

Research Article

Nonsmooth Recursive Identification of Sandwich Systems with Backlash-Like Hysteresis

Ruili Dong,¹ Yonghong Tan,¹ Hui Chen,² and Yangqiu Xie³

¹ College of Mechanical and Electronic Engineering, Shanghai Normal University, Shanghai 200234, China

² Department of Automation, Southeast University, Nanjing 210096, China

³ School of Electronic Engineering, Xi'dian University, Xi'an 710071, China

Correspondence should be addressed to Yonghong Tan, tany@shnu.edu.cn

Received 29 March 2012; Revised 10 June 2012; Accepted 10 June 2012

Academic Editor: Zhiwei Gao

Copyright © 2012 Ruili Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A recursive gradient identification algorithm based on the bundle method for sandwich systems with backlash-like hysteresis is presented in this paper. In this method, a dynamic parameter estimation scheme based on a subgradient is developed to handle the nonsmooth problem caused by the backlash embedded in the system. The search direction of the algorithm is estimated based on the so-called bundle method. Then, the convergence of the algorithm is discussed. After that, simulation results on a nonsmooth sandwich system are presented to validate the proposed estimation algorithm. Finally, the application of the proposed method to an X-Y moving positioning stage is illustrated.

1. Introduction

Usually, a sandwich system with backlash-like hysteresis is defined as the system that a backlash-like hysteresis is sandwiched between two linear dynamic subsystems. In engineering applications, many mechanical systems such as mechanical transmission systems, servo control systems, and hydraulic valve systems can be described by the so-called sandwich systems with backlash-like hysteresis. The reason to cause the backlash-like hysteresis phenomenon is mainly due to the gaps existing in transmission mechanism systems such as gearbox and ball screw.

Recently, identification of sandwich systems has become one of the interesting issues in the domain of modeling and control for complex systems. References [1–3] proposed the recursive identification methods for the sandwich system with smooth nonlinearities. The main ideas of those approaches are to extend the linear system identification methods to smooth nonlinear cases. Moreover, there have been some methods for the identification of

Hammerstein or Wiener systems with backlash-like hysteresis [4–8], most of which are the modified linear system identification methods.

However, until today, there have been very few publications concerning the identification of the sandwich systems with backlash-like hysteresis. Reference [9] proposed a method to identify the sandwich systems with backlash-like hysteresis, but the approach is still based on idea to extend the linear system identification method to nonlinear cases. On the other hand, the switching functions in that method have significant influence on the convergence speed of the algorithm.

In this paper, a recursive gradient algorithm based on the bundle method is proposed to identify parameters of the sandwich model. In this algorithm, the effect of the nonsmoothness caused by the backlash-like hysteresis in sandwich system is considered. In order to obtain the optimizing search direction at the nonsmooth points of the system, the Clarke subgradient technique is utilized based on the idea of the bundle method [10–12]. By comparing with the above-mentioned available methods, the proposed method employs the nonsmooth optimization technique to identify the nonsmooth sandwich systems with backlash-like hysteresis. Thus, it will provide us with a new approach for dealing with on-line modeling of nonsmooth dynamic systems. A numerical example will be presented to evaluate the performance of the proposed approach. Finally, experimental results on an X - Y moving positioning stage are illustrated.

2. Brief Description of Sandwich Systems with Backlash

The structure of a sandwich system with backlash-like hysteresis is shown in Figure 1, in which a backlash-like hysteresis is embedded between the input and output linear subsystems, that is, $L_1(\cdot)$ and $L_2(\cdot)$. It is assumed that input $u(k)$ and output $y(k)$ can be measured directly, but the internal variables $x(k)$ and $v(k)$ are not measurable.

Suppose that both linear subsystems are stable, and the time delays q_1 and q_2 in $L_1(\cdot)$ and $L_2(\cdot)$ are known, respectively. The corresponding discrete-time models of $L_1(\cdot)$ and $L_2(\cdot)$ are, respectively, written as

$$\begin{aligned} x(k) &= -\sum_{i_2=1}^{n_a} a_{i_2} x(k-i_2) + \sum_{j_2=0}^{n_b} b_{j_2} u(k-q_1-j_2), \\ y(k) &= -\sum_{i_1=1}^{n_c} c_{i_1} y(k-i_1) + \sum_{j_1=0}^{n_d} d_{j_1} v(k-q_2-j_1), \end{aligned} \quad (2.1)$$

where n_a and n_b are the orders of $L_1(\cdot)$, q_1 is the time delay, and a_{i_2} as well as b_{j_2} are the coefficients of $L_1(\cdot)$; n_c and n_d are the orders of $L_2(\cdot)$, q_2 is the time delay, and c_{i_1} and d_{j_1} are the coefficients of $L_2(\cdot)$. Let both b_0 and d_0 be equal to unity for unique representation.

Note that the backlash-like hysteresis shown in Figure 1 is specified by the slopes m_1 and m_2 as well as the absolute thresholds, D_1 and D_2 , where $0 < m_1 < \infty$, $0 < m_2 < \infty$,

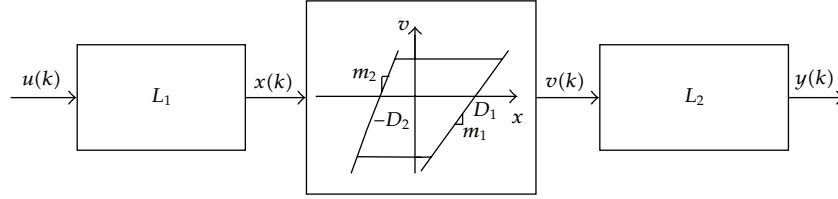


Figure 1: The structure of the sandwich system with backlash-like hysteresis.

$0 < D_1 < \infty$, and $0 < D_2 < \infty$. Hence, the discrete-time model of the backlash-like hysteresis is described as

$$v(k) = \begin{cases} m_1(x(k) - D_1), & x(k) > \frac{v(k-1)}{m_1} + D_1, \quad x(k) > x(k-1), \text{ increase zone,} \\ v(k-1), & \frac{v(k-1)}{m_2} - D_2 \leq x(k) \leq \frac{v(k-1)}{m_1} + D_1, \text{ memory zone,} \\ m_2(x(k) + D_2), & x(k) < \frac{v(k-1)}{m_2} - D_2, \quad x(k) < x(k-1), \text{ decrease zone.} \end{cases} \quad (2.2)$$

For the convenience to describe the system, the discrete-time model of the backlash-like hysteresis can be rewritten as

$$\begin{aligned} m(k) &= m_1 + (m_2 - m_1)g(k), \\ v1(k) &= m(k)(x(k) + g(k)x(k) - D_1g_1(k) + D_2g_2(k)), \\ v(k) &= v1(k) + [v(k-1) - v1(k)](g_1(k) - 1)(g_2(k) - 1), \end{aligned} \quad (2.3)$$

where the switching functions $g(k)$, $g_1(k)$, and $g_2(k)$ are, respectively, defined as

$$\begin{aligned} g(k) &= \begin{cases} 0, & \Delta x(k) > 0 \\ 1, & \Delta x(k) \leq 0, \end{cases} \\ g_1(k) &= \begin{cases} 1, & x(k) > \frac{v(k-1)}{m_1} + D_1, \quad x(k) > x(k-1), \\ 0, & \text{else,} \end{cases} \\ g_2(k) &= \begin{cases} 1, & x(k) < \frac{v(k-1)}{m_2} - D_2, \quad x(k) < x(k-1), \\ 0, & \text{else,} \end{cases} \end{aligned} \quad (2.4)$$

where $\Delta x(k) = x(k) - x(k-1)$.

Thus, (2.1)–(2.3) present the model to describe the sandwich system with backlash-like hysteresis. Hence, the unknown parameter vector of the model can be written as $\theta \in R^{n_a+n_b+n_c+n_d+4}$, where

$$\theta = [c_1, \dots, c_{n_c}, a_1, \dots, a_{n_a}, m_1, m_2, D_1, D_2, b_1, \dots, b_{n_b}, d_1, \dots, d_{n_d}]^T. \quad (2.5)$$

According to concept of the gradient algorithm, define the objective function as

$$Q(k, \hat{\theta}(k)) = \sum_{k=1}^n \frac{[y(k) - \hat{y}(k, \hat{\theta}(k))]^2}{2} = \frac{1}{2} \sum_{k=1}^n f(k, \hat{\theta}(k)), \quad (2.6)$$

where $\hat{\theta}$ is the estimate of θ , and $\hat{y}(k, \hat{\theta}(k))$ is the output of system model. The optimal estimate of $\hat{\theta}$ can be obtained by minimizing the above-mentioned criterion.

3. The Nonsmooth Estimation of the Sandwich Model with Backlash-Like Hysteresis

In this section, a gradient-based identification algorithm is proposed for identification of the sandwich system with backlash-like hysteresis. Due to the nonsmoothness of the backlash, the gradients of the system output with respect to the parameters of the backlash at nonsmooth points will not exist. The smooth gradient-based methods directly applied to nonsmooth systems may fail in convergence [13]. On the other hand, the genetic algorithms [14] or Powell's method [15], which are based on derivative-free techniques, may be unreliable and become inefficient when the system structure is complicated. Thus, we should find a special way for solving this problem. The simplest way to solve the problem is to apply the Clarke subgradients [11] to the approximation of the gradients at the nonsmooth points.

The basic idea of the bundle method is to approximate the subdifferential of $Q(k, \hat{\theta}(k))$ with respect to $\hat{\theta}(k)$ by gathering the subgradients from previous iterations into a bundle for the nonsmooth objective function $Q(k, \hat{\theta}(k))$. The gradient $\nabla Q(k, \hat{\theta}(k))$ can change discontinuously, and some change of the gradient may not be small in the neighborhood of the minimum of the function. So the values of $Q(k, \hat{\theta}(k))$ and $\partial Q(k, \hat{\theta}(k))$ at a single point $\hat{\theta}(k)$ do not offer sufficient information of the local behavior of $Q(k, \hat{\theta}(k))$. The detail of the bundle method can be found in [10–12] and reference therein.

Considering that the sandwich system with backlash-like hysteresis is locally Lipschitz continuous, we have the following definition.

Definition 3.1 (see [11]). Let $F: R^n \times R \rightarrow R$ be locally Lipschitz continuous. This allows one to define a Clarke subgradient of F at ξ as $dF(\xi)$:

$$dF(\xi) \in \partial F(\xi), \quad \text{subject to } \partial F(\xi) = \text{conv} \left\{ \nabla F(\xi^i) \mid \xi^i \rightarrow \xi, \nabla F(\xi^i) \text{ exists} \right\}, \quad (3.1)$$

where “conv” denotes the convex hull of a set.

The set of all the Clarke subgradients is the Clarke subdifferential of F at ξ which is denoted by $\partial F(\xi)$ [11].

Considering that backlash-like hysteresis is a nonsmooth mapping, the gradients of parameters in $L_1(\cdot)$ with respect to $v(k)$ do not exist at a nonsmooth point. Hence, we define the parameters of the backlash-like hysteresis and $L_1(\cdot)$ as $\sigma = \{m_1, m_2, D_1, D_2, a_1 \cdots a_{na}, b_1 \cdots b_{nb}\} \in R^{4+na+nb}$. Considering the cost function described by (2.6), the gradients of $f(\cdot)$ with respect to σ will not exist at the nonsmooth points. Hence,

at the nonsmooth points of $Q(\cdot)$, the Clarke subdifferential of $f(\cdot)$ with respect to σ , that is, $\partial f(\sigma)$, can be obtained by

$$\partial f(\sigma) = -\text{conv} \left\{ \left[y(k) - \hat{y}(k, \hat{\theta}(k)) \right] \left(\sum_{j_1=0}^{n_b} d_{j_1} \partial \hat{v}_{(k-j_1-q_2)}(\sigma) \right) \right\}, \quad (3.2)$$

where $\partial \hat{v}_{(k-j_1-q_2)}(\sigma) = \text{conv} \{ \nabla \hat{v}_{(k-j_1-q_2)}(\sigma) \}$, and $\nabla \hat{v}_{(k-j_1-q_2)}(\sigma)$ is the gradient of \hat{v} , the output of backlash-like hysteresis, with respect to σ at the smooth points. Thus, the corresponding gradients of $\hat{v}(k - j_1 - q_2)$ with respect to σ at the smooth points are

$$\nabla \hat{v}_{(k-j_1-q_2)}(\sigma) = \begin{cases} \left[\begin{array}{l} u(k - j_1 - q_2 - q_1) - \hat{D}_1(k-1), 0, \\ -\hat{m}_1(k-1), 0, -\hat{m}_1(k-1)\hat{x}(k-1-j_1-q_2), \dots, \\ -\hat{m}_1(k-1)\hat{x}(k-n_a-j_1-q_2), \\ \hat{m}_1(k-1)u(k-1-j_1-q_2-q_1), \dots, \\ \hat{m}_1(k-1)u(k-n_b-j_1-q_2-q_1) \end{array} \right]^T, & \text{in increase zones,} \\ [0, 0, 0, 0, \dots, 0, 0, \dots, 0]^T, & \text{in memory zones,} \\ \left[\begin{array}{l} 0, u(k - j_1 - q_2 - q_1) + \hat{D}_2(k-1), 0, \\ \hat{m}_2(k-1), -\hat{m}_2(k-1)\hat{x}(k-1-j_1-q_2), \dots, \\ -\hat{m}_2(k-1)\hat{x}(k-n_a-j_1-q_2), \\ \hat{m}_2(k-1)u(k-1-j_1-q_2-q_1), \dots, \hat{m}_2(k-1) \\ u(k-n_b-j_1-q_2-q_1) \end{array} \right]^T, & \text{in decrease zones,} \end{cases} \quad (3.3)$$

where $\hat{x}(k) = -\sum_{i_2=1}^{n_a} \hat{a}_{i_2} \hat{x}(k-i_2) + \sum_{j_2=0}^{n_b} \hat{b}_{j_2} u(k-q_1-j_2)$, and the coefficients \hat{a}_{i_2} and \hat{b}_{j_2} are the corresponding estimated values at the previous step.

Hence, based on (3.2) and (3.3), the Clarke subdifferential of $f(\cdot)$ with respect to σ can be obtained at nonsmooth points of the system. Besides, as $L_2(\cdot)$ is a smooth function, the gradients of $f(\cdot)$ with respect to the parameters of the linear subsystems $L_2(\cdot)$ always exist. So, the Clarke subdifferential of $f(\cdot)$ with respect to all the unknown parameters of the sandwich system can be determined.

The proper Clarke subgradient direction $\mathbf{t}(k, \hat{\theta}(k))$ of $f(\cdot)$ with respect to the parameters to be estimated at nonsmooth points can be derived based on

$$\begin{aligned} \min_{\varphi, \mathbf{d}} & \left(\varphi(k) + \frac{1}{2} \|\mathbf{t}(k, \hat{\theta}(k))\|^2 \right) \\ \text{s.t.} & -\beta_j(k) + \langle \mathbf{h}_j(k), \mathbf{t}(k, \hat{\theta}(k)) \rangle \leq \varphi(k), \quad \forall j \in J_k, \end{aligned} \quad (3.4)$$

where $\|\cdot\|$ denotes the Euclidean norm; J_k is a nonempty subset of $\{1, \dots, k\}$; set $\varphi(k)$ is the predicted amount of descent; $\mathbf{h}_j(k) \in \partial f(k, \hat{\theta}_j^*(k))$ for $j \in J_k$, and $\hat{\theta}_j^*(k)$ are some trail points (from the past iterations); $\beta_j(k) = \max\{|\alpha_j(k)|, \gamma(s_j(k))^2\}$ is the locality measure of subgradient; $\gamma \geq 0$ is the distance measure parameter ($\gamma = 0$ if $f(k, \hat{\theta}(k))$ is convex),

$\alpha_j(k) = f(k, \hat{\boldsymbol{\theta}}(k)) - f(k, \hat{\boldsymbol{\theta}}_j^*(k)) - \mathbf{h}_j(k)(\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}_j^*(k))$ is the linearization error; $s_j(k) = \|\hat{\boldsymbol{\theta}}_j(k) - \hat{\boldsymbol{\theta}}_j^*(k)\| + \sum_{i=j}^{k-1} \|\hat{\boldsymbol{\theta}}_{i+1}(k) - \hat{\boldsymbol{\theta}}_i(k)\|$ is the distance measure to estimate $\|\hat{\boldsymbol{\theta}}(k) - \hat{\boldsymbol{\theta}}_j^*(k)\|$ without the requirement to store the trial point $\hat{\boldsymbol{\theta}}^*(k)$.

According to formula (3.4), $\mathbf{t}(k)$ and $\varphi(k)$ are obtained, that is,

$$\mathbf{t}(k, \hat{\boldsymbol{\theta}}(k)) = -\sum_{j \in J_k} \lambda_j^k \mathbf{h}_j(k) = \sum_{j \in J_k} \left[y(k) - \hat{y}(k, \hat{\boldsymbol{\theta}}_j(k)) \right] \lambda_j^k \mathbf{w}_j(k) = e(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}_j(k), \quad (3.5)$$

$$\varphi(k) = -\left\| \mathbf{t}(k, \hat{\boldsymbol{\theta}}(k)) \right\|^2 - \sum_{j \in J_k} \lambda_j^k \beta_j(k), \quad (3.6)$$

where $\mathbf{w}_j(k) = \partial \hat{y}(k, \hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta} |_{\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_j(k)}$, $e(k, \hat{\boldsymbol{\theta}}(k)) = y(k) - \hat{y}(k, \hat{\boldsymbol{\theta}}(k))$, $\bar{\mathbf{h}}_j(k) = \sum_{j \in J_k} \lambda_j^k \mathbf{w}_j(k)$, $\lambda_j^k \geq 0$, and $\sum_{j \in J_k} \lambda_j^k = 1$.

Remark 3.2. If $f(\cdot)$ is convex, the model $f(k, \hat{\boldsymbol{\theta}}(k))$ is an underestimate for $f(\cdot)$, and the nonnegative linearization error $\alpha_j(k)$ measures the performance of an approximation of the model to the original cost function. If $f(\cdot)$ is nonconvex, these facts are not valid anymore because $\alpha_j(k)$ may have a small or even negative value, although the trial point $\hat{\boldsymbol{\theta}}_j^*(k)$ locates far away from the current iteration point $\hat{\boldsymbol{\theta}}(k)$, and thus, the corresponding subgradient $\mathbf{h}_j(k)$ is worthless. For these reasons, the locality measure of subgradient $\beta_j(k)$ is introduced.

Therefore, the proposed recursive gradient estimation algorithm based on bundle method for the sandwich model with backlash-like hysteresis is shown as follows.

Step 1. Select starting point $\boldsymbol{\theta}_0 \in R^{n_a+n_b+n_c+n_d+4}$ and stopping parameter $\delta > 0$. Calculate $f(k, \boldsymbol{\theta}_0)$ and vector $\mathbf{h}_j(k) \in \partial f(k, \boldsymbol{\theta}_0)$, where $j \in J_k$, $J_k = \{k_0\}$, $|J_k| \leq k_1$, $|J_k|$ is the element number of J_k , and k_1 is a given positive number. Set $\beta_j(k) = 0$, $k = k_0$ and the line search parameters

$$q \in (0, 0.5), \quad q^* \in (q, 1), \quad \eta(0) \in (0, 1]. \quad (3.7)$$

Step 2. Calculate optimal solution $(\varphi(k), \mathbf{t}(k, \hat{\boldsymbol{\theta}}(k)))$ based on formulas (3.2)–(3.6). If $\varphi(k) \geq -\delta$, then stop.

Step 3. Search for the largest step size $\eta(k) \in [0, 1]$ such that $\eta(k) \geq \eta(0)$ and if

$$f(k, \hat{\boldsymbol{\theta}}(k) + \eta(k) \mathbf{t}(k)) \leq f(k, \hat{\boldsymbol{\theta}}(k)) + q \eta(k) \varphi(k), \quad (3.8)$$

it holds

$$\varphi(k) = f(k, \hat{\boldsymbol{\theta}}(k) + \mathbf{t}(k)) - f(k, \hat{\boldsymbol{\theta}}(k)) < 0. \quad (3.9)$$

Then, we take a long step and set $\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \eta(k) \mathbf{t}(k)$ and $\hat{\boldsymbol{\theta}}^*(k+1) = \hat{\boldsymbol{\theta}}(k+1)$; go to Step 4.

Otherwise, if $0 < \eta(k) < \eta(0)$, and formula (3.8) holds, then we take a short step and set $\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \eta(k) \mathbf{t}(k)$, and $\hat{\boldsymbol{\theta}}^*(k+1) = \hat{\boldsymbol{\theta}}(k) + \eta^*(k) \mathbf{t}(k)$ where $\eta^*(k) > \eta(k)$. Go to Step 5.

If $\eta(k) = 0$, and formula (3.8) holds, we take a null step, and namely set $\widehat{\boldsymbol{\theta}}(k+1) = \widehat{\boldsymbol{\theta}}(k)$ and $\widehat{\boldsymbol{\theta}}^*(k+1) = \widehat{\boldsymbol{\theta}}(k) + \eta^*(k)\mathbf{t}(k)$; go to Step 5.

Step 4. Let $J_k = J_k \cup \{k+1\}$, $k = k+1$; if $k \leq k_1$, then $J_k = \{1, \dots, k\}$, and if $k > k_1$, then $J_k = J_{k-1} \cup \{k\} \setminus \{k-k_1\}$, then go to Step 2.

Step 5. $J_k = J_k \cup \{k+1\}$, $k = k+1$; if $k \leq k_1$, then $J_k = \{1, \dots, k\}$, and if $k > k_1$, then $J_k = J_{k-1} \cup \{k\} \setminus \{k-k_1\}$, and the proper Clarke subgradient $\mathbf{h}_j(k)$ satisfies

$$-\beta_j \left(k, \widehat{\boldsymbol{\theta}}(k) \right) + \mathbf{h}_j^T \left(k, \widehat{\boldsymbol{\theta}}(k) \right) \mathbf{t} \left(k, \widehat{\boldsymbol{\theta}}(k-1) \right) \geq q^* \varphi \left(k, \widehat{\boldsymbol{\theta}}(k-1) \right), \quad (3.10)$$

then go to Step 2.

Remark 3.3. In long step, there is an obvious decrease in the value of the objective function. Hence, it is unnecessary to detect discontinuities in the gradient of $f(\cdot)$. Thus, we just set $\mathbf{h}_j(k) \in \partial f(k, \widehat{\boldsymbol{\theta}}(k))$. On the other hand, in short steps and null steps, the gradient of $f(\cdot)$ is discontinuous. Then, based on (3.10), both $\widehat{\boldsymbol{\theta}}(k)$ and $\widehat{\boldsymbol{\theta}}^*(k)$ located on the opposite sides of this discontinuity are guaranteed, and the new subgradient $\mathbf{h}_j(k) \in \partial f(k, \widehat{\boldsymbol{\theta}}^*(k))$ will force an obvious modification of the next search direction. Hence, the algorithm approximates the effectively searching direction at nonsmooth points based on the bundle method, which cannot be realized by the smooth optimization techniques.

Remark 3.4. If the value of $\eta(0)$ is too small, the convergence speed will be very sluggish, while $\eta(0)$ is too large, and the algorithm may not be convergent. Hence, it is important for $\eta(0)$ to be chosen properly. Usually, $\eta(0)$ is chosen based on an empirical method.

Remark 3.5. If all the Clarke subgradients are included in J_k , the corresponding storage capacity is infinite. Hence, the number of the subgradients in J_k must be constrained. In the proposed algorithm, we give the upper bound of $|J_k| \leq k_1$, and the upper bound k_1 is specified by empirical method.

4. Convergence of the Estimation

For the convergence of the above-mentioned estimation algorithm, we have the following

Theorem 4.1. *Suppose that $\eta(k)$ and $\beta_j(k)$ satisfy*

$$0 \leq \eta(k) \leq \frac{2e^2(k)\bar{\mathbf{h}}(k)\bar{\mathbf{h}}^T(k) - \beta_j(k)}{e^2(k)\bar{\mathbf{h}}^T(k)\bar{\mathbf{h}}(k) \left(1 + \left[\bar{\mathbf{h}}(k)\bar{\mathbf{h}}^T(k) \right] \right)}, \quad (4.1)$$

$$\beta_j(k) \leq 2e^2(k) \left[\bar{\mathbf{h}}(k)\bar{\mathbf{h}}^T(k) \right], \quad (4.2)$$

respectively, then the parameters $\boldsymbol{\theta}$ can be convergent to a local optimal value.

Proof. The proof of this theorem can be found in Appendix. \square

5. Simulation

The proposed approach is used to identify a numerical sandwich system with backlash-like hysteresis based on the measured system input and output. Suppose that the parameters of the backlash-like hysteresis in the system are $m_1 = 1$, $m_2 = 1.2$, $D_1 = 0.5$, and $D_2 = 0.6$. The linear subsystems $L_1(\cdot)$ and $L_2(\cdot)$ are

$$x(k) = -0.1x(k-1) - 0.2x(k-2) + 1.5u(k-1), \quad (5.1)$$

$$y(k) = -1.2y(k-1) - 0.32y(k-2) + 2v(k-1) - 0.1v(k-2), \quad (5.2)$$

respectively.

That implies $a_1 = 0.1$, $a_2 = 0.2$, $b_0 = 1.5$, $c_1 = 1.2$, $c_2 = 0.32$, $d_0 = 2$, and $d_1 = -0.1$. In the simulation, both b_0 and d_0 are assumed to be equal to unity for model uniqueness, which implies that the corresponding equivalent true values of the coefficients are $\tilde{a}_1 = 0.1$, $\tilde{a}_2 = 0.2$, $\tilde{c}_1 = 1.2$, $\tilde{c}_2 = 0.32$, $\tilde{d}_1 = -0.05$, $\tilde{m}_1 = 3$, $\tilde{m}_2 = 3.6$, $\tilde{D}_1 = 0.33$, and $\tilde{D}_2 = 0.4$, respectively, but this does not affect the properties of the whole system.

In the simulation, the signal to excite the system is a random sequence with variance $\sigma^2 = 0.49$. Choose $\delta = 1.0 \times 10^{-4}$. In the proposed algorithm, based on Remark 3.4, select $\eta(0) = 0.015$, $k_1 = 6$, $\theta_0 = [0, 0, 0, 0, 0.1, 0.1, 0.1, 0.1, 0]^T$, and $\beta_1(k_0) = 0$, respectively. For comparison, the traditional gradient method is also used to estimate the parameters of the system. In this method, the nonsmooth points of the system are omitted for the gradients of the system do not exist at nonsmooth points. The initialized values of the parameters are the same as those used in the proposed method. The optimizing step is chosen as 0.009.

Figure 2 illustrates the comparison of the estimated parameter convergence procedures between the proposed method and the traditional gradient method. In Figure 2, blue and solid lines denote the convergence procedures of the parameters estimated by the proposed method, while red and dotted lines show the convergence procedures of the parameters determined by the traditional gradient method. From Figure 2, we note that the parameters of the backlash-like hysteresis converge slower than those of the linear submodels especially the input linear submodel. Moreover, the proposed method has achieved faster convergence than that of the traditional gradient method. It is noticed that the oscillation and sharp jumps happened in the estimation procedure of the traditional gradient approach.

In the case that the system is affected by random noise, the proposed strategy can still obtain better convergence of parameter estimation. In the simulation with noise, the signal to noise rate (SNR) is equal to 46.5. All the initial values of the parameters are the same as those in the noise-free case.

Figure 3 shows the comparison of the convergence procedures of the estimated parameters in the case with noise between the proposed method and the traditional gradient approach. Similar to the noise-free case, the blue and solid lines denote the convergence procedures of the parameters estimated by the proposed method, while the red and dotted lines show the convergence procedures of the parameters estimated by the traditional gradient method. Obviously, the proposed method has obtained faster convergent results than the traditional gradient method.

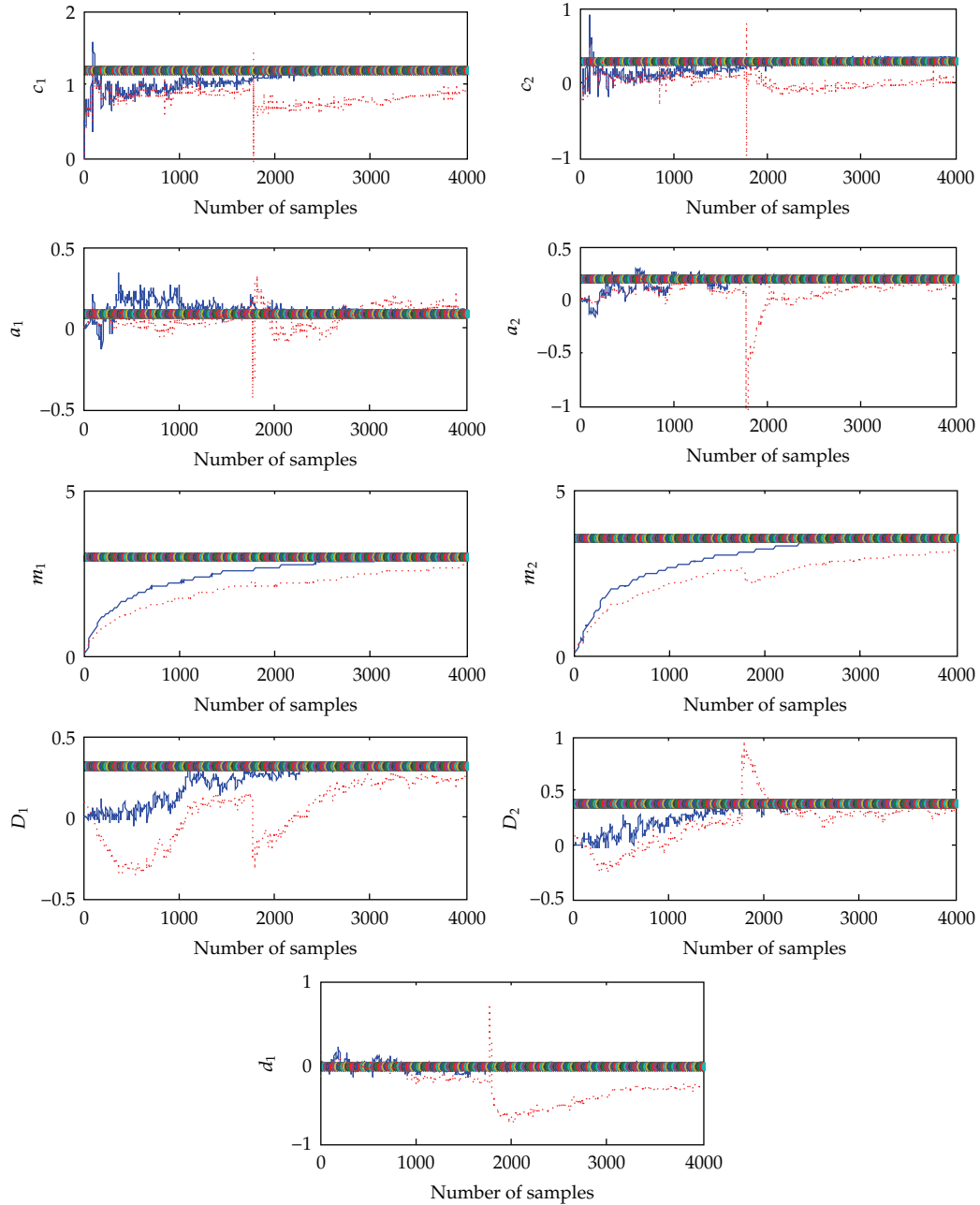


Figure 2: The convergence of the estimated parameters (noise-free case): proposed method: blue and solid line; traditional gradient method: red and dotted lines.

6. Application to an X-Y Moving Positioning Stage

The proposed identification approach is also applied to the modeling of an X-Y moving positioning stage with the architecture shown in Figure 4. In this equipment, the movement of the work platform of each axis is driven by a DC servomotor through a ball-screw-nut

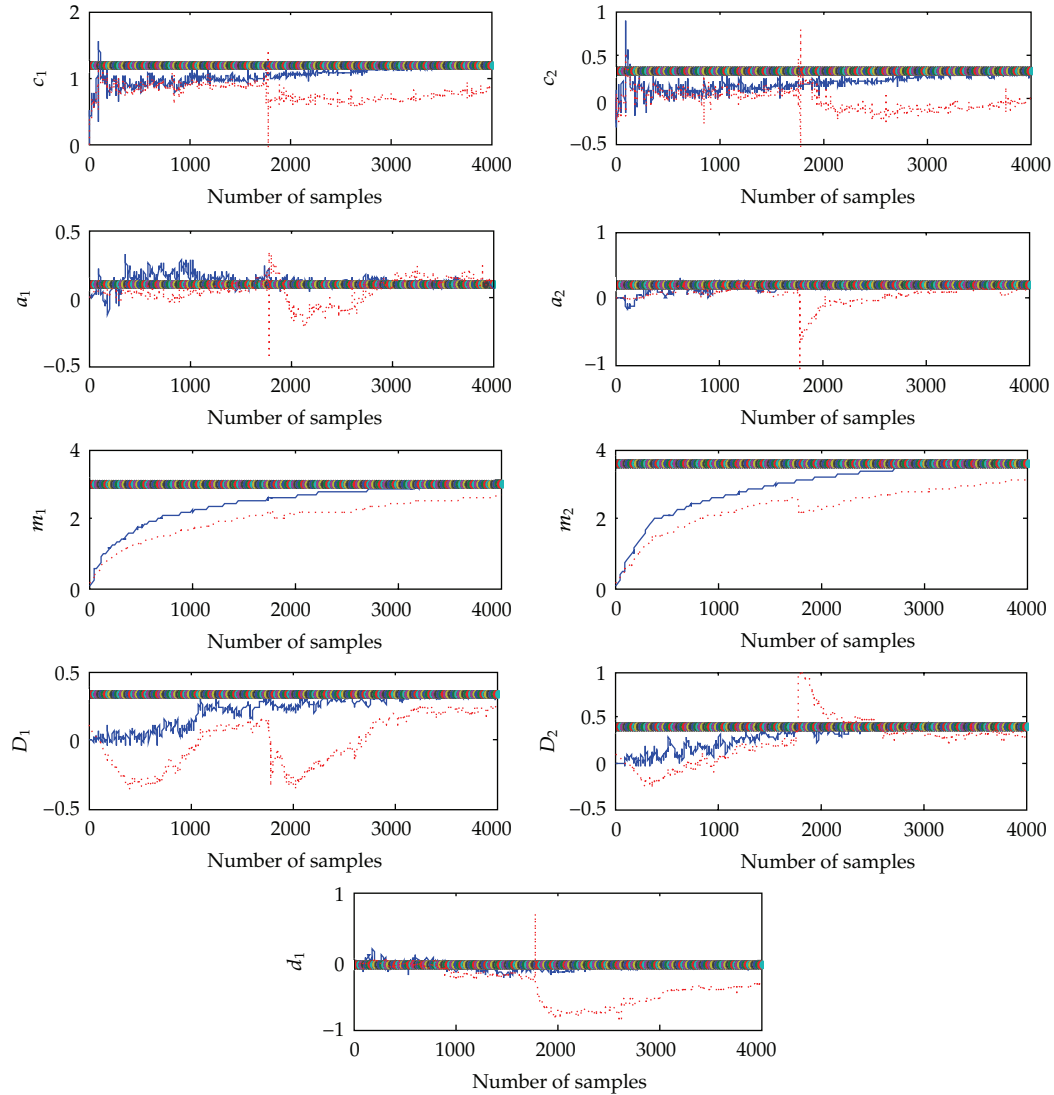


Figure 3: The convergence of the estimated parameters (noisy case): proposed method: blue and solid lines; traditional gradient method: red and dotted line.

mechanism which transforms the rotational shaft movement into linear displacement. The servomotor is controlled by a digital signal processor (TMS320LF-2407A). The displacement of each axis is measured by a linear encoder (RGF2000H125B). The signals of both phase A and phase B encoders are decoded by a quadrature decoding circuit which is based on the decoding chip (Agilent HCTL-2020).

In this system, the servomotor can be considered as a second-order linear dynamic subsystem. The movement of the work platform is also described by a linear second-order dynamic model. Due to the inherent characteristic, both dead zone and backlash-like hysteresis exist in this system. In order to simplify the identification procedure, the dead zone is compensated by a dead zone inverse model-based compensator. Thus, in the identification, only the effect of backlash-like hysteresis existing in the ball-screw-nut mechanism will

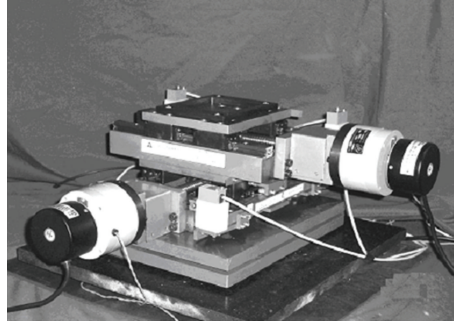


Figure 4: The X-Y moving positioning stage.

be considered. Therefore, the identified system is actually a typical sandwich system with backlash-like hysteresis. In this section, only the identification procedure of axis A will be presented due to the limited space. The corresponding models used to describe the behavior of axis A are shown as follows:

(1) the input linear model (L_1):

$$x(k) = -a_1x(k-1) - a_2x(k-2) + b_0u(k-1), \quad (6.1)$$

(2) the model of the backlash-like hysteresis:

$$v(k) = \begin{cases} m_1(x(k) - D_1), & x(k) > \frac{v(k-1)}{m_1} + D_1, \quad x(k) > x(k-1), \\ v(k-1), & \frac{v(k-1)}{m_2} - D_2 \leq x(k) \leq \frac{v(k-1)}{m_1} + D_1, \\ m_2(x(k) + D_2), & x(k) < \frac{v(k-1)}{m_2} - D_2, \quad x(k) < x(k-1), \end{cases} \quad (6.2)$$

(3) the output linear model (L_2):

$$y(k) = -c_1y(k-1) - c_2y(k-2) + d_0v(k) + d_1v(k-1), \quad (6.3)$$

where $y(k)$ is the moving speed of the work platform.

Based on the operating requirement, a sequence of square wave plus sinusoidal wave is used to excite the system within the operating range. The corresponding amplitude of the input varies in the range between -1.09 V and 1.05 V, and the sample period is 0.5 ms.

In this model, both b_0 and d_0 are set to one. The initial values of the other parameters are chosen as $\eta(0) = 0.00116$, $\mu = 1$, $\theta_0 = [0, 0, 0, 0, 1, 1, 0.001, 0.001, 0]^T$, and $\beta(k_0) = 0$. After 6700 steps, the convergence of the estimation is achieved. Figure 5 illustrates the corresponding procedure of the parameter estimation. It shows that the estimate procedure converges quickly. Figure 6 shows the corresponding mean square error (MSE) of the parameter estimation. We can see that the MSE is decreased sharply in the beginning, at the 180th step, and a local minimum can be found. After that, the algorithm jumps out of the local minimum, and the corresponding MSE gradually converges to a constant of about 0.4.

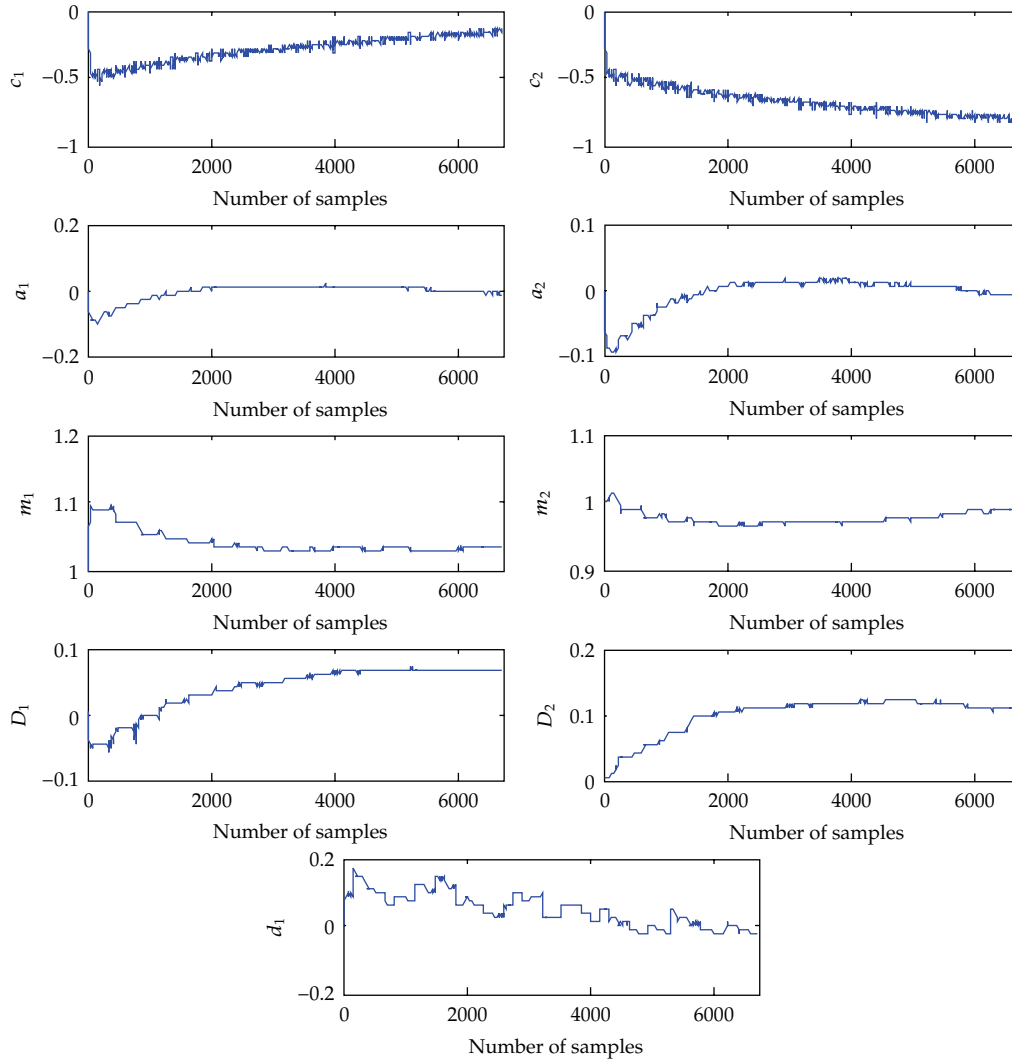


Figure 5: Convergence of parameters of the model of the X-Y moving positioning stage.

Then, the corresponding model validation result is shown in Figure 7(a), while Figure 7(b) shows the comparison of the input-output plots between the proposed model and the real data. The maximum relative modeling error is less than 11%. Moreover, it is obvious that the obtained model can accurately approximate the behavior of the X-Y moving positioning stage. Hence, it can be concluded that the proposed identification method is rather promising in engineering application.

7. Conclusion

In this paper, a recursive gradient-based identification algorithm for the sandwich system with backlash-like hysteresis is proposed. The subgradient is applied to the search of gradient direction at the nonsmooth points of the system. In order to find the proper search

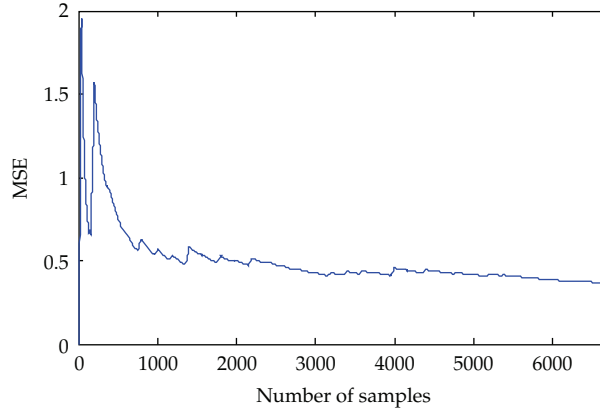


Figure 6: The MSE plot.

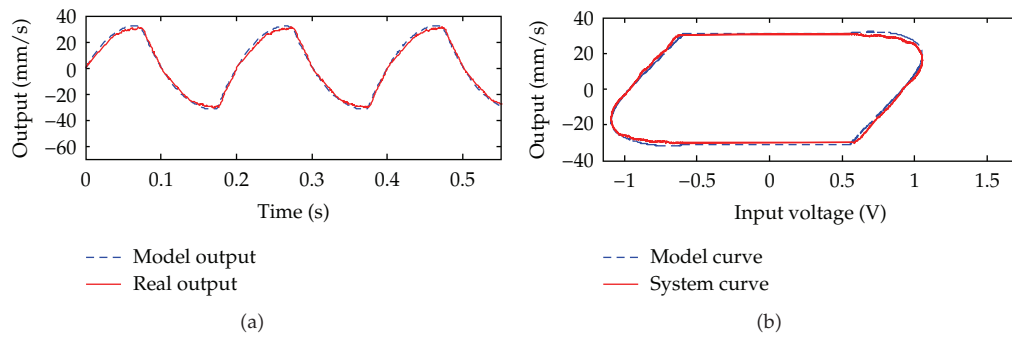


Figure 7: Model validation result.

direction at the nonsmooth points, the technique of so-called bundle method is utilized. Simulation results have shown that the proposed algorithm has provided us with an option for identification of nonsmooth dynamic systems, and it provides a novel method to identify the more complicated nonsmooth systems. The experimental results of X-Y stage also show that the proposed method has potential in engineering applications.

Appendix

Based on (3.5) and Step 3 of the algorithm, it is obtained

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \eta(k)e(k, \hat{\theta}(k))\bar{h}(k). \quad (\text{A.1})$$

Subtracting the local optimal value θ_1 from both sides of (A.1), it leads to

$$\hat{\theta}(k+1) - \theta_1 = \hat{\theta}(k) - \theta_1 + \eta(k)e(k, \hat{\theta}(k))\bar{h}(k). \quad (\text{A.2})$$

Rewrite (A.2) as

$$\tilde{\boldsymbol{\theta}}(k+1) = \tilde{\boldsymbol{\theta}}(k) + \eta(k)e(k, \hat{\boldsymbol{\theta}}(k))\bar{\mathbf{h}}(k), \quad (\text{A.3})$$

where $\tilde{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k+1) - \boldsymbol{\theta}_1$.

Choose the quadratic function as

$$L(k+1) = \tilde{\boldsymbol{\theta}}^T(k+1)\tilde{\boldsymbol{\theta}}(k+1) + e^2(k, \hat{\boldsymbol{\theta}}(k+1)). \quad (\text{A.4})$$

According to (A.3), it leads to

$$\begin{aligned} \tilde{\boldsymbol{\theta}}^T(k+1)\tilde{\boldsymbol{\theta}}(k+1) - \tilde{\boldsymbol{\theta}}^T(k)\tilde{\boldsymbol{\theta}}(k) &= 2\eta(k)e(k, \hat{\boldsymbol{\theta}}(k))\tilde{\boldsymbol{\theta}}^T(k)\bar{\mathbf{h}}(k) \\ &\quad + \eta^2(k)e^2(k, \hat{\boldsymbol{\theta}}(k))\bar{\mathbf{h}}^T(k)\bar{\mathbf{h}}(k). \end{aligned} \quad (\text{A.5})$$

As $f(k, \boldsymbol{\theta}) = [y(k) - \hat{y}(k, \boldsymbol{\theta})]^2$, if (3.5) holds, then the cutting-plane model is

$$\hat{e}_1(k, \boldsymbol{\theta}) = \max \left\{ [y(k) - \hat{y}(k, \hat{\boldsymbol{\theta}}(k))]^2 + 2e(k, \hat{\boldsymbol{\theta}}(k)) \left(\hat{\boldsymbol{\theta}}^T(k) - \boldsymbol{\theta}^T \right) \bar{\mathbf{h}}(k) - \beta_j(k) \right\}. \quad (\text{A.6})$$

Based on the definitions of $\beta_j(k)$ and $\alpha_j(k)$, as well as the idea of bundle method, we know that $\hat{e}_1(k, \boldsymbol{\theta}) \leq f(k, \hat{\boldsymbol{\theta}}(k))$. Thus, considering (A.6) yields

$$2e(k, \hat{\boldsymbol{\theta}}(k)) \left(\hat{\boldsymbol{\theta}}^T(k) - \boldsymbol{\theta}^T \right) \bar{\mathbf{h}}(k) \leq \beta_j(k). \quad (\text{A.7})$$

Choosing $\boldsymbol{\theta} = \boldsymbol{\theta}_1$ and holding (A.7) yield

$$2e(k, \hat{\boldsymbol{\theta}}(k)) \left(\hat{\boldsymbol{\theta}}^T(k) - \boldsymbol{\theta}_1^T \right) \bar{\mathbf{h}}(k) \leq \beta_j(k). \quad (\text{A.8})$$

Based on (A.5) and (A.8), we obtain

$$\tilde{\boldsymbol{\theta}}(k+1)\tilde{\boldsymbol{\theta}}^T(k+1) - \tilde{\boldsymbol{\theta}}(k)\tilde{\boldsymbol{\theta}}^T(k) \leq \eta(k)\beta_j(k) + \eta^2(k)e^2(k, \hat{\boldsymbol{\theta}}(k))\bar{\mathbf{h}}^T(k)\bar{\mathbf{h}}(k). \quad (\text{A.9})$$

According to the idea of the gradient algorithm and [16], we know that

$$e(k, \hat{\boldsymbol{\theta}}(k+1)) = e(k, \hat{\boldsymbol{\theta}}(k)) + \Delta e(k, \hat{\boldsymbol{\theta}}(k)). \quad (\text{A.10})$$

Hence, the change of $e(k)$ is written as

$$\Delta e(k, \hat{\boldsymbol{\theta}}(k)) = \left[\frac{\partial e(k, \hat{\boldsymbol{\theta}}(k))}{\partial \boldsymbol{\theta}(k)} \right]^T \Delta \boldsymbol{\theta}(k) = -\bar{\mathbf{h}}(k) \Delta \boldsymbol{\theta}(k). \quad (\text{A.11})$$

According to (3.5) and Step 3 of the algorithm, it yields:

$$\Delta \boldsymbol{\theta}(k) = \eta(k) \mathbf{t}(k) = \eta(k) e(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}^T(k). \quad (\text{A.12})$$

From (A.11) and (A.12), we get

$$\Delta e(k, \hat{\boldsymbol{\theta}}(k)) = -\eta(k) e(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k). \quad (\text{A.13})$$

According to (A.10) and (A.13), we obtain

$$e^2(k, \hat{\boldsymbol{\theta}}(k+1)) - e^2(k, \hat{\boldsymbol{\theta}}(k)) = -2\eta(k) e^2(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k) + \eta^2(k) e^2(k, \hat{\boldsymbol{\theta}}(k)) \left[\bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k) \right]^2. \quad (\text{A.14})$$

Based on (A.4), (A.9), and (A.14), it leads to the following:

$$\begin{aligned} L(k+1) - L(k) &\leq \eta(k) \beta_j(k) | \eta^2(k) e^2(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}^T(k) \bar{\mathbf{h}}(k) \left(1 + \left[\bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k) \right] \right) \\ &\quad - 2\eta(k) e^2(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k) \end{aligned} \quad (\text{A.15})$$

if $0 < \eta(k) \leq (2e^2(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k) - \beta_j(k)) / e^2(k, \hat{\boldsymbol{\theta}}(k)) \bar{\mathbf{h}}^T(k) \bar{\mathbf{h}}(k) (1 + [\bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k)])$ and $\beta_j(k) < 2e^2(k, \hat{\boldsymbol{\theta}}(k)) [\bar{\mathbf{h}}(k) \bar{\mathbf{h}}^T(k)]$, we have

$$L(k+1) - L(k) \leq 0. \quad (\text{A.16})$$

Hence, the parameters $\boldsymbol{\theta}$ can be convergent to a local optimal value.

Acknowledgments

This work was supported by the projects of Shanghai Normal University (DZL811, DRL904, and DYL201005); the projects of Shanghai Education Commission (11YZ92); the project of NSFC (Grant nos. 60971004 and 61171088); the projects of Science and Technology Commission of Shanghai (09220503000, 10JC1412200, and 09ZR1423400).

References

- [1] N. J. Bershad, P. Celka, and S. McLaughlin, "Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with Hermite polynomial expansions," *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 1060–1071, 2001.
- [2] A. H. Tan and K. Godfrey, "Identification of Wiener-Hammerstein models using linear interpolation in the frequency domain (LIFRED)," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 3, pp. 509–521, 2002.
- [3] M. Boutayeb and M. Darouach, "Recursive identification method for MISO Wiener-Hammerstein model," *IEEE Transactions on Automatic Control*, vol. 40, no. 2, pp. 287–291, 1995.
- [4] V. Cerone and D. Regruto, "Bounding the parameters of linear systems with input backlash-like hysteresis," in *Proceedings of the American Control Conference*, pp. 4476–4481, Portland, Ore, USA, June 2005.
- [5] E.-W. Bai, "Identification of linear systems with hard input nonlinearities of known structure," *Automatica*, vol. 38, no. 5, pp. 853–860, 2002.
- [6] F. Giri, Y. Rochdi, F. Z. Chaoui, and A. Brouri, "Identification of Hammerstein systems in presence of hysteresis-backlash and hysteresis-relay nonlinearities," *Automatica*, vol. 44, no. 3, pp. 767–775, 2008.
- [7] R. Dong, Y. Tan, and H. Chen, "Recursive identification for dynamic systems with backlash," *Asian Journal of Control*, vol. 12, no. 1, pp. 26–38, 2010.
- [8] R. Dong, Q. Tan, and Y. Tan, "Recursive identification for dynamic systems with output backlash-like hysteresis and its convergence," *International Journal of Applied Mathematics and Computer Science*, vol. 19, no. 4, pp. 631–638, 2009.
- [9] R. Dong and R. Tan, "Online identification algorithm and convergence analysis for sandwich systems with backlash," *International Journal of Control, Automation and Systems*, vol. 9, no. 3, pp. 1–7, 2011.
- [10] C. Eitzinger, "Nonsmooth training of fuzzy neural networks," *Soft Computing*, vol. 8, pp. 443–448.
- [11] M. M. Mäkelä, M. Miettinen, L. Lukšan, and J. Vlček, "Comparing nonsmooth nonconvex bundle methods in solving hemivariational inequalities," *Journal of Global Optimization*, vol. 14, no. 2, pp. 117–135, 1999.
- [12] S. A. Miller, *An inexact bundle method for solving large structured linear matrix inequality [Doctoral dissertation]*, University of California, Santa Barbara, Calif, USA, 2001.
- [13] C. Lemaréchal, "Nondifferentiable optimization," in *Optimization*, G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, Eds., vol. 1, pp. 529–572, North-Holland, Amsterdam, The Netherlands, 1989.
- [14] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1998.
- [15] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, 2nd edition, 1987.
- [16] C.-C. Ku and K. Y. Lee, "Diagonal recurrent neural networks for dynamic systems control," *IEEE Transactions on Neural Networks*, vol. 6, no. 1, pp. 144–156, 1995.