Article

A CS Recovery Algorithm for Model and Time Delay Identification of MISO-FIR Systems

Yanjun Liu [1,2,]* **and Taiyang Tao** [2]

[1] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China

[2] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; E-Mail: taiyangtao@hotmail.com

* Author to whom correspondence should be addressed; E-Mail: yjl@jiangnan.edu.cn; Tel.: +86-135-8415-9252; Fax: +86-510-8591-0633.

Academic Editor: Paul M. Goggans

Abstract: This paper considers identifying the multiple input single output finite impulse response (MISO-FIR) systems with unknown time delays and orders. Generally, parameters, orders and time delays of an MISO system are separately identified from different algorithms. In this paper, we aim to perform the model identification and time delay estimation simultaneously from a limited number of observations. For an MISO-FIR system with many inputs and unknown input time delays, the corresponding identification model contains a large number of parameters, requiring a great number of observations for identification and leading to a heavy computational burden. Inspired by the compressed sensing (CS) recovery theory, a threshold orthogonal matching pursuit algorithm (TH-OMP) is presented to simultaneously identify the parameters, the orders and the time delays of the MISO-FIR systems. The proposed algorithm requires only a small number of sampled data compared to the conventional identification methods, such as the least squares method. The effectiveness of the proposed algorithm is verified by simulation results.

Keywords: compressed sensing; sparse; parameter identification; gradient projection pursuit algorithm; time delay estimation

1. Introduction

In many industry processes, time delay is an unavoidable behavior in their dynamical models. For example, in a chemical industry, analyzers need a certain amount of time to process an analysis, which is used as a part of the control loop [1]. Analyzing and designing a controller for a system with time delays depend on an effective dynamical system model. System identification is an effective way to build mathematical models of dynamical systems from observed input-output data. Conventional identification methods, such as the least squares methods [2], the stochastic gradient methods [3] and the maximum likelihood estimation methods [4], usually require a large number of sampled data and that the model structures, including time delays, are *a priori* known [5,6]. However, this is not the case in many situations in practice for which the time delays and systems orders have to be estimated together with the parameters. Most of the process industries, such as distillation columns, heat exchangers and reactors, can be modeled as multiple input multiple output (MIMO) or multiple input single output (MISO) systems with long time delays [1]. For a multiple input single output finite impulse response (MISO-FIR) system with many inputs and unknown input time delays, the parameterized model contains a large number of parameters, requiring a great number of observations and a heavy computational burden for identification. Furthermore, the measuring cost is an issue. Especially for chemical processes, it will take a long time to get enough data because of the long analysis time. Therefore, it is a challenging work to find new methods to identify the time delays and parameters of multivariable systems with less computational burden using a small number of sampled data.

System models of practical interest, especially the control-oriented models, often require low order, which implies that the systems contain only a few non-zero parameters. Because the time delays and orders are unknown, we can parameterize such an MISO-FIR system by a high-dimensional parameter vector, with only a few non-zero elements. These systems can be termed as sparse systems. On the other hand, a sparse MISO-FIR system contains a long impulse response, but only a few terms are non-zeros. The sparse systems are widely discussed in many fields [7–11]. For such sparse systems, the identification objective is to estimate the sparse parameter vector and to read the time delays and orders from the structure of the estimated parameter vector.

In this paper, we consider the identification of the sparse MISO-FIR systems by using the compressed sensing (CS) recovery theory [12–15]. CS has been widely applied in signal processing and communication systems, which enables the recovery of an unknown vector from a set of measurements under the assumption that the signal is sparse and certain conditions on the measurement matrices [16–19]. The identification of a sparse system can be viewed as a CS recovery problem. The sparse recovery problem can be solved by the l_1-norm convex relaxation. However, the l_1 regularization schemes are always complex, requiring heavy computational burdens [20–22]. The greedy methods have speed advantages and are easily implemented. In this literature, a block orthogonal matching pursuit (BOMP) algorithm has been applied to identify the determined linear time invariant (LTI) and linear time variant (LTV) auto-regressive with external (ARX) models without disturbances [9]. In this paper, we will consider identifying the parameters, orders and time delays of the MISO-FIR systems from a small number of noisy measured data.

Briefly, the structure of this paper is as follows. Section 2 gives the model description of MISO-FIR systems and describes the identification problem. Section 3 presents a threshold orthogonal matching pursuit (TH-OMP) algorithm based on the compressed sensing theory. Section 4 provides a simulation example to show the effectiveness of the proposed algorithm. Finally, some concluding remarks are given in Section 5.

2. Problem Description

Consider an MISO-FIR system with time delays:

$$y(t) = \sum_{i=1}^{r} z^{-d_i} B_i(z) u_i(t) + v(t) \tag{1}$$

where $u_i(t)$ is the i-th input, $y(t)$ the output, d_i the time delay of the i-th input channel and $v(t)$ a white noise with zero mean and variance σ^2, $B_i(z)$, $i = 1, 2, \cdots, r$ are the polynomials in the unit backward shift operator z^{-1} (*i.e.*, $z^{-1}y(t) = y(t-1)$) and:

$$B_i(z) := b_{i1}z^{-1} + b_{i2}z^{-2} + \cdots + b_{in_{bi}}z^{-n_{bi}}$$

Assume that $y(t) = 0, u(t) = 0$ and $v(t) = 0$ for $t \leqslant 0$ and the orders n_{bi}, the time delays d_i, as well as the parameters b_{ij} are unknown. Thus, the identification goal is to estimate the unknown parameters b_{ij}, orders n_{bi} and the time delays d_i from the observations.

Define the information vector:

$$\boldsymbol{\varphi}(t) := [u_1(t-1), \cdots, u_1(t-d_1), u_1(t-d_1-1), \cdots, u_1(t-d_1-n_{b_1}), \cdots, u_1(t-l), \cdots,$$
$$u_r(t-1), \cdots, u_r(t-d_r), u_r(t-d_r-1), \cdots, u_r(t-d_r-n_{b_r}), \cdots, u_r(t-l)]^{\mathrm{T}} \in \mathbb{R}^N$$

where l is the input maximum regression length. l must be chosen to capture all of the time delays and the degrees of the polynomials $B_i(z)$. Thus, $l \geqslant \max(d_i + n_{bi})$. The dimension of $\boldsymbol{\varphi}(t)$ is $N = lr$. The corresponding parameter vector $\boldsymbol{\theta}$ is:

$$\boldsymbol{\theta} := [0, \cdots, 0, b_{11}, \cdots, b_{1n_{b_1}}, 0, \cdots, 0, b_{i1}, \cdots, b_{in_{b_i}}, 0, \cdots, 0, b_{r1}, \cdots, b_{rn_{b_r}}, 0, \cdots, 0]^{\mathrm{T}} \in \mathbb{R}^N$$

Then, Equation (1) can be written as:

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta} + v(t) \tag{2}$$

Because of the unknown orders and time delays, the parameter vector $\boldsymbol{\theta}$ contains many zeros, but only a few non-zeros. According to the CS theory, the parameter vector $\boldsymbol{\theta}$ can be viewed as a sparse signal, and the number of non-zero entries $K = \sum_{i=1}^{r} n_{bi}$ is the sparsity level of $\boldsymbol{\theta}$ [8]. Let $\|\boldsymbol{\theta}\|_0$ denote the number of non-zero entries in $\boldsymbol{\theta}$. Then, the identification problem can be described as:

$$\hat{\boldsymbol{\theta}} = \arg\min \|\boldsymbol{\theta}\|_0, \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}\|_2 < \varepsilon, \tag{3}$$

where $\hat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$ and $\varepsilon > 0$ is the error tolerance.

For m observations, define the stacked matrix and vectors as:

$$\boldsymbol{\Phi}_m := \begin{bmatrix} \boldsymbol{\varphi}^{\mathrm{T}}(1) \\ \boldsymbol{\varphi}^{\mathrm{T}}(2) \\ \vdots \\ \boldsymbol{\varphi}^{\mathrm{T}}(m) \end{bmatrix} \in \mathbb{R}^{m \times N}, \; \boldsymbol{y}_m := \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(m) \end{bmatrix} \in \mathbb{R}^m, \; \boldsymbol{v}_m := \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(m) \end{bmatrix} \in \mathbb{R}^m$$

Then, we have:

$$\boldsymbol{y}_m = \boldsymbol{\Phi}_m \boldsymbol{\theta} + \boldsymbol{v}_m \tag{4}$$

If there are enough measurements, *i.e.*, $m \gg N$, according to the least squares principle, we can get the least squares estimate of $\boldsymbol{\theta}$,

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\boldsymbol{\Phi}_m^{\mathrm{T}} \boldsymbol{\Phi}_m)^{-1} \boldsymbol{\Phi}_m^{\mathrm{T}} \boldsymbol{y}_m$$

However, from Equation (2), it is obvious that N is a large number, especially for systems with a large number of inputs. Therefore, it will take a lot of time and effort to get enough measurements. In order to improve the identification efficiency, in this paper, we aim to identify the parameter vector $\boldsymbol{\theta}$ using less observations ($K < m < N$) based on the CS recovery theory.

3. Identification Algorithm

The CS theory, which was introduced by Candès, Romberg and Tao [12,13] and Donoho [14], has been widely used in signal processing. It indicates that an unknown signal vector can be recovered from a small set of measurements under the assumptions that the signal vector is sparse and some conditions on the measurement matrix. The identification problem can be viewed as a CS recovery problem in that the parameter vector $\boldsymbol{\theta}$ is sparse.

From Equation (4), we can see that the output vector $\boldsymbol{y}_m$ can be written as a linear combination of the columns of $\boldsymbol{\Phi}_m$ plus the noise vector $\boldsymbol{v}_m$, *i.e.*,

$$\boldsymbol{y}_m = \boldsymbol{\phi}_1 \theta_1 + \boldsymbol{\phi}_2 \theta_2 + \cdots + \boldsymbol{\phi}_i \theta_i + \cdots + \boldsymbol{\phi}_N \theta_N + \boldsymbol{v}_m \tag{5}$$

where θ_i is the i-th element of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}_i$ is the i-th column of $\boldsymbol{\Phi}_m$. Because there are only K non-zero parameters in $\boldsymbol{\theta}$, the main idea is to find the K non-zero items at the right-hand side of Equation (5).

In this paper, we propose a TH-OMP algorithm to do the identification. First, we give some notations. Let $k = 1, 2, \cdots$ be the iterative number and λ_k be the solution support of the k-th iterative number; Λ_k is a set composed of λ_i, $i = 1, 2, \cdots, k$; r_k denotes the residual of the k-th iterative number; $\boldsymbol{\Phi}_{\Lambda_k}$ is the sub-information matrix composed by the k columns of $\boldsymbol{\Phi}_m$ indexed by Λ_k. $\hat{\boldsymbol{\theta}}_k$ is the parameter estimation of the k-th iterative number. The TH-OMP algorithm is initialized as $\boldsymbol{r}_0 = \boldsymbol{y}_m$, $\Lambda_0 = \varnothing$ and $\hat{\boldsymbol{\theta}}_{\Lambda_0} = \boldsymbol{0}$. At the k-th iteration, minimizing the criteria function:

$$\varepsilon(i) = \min \|\boldsymbol{r}_{k-1} - \boldsymbol{\phi}_i \theta_i\|^2, \; i = 1, 2, \cdots, N$$

with respect to θ_i yields:

$$\boldsymbol{\theta}_i = \frac{\boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{r}_{k-1}}{\|\boldsymbol{\theta}_i\|^2}$$

Plugging it back into $\varepsilon(i)$, we have:

$$
\begin{aligned}
\varepsilon(i) &= \min_{\theta_i} \|\boldsymbol{\phi}_i\theta_i - \boldsymbol{r}_{k-1}\|^2 \\
&= \left\| \frac{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1}}{\|\boldsymbol{\phi}_i\|^2}\boldsymbol{\phi}_i - \boldsymbol{r}_{k-1} \right\|^2 \\
&= \|\boldsymbol{r}_{k-1}\|^2 - 2\frac{(\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1})^2}{\|\boldsymbol{\phi}_i\|^2} + \frac{(\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1})^2}{\|\boldsymbol{\phi}_i\|^2} \\
&= \|\boldsymbol{r}_{k-1}\|^2 - \frac{(\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1})^2}{\|\boldsymbol{\phi}_i\|^2} \\
&= \|\boldsymbol{r}_{k-1}\|^2 - (\frac{1}{\|\boldsymbol{\phi}_i\|}\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{r}_{k-1})^2
\end{aligned}
\tag{6}
$$

From the above equation, we can see that the quest for the smallest error is equivalent to the quest for the largest inner product between the residual $\mathbf{r}_{k-1}$ and the normalized column vectors of $\boldsymbol{\Phi}_m$. Thus, the k-th solution support can be obtained by:

$$
\lambda_k = \arg \max_{i=1,2,\cdots,N} |\langle \boldsymbol{r}_{k-1}, \frac{\boldsymbol{\phi}_i}{\|\boldsymbol{\phi}_i\|}\rangle|
\tag{7}
$$

Update the support set Λ_k and the sub-information matrix $\boldsymbol{\Phi}_{\Lambda_k}$ by:

$$
\Lambda_k := \Lambda_{k-1} \cup \lambda_k, \quad \boldsymbol{\Phi}_{\Lambda_k} := \boldsymbol{\Phi}_{\Lambda_{k-1}} \cup \boldsymbol{\phi}_{\lambda_k}
\tag{8}
$$

The next step is to find the k-th provisional parameter estimate from the sub-information matrix $\boldsymbol{\Phi}_{\Lambda_k}$ and measurement vector $\boldsymbol{y}_m$. Minimizing the cost function:

$$
J(\boldsymbol{\theta}_{\Lambda_k}) = \|\boldsymbol{y}_m - \boldsymbol{\Phi}_{\Lambda_k}\boldsymbol{\theta}_{\Lambda_k}\|^2
\tag{9}
$$

leads to the k-th parameter estimate:

$$
\hat{\boldsymbol{\theta}}_{\Lambda_k} = (\boldsymbol{\Phi}_{\Lambda_k}^{\mathrm{T}}\boldsymbol{\Phi}_{\Lambda_k})^{-1}\boldsymbol{\Phi}_{\Lambda_k}^{\mathrm{T}}\boldsymbol{y}_m
\tag{10}
$$

Then, the k-th residual can be computed by:

$$
\boldsymbol{r}_k = \boldsymbol{y}_m - \boldsymbol{\Phi}_{\Lambda_k}\hat{\boldsymbol{\theta}}_{\Lambda_k}
\tag{11}
$$

If the system is noise free, then the residual will be zero after K iterations, and the non-zero parameters, as well as their positions can be exactly determined. This is the so-called orthogonal matching pursuit (OMP) algorithm. However, for systems disturbed by noises, the OMP algorithm cannot give the sparsest solution, because the residual $\boldsymbol{r}_K$ is non-zero after K iterations, and the iteration continues. In order to solve this problem, an appropriate small threshold ϵ can be set to filter the parameter estimate $\hat{\boldsymbol{\theta}}_{\Lambda_k}$. If $|\hat{\theta}_{\Lambda_\epsilon}| < \epsilon$, let $\hat{\theta}_{\Lambda_\epsilon} = 0$, where $\hat{\theta}_{\Lambda_\epsilon}$ is the i-th element of $\hat{\boldsymbol{\theta}}_{\Lambda_k}$. The choice of ϵ will be discussed in the simulation part. Denote $\hat{\boldsymbol{\theta}}_{\Lambda_{k\epsilon}}$ as the parameter estimate after the filtering. Then, the residual can be computed by:

$$
\boldsymbol{r}_k = \boldsymbol{y}_m - \boldsymbol{\Phi}_{\Lambda_k}\hat{\boldsymbol{\theta}}_{\Lambda_{k\epsilon}}
\tag{12}
$$

if $\|\mathbf{r}_k\| < \varepsilon$, then stop the iteration; otherwise, go to apply another iteration. If the iteration stops at k, we can recover the parameter estimate $\hat{\boldsymbol{\theta}} \in \mathbb{R}^N$ from $\hat{\boldsymbol{\theta}}_{\Lambda_{k\epsilon}} \in \mathbb{R}^k$ based on the support set Λ_k.

It is worth noting that taking the derivative of $J(\boldsymbol{\theta}_{\Lambda_k})$ with respect to $\hat{\boldsymbol{\theta}}_{\Lambda_k}$ gives:

$$
\begin{aligned}
\frac{\partial J(\hat{\boldsymbol{\theta}}_{\Lambda_k})}{\partial \hat{\boldsymbol{\theta}}_{\Lambda_k}} &= -\boldsymbol{\Phi}_{\Lambda_k}^{\mathrm{T}}(\boldsymbol{y}_m - \boldsymbol{\Phi}_{\Lambda_k}\hat{\boldsymbol{\theta}}_{\Lambda_k}) \\
&= \boldsymbol{\Phi}_{\Lambda_k}^{\mathrm{T}}\boldsymbol{r}_k = \mathbf{0}
\end{aligned}
\tag{13}
$$

The above equation shows that the residual $\mathbf{r}_k$ is orthogonal to the sub-information matrix $\boldsymbol{\Phi}_{\Lambda_k}$. Therefore, it does not require computing the inner products of $\mathbf{r}_k$ and the columns in $\boldsymbol{\Phi}_{\Lambda_k}$ during the $(k+1)$-th iteration. Modify Equation (7) as:

$$
\lambda_k = \arg \max_{i \in \Omega \backslash \Lambda_{k-1}} |\langle \boldsymbol{r}_{k-1}, \frac{\phi_i}{\|\phi_i\|}\rangle|, \quad \Omega = \{1, 2, \cdots, N\}
\tag{14}
$$

Then, the computational burden can be reduced.

From the recovered parameter estimation $\hat{\boldsymbol{\theta}}$, the orders and time delays of each input channel can be estimated. The orders n_{bi} can be determined by the number of elements of each non-zero block. Then, the time delay estimates can be computed from the orders n_{bi} and the input maximum regression length l. It can be seen from Equation (2) that $\boldsymbol{\theta}$ contains $(r+1)$ zero blocks. Denote the number of zeros of each zero-block as n_i, $i = 1, 2, \cdots, r+1$. Then, the time delay estimates can be computed by:

$$
\begin{aligned}
\hat{d}_1 &= n_1, \\
\hat{d}_i &= n_i - (l - \hat{d}_{i-1} - n_{b_i}), \ i = 2, 3, \cdots, r
\end{aligned}
\tag{15}
$$

The proposed TH-OMP algorithm can be implemented as the following steps.

1. Set the initial values: let $k = 0$, $\boldsymbol{r}_0 = \boldsymbol{y}_m$, $\Lambda_0 = \varnothing$ and $\hat{\boldsymbol{\theta}}_{\Lambda_0} = \mathbf{0}$. Give the error tolerant ε.
2. Increment k by one; compute the solution support λ_k using Equation (14).
3. Update the support set Λ_k and the sub-information matrix $\boldsymbol{\Phi}_{\Lambda_k}$ using Equation (8).
4. Compute the parameter estimate $\hat{\boldsymbol{\theta}}_{\Lambda_k}$ by using Equation (10).
5. Using the threshold ϵ to filter $\hat{\boldsymbol{\theta}}_{\Lambda_k}$ to get $\hat{\boldsymbol{\theta}}_{\Lambda_{k\epsilon}}$.
6. Compute the residual using Equation (12). If $\|\mathbf{r}_k\| < \varepsilon$, stop the iteration; otherwise, go back to Step 2 to apply another iteration.
7. Recover the parameter estimate $\hat{\boldsymbol{\theta}} \in \mathbb{R}^N$ from $\hat{\boldsymbol{\theta}}_{\Lambda_{k\epsilon}} \in \mathbb{R}^k$ based on the support set Λ_k.
8. Read the orders n_{bi} from $\hat{\boldsymbol{\theta}} \in \mathbb{R}^N$ and estimate the time delays using Equation (15).

4. Simulation Example

Consider the steam water heat exchanger system in Figure 1, where the steam flow u_1 and the water flow u_2 are two manipulated variables to control the water temperature y. The discrete input-output representation of the process is expressed as the following equation.
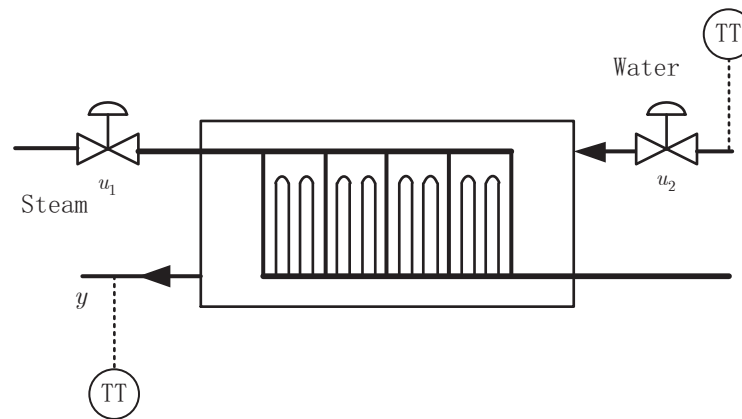
Figure 1. The steam water heat exchanger.

$$y(t) = z^{-d_1} B_1(z) u_1(t) + z^{-d_2} B_2(z) u_2(t) + v(t)$$
$$B_1(z) = 0.95 z^{-1} + 0.72 z^{-2} + 0.40 z^{-3},$$
$$B_2(z) = -0.63 z^{-1} - 0.47 z^{-2} - 0.25 z^{-3},$$
$$d_1 = 42, d_2 = 20$$

Let $l = 50$; then the parameter vector to be identified is:

$$\boldsymbol{\theta} = [\mathbf{0}_{42}, 0.95, 0.72, 0.40, \mathbf{0}_{25}, -0.63, -0.47, -0.25, \mathbf{0}_{27}]^{\mathrm{T}} \in \mathbb{R}^N$$

It can be seen from the above equation that the dimension of the parameter vector $\boldsymbol{\theta}$ is $N = 100$, but only six of the parameters are non-zeros.

In the identification, the inputs $\{u_i(t)\}$, $i = 1, 2$ are taken as independent persistent excitation signal sequences with zero mean and unit variances and $\{v(t)\}$ as white noise sequences with zero mean and variance $\sigma^2 = 0.10^2$. Let $\epsilon = 0.05$, applying the proposed TH-OMP algorithm and the least squares (LS) algorithm to estimate the system with m measurements. The parameter estimation errors $\delta := (\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|)/\|\boldsymbol{\theta}\|$ *versus* m are shown in Figure 2.

From Figure 2, we can see that under a limited dataset ($m < 100$), the parameter estimation error of the TH-OMP algorithm approaches zero, and the estimation accuracy is much higher than that of the LS algorithm.

Let $m = 80$; using the TH-OMP algorithm to estimate $\boldsymbol{\theta}$, the parameter estimation errors δ *versus* iteration k are shown in Figure 3; the non-zero parameter estimates *versus* iteration k are shown in Table 1 and Figure 4, where the dash dot lines are true parameters and the solid lines are parameter estimates.
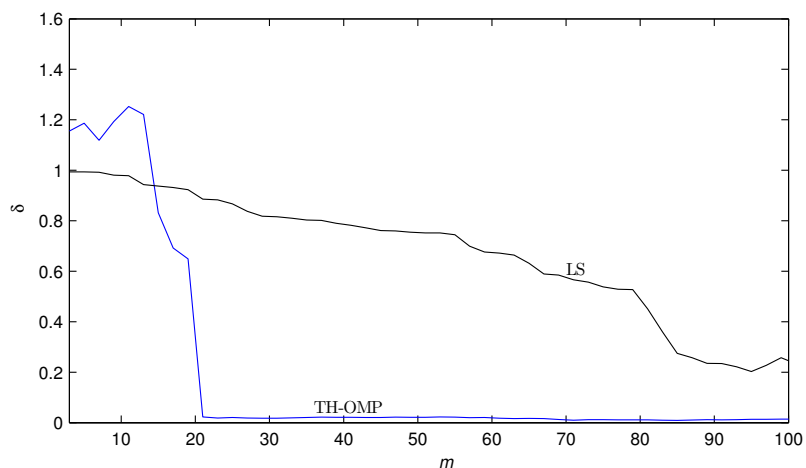
Figure 2. The estimation error δ *versus* m measurements.
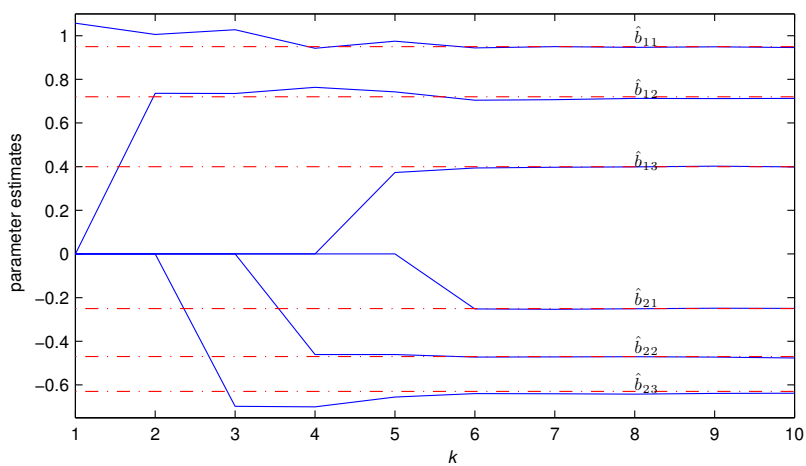
Figure 3. The estimation error δ *versus* iteration k ($m = 80$, $\sigma^2 = 0.10^2$).
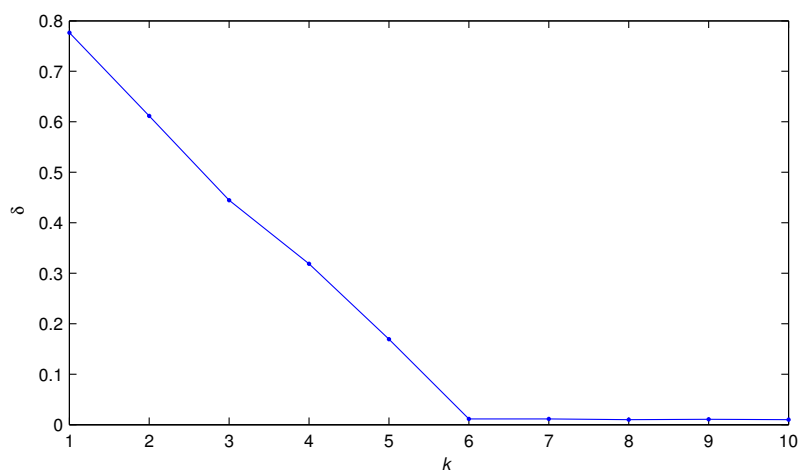
Figure 4. The parameters estimates *versus* iteration k.

Table 1. The estimation error δ *versus* iteration k (m=80, $\sigma^2 = 0.10^2$).

k	b_{11}	b_{12}	b_{13}	b_{21}	b_{22}	b_{23}	$\delta(\%)$
1	1.0567	0.0000	0.0000	0.0000	0.0000	0.0000	77.8404
2	1.0056	0.7357	0.0000	0.0000	0.0000	0.0000	61.0813
3	1.0274	0.7349	0.0000	−0.6982	0.0000	0.0000	44.8210
4	0.9422	0.7635	0.0000	−0.7006	−0.4604	0.0000	31.8601
5	0.9748	0.7429	0.3734	−0.6551	−0.4605	0.0000	16.9639
6	0.9436	0.7041	0.3936	−0.6397	−0.4727	−0.2522	1.3977
7	0.9495	0.7071	0.3973	−0.6408	−0.4715	−0.2537	1.1650
8	0.9466	0.7127	0.3986	−0.6422	−0.4707	−0.2509	0.9755
9	0.9487	0.7120	0.4022	−0.6386	−0.4722	−0.2490	0.8137
10	0.9458	0.7129	0.3988	−0.6380	−0.4766	−0.2498	0.8841
True values	0.9500	0.7200	0.4000	−0.6300	−0.4700	−0.2500	

After 10 iterations, the estimation error is $\delta = 0.8841\%$, and the estimated parameter vector is:

$$\hat{\boldsymbol{\theta}}_{\epsilon=0.05} = [\mathbf{0}_{42}, 0.9472, 0.7105, 0.3921, \mathbf{0}_{25}, 0.6246, 0.4751, 0.2524, \mathbf{0}_{27}]^{\mathrm{T}} \in \mathbb{R}^N \qquad (16)$$

If we choose $\epsilon = 0.01$, then we can get the estimated parameter vector as:

$$\hat{\boldsymbol{\theta}}_{\epsilon=0.01} = [\mathbf{0}_4, 0.0217, \mathbf{0}_{25}, 0.0303, \mathbf{0}_{11}, 0.9458, 0.7129, 0.3988, \mathbf{0}_{25}$$
$$0.6220, 0.4634, 0.2502, \mathbf{0}_{13}, -0.0244, \mathbf{0}_{13}]^{\mathrm{T}} \in \mathbb{R}^N$$

It is obvious that the parameter vector $\boldsymbol{\theta}$ has only three zero blocks in this example. However, we can see from the above equation that there exist three undesirable parameter estimates $\hat{\theta}_5$, $\hat{\theta}_{31}$ and $\hat{\theta}_{87}$. Moreover, the three parameter estimates are much smaller than other parameters. Therefore, according to the structure of $\boldsymbol{\theta}$, the parameters $\hat{\theta}_5$, $\hat{\theta}_{31}$ and $\hat{\theta}_{87}$ can be set to zeros. Then, the estimation result is the same as the case when $\epsilon = 0.05$. This implies that even if a too small ϵ is chosen, we can still obtain the effective parameter estimates based on the model structures.

From Equations (15) and (16), we can obtain the orders $n_{b_1} = n_{b_2} = 3$ and get the time delay estimates as:

$$\hat{d}_1 = 42, \quad \hat{d}_2 = n_2 - (l - d_1 - n_{b1}) = 25 - (50 - 42 - 3) = 20 \qquad (17)$$

From Figures 2–4, Table 1 and Equations (16) and (17), we can conclude that the TH-OMP algorithm converges faster than the LS algorithm, can achieve a high estimation accuracy by K iterations with finite measurements ($m < N$) and can effectively estimate the time delay of each channel.

5. Conclusions

This paper presents a TH-OMP algorithm for MISO-FIR systems with unknown orders and time delays. The parameter estimates, orders and time delays can be simultaneously estimated from a small number of observation data. The proposed method can be simply implemented and can reduce the

measuring cost, as well as improve the identification efficiency. The simulation results are given to demonstrate the performance of the proposed method.

Acknowledgments

Author Contributions

Yanjun Liu worked on the main details of the algorithm and was responsible for writing the manuscript. Taiyang Tao helped with the MATLAB simulation.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Normey-Rico, J.E.; Camacho, E.F. *Control of Dead-Time Processes*; Spinger: Berlin, Germany, 2007.
2. Liu, Y.J.; Ding, F. Convergence properties of the least squares estimation algorithm for multivariable systems. *Appl. Math. Model.* **2013**, *37*, 476–483.
3. Liu, Y.J.; Ding, R. Consistency of the extended gradient identification algorithm for multi-input multi-output systems with moving average noises. *Int. J. Comput. Math.* **2013**, *90*, 1840–1852.
4. Li, J.H.; Ding, F.; Hua, L. Maximum likelihood newton recursive and the newton iterative estimation algorithms for hammerstein CARAR systems. *Nonlinear Dyn.* **2014**, *75*, 235–245.
5. Ljung, L. *System Identification, Theory for the User*, 2nd ed.; Prentince Hall: Englewood Cliffs, NJ, USA, 1999.
6. Ding, F. *System Identification-New Theory and Methods*; Science Press: Beijing, China, 2013.
7. Haupt, J.; Bajwa, W.U.; Raz, G.; Nowak, R. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE Trans. Inf. Theory* **2010**, *56*, 5862–5875.
8. Sanandaji, B.M.; Vincent, T.L.; Wakin, M.B. Exact topology identification of large-scale interconnected dynamical systems from compressive observations. In Proceedings of the 2011 American Control Conference, San Francisco, CA, USA, 29 June–1 July 2011; pp. 649–656.
9. Sanandaji, B.M.; Vincent, T.L.; Wakin, M.B.; Toth, R. Compressive system identification of LTI and LTV ARX models. In Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, USA, 12–15 December 2011; pp. 791–798.
10. Gui, G.; Fumiyuki, A. Stable adaptive sparse filtering algorithms for estimating multiple-input multiple-output channels. *IET Commun.* **2014**, *8*, 1032–1040.
11. Gui, G.; Xu, L.; Fumiyuki, A. Normalized least mean square-based adaptive sparse filtering algorithms for estimating multiple-input multiple-output channels. *Wirel. Commun. Mob. Comput.* **2015**, *15*, 1079–1088.

12. Candès, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509.

13. Candès, E.J. Compressive sampling. In *Proceedings of the 2006 International Congress of Mathematics*; The European Mathematical Society: Madrid, Spain, 2006; pp. 1433–1452.

14. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306.

15. Elad, M. Optimized projections for compressed sensing. *IEEE Trans. Signal Process.* **2007**, *55*, 5695–5702.

16. Baraniuk, R. A lecture on compressive sensing. *IEEE Signal Process. Mag.* **2007**, *24*, 118–121.

17. Candès, E.J.; Wakin, M.B. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 21–30.

18. Tropp, A.J. Just relax: Convex programming methods for ifentifying sparse signals in noise. *IEEE Trans. Inf. Theory* **2006**, *52*, 1030–1051.

19. Candès, E.J. Robust uncertainty principles: Extra signal reconstruction from highly frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509.

20. Tóth, R.; Sanandaji, B.M.; Poolla, K.; Vincent, T.L. Compressive system identification in the linear time-invariant framework. In Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, USA, 12–15 December 2011; pp. 783–790.

21. Yang, Z.; Zhang, C.S.; Deng, J.; Lu, W.M. Orthonormal expansion ℓ_1 minimization algorithms for compressed sensing. *IEEE Trans. Signal Process.* **2011**, *59*, 6285–6290.

22. Le, V.L.; Lauer, F.; Bloch, G. Selective ℓ_1 minimization for sparse recovery. *IEEE Trans. Autom. Control* **2014**, *59*, 3008–3013.