Journal of Computational and Applied Mathematics 345 (2019) 135–145

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Model recovery for Hammerstein systems using the hierarchical orthogonal matching pursuit method☆

Dongqing Wang [a,c,*], Yaru Yan [a], Yanjun Liu [b], Junhang Ding [a,c]

[a] College of Automation and Electrical Engineering, Qingdao University, Qingdao 266071, China
[b] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China
[c] Collaborative Innovation Center for Eco-Textiles of Shandong Province, Qingdao 266071, China

ARTICLE INFO

Article history:
Received 13 October 2016
Received in revised form 21 May 2017

Keywords:
Hierarchical identification principle
Hammerstein system
Orthogonal matching pursuit (OMP)
Compressed sensing (CS)
Parameter estimation

ABSTRACT

Most papers concentrate on the parameter identification of Hammerstein systems with known orders. This paper, motivated by the recent developments in sparse approximations, investigates the combined parameter and order determination of Hammerstein systems. The methodology used relies on greedy schemes—the orthogonal matching pursuit (OMP) algorithm in the compressive sensor (CS) theory. In particular, the first step recasts a bilinear Hammerstein system into two fictitious pseudo-regressive sub-systems which respectively contain the parameters of the nonlinear part or the parameters of the linear part by the hierarchical identification principle. The second step adopts a hierarchical orthogonal matching pursuit (H-OMP) selection procedure to interactively select the parameters and orders of the two sub-systems under the frame of the compressive sensor. Finally, the proposed algorithm is tested on a simulation example.

1. Introduction

Nowadays system modeling and identification are very important for nonlinear systems and complex systems [1–5]. The traditional identification methods for Hammerstein nonlinear systems are popular for the last two decades, including the over-parametrization model based methods [6,7], the iterative/recursive identification methods [8–11], the key term separation principle based identification methods [12–15], the hierarchical identification methods [16–18], and the maximum likelihood estimation methods [19–21], etc. Recently, Li used the Levenberg–Marquardt optimization method to estimate parameters for a Hammerstein output error system [22]; Chen adopted a particle swarm optimization algorithm to estimate unknown parameters for a Hammerstein system [23]. But the above conventional identification methods need process thousands of input and output data, this costs a lot of time in data sampling and parameter estimating.

In the past decade, the compressive sensing method based on the sparsity principle has aroused much attention in signal processing field [24–26], it has an advantage of saving computation in recovering parameters of a system, by collecting only a few data. Generally, the compressive sensing method can be described as to reconstruct a S-sparse vector $\boldsymbol{\Theta} \in \mathbb{R}^n$ from linear measurements in $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ and observations in $\boldsymbol{Y} \in \mathbb{R}^m$ under the form: $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{\Theta}$. By definition, the number of sparsity, observations, components of the unknown signal are decreased, i.e., $S < m < n$. Accurate reconstruction can be achieved by two types of approaches: the greedy algorithms like the thresholding algorithms [27–29] or the orthogonal matching

☆ This work was supported by the National Natural Science Foundation of China under grant 61573205, 61403217, and the Shandong Provincial Natural Science Foundation of China under grant ZR2015FM017.

* Corresponding author.
E-mail addresses: dqwang64@qdu.edu.cn (D.Q. Wang), yryan92@126.com (Y. Yan), yjl@jiangnan.edu.cn (Y. Liu), dingjunhang@163.com (J. Ding).

https://doi.org/10.1016/j.cam.2018.06.016
0377-0427/

pursuit (OMP) algorithms [30–32], and the basis pursuit (BP) algorithms [33–35]. The greedy OMP algorithm imposes an L_0-norm on the sparse vector $\boldsymbol{\Theta}$, works iteratively by picking up the support columns (atoms) in the measurement matrix in a greedy fashion, and has simple and rapid advantages over the BP method which solves its convex relaxation by the standard L_1-norm technique.

In this paper, a hierarchical orthogonal matching pursuit (H-OMP) algorithm is investigated to simultaneously select the orders and parameters of a Hammerstein system. The explorations lie in three main aspects:

- The first is to recast the system into two fictitious pseudo-regressive sub-systems: one contains the parameters of the nonlinear part and the other contains the parameters of the linear part by the hierarchical identification principle.
- The second is to adopt the H-OMP method to interactively select the parameters and orders of these two sub-systems under the framework of the CS theory.
- Comparing with the existing hierarchical least squares (H-LS) method, the proposed H-OMP algorithm is not necessary to collect a lot of data and invest a lot of power on the parameter identification.

The rest of the paper is organized as follows. Section 2 demonstrates the problem formulation of a Hammerstein system. Section 3 presents the H-OMP identification algorithm to interactively select the orders and parameters of these two sub-systems. Section 4 derives the existing hierarchical least squares (H-LS) algorithm for comparison. Section 5 provides a numerical example for the proposed algorithm. Finally, the concluding remarks are involved in Section 6.

2. The problem formulation

The input nonlinear and output linear functions of a Hammerstein system are expressed as

$$x(t) = f[u(t)] = \sum_{k=1}^{n_c} c_k f_k[u(t)], \tag{1}$$

$$y(t) = B(z)x(t) + v(t), \tag{2}$$

where $B(z) = b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b}$, $u(t)$ and $y(t)$ are the system input and output, $x(t)$ is an internal variable, $v(t)$ is stochastic white noise with zero mean, the input nonlinearity f is modeled as a linear combination of basis functions f_i.

Assume that the orders n_b and n_c are unknown, we set sufficient length l as orders of the nonlinear/linear functions ($l > n_b, l > n_c$). From the $y - x$ relationship in (2), we get

$$y(t) = b_1 x(t-1) + b_2 x(t-2) + \cdots + b_{n_b} x(t-n_b) + \cdots + b_l x(t-l) + v(t). \tag{3}$$

Substituting Eq. (1) into $x(t-i)$ into Eq. (3) gives:

$$y(t) = b_1 \sum_{k=1}^{n_c,\ldots,l} c_k f_k[u(t-1)] + \cdots + b_{n_b} \sum_{k=1}^{n_c,\ldots,l} c_k f_k[u(t-n_b)] + \cdots + b_l \sum_{k=1}^{n_c,\ldots,l} c_k f_k[u(t-l)] + v(t).$$

Define the information matrix and the parameter vectors:

$$\boldsymbol{F}(t) = \begin{bmatrix} f_1[u(t-1)], \ldots, f_{n_c}[u(t-1)], \ldots, f_l[u(t-1)] \\ \vdots \\ f_1[u(t-n_b)], \ldots, f_{n_c}[u(t-n_b)], \ldots, f_l[u(t-n_b)] \\ \vdots \\ f_1[u(t-l)], \ldots, f_{n_c}[u(t-l)], \ldots, f_l[u(t-l)] \end{bmatrix} \in \mathbb{R}^{l\times l},$$

$$\boldsymbol{b} = [b_1, b_2, \ldots, b_{n_c}, \underbrace{0, \ldots, 0}_{l-n_b}]^{\mathrm{T}} \in \mathbb{R}^l, \ \boldsymbol{c} = [c_1, c_2, \ldots, c_{n_c}, \underbrace{0, \ldots, 0}_{l-n_c}]^{\mathrm{T}} \in \mathbb{R}^l,$$

then we have

$$y(t) = \boldsymbol{b}^{\mathrm{T}} \boldsymbol{F}(t) \boldsymbol{c} + v(t). \tag{4}$$

It seems difficult to recast the above obtained bilinear system under the CS framework. In this letter, based on the hierarchical identification principle, we recast the bilinear system into two simple pseudo-regressive sub-systems as follows.

By multiplying $\boldsymbol{F}(t)$ with $\boldsymbol{c}$ ($\boldsymbol{F}(t)\boldsymbol{c} := \boldsymbol{F}_c(t) \in \mathbb{R}^{l\times 1}$), the bilinear model (4) is transformed into a pseudo-regressive **sub-system I** about the parameter vector $\boldsymbol{b}$ of the linear part,

$$\textbf{Sub-system I}: \ y(t) = \boldsymbol{F}_c^{\mathrm{T}}(t)\boldsymbol{b} + v(t). \tag{5}$$

Similarly, by multiplying $\boldsymbol{b}^{\mathrm{T}}$ with $\boldsymbol{F}(t)$ ($\boldsymbol{b}^{\mathrm{T}}\boldsymbol{F}(t) := \boldsymbol{F}_b(t) \in \mathbb{R}^{1\times l}$), the bilinear model (4) is transformed into a pseudo-regressive **sub-system II** about the parameter vector $\boldsymbol{c}$ of the nonlinear part,

$$\textbf{Sub-system II}: \ y(t) = \boldsymbol{F}_b(t)\boldsymbol{c} + v(t). \tag{6}$$

Substituting m sampled data into (5) and (6) gives

$$\begin{cases} y(1) = \boldsymbol{F}_c^{\mathrm{T}}(1)\boldsymbol{b} + v(1), \\ y(2) = \boldsymbol{F}_c^{\mathrm{T}}(2)\boldsymbol{b} + v(2), \\ \quad \vdots \\ y(m) = \boldsymbol{F}_c^{\mathrm{T}}(m)\boldsymbol{b} + v(m) \end{cases}$$

and

$$\begin{cases} y(1) = \boldsymbol{F}_b(1)\boldsymbol{c} + v(1), \\ y(2) = \boldsymbol{F}_b(2)\boldsymbol{c} + v(2), \\ \quad \vdots \\ y(m) = \boldsymbol{F}_b(m)\boldsymbol{c} + v(m). \end{cases}$$

Define

$$\boldsymbol{Y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(m) \end{bmatrix} \in \mathbb{R}^m, \quad \boldsymbol{V} = \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(m) \end{bmatrix} \in \mathbb{R}^m,$$

$$\boldsymbol{\Xi}_1 = \begin{bmatrix} \boldsymbol{F}_c^{\mathrm{T}}(1) \\ \boldsymbol{F}_c^{\mathrm{T}}(2) \\ \vdots \\ \boldsymbol{F}_c^{\mathrm{T}}(m) \end{bmatrix} \in \mathbb{R}^{m\times l}, \quad \boldsymbol{\Xi}_2 := \begin{bmatrix} \boldsymbol{F}_b(1) \\ \boldsymbol{F}_b(2) \\ \vdots \\ \boldsymbol{F}_b(m) \end{bmatrix} \in \mathbb{R}^{m\times l}.$$

Then two sets of equations can be rewritten as two pseudo-regressive matrix equations, respectively,

$$\boldsymbol{Y} = \boldsymbol{\Xi}_1\boldsymbol{b} + \boldsymbol{V}, \tag{7}$$

$$\boldsymbol{Y} = \boldsymbol{\Xi}_2\boldsymbol{c} + \boldsymbol{V}. \tag{8}$$

Eqs. (7) and (8) are two pseudo-regressive sub-systems meeting the CS framework. According to the least squares principle, if there are enough measurements, i.e., $m \gg l$, we can interactively get the least squares estimates of $\boldsymbol{b}$ and $\boldsymbol{c}$,

$$\hat{\boldsymbol{b}}_{\mathrm{LS}} = (\boldsymbol{\Xi}_1^{\mathrm{T}}\boldsymbol{\Xi}_1)^{-1}\boldsymbol{\Xi}_1^{\mathrm{T}}\boldsymbol{Y},$$
$$\hat{\boldsymbol{c}}_{\mathrm{LS}} = (\boldsymbol{\Xi}_2^{\mathrm{T}}\boldsymbol{\Xi}_2)^{-1}\boldsymbol{\Xi}_2^{\mathrm{T}}\boldsymbol{Y}.$$

But the above traditional least squares will take a large computation and a lot of time to get the estimates $\hat{\boldsymbol{b}}_{\mathrm{LS}}$ and $\hat{\boldsymbol{c}}_{\mathrm{LS}}$. Because the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ contain only a few non-zeros, but many zeros. According to the CS theory, the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ can be viewed as a sparse signal. Let $S = \|\boldsymbol{b}\|_0 + \|\boldsymbol{c}\|_0 = n_b + n_c$ denote the numbers of non-zero entries in $\boldsymbol{b}$ and $\boldsymbol{c}$, then the identification problem can be described as an orthogonal matching pursuit (OMP):

$$\hat{\boldsymbol{b}} = \mathrm{argmin}\|\boldsymbol{b}\|_0, \quad \text{s.t.} \quad \|y(t) - \boldsymbol{F}_c^{\mathrm{T}}(t)\boldsymbol{b}\|_2 < \varepsilon,$$
$$\hat{\boldsymbol{c}} = \mathrm{argmin}\|\boldsymbol{c}\|_0, \quad \text{s.t.} \quad \|y(t) - \boldsymbol{F}_b(t)\boldsymbol{c}\|_2 < \varepsilon,$$

where $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{c}}$ are the estimates of $\boldsymbol{b}$ and $\boldsymbol{c}$ and ε ($\varepsilon > 0$) is the error tolerance.

3. The H-OMP identification method

3.1. The CS background and the OMP method

3.1.1. The CS background

Suppose $\boldsymbol{X}$ meet the following equation,

$$\boldsymbol{Y} = \boldsymbol{\phi}\boldsymbol{X},$$

where the output $\boldsymbol{Y} \in \mathbb{R}^M$ and $\boldsymbol{\phi} \in \mathbb{R}^{M\times N}$ ($M < N$), naturally $\boldsymbol{X} \in \mathbb{R}^N$ is usually not sparse, solving these underdetermined equations to get unknown $\boldsymbol{X}$ is impossible. CS technique represents $\boldsymbol{X} = \boldsymbol{\Omega\Theta}$ with a transform $\boldsymbol{\Omega} \in \mathbb{R}^{N\times N}$ of a K-sparse signal $\boldsymbol{\Theta} \in \mathbb{R}^N$. Then the CS equation or the CS framework is expressed as

$$\boldsymbol{Y} = \boldsymbol{\Phi\Theta},$$

with the measurement matrix $\boldsymbol{\Phi} = \boldsymbol{\phi\Omega} \in \mathbb{R}^{M\times N}$, and the K-sparse signal $\boldsymbol{\Theta} \in \mathbb{R}^N$ ($K \le M < N$).

In CS theory, accurately recovering a K-sparse vector $\boldsymbol{\Theta} \in \mathbb{R}^N$ under the form $\boldsymbol{Y} = \boldsymbol{\Phi\Theta}$ should satisfy the following two requirements.

Theorem 1. *Refer to [36], let $K \leq M < N$ and every set of $2K$ columns of the matrix $\Phi \in \mathbb{R}^{M\times N}$ be linearly independent. Then a K-sparse vector $\Theta \in \mathbb{R}^N$ can be reconstructed uniquely from $\boldsymbol{Y} = \Phi\Theta$. This unique solution is in fact the sparsest solution, i.e.,*

$$\hat{\Theta} = \text{argmin}\|\Theta\|_0, \quad \text{s.t.} \quad \|\boldsymbol{Y} = \Phi\Theta\|_2 < \varepsilon.$$

Theorem 2. *Refer to [37], an $M \times N$ matrix $\Phi \in \mathbb{R}^{M\times N}$ obey the K-restricted isometry property (K-RIP) condition with constant δ_K. More precisely, let $\Lambda \subset \{1, 2, \ldots, K\}$ and Φ_Λ be the submatrix of Φ consisting of the columns indexed by $\boldsymbol{\Lambda}$, the local isometry constant δ_K is the smallest number satisfying $(1-\delta_K)\|\Theta_\Lambda\|_2^2 \leq \|\Phi_\Lambda\Theta_\Lambda\|_2^2 \leq (1+\delta_K)\|\Theta_\Lambda\|_2^2$.*

3.1.2. The OMP method

In the CS framework $\boldsymbol{Y} = \Phi\Theta$, suppose $\Phi = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]$ ($\boldsymbol{\phi}_i$ is called an atom of Φ), $\Theta_k = [\theta_1, \theta_2, \ldots, \theta_N]^{\mathrm{T}}$, $k = 1, 2, \ldots$ be the iterative number and λ_k be the index of the solution support at the kth iteration, Λ_k is a set composed of λ_i, $i = 1, 2, \ldots, k$, r_k denotes the residual at the kth iteration, Φ_{Λ_k} is the sub-measurement matrix composed of the k columns of Φ indexed by Λ_k, $\hat{\Theta}_k$ is the estimated parameter vector at the kth iteration.

The orthogonal matching pursuit (OMP) method is a kind of greedy algorithms, it works iteratively by picking the support atoms in a greedy fashion. In each step it finds the atom with highest absolute inner product with the residual and adds it to the already found support atoms. Then it calculates a new approximant by projecting the signal on the linear span of the already found support atoms and a new residual by subtracting the approximant from the signal.

This section narrates the derivation of the OMP recovery algorithm. Define a cost function at the kth iteration

$$J(\theta_i) := \|\boldsymbol{r}_{k-1} - \boldsymbol{\phi}_i\theta_i\|^2, \ i = 1, 2, \ldots, N. \tag{9}$$

Remark 1. Minimizing $J(\theta_i)$ with respect to θ_i means that the derivation of $J(\theta_i)$ with respect to θ_i equals 0, that is,

$$\frac{\partial J(\theta_i)}{\partial\theta_i} = -2\boldsymbol{\phi}_i^{\mathrm{T}}[\boldsymbol{r}_{k-1} - \boldsymbol{\phi}_i\theta_i] = 0,$$

and we get

$$\theta_i = \frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2}. \tag{10}$$

Substitute the above θ_i into Eq. (9) to get a minimized $J(\theta_i)$ as

$$\begin{aligned}
J(\theta_i)_{min} &= \left\|\boldsymbol{\phi}_i\frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2} - \boldsymbol{r}_{k-1}\right\|^2 \\
&= \left(\frac{\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2} - \boldsymbol{r}_{k-1}\right)^{\mathrm{T}}\left(\frac{\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2} - \boldsymbol{r}_{k-1}\right) \\
&= \left(\frac{\boldsymbol{r}_{k-1}^{\mathrm{T}}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}}{\|\boldsymbol{\phi}_i\|^2} - \boldsymbol{r}_{k-1}^{\mathrm{T}}\right)\left(\frac{\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2} - \boldsymbol{r}_{k-1}\right) \\
&= \frac{\boldsymbol{r}_{k-1}^{\mathrm{T}}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2\|\boldsymbol{\phi}_i\|^2} - \frac{\boldsymbol{r}_{k-1}^{\mathrm{T}}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2} - \frac{\boldsymbol{r}_{k-1}^{\mathrm{T}}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2} + \boldsymbol{r}_{k-1}^{\mathrm{T}}\boldsymbol{r}_{k-1} \\
&= \|\boldsymbol{r}_{k-1}\|^2 - \frac{(\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1})^2}{\|\boldsymbol{\phi}_i\|^2}.
\end{aligned} \tag{11}$$

Remark 2. The result $\left[J(\theta_i)_{min} = \|\boldsymbol{r}_{k-1}\|^2 - \frac{(\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1})^2}{\|\boldsymbol{\phi}_i\|^2}\right]$ says that minimizing $J(\theta_i)$ is equivalent to the quest for the largest inner product between the residual $\boldsymbol{r}_{k-1}$ and the normalized column vector $\boldsymbol{\phi}_i$ of Φ.

In summary, the steps of the OMP algorithm iterate from $k = 1$ as:

Step 1. Find the index of the largest inner product,

$$\lambda_k = arg\max_{i=1,2,\ldots,N}|\langle\boldsymbol{r}_{k-1}, \frac{\boldsymbol{\phi}_i}{\|\boldsymbol{\phi}_i\|}\rangle|;$$

Step 2. Update the supporting index set,

$$\Lambda_k = \Lambda_{k-1} \cup \lambda_k;$$

Step 3. Update the supporting sub-measurement matrix,

$$\Phi_{\Lambda_k} = \Phi_{\Lambda_{k-1}} \cup \Phi_{\lambda_k};$$

Step 4. Update the estimated parameter vector,

$$minJ(\Theta_{\Lambda_k}) = \|\boldsymbol{Y} - \Phi_{\Lambda_k}\Theta_{\Lambda_k}\|^2;$$

Step 5. Compute the residual,

$$\boldsymbol{r}_k = \boldsymbol{Y} - \Phi\Theta_k.$$

until the $\boldsymbol{r}_k < \varepsilon$, ε is a small positive constant.

3.2. The H-OMP algorithm

The above mentioned two sub-systems $\boldsymbol{Y} = \Xi_1\boldsymbol{b} + \boldsymbol{V}$ and $\boldsymbol{Y} = \Xi_2\boldsymbol{c} + \boldsymbol{V}$ in (7) and (8) meet the CS framework, due to $\boldsymbol{b}$ and $\boldsymbol{c}$ being S_b-sparse and S_c-sparse vectors, respectively. With the setup of $S_b \leq M < l$ and $S_c \leq M < l$, the task is to reconstruct $\boldsymbol{b}$ from $\boldsymbol{Y}$ and Ξ_1, and $\boldsymbol{c}$ from $\boldsymbol{Y}$ and Ξ_2. A column Ξ_{1i}/Ξ_{2i} of Ξ_1/Ξ_2 is also called an atom as defined by,

$$\Xi_1 = [\Xi_{11}, \Xi_{12}, \ldots, \Xi_{1l}],$$
$$\Xi_2 = [\Xi_{21}, \Xi_{22}, \ldots, \Xi_{2l}].$$

Thus the output observation vector $\boldsymbol{Y}$ is the linear combination of l atoms Ξ_{1i} or Ξ_{2i},

$$\boldsymbol{Y} = \sum_{i=1}^{l} \Xi_{1i} b_i + \boldsymbol{V} \tag{12}$$

$$= \sum_{i=1}^{l} \Xi_{2i} c_i + \boldsymbol{V}. \tag{13}$$

Assume $k = 1, 2, \ldots$ be the iterative number, let us give some notations at the kth iteration: λ_{bk} and λ_{ck} be the column indexes of the solution support in Ξ_1 and Ξ_2; Λ_{bk} and Λ_{ck} are solution sets composed of λ_{bk} and λ_{ck}, respectively; r_{bk} and r_{ck} denote the residuals; $\Xi_{1,\Lambda_{bk}}$ and $\Xi_{2,\Lambda_{ck}}$ are the sub-measurement matrix composed of the n_b and n_c columns of Ξ_1 and Ξ_2 indexed by Λ_{bk} and Λ_{ck}; $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{c}}_k$ are the estimated parameter vectors. The algorithms are initialized as $\boldsymbol{r}_{b0} = \boldsymbol{r}_{c0} = \boldsymbol{r}_0 = \boldsymbol{Y}$ and $\Lambda_{b0} = \Lambda_{c0} = \emptyset$. Define two cost functions,

$$\varepsilon_b(b_i) = \|\boldsymbol{r}_{b,k-1} - \Xi_{1i} b_i\|^2, \ i = 1, 2, \ldots, l, \tag{14}$$

$$\varepsilon_c(c_i) = \|\boldsymbol{r}_{c,k-1} - \Xi_{2i} c_i\|^2. \tag{15}$$

According to Remarks 1 and 2, minimizing $\varepsilon_b(b_i)$ and $\varepsilon_c(c_i)$ with respect to b_i and c_i yields:

$$min\varepsilon_b(b_i) = \|\boldsymbol{r}_{b,k-1}\|^2 - \left(\frac{\Xi_{1i}^{\mathrm{T}}\boldsymbol{r}_{b,k-1}}{\|\Xi_{1i}\|}\right)^2, \tag{16}$$

$$min\varepsilon_c(c_i) = \|\boldsymbol{r}_{c,k-1}\|^2 - \left(\frac{\Xi_{2i}^{\mathrm{T}}\boldsymbol{r}_{c,k-1}}{\|\Xi_{2i}\|}\right)^2. \tag{17}$$

Remark 3. Eqs. (16) and (17) indicate that the pursuit for the smallest error is equivalent to the pursuit for the largest inner product between the residual $\boldsymbol{r}_{b,k-1}$ and the normalized column vectors Ξ_{1i} in Ξ_1, or between $\boldsymbol{r}_{c,k-1}$ and Ξ_{2i} in Ξ_2.

Thus, the kth solution support can be obtained by:

$$\lambda_{bk} = \mathrm{argmax}_{(i=1,2,\ldots,l)} |\langle \boldsymbol{r}_{b,k-1}, \frac{\Xi_{1i}}{\|\Xi_{1i}\|}\rangle|,$$
$$\lambda_{ck} = \mathrm{argmax}_{(i=1,2,\ldots,l)} |\langle \boldsymbol{r}_{c,k-1}, \frac{\Xi_{2i}}{\|\Xi_{2i}\|}\rangle|.$$

Update the support sets Λ_{bk} and Λ_{ck}, and the sub-measurement matrices $\Xi_{1,\Lambda_{bk}}$ and $\Xi_{2,\Lambda_{ck}}$ by

$$\Lambda_{bk} = \Lambda_{b,k-1} \cup \lambda_{bk},$$
$$\Lambda_{ck} = \Lambda_{c,k-1} \cup \lambda_{ck},$$
$$\Xi_{1,\Lambda_{bk}} = \Xi_{1,\Lambda_{b,k-1}} \cup \Xi_{1,\lambda_{bk}},$$
$$\Xi_{2,\Lambda_{ck}} = \Xi_{2,\Lambda_{c,k-1}} \cup \Xi_{2,\lambda_{ck}}.$$

Define two cost functions:

$$J_1(\boldsymbol{b}_{\Lambda_{bk}}) = \|\boldsymbol{Y} - \Xi_{1,\Lambda_{bk}}\boldsymbol{b}_{\Lambda_{bk}}\|^2,$$
$$J_2(\boldsymbol{c}_{\Lambda_{ck}}) = \|\boldsymbol{Y} - \Xi_{2,\Lambda_{ck}}\boldsymbol{c}_{\Lambda_{ck}}\|^2.$$

There exist two difficulties in minimizing the two cost functions to get the estimated parameter vectors $\hat{\boldsymbol{b}}_{\Lambda_{bk}}$ and $\hat{\boldsymbol{c}}_{\Lambda_{ck}}$:

- When minimizing $J_1(\boldsymbol{b}_{\Lambda_{bk}})$ about $\boldsymbol{b}_{\Lambda_{bk}}$, there exist the unknown parameter vectors $\boldsymbol{c}$ in $\boldsymbol{F}_c(t)$ in $\Xi_{1,\Lambda_{bk}}$;
- When minimizing $J_2(\boldsymbol{c}_{\Lambda_{ck}})$ about $\boldsymbol{c}_{\Lambda_{ck}}$, there exist the unknown parameter vectors $\boldsymbol{b}$ in $\boldsymbol{F}_b(t)$ in $\Xi_{2,\Lambda_{ck}}$.

The solutions are to use their estimates $\hat{\boldsymbol{c}}_{k-1}$ and $\hat{\boldsymbol{b}}_k$ at iteration k to replace $\boldsymbol{c}$ and $\boldsymbol{b}$, respectively; and the corresponding estimates of $\boldsymbol{F}_c(t)$ and Ξ_1, and $\boldsymbol{F}_b(t)$ and Ξ_2 at iteration k can be written as,

$$\hat{\boldsymbol{F}}_{c,k}(t) = \boldsymbol{F}(t)\hat{\boldsymbol{c}}_{k-1},\ \hat{\boldsymbol{F}}_{b,k}(t) = \hat{\boldsymbol{b}}_k^{\mathrm{T}}\boldsymbol{F}(t),$$
$$\hat{\Xi}_{1,k} = \begin{bmatrix} \hat{\boldsymbol{F}}_{c,k}^{\mathrm{T}}(1) \\ \hat{\boldsymbol{F}}_{c,k}^{\mathrm{T}}(2) \\ \vdots \\ \hat{\boldsymbol{F}}_{c,k}^{\mathrm{T}}(m) \end{bmatrix} = [\hat{\Xi}_{11,k}, \hat{\Xi}_{12,k}, \ldots, \hat{\Xi}_{1l,k}],\ \hat{\Xi}_{2,k} = \begin{bmatrix} \hat{\boldsymbol{F}}_{b,k}(1) \\ \hat{\boldsymbol{F}}_{b,k}(2) \\ \vdots \\ \hat{\boldsymbol{F}}_{b,k}(m) \end{bmatrix} = [\hat{\Xi}_{21,k}, \hat{\Xi}_{22,k}, \ldots, \hat{\Xi}_{2l,k}].$$

Replacing unknown variables with their estimates, minimizing $J_1(\boldsymbol{b}_{\Lambda_{bk}})$ and $J_2(\boldsymbol{c}_{\Lambda_{ck}})$, and combining the hierarchical identification principle with the least squares identification principle, the least squares estimates $\hat{\boldsymbol{b}}_{\Lambda_{bk}}$ and $\hat{\boldsymbol{c}}_{\Lambda_{ck}}$ at the k step are given,

$$\hat{\boldsymbol{b}}_{\Lambda_{bk}} = (\hat{\Xi}_{1,\Lambda_{bk}}^{\mathrm{T}}\hat{\Xi}_{1,\Lambda_{bk}})^{-1}\hat{\Xi}_{1,\Lambda_{bk}}^{\mathrm{T}}\boldsymbol{Y},$$
$$\hat{\boldsymbol{c}}_{\Lambda_{ck}} = (\hat{\Xi}_{2,\Lambda_{ck}}^{\mathrm{T}}\hat{\Xi}_{2,\Lambda_{ck}})^{-1}\hat{\Xi}_{2,\Lambda_{ck}}^{\mathrm{T}}\boldsymbol{Y}.$$

The residuals at the k step can be computed by:

$$\boldsymbol{r}_{bk} = \boldsymbol{Y} - \hat{\Xi}_{1,\Lambda_{bk}}\hat{\boldsymbol{b}}_{\Lambda_{bk}}, \tag{18}$$
$$\boldsymbol{r}_{ck} = \boldsymbol{Y} - \hat{\Xi}_{2,\Lambda_{ck}}\hat{\boldsymbol{c}}_{\Lambda_{ck}}. \tag{19}$$

Remark 4. At the beginning, the hierarchical orthogonal matching pursuit (H-OMP) algorithm with the estimates $\hat{\boldsymbol{c}}_{k-1}$ and $\hat{\boldsymbol{b}}_k$ in the sub-information matrices $\hat{\Xi}_{1,\Lambda_{bk}}$ and $\hat{\Xi}_{2,\Lambda_{ck}}$ causes an inaccurate support atom selection. With the iteration k increasing, these estimated vectors become more accurate, and the mis-selected support atoms certainly un-meet the threshold requirement, the corresponding elements in the estimated parameter support set will be a small non-zero value.

Because practical systems always along with some noises, we set an appropriate small threshold ε to filter the parameter estimates $\hat{\boldsymbol{b}}_{\Lambda_{bk}}$ and $\hat{\boldsymbol{c}}_{\Lambda_{ck}}$. If $|\hat{b}_{h_b,k}| < \varepsilon$ and/or $|\hat{c}_{h_c,k}| < \varepsilon$ (where $\hat{b}_{h_b,k}$ is the h_bth element of $\hat{\boldsymbol{b}}_{\Lambda_{bk}}$, and $\hat{c}_{h_c,k}$ is the h_cth element of $\hat{\boldsymbol{c}}_{\Lambda_{ck}}$), eliminate $\hat{b}_{h_b,k}$ and/or $\hat{c}_{h_c,k}$ from $\hat{\boldsymbol{b}}_{\Lambda_{bk}}$ and/or $\hat{\boldsymbol{c}}_{\Lambda_{ck}}$, and the corresponding $\hat{\Xi}_{1h_b,k}$ and/or $\hat{\Xi}_{2h_c,k}$ from $\hat{\Xi}_{1,k}$ and/or $\hat{\Xi}_{2,k}$. Then we have

$$\hat{\boldsymbol{b}}_{\Lambda_{bk\varepsilon}} = [\hat{\boldsymbol{b}}_{\Lambda_{bk}}/\hat{b}_{h_b,k} \notin \hat{\boldsymbol{b}}_{\Lambda_{bk}}],$$
$$\hat{\Xi}_{1,\Lambda_{bk\varepsilon}} = [\hat{\Xi}_{1,\Lambda_{bk}}/\hat{\Xi}_{1h_b,k} \notin \hat{\Xi}_{1,\Lambda_{bk}}];$$
$$\hat{\boldsymbol{c}}_{\Lambda_{ck\varepsilon}} = [\hat{\boldsymbol{c}}_{\Lambda_{ck}}/\hat{c}_{h_c,k} \notin \hat{\boldsymbol{c}}_{\Lambda_{ck}}],$$
$$\hat{\Xi}_{2,\Lambda_{ck\varepsilon}} = [\hat{\Xi}_{2,\Lambda_{ck}}/\hat{\Xi}_{h_c,k} \notin \hat{\Xi}_{2,\Lambda_{ck}}].$$

Because $\hat{b}_{h_b,k}$ and $\hat{c}_{h_c,k}$ are h_bth and h_cth elements in $\hat{\boldsymbol{b}}_{\Lambda_k}$ and $\hat{\boldsymbol{c}}_{\Lambda_k}$, and $\hat{\Xi}_{1h_b,k}$ and $\hat{\Xi}_{2h_c,k}$ are h_bth and h_cth columns in $\hat{\Xi}_{1,\Lambda_k}$ and $\hat{\Xi}_{2,\Lambda_k}$; the expressions of eliminating $\hat{b}_{h_b,k}$ from $\hat{\boldsymbol{b}}_{\Lambda_{bk}}$ and $\hat{\Xi}_{1h_b,k}$ from $\hat{\Xi}_{1,\Lambda_{bk}}$ in Matlab are

$$\hat{\boldsymbol{b}}_{\Lambda_{bk\varepsilon}} : \hat{\boldsymbol{b}}_{\Lambda_{bk}}(h_b) = [\,],$$
$$\hat{\Xi}_{1,\Lambda_{ck\varepsilon}} : \hat{\Xi}_{1,\Lambda_{ck}}(h_b) = [\,];$$

the expressions of eliminating $\hat{c}_{h_c,k}$ from $\hat{\boldsymbol{c}}_{\Lambda_{ck}}$ and $\hat{\Xi}_{2h_c,k}$ from $\hat{\Xi}_{2,\Lambda_{ck}}$ in Matlab are

$$\hat{\boldsymbol{c}}_{\Lambda_{k\varepsilon}} : \hat{\boldsymbol{c}}_{\Lambda_k}(h_c) = [\,],$$
$$\hat{\Xi}_{2,\Lambda_{k\varepsilon}} : \hat{\Xi}_{2,\Lambda_k}(h_c) = [\,].$$

Then the residual can be computed by

$$\boldsymbol{r}_{bk\varepsilon} = \boldsymbol{Y} - \hat{\Xi}_{1,\Lambda_{bk\varepsilon}}\hat{\boldsymbol{b}}_{\Lambda_{bk\varepsilon}}, \tag{20}$$
$$\boldsymbol{r}_{ck\varepsilon} = \boldsymbol{Y} - \hat{\Xi}_{2,\Lambda_{ck\varepsilon}}\hat{\boldsymbol{c}}_{\Lambda_{ck\varepsilon}}. \tag{21}$$

The steps of the H-OMP algorithm are summarized as:

The H-OMP for CS Recovery

Measurements: $\Xi_1 = [\Xi_{c1}, \Xi_{c2}, \cdots, \Xi_{cl}],\ \Xi_2 = [\Xi_{b1}, \Xi_{b2}, \cdots, \Xi_{bl}]$;
Output vectors: $\boldsymbol{Y}$;
Sparsity level: $S_b = n_b,\ S_c = n_c$;
Parameter vectors $\boldsymbol{b} = [b_1, b_2, \cdots, b_l]^{\mathrm{T}}, \boldsymbol{c} = [c_1, c_2, \cdots, c_l]^{\mathrm{T}}$;
Initialization: $k = 1, \boldsymbol{r}_{b0} = \boldsymbol{r}_{c0} = \boldsymbol{Y}, \Lambda_{b0} = \Lambda_{c0} = \varnothing$;
$\hat{\boldsymbol{b}}_{\Lambda_0} = \boldsymbol{0}, \hat{\boldsymbol{c}}_{\Lambda_0}$=a constant vector.

Repeat

$$\boldsymbol{F}(t) = \begin{bmatrix} f_1[u(t-1)], \cdots, f_{n_c}[u(t-1)], \cdots, f_l[u(t-1)] \\ \vdots \\ f_1[u(t-n_b)], \cdots, f_{n_c}[u(t-n_b)], \cdots, f_l[u(t-n_b)] \\ \vdots \\ f_1[u(t-l)], \cdots, f_{n_c}[u(t-l)], \cdots, f_l[u(t-l)] \end{bmatrix};$$

Estimating $\boldsymbol{b}$:

$$\Xi_{1,k} = \begin{bmatrix} (\boldsymbol{F}(1)\hat{\boldsymbol{c}}_{k-1})^{\mathrm{T}} \\ (\boldsymbol{F}(2)\hat{\boldsymbol{c}}_{k-1})^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{F}(m)\hat{\boldsymbol{c}}_{k-1})^{\mathrm{T}} \end{bmatrix} = [\Xi_{11,k}, \Xi_{12,k}, \cdots, \Xi_{1l,k}];$$

λ_{bk}= index of the highest amplitude component of $\Xi_{1i,k}^{\mathrm{T}}\boldsymbol{r}_{b,k-1}$;
$\Lambda_{bk} = \Lambda_{b,k-1} \bigcup \lambda_{bk}$;
$\hat{\boldsymbol{b}}_{\Lambda_{bk}} = (\Xi_{1,\Lambda_{bk}}^{\mathrm{T}}\Xi_{1,\Lambda_{bk}})^{-1}\Xi_{1,\Lambda_{bk}}^{\mathrm{T}}\boldsymbol{Y}$;
If $|\hat{b}_{h_b,k}| < \varepsilon$,
let $\hat{\boldsymbol{b}}_{\Lambda_{bk\varepsilon}} : \hat{\boldsymbol{b}}_{\Lambda_{bk}}(h_b) = [\ ]$,
$\hat{\Xi}_{1,\Lambda_{ck\varepsilon}} : \hat{\Xi}_{1,\Lambda_{ck}}(h_b) = [\ ]$;
$\boldsymbol{r}_{bk\varepsilon} = \boldsymbol{Y} - \Xi_{1,\Lambda_{bk\varepsilon}}\hat{\boldsymbol{b}}_{\Lambda_{bk\varepsilon}}$;

Estimating $\boldsymbol{c}$:

$$\Xi_{2,k} = \begin{bmatrix} \hat{\boldsymbol{b}}_k^{\mathrm{T}}\boldsymbol{F}(1) \\ \hat{\boldsymbol{b}}_k^{\mathrm{T}}\boldsymbol{F}(2) \\ \vdots \\ \hat{\boldsymbol{b}}_k^{\mathrm{T}}\boldsymbol{F}(m) \end{bmatrix} = [\Xi_{21,k}, \Xi_{22,k}, \cdots, \Xi_{2l,k}];$$

λ_{ck}= index of the highest amplitude component of $\Xi_{2i,k}^{\mathrm{T}}\boldsymbol{r}_{c,k-1}$;
$\Lambda_{ck} = \Lambda_{c,k-1} \bigcup \lambda_{ck}$;
$\hat{\boldsymbol{c}}_{\Lambda_{ck}} = (\Xi_{2,\Lambda_{ck}}^{\mathrm{T}}\Xi_{2,\Lambda_{ck}})^{-1}\Xi_{2,\Lambda_{ck}}^{\mathrm{T}}\boldsymbol{Y}$;
If $|\hat{c}_{h_c,k}| < \varepsilon$,
let $\hat{\boldsymbol{c}}_{\Lambda_{k\varepsilon}} : \hat{\boldsymbol{c}}_{\Lambda_k}(h_c) = [\ ]$,
$\hat{\Xi}_{2,\Lambda_{k\varepsilon}} : \hat{\Xi}_{2,\Lambda_k}(h_c) = [\ ]$;
$\boldsymbol{r}_{ck\varepsilon} = \boldsymbol{Y} - \Xi_{b,\Lambda_{ck\varepsilon}}\hat{\boldsymbol{c}}_{\Lambda_{ck\varepsilon}}$;
if $\|\boldsymbol{r}_{bk\varepsilon}\| < \varepsilon_1$ and $\|\boldsymbol{r}_{ck\varepsilon}\| < \varepsilon_1$, stop the iteration;
Otherwise

Return with k=$k+1$.

Estimated parameter vectors: $\hat{\boldsymbol{b}}_{\Lambda_{bk\varepsilon}}$ ($\hat{\boldsymbol{b}}_k$) and $\hat{\boldsymbol{c}}_{\Lambda_{ck\varepsilon}}$ ($\hat{\boldsymbol{c}}_k$).

4. The existing hierarchical least squares (H-LS) method

For comparison, we describe the existing H-LS algorithm. Refer to [16,17], the existing H-LS algorithm for Hammerstein system is under the condition of known n_b and n_c.

Define the information matrix and the parameter vectors:

$$\boldsymbol{F}_E(t) = \begin{bmatrix} f_1[u(t-1)], f_2[u(t-1)], \ldots, f_{n_c}[u(t-1)] \\ f_1[u(t-2)], f_2[u(t-2)], \ldots, f_{n_c}[u(t-2)] \\ \vdots \\ f_1[u(t-n_b)], f_2[u(t-n_b)], \ldots, f_{n_c}[u(t-n_b)] \end{bmatrix} \in \mathbb{R}^{n_b \times n_c},$$

$$\boldsymbol{b}_E = [b_1, b_2, \ldots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_b},$$
$$\boldsymbol{c}_E = [c_1, c_2, \ldots, c_{n_c}]^{\mathrm{T}} \in \mathbb{R}^{n_c},$$

then we have

$$y(t) = \boldsymbol{b}_E^{\mathrm{T}} \boldsymbol{F}_E(t) \boldsymbol{c}_E + v(t). \tag{22}$$

Define a quadratic cost function containing two parameter vectors,

$$J_E(\boldsymbol{b}_E, \boldsymbol{c}_E) = \sum_{t=1}^{L} [y(t) - \boldsymbol{b}_E^{\mathrm{T}} \boldsymbol{F}_E(t) \boldsymbol{c}_E]^2,$$

Under the known n_b and n_c, according to the least squares principle and the hierarchical identification principle [16,17], minimizing the cost functions $J_E(\boldsymbol{b}_E, \boldsymbol{c}_E)$ with respect to $\boldsymbol{b}_E$ and $\boldsymbol{c}_E$ gives

$$\hat{\boldsymbol{b}}_E(t) = \hat{\boldsymbol{b}}_E(t-1) + \boldsymbol{L}_{1E}(t)[y(t) - \hat{\boldsymbol{b}}_E^{\mathrm{T}}(t-1)\boldsymbol{F}_E(t)\hat{\boldsymbol{c}}_E(t-1)]^{\mathrm{T}}, \tag{23}$$

$$\boldsymbol{L}_{1E}(t) = \boldsymbol{P}_{1E}(t-1)\boldsymbol{F}_E(t)\hat{\boldsymbol{c}}_E(t-1) \times \{1 + [\boldsymbol{F}_E(t)\hat{\boldsymbol{c}}(t-1)]^{\mathrm{T}} \boldsymbol{P}_{1E}(t-1)\boldsymbol{F}_E(t)\hat{\boldsymbol{c}}_E(t-1)\}^{-1}, \tag{24}$$

$$\boldsymbol{P}_{1E}(t) = \{\boldsymbol{I} - \boldsymbol{L}_{1E}(t)[\boldsymbol{F}_E(t)\hat{\boldsymbol{c}}_E(t-1)]^{\mathrm{T}}\}\boldsymbol{P}_{1E}(t-1), \tag{25}$$

$$\hat{\boldsymbol{c}}_E(t) = \hat{\boldsymbol{c}}_E(t-1) + \boldsymbol{L}_{2E}(t)[y(t) - \hat{\boldsymbol{b}}_E^{\mathrm{T}}(t)\boldsymbol{F}_E(t)\hat{\boldsymbol{c}}_E(t-1)], \tag{26}$$

$$\boldsymbol{L}_{2E}(t) = \boldsymbol{P}_{2E}(t-1)[\hat{\boldsymbol{b}}_E^{\mathrm{T}}(t)\boldsymbol{F}_E(t)]^{\mathrm{T}} \times \{1 + \hat{\boldsymbol{b}}_E^{\mathrm{T}}(t)\boldsymbol{F}_E(t)\boldsymbol{P}_{2E}(t-1)[\hat{\boldsymbol{b}}_E^{\mathrm{T}}(t)\boldsymbol{F}_E(t)]^{\mathrm{T}}\}^{-1}, \tag{27}$$

$$\boldsymbol{P}_{2E}(t) = [\boldsymbol{I} - \boldsymbol{L}_{2E}(t)\hat{\boldsymbol{b}}_E^{\mathrm{T}}(t)\boldsymbol{F}_E(t)]\boldsymbol{P}_{2E}(t-1), \tag{28}$$

$$\boldsymbol{F}_E(t) = \begin{bmatrix} f_1[u(t-1)] & f_2[u(t-1)] & \cdots & f_{n_c}[u(t-1)] \\ f_1[u(t-2)] & f_2[u(t-2)] & \cdots & f_{n_c}[u(t-2)] \\ \vdots & \vdots & & \vdots \\ f_1[u(t-n_b)] & f_2[u(t-n_b)] & \cdots & f_{n_c}[u(t-n_b)] \end{bmatrix}. \tag{29}$$

With the known n_b and n_c, the above hierarchical least squares will take a large computation and a lot of time to get the estimated vectors $\hat{\boldsymbol{b}}_E$ and $\hat{\boldsymbol{c}}_E$.

5. Examples

5.1. Example 1

Consider the following Hammerstein system

$$y(t) = B(z)x(t) + v(t),$$
$$x(t) = c_1 u(t) + c_2 u^2(t) = 0.95u(t) - 0.60u^2(t),$$
$$B(z) = b_1 z^{-1} + b_2 z^{-2} = 1 + 0.80z^{-1} + 0.20z^{-2},$$
$$\boldsymbol{b} = [0.80, 0.20, \underbrace{0, \ldots, 0}_{4\ zeros}]^{\mathrm{T}},\ \boldsymbol{c} = [0.95, -0.60, \underbrace{0, \ldots, 0}_{4\ zeros}]^{\mathrm{T}}.$$

In simulation, the input $\{u(t)\}$ is taken as an uncorrelated persistently excited signal vector sequence with zero mean and unit variance, and $\{v(t)\}$ is taken as a white noise sequence with zero mean and variance $\sigma^2 = 0.02^2$.

For the proposed H-OMP algorithm with unknown orders n_b and n_c: sample 12 set of input–output data, apply the proposed H-OMP algorithm to estimate the parameters of this system, the parameter estimates and their errors are shown in Table 1 . The estimation error of the parameters is

$$\delta_k := \sqrt{\frac{\|\hat{\boldsymbol{b}}_k - \boldsymbol{b}\|^2 + \|\hat{\boldsymbol{c}}_k - \boldsymbol{c}\|^2}{\|\boldsymbol{b}\|^2 + \|\boldsymbol{c}\|^2}} \times 100\%.$$

For the existing H-LS algorithm with known orders $n_b = 2$ and $n_c = 2$: sample 5000 set of input–output data, apply the H-LS algorithm to estimate the parameters of this system, the parameter estimates and their errors are shown in Table 2. The estimation error of the parameters is

$$\delta_E := \sqrt{\frac{\|\hat{\boldsymbol{b}}_E - \boldsymbol{b}\|^2 + \|\hat{\boldsymbol{c}}_E - \boldsymbol{c}\|^2}{\|\boldsymbol{b}\|^2 + \|\boldsymbol{c}\|^2}} \times 100\%.$$

Table 1
The parameter estimates and errors of the H-OMP algorithm.

k	1	2	3	4	5	6
$b_1 = 0.85$	0.850	0.840	0.876	0.879	0.865	0.865
$b_2 = 0.20$	0.000	−0.176	−0.187	−0.187	−0.175	−0.178
$b_3 = 0.00$	0.000	0.000	0.000	0.000	0.000	0.000
$b_4 = 0.00$	0.000	0.000	0.000	0.000	0.000	0.000
$b_5 = 0.00$	0.000	0.000	0.000	0.020	0.004	0.003
$b_6 = 0.00$	0.000	0.000	0.000	0.000	0.000	0.000
$c_1 = 0.70$	0.000	0.000	0.580	0.687	0.690	0.692
$c_2 = -0.40$	0.000	0.000	0.000	−0.392	−0.385	−0.394
$c_3 = 0.00$	0.000	0.000	0.000	0.000	0.000	0.000
$c_4 = 0.00$	0.000	0.000	0.000	0.000	0.000	0.000
$c_5 = 0.00$	0.104	0.147	0.000	0.000	−0.001	0.000
$c_6 = 0.00$	0.000	−0.083	−0.080	0.000	0.000	0.000
$\delta_k(\%)$	70.436	69.340	35.860	3.368	2.917	2.473

Table 2
The parameter estimates and errors of the H-LS algorithm.

t	100	200	500	1000	2000	5000
$b_1 = 0.85$	0.014	0.025	0.291	0.816	0.818	0.817
$b_2 = -0.20$	0.007	0.004	0.019	−0.186	−0.193	−0.192
$c_1 = 0.70$	0.015	0.024	0.217	0.724	0.731	0.729
$c_2 = -0.40$	0.004	−0.005	−0.187	−0.419	−0.413	−0.414
$\delta_E(\%)$	98.654	97.237	67.280	3.997	3.940	3.934

The simulations are summarized as follows:

(1) Tables 1 and 2 show that the parameter estimation errors become (generally) smaller and smaller with the iteration k increasing or with the recursion t increasing.

(2) Table 1 shows that 4 $(n_b + n_c)$ support atoms are selected at k= 4, where 2 wrongly selected atoms corresponding to parameters c_5 and c_6 are deleted with iteration k increasing.

(3) The H-OMP algorithm with 12 set of data and the existing H-LS algorithm with 5000 set of data perform parameter estimation algorithm to obtain similar parameter estimation accuracy.

5.2. Example 2

Consider a simple experimental setup of a water tank system in Fig. 1, where $u(t)$ is the valve opening, $x(t)$ is the water inlet flow, and $y(t)$ is the liquid level. This system from $u(t)$ to $y(t)$ can be modeled by a simple Hammerstein model,

$$\begin{aligned} x(t) &= \gamma_1 u^2(t) + \gamma_2 u^2(t) + \cdots + \gamma_l u^l(t), \\ y(t) &= \beta_1 x(t-1) + \beta_2 x(t-2) + \cdots + \beta_l x(t-l) + v(t), \end{aligned} \tag{30}$$

then $y(t)$ can be expressed as

$$\begin{aligned} y(t) &= \sum_{i=1}^{l} [\beta_i(\gamma_1 u(t-i) + \gamma_2 u^2(t-i) + \cdots + \gamma_l u^l(t-i))] + v(t), \\ &= \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{F} \boldsymbol{\gamma} + v(t). \end{aligned}$$

Suppose the enough length $l = 6$ for orders β_l and γ_l, the parameter vectors $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_l]^{\mathrm{T}}$ and $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_l]^{\mathrm{T}}$ are sparse. Then we can apply the H-OMP algorithm to estimate the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Due to the different variable sampled with different scales, all the input and output variables need be re-scaled to be included within the interval $[-1, 1]$ by using the following equation:

$$\tilde{h}(t) = 2\frac{h(t) - h_{min}(t)}{h_{max}(t) - h_{min}(t)} - 1.$$

where $h(t)$ and $\tilde{h}(t)$ are the old and new values of a sampling point respectively. $h_{min}(t)$ and $h_{max}(t)$ are their minimum and maximum values. The fitting accuracy of the model is evaluated in terms of their mean square errors (MSE), defined as

$$\mathrm{MSE} = \frac{1}{L}\sum_{t=1}^{L} [\hat{y}(t) - y(t)]^2.$$

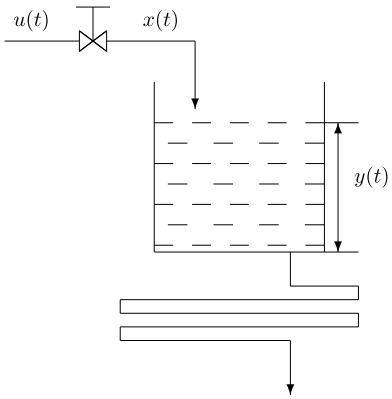

Fig. 1. An experimental setup of a water tank system.

Table 3
The parameter estimates and mean square errors of the PV system.

k	β_1	$\beta_2 \sim \beta_6$	γ_1	γ_2	$\gamma_3 \sim \gamma_6$	MSE
10	0.423	0	0.353	1.070	0	3.23%

where $\hat{y}(t)$ is an estimate of the real output $y(t)$. Applying the proposed H-OMP algorithm to estimate the parameters of this system, the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and their mean square errors (MSE) are shown in Table 3. It is clear that the MSE is small, the Hammerstein model and the H-OMP method is effective for modeling a water tank system.

6. Conclusions

This paper investigates the identification problems of a block-oriented Hammerstein system by using the sparsity-seeking hierarchical orthogonal matching pursuit (H-OMP) optimization algorithm of the compressed sensing theory. The idea is to recast the block-oriented Hammerstein system into two pseudo-regressive sub-systems by using the hierarchical identification principle. Each sub-system contains the parameters of the nonlinear part or the parameters of the linear part. Then the H-OMP algorithm is investigated to interactively select the parameters and orders of the nonlinear and linear parts. The characteristics of the proposed algorithm lie in three aspects:

- In contrast to the standard OMP algorithm with known measurement matrix, there exist unknown variables in the measurement matrix Ξ_1 and Ξ_2 in the H-OMP algorithm.
- Due to replacing the unknown variables with their estimates in Ξ_1 and Ξ_2, there exists the correcting process of picking up and deleting wrongly selected atoms with iteration k increasing.
- The advantage of the H-OMP method over the traditional identification methods for the Hammerstein system is that it is not necessary to collect a lot of data and invest a lot of power on the parameter identification.

The proposed identification method can be extended to various systems, such as time-delay systems [38,39], multivariate systems [40,41], network systems [42,43] and process systems [44,45], etc.

References

[1] G. Dimitriu, R. Ştefănescu, I.M. Navon, Comparative numerical analysis using reduced-order modeling strategies for nonlinear large-scale systems, J. Comput. Appl. Math. 310 (2017) 32–43.
[2] R.F. Reis, F.dos S. Loureiro, M. Lobosco, 3D numerical simulations on GPUs of hyperthermia with nanoparticles by a nonlinear bioheat model, J. Comput. Appl. Math. 295 (2016) 35–47.
[3] J. Chovan, M. Slodička, Induction hardening of steel with restrained Joule heating and nonlinear law for magnetic induction field: Solvability, J. Comput. Appl. Math. 311 (2017) 630–644.
[4] F. Ding, F.F. Wang, L. Xu, T. Hayat, A. Alsaedi, Parameter estimation for pseudo-linear systems using the auxiliary model and the decomposition technique, IET Control Theory Appl. 11 (3) (2017) 390–400.
[5] L. Xu, F. Ding, Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling, Circuits, Syst. Signal Process. 36 (4) (2017) 1735–1753.
[6] H. Salhi, S. Kamoun, A recursive parametric estimation algorithm of multivariable nonlinear systems described by Hammerstein mathematical models, Appl. Math. Model. 39 (16) (2015) 4951–4962.
[7] M. Jafari, M. Salimifard, M. Dehghani, Identification of multivariable nonlinear systems in the presence of colored noises using iterative hierarchical least squares algorithm, ISA Trans. 53 (4) (2014) 1243–1252.
[8] G.Q. Li, C.Y. Wen, W.X. Zheng, G.S. Zhao, Iterative identification of block-oriented nonlinear systems based on biconvex optimization, Systems Control Lett. 79 (2015) 68–75.

[9] E.W. Bai, K. Li, Convergence of the iterative algorithm for a general Hammerstein system identification, Automatica 46 (11) (2010) 1891–1896.
[10] J. Chen, Y.J. Liu, X.H. Wang, Recursive least squares algorithm for nonlinear dual-rate systems using missing-output estimation model, Circuits Systems Signal Process. 36 (4) (2017) 1406–1425.
[11] D.Q. Wang, W. Zhang, Improved least squares identification algorithm for multivariable Hammerstein systems, J. Franklin Inst. B 352 (11) (2015) 5292–5307.
[12] J. Vörös, Identification of nonlinear dynamic systems with input saturation and output backlash using three-block cascade models, J. Franklin Inst.–Eng. Appl. Math. 351 (12) (2014) 5455–5466.
[13] J. Vörös, Identification of nonlinear cascade systems with time-varying backlash, J. Electr. Eng. 62 (2) (2011) 87–92.
[14] Y.W. Mao, F. Ding, Adaptive filtering parameter estimation algorithms for Hammerstein nonlinear systems, Signal Process. 128 (2016) 417–425.
[15] D.Q. Wang, H.B. Liu, F. Ding, Highly efficient identification methods for dual-rate Hammerstein systems, IEEE Trans. Control Syst. Technol. 23 (5) (2015) 1952–1960.
[16] D.Q. Wang, F. Ding, Parameter estimation algorithms for multivariable Hammerstein CARMA systems, Inform. Sci. 355–356 (2016) 237–248.
[17] D.Q. Wang, L. Mao, F. Ding, Recasted models based hierarchical extended stochastic gradient method for MIMO nonlinear systems, IET Control Theory Appl. 11 (4) (2017) 476–485.
[18] F. Ding, X.H. Wang, Hierarchical stochastic gradient algorithm and its performance analysis for a class of bilinear-in-parameter systems, Circuits Systems Signal Process. 36 (4) (2017) 1393–1405.
[19] L. Ma, X.G. Liu, Recursive maximum likelihood method for the identification of Hammerstein ARMAX system, Appl. Math. Model. 40 (13–14) (2016) 6523–6535.
[20] D.Q. Wang, Z. Zhang, J.Y. Yuan, Maximum likelihood estimation method for dual-rate Hammerstein systems, Int. J. Control Autom. Syst. 15 (2) (2017) 698–705.
[21] J.H. Li, Parameter estimation for Hammerstein CARARMA systems based on the Newton iteration, Appl. Math. Lett. 26 (1) (2013) 91–96.
[22] J.H. Li, W.X. Zheng, J.P. Gu, L. Hua, Parameter estimation algorithms for Hammerstein output error systems using Levenberg–Marquardt optimization method with varying interval measurements, J. Franklin Inst. 354 (1) (2017) 316–331.
[23] J. Chen, X.P. Wang, Identification of Hammerstein systems with continuous nonlinearity, Inform. Process. Lett. 115 (11) (2015) 822–827.
[24] X.Y. Shen, L.M. Chen, Y.T. Gu, H.C. So, Square-root Lasso with nonconvex regularization: an ADMM approach, IEEE Signal Process. Lett. 23 (7) (2016) 934–938.
[25] G. Hua, Y. Xiang, G.A. Bi, When compressive sensing meets data hiding, IEEE Signal Process. Lett. 23 (4) (2016) 473–477.
[26] D.J. Bi, Y.L. Xie, X.F. Li, Y.R. Zheng, A sparsity basis selection method for compressed sensing, IEEE Signal Process. Lett. 22 (10) (2016) 1738–1742.
[27] K. Schnass, P. Vandergheynst, Average performance analysis for thresholding, IEEE Signal Process. Lett. 14 (11) (2016) 828–831.
[28] B.E. Hansen, Sample splitting and threshold estimation, Econometrica 68 (3) (2000) 575–603.
[29] T. Blumensath, M.E. Davies, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. 14 (5) (2008) 629–654.
[30] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.
[31] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.
[32] D.Q. Wang, L.W. Li, Y. Ji, Y.R. Yan, Model recovery for Hammerstein systems using the auxiliary model based orthogonal matching pursuit method, Applied Mathematical Modelling 54 (2018) 537–550.
[33] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1999) 33–61.
[34] J.F.C. Mota, J. Xavier, P.M.Q. Aguiar, M. Puschel, Distributed basis pursuit, IEEE Trans. Signal Process. 60 (4) (2012) 1942–1956.
[35] D. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory 52 (1) (2006) 6–18.
[36] Y.W. Mao, F. Ding, Y.J. Liu, Parameter estimation algorithms for Hammerstein time-delay systems based on the OMP scheme, IET Signal Process. 11 (3) (2017) 265–274.
[37] H. Rauhut, K. Schnass, P. Vandergheynst, Compressed sensing and redundant dictionaries, IEEE Trans. Inform. Theory 54 (5) (2008) 2210–2219.
[38] L. Xu, F. Ding, Y. Gu, A. Alsaedi, T. Hayat, A multi-innovation state and parameter estimation algorithm for a state space system with d-step state-delay, Signal Process. 140 (2017) 97–103.
[39] F. Ding, X.H. Wang, L. Mao, L. Xu, Joint state and multi-innovation parameter estimation for time-delay linear systems and its convergence based on the Kalman filtering, Digit. Signal Process. 62 (2017) 211–223.
[40] F. Ding, F.F. Wang, L. Xu, M.H. Wu, Decomposition based least squares iterative identification algorithm for multivariate pseudo-linear ARMA systems using the data filtering, J. Franklin Inst. B 354 (3) (2017) 1321–1339.
[41] D.Q. Wang, Y.P. Gao, Recursive maximum likelihood identification method for a multivariable controlled autoregressive moving average system, IMA J. Math. Control Inform. 33 (4) (2016) 1015–1031.
[42] L. Qiu, Y. Shi, J.F. Pan, G. Xu, Networked H_∞, Controller design for a direct-drive linear motion control system, IEEE Trans. Ind. Electron. 63 (10) (2016) 6281–6291.
[43] L. Qiu, Y. Shi, F.Q. Yao, G. Xu, B.G. Xu, Network-nased robust H_2/H_∞ control for linear systems with two-channel random packet dropouts and time delays, IEEE Trans. Cybern. 45 (8) (2015) 1450–1462.
[44] Y. Fu, X.G. Liu, Nonlinear dynamic behaviors and control based on simulation of high-purity heat integrated air separation column, ISA Trans. 55 (2015) 145–153.
[45] Y. Fu, X.G. Liu, Nonlinear wave modeling and dynamic analysis of high-purity heat integrated air separation column, Sep. Purif. Technol. 151 (2015) 14–22.