Recursive Identification for Hammerstein Systems with State-space Model

CHEN Xi[1] FANG Hai-Tao[1]

Abstract In this paper, the subspace methods for multi-input multi-output (MIMO) Hammerstein systems in state-space form are considered, for the reason that relation of input and output in MIMO case is appropriate to indicate in this form. In open-loop case, a subspace identification method is given for Hammerstein systems, in which the nonlinear static function can be represented as a linear combination of finite basis functions, and the recursive version of this method is also given. We show that, in mild conditions, the method given in this paper is consistent in some sense. A numerical example is given to show the performance of this method.

Key words Subspace method, Hammerstein system, recursive method, principal component analysis (PCA), multi-input multi-output (MIMO), stochastic approximation

DOI 10.3724/SP.J.1004.2010.01460

Generally, system identification is carried out to figure out the structure of the dynamic system by making use of the system input and output corrupted by additive noise. Most of the existing methods focus on the parametric model-based techniques to find a well-suited model described by a finite number of data. Examples of such methods include the prediction error method (PEM) to estimate the free parameters based on an optimization problem, the instrumental variable (IV) method to find a good principal random variable uncorrelated with the noise and sufficiently correlated with the states and inputs, and the maximum likelihood method and the subspace identification method (SIM)[1−4], and so on, respectively, in various model structure.

The subspace identification method is designed especially for the time-invariant systems, the beginning of which could be traced back to the 1960s. It is naturally applicable to the multi-input multi-output (MIMO) systems in the form of state space model[5], which is convenient for estimation, filtering, prediction, and control. Much attention has been attracted from researchers because of its efficient, convenient, and stable computational characteristics. As the subspace method has been developed in several decades, it has attained maturity, especially against linear state space model.

As for the nonlinear system, the state space model description is applied widely in various fields. General form of the dynamic system could be described as follows:

$$x_{t+1} = f(t, x_t, u_t, \theta_t) + v_{t+1} \quad (1)$$

$$y_t = h(t, x_t, u_t, \theta_t) + w_t \quad (2)$$

in which x_t, u_t, and y_t are the system state, input, and output respectively, and θ_t denotes the parameter of the system, v_t and w_t represent the process noise and the observed noise.

In the nonlinear system field, there is a simple and typical kind of system, which consists of a linear dynamic part cascaded with a static nonlinear part at the input or output. The system is called Hammerstein system, if the nonlinearity is prior to the linear part. And the same is called Wiener system, if the cascade order alters. The Hammerstein system is used for modeling biological process, chemical process, and in signal processing applications, which can be shown in the state-space form as

$$x_{k+1} = Ax_k + Bg(u_k) + v_{k+1} \quad (3)$$

$$y_{k+1} = Cx_{k+1} + Du_{k+1} + w_{k+1} \quad (4)$$

Nonlinear identification[6−7] is an increasingly active research topic, which includes parametric and nonparametric methods. As is widely known, no omnipotent identification method exists. Extensions to nonlinear systems though desirable are much harder in general. There are plenty of methods for Hammerstein system identification[8−15], but they are mainly designed against the single-input and single-output (SISO) model[16−17]. These days, researches on the subspace methods have been extended from linear systems to nonlinear systems. And some results of the nonlinear system identification methods based on subspace identification have appeared in succession[18−21]. However, the nonlinear functions of these existing subspace identification methods are in particular forms and little theoretical analysis about convergence, consistency[8−9, 22−23], rate, and so on, has been given.

This paper is targeted to give a subspace identification method of the Hammerstein system in open-loop case, based on the state space representation. And the recursive algorithm[24−26] of the method is also shown. In order to obtain faster computation, we have utilized the averaging technique in view of the principal analysis based on stochastic approximation[27−28] to identify the extended observable matrix. In addition, some asymptotic quality is analyzed under mild condition. And a simulated example is given to show that the average technique is indeed a favorable alternative to speed up the computation process.

The rest of this paper is organized as follows: Section 1 sets the model and its assumptions, which hold throughout the whole paper. Section 2 gives the recursive version of identification, and its consistency analysis is also followed. Finally, a simulated example is given in Section 3 to evaluate the given method.

Notations.

E : Expectation;

$\mathbf{R}$: Real field;

$\mathbf{R}^n$: Real space in n dimensions;

A^{T} : Transpose of A;

$\lambda_{\max}(A)$: Largest eigenvalue of A;

$\mathrm{span}(A)$: Space spanned by column of A;

$\mathrm{Adj}(A)$: Adjoint matrix of A.

Manuscript received May 18, 2010; accepted July 20, 2010
Supported by National Natural Science Foundation of China (60221301, 60774020) and the National Laboratory of Space Intelligent Control

1. Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P. R. China

1 Identification

1.1 Model and assumptions

We consider the discrete MIMO Hammerstein system in state space form as follows:

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + B\boldsymbol{h}(\boldsymbol{u}_k) \quad (5)$$

$$\boldsymbol{y}_{k+1} = C\boldsymbol{x}_{k+1} + \boldsymbol{e}_{k+1} \quad (6)$$

where $\boldsymbol{x}_k \in \mathbf{R}^n$, $\boldsymbol{u}_k \in \mathbf{R}^m$, $\boldsymbol{y}_k \in \mathbf{R}^l$, $\boldsymbol{e}_k \in \mathbf{R}^l$ are the system state, input, measured output, and observed noise, respectively. Here, $A \in \mathbf{R}^{n\times n}, B \in \mathbf{R}^{n\times m}, C \in \mathbf{R}^{l\times n}$, and $\boldsymbol{h} : \mathbf{R}^m \to \mathbf{R}^m$ is the unknown static nonlinear function. We assume the order of the system has been known beforehand, denoted as n.

For system (5) and (6), we make the following assumptions:

A 1. $(A,\ C)$ is observable and $(A,\ B)$ is controllable;

A 2. The system (5) is asymptotically stable, i.e., $|\lambda_{\max}(A)| < 1$;

A 3. The input $\boldsymbol{u}_k$ is a sequence of independently and identically distributed random variables with m dimensions;

A 4. Let $\boldsymbol{\varphi}_i(\boldsymbol{u}_k)$, $i = 1, \cdots, N$ be N bounded functions of m dimensions, satisfying $\mathrm{E}\{\boldsymbol{\varphi}_i(\boldsymbol{u}_k)\boldsymbol{\varphi}_j(\boldsymbol{u}_k)^{\mathrm{T}}\} = \delta_{ij}$, where

$$\delta_{ij} = \begin{cases} I, & i = j \\ 0, & i \neq j \end{cases}$$

and $\boldsymbol{h}$ could be constructed as the sum of $\boldsymbol{\varphi}_i(\boldsymbol{u}_k)$ with proper coefficients, namely,

$$\boldsymbol{h}(\boldsymbol{u}) = \sum_{i=1}^{N} D_i\boldsymbol{\varphi}_i(\boldsymbol{u})$$

where $D_1 = I$;

A 5. $\{\boldsymbol{e}_k\}$ is a sequence of bounded white noise, independent from $\{\boldsymbol{u}_k\}$ satisfying

$$\mathrm{E}\{\boldsymbol{e}_k\} = \boldsymbol{0},\ \mathrm{E}\{\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{T}}\} = \sigma_e I$$

where $\sigma_e > 0$ and I is the unit matrix.

Remark 1. Since N could be taken large enough, A 4 is reasonably acceptable, if $\boldsymbol{h}$ has a wide range. Any nonlinear function could be approximated as $\boldsymbol{h}(\boldsymbol{u}) = \sum_{i=1}^{N} D_i\boldsymbol{\varphi}_i(\boldsymbol{u})$. This condition makes sure that $\frac{1}{L}\sum_{k=l}^{L} \mathrm{E}\{\boldsymbol{h}(\boldsymbol{u}_k)\boldsymbol{h}(\boldsymbol{u}_{k-l})^{\mathrm{T}}\} = \sum_{i=1}^{N} D_iD_i^{\mathrm{T}} = I + \sum_{i=2}^{N} D_iD_i^{\mathrm{T}} > 0$, which can be considered as some kind of persistently excited restriction. Furthermore, the limitation on $D_1 = I$ could be broadened to nonsingular matrix. In this case, we could multiply it to B from the right, changing B into BD, and $\boldsymbol{h}(\boldsymbol{u})$ could still be approximated by a sum of the basic functions with proper coefficients and I as the first coefficient.

Remark 2. The problem of approximation error of the nonlinear function $\boldsymbol{h}$ on the result of the estimation is $O(\epsilon)$, which can be obtained from the theory of robustness of stochastic approximation-based principal component analysis (SABPCA) algorithm later introduced. Certain conditions have been fulfilled in order to guarantee its estimation accuracy. And further knowledge on this topic can be found in [29].

1.2 Denotation

The purpose of identification is to estimate the system matrices A, B, C, and the coefficients $\{D_i, i = 2, \cdots, N\}$ of the unknown nonlinear function $\boldsymbol{h}$.

In order to identify the extended state space model corresponding to the system (5) and (6), let $D = [I,\ D_2, \cdots, D_N] \in \mathbf{R}^{m\times Nm}$, $\tilde{\boldsymbol{u}}_k = [\boldsymbol{\varphi}_1^{\mathrm{T}}(\boldsymbol{u}_k), \cdots, \boldsymbol{\varphi}_N^{\mathrm{T}}(\boldsymbol{u}_k)]^{\mathrm{T}} \in \mathbf{R}^{Nm}$, and then the state space model can be described as

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + BD\tilde{\boldsymbol{u}}_k \quad (7)$$

$$\boldsymbol{y}_{k+1} = C\boldsymbol{x}_{k+1} + \boldsymbol{e}_{k+1} \quad (8)$$

Let

$$\boldsymbol{y}_f(k) = \begin{bmatrix} \boldsymbol{y}_{k-f+1}^{\mathrm{T}} & \boldsymbol{y}_{k-f+2}^{\mathrm{T}} & \cdots & \boldsymbol{y}_k^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbf{R}^{lf}$$

and we denote $\boldsymbol{e}_f(k)$ and $\boldsymbol{u}_f(k)$ in the same form as $\boldsymbol{y}_f(k)$, the components of which are respectively consisted of $\boldsymbol{e}_{k-f+1}, \cdots, \boldsymbol{e}_k$ and $\boldsymbol{u}_{k-f+1}, \cdots, \tilde{\boldsymbol{u}}_k$ in stead of $\boldsymbol{y}_{k-f+1}, \cdots, \boldsymbol{y}_k$.

Denote

$$Y_k = \begin{bmatrix} \boldsymbol{y}_f(1) & \boldsymbol{y}_f(2) & \cdots & \boldsymbol{y}_f(k) \end{bmatrix} \in \mathbf{R}^{lf\times k}$$

and Y_k is the Hankle matrix of all the output data before the time k. f is an arbitrary integer satisfying $f > n$.

1.3 Identification of O_f

Having denoted all the symbols we need, it is easy to gain the extended state space model by the state space system (7) and (8) as follows:

$$\boldsymbol{y}_f(k) = O_f\boldsymbol{x}_{k-f+1} + \Phi_f\boldsymbol{u}_f(k) + \boldsymbol{e}_f(k) \quad (9)$$

where

$$O_f = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix} \in \mathbf{R}^{lf\times n}$$

is an extended observable matrix of rank n, and

$$\Phi_f = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ CBD & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ CA^{f-2}BD & \cdots & CBD & 0 \end{bmatrix} \in \mathbf{R}^{lf\times Nmf} \quad (10)$$

is a block lower triangular Toeplitz matrix.

In order to estimate O_f and noticing (9), we intend to remove the input term $\boldsymbol{u}_f(k)$.

Lemma 1. If A 2 $\sim$ A 5 hold, then $\{(\boldsymbol{y}_k^{\mathrm{T}},\ \boldsymbol{u}_k^{\mathrm{T}})^{\mathrm{T}}\}$ is an asymptotically stable and ergodic sequence; and so is $(\boldsymbol{y}_f(k)^{\mathrm{T}}, \boldsymbol{u}_f(k)^{\mathrm{T}})^{\mathrm{T}}$.

From Lemma 1, by the ergodic characteristic of $\boldsymbol{u}_f(k)$ and $\boldsymbol{y}_f(k)$ we could conclude that:

$$\frac{1}{k}Y_kU_k^{\mathrm{T}} \xrightarrow[k\to\infty]{\text{a.s.}} R_{yu} = \mathrm{E}\{\boldsymbol{y}_f(k)\boldsymbol{u}_f^{\mathrm{T}}(k)\}$$

$$\frac{1}{k}U_kU_k^{\mathrm{T}} \xrightarrow[k\to\infty]{\text{a.s.}} R_{uu} = \mathrm{E}\{\boldsymbol{u}_f(k)\boldsymbol{u}_f^{\mathrm{T}}(k)\}$$

A 3 and A 4 assure $R_{uu} > 0$, so $U_kU_k^{\mathrm{T}}/k$ is reversible for large enough k, and $(U_kU_k^{\mathrm{T}}/k)^{-1} \xrightarrow[k\to\infty]{\text{a.s.}} R_{uu}^{-1}$.

Considering

$$\boldsymbol{z}_f(k+1) = \boldsymbol{y}_f(k+1) - R_{yu}R_{uu}^{-1}\boldsymbol{u}_f(k+1)$$

$$R = \mathrm{E}\{\boldsymbol{z}_f(k+1)\boldsymbol{z}_f^{\mathrm{T}}(k+1)\} = R_{yy} - R_{yu}R_{uu}^{-1}R_{uy} \quad (11)$$

and

$$r_{xx} = \mathrm{E}\{\boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}}\}$$
$$r_{xu} = \mathrm{E}\{\boldsymbol{x}_{k-f+1} \boldsymbol{u}_f^{\mathrm{T}}(k)\}$$

by the definition (11) of R, the extended state space model (9) and the assumptions, we have

$$R = O_f(r_{xx} - r_{xu}R_{uu}^{-1}r_{ux})O_f^{\mathrm{T}} + \sigma_e I \tag{12}$$

Since the extended observable matrix O_f is full column rank (rank $O_f = n$), by (12), if

$$r_{xx} - r_{xu}R_{uu}^{-1}r_{ux} > 0 \tag{13}$$

it is concluded that n of the eigenvalues of R are larger than σ_e, and the rest of $lf - n$ ones exactly equal to σ_e. From the singular value decomposition, we know that under condition (13), R can be decomposed as follows:

$$R = [R_s \ R_n] \begin{bmatrix} \Lambda_s & 0 \\ 0 & \Lambda_n \end{bmatrix} \begin{bmatrix} R_s^{\mathrm{T}} \\ R_n^{\mathrm{T}} \end{bmatrix} \tag{14}$$

where $\Lambda_n = \sigma_e I$. R_s includes the n principal eigenvectors corresponding to the n principal eigenvalues and R_n contains $lf - n$ eigenvectors corresponding to the eigenvalue σ_e. $[R_s \ R_n]$ is orthogonal matrix.

Comparing (12) and (14), the space spanned by the columns of O_f is identical to the space spanned by the columns of R_s, namely,

$$\mathrm{span}(O_f) = \mathrm{span}(R_s)$$

Therefore, the problem turns to be condition under which the inequality (13) will hold. And the following proposition will give an answer to this problem.

Proposition 1. If A 1 $\sim$ A 5 hold, then

$$r_{xx} - r_{xu}R_{uu}^{-1}r_{ux} > 0$$

Proof. The proof is given in Appendix A. □

Proposition 1 assures $\mathrm{span}(O_f) = \mathrm{span}(R_s)$. Therefore, in order to estimate O_f, we just have to estimate the first n eigenvectors of R in (14).

2 Recursive algorithm

In this section, we will give a version of recursive subspace identification method of Hammerstein system. First of all, a recursive estimation of O_f will be shown.

2.1 Estimation of O_f

In the previous section, we have shown that O_f can be estimated from the n principal eigenvectors $\boldsymbol{v}^{(i)}$ of R. And the principal eigenvectors can be calculated from the stochastic approximation-based principal component analysis algorithm. Here, to accelerate the computation speed, we will utilize the averaging technique of stochastic approximation[16].

Denote $\boldsymbol{v}^{(i)}$ as a unit eigenvector of R corresponding to the eigenvalue $\lambda^{(i)}$, $i = 1, \cdots, lf$. In recursively estimating $\boldsymbol{v}^{(i)}$ and $\lambda^{(i)}$, $i = 1, \cdots, lf$, we will compute R, defined in (11). So, we perform the data projection as analyzed in Subsection 1.3:

$$\hat{\boldsymbol{z}}_f(k+1) = \boldsymbol{y}_f(k+1) - R_{yu}(k+1)R_{uu}^{-1}(k+1)\boldsymbol{u}_f(k+1) \tag{15}$$

where

$$R_{yu}(k+1) = R_{yu}(k) - \frac{1}{k+1}\left[R_{yu}(k) - \boldsymbol{y}_f(k+1)\boldsymbol{u}_f{}^{\mathrm{T}}(k+1)\right]$$
$$R_{uu}^{-1}(k+1) = N(k+1) = N(k) + \frac{1}{k}\left[N(k) - (k+1)\frac{N(k)\boldsymbol{u}_f(k)\boldsymbol{u}_f(k)N(k)}{k + \boldsymbol{u}_f{}^{\mathrm{T}}(k)N(k)\boldsymbol{u}_f(k)}\right]$$

The initial value $R_{yu}(0) = 0$ and $N(0)$ could be taken as arbitrary positive matrix.

After obtaining the data projection, we will calculate the component eigenvectors of $\hat{\boldsymbol{z}}_f(k+1)\hat{\boldsymbol{z}}_f^{\mathrm{T}}(k+1)$ by principal component analysis (PCA) based on the stochastic approximation.

1)
$$\tilde{\boldsymbol{v}}_{k+1}^{(1)} = \boldsymbol{v}_k^{(1)} + a_k \hat{\boldsymbol{z}}_f(k+1)\hat{\boldsymbol{z}}_f^{\mathrm{T}}(k+1)\boldsymbol{v}_k^{(1)}$$
$$\boldsymbol{v}_{k+1}^{(1)} = \frac{\tilde{\boldsymbol{v}}_{k+1}^{(1)}}{\|\tilde{\boldsymbol{v}}_{k+1}^{(1)}\|}, \qquad \text{if } \|\tilde{\boldsymbol{v}}_{k+1}^{(1)}\| \neq 0$$

If $\|\tilde{\boldsymbol{v}}_{k+1}^{(1)}\| = 0$, then reset $\boldsymbol{v}_k^{(1)}$ to be the unit vector. $\boldsymbol{v}_k^{(1)}$ should be taken as the one of the k-th R's unit characteristic vector estimation.

2) Assume that $\boldsymbol{v}_k^{(i)}$, $i = 1, \cdots, j$ has been defined. Denote $V_k^{(j)} = [\boldsymbol{v}_k^{(1)}, P_k^{(1)}\boldsymbol{v}_k^{(2)}, \cdots, P_k^{(j-1)}\boldsymbol{v}_k^{(j)}]$ be a matrix of $lf \times j$, where

$$P_k^{(i)} = I - V_k^{(i)}V_k^{(i)\dagger}, \quad i = 1, 2, \cdots, j-1$$

$V_k^{(j)\dagger}$ here represents the Moore-Penrose pseudoinverse of $V_k^{(j)}$. When k is large enough, $V_k^{(j)}$ is a full rank matrix of order $lf \times j$ $(lf > j)$

$$V_k^{(j)\dagger} = (V_k^{(j)\mathrm{T}}V_k^{(j)})^{-1}V_k^{(j)\mathrm{T}}$$

Furthermore, we could estimate the j-th unit characteristic vector:

$$\tilde{\boldsymbol{v}}_{k+1}^{(j+1)} = P_k^{(j)}\boldsymbol{v}_k^{(j+1)} + a_k P_k^{(j)}\hat{\boldsymbol{z}}_f(k+1)\hat{\boldsymbol{z}}_f^{\mathrm{T}}(k+1)P_k^{(j)}\boldsymbol{v}_k^{(j+1)}$$
$$\boldsymbol{v}_{k+1}^{(j+1)} = \frac{\tilde{\boldsymbol{v}}_{k+1}^{(j+1)}}{\|\tilde{\boldsymbol{v}}_{k+1}^{(j+1)}\|}, \qquad \text{if } \|\tilde{\boldsymbol{v}}_{k+1}^{(j+1)}\| \geq \epsilon$$

where $0 < \epsilon < 1/4$, $j = 1, 2, \cdots, lf - 1$.

If $\|\tilde{\boldsymbol{v}}_{k+1}^{(j+1)}\| < \epsilon$, then reset $\boldsymbol{v}_k^{(j+1)}$, so that $\|\boldsymbol{v}_k^{(j+1)}\| = 1$ and $\|P_k^{(j)}\boldsymbol{v}_k^{(j+1)}\| = 1$.

Then, we will give the algorithm as follows.

Step 0. Let $a_k = 1/k$ and $0 < M_k \longrightarrow \infty$ and take $\{\boldsymbol{v}_0^{(j)}, \ j = 1, \cdots, lf\}$ as arbitrary unit orthogonal group, $\lambda_0^{(j)} = 0$, $j = 1, \cdots, lf$, $\sigma_0 = 0$, $k = 0$.

Step 1. Data projection to calculate $\boldsymbol{z}_f(k)$ using (15):

$$\hat{\boldsymbol{z}_f}(k+1) = \boldsymbol{y}_f(k+1) - R_{yu}(k+1)N(k+1)\boldsymbol{u}_f(k+1)$$

Step 2. Use SABPCA algorithm to calculate the principal characteristic vectors $\boldsymbol{v}_{k+1}^{(i)}, i = 1, \cdots, lf$ as shown above.

Step 3. Estimate the eigen-roots

$$\lambda_{k+1}^{(j)} = [\lambda_k^{(j)} - a_k(\lambda_k^{(j)} - \|\hat{\boldsymbol{z}}_f^{\mathrm{T}}(k+1)\boldsymbol{v}_k^{(j)}\|^2)]\times \mathbf{1}_{[|\lambda_k^{(j)} - a_k(\lambda_k^{(j)} - \|\hat{\boldsymbol{z}}_f^{\mathrm{T}}(k+1)\boldsymbol{v}_k^{(j)}\|^2)| \leq M_{\sigma_k}]}, \quad j = 1, 2, \cdots, lf$$

where

$$\sigma_{k+1} = \sigma_k + \mathbf{1}_{[|\lambda_k^{(j)} - a_k(\lambda_k^{(j)} - \|\hat{\boldsymbol{z}}_f^{\mathrm{T}}(k+1)\boldsymbol{v}_k^{(j)}\|^2)| > M\sigma_k]}$$

Step 4. if $k \le N_0$, let $k = k+1$ and go to Step 1, else use fast average technique to proceed the step:

$$\boldsymbol{v}_{k+1}^{(j)} = \frac{1}{k+1}(\boldsymbol{v}_k^{(j)} + k \times \boldsymbol{v}_{k+1}^{(j)})$$
$$\lambda_{k+1}^{(j)} = \frac{1}{k+1}(\lambda_k^{(j)} + k \times \lambda_k^{(j)})$$

and perform normalization to $\boldsymbol{v}_k^{(j)}$, $j = 1, \cdots, lf$;

Step 5. Rearrange $\{\lambda_{k+1}^{(j)}, \boldsymbol{v}_{k+1}^{(j)}\}$, so that $\lambda_{k+1}^{(1)} \ge \lambda_{k+1}^{(2)} \ge \cdots \ge \lambda_{k+1}^{(lf)}$, and let

$$\hat{O}_f(k+1) = [\boldsymbol{v}_{k+1}^{(1)}\ \boldsymbol{v}_{k+1}^{(2)}\ \cdots\ \boldsymbol{v}_{k+1}^{(n)}] \tag{16}$$

Step 6. Take $k = k+1$ and return to Step 1.

Remark 1. From the analysis and algorithm, $\lambda^{(j)}$ is constant when $j > n$, and $\lambda^{(n)} > \lambda^{(n+1)}$, so the order n of the system matrix can be estimated from $\lambda_k^{(j)}$ if the order is unknown.

Remark 2. Averaging technique will make the noise converge in equilibrium, however, the state deviation in the beginning is too large to use averaging technique, since it will eliminate the deviation too slowly. Therefore, in the algorithm above, we do not utilize fast average technique in the first N_0 steps. And N_0 will be taken large enough so that the averaging technique plays its role more efficiently.

The following theorem indicates O_f to be strongly consistent up to some extent.

Theorem 1. If A 1 $\sim$ A 5 hold, then the estimation $\hat{O}_f(k)$ obtained as in (16), converges to O_f almost everywhere, i.e., there exists a nonsingular matrix T of order $n \times n$, so that

$$\hat{O}_f(k) \xrightarrow[k\to\infty]{\text{a.s.}} O_f T$$

Proof. The proof of Theorem 1 is also attached in Appendix B. □

2.2 Recursive estimation of A, B, C, and D

2.2.1 Estimation of A and C

Let $\hat{A}(k)$, $\hat{B}(k)$, $\hat{C}(k)$, and $\hat{D}(k)$ be the k-th estimation of A, B, C, and D, respectively. Once O_f has been known, the system matrix C can be directly observed from O_f.

As for the estimation of A, we will utilize the transfer invariance of O_f:

$$O_f(l+1 : lf, :) = O_f(1 : lf - l, :)A$$

Denote

$$\underline{O}(k) = \hat{O}_f(k)((l+1) : lf, :)$$
$$\overline{O}(k) = \hat{O}_f(k)(1 : (f-1)l, :)$$

utilizing the transfer invariance of O_f, then

$$\underline{O}(k+1) = \overline{O}(k) \times A + \epsilon_k$$

where ϵ_k is the noise in the process of estimation.

Using recursive least square estimation method, $\hat{A}(k)$ could be recursively calculated as follows:

$$\begin{cases} \hat{A}(k+1) = \hat{A}(k) + \gamma(k)G(k)\overline{O}(k)[\underline{O}(k+1) - \overline{O}(k) \times \hat{A}(k)] \\ G(k) = G(k) - \gamma(k)G(k)\overline{O}(k)\overline{O}(k)^{\mathrm{T}}G(K) \end{cases} \tag{17}$$

where

$$G(k) = \left(\sum_{i=0}^{k-1} \overline{O}(i)\overline{O}(i)^{\mathrm{T}}\right)^{-1}$$
$$\gamma(k) = (1 + \overline{O}(k)^{\mathrm{T}}G(k)\overline{O}(k))^{-1}$$

2.2.2 Estimation of B and D

Let $O_f^{\perp} \in \mathbf{R}^{lf \times (lf-n)}$ be denoted as the arbitrary orthogonal complement of O_f, i.e. $O_f^{\mathrm{T}} O_f^{\perp} = 0$ and $O_f^{\perp}$ is full rank in column.

To estimate matrix B and D, we pre-multiply (9) by $(O_f^{\perp})^{\mathrm{T}}$, then

$$(O_f^{\perp})^{\mathrm{T}}\boldsymbol{y}_f(k) = (O_f^{\perp})^{\mathrm{T}}\Phi_f \boldsymbol{u}_f(k) + (O_f^{\perp})^{\mathrm{T}}\boldsymbol{e}_f(k) \tag{18}$$

in which $\boldsymbol{u}_f(k)$ and $\boldsymbol{e}_f(k)$ are uncorrelated and Φ_f is a linear matrix of BD.

Similarly, we will also use the recursive least square estimation method to calculate BD recursively. Let $\boldsymbol{z}_{\mathbf{f}} = [\boldsymbol{z}_1^{\mathrm{T}}\ \boldsymbol{z}_2^{\mathrm{T}}\ \cdots\ \boldsymbol{z}_f^{\mathrm{T}}]^{\mathrm{T}}$ be a vector as the form of $\boldsymbol{u}_f(k)$, namely, $\boldsymbol{z}_j \in \mathbf{R}^m$, $j = 1, 2, \cdots, f$.

Let

$$\boldsymbol{\theta} = \mathbf{Vec}(BD)$$

where BD is the product of B and D.

Denote

$$M(C, A, \boldsymbol{z}_f) = \begin{bmatrix} 0 \\ \boldsymbol{z}_1^{\mathrm{T}} \otimes C \\ \boldsymbol{z}_1^{\mathrm{T}} \otimes CA + \boldsymbol{z}_2^{\mathrm{T}} \otimes C \\ \vdots \\ \sum_{i=1}^{f-1} \boldsymbol{z}_i^{\mathrm{T}} \otimes CA^{f-1-i} \end{bmatrix}$$

where $\otimes$ is Kronecker product. For any vector $\boldsymbol{u}_f(k)$ as defined in Section 1,

$$\Phi_f \boldsymbol{u}_f(k) = M(C, A, \boldsymbol{u}_f(k))\boldsymbol{\theta} + \boldsymbol{\omega}_k \tag{19}$$

where $\boldsymbol{\omega}_k$ is the estimation noise.

Denote

$$\widetilde{O}_f(k) = \begin{bmatrix} \hat{C}(k) \\ \hat{C}(k)\hat{A}(k) \\ \vdots \\ \hat{C}(k)\hat{A}^{f-1}(k) \end{bmatrix}$$
$$\widehat{O_f^{\perp}}(k) = (\widetilde{O}_f(k))^{\perp}$$
$$M(k) = M(\hat{C}(k), \hat{A}(k), u_f(k))$$
$$\psi(k) = \left[(\widehat{O_f^{\perp}}(k))^{\mathrm{T}} M(k)\right]^{\mathrm{T}} = M^{\mathrm{T}}(k)\widehat{O_f^{\perp}}(k)$$

where $\widehat{O_f^{\perp}}(k)$ is the orthogonal complement matrix of k-th extended observable estimated matrix $\widetilde{O}_f(k)$ generated by $\hat{C}(k)$ and $\hat{A}(k)$, i.e., satisfying $\widetilde{O}_f(k)^{\mathrm{T}}\widehat{O_f^{\perp}}(k) = 0$, and $\widehat{O_f^{\perp}}(k) \in \mathbf{R}^{lf \times lf-n}$ is full column rank. $M(k)$ is the k-th matrix substituting $\hat{C}(k)$ and $\hat{A}(k)$ into $M(C, A, \boldsymbol{u}_f(k))$ instead of C and A.

Next, we put (19) into (18) and gain the following result:

$$(O_f^{\perp})^{\mathrm{T}}\boldsymbol{y}_f(k+1) = (O_f^{\perp})^{\mathrm{T}}M(C,A,\boldsymbol{u}_f(k))\boldsymbol{\theta} + (O_f^{\perp})^{\mathrm{T}}\boldsymbol{e}_f(k) \tag{20}$$

Then $\boldsymbol{\theta}$, namely, $\mathbf{Vec}(BD)$ can be estimated using recursive least square method as follows:

$$\begin{cases} \psi(k) = (O_f^{\perp})^{\mathrm{T}}M(C,A,\boldsymbol{u}_f(k)) \\ \hat{\boldsymbol{\theta}}(k+1) = \boldsymbol{\theta}(k) + \kappa(k)P(k)\psi(k)[(O_f^{\perp})^{\mathrm{T}}\boldsymbol{y}_f(k+1) - \\ \qquad\qquad (O_f^{\perp})^{\mathrm{T}}M(C,A,\boldsymbol{u}_f(k))\boldsymbol{\theta}] \\ P(k+1) = P(k) + \kappa(k)P(k)\psi(k)\psi(k)^{\mathrm{T}}P(k) \end{cases}$$

where

$$P(k) = \left(\sum_{i=0}^{k-1}\psi(k)\psi(k)^{\mathrm{T}}\right)^{-1}$$
$$\kappa(k) = (1+\psi(k)^{\mathrm{T}}G(k)\psi(k))^{-1}$$

The initial values $P(0), G(0)$ could be chosen as I with proper dimensions, and $A(0), \boldsymbol{\theta}(0)$ could be taken as any deterministic matrix of compatible dimensions.

Once $\mathbf{Vec}(BD)$ is obtained, since $D_1 = I$ by the definition of D we could observe B directly from BD, i.e., $B = BD(1:n, 1:m)$. And then by the property of the Kronecker product $ABC = C^{\mathrm{T}} \otimes B^{\mathrm{T}} \otimes A^{\mathrm{T}}$ with arbitrary suitable dimension, respectively, D could be estimated uniquely under proper restriction on B.

3 Simulation results

In this section, we will give a simulated example for the case of MIMO Hammerstein system to evaluate the proposed algorithm in the previous section:

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + Bh(\boldsymbol{u}_k) \tag{21}$$
$$\boldsymbol{y}_{k+1} = C\boldsymbol{x}_{k+1} + \boldsymbol{e}_{k+1} \tag{22}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -0.5625 & -1.5 & 0 \\ 0 & 0 & 0.85 \end{bmatrix}$$
$$B = \begin{bmatrix} 1.5 & 0.5 \\ 0 & 1.5 \\ 0.1 & 0.2 \end{bmatrix}$$
$$C = \begin{bmatrix} 0 & 0.3 & 0 \\ 0.5 & 0 & 2 \end{bmatrix}$$
$$D_1 = I_{2\times 2}$$
$$D_2 = I_{2\times 2}$$

The static nonlinearity is chosen as

$$\boldsymbol{h}(u) = \begin{bmatrix} 1+\sin\ u(1) \\ \cos\left(\dfrac{u(2)}{2}\right) + 2\sin\left(\dfrac{u(2)}{2}\right) \end{bmatrix}$$

and the inputs $u(i), i = 1, 2$ are taken uniformly on the interval $[-1, 1]$. The observed noise is also chosen as a sequence of Gaussian white noise with standard deviation $\sigma = 0.1 \times I$, where data set has 3 000 input and output measurements respectively. Let $N_0 = 1\,000$.

To evaluate the gap between the spanned subspace by the column vector of $\hat{O}_f$ and O_f, we introduce the following variant:

$$\Theta(\mathrm{span}(O_f), \mathrm{span}(\hat{O}_f(k))) = \|\Pi_{O_f} - \Pi_{\hat{O}_f(k)}\|$$

where $\prod_X$ denotes the orthogonal projection matrix on the spanned space of column of X. By the definition of Θ, it can be considered as an angle between the two subspaces. The smaller the angle is, the closer the two subspaces will be.

As shown in Fig. 1, the gap between the true value O_f and its estimation $\hat{O}_f(k)$ has approached to zero as soon as it turns to the fast averaging computation on the 1 000-th step.

Fig. 2 has given a comparison in the end of steps from 1 500 to 2 000 between the normal PCA method in the winding line and method using fast averaging technique underneath in the straight line. We can observe that the line in the fast averaging method goes to zero more smoothly and quickly than the normal method. It is indicated that fast averaging technique is a more satisfying choice in identifying the extended observable matrix.

The true eigenvalues of A are 0.85, −0.75, and −0.75. Fig. 3 has shown that the estimation of A's eigenvalues converge to its true value gradually as the computation process goes to the end.

Fig. 4 has indicated that the least square estimation of B and D could be rather accurate.

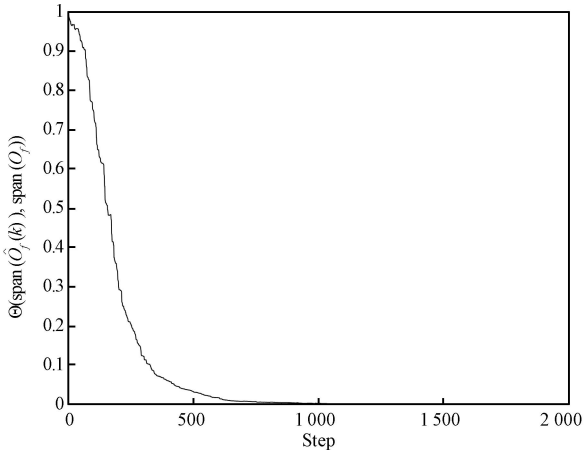

Fig. 1 Gap between $\hat{O}_f(k)$ and O_f

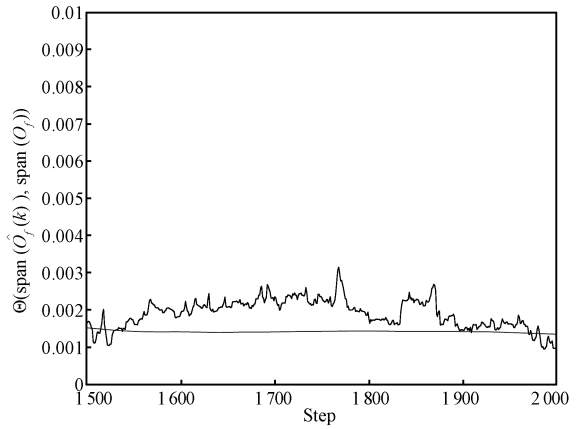

Fig. 2 Comparison between fast averaging method and normal method

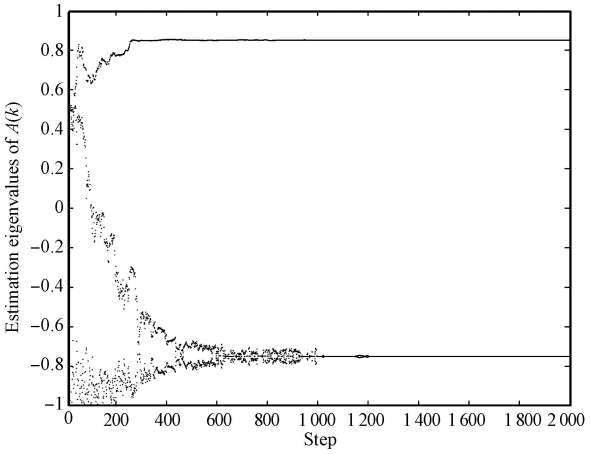

Fig. 3 Estimation eigenvalues of $A(k)$

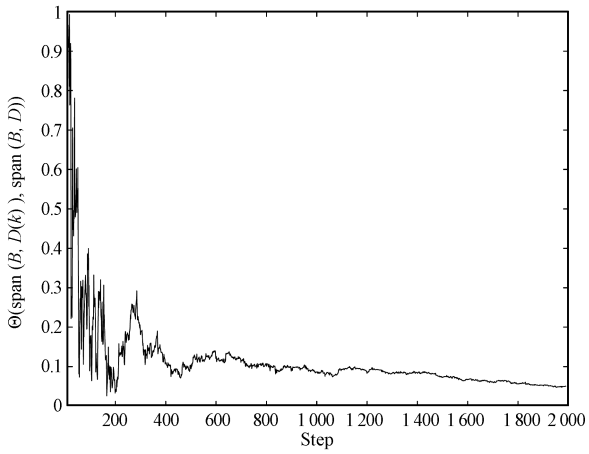

Fig. 4 Gap between the two subspaces spanned by B, D and their true values

4 Conclusion

In this paper, a method for estimating the parameters of MIMO Hammerstein system in state-space form applying subspace identification method was proposed. Extension of Hammerstein system to the MIMO case has brought in many difficulties. The existing methods mainly deal with SISO case, among which the relation between the input and output is indicated in the transfer function form. However, the transfer function in the MIMO form is difficult to show and the relationship between the parameters and the inputs is troublesome to indicate. So, the state space model is applied. In particular, the presented recursive algorithm based on PCA in view of the stochastic approximation methods has been given. This is an important extension of the previous works, which have been restricted to the area of SISO Hammerstein system. With a purpose to speed up the computation rate, we have applied the PCA algorithm based on the stochastic approximation using fast averaging technique.

Although the results obtained here are satisfying, to extend the subspace identification methods on the general nonlinear system, further research need to be performed. To recursively identify this general nonlinear system is also our future focus of research.

Appendix A The proof of Proposition 1

To show $r_{xx} - r_{xu}R_{uu}^{-1}r_{ux} > 0$, it is equivalent to prove that

$$\mathrm{E}\left\{\begin{bmatrix}\boldsymbol{x}_k\\ \boldsymbol{u}_f(k)\end{bmatrix} \times \begin{bmatrix}\boldsymbol{x}_k\\ \boldsymbol{u}_f(k)\end{bmatrix}^{\mathrm{T}}\right\} = \begin{bmatrix}r_{xx} & r_{xu}\\ r_{uu} & r_{uu}\end{bmatrix} > 0 \qquad \text{(A1)}$$

By the state space model (7) and (8). Let $\boldsymbol{x}_k = \frac{F^{\boldsymbol{u}}(q^{-1})}{a(q^{-1})}\tilde{\boldsymbol{u}}_k$, where q^{-1} is the backward shift operator, i.e., $q^{-1}\tilde{\boldsymbol{u}}_k = \tilde{\boldsymbol{u}}_{k-1}$,

$$\begin{aligned} F^{\boldsymbol{u}}(q^{-1}) &= \frac{\mathrm{Adj}(qI-A)BD}{q^n} = \\ &\quad F_1^{\boldsymbol{u}}q^{-1} + F_2^{\boldsymbol{u}}q^{-2} + \cdots + F_n^{\boldsymbol{u}}q^{-n} \\ a(q^{-1}) &= \frac{\det(qI-A)}{q^n} = \\ &\quad a_0 + a_1q^{-1} + a_2q^{-2} + \cdots + a_nq^{-n} \end{aligned}$$

where $F_i^{\boldsymbol{u}} \in \mathbf{R}^{n\times Nm}$, $a_0 = 1$.

$$\begin{bmatrix}\boldsymbol{x}_k\\ \boldsymbol{u}_f(k)\end{bmatrix} = \frac{\Xi}{a(q^{-1})}\begin{bmatrix}\tilde{\boldsymbol{u}}_{k-n}\\ \vdots\\ \tilde{\boldsymbol{u}}_{k+f-1}\end{bmatrix}$$

where

$$\Xi = \begin{bmatrix} F_n^{\boldsymbol{u}} & F_{n-1}^{\boldsymbol{u}} & \cdots & F_1^{\boldsymbol{u}} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & a_nI & a_{n-1}I & \cdots & a_0I \\ 0 & 0 & \cdots & a_nI & a_{n-1}I & a_{n-2}I & \cdots & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots & \vdots & \iddots & \vdots \\ 0 & a_nI & \cdots & a_2I & a_1I & a_0I & \cdots & 0 \\ a_nI & a_{n-1}I & \cdots & a_1I & a_0I & 0 & \cdots & 0 \end{bmatrix}$$

If Ξ has full row rank under A 1 $\sim$ A 4, and

$$\mathrm{E}\left\{\frac{1}{a(q^{-1})}\begin{bmatrix}\tilde{\boldsymbol{u}}_{k-n}\\ \vdots\\ \tilde{\boldsymbol{u}}_{k+f-1}\end{bmatrix} \times \frac{1}{a(q^{-1})}\begin{bmatrix}\tilde{\boldsymbol{u}}_{k-n}\\ \vdots\\ \tilde{\boldsymbol{u}}_{k+f-1}\end{bmatrix}^{\mathrm{T}}\right\} \qquad \text{(A2)}$$

is positive definite, then (A1) is true. Note that

$$\begin{bmatrix}\tilde{\boldsymbol{u}}_{k-n}\\ \vdots\\ \tilde{\boldsymbol{u}}_{k+f-1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{u}_{k-n})\\ \vdots\\ \boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{u}_{k+f-1})\end{bmatrix}$$

is persistently excited as stated in the remark after the assumptions A 1 $\sim$ A 4.

At last, we will show that Ξ is full row rank. Equivalently, we can show that $[F_n^{\boldsymbol{u}} \ \cdots \ F_1^{\boldsymbol{u}}]$ is full rank. Define

$$\begin{aligned} F^{\boldsymbol{u}}(q^{-1}) &= \frac{\mathrm{Adj}(qI-A)BD}{q^n} = \\ &\quad a(q^{-1})(qI-A)^{-1}BD = \\ &\quad (a_0 + a_1q^{-1} + \cdots + a_nq^{-n})q^{-1} \\ &\quad (I + q^{-1}A + q^{-2}A^2 + \cdots)BD \end{aligned}$$

Comparing the coefficients of the two sides of this equation,

we have

$$[F_n^{\boldsymbol{u}} \ \cdots \ F_1^{\boldsymbol{u}}] = [A^{n-1}BD \ A^{n-2}BD \ \cdots \ BD]\times \begin{bmatrix} a_0 I & 0 & \cdots & 0 \\ a_1 I & a_0 I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n-1} I & a_{n-2} I & \cdots & a_0 I \end{bmatrix} \tag{A3}$$

Since $[A^{n-1}BD, A^{n-2}BD, \cdots, BD]$ is full rank in column by the controllability of (A; B) in assumption A 1, $D = [I, D_2, \cdots, D_N]$, and the right matrix consisting of $a_0, a_1, \cdots, a_N$ of the right equality of (A3) is nonsingular, $[F_n^{\boldsymbol{u}}, \cdots, F_1^{\boldsymbol{u}}]$ is also full rank. □

Appendix B The proof of Theorem 1

From the analysis in Subsection 1.3, we know $\text{span}\{O_f\} = \text{span}\{R_s\}$, by Theorem 1, under the assumptions A 1 ∼ A 5, then we will show $\hat{O}_f(k) \xrightarrow[k\to\infty]{\text{a.s.}} O_f T$, under the assumptions A 1 ∼ A 5.

Since $\{\boldsymbol{u}_k\}$ is an independent identically distributed process, $\{\boldsymbol{\varphi}_i, i = 1, \cdots, N\}$ are bounded, $\{\tilde{\boldsymbol{u}}_k\}$ is a bounded process, the observed noise $\{\boldsymbol{e}_k\}$ is bounded, Y_k is bounded. By the definition of $\hat{\boldsymbol{z}}_f(k)$, then $R_k = \text{E}\{\hat{\boldsymbol{z}}_f(k)\hat{\boldsymbol{z}}_f(k)^{\text{T}}\}$ is also bounded. In view of Chapter 3.3 of [15], we just have to show the noise condition:

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{i=k}^{m(k,T)} a_k(\hat{\boldsymbol{z}}_f(i+1)\hat{\boldsymbol{z}}_f^{\text{T}}(i+1) - R)\right\| = 0, \quad \forall T_k \in (0, T]$$

where $m(k,T) = \max\{m, \sum_{i=k}^{m} \le T\}$.

Without loss of generality, we assume $a_k = 1/k$. By the result of Lemma 1,

$$\frac{1}{k}(U_k U_k^{\text{T}})^{-1} \xrightarrow[k\to\infty]{\text{a.s.}} R_{uu}^{-1}$$

$$\frac{1}{k}(Y_k U_k^{\text{T}}) \xrightarrow[k\to\infty]{\text{a.s.}} R_{yu}$$

Then, we use the second equation to subtract the first one and it turns out to be:

$$\Delta(k) = R_{yu}R_{uu}^{-1} - (\frac{1}{k}Y_k U_k^{\text{T}})\frac{1}{k}(U_k U_k^{\text{T}})^{-1} \xrightarrow[k\to\infty]{\text{a.s.}} 0$$

$$\begin{aligned}
&\hat{\boldsymbol{z}}_f(i+1)\hat{\boldsymbol{z}}_f^{\text{T}}(i+1) - R = \\
&[\boldsymbol{y}_f(i+1) - Y_{i+1}U_{i+1}^{\text{T}}(U_k U_k^{\text{T}})^{-1}\boldsymbol{u_f}(i+1)]\times \\
&[\boldsymbol{y}_f(i+1) - Y_{i+1}U_{i+1}^{\text{T}}(U_k U_k^{\text{T}})^{-1}\boldsymbol{u}_f(i+1)]^{\text{T}} - R = \\
&[\boldsymbol{y}_f(i+1) - Y_{i+1}U_{i+1}^{\text{T}}(U_k U_k^{\text{T}})^{-1}\boldsymbol{u}_f(i+1) + \Delta(k)\boldsymbol{u}_f(i+1)]\times \\
&[\boldsymbol{y}_f(i+1) - Y_{i+1}U_{i+1}^{\text{T}}(U_k U_k^{\text{T}})^{-1}\boldsymbol{u}_f(i+1)]^{\text{T}} + \\
&\Delta(k)\boldsymbol{u}_f(i+1)^{\text{T}} - R = \\
&[\hat{\boldsymbol{z}}_f(i+1) + \Delta(i+1)\boldsymbol{u}_f(i+1)]\times \\
&[\hat{\boldsymbol{z}}_f(i+1) + \Delta(i+1)\boldsymbol{u}_f(i+1)]^{\text{T}} - R = \\
&\hat{\boldsymbol{z}}_f(i+1)\hat{\boldsymbol{z}}_f(i+1)^{\text{T}} - R + \hat{\boldsymbol{z}}_f(i+1)\boldsymbol{u}_f^{\text{T}}(i+1)\Delta^{\text{T}}(i+1) + \\
&\Delta(i+1)\boldsymbol{u}_f(i+1)\boldsymbol{z}_f^{\text{T}}(i+1) + \Delta^{\text{T}}(i+1)\boldsymbol{u}_f^{\text{T}}(i+1)\Delta^{\text{T}}(i+1)
\end{aligned}$$

As we have known that $\Delta(i+1) \xrightarrow[i\to\infty]{\text{a.s.}} 0$ and $[\boldsymbol{u}_f(i+1), \hat{\boldsymbol{z}}_f(i+1)]$ is bounded, then

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\Big\|\sum_{i=k}^{m(k,T)}\frac{1}{i}[\hat{\boldsymbol{z}}_f(i+1)\boldsymbol{u}_f^{\text{T}}(i+1)\Delta^{\text{T}}(i+1) + \Delta(i+1)\boldsymbol{u}_f(i+1)\boldsymbol{z}_f^{\text{T}}(i+1) + \Delta^{\text{T}}(i+1)\boldsymbol{u}_f^{\text{T}}(i+1)\Delta^{\text{T}}(i+1)]\Big\| = 0$$

a.s. $\forall\ T_k \in (0, T]$

and we just have to show that

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\parallel\sum_{i=k}^{m(k,T)}\frac{1}{i}[\hat{\boldsymbol{z}}_f(i+1)\hat{\boldsymbol{z}}_f^{\text{T}}(i+1) - R]\parallel = 0, \quad \text{a.s.} \quad \forall T_k \in (0, T]$$

From Lemma 1, we have come to know that $\{(\boldsymbol{y}_k^{\text{T}}, \boldsymbol{u}_k^{\text{T}})^{\text{T}}\}$ has the stable ergodic characteristic, so does $\{\boldsymbol{z}_f(k)\boldsymbol{z}_f(k)^{\text{T}}\}$. By (11), $\text{E}\{\tilde{R}(k)\} = \text{E}\{\hat{\boldsymbol{z}}_f(k)\hat{\boldsymbol{z}}_f(k)^{\text{T}}\} = R$ and $\frac{1}{k}\sum_{i=1}^{k}\tilde{R}(k) - R \xrightarrow[i\to\infty]{\text{a.s.}} 0$. Let $\eta_k = \sum_{i=1}^{k}(\tilde{R}(k) - R)$, i.e., $\eta_k \xrightarrow[i\to\infty]{\text{a.s.}} 0$,

$$\begin{aligned}
&\sum_{i=k}^{m(k,T)}\frac{1}{i}(\hat{\boldsymbol{z}}_f(i+1)\hat{\boldsymbol{z}}_f^{\text{T}}(i+1) - R) = \\
&\sum_{i=k}^{m(k,T)}\frac{1}{i}(\eta_{i+1} - \eta_i) = \\
&\frac{1}{m(k,T)}\eta_{m(k,T)+1} - \frac{1}{k}\eta_k + \sum_{i=k+1}^{m(k,T)}\eta_k(\frac{1}{i+1} - \frac{1}{i}) = \\
&\frac{1}{m(k,T)}\eta_{m(k,T)+1} - \frac{1}{k}\eta_k + \sum_{i=k+1}^{m(k,T)}\eta_k(\frac{1}{(i+1)i}) \\
&\xrightarrow[i\to\infty]{\text{a.s.}} 0
\end{aligned}$$

So, (4) holds naturally.

□

References

1 Bauer D. Asympotic properties of subspace estimators. *Automatica*, 2005, **41**(3): 359−376

2 Katayama T. *Subspace Methods for System Identification*. London: Springer-Verlag, 2005

3 van Overschee P, de Moor B. *Subspace Identification for Linear Systems*. Boston: Kluwer Academic Publisher, 1996

4 Viberg M. Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, 1995, **31**(12): 1835−1851

5 Lovera M, Gustafsson T, Verhaegen M. Recursive subspace identification of linear and non-linear Wiener state-space models. *Automatica*, 2000, **36**(11): 1639−1650

6 Bendat J S. *Nonlinear System Analysis and Identification from Random Data*. New York: Wiley, 1999

7 Ljung L. *System Identification: Theory for the User (Second Edition)*. New York: Prentice Hall, 1999

8 Chen H F. Pathwise convergence of recursive identification algorithms for Hammerstein systems. *IEEE Transactions on Automatic Control*, 2004, **49**(10): 1641−1649

9 Chen H F. Strong consistency of recursive identification for Hammerstein systems with discontinuous piecewise-linear memoryless block. *IEEE Transactions on Automatic Control*, 2005, **50**(10): 1612−1617

10 Greblicki W. Recursive identification of Wiener system. *International Journal of Applied Mathematics and Computer Science*, 2001, **11**(4): 977−991

11 Hu X L, Chen H F. Strong consistence of recursive identification for Wiener systems. *Automatica*, 2005, **41**(11): 1905−1916

12 Hunter I W, Korenberg M J. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, 1986, **55**(2-3): 135−144

13 Zhao Y L, Wang L Y, Yin G G, Zhang J F. Identification of Wiener systems with binary-valued output observations. *Automatica*, 2007, **43**(10): 1752−1765

14 Ding F, Chen T W. Identification of Hammerstein nonlinear ARMAX systems. *Automatica*, 2005, **41**(9): 1479−1489

15 Boko L, Mercere G, Lecoeuche S, Lovera M. Recursive subspace identification of Hammerstein models based on least squares support vector machines. *IET Control Theory and Applications*, 2009, **3**(9): 1209−1216

16 Palanthandalam-Madapusi H J, Ridley A J, Bernstein D S. Identification and prediction of ionospheric dynamics using a Hammerstein-Wiener model with radial basis function. In: Proceedings of the American Control Conference. Portland, USA: IEEE, 2005. 5052−5057

17 Zhao W X. Parametric identification of Hammerstein systems with consistency results using stochastic inputs. *IEEE Transactions on Automatic Control*, 2010, **55**(2): 474−480

18 Felici F, van Wingerden J W, Verhaegen M. Subspace identification of MIMO LPV systems using a periodic scheduling sequence. *Automatica*, 2007, **43**(10): 1684−1697

19 Palanthandalam-Madapusi H J, Lacy S, Hoagg J B, Bernstein D S. Subspace-based identification for linear and nonlinear systems. In: Proceedings of the American Control Conference. Portland, USA: IEEE, 2005. 2320−2334

20 Palanthandalam-Madapusi H J, Bernstein D S, Ridley A J. Space weather forecasting. *IEEE Control Systems Magazine*, 2007, **27**(5): 109−123

21 Westwick D, Verhaegen M. Identifying MIMO wiener systems using subspace model identification methods. *Signal Processing*, 1996, **52**(2): 235−258

22 Chen H F, Guo L. *Identification and Stochastic Adaptive Control*. Boston: Birkhauser, 1991

23 Chiuso A, Picci G. Consistency analysis of some closed-loop subspace identification methods. *Automatica*, 2005, **41**(3): 377−391

24 Jiang Y P, Fang H T. Recursive identification of MIMO Wiener systems with general inputs. In: Proceedings of the 48th IEEE Conference on Decision and Control and the 28th Chinese Control Conference. Shanghai, China: IEEE, 2009. 3626−3631

25 Jiang Y P, Fang H T. Recursive subspace identification based on principal component analysis. In: Proceedings of the 25th Chinese Control Conference. Harbin, China: IEEE, 2006. 444−449

26 Jiang Y P, Fang H T. Recursive subspace identification algorithm for closed-loop stochastic system. In: Proceedings of the 15th IFAC Symposium on System Identification. Saint-Malo, France: Elsevier, 2009. 116−121

27 Chen H F. *Stochastic Approximation and Its Applications*. Berlin: Springer, 2002

28 Polyak B T, Juditsky A B. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*,1992, **30**(4): 838−858

29 Chen H F, Guo L, Gao A J. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Processes and Their Applications*, 1988, **27**: 217−231

CHEN Xi Ph. D. candidate at the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences. She received her bachelor degree in applied mathematics from Beijing Normal University in 2008. Her research interest covers system modeling and identification. Corresponding author of this paper.
E-mail: chenxi@amss.ac.cn

FANG Hai-Tao Professor at the Key Laboratory of Systems and Control, Institute of Systems Science, Chinese Academy of Sciences. He received his B. S. degree in probability and statistics from Peking University in 1990, his M. S. degree in applied mathematics from Tsinghua University in 1993, and his Ph. D. degree from Peking University in 1996, respectively. From 1996 to 1998, he was a postdoctoral fellow at the Institute of Systems Science and joined the institute as an assistant professor in 1998. During 1998, 1999, and 2001, he was with Hong Kong University of Science and Technology as a research associate. His research interest covers optimization, identification, and control in stochastic systems and their applications in signal processing and communication.
E-mail: htfang@iss.ac.cn