N)$. Since the auxiliary variable information vector is irrelevant with the noise vector, the expectation

$$\mathrm{E}\{[\Psi_w^*(N)\Psi_w^{\mathrm{T}}(N)]^{-1}\Psi_w^*(N)W(N)\}=0. \tag{32}$$

We get the expectation of the estimate $\hat{\vartheta}^*$ as

$$\begin{aligned}\mathrm{E}(\hat{\vartheta}_w^*)=&E\{[\Psi_w^*(N)\Psi_w^{\mathrm{T}}(N)]^{-1}\Psi_w^*(N)Y(N)\}\\=&\mathrm{E}\{[\Psi_w^*(N)\Psi_w^{\mathrm{T}}(N)]^{-1}\Psi_w^*(N)\\&\times[\Psi_w^{\mathrm{T}}(N)\vartheta_w+W(N)]\}\\=&\vartheta_w+\mathrm{E}\{[\Psi_w^*(N)\Psi_w^{\mathrm{T}}(N)]^{-1}\Psi_w^*(N)W(N)\}\\=&\vartheta_w^*,\end{aligned}$$

which means that $\hat{\vartheta}_w^*$ is unbiased.

Remark 3: Although the ML-LSAV algorithm can get the unbiased estimates of $c_1,c_2,\cdots,c_{n_c}$ and $d_1,d_2,\cdots,d_{n_d}$, it cannot estimate the parameters of the colored noise $g_1,g_2,\cdots,g_{n_g}$.

4.2. Maximum likelihood bias compensation gradient-based iterative algorithm

To overcome the limitation of the ML-LSAV algorithm, a maximum likelihood bias compensation gradient-based iterative (ML-BCGI) algorithm is developed to identify all the parameter estimates of the nonlinear system with colored noise. Since the unmeasurable noise variables $v(\tau-i)$ are contained in the information vector $\psi(\tau)$, the unknown noise variables can be replaced by their estimates in the information vector. Let $\hat{v}(\tau)$ and $\hat{\psi}(\tau)$ denote the estimates of $v(\tau)$ and $\psi(\tau)$ at time τ, respectively. Let $\hat{\vartheta}_k$ be the iterative estimate of ϑ at iteration k, and computing the estimates of $v_k(\tau)$ by using the parameter estimate at iteration $k-1$ gets

$$\hat{v}_k(\tau)=y(\tau)-\hat{\psi}^{\mathrm{T}}_{k-1}(\tau)\hat{\vartheta}_{k-1},\ \ \tau=1,2,\cdots,N.$$

The estimate of $\hat{\vartheta}$ using the maximum likelihood gradient-based iterative algorithm is as follows:

$$\begin{aligned}\hat{\vartheta}_k=&\hat{\vartheta}_{k-1}-\frac{\mu_k}{2}\operatorname{grad}[J(\hat{\vartheta}_{k-1})]\\=&\hat{\vartheta}_{k-1}+\mu_k\hat{\Psi}_{k-1}(N)[Y(N)-\hat{\Psi}^{\mathrm{T}}_{k-1}(N)\hat{\vartheta}_{k-1}]\\=&\hat{\vartheta}_{k-1}+\mu_k\hat{\Psi}_{k-1}(N)[\Psi^{\mathrm{T}}(N)-\hat{\Psi}^{\mathrm{T}}_{k-1}(N)]\vartheta\\&+\mu_k\hat{\Psi}_{k-1}(N)\hat{V}(N),\end{aligned}\tag{33}$$

where $\mu_k>0$ is the convergence factor. Let $\hat{e}_k=\hat{\vartheta}_k-\vartheta$ and $\hat{e}_{k-1}=\hat{\vartheta}_{k-1}-\vartheta$. From Equation (33), we have

$$\begin{aligned}\hat{e}_k=&\hat{e}_{k-1}+\mu_k\hat{\Psi}_{k-1}(N)\hat{\Psi}^{\mathrm{T}}_{k-1}(N)\hat{e}_{k-1}\\&+\mu_k\hat{\Psi}_{k-1}(N)\hat{V}(N),\\=&[I-\mu_k\hat{\Psi}_{k-1}(N)\hat{\Psi}^{\mathrm{T}}_{k-1}(N)]\hat{e}_{k-1}\\&+\mu_k\hat{\Psi}_{k-1}(N)\hat{V}(N),\end{aligned}\tag{34}$$

where $I\in\mathbb{R}^{n_0\times n_0}$ is the identity matrix.

From (34), the estimate $\hat{\vartheta}_k$ in (33) is biased. The idea of the ML-BCGI algorithm is to compensate the bias $\mu_k\hat{\Psi}_{k-1}(N)\hat{V}(N)$ of the estimate $\hat{\vartheta}_k$ into the gradient-based iterative algorithm, and $\hat{\vartheta}_k$ can be expressed as

$$\begin{aligned}\hat{\vartheta}_k=&\hat{\vartheta}_{k-1}+\mu_k\hat{\Psi}_{k-1}(N)[Y(N)-\hat{\Psi}^{\mathrm{T}}_{k-1}(N)\hat{\vartheta}_{k-1}]\\&-\mu_k\hat{\Psi}_{k-1}(N)\hat{V}_{k-1}(N).\end{aligned}$$

Then, the estimate of the ML-BCGI algorithm is unbiased. Furthermore, the parameters of the colored noise can also be estimated.

In summary, the ML-BCGI algorithm for the Hammerstein system with colored noise can be expressed as

$$\begin{aligned}\hat{\vartheta}_k=&\hat{\vartheta}_{k-1}-\mu_k\hat{\Psi}_{k-1}(N)\hat{V}_{k-1}(N)+\mu_k\\&\times\hat{\Psi}_{k-1}(N)[Y(N)-\hat{\Psi}^{\mathrm{T}}_{k-1}(N)\hat{\vartheta}_{k-1}],\end{aligned}\tag{35}$$

$$Y(N)=[y(1),y(2),\cdots,y(N)]^{\mathrm{T}},\tag{36}$$

$$\hat{V}_{k-1}(N)=[\hat{v}_{k-1}(1),\hat{v}_{k-1}(2),\cdots,\hat{v}_{k-1}(N)]^{\mathrm{T}},\tag{37}$$

$$\hat{\Psi}_{k-1}(N)=[\hat{\psi}_{k-1}(1),\hat{\psi}_{k-1}(2),\cdots,\hat{\psi}_{k-1}(N)],\tag{38}$$

$$\begin{aligned}\hat{\psi}_{k-1}(\tau)=&[-y(\tau-1),-y(\tau-2),\cdots,-y(\tau-n_c),\\&\hat{f}_1^{k-1}(\tau),\hat{f}_2^{k-1}(\tau),\cdots,\hat{f}_n^{k-1}(\tau),\\&u(\tau)\hat{g}^{k-1}(\tau),u^2(\tau)\hat{g}^{k-1}(\tau),\cdots,\\&u^n(\tau)\hat{g}^{k-1}(\tau),\hat{\hat{x}}_{k-1}(\tau-1),\\&\hat{\hat{x}}_{k-1}(\tau-2),\cdots,\hat{\hat{x}}_{k-1}(\tau-n_d),\\&\hat{v}_{k-1}(\tau-1),\hat{v}_{k-1}(\tau-2),\cdots,\\&\hat{v}_{k-1}(\tau-n_g)]^{\mathrm{T}},\end{aligned}\tag{39}$$

$$\begin{aligned}\hat{\hat{x}}_{k-1}(\tau)=&\sum_{j=1}^{n}\hat{\gamma}_j^{k-1}\hat{r}_{k-1}^j\hat{f}_j^{k-1}(\tau)\\&+\sum_{j=1}^{n}\hat{\gamma}_j^{k-1}u^j(\tau)\hat{g}^{k-1}(\tau),\end{aligned}\tag{40}$$

$$\hat{v}_{k-1}(\tau)=y(\tau)-\hat{\psi}^{\mathrm{T}}_{k-1}(\tau)\hat{\vartheta}_{k-1},\tag{41}$$

$$\begin{aligned}\hat{\vartheta}_{k-1}=&[c_1^{k-1},c_2^{k-1},\cdots,c_{n_c}^{k-1},\gamma_1^{k-1}r_{k-1},\\&\gamma_2^{k-1}r_{k-1}^2,\cdots,\gamma_n^{k-1}r_{k-1}^n,\gamma_1^{k-1},\\&\gamma_2^{k-1},\cdots,\gamma_n^{k-1},d_1^{k-1},d_2^{k-1},\cdots,d_{n_d}^{k-1},\\&g_1^{k-1},g_2^{k-1},\cdots,g_{n_g}^{k-1}]^{\mathrm{T}},\end{aligned}\tag{42}$$

$$\mu_k<2\{\lambda_{\max}[\hat{\Psi}_{k-1}(N)\hat{\Psi}^T_{k-1}(N)]\}^{-1}.\tag{43}$$

Remark 4: The ML-BCGI algorithm can not only estimate the parameters c_1, c_2, $\cdots$, c_{nc}, d_1, d_2, $\cdots$, d_{nd}, but also obtain the parameters g_1, g_2, $\cdots$, g_{ng}, which means that it is more effective than the ML-LSAV algorithm.

Remark 5: The ML-BCGI algorithm has heavy computation efforts. We can use the stochastic average gradient algorithm to reduce the computation burden and keep the convergence rate unchanged in our future research.

Remark 6: Compared with the traditional LS and GI algorithms, both the ML-LSAV and the ML-BCGI algorithms proposed in this paper can obtain the unbiased parameter estimates. The proposed algorithm in this paper can combine other recursive schemes [20–22], the particle filtering algorithms [23, 24], and the iterative schemes [25] to study the identification problems other linear and nonlinear systems [26–30].

The implementation of the ML-BCGI algorithm involves the following steps.

1) Let $k=1$, $\hat{\hat{x}}_0(\tau)=1/p_0$, $\hat{\vartheta}_0=\mathbf{1}/p_0$ with $\mathbf{1}$ being a column vector whose entries are all unity and $p_0=10^6$.

2) Let $\tau=1$, $y(\tau)=0,u(\tau)=0$, $\tau\leqslant0$, and give a small positive number ε.

3) Collect the input-output data $u(\tau)$ and $y(\tau)$, and form $Y(N)$ by (36).

4) Compute $\hat{\hat{x}}_{k-1}(\tau)$ by (40) and $\hat{v}_{k-1}(\tau)$ by (41), $\tau=1$, 2, $\cdots$, N. Form $\hat{V}_{k-1}(N)$ by (37).

5) Form $\hat{\psi}_{k-1}(\tau)$ by (39), $\tau=1$, 2, $\cdots$, N. Form $\hat{\Psi}_N^{k-1}$ by (38). Choose μ_k according to (43).

6) Update the parameter estimate $\hat{\vartheta}_k$ by (35).

7) Compare $\hat{\vartheta}_k$ with $\hat{\vartheta}_{k-1}$, if the value is less than or

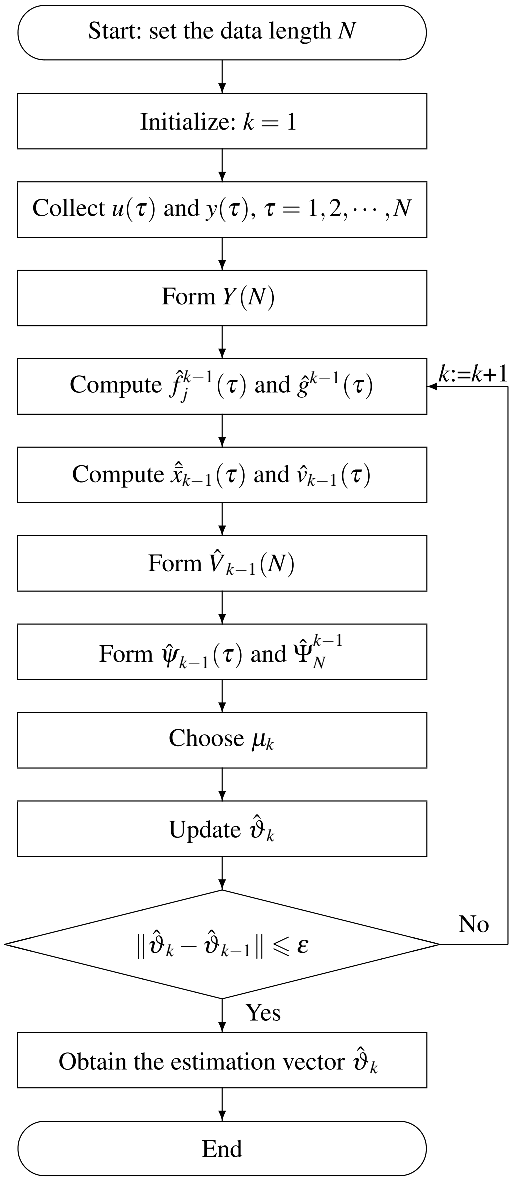


Fig. 4. The flowchart of the ML-BCGI algorithm for $\hat{\vartheta}_k$.

equal to ε, then obtain the $\hat{\vartheta}_k$ and terminate the procedure; otherwise, let $k = k+1$ and go to Step 4.

The flowchart of calculating the parameter estimate $\hat{\vartheta}_k$ in the ML-BCGI algorithm for the Hammerstein saturation nonlinear system is presented in Fig. 4.

5. NUMERICAL EXAMPLES

Example 1: Consider the following Hammerstein system with saturation nonlinearities:

$$y(\tau) = \frac{D(\zeta)}{C(\zeta)}\bar{x}(\tau) + \frac{1}{C(\zeta)}v(\tau),$$

the saturation nonlinearities are shown in Fig. 2 with $r = 0.4$, $\gamma_1 = 0.8$, $\gamma_2 = -0.5$, and $\gamma_3 = 0.95$, and

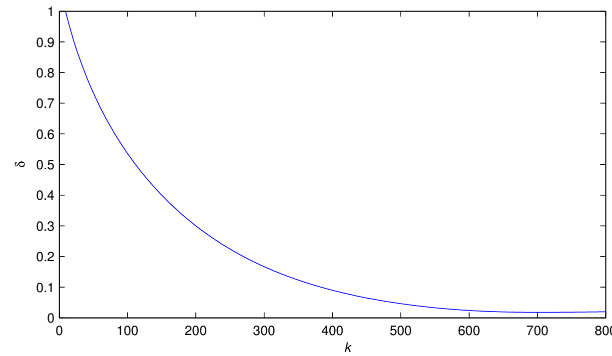


Fig. 5. The estimation errors δ versus k of Example 1.

$$\begin{aligned}
C(\zeta) &= 1 + c_1 z^{-1} = 1 + 0.65z^{-1},\\
D(\zeta) &= d_0 + d_1 z^{-1} = 1 + 0.023z^{-1},\\
\bar{x}(\tau) &= \gamma_1 u(\tau) + \gamma_2 u^2(\tau) + \gamma_3 u^3(\tau)\\
&= 0.80u(\tau) - 0.50u^2(\tau) + 0.95u^3(\tau),\\
\theta &= [c_1, d_1, \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3]^{\mathrm{T}},\\
\beta_1 &= -\gamma_1 r + \gamma_2 r^2 - \gamma_3 r^3,\\
\beta_2 &= \gamma_1 r + \gamma_2 r^2 + \gamma_3 r^3,\\
\varphi(\tau) &= [y(\tau-1), x(\tau-1), \xi(r+u(\tau)), \xi(r-u(\tau)),\\
&\quad u(\tau)\xi(-r-u(\tau))\xi(-r+u(\tau)),\\
&\quad u(\tau)^2\xi(-r-u(\tau))\xi(u(\tau)-r),\\
&\quad u(\tau)^3\xi(-r-u(\tau))\xi(-r+u(\tau))]^{\mathrm{T}}.
\end{aligned}$$

The input $\{u(\tau)\}$ is taken as a persistent excitation signal sequence and satisfies $N(0,1)$, and $\{v(\tau)\}$ is taken as a white noise sequence and satisfies $N(0, 0.1^2)$. Table 1 lists the numerical results of the Hammerstein system with saturation nonlinearities. The parameter estimation errors $\delta := \|\hat{\theta} - \theta\|/\|\theta\|$ are shown in Fig. 5.

Example 2: Consider the Hammerstein system proposed in Example 1 with colored noise:

$$\begin{aligned}
y(\tau) &= \frac{D(\zeta)}{C(\zeta)}\bar{x}(\tau) + \frac{G(\zeta)}{C(\zeta)}v(\tau),\\
G(\zeta) &= 1 + g_1 z^{-1} = 1 + 0.70z^{-1},\\
w(\tau) &= 0.70v(\tau-1) + v(\tau),\\
\vartheta_w^* &= [c_1, d_1, \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3, g_1]^{\mathrm{T}},\\
\psi_w^*(\tau) &= [y(\tau-1), x(\tau-1), \xi(u(\tau)+r), \xi(r-u(\tau)),\\
&\quad u(\tau)\xi(-r-u(\tau))\xi(-r+u(\tau)),\\
&\quad u(\tau)^2\xi(-u(\tau)-r)\xi(u(\tau)-r),\\
&\quad u(\tau)^3\xi(-r-u(\tau))\xi(-r+u(\tau)), v(\tau-1)]^{\mathrm{T}}.
\end{aligned}$$

The input $\{u(\tau)\}$ and $\{v(\tau)\}$ are the same as those in Example 1. Applying the traditional LS algorithm and the ML-LSAV algorithm to identify the parameters of the Hammerstein saturation system with colored noise, the parameter estimates and their errors are shown in Table 2.

Table 1. The Hammerstein system with saturation nonlinearities based ML-GI algorithm estimates and errors.

k	c_1	d_1	β_1	β_2	γ_1	γ_2	γ_3	$\delta(\%)$
10	0.53710	-0.64422	-0.40545	0.25225	0.89985	0.91658	0.99029	99.26874
90	0.63087	-0.10716	-0.46148	0.29853	0.77206	0.39264	0.97708	56.91015
180	0.64987	0.01579	-0.46019	0.29995	0.77628	0.03292	0.97632	33.65879
270	0.65303	0.03384	-0.45998	0.30018	0.77888	-0.18557	0.97551	19.93603
360	0.65316	0.03266	-0.46000	0.30016	0.78036	-0.31959	0.97471	11.55901
450	0.65283	0.02914	-0.46005	0.30012	0.78127	-0.40213	0.97394	6.47230
540	0.65253	0.02630	-0.46009	0.30008	0.78185	-0.45304	0.97320	3.50328
630	0.65231	0.02437	-0.46012	0.30005	0.78223	-0.48448	0.97248	2.06142
720	0.65217	0.02313	-0.46014	0.30003	0.78250	-0.50389	0.97177	1.78378
True values	0.65000	0.02300	-0.46080	0.30080	0.80000	-0.50000	0.95000	

Table 2. The Hammerstein saturation system with colored noise based LS and ML-LSAV algorithm estimates and errors.

	k	c_1	d_1	β_1	β_2	γ_1	γ_2	γ_3	$\delta(\%)$
LS	2	0.70588	-0.08667	-0.46238	0.29878	0.78882	-0.51251	0.97608	8.00021
	4	0.70670	-0.08602	-0.46243	0.29873	0.78885	-0.51301	0.97587	7.98827
	20	0.70672	-0.08600	-0.46243	0.29873	0.78885	-0.51303	0.97587	7.98797
ML-LSAV	2	0.64957	0.02300	-0.46237	0.29878	0.78881	-0.51248	0.97601	1.95645
	4	0.65033	0.02282	-0.46243	0.29873	0.78884	-0.51299	0.97580	1.95850
	20	0.65033	0.02282	-0.46243	0.29873	0.78884	-0.51300	0.97580	1.95853
True values		0.65000	0.02300	-0.46080	0.30080	0.80000	-0.50000	0.95000	

Table 3. The Hammerstein saturation system with colored noise based GI algorithm estimates and errors.

k	c_1	d_1	β_1	β_2	γ_1	γ_2	γ_3	g_1	$\delta(\%)$
10	0.64022	-0.63714	-0.05252	0.28206	0.92393	0.94303	0.99270	0.99897	96.33431
90	0.61356	-0.52717	-0.46057	0.29830	0.78405	0.54962	0.97965	0.98906	70.38676
180	0.62094	-0.40320	-0.46010	0.29876	0.78252	0.23208	0.98030	0.97854	51.47935
270	0.62654	-0.30827	-0.45969	0.29916	0.78530	0.00833	0.98156	0.96840	38.32828
360	0.63084	-0.23523	-0.45935	0.29950	0.78737	-0.14921	0.98288	0.95854	29.29608
450	0.63413	-0.17885	-0.45907	0.29978	0.78883	-0.26003	0.98423	0.94892	23.19423
540	0.63668	-0.13520	-0.45885	0.30000	0.78984	-0.33791	0.98560	0.93947	19.13919
630	0.63865	-0.10132	-0.45867	0.30018	0.79054	-0.39259	0.98698	0.93018	16.46585
720	0.64018	-0.07495	-0.45853	0.30031	0.79100	-0.43094	0.98837	0.92101	14.68384
True values	0.65000	0.02300	-0.46080	0.30080	0.80000	-0.50000	0.95000	0.70000	

Applying the GI algorithm and the ML-BCGI algorithm to identify the parameters of this system, the parameter estimates and their errors are shown in Tables 3 and 4, and the parameter estimation errors $\delta := \|\hat{\vartheta}_w^* - \vartheta_w^*\|/\|\vartheta_w^*\|$ are shown in Figs. 6 and 7.

From Tables 2-4 and Figs. 6-7, we can get the following conclusions:

1) The ML-LSAV algorithm can get the unbiased system parameters $c_1, c_2, \cdots, c_{n_c}$ and $d_1, d_2, \cdots, d_{n_d}$, but it cannot estimate the parameters of the colored noise $g_1, g_2, \cdots, g_{n_g}$.
2) The ML-BCGI algorithm can get the unbiased estimate of the Hammerstein system with colored noise.
3) Although the ML-LSAV algorithm has heavy computational burden, it can get the estimate quickly.
4) Comparing with the traditional LS and GI algorithms, the proposed algorithms are more effective.

6. CONCLUSIONS

In this paper, some identification algorithms are proposed for Hammerstein saturation systems with white noise and colored noise. The ML-GI algorithm is introduced for the Hammerstein systems with white noise. For

Table 4. The Hammerstein saturation system with colored noise based ML-BCGI algorithm estimates and errors.

k	c_1	d_1	β_1	β_2	γ_1	γ_2	γ_3	g_1	$\delta(\%)$
10	0.62766	-0.14471	-0.45009	0.30053	0.91672	0.93370	0.99196	0.99590	85.26650
90	0.63630	-0.06327	-0.45719	0.30089	0.78474	0.48721	0.97865	0.96013	59.10552
180	0.64156	-0.01329	-0.45692	0.30118	0.78551	0.14309	0.97833	0.92158	39.31867
270	0.64430	0.01172	-0.45677	0.30134	0.78813	-0.08797	0.97832	0.88436	26.09624
360	0.64561	0.02279	-0.45669	0.30141	0.78990	-0.24327	0.97830	0.84837	17.18695
450	0.64614	0.02645	-0.45667	0.30143	0.79110	-0.34776	0.97829	0.81356	11.08946
540	0.64627	0.02641	-0.45667	0.30143	0.79192	-0.41816	0.97826	0.77987	6.81799
630	0.64621	0.02477	-0.45668	0.30141	0.79248	-0.46565	0.97822	0.74728	3.78162
720	0.64606	0.02259	-0.45670	0.30138	0.79288	-0.49774	0.97816	0.71573	1.93714
True values	0.65000	0.02300	-0.46080	0.30080	0.80000	-0.50000	0.95000	0.70000	

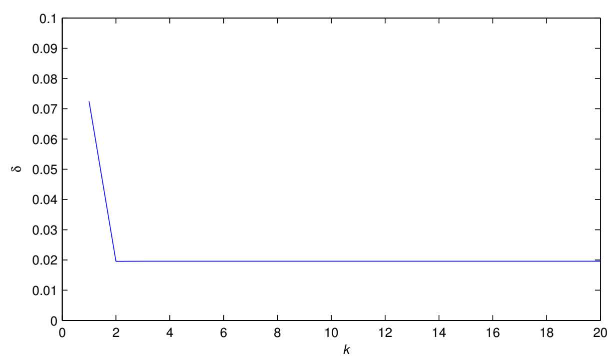


Fig. 6. The ML-LSAV estimation errors δ versus k.

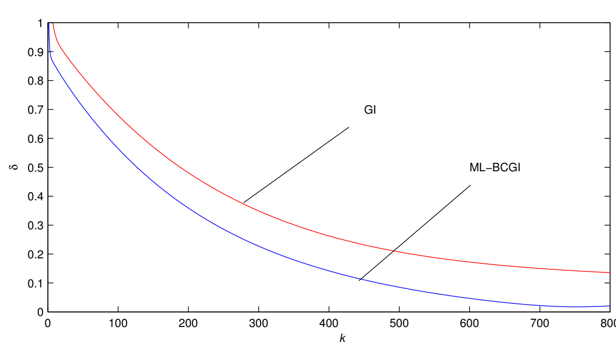


Fig. 7. The GI and ML-BCGI estimation errors δ versus k.

the systems with colored noise, the ML-LSAV algorithm is proposed, by which the parameters (not including the parameters of the colored noise) of the nonlinear systems can be obtained. To overcome this weakness, the ML-BCGI algorithm is developed. It can identify not only the parameters of the nonlinear parts but also the parameters of the noise model. For further research, The convergence analysis of these algorithms is also an open and challenging topic, which deserves further study. Some coupled parameter identification methods [31, 32] can be extended to Hammerstein nonlinear systems and can be applied to other literatures [33–37] such as engineering systems.

REFERENCES

[1] F. Ding, X. P. Liu, and G. Liu, "Identification methods for Hammerstein nonlinear systems," *Digital Signal Processing*, vol. 21, no. 2, pp. 215-238, March 2011.

[2] G. Y. Chen, M. Gan, and G. L. Chen, "Generalized exponential autoregressive models for nonlinear time series: Stationarity, estimation and applications," *Information Sciences*, vol. 438, pp. 46-57, April 2018.

[3] X. Zhang, F. Ding, and E. F. Yang, "State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1157-1173, July 2019.

[4] X. Zhang and F. Ding, "Hierarchical parameter and state estimation for bilinear systems," *International Journal of Systems Science*, vol. 51, no. 2, 275-290, 2020.

[5] E. W. Bai and K. Li, "Convergence of the iterative algorithm for a general Hammerstein system identification," *Automatica,*, vol. 46, no. 11, pp. 1891-1896, November 2010.

[6] T. Wigren, "Recursive prediction error identification using the nonlinear Wiener Model," *Automatica*, vol. 29, no. 4, pp. 1011-1029, April 1993.

[7] J. Voros, "Parameter identification of discontinuous Hammerstein systems," *Automatica*, vol. 33, no. 6, pp. 1141-146, June 1997.

[8] E. W. Bai, "Identification of linear systems with hard input nonlinearities of known structure," *Automatica*, vol. 38, no. 5, pp. 853-860, May 2002.

[9] B. M. Chen, T. H. Lee, K. M. Peng, and V. Venkataramanan, "Composite nonlinear feedback control for linear systems with input saturation: Theory and application," *IEEE Transactions on Automatic Control*, vol. 48, no. 3, pp. 427-439, April 2003.

[10] C. Wen, J. Zhou, Z. Liu, and H. Su, "Robust adaptive control of uncertain nonlinear systems in the presence of input saturation and external disturbance," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1672-1678, August 2011.

[11] J. Chen, X. P. Wang, and R. F. Ding, "Gradient based estimation algorithm for Hammerstein systems with saturation and dead-zone nonlinearities," *Applied Mathematical Modelling*, vol. 36, no. 1, pp. 238-243, January 2012.

[12] F. Ding, "Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling," *Applied Mathematical Modelling*, vol. 37, no. 4, pp. 1694-1704,. April 2013.

[13] M. H. Li, X. M. Liu, and F. Ding, "The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1189-1211, July 2019.

[14] J. Pan, X. Jiang, X. K. Wan, and W. Ding, "A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems," *International Journal of Control Automation and Systems*, vol. 15, no. 3, pp. 1189-1197, June 2017.

[15] A. Wills, B. Ninness, and S. Gibson, "Maximum likelihood estimation of state space models from frequency domain data," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 19-33, January 2009.

[16] H. F. Xia, Y. Ji, Y. J. Liu, and L. Xu, "Maximum likelihood-based multi-innovation stochastic gradient method for multivariable systems," *International Journal of Control Automation and Systems*, vol. 17, no. 3, pp. 565-574, March 2019.

[17] T. B. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39-49, January 2011.

[18] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind maximum likelihood identification of Hammerstein systems," *Automatica*, vol. 44, no. 12, pp. 3139-3146, December 2008.

[19] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind maximum-likelihood identification of Wiener systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3017-3029, August 2009.

[20] J. Ding, F. Ding, X. P. Liu, and G. Liu, "Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data," *IEEE Transactions on Automatic Control*, vol. 56, no. 11, pp. 2677-2683, November 2011.

[21] F. Ding, G. Liu, and X. P. Liu, "Parameter estimation with scarce measurements," *Automatica*, vol. 47, no. 8, pp. 1646-1655, August 2011.

[22] Y. J. Liu, F. Ding, and Y. Shi, "An efficient hierarchical identification method for general dual-rate sampled-data systems," *Automatica*, vol. 50, no. 3, pp. 962-970, March 2014.

[23] J. Ding, J. Z. Chen, J. X. Lin, and G. P. Jiang, "Particle filtering-based recursive identification for controlled autoregressive systems with quantised output," *IET Control Theory and Applications*, vol. 13, no. 14, pp. 2181-2187, September 2019.

[24] J. Ding, Z. X. Cao, J. Z. Chen, and G. P. Jiang, "Weighted parameter estimation for Hammerstein nonlinear ARX systems," *Circuits Systems and Signal Processing*, vol. 39, pp. 2178-2192, 2020.

[25] L. J. Wan and F. Ding, "Decomposition- and gradient-based iterative identification algorithms for multivariable systems using the multi-innovation theory," *Circuits Systems and Signal Processing*, vol. 38, no. 7, pp. 2971-2991, July 2019.

[26] J. Pan, W. Li, and H. P. Zhang, "Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control," *International Journal of Control Automation and Systems*, vol. 16, no. 6, pp. 2878-2887, December 2018.

[27] X. K. Wan, Y. Li, C. Xia, M. H. Wu, J. Liang, and N. Wang, "A T-wave alternans assessment method based on least squares curve fitting technique," *Measurement*, vol. 86, pp. 93-100, May 2016.

[28] H. Ma, J. Pan, F. Ding, L. Xu, and W. Ding, "Partially-coupled least squares based iterative parameter estimation for multi-variable output-error-like autoregressive moving average systems," *IET Control Theory and Applications*, vol. 13, no. 18, pp. 3040-3051, December 2019.

[29] H. Liu, Q. X. Zou, and Z. P. Zhang, "Energy disaggregation of appliances consumptions using ham approach," *IEEE Access*, vol. 7, pp. 185977-185990, 2019.

[30] L. Wang, H. Liu, L. V. Dai, and Y. W. Liu, "Novel method for identifying fault location of mixed lines," *Energies*, vol. 11, no. 6, Article Number: 1529, June 2018.

[31] F. Ding, "Coupled-least-squares identification for multivariable systems," *IET Control Theory and Applications*, vol. 7, no. 1, pp. 68-79, January 2013.

[32] F. Ding, G. Liu, and X. P. Liu, "Partially coupled stochastic gradient identification methods for non-uniformly sampled systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1976-1981, August 2010.

[33] J. G. Zhou, H. L. Xiao, W. W. Jiang, W. F. Bai, and G. L. Liu, "Automatic subway tunnel displacement monitoring using robotic total station," *Measurement*, vol. 151, p. 107251, February 2020.

[34] F. Ding, L. Xu, D. D. Meng, X. B. Jin, A. Alsaedi, and T. Hayat, "Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model," *Journal of Computational and Applied Mathematics*, vol. 369, p. 112575, May 2020.

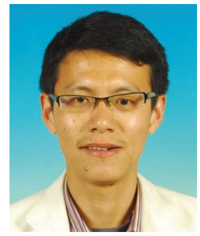
[35] M. Y. Liu, H. F. Ku, J. X. Zhang, P. D. Xu, and C. H. Wu, "Predicting fatigue life for finite line contact under starved elastohydrodynamic lubrication condition," *Mathematical Problems in Engineering*, vol. 2020, p. 5928621, January 2020.

[36] F. Ding, L. Lv, J. Pan, X. K. Wan, and X. B. Jin, "Two-stage gradient-based iterative estimation methods for controlled autoregressive systems using the measurement data," *International Journal of Control Automation and Systems*, 2020. DOI: 10.1007/s12555-019-0140-3

[37] M. Y. Liu, P. D. Xu, J. X. Zhang, and H. F. Ding, "Analysis of mechanical power loss of a helical gear pair based on the starved thermal-elastohydrodynamic lubrication model," *Industrial Lubrication and Tribology*, 2020. DOI: 10.1108/ILT-06-2019-0225

Yan Pu received her B.Sc. degree from the School of Mathematical Science, Soochow University (Suzhou, China) in 2000, and her M.Sc. degree from the Department of Mathematics, Nanjing University (Nanjing, China) in 2004. She is currently a Ph.D. student at the School of Internet of Things Engineering, Jiangnan University (Wuxi, China). Her research interests include parameter estimation, processing control and system identification.

Yongqing Yang received his B.Sc. degree from Anhui Normal University (Wuhu, China), an M.Sc. degree from Anhui University of Science and Technology (Huainan, China) and a Ph.D. degree from Southeast University (Nanjing, China), in 1985, 1992, and 2007, respectively. He is currently a Professor of Jiangnan University. He is the author or coauthor of more than 50 journal papers. His research interests include nonlinear systems, neural networks and optimization.



Jing Chen received his B.Sc. and M.Sc. degrees from the College of Mathematical Sciences, Yangzhou University (Yangzhou, China), in 2003 and 2006, respectively. He received his Ph.D. degree from the School of Internet of Things Engineering, Jiangnan University (Wuxi, China) in 2013. He is an Assistant Professor at the School of Science, Jiangnan University. His research interests include processing control and system identification.