

Research Article

Flexible Least Squares Algorithm for Switching Models

Yunxia Ni, Lixing Lv, and Yuejiang Ji 

Wuxi Vocational College of Science and Technology, Wuxi 214122, China

Correspondence should be addressed to Yuejiang Ji; yyj1981917@126.com

Received 10 July 2022; Revised 29 August 2022; Accepted 15 September 2022; Published 26 October 2022

Academic Editor: Qingling Wang

Copyright © 2022 Yunxia Ni 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.

The self-organizing model and expectation-maximization method are two traditional identification methods for switching models. They interactively update the parameters and model identities based on offline algorithms. In this paper, we propose a flexible recursive least squares algorithm which constructs the cost function based on two kinds of errors: the neighboring two-parameter estimation errors and the output estimation errors. Such an algorithm has several advantages over the two traditional identification algorithms: it (1) can estimate the parameters of all the sub-models without prior knowledge of the model identities; (2) has less computational efforts; and (3) can update the parameters with newly arrived data. The convergence properties and simulation examples are provided to illustrate the efficiency of the algorithm.

1. Introduction

Least squares (LS) algorithm is the most widely used method in parameter estimation [1–3]. It defines a cost function which is composed of the errors between the true outputs and the predicted outputs. Then, the estimations can be obtained by solving the derivative function of the cost function. The LS algorithm has fast convergence rates but with the cost of heavy computational efforts [4, 5]. In addition, the LS algorithm needs to compute the inverse of a matrix. If the matrix has a high order or is ill-conditioned, the LS algorithm is inefficient [6–9].

To reduce the computational efforts and to avoid the matrix inversion, the recursive least squares (RLS) algorithm is a good choice. The basic idea of the RLS algorithm is to update the parameter estimations using the newly arrived data, that is, the cost function of the RLS algorithm is composed of only one set of data rather than all the collected data [10–12]. Therefore, the RLS algorithm has less computational efforts, and it does not require calculating the matrix inverse. However, the RLS algorithm has slow convergence rates when compared with the LS algorithm [13, 14]. With the aim to increase the convergence rates, many modified RLS algorithms are developed, e.g., the

multi-innovation RLS algorithm [15, 16] and the hierarchical RLS algorithm [17, 18].

Although the RLS algorithm and its modified counterparts can identify systems with less computational efforts and fast convergence rates, they have the assumption that the considered model is a single model. If the system is described by a switching model, those algorithms are inefficient. Switching models are widely used in engineering practices [19, 20]. Such models have several modes with different dynamical properties, and the modes are associated with various operating conditions [21, 22]. The difficulty in switching system identification is that the times of the operating points (model identities) may be unknown. To identify the switching models, one should first determine the operating points/model identities.

The self-organizing model (SOM)-based method and the expectation-maximization (EM) method are two classical identification algorithms which are usually used for switching models [23–26]. The SOM method introduces several cost functions which correspond to each sub-model in each iteration, and the smallest cost function is associated with the true model in this sampling instant [23]. The EM algorithm regards the model identities as hidden variables and updates these identities in the EM-E step; once the identity estimations are obtained, the parameter estimations

are computed in the EM-M step. These two steps run interactively until these two kinds of estimations converge to their true values [27, 28]. The SOM method should compute several cost functions in each iteration, and the EM algorithm needs to compute the model identities first. In addition, both these algorithms are offline algorithms. That is, they have heavy computational efforts and cannot update the parameters based on newly arrived data.

The flexible least squares (FLS) algorithm, first developed by Kalaba and Tesfatsion [29], is used for time-varying system identification. Its cost function contains two parts: one is the error of the two parameters in the two neighboring instants, and the other is the error between the true outputs and predicted outputs [30–32]. Due to the first error, the parameter estimations can catch the varying parameters. Inspired by the FLS algorithm, we develop a novel FLS algorithm for switching models. This algorithm is termed as flexible recursive least squares (FRLS) algorithm. Compared with the SOM, EM, and FLS algorithms, this algorithm has the following advantages: (1) the FRLS algorithm is an online algorithm, and thus it can update the parameters with newly arrived data; (2) the FRLS algorithm has less computational efforts; and (3) the FRLS algorithm can estimate the parameters of all the sub-models without prior knowledge of the model identities.

The remainder of the paper is organized as follows. Section 2 explains the switching model and traditional identification algorithms. Section 3 proposes the offline FLS algorithm and online FLS algorithm. Section 4 provides several simulation examples. Finally, Section 5 summarizes the paper and gives some future directions.

2. Problem Statement

Let us define some notations first: \mathbf{I} means an identity matrix of the appropriate sizes; the superscript T stands for the matrix transpose; the norm of a matrix \mathbf{X} is defined as $\|\mathbf{X}\| = \sqrt{\lambda_{\max}[\mathbf{X}\mathbf{X}^{\text{T}}]}$; $\lambda_{\max}[\mathbf{X}\mathbf{X}^{\text{T}}]$ means the maximum eigenvalue of matrix $\mathbf{X}\mathbf{X}^{\text{T}}$; and the norm of a vector $\mathbf{z} = [z_1, z_2, \dots, z_n]^{\text{T}} \in \mathbb{R}^n$ is defined as $\|\mathbf{z}\| = (\sum_{i=1}^n z_i^2)^{1/2}$.

2.1. Switching Model. Consider the following switching model:

$$y_i(t) = \varphi_i^{\text{T}}(t)\vartheta_i + v_i(t), \quad i = 1, \dots, N, \quad (1)$$

where $y_i(t)$ is the output of the i -th model; $\varphi_i(t) \in \mathbb{R}^{m_i}$ is the information vector of the i -th model, which is composed of the input and output data before the sampling instant t ; ϑ_i is the parameter vector of the i -th model; $v_i(t)$ is a Gaussian white noise and satisfies $v_i(t) \sim N(0, \sigma_i^2)$; and N is the number of the sub-models.

In the sampling instant t , there is no knowledge of the identity of the model. We aim to estimate the parameter vectors $\vartheta_i, i = 1, \dots, N$ based on the collected data.

Collect L sets of input and output data, and define the following cost function:

$$J(\vartheta_1, \vartheta_2, \dots, \vartheta_N) = \sum_{t=1}^L \sum_{i=1}^N w_i(t) [y_i(t) - \varphi_i^{\text{T}}(t)\vartheta_i]^2, \quad (2)$$

where $w_i(t)$ is the model identity in the sampling instant t . For example, in the sampling instant t , the true model is the s -th model, and then the true values of the identities of all the sub-models are $w_1(t) = 0, w_2(t) = 0, \dots, w_{s-1}(t) = 0, w_s(t) = 1, w_{s+1}(t) = 0, \dots, w_N(t) = 0$. To estimate the parameters, the following assumptions are introduced.

Assumption 1. The number of the collected data is larger than the number of the unknown parameters, that is,

$$L > m_1 + m_2 + \dots + m_N. \quad (3)$$

In addition, assume that the number of data of the i -th sub-model is L_i ; then,

$$L_i > m_i. \quad (4)$$

Assumption 2. For the switching model proposed in (1), all the input data are taken as persistent excited.

Assumption 3. All the sub-models have the same information vector but different parameter vectors, that is, the switching model can be written by

$$y_i(t) = \varphi^{\text{T}}(t)\vartheta_i + v_i(t), \quad i = 1, \dots, N, \quad \varphi(t) \in \mathbb{R}^m. \quad (5)$$

Remark 1. Assumptions 1 and 2 can ensure that the information matrices of all the sub-models are nonsingular [6]. Assumption 3 can also be easily obtained [33]. For example, for a switching model with unknown structures, we can use the kernel method to describe the model, and all the sub-models approximated by using the kernel method can have the same structure.

2.2. Traditional Identification Algorithms. Rewrite the cost function of the switching model as follows:

$$J(\vartheta_1, \vartheta_2, \dots, \vartheta_N) = \sum_{t=1}^L \sum_{i=1}^N w_i(t) [y(t) - \varphi^{\text{T}}(t)\vartheta_i]^2. \quad (6)$$

Assume that the parameter estimations and identity estimations in iteration $k-1$ are $\hat{\vartheta}_i^{k-1}, i = 1, 2, \dots, N$ and $\hat{w}_i^{k-1}(t), t = 1, 2, \dots, L$.

Both the SOM and EM algorithms estimate the parameters through two steps:

- (1) Estimate the model identity estimations $\hat{w}_i^k(t), i = 1, 2, \dots, N, t = \bar{1}, 2, \dots, L$, based on the parameter estimations $\hat{\vartheta}_i^{k-1}, i = 1, 2, \dots, N$.
- (2) Update the parameter estimations $\hat{\vartheta}_i^k, i = 1, 2, \dots, N$ based on the model identity estimations $\hat{w}_i^k(t), i = 1, 2, \dots, N, t = 1, 2, \dots, L$.

The difference between the SOM and EM algorithms is in the first step. In the SOM algorithm, the model identity estimate $\hat{w}_i^k(t)$ is 1 or 0. For example, in iteration k of the sampling instant t , let

$$\hat{\vartheta}_s^{k-1} = \operatorname{argmin}_{\hat{\vartheta}_i^{k-1}(t)} \left[\left[y(t) - \varphi^T(t) \hat{\vartheta}_i^{k-1} \right]^2, \quad i = 1, 2, \dots, N \right]. \quad (7)$$

Then, $w_s^k(t) = 1$ and the other identity estimations $w_i^k(t) = 0, i \neq s$.

On the other hand, in the EM algorithm, let

$$\varepsilon_i(t) = \left[y(t) - \varphi^T(t) \hat{\vartheta}_i^{k-1} \right]^2, \quad i = 1, 2, \dots, N. \quad (8)$$

Then, the identity estimate $w_j^k(t)$ can be computed by

$$w_j^k(t) = \frac{1/\varepsilon_j(t)}{\sum_{i=1}^N 1/\varepsilon_i(t)}. \quad (9)$$

Remark 2. Both the SOM and EM algorithms are offline algorithms; if the order of the system is large, their computational efforts are heavy. In addition, they cannot update the parameters with newly arrived data [23, 34].

3. Flexible Recursive Least Squares Algorithm

The SOM and EM algorithms update the parameters through two steps, and these two steps are related to each other. If one kind of estimations has poor estimation accuracy, the other may be also poor or divergent. In this section, we use the FLS algorithm for the switching models, which can estimate the parameters without prior knowledge of the model identities.

3.1. Offline FLS Algorithm. Define

$$\begin{aligned} Y(L) &= [y(L), y(L-1), \dots, y(1)]^T, \\ \Phi(L) &= \begin{bmatrix} \varphi(L) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \varphi(1) \end{bmatrix} \in \mathbb{R}^{Lm \times L}, \\ \Theta(L) &= [\vartheta^T(L), \dots, \vartheta^T(1)] \in \mathbb{R}^{Lm}, \\ V(L) &= [v(L), v(L-1), \dots, v(1)]^T. \end{aligned} \quad (10)$$

Then, the switching model can be written as

$$Y(L) = \Phi^T(L) \Theta(L) + V(L). \quad (11)$$

Let

$$\Omega = \begin{bmatrix} \mu I & -\mu I & 0 & \dots & 0 & 0 \\ -\mu I & 2\mu I & -\mu I & \dots & 0 & 0 \\ 0 & -\mu I & 2\mu I & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2\mu I & -\mu I \\ 0 & 0 & 0 & \dots & -\mu I & \mu I \end{bmatrix}. \quad (12)$$

Unlike the SOM and EM algorithms, the cost function of the offline FLS algorithm is written by

$$J(\Theta(L)) = \|Y(L) - \Phi^T(L) \Theta(L)\|^2 + \Theta^T(L) \Omega \Theta(L). \quad (13)$$

Using the FLS algorithm to update the parameters yields

$$\hat{\Theta}(L) = [\Omega + \Phi(L) \Phi^T(L)]^{-1} \Phi(L) Y(L). \quad (14)$$

Remark 3. From equation (5), we can get that the offline FLS (O-FLS) algorithm can estimate the parameters in only one iteration and does not require the model identify estimation. However, it increases the order of the information matrix intensively. For example, the order of the information matrix is Lm in the O-FLS algorithm, while in the EM and SOM methods, the order is m .

Remark 4. The offline FLS (O-FLS) algorithm requires a high-order matrix inverse calculation, e.g., $[\Omega + \Phi(L) \Phi^T(L)]^{-1}$. When the high-order matrix is singular or ill-conditioned, computing its inverse is impossible.

3.2. Flexible Recursive Least Squares Algorithm. To reduce the computational efforts and to avoid a high-order matrix inversion, this section proposes an online FLS algorithm which is termed as flexible recursive least squares (FRLS) algorithm.

Assume that the parameter vector in the sampling instant $t-1$ is $\hat{\vartheta}(t-1)$. Define the following cost function:

$$\begin{aligned} J(\vartheta(t)) &= \sum_{s=1}^t [y(s) - \varphi^T(s) \vartheta(s)]^2 \\ &+ \mu \sum_{s=1}^t [\vartheta(s) - \vartheta(s-1)]^T [\vartheta(s) - \vartheta(s-1)]. \end{aligned} \quad (15)$$

In the sampling instant t , all the parameter estimations before t have been obtained, and thus (15) is simplified as

$$J(\vartheta(t)) = \left[y(t) - \varphi^T(t)\vartheta(t) \right]^2 + \mu [\vartheta(t) - \widehat{\vartheta}(t-1)]^T [\vartheta(t) - \widehat{\vartheta}(t-1)]. \quad (16)$$

Taking the derivative of $J(\vartheta(t))$ with respect to $\vartheta(t)$ yields

$$\widehat{\vartheta}(t) = \left[\varphi(t)\varphi^T(t) + \mu I \right]^{-1} [\varphi(t)y(t) + \mu \widehat{\vartheta}(t-1)]. \quad (17)$$

Next, we use the recursive method to obtain the relationships between $\widehat{\vartheta}(t)$ and $\widehat{\vartheta}(t-1)$. (17) is transformed into

$$\begin{aligned} \widehat{\vartheta}(t) &= \left[\varphi(t)\varphi^T(t) + \mu I \right]^{-1} \\ &\quad \left[\varphi(t)y(t) + \mu \widehat{\vartheta}(t-1) + \varphi(t)\varphi^T(t)\widehat{\vartheta}(t-1) - \varphi(t)\varphi^T(t)\widehat{\vartheta}(t-1) \right] \\ &= \left[\varphi(t)\varphi^T(t) + \mu I \right]^{-1} \\ &\quad \left[\mu \widehat{\vartheta}(t-1) + \varphi(t)\varphi^T(t)\widehat{\vartheta}(t-1) + \varphi(t)y(t) - \varphi(t)\varphi^T(t)\widehat{\vartheta}(t-1) \right] \\ &= \widehat{\vartheta}(t-1) + \left[\varphi(t)\varphi^T(t) + \mu I \right]^{-1} \varphi(t) \times \left[y(t) - \varphi^T(t)\widehat{\vartheta}(t-1) \right]. \end{aligned} \quad (18)$$

Then, the FRLS algorithm can be summarized as follows:

$$\begin{aligned} \widehat{\vartheta}(t) &= \widehat{\vartheta}(t-1) + Q(t)\varphi(t)E(t), \\ E(t) &= \left[y(t) - \varphi^T(t)\widehat{\vartheta}(t-1) \right], \\ Q(t) &= \left[\varphi(t)\varphi^T(t) + \mu I \right]^{-1}. \end{aligned} \quad (19)$$

Remark 5. Compared with the O-FLS algorithm, the FRLS algorithm performs a low-order (m - order) matrix inversion rather than a high-order (Lm - order) matrix inversion. Therefore, the FRLS algorithm has less computational efforts than the O-FLS algorithm (Algorithm1).

Then, the steps of the FRLS algorithm are listed as follows.

In the FRLS algorithm, there exists a dense matrix inversion, which leads to heavy computational efforts. To further reduce the computational efforts, the following lemma is introduced.

Lemma 1. For the matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times r}$, and $C \in \mathbb{R}^{r \times n}$, if the matrix A is nonsingular, the following equality holds:

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}. \quad (20)$$

Proof. For the matrix $(A + BC)$, we have

$$\begin{aligned} &(A + BC) \left[A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1} \right] \\ &= I - B(I + CA^{-1}B)^{-1}CA^{-1} + BCA^{-1} - BCA^{-1}B(I + CA^{-1}B)^{-1}CA^{-1} \\ &= I - [B + BCA^{-1}B](I + CA^{-1}B)^{-1}CA^{-1} + BCA^{-1} \\ &= I - BI + CA^{-1}B^{-1}CA^{-1} + BCA^{-1} \\ &= I. \end{aligned} \quad (21)$$

In addition, we can obtain

```

(i) Initialise  $\hat{\vartheta}(0) = 1/p_0$ ,  $p_0 = 10^6$ ,  $\mathbf{1}$  is a vector whose
(ii)   entries all equal to 1, and assign a positive constant  $\mu$ 
(iii) repeat
(iv)   for  $t = 1, 2, \dots$ , do
(v)     Collect the input-output data  $u(t), y(t)$ 
(vi)    Form  $\varphi(t)$ 
(vii)   Compute  $E(t)$ 
(viii)  Compute  $Q(t)$ 
(ix)    Update  $\hat{\vartheta}(t)$ 
(x)     end
(xi)   until convergence

```

ALGORITHM 1: FRLS algorithm.

$$\begin{aligned}
& \left[A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1} \right] (A + BC) \\
&= I + A^{-1}BC - A^{-1}B(I + CA^{-1}B)^{-1}C - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}BC \\
&= I + A^{-1}BC - A^{-1}B(I + CA^{-1}B)^{-1}(C + CA^{-1}BC) \\
&= I + A^{-1}BC - A^{-1}B(I + CA^{-1}B)^{-1}(I + CA^{-1}B)C \\
&= I.
\end{aligned} \tag{22}$$

Then, the proof is completed.

According to Lemma 1, the matrix $Q(t)$ is simplified as

$$\begin{aligned}
Q(t) &= [\varphi(t)\varphi^T(t) + \mu I]^{-1} \\
&= \frac{1}{\mu} I - \frac{1/\mu^2 \varphi(t)\varphi^T(t)}{1 + 1/\mu \varphi^T(t)\varphi(t)} \\
&= \frac{1}{\mu} I - \frac{\varphi(t)\varphi^T(t)}{\mu^2 + \mu \varphi^T(t)\varphi(t)}.
\end{aligned} \tag{23}$$

Remark 6. Based on equation (15), in each sampling instant, a dense matrix inversion is transformed into vector multiplication. Therefore, the computational efforts are reduced.

3.3. Convergence Properties of the Two Kinds of FLS Algorithms. The convergence properties of the O-FLS and FRLS algorithms are given in this section which can help the researchers follow these two algorithms.

3.3.1. Convergence Property of the O-FLS Algorithm

Theorem 1. For the switching model proposed in (1), the parameter estimations $\hat{\Theta}(L)$ updated by the O-FLS algorithm are expressed by (5). Then, $\hat{\Theta}(L)$ are biased.

Proof. Rewrite the O-FLS algorithm as follows:

$$\hat{\Theta}(L) = [\Omega + \Phi(L)\Phi^T(L)]^{-1} \Phi(L)Y(L). \tag{24}$$

Substituting $Y(L) = \Phi^T(L)\Theta(L) + V(L)$ into the above equation yields

$$\hat{\Theta}(L) = [\Omega + \Phi(L)\Phi^T(L)]^{-1} \Phi(L) [\Phi^T(L)\Theta(L) + V(L)], \tag{25}$$

where $V(L)$ is Gaussian white and independent on $\Phi(L)$, and the above equation can be written by

$$\hat{\Theta}(L) = [\Omega + \Phi(L)\Phi^T(L)]^{-1} \Phi(L)\Phi^T(L)\Theta(L). \tag{26}$$

Since $\Phi(L) \in \mathbb{R}^{Lm \times L}$, the matrix $\Phi(L)\Phi^T(L)$ is singular, and the matrix Ω cannot be a zero matrix. Therefore, the O-FLS algorithm is a biased algorithm.

Remark 7. A small μ can get more accurate parameter estimations. However, a small μ may lead to slow convergence rates between the two neighboring sub-models. Therefore, we should assign different values for μ . For example, in the fixed interval, a small μ is better, while near the switching points, a larger one is better.

3.3.2. Convergence Property of the FRLS Algorithm

Theorem 2. For the switching model proposed in (1), the parameter estimations $\hat{\vartheta}(t)$ updated by the FRLS algorithm are expressed by (10)–(12). Then, the sequence $\{\hat{\vartheta}(t)\}$ is convergent.

Proof. The FRLS algorithm is written by

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + [\varphi(t)\varphi^T(t) + \mu I]^{-1} \varphi(t) [y(t) - \varphi^T(t)\hat{\vartheta}(t-1)]. \tag{27}$$

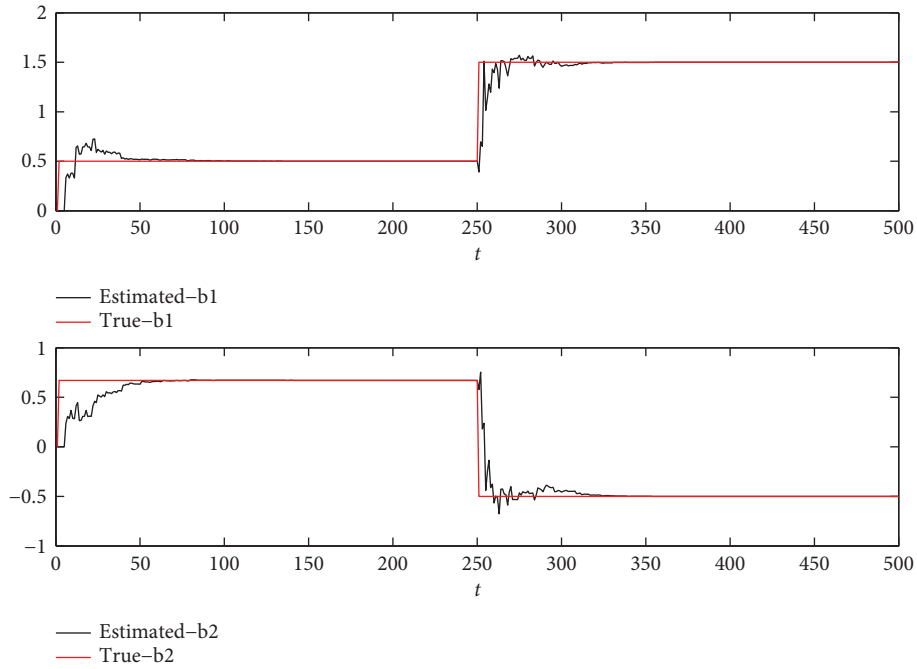


FIGURE 1: The parameter estimations and true parameters b_1 and b_2 .

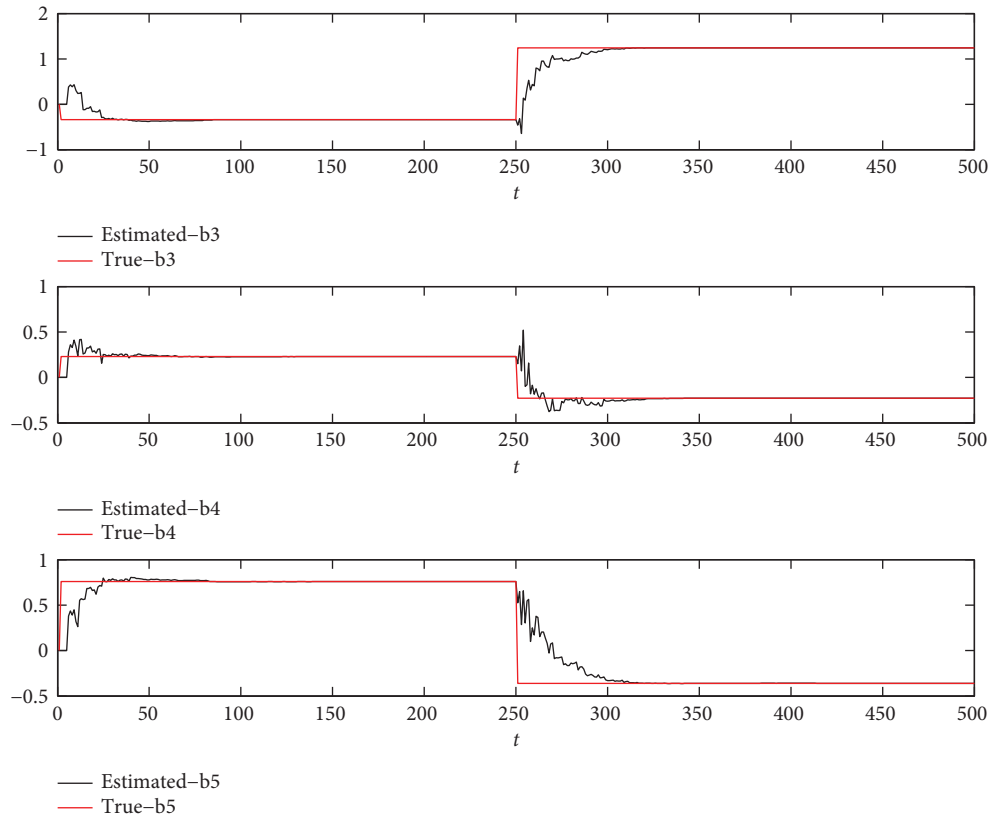


FIGURE 2: The parameter estimations and true parameters $b_3 - b_5$.

TABLE 1: The elapsed times and estimation errors of these three algorithms.

Algorithms	FRLS	EM	SOM
Elapsed times (second)	1.704	2.257	2.038
Estimation errors	1.201%	1.145%	1.273%

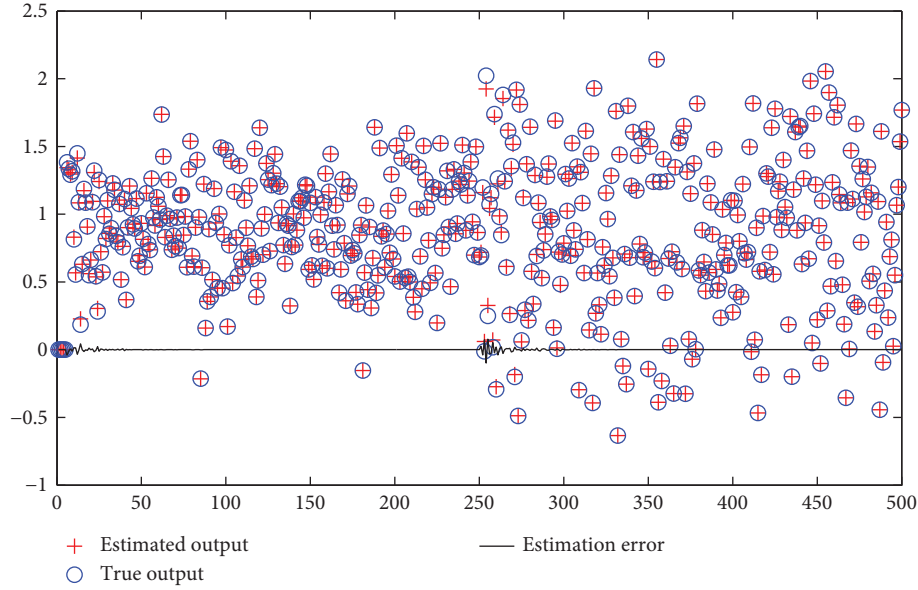


FIGURE 3: The output estimations and their estimation errors.

Assume that the data from $1 \rightarrow L_1, L_1 \gg m$ belong to model 1, and subtracting the true value ϑ_1 on both sides of the above equation yields

$$\begin{aligned} \hat{e}(t) &= \hat{e}(t-1) - [\varphi(t)\varphi^T(t) + \mu I]^{-1} \varphi(t)\varphi^T \hat{e}(t-1) \\ &= [\mathbf{I} - [\varphi(t)\varphi^T(t) + \mu I]^{-1} \varphi(t)\varphi^T(t)] \hat{e}(t-1). \end{aligned} \quad (28)$$

For the reason that

$$\left\| \left[\mathbf{I} - [\varphi(t)\varphi^T(t) + \mu I]^{-1} \varphi(t)\varphi^T(t) \right] \right\| < 1, \quad (29)$$

we have

Sub – model1:

$$y(t) = 0.5u(t-1) + 0.67u(t-2) - 0.34u(t-3) + 0.23u(t-4) + 0.76u(t-5) + v(t),$$

Sub – model2:

$$\begin{aligned} y(t) &= 1.5u(t-1) - 0.5u(t-2) + 1.24u(t-3) - 0.23u(t-4) - 0.36u(t-5) + v(t), \\ u(t) &\sim N(0, 1), v(t) \sim N(0, 0.1^2). \end{aligned} \quad (31)$$

Let

$$\begin{aligned} \vartheta &= [b_1, b_2, b_3, b_4, b_5]^T, \\ \vartheta_1 &= [b_1, b_2, b_3, b_4, b_5]^T = [0.5, 0.67, -0.34, 0.23, 0.76]^T, \\ \vartheta_2 &= [b_1, b_2, b_3, b_4, b_5]^T = [1.5, -0.5, 1.24, -0.23, -0.36]^T. \end{aligned} \quad (32)$$

$$\|\hat{e}(t)\| < \|\hat{e}(t-1)\|. \quad (30)$$

Therefore, the FRLS algorithm is convergent. \square

Remark 8. The FRLS algorithm has the assumption that the identities of the data are unchanging in a fixed interval. If the identities are changing continually, the FRLS algorithm is divergent.

4. Examples

4.1. *Example 1.* Consider the following switching model:

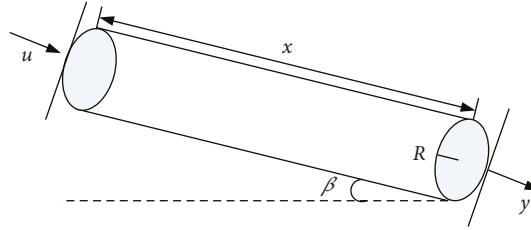


FIGURE 4: The open channel system.

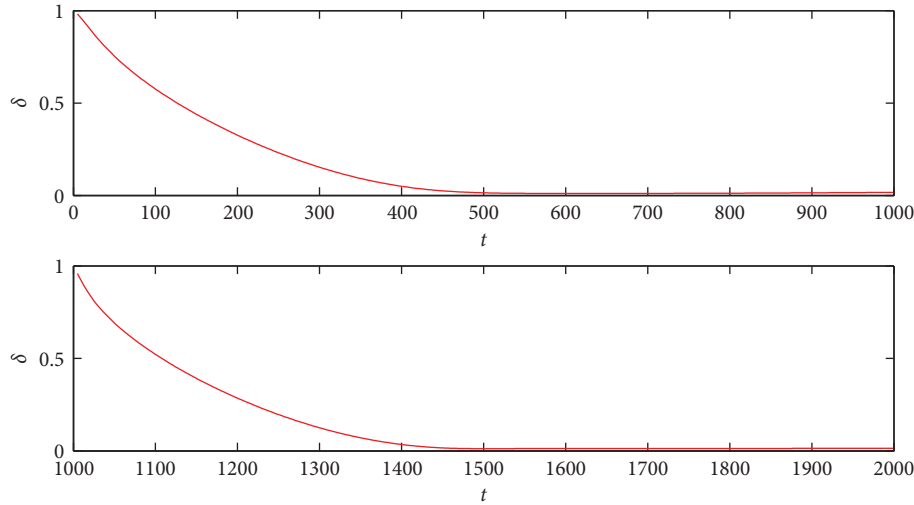


FIGURE 5: The parameter estimation errors (FRLS).

In simulation, we collect 500 sets of input and output data, where the data from 1: 250 belong to model 1, and those from 251: 500 belong to model 2.

Use the FRLS algorithm for this switching model. The parameter estimations are shown in Figures 1 and 2. The predicted outputs and the true outputs, and their errors are shown in Figure 3. In addition, apply the EM and SOM algorithms for the switching model, where the initial identities for each sub-model are $\omega_j^0(t) = 1/2$, $j = 1, 2$ and $t = 1, 2, \dots, 500$. The estimation errors and elapsed times of the three algorithms are shown Table 1.

From this simulation, we can get the following findings:

- (1) The parameter estimations using the FRLS algorithm can asymptotically converge to the true values (see Figures 1 and 2).
- (2) The predicted outputs using the FRLS algorithm can catch the true outputs (see Figure 3).
- (3) The number of the data in a fixed interval must be larger than the number of the unknown parameters.
- (4) All the FRLS, EM, and SOM algorithms are effective for the switching model, but the FRLS algorithm has the smallest elapsed times, that is, the FRLS

TABLE 2: The parameter estimations of sub-model 1.

t	$\vartheta_{1,1}$	$\vartheta_{1,2}$	$\vartheta_{1,3}$	$\vartheta_{1,4}$	$\vartheta_{1,5}$	δ_1 (%)
100	0.17234	0.23763	0.01434	0.02163	0.00859	57.58089
300	0.34494	0.47465	0.01688	0.02067	0.00220	15.25782
500	0.40302	0.55258	0.01768	0.02042	0.00016	1.40728
700	0.40657	0.55271	0.01770	0.02038	0.00008	1.10091
800	0.40818	0.55113	0.01770	0.02036	0.00006	1.20170
1000	0.41006	0.54930	0.01770	0.02032	0.00004	1.40738
True values	0.40970	0.55790	0.01830	0.02020	-0.0045	

algorithm has the least computational efforts among these three algorithms, as shown in Table 1.

4.2. Example 2: A Switching Open Channel System. In this section, we consider an open channel system, which is shown in Figure 4. The radius of the channel is R , the length of the channel is x , $u(t)$ is the discharge at the upstream end, $y(t)$ is the discharge at the downstream end, and the slope is β . To ensure the discharge $y(t)$ to flow in a fixed speed, we should control $u(t)$. The relationship between $u(t)$ and $y(t)$ can be expressed by a linear model. In simulation, two slopes $\beta_1 = 10$ and $\beta_2 = 15$ degrees are assigned to the open

TABLE 3: The parameter estimations of sub-model 2.

t	$\vartheta_{2,1}$	$\vartheta_{2,2}$	$\vartheta_{2,3}$	$\vartheta_{2,4}$	$\vartheta_{2,5}$	δ_2 (%)
1100	0.38576	0.07080	0.03194	0.03643	0.03259	52.18099
1300	0.80948	0.14792	0.03312	0.02342	0.07819	1.29770
1500	0.80628	0.14756	0.03302	0.02351	0.07556	1.20113
1700	0.80948	0.14792	0.03312	0.02342	0.07819	1.29770
1800	0.80960	0.14778	0.03313	0.02342	0.07823	1.31978
1000	0.81009	0.14728	0.03313	0.02341	0.07822	1.39991
True values	0.80470	0.15730	0.03300	0.02310	0.07630	

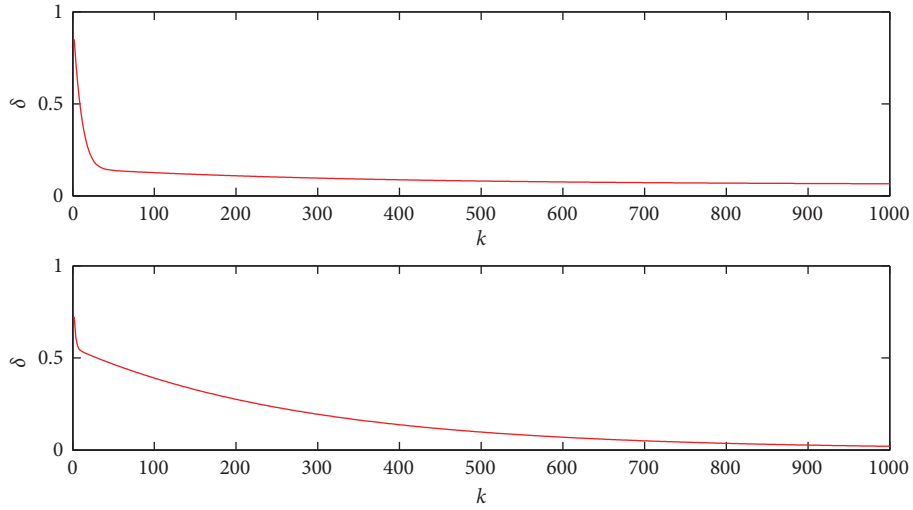


FIGURE 6: The parameter estimation errors (EM).

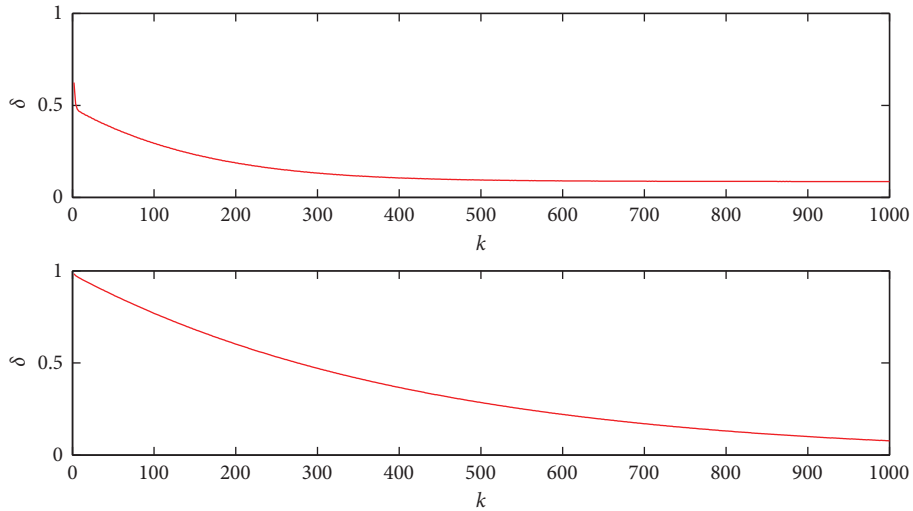


FIGURE 7: The parameter estimation errors (SOM).

TABLE 4: The elapsed times of these three algorithms.

Algorithms	FRLS	EM	SOM
Elapsed times (second)	1.952	2.59	2.33

channel system. These two slopes lead to two different dynamics which should be described by two models: [14]:

Sub – model1:

$$y(t) = 0.4097y(t-1) + 0.5579y(t-2) + 0.0183u(t-1) \\ + 0.0202u(t-2) - 0.0045u(t-3) + v(t), \\ t = 1, 2, \dots, 1000,$$

Sub – model2:

$$y(t) = 0.8047y(t-1) + 0.1573y(t-2) + 0.0330u(t-1) \\ + 0.0231u(t-2) + 0.0763u(t-3) + v(t), \\ t = 1001, 1002, \dots, 2000.$$

(33)

We collect 2000 sets of input-output data using Matlab software, where the sequence $\{u(t)\}$ is generated by

$$u = \text{idinput}(L', RBS', [0, 1], [-1, 1]). \quad (34)$$

The data from $t = 1: 1000$ belong to model 1, and those from 1001: 2000 belong to model 2.

Apply the FRLS algorithm for the considered switching model. The parameter estimations and their estimation errors $\delta = \|\hat{\vartheta}_i(t) - \vartheta\|/\|\vartheta\|$, $i = 1, 2$, are shown in Figure 5 and Tables 2 and 3.

Furthermore, we use the traditional EM and SOM algorithms for the switching open channel system ($\omega_j^0(t) = 1/2$, $j = 1, 2$ and $t = 1, 2, \dots, 2000$). The parameter estimations and their estimation errors are shown in Figures 6 and 7. The elapsed times of these three algorithms are shown Table 4.

This example shows that (1) all the FRLS, EM, and SOM algorithms are convergent, as shown in Figures 5–7; (2) the FRLS algorithm has the smallest elapsed times among these three algorithms, and this is shown in Table 4, that is, the FRLS algorithm has the least computational efforts among these three algorithms.

5. Conclusions

An online FLS algorithm, termed as flexible recursive least squares (FRLS) algorithm, is proposed for switching models in this study. Its cost function is composed of the errors between the two neighboring parameter estimations and the errors between the true outputs and the predicted outputs. With the help of the two neighboring parameter estimation errors, the operating points of the switching models can be determined, and the parameters of each sub-model can also be obtained. Compared with the SOM and EM algorithms, the FRLS algorithm can estimate the parameter estimations without prior knowledge of the model identities. In addition, the FRLS algorithm is an online algorithm, which has less computational efforts and can update the parameters with newly arrived data.

Although the FRLS algorithm has several advantages over the traditional identification algorithms, several challenging issues about the FRLS algorithm need to be considered in future. For example, if the sub-models switch continually, how to apply the FRLS algorithm to the switching model? How to choose a suitable μ to make the FRLS algorithm converge quickly to the true values? These topics remain as open problems.

Data Availability

All data generated or analyzed during this study are included in this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Natural Science Foundation of Jiangsu Province (no. BK20131109).

References

- [1] C. P. Yu, L. Ljung, A. Wills, and M. Verhaegen, “Constrained subspace method for the identification of structured state-space models (COSMOS),” *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 4201–4214, 2020.
- [2] B. Q. Mu, E. W. Bai, W. X. Zheng, and Q. M. Zhu, “A globally consistent nonlinear least squares estimator for identification of nonlinear rational systems,” *Automatica*, vol. 77, pp. 322–335, 2017.
- [3] 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, 2011.
- [4] F. Ding, H. Ma, J. Pan, and E. F. Yang, “Hierarchical gradient and least squares-based iterative algorithms for input nonlinear output-error systems using the key term separation,” *Journal of the Franklin Institute*, vol. 358, no. 9, pp. 5113–5135, 2021.
- [5] J. M. Li and F. Ding, “Identification methods of nonlinear systems based on the kernel functions,” *Nonlinear Dynamics*, vol. 104, no. 3, pp. 2537–2552, 2021.
- [6] T. Soderstrom and P. Stoica, *System Identification*, Prentice-Hall, Hoboken, New Jersey, USA, 1989.
- [7] J. Chen, J. Ma, M. Gan, and Q. Zhu, “Multi-direction gradient iterative algorithm: a unified framework for gradient iterative and least squares algorithms,” *IEEE Transactions on Automatic Control*, p. 1, 2021.
- [8] J. Chen, B. Huang, M. Gan, and C. P. Chen, “A novel reduced-order algorithm for rational models based on Arnoldi process and Krylov subspace,” *Automatica*, vol. 129, Article ID 109663, 2021.
- [9] G. Y. Chen, M. Gan, C. L. P. Chen, and H. X. Li, “A regularized variable projection algorithm for separable nonlinear

- least-squares problems,” *IEEE Transactions on Automatic Control*, vol. 64, no. 2, pp. 526–537, 2019.
- [10] M. Jiao, D. Q. Wang, Y. Yang, and F. Liu, “More intelligent and robust estimation of battery state-of-charge with an improved regularized extreme learning machine,” *Engineering Applications of Artificial Intelligence*, vol. 104, no. 2, Article ID 104407, 2021.
- [11] Q. L. Liu, F. Y. Chen, F. Ding, A. Alsaedi, and T. Hayat, “Recursive least squares estimation methods for a class of nonlinear systems based on non-uniform sampling,” *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 8, pp. 1612–1632, 2021.
- [12] M. Jiao and D. Q. Wang, “The Savitzky-Golay filter based bidirectional long short-term memory network for SOC estimation,” *International Journal of Energy Research*, vol. 45, no. 13, pp. 19467–19480, 2021.
- [13] L. Xu, F. Y. Chen, F. Ding, A. Alsaedi, and T. Hayat, “Hierarchical recursive signal modeling for multi-frequency signals based on discrete measured data,” *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 5, pp. 676–693, 2021.
- [14] L. Bako, K. Boukharouba, E. Duviella, and S. Lecoeuche, “A recursive identification algorithm for switched linear/affine models,” *Nonlinear Analysis: Hybrid Systems*, vol. 5, no. 2, pp. 242–253, 2011.
- [15] S. J. Fan, L. Xu, F. Ding, A. Alsaedi, and T. Hayat, “Correlation analysis-based stochastic gradient and least squares identification methods for errors-in-variables systems using the multi-innovation,” *International Journal of Control, Automation and Systems*, vol. 19, no. 1, pp. 289–300, 2021.
- [16] H. F. Xia, Y. Ji, L. Xu, A. Alsaedi, and T. Hayat, “Maximum likelihood-based gradient estimation for multivariable nonlinear systems using the multi-innovation identification theory,” *International Journal of Robust and Nonlinear Control*, vol. 30, no. 14, pp. 5446–5463, 2020.
- [17] M. Gan, H. T. Zhu, G. Y. Chen, and C. L. P. Chen, “Weighted generalized cross validation based regularization for broad learning system,” *IEEE Transactions on Cybernetics*, vol. 52, no. 5, pp. 4064–4072, 2022.
- [18] Y. H. Zhou, X. Zhang, and F. Ding, “Hierarchical estimation approach for RBF-AR models with regression weights based on the increasing data length,” *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 68, no. 12, pp. 3597–3601, 2021.
- [19] H. Y. Yan and Y. G. Zhu, “Bang-bang control model for uncertain switched systems,” *Applied Mathematical Modelling*, vol. 39, no. 10–11, pp. 2994–3002, 2015.
- [20] F. Y. Wu and J. Lian, “Stabilization of constrained switched systems via multiple Lyapunov R-functions,” *Systems & Control Letters*, vol. 139, p. 104686, 2020.
- [21] F. Bianchi, V. Breschi, D. Piga, and L. Piroddi, “Model structure selection for switched NARX system identification: a randomized approach,” *Automatica*, vol. 125, Article ID 109415, 109415 pages, 2021.
- [22] F. Lauer and G. Bloch, *Hybrid System Identification*, Springer, Berlin, Germany, 2019.
- [23] J. E. Saxen, H. Saxen, and H. T. Toivonen, “Identification of switching linear systems using self-organizing models with application to silicon prediction in hot metal,” *Applied Soft Computing*, vol. 47, pp. 271–280, 2016.
- [24] J. Chen, J. X. Ma, Y. J. Liu, and F. Ding, “Identification methods for time-delay systems based on the redundant rules,” *Signal Processing*, vol. 137, pp. 192–198, 2017.
- [25] Y. J. Lu, B. Huang, and S. Khatibisepehr, “A variational Bayesian approach to robust identification of switched ARX models,” *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 3195–3208, 2016.
- [26] X. P. Liu and X. Q. Yang, “Identification of nonlinear state-space systems with skewed measurement noises,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, pp. 1–9, 2022.
- [27] J. Chen, B. Huang, F. Ding, and Y. Gu, “Variational Bayesian approach for ARX systems with missing observations and varying time-delays,” *Automatica*, vol. 94, pp. 194–204, 2018.
- [28] D. Q. Wang, S. Zhang, M. Gan, and J. Qiu, “A novel EM identification method for Hammerstein systems with missing output data,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2500–2508, 2020.
- [29] R. Kalaba and L. Tesfatsion, “Time-varying linear regression via flexible least squares,” *Computers & Mathematics with Applications*, vol. 17, no. 8–9, pp. 1215–1245, 1989.
- [30] R. Kalaba and L. Tesfatsion, “Flexible least squares for approximately linear systems,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 5, pp. 978–989, 1990.
- [31] G. Montana, K. Triantafyllopoulos, and T. Tsagaris, “Flexible least squares for temporal data mining and statistical arbitrage,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 2819–2830, 2009.
- [32] J. Chen, L. X. Guo, M. F. Hu, M. Gan, and Q. M. Zhu, “Augmented flexible least squares algorithm for time-varying parameter systems,” *International Journal of Robust and Nonlinear Control*, vol. 32, no. 6, pp. 3549–3567, 2021.
- [33] T. S. Chen, M. S. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto, “System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques,” *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 2933–2945, 2014.
- [34] N. Sammaknejad, Y. Zhao, and B. Huang, “A review of the expectation maximization algorithm in data-driven process identification,” *Journal of Process Control*, vol. 73, pp. 123–136, 2019.