

Least Squares-Based Iterative Identification Methods for Linear-in-Parameters Systems Using the Decomposition Technique

Feifei Wang¹ · Yanjun Liu¹ · Erfu Yang²

Received: 28 October 2015 / Revised: 16 December 2015 / Accepted: 17 December 2015 /
Published online: 6 January 2016
© Springer Science+Business Media New York 2015

Abstract By extending the least squares-based iterative (LSI) method, this paper presents a decomposition-based LSI (D-LSI) algorithm for identifying linear-in-parameters systems and an interval-varying D-LSI algorithm for handling the identification problems of missing-data systems. The basic idea is to apply the hierarchical identification principle to decompose the original system into two fictitious sub-systems and then to derive new iterative algorithms to estimate the parameters of each sub-system. Compared with the LSI algorithm and the interval-varying LSI algorithm, the decomposition-based iterative algorithms have less computational load. The numerical simulation results demonstrate that the proposed algorithms work quite well.

Keywords Parameter estimation · Iterative identification · Decomposition technique · Missing data · Linear-in-parameters system

✉ Feifei Wang
ffwang12@126.com

Yanjun Liu
yanjunliu_1983@126.com

Erfu Yang
erfu.yang@strath.ac.uk

¹ Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, People's Republic of China

² Department of Design, Manufacture and Engineering Management, Space Mechatronic Systems Technology Laboratory, Strathclyde Space Institute, University of Strathclyde, Glasgow G1 1XJ, Scotland, UK

1 Introduction

Parameter estimation and mathematical models are essential for system identification [13,31,33], system optimization [16,24] and state and data filtering [14,19,32]. Exploring new parameter estimation methods is an eternal theme of system identification [5,6], and many identification methods have been developed for linear and nonlinear systems [1,25,38,40], dual-rate sampled systems [9,11,36] and state-delay systems [28]. Iterative methods can be used for estimating parameters and solving matrix equations [4]. The iterative identification algorithms make full use of the measured data at each iteration and thus can produce more accurate parameter estimates than the existing recursive identification algorithms [29]. For decades, many iterative methods have been applied in the parameter estimation, such as the Newton iterative method [7,26,41,42], the gradient-based iterative methods [39] and the least squares-based iterative (LSI) method [17]. Jin et al. [23] studied the LSI identification methods for multivariable integrating and unstable processes in closed loop; Wang et al. [37] derived several gradient-based iterative estimation algorithms for a class of nonlinear systems with colored noises using the filtering technique.

The least squares identification method involves matrix inversion, and its computational complexity depends on the dimensions of the covariance matrices [18]. In order to reduce the computational complexity, the decomposition technique is usually taken to transform a large-scale system into several sub-systems with small sizes, which can be easier to identify. Chen et al. [2] developed a decomposition-based least squares identification algorithm for input nonlinear systems by adopting the key term separation technique; Zhang [43] proposed a decomposition-based LSI identification algorithm for output error moving average systems based on the hierarchical identification principle.

In the field of system identification, missing-data systems have received much attention. Dual-rate sampled systems and multirate (non-uniformly) sampled systems can be regarded as a class of the systems with missing data [10]. In recent years, different identification methods for missing-data systems have been reported in the literature, e.g., the interval-varying auxiliary model-based recursive least squares method [8], the filtering-based multiple-model method [27] and the interval-varying auxiliary model-based multi-innovation stochastic gradient (V-AM-MISG) identification method [8,12]. Recently, Jin et al. [22] extended the V-AM-MISG method to multivariable output error systems with scarce measurements by means of the interval-varying and multi-innovation methods in [8,12]; Raghavan et al. [30] studied the expectation maximization-based state-space model identification problems with irregular output sampling.

This paper applies the decomposition technique to study the parameter identification problems of linear-in-parameters systems for improving computational efficiency. The key is to decompose the information vector into two sub-information vectors and the parameter vector into two sub-parameter vectors with smaller dimensions and fewer variables and then to estimate the parameters of each sub-system, respectively. The main contributions are as follows.

- A decomposition-based LSI (D-LSI) algorithm is developed for linear-in-parameters systems by employing the hierarchical identification principle.
- An interval-varying D-LSI algorithm is derived for estimating the parameters of the systems with missing data.
- The proposed algorithms have higher computational efficiency than the LSI algorithm and the interval-varying LSI algorithm.

This paper is organized as follows: Section 2 introduces the identification model of the linear-in-parameters systems. Section 3 gives an LSI algorithm for comparisons. A D-LSI algorithm for the linear-in-parameters systems is developed in Sect. 4. Section 5 describes the parameter estimation problem with missing data and proposes an interval-varying LSI algorithm. Section 6 derives an interval-varying D-LSI algorithm to reduce computational load. The effectiveness of the proposed algorithms is illustrated by two simulation examples in Sect. 7. Finally, Sect. 8 gives some conclusions.

2 System Description and Identification Model

Let us introduce some notation. “ $A =: X$ ” or “ $X := A$ ” stands for “ A is defined as X ”; $\hat{\boldsymbol{\vartheta}}(t)$ denotes the estimate of $\boldsymbol{\vartheta}$ at time t ; the norm of a matrix (or a column vector) \mathbf{X} is defined by $\|\mathbf{X}\|^2 := \text{tr}[\mathbf{X}\mathbf{X}^T]$; $\mathbf{1}_n$ stands for an n -dimensional column vector whose elements are all 1; the superscript T denotes the matrix transpose.

Consider the linear-in-parameters system which can be expressed as

$$A(z)y(t) = \frac{\boldsymbol{\phi}^T(t)}{F(z)}\boldsymbol{\theta} + v(t), \quad (1)$$

where $y(t) \in \mathbb{R}$ is the measured output, $\boldsymbol{\phi}(t) \in \mathbb{R}^m$ is the information vector consisting of the system input–output data, $\boldsymbol{\theta} \in \mathbb{R}^m$ is the parameter vector to be estimated, $v(t) \in \mathbb{R}$ is the random white noise with zero mean and variance σ^2 , $A(z)$ and $F(z)$ with known orders n_a and n_f are polynomials in the unit backward shift operator z^{-1} with the property $z^{-1}y(t) = y(t-1)$, and defined by

$$\begin{aligned} A(z) &:= 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_{n_a}z^{-n_a}, \quad a_i \in \mathbb{R}, \\ F(z) &:= 1 + f_1z^{-1} + f_2z^{-2} + \cdots + f_{n_f}z^{-n_f}, \quad f_i \in \mathbb{R}. \end{aligned}$$

The objective of this paper is to use the decomposition technique to derive iterative methods for estimating the parameters $\boldsymbol{\theta}$, a_i and f_i in (1) from observation data for reducing the computational load. Without loss of generality, assume that $\boldsymbol{\phi}(t) = \mathbf{0}$, $y(t) = 0$ and $v(t) = 0$ for $t \leq 0$.

Define the parameter vectors and the information vectors,

$$\begin{aligned} \boldsymbol{\vartheta} &:= [\mathbf{a}^T, \mathbf{f}^T, \boldsymbol{\theta}^T]^T \in \mathbb{R}^n, \quad n := n_a + n_f + m, \\ \mathbf{a} &:= [a_1, a_2, \dots, a_{n_a}]^T \in \mathbb{R}^{n_a}, \\ \mathbf{f} &:= [f_1, f_2, \dots, f_{n_f}]^T \in \mathbb{R}^{n_f}, \end{aligned}$$

$$\begin{aligned}\boldsymbol{\varphi}(t) &:= \left[\boldsymbol{\varphi}_y^T(t), \boldsymbol{\varphi}_x^T(t), \boldsymbol{\phi}^T(t) \right]^T \in \mathbb{R}^n, \\ \boldsymbol{\varphi}_y(t) &:= [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T \in \mathbb{R}^{n_a}, \\ \boldsymbol{\varphi}_x(t) &:= [-x(t-1), -x(t-2), \dots, -x(t-n_f)]^T \in \mathbb{R}^{n_f}.\end{aligned}$$

Define the intermediate variable

$$\begin{aligned}x(t) &:= \frac{\boldsymbol{\phi}^T(t)\boldsymbol{\theta}}{F(z)} \\ &= [1 - F(z)]x(t) + \boldsymbol{\phi}^T(t)\boldsymbol{\theta} \\ &= \boldsymbol{\varphi}_x^T(t)\boldsymbol{f} + \boldsymbol{\phi}^T(t)\boldsymbol{\theta}.\end{aligned}\quad (2)$$

Then, System (1) can be rewritten as

$$\begin{aligned}y(t) &= [1 - A(z)]y(t) + x(t) + v(t) \\ &= \boldsymbol{\varphi}_y^T(t)\boldsymbol{a} + \boldsymbol{\varphi}_x^T(t)\boldsymbol{f} + \boldsymbol{\phi}^T(t)\boldsymbol{\theta} + v(t) \\ &= \boldsymbol{\varphi}^T(t)\boldsymbol{\vartheta} + v(t).\end{aligned}\quad (3)$$

$$(4)$$

Equation (4) is the identification model of System (1), and its parameter vector $\boldsymbol{\vartheta}$ contains all the parameters $\boldsymbol{\theta}$, a_i and f_i of the system.

3 The Least Squares-Based Iterative Algorithm

In this section, we give a least squares-based iterative algorithm for comparisons.

Consider the newest p data from $j = t - p + 1$ to $j = t$ (p represents the data length). According to the identification model in (4), define a quadratic function:

$$J(\boldsymbol{\vartheta}) := \sum_{j=0}^{p-1} \left[y(t-j) - \boldsymbol{\varphi}^T(t-j)\boldsymbol{\vartheta} \right]^2.$$

Assume that the information matrix $\boldsymbol{\varphi}(t)$ is persistently exciting for large p . Minimizing the function $J(\boldsymbol{\vartheta})$, we can obtain the least squares estimate of the parameter vector $\boldsymbol{\vartheta}$:

$$\hat{\boldsymbol{\vartheta}}(t) = \left[\sum_{j=0}^{p-1} \boldsymbol{\varphi}(t-j)\boldsymbol{\varphi}^T(t-j) \right]^{-1} \sum_{j=0}^{p-1} \boldsymbol{\varphi}(t-j)y(t-j).\quad (5)$$

Notice that the estimate $\hat{\boldsymbol{\vartheta}}(t)$ in (5) is impossible to obtain directly because the information vector $\boldsymbol{\varphi}(t-j)$ contains the unknown term $x(t-i)$. Here, the approach is based on the hierarchical identification principle: let $k = 1, 2, 3, \dots$ be an iterative

variable, $\hat{\boldsymbol{\vartheta}}_k(t) := \begin{bmatrix} \hat{\boldsymbol{a}}_k(t) \\ \hat{\boldsymbol{f}}_k(t) \\ \hat{\boldsymbol{\theta}}_k(t) \end{bmatrix} \in \mathbb{R}^n$ be the iterative estimate of $\boldsymbol{\vartheta}$ at iteration k , use the estimate $\hat{x}_{k-1}(t-i)$ of $x(t-i)$ to construct the estimate $\hat{\boldsymbol{\varphi}}_{x,k}(t)$ of $\boldsymbol{\varphi}_x(t)$ at iteration k :

$$\hat{\boldsymbol{\varphi}}_{x,k}(t) := [-\hat{x}_{k-1}(t-1), -\hat{x}_{k-1}(t-2), \dots, -\hat{x}_{k-1}(t-n_f)]^T \in \mathbb{R}^{n_f},$$

and define the estimate of $\boldsymbol{\varphi}(t)$:

$$\hat{\boldsymbol{\varphi}}_k(t) := [\boldsymbol{\varphi}_y^T(t), \hat{\boldsymbol{\varphi}}_{x,k}^T(t), \boldsymbol{\phi}^T(t)]^T \in \mathbb{R}^n.$$

Replacing $\boldsymbol{\varphi}_x(t)$, $\boldsymbol{\theta}$ and \boldsymbol{f} in (2) with $\hat{\boldsymbol{\varphi}}_{x,k}(t)$, $\hat{\boldsymbol{\theta}}_k(t)$ and $\hat{\boldsymbol{f}}_k(t)$, respectively, the estimate $\hat{x}_k(t)$ of $x(t)$ can be computed by

$$\hat{x}_k(t) = \hat{\boldsymbol{\varphi}}_{x,k}^T(t) \hat{\boldsymbol{f}}_k(t) + \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}_k(t).$$

Replacing $\boldsymbol{\varphi}(t-j)$ in (5) with $\hat{\boldsymbol{\varphi}}_k(t-j)$, we can obtain the following least squares-based iterative (LSI) algorithm for estimating $\boldsymbol{\vartheta}$:

$$\hat{\boldsymbol{\vartheta}}_k(t) = \hat{\boldsymbol{S}}_k^{-1}(t) \sum_{j=0}^{p-1} \hat{\boldsymbol{\varphi}}_k(t-j) y(t-j), \tag{6}$$

$$\hat{\boldsymbol{S}}_k(t) := \sum_{j=0}^{p-1} \hat{\boldsymbol{\varphi}}_k(t-j) \hat{\boldsymbol{\varphi}}_k^T(t-j), \tag{7}$$

$$\hat{\boldsymbol{\varphi}}_k(t) = [\boldsymbol{\varphi}_y^T(t), \hat{\boldsymbol{\varphi}}_{x,k}^T(t), \boldsymbol{\phi}^T(t)]^T, \tag{8}$$

$$\boldsymbol{\varphi}_y(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \tag{9}$$

$$\hat{\boldsymbol{\varphi}}_{x,k}(t) = [-\hat{x}_{k-1}(t-1), -\hat{x}_{k-1}(t-2), \dots, -\hat{x}_{k-1}(t-n_f)]^T, \tag{10}$$

$$\hat{x}_k(t) = \hat{\boldsymbol{\varphi}}_{x,k}^T(t) \hat{\boldsymbol{f}}_k(t) + \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}_k(t), \tag{11}$$

$$\hat{\boldsymbol{\vartheta}}_k(t) = \begin{bmatrix} \hat{\boldsymbol{a}}_k(t) \\ \hat{\boldsymbol{f}}_k(t) \\ \hat{\boldsymbol{\theta}}_k(t) \end{bmatrix}. \tag{12}$$

The LSI parameter estimation algorithm is able to make full use of all the input–output data at each iteration, and thus, the parameter estimation accuracy can be greatly improved.

4 The Decomposition-Based LSI Algorithm

The LSI algorithm can improve the parameter estimation accuracy, but the disadvantage is that it needs heavy computational load for large-scale systems. By means of

the hierarchical identification principle, the following derives a D-LSI algorithm to improve the computational efficiency.

The identification model in (3) includes the known information vectors $\varphi_y(t)$ and $\phi(t)$, and the unknown information vector $\varphi_x(t)$. Define a new information vector

$$\varphi_1(t) := \left[\varphi_y^T(t), \phi^T(t) \right]^T \in \mathbb{R}^{n_a+m}, \quad (13)$$

and the corresponding parameter vector

$$\theta_1 := \left[a^T, \theta^T \right]^T \in \mathbb{R}^{n_a+m}.$$

Based on the hierarchical identification principle [3], by defining two intermediate variables

$$y_1(t) := y(t) - \varphi_x^T(t)f, \quad (14)$$

$$y_2(t) := y(t) - \varphi_1^T(t)\theta_1, \quad (15)$$

we can decompose the identification model in (3) into the following two fictitious sub-models:

$$y_1(t) = \varphi_1^T(t)\theta_1 + v(t), \quad (16)$$

$$y_2(t) = \varphi_x^T(t)f + v(t). \quad (17)$$

The parameter vectors $\theta_1 = \begin{bmatrix} a \\ \theta \end{bmatrix}$ and f to be identified are included in the two sub-models, respectively.

According to Eqs. (16) and (17), minimizing the quadratic functions

$$J_1(\theta_1) := \sum_{j=0}^{p-1} \left[y_1(t-j) - \varphi_1^T(t-j)\theta_1 \right]^2,$$

$$J_2(f) := \sum_{j=0}^{p-1} \left[y_2(t-j) - \varphi_x^T(t-j)f \right]^2,$$

we can obtain the following least squares estimates of the parameter vectors θ and f :

$$\hat{\theta}_1(t) = \left[\sum_{j=0}^{p-1} \varphi_1(t-j)\varphi_1^T(t-j) \right]^{-1} \sum_{j=0}^{p-1} [\varphi_1(t-j)y_1(t-j)], \quad (18)$$

$$\hat{f}(t) = \left[\sum_{j=0}^{p-1} \varphi_x(t-j)\varphi_x^T(t-j) \right]^{-1} \sum_{j=0}^{p-1} [\varphi_x(t-j)y_2(t-j)]. \quad (19)$$

Here, we have used the assumption that the information vectors $\varphi_1(t)$ and $\varphi_x(t)$ are persistently exciting for large p . Substituting (14) and (15) into (18) and (19), respectively, we have

$$\hat{\theta}_1(t) = \left[\sum_{j=0}^{p-1} \varphi_1(t-j)\varphi_1^T(t-j) \right]^{-1} \sum_{j=0}^{p-1} \varphi_1(t-j) \left[y(t-j) - \varphi_x^T(t-j)\mathbf{f} \right], \tag{20}$$

$$\hat{\mathbf{f}}(t) = \left[\sum_{j=0}^{p-1} \varphi_x(t-j)\varphi_x^T(t-j) \right]^{-1} \sum_{j=0}^{p-1} \varphi_x(t-j) \left[y(t-j) - \varphi_1^T(t-j)\theta_1 \right]. \tag{21}$$

However, the information vector $\varphi_x(t)$ contains the unknown term $x(t-i)$, thus the algorithm in (20) and (21) cannot be implemented. Similarly, we use the hierarchical identification principle to solve this problem: let $\hat{\theta}_{1,k}(t) := [\hat{\mathbf{a}}_k^T(t), \hat{\theta}_k^T(t)]^T \in \mathbb{R}^{n_a+m}$ be the iterative estimate of θ_1 at iteration k , $\hat{\varphi}_{x,k}(t)$ be the estimate of $\varphi_x(t)$ by replacing $x(t-i)$ with its estimate $\hat{x}_{k-1}(t-i)$ at iteration $k-1$.

Replacing $\varphi_x(t)$, \mathbf{f} and θ_1 in (20) and (21) with their corresponding estimates $\hat{\varphi}_{x,k}(t)$, $\hat{\mathbf{f}}_{k-1}(t)$ and $\hat{\theta}_{1,k-1}(t)$, respectively, we can summarize the decomposition-based LSI (D-LSI) algorithm of the linear-in-parameters systems as

$$\hat{\theta}_{1,k}(t) = \mathbf{S}_1^{-1}(t) \sum_{j=0}^{p-1} \varphi_1(t-j) \left[y(t-j) - \hat{\varphi}_{x,k}^T(t-j)\hat{\mathbf{f}}_{k-1}(t) \right], \tag{22}$$

$$\mathbf{S}_1(t) := \sum_{j=0}^{p-1} \varphi_1(t-j)\varphi_1^T(t-j), \tag{23}$$

$$\hat{\mathbf{f}}_k(t) = \hat{\mathbf{S}}_{2,k}^{-1}(t) \sum_{j=0}^{p-1} \hat{\varphi}_{x,k}(t-j) \left[y(t-j) - \varphi_1^T(t-j)\hat{\theta}_{1,k-1}(t) \right], \tag{24}$$

$$\hat{\mathbf{S}}_{2,k}(t) := \sum_{j=0}^{p-1} \hat{\varphi}_{x,k}(t-j)\hat{\varphi}_{x,k}^T(t-j), \tag{25}$$

$$\varphi_1(t) = \left[\varphi_y^T(t), \boldsymbol{\phi}^T(t) \right]^T, \tag{26}$$

$$\varphi_y(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \tag{27}$$

$$\hat{\varphi}_{x,k}(t) = [-\hat{x}_{k-1}(t-1), -\hat{x}_{k-1}(t-2), \dots, -\hat{x}_{k-1}(t-n_f)]^T, \tag{28}$$

$$\hat{x}_k(t) = \hat{\varphi}_x^T(t)\hat{\mathbf{f}}_k(t) + \boldsymbol{\phi}^T(t)\hat{\theta}_k(t), \tag{29}$$

$$\hat{\theta}_{1,k}(t) = \begin{bmatrix} \hat{\mathbf{a}}_k(t) \\ \hat{\theta}_k(t) \end{bmatrix}, \tag{30}$$

$$\hat{\mathbf{a}}_k(t) := [\hat{a}_{1,k}(t), \hat{a}_{2,k}(t), \dots, \hat{a}_{n_a,k}(t)]^T, \tag{31}$$

$$\hat{\mathbf{f}}_k(t) := \left[\hat{f}_{1,k}(t), \hat{f}_{2,k}(t), \dots, \hat{f}_{n_f,k}(t) \right]^T. \quad (32)$$

In the D-LSI algorithm, the dimensions of the covariance matrices $\mathbf{S}_1^{-1}(t)$ and $\hat{\mathbf{S}}_{2,k}^{-1}(t)$ in (22) and (24) are $(n_a + m) \times (n_a + m)$ and $n_f \times n_f$. In the LSI algorithm, the dimension of the covariance matrix $\hat{\mathbf{S}}_k^{-1}(t)$ in (6) is $(n_a + m + n_f) \times (n_a + m + n_f)$. Thus, the D-LSI algorithm requires less computational cost than the LSI algorithm.

The steps involved in the D-LSI algorithm to compute the parameter estimation vectors $\hat{\boldsymbol{\theta}}_{1,k}(t)$ and $\hat{\mathbf{f}}_k(t)$ are listed in the following.

1. Set the data length p , let $t = p$, collect the observation data $\{y(i), \boldsymbol{\phi}(i) : i = 0, 1, \dots, p - 1\}$, and set a small positive number ε .
2. Collect the observation data $y(t)$ and $\boldsymbol{\phi}(t)$ and form $\boldsymbol{\varphi}_y(t)$ using (27) and $\boldsymbol{\varphi}_1(t)$ using (26).
3. Let $k = 1$, set the initial values $\hat{\boldsymbol{\theta}}_{1,0}(0) = \mathbf{1}_{n_a+m}/p_0$, $\hat{\mathbf{f}}_0(0) = \mathbf{1}_{n_f}/p_0$, $\hat{x}_0(t-i) = 1/p_0$ ($i = 1, 2, \dots, n_f$), $p_0 = 10^6$.
4. Form $\hat{\boldsymbol{\varphi}}_{x,k}(t)$ using (28), compute $\mathbf{S}_1(t)$ and $\hat{\mathbf{S}}_{2,k}(t)$ using (23) and (25).
5. Update the parameter estimation vectors $\hat{\boldsymbol{\theta}}_{1,k}(t)$ and $\hat{\mathbf{f}}_k(t)$ using (22) and (24), respectively.
6. Read $\hat{\boldsymbol{\theta}}_k(t)$ from $\hat{\boldsymbol{\theta}}_{1,k}(t)$ using (30), and compute $\hat{x}_k(t)$ using (29).
7. Compare $\hat{\boldsymbol{\theta}}_{1,k}(t)$ with $\hat{\boldsymbol{\theta}}_{1,k-1}(t)$ and $\hat{\mathbf{f}}_k(t)$ with $\hat{\mathbf{f}}_{k-1}(t)$: if

$$\|\hat{\boldsymbol{\theta}}_{1,k}(t) - \hat{\boldsymbol{\theta}}_{1,k-1}(t)\| + \|\hat{\mathbf{f}}_k(t) - \hat{\mathbf{f}}_{k-1}(t)\| \leq \varepsilon,$$

obtain k , $\hat{\boldsymbol{\theta}}_{1,k}(t)$ and $\hat{\mathbf{f}}_k(t)$, increase t by 1, and go to Step 2; otherwise, increase k by 1, and go to Step 4.

5 The Interval-Varying LSI Algorithm

This section derives an interval-varying LSI algorithm to solve the identification problems of systems with missing data.

In many applications, there are many reasons for missing sampled data to arise. In general, a missing-data system implies that most data are available and few data are missing over a period of time. The following considers such a system with missing data that the inputs are normally available at every instant t because the input signals are usually generated by digital computers in practice, and only a small number of data are missing, as shown in Fig. 1 [8, 12], where “+” stands for missing data or bad data (outliers or unbelievable data), e.g., the outputs $y(3)$, $y(8)$, $y(9)$, $y(23)$, ... are missing samples and $y(15)$, ... are unbelievable samples.

For convenience, we define an integer sequence $\{t_s, s = 0, 1, 2, \dots\}$ satisfying

$$0 = t_0 < t_1 < t_2 < t_3 < \dots < t_{s-1} < t_s < \dots$$

with $t_s^* := t_s - t_{s-1} \geq 1$, such that $y(t)$ and $\boldsymbol{\varphi}_y(t)$ are available only when $t = t_s$ ($s = 0, 1, 2, \dots$), or equivalently, the data set $\{y(t_s), \boldsymbol{\varphi}_y(t_s) : s = 0, 1, 2, \dots\}$ contains

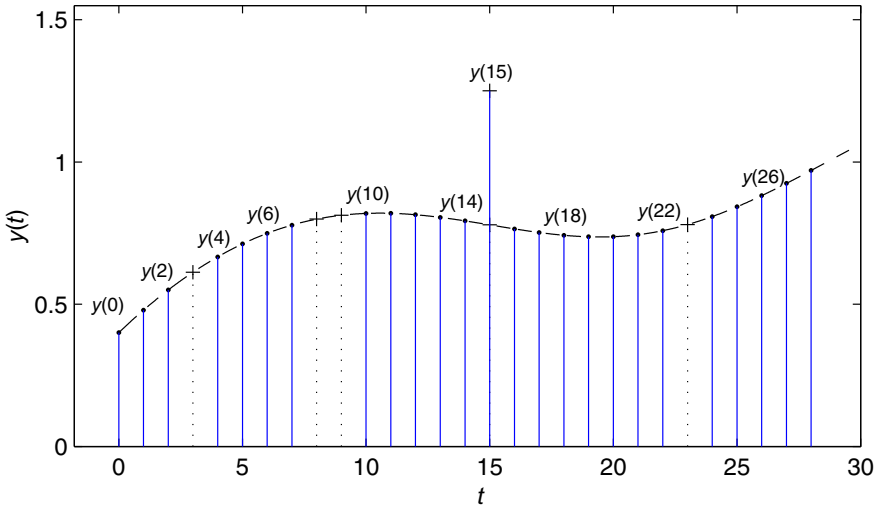


Fig. 1 A missing output data pattern

all available outputs. For instance, for the missing-data pattern in Fig. 1, when the order $n_a = 3$, define the integer sequence $\{t_0, t_1, t_2, \dots, t_9, \dots\}$, for $t_0 = 0, t_1 = 7, t_2 = 13, \dots, t_9 = 28, \dots$, i.e., $\{y(t_0), \varphi_y(t_0)\}, \{y(t_1), \varphi_y(t_1)\}, \{y(t_2), \varphi_y(t_2)\}, \dots, \{y(t_9), \varphi_y(t_9)\}, \dots$ are available.

Replacing t in (4) with t_s gives

$$y(t_s) = \varphi^T(t_s) \vartheta + v(t_s) \tag{33}$$

with

$$\begin{aligned} \varphi(t_s) &= [\varphi_y^T(t_s), \varphi_x^T(t_s), \phi^T(t_s)]^T, \\ \varphi_y(t_s) &= [-y(t_s - 1), -y(t_s - 2), \dots, -y(t_s - n_a)]^T, \\ \varphi_x(t_s) &= [-x(t_s - 1), -x(t_s - 2), \dots, -x(t_s - n_f)]^T. \end{aligned} \tag{34}$$

Consider p data from $i = t_{s-p+1}$ to $i = t_s$. Define the stacked output vector $Y(t_s)$ and the stacked information matrix $\Psi(t_s)$ as

$$Y(t_s) := \begin{bmatrix} y(t_s) \\ y(t_{s-1}) \\ \vdots \\ y(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^p, \quad \Psi(t_s) := \begin{bmatrix} \varphi^T(t_s) \\ \varphi^T(t_{s-1}) \\ \vdots \\ \varphi^T(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^{p \times n}.$$

Assume that the information vector $\varphi(t_s)$ is persistently exciting for large p , that is, $[\Psi^T(t_s)\Psi(t_s)]$ is non-singular. The difficulty is that the information vector $\varphi_x(t_s)$ in $\Psi(t_s)$ contains the unknown variable $x(t_s - i)$. Replacing $x(t_s - i)$ in (34) with their estimates $\hat{x}_{k-1}(t_s - i)$ at iteration $k - 1$, and minimizing the quadratic function

$$J(\vartheta) := \|\mathbf{Y}(t_s) - \Psi(t_s)\vartheta\|^2,$$

we can obtain the following interval-varying least squares-based iterative (V-LSI) algorithm for estimating the parameter vector ϑ :

$$\hat{\vartheta}_k(t_s) = \left[\hat{\Psi}_k^T(t_s) \hat{\Psi}_k(t_s) \right]^{-1} \hat{\Psi}_k^T(t_s) \mathbf{Y}(t_s), \tag{35}$$

$$\hat{\vartheta}_k(t) = \hat{\vartheta}_k(t_s), \quad t \in T_s := \{t_s, t_s + 1, \dots, t_{s+1} - 1\}, \tag{36}$$

$$\mathbf{Y}(t_s) = [y(t_s), y(t_{s-1}), \dots, y(t_{s-p+1})]^T, \tag{37}$$

$$\hat{\Psi}_k(t_s) = [\hat{\phi}_k(t_s), \hat{\phi}_k(t_{s-1}), \dots, \hat{\phi}_k(t_{s-p+1})]^T, \tag{38}$$

$$\hat{\phi}_k(t_s) = [\phi_y^T(t_s), \hat{\phi}_{x,k}^T(t_s), \phi^T(t_s)]^T, \tag{39}$$

$$\phi_y(t_s) = [-y(t_s - 1), -y(t_s - 2), \dots, -y(t_s - n_a)]^T, \tag{40}$$

$$\hat{\phi}_{x,k}(t_s) = [-\hat{x}_{k-1}(t_s - 1), -\hat{x}_{k-1}(t_s - 2), \dots, -\hat{x}_{k-1}(t_s - n_f)]^T, \tag{41}$$

$$\hat{\vartheta}_k(t_s) = [\hat{a}_k^T(t_s), \hat{f}_k^T(t_s), \hat{\theta}_k^T(t_s)]^T, \tag{42}$$

$$\begin{aligned} \hat{x}_k(j) &= \hat{\phi}_{x,k}^T(j) \hat{f}_k(t_s) + \phi^T(j) \hat{\vartheta}_k(t_s), \quad j \in [t_1, t_{s+1}], \\ \hat{x}_k(i) &= 1/p_0, \quad i \leq t_1 - 1. \end{aligned} \tag{43}$$

We simply hold the parameter estimate $\hat{\vartheta}_k(t)$ remains unchanged over the interval $[t_s, t_{s+1} - 1]$.

6 The Interval-Varying D-LSI Algorithm

In the following, we study an interval-varying D-LSI algorithm based on the decomposition technique to reduce computational cost.

Replacing t in (14)–(17) with t_s gives

$$\begin{aligned} y_1(t_s) &= y(t_s) - \phi_x^T(t_s) \mathbf{f} \\ &= \phi_1^T(t_s) \theta_1 + v(t_s), \\ y_2(t_s) &= y(t_s) - \phi_1^T(t_s) \theta_1 \\ &= \phi_x^T(t_s) \mathbf{f} + v(t_s). \end{aligned}$$

Define the stacked output vectors $\mathbf{Y}(t_s)$, $\mathbf{Y}_1(t_s)$ and $\mathbf{Y}_2(t_s)$ and the stacked information matrices $\Psi_1(t_s)$ and $\Psi_x(t_s)$ as

$$\mathbf{Y}(t_s) := \begin{bmatrix} y(t_s) \\ y(t_{s-1}) \\ \vdots \\ y(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^p, \quad \mathbf{Y}_1(t_s) := \begin{bmatrix} y_1(t_s) \\ y_1(t_{s-1}) \\ \vdots \\ y_1(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^p,$$

$$\begin{aligned}
 \mathbf{Y}_2(t_s) &:= \begin{bmatrix} y_2(t_s) \\ y_2(t_{s-1}) \\ \vdots \\ y_2(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^p, \quad \boldsymbol{\Psi}_1(t_s) := \begin{bmatrix} \boldsymbol{\varphi}_1^T(t_s) \\ \boldsymbol{\varphi}_1^T(t_{s-1}) \\ \vdots \\ \boldsymbol{\varphi}_1^T(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^{p \times (n_a+m)}, \\
 \boldsymbol{\Psi}_x(t_s) &:= \begin{bmatrix} \boldsymbol{\varphi}_x^T(t_s) \\ \boldsymbol{\varphi}_x^T(t_{s-1}) \\ \vdots \\ \boldsymbol{\varphi}_x^T(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^{p \times n_f}.
 \end{aligned}$$

Define two quadratic functions:

$$\begin{aligned}
 J_1(\boldsymbol{\theta}_1) &:= \|\mathbf{Y}_1(t_s) - \boldsymbol{\Psi}_1(t_s)\boldsymbol{\theta}_1\|^2, \\
 J_2(\mathbf{f}) &:= \|\mathbf{Y}_2(t_s) - \boldsymbol{\Psi}_x(t_s)\mathbf{f}\|^2.
 \end{aligned}$$

Assume that the information vectors $\boldsymbol{\varphi}_1(t_s)$ and $\boldsymbol{\varphi}_x(t_s)$ are persistently exciting for large p , that is, $[\boldsymbol{\Psi}_1^T(t_s)\boldsymbol{\Psi}_1(t_s)]$ and $[\boldsymbol{\Psi}_x^T(t_s)\boldsymbol{\Psi}_x(t_s)]$ are non-singular. Letting the partial derivatives of $J_1(\boldsymbol{\theta}_1)$ and $J_2(\mathbf{f})$ with respect to $\boldsymbol{\theta}_1$ and \mathbf{f} be zero leads to the following least squares estimates of the parameter vectors $\boldsymbol{\theta}_1$ and \mathbf{f} :

$$\begin{aligned}
 \hat{\boldsymbol{\theta}}_1(t_s) &= [\boldsymbol{\Psi}_1^T(t_s)\boldsymbol{\Psi}_1(t_s)]^{-1} \boldsymbol{\Psi}_1^T(t_s)\mathbf{Y}_1(t_s) \\
 &= [\boldsymbol{\Psi}_1^T(t_s)\boldsymbol{\Psi}_1(t_s)]^{-1} \boldsymbol{\Psi}_1^T(t_s)[\mathbf{Y}(t_s) - \boldsymbol{\Psi}_x(t_s)\mathbf{f}], \tag{44}
 \end{aligned}$$

$$\begin{aligned}
 \hat{\mathbf{f}}(t_s) &= [\boldsymbol{\Psi}_x^T(t_s)\boldsymbol{\Psi}_x(t_s)]^{-1} \boldsymbol{\Psi}_x^T(t_s)\mathbf{Y}_2(t_s) \\
 &= [\boldsymbol{\Psi}_x^T(t_s)\boldsymbol{\Psi}_x(t_s)]^{-1} \boldsymbol{\Psi}_x^T(t_s)[\mathbf{Y}(t_s) - \boldsymbol{\Psi}_1(t_s)\boldsymbol{\theta}_1]. \tag{45}
 \end{aligned}$$

However, we can see that the right-hand sides of Eqs. (44) and (45) contain the unknown parameters $\boldsymbol{\theta}_1$ and \mathbf{f} , and the information vector $\boldsymbol{\varphi}_x(t_s)$ in $\boldsymbol{\Psi}_x(t_s)$ contains the unknown term $x(t_s - i)$, so the estimates $\hat{\boldsymbol{\theta}}_1(t_s)$ and $\hat{\mathbf{f}}(t_s)$ are impossible to compute directly. Here, we use the hierarchical identification principle to solve this problem: let

$\hat{\boldsymbol{\theta}}_{1,k}(t_s) := \begin{bmatrix} \hat{\mathbf{a}}_k(t_s) \\ \hat{\boldsymbol{\theta}}_k(t_s) \end{bmatrix}$ and $\hat{\mathbf{f}}_k(t_s)$ be the iterative estimates of $\boldsymbol{\theta}_1 = \begin{bmatrix} \mathbf{a} \\ \boldsymbol{\theta} \end{bmatrix}$ and \mathbf{f} at iteration k , respectively, and $\hat{x}_k(t_s - i)$ be the estimate of $x(t_s - i)$. Define

$$\begin{aligned}
 \hat{\boldsymbol{\varphi}}_{x,k}(t_s) &:= [-\hat{x}_{k-1}(t_s - 1), -\hat{x}_{k-1}(t_s - 2), \dots, -\hat{x}_{k-1}(t_s - n_f)]^T \in \mathbb{R}^{n_f}, \\
 \hat{\boldsymbol{\Psi}}_{x,k}(t_s) &:= \begin{bmatrix} \hat{\boldsymbol{\varphi}}_{x,k}^T(t_s) \\ \hat{\boldsymbol{\varphi}}_{x,k}^T(t_{s-1}) \\ \vdots \\ \hat{\boldsymbol{\varphi}}_{x,k}^T(t_{s-p+1}) \end{bmatrix} \in \mathbb{R}^{p \times n_f}.
 \end{aligned}$$

Replacing θ , f and $\varphi_x(t_s)$ in (2) with $\hat{\theta}_k(t_s)$, $\hat{f}_k(t_s)$ and $\hat{\varphi}_{x,k}(j)$, the estimate $\hat{x}_k(j)$ of $x(j)$ can be computed by

$$\hat{x}_k(j) = \hat{\varphi}_{x,k}^T(j) \hat{f}_k(t_s) + \phi^T(j) \hat{\theta}_k(t_s).$$

Replacing $\Psi_x(t_s)$, θ_1 and f in (44) and (45) with their corresponding estimates $\hat{\Psi}_{x,k}(t_s)$, $\hat{\theta}_{1,k-1}(t_s)$ and $\hat{f}_{k-1}(t_s)$, respectively, we can summarize the interval-varying D-LSI algorithm of computing $\hat{\theta}_{1,k}(t_s)$ and $\hat{f}_k(t_s)$ as

$$\hat{\theta}_{1,k}(t_s) = [\Psi_1^T(t_s) \Psi_1(t_s)]^{-1} \Psi_1^T(t_s) [Y(t_s) - \hat{\Psi}_{x,k}(t_s) \hat{f}_{k-1}(t_s)], \tag{46}$$

$$\hat{\theta}_{1,k}(t) = \hat{\theta}_{1,k}(t_s), \quad t \in T_s := \{t_s, t_s + 1, \dots, t_{s+1} - 1\}, \tag{47}$$

$$\hat{f}_k(t_s) = [\hat{\Psi}_{x,k}^T(t_s) \hat{\Psi}_{x,k}(t_s)]^{-1} \hat{\Psi}_{x,k}^T(t_s) [Y(t_s) - \Psi_1(t_s) \hat{\theta}_{1,k-1}(t_s)], \tag{48}$$

$$\hat{f}_k(t) = \hat{f}_k(t_s), \tag{49}$$

$$Y(t_s) = [y(t_s), y(t_{s-1}), \dots, y(t_{s-p+1})]^T, \tag{50}$$

$$\Psi_1(t_s) = [\varphi_1(t_s), \varphi_1(t_{s-1}), \dots, \varphi_1(t_{s-p+1})]^T, \tag{51}$$

$$\hat{\Psi}_{x,k}(t_s) = [\hat{\varphi}_{x,k}(t_s), \hat{\varphi}_{x,k}(t_{s-1}), \dots, \hat{\varphi}_{x,k}(t_{s-p+1})]^T, \tag{52}$$

$$\varphi_1(t_s) = \begin{bmatrix} \varphi_y(t_s) \\ \phi(t_s) \end{bmatrix}, \tag{53}$$

$$\varphi_y(t_s) = [-y(t_s - 1), -y(t_s - 2), \dots, -y(t_s - n_a)]^T, \tag{54}$$

$$\hat{\varphi}_{x,k}(t_s) = [-\hat{x}_{k-1}(t_s - 1), -\hat{x}_{k-1}(t_s - 2), \dots, -\hat{x}_{k-1}(t_s - n_f)]^T, \tag{55}$$

$$\hat{x}_k(j) = \hat{\varphi}_{x,k}^T(j) \hat{f}_k(t_s) + \phi^T(j) \hat{\theta}_k(t_s), \quad j \in [t_1, t_{s+1} - 1], \tag{56}$$

$$\hat{\theta}_{1,k}(t_s) = \begin{bmatrix} \hat{a}_k(t_s) \\ \hat{\theta}_k(t_s) \end{bmatrix}, \tag{57}$$

$$\hat{a}_k(t_s) = [\hat{a}_{1,k}(t_s), \hat{a}_{2,k}(t_s), \dots, \hat{a}_{n_a,k}(t_s)]^T, \tag{58}$$

$$\hat{f}_k(t_s) = [\hat{f}_{1,k}(t_s), \hat{f}_{2,k}(t_s), \dots, \hat{f}_{n_f,k}(t_s)]^T. \tag{59}$$

To initialize the algorithm, we take $\hat{\theta}_{1,0}(t_0)$ and $\hat{f}_0(t_0)$ as real vectors with small positive entries, e.g., $\hat{\theta}_{1,0}(t_0) = \mathbf{1}_{n_a+m}/p_0$, $\hat{f}_0(t_0) = \mathbf{1}_{n_f}/p_0$, $\hat{x}_0(i) = 1/p_0$ ($i \leq t_1 - 1$), $p_0 = 10^6$. The parameter estimates $\hat{\theta}_{1,k}(t)$ and $\hat{f}_k(t)$ in (47) and (49) remain unchanged over the interval $[t_s, t_{s+1} - 1]$.

The interval-varying D-LSI algorithm (which is abbreviated as the V-D-LSI algorithm) uses the data over the finite data window with the length p , thus the V-D-LSI algorithm can track time-varying parameters and be used for online identification. The interval-varying identification algorithms are proposed for missing-data systems but can be extended to systems with scarce measurements.

7 Examples

Example 1 Consider the following linear-in-parameters system:

$$A(z)y(t) = \frac{\phi^T(t)}{F(z)}\theta + v(t),$$

$$A(z) = 1 + a_1z^{-1} + a_2z^{-2} = 1 + 0.27z^{-1} + 0.75z^{-2},$$

$$F(z) = 1 + f_1z^{-1} + f_2z^{-2} = 1 - 0.31z^{-1} - 0.44z^{-2},$$

$$\theta = [b_1, b_2]^T = [-0.56, 0.91]^T,$$

$$\vartheta = [a_1, a_2, f_1, f_2, b_1, b_2]^T = [0.27, 0.75, -0.31, 0.44, -0.56, 0.91]^T.$$

In simulation, $\{\phi(t)\}$ is taken as an uncorrelated persistent excitation vector sequence with zero mean and unit variance, $\{v(t)\}$ as a white noise sequence with zero mean and different variances $\sigma^2 = 0.10^2$ and $\sigma^2 = 0.50^2$, respectively. Take the data length $t = p = L_e = 3000$ data and apply the LSI algorithm and the D-LSI algorithm to identify this example system, the parameter estimates and their errors δ versus iteration k are given in Tables 1 and 2 and Figs. 2 and 3 where the parameter estimation error is defined as $\delta := \|\hat{\vartheta}_k - \vartheta\|/\|\vartheta\| \times 100\%$.

From Tables 1 and 2 and Figs. 2 and 3, we can draw the following conclusions.

- The estimation errors given by the LSI algorithm and the D-LSI algorithm become smaller (in general) as iteration k increases or the noise variance σ^2 decreases—see Tables 1 and 2 and Figs. 2 and 3.
- The parameter estimates given by the LSI algorithm and the D-LSI algorithm are very close to the true parameters for large k —see Tables 1 and 2.

Table 1 The LSI parameter estimates and errors versus iteration k

σ^2	k	a_1	a_2	f_1	f_2	b_1	b_2	δ (%)
0.10 ²	1	0.12130	0.86000	-0.00337	0.00290	-0.56235	0.82223	39.77604
	2	0.14538	0.77527	-0.22870	0.29238	-0.56294	0.93341	14.77629
	5	0.26612	0.74741	-0.30777	0.43972	-0.55836	0.90819	0.39812
	10	0.27048	0.74969	-0.31134	0.44190	-0.55829	0.90758	0.26492
	20	0.27048	0.74970	-0.31134	0.44191	-0.55829	0.90758	0.26512
0.50 ²	1	0.13918	0.84802	-0.00655	0.00878	-0.55630	0.80354	39.10793
	2	0.16413	0.77135	-0.26148	0.29224	-0.55690	0.93013	13.24651
	5	0.26898	0.74869	-0.31399	0.44738	-0.55145	0.89849	1.16127
	10	0.27235	0.74989	-0.31740	0.44807	-0.55140	0.89829	1.27600
	20	0.27234	0.74990	-0.31737	0.44809	-0.55140	0.89828	1.27613
True values		0.27000	0.75000	-0.31000	0.44000	-0.56000	0.91000	

Table 2 The D-LSI parameter estimates and errors versus iteration k

σ^2	k	a_1	a_2	f_1	f_2	b_1	b_2	δ (%)
0.10^2	1	0.12177	0.86005	-0.00401	0.00402	-0.56182	0.82139	39.69414
	2	0.12224	0.85881	-0.11861	0.25627	-0.56184	0.82332	23.23453
	5	0.22099	0.73968	-0.25638	0.41860	-0.55949	0.88763	5.53938
	10	0.26770	0.74829	-0.30816	0.44128	-0.55843	0.90584	0.39957
	20	0.27050	0.74976	-0.31143	0.44184	-0.55835	0.90758	0.26330
0.50^2	1	0.13974	0.84813	-0.00767	0.00955	-0.55586	0.80239	39.03257
	2	0.14043	0.84569	-0.13753	0.27100	-0.55590	0.80614	21.44534
	5	0.24006	0.74297	-0.28309	0.43294	-0.55289	0.88201	3.51513
	10	0.27174	0.74971	-0.31695	0.44783	-0.55176	0.89762	1.27090
	20	0.27254	0.75015	-0.31788	0.44791	-0.55173	0.89826	1.27761
True values		0.27000	0.75000	-0.31000	0.44000	-0.56000	0.91000	

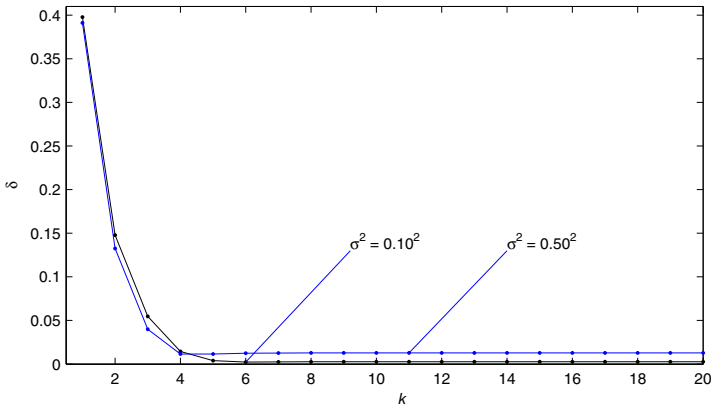
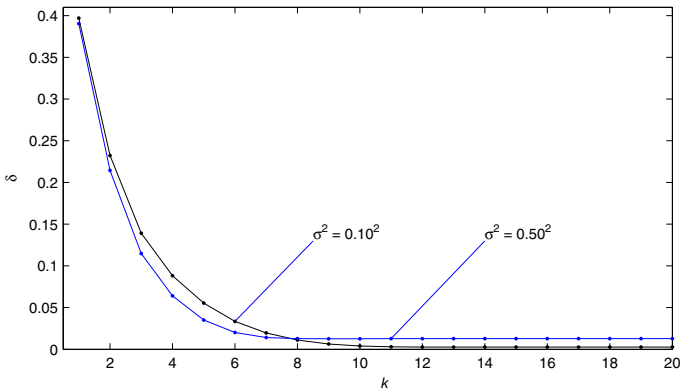
**Fig. 2** The LSI estimation errors versus k with $\sigma^2 = 0.10^2$ and $\sigma^2 = 0.50^2$ **Fig. 3** The D-LSI estimation errors versus k with $\sigma^2 = 0.10^2$ and $\sigma^2 = 0.50^2$

Table 3 The V-LSI parameter estimates and errors versus iteration k

σ^2	k	a_1	a_2	f_1	f_2	b_1	b_2	δ (%)
0.50 ²	1	-1.10186	0.53319	-0.00827	0.00558	0.53791	1.16373	37.75946
	2	-1.01136	0.42678	-0.42767	0.25248	0.54744	0.98116	18.37233
	5	-1.11555	0.43635	-0.40614	0.43932	0.55198	0.93555	6.40856
	10	-1.16423	0.45656	-0.33473	0.48704	0.55060	0.94772	2.39188
	15	-1.16632	0.45156	-0.33942	0.51080	0.55093	0.94364	1.23667
	20	-1.16497	0.44977	-0.34347	0.51439	0.55107	0.94211	1.01806
1.00 ²	1	-1.11442	0.51426	-0.01346	0.00942	0.53073	1.15876	37.17634
	2	-1.11032	0.45120	-0.34239	0.38281	0.54239	0.97659	8.90486
	5	-1.15952	0.44729	-0.35694	0.51329	0.54587	0.94197	1.32722
	10	-1.16450	0.44895	-0.34075	0.51349	0.54520	0.94783	1.48817
	15	-1.16474	0.44887	-0.34008	0.51416	0.54518	0.94804	1.49924
	20	-1.16474	0.44885	-0.34008	0.51424	0.54518	0.94803	1.49831
True values		-1.17000	0.45000	-0.35000	0.52000	0.56000	0.93000	

Table 4 The V-D-LSI parameter estimates and errors versus iteration k

σ^2	k	a_1	a_2	f_1	f_2	b_1	b_2	δ (%)
0.50 ²	1	-1.10204	0.53312	-0.00790	0.00396	0.53969	1.16329	37.82652
	2	-1.10010	0.53112	-0.26760	0.21451	0.53985	1.16005	22.90660
	5	-1.03524	0.36156	-0.42477	0.47357	0.55009	0.97497	10.68003
	10	-1.13298	0.41964	-0.36772	0.51205	0.55198	0.94033	3.00632
	15	-1.15968	0.44481	-0.34880	0.51371	0.55235	0.93979	1.02295
	20	-1.16407	0.44906	-0.34565	0.51391	0.55241	0.93984	0.88562
1.00 ²	1	-1.11466	0.51423	-0.01346	0.00730	0.53353	1.15746	37.23572
	2	-1.11209	0.51143	-0.26048	0.26990	0.53380	1.15163	20.14157
	5	-1.12652	0.42035	-0.35880	0.50256	0.54661	0.95627	3.57665
	10	-1.16391	0.44763	-0.34240	0.51464	0.54738	0.94437	1.25490
	15	-1.16501	0.44878	-0.34155	0.51445	0.54740	0.94468	1.26483
	20	-1.16503	0.44880	-0.34153	0.51444	0.54740	0.94469	1.26540
True values		-1.17000	0.45000	-0.35000	0.52000	0.56000	0.93000	

Example 2 Consider the following linear-in-parameters system with missing data:

$$A(z)y(t) = \frac{\phi^T(t)}{F(z)}\theta + v(t),$$

$$A(z) = 1 + a_1z^{-1} + a_2z^{-2} = 1 - 1.17z^{-1} + 0.45z^{-2},$$

$$F(z) = 1 + f_1z^{-1} + f_2z^{-2} = 1 - 0.35z^{-1} + 0.52z^{-2},$$

$$\theta = [b_1, b_2]^T = [0.56, 0.93]^T,$$

$$\vartheta = [a_1, a_2, f_1, f_2, b_1, b_2]^T = [-1.17, 0.45, -0.35, 0.52, 0.56, 0.93]^T.$$

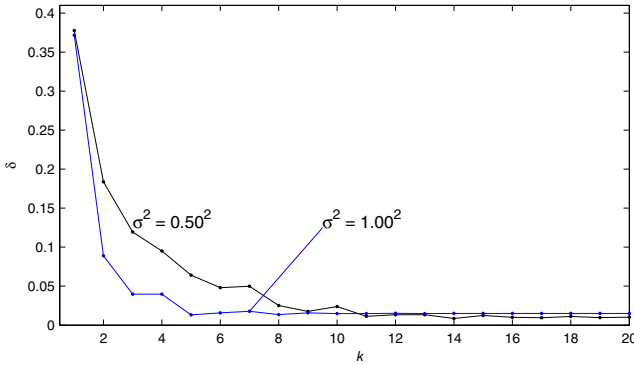


Fig. 4 The V-LSI estimation errors versus k with $\sigma^2 = 0.50^2$ and $\sigma^2 = 1.00^2$

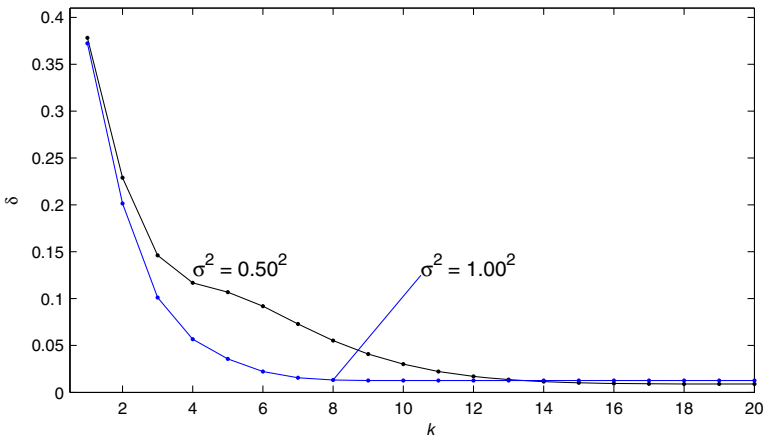


Fig. 5 The V-D-LSI estimation errors versus k with $\sigma^2 = 0.50^2$ and $\sigma^2 = 1.00^2$

The simulation conditions are similar to those of Example 1, here the noise variances $\sigma^2 = 0.50^2$ and $\sigma^2 = 1.00^2$, respectively. Take $s = p = L_e = 3000$ and $t_s^* = 3$, collect the input–output data $\{\phi(t), y(t_s)\}$. Apply the V-LSI algorithm and the V-D-LSI algorithm to identify this example system, the parameter estimates and their estimation errors $\delta := \|\hat{\boldsymbol{\theta}}_k(t_s) - \boldsymbol{\theta}\| / \|\boldsymbol{\theta}\| \times 100\%$ versus k are given in Tables 3 and 4 and Figs. 4 and 5.

From Tables 3 and 4 and Figs. 4 and 5, it is clear that as the iteration k increases, the parameter estimates given by the V-LSI algorithm and the V-D-LSI algorithm converge to their true values, and the estimation errors become smaller (generally); under the same data length and noise variance, the estimation accuracies of the V-LSI algorithm and the V-D-LSI algorithm are close.

8 Conclusions

A D-LSI algorithm and a V-D-LSI algorithm are derived for identifying the linear-in-parameters systems by means of the least squares search and the decomposition

technique. The analysis shows that under the same noise level and iteration, the D-LSI algorithm and the V-D-LSI algorithm give almost same parameter estimation accuracy. Compared with the LSI algorithm and the V-LSI algorithm, the decomposition-based iterative algorithms require less computational cost. The simulation results indicate that the proposed algorithms can generate highly accurate parameter estimates. The identification idea can be extended to study the parameter estimation problems of other linear systems and nonlinear systems with colored noises, missing-data systems and scarce measurement systems [34,35], hybrid networks and uncertain chaotic delayed systems [20,21], and can be applied to other fields [15].

Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 61304138), the Natural Science Foundation of Jiangsu Province (China, BK20130163), the 111 Project (B12018) and the University Graduate Practice Innovation Program of Jiangsu Province (SJZZ15_0151). The first author is grateful to her supervisor—Professor Feng Ding, and the main idea of this paper is from him and his books “System Identification—Multi-Innovation Identification Theory and Methods, Beijing: Science Press, 2016” and Ding [5,6].

References

1. M. Amairi, Recursive set-membership parameter estimation using fractional model. *Circuits Syst. Signal Process.* **34**(12), 3757–3788 (2015)
2. H.B. Chen, F. Ding, Y.S. Xiao, Decomposition-based least squares parameter estimation algorithm for input nonlinear systems using the key term separation technique. *Nonlinear Dyn.* **79**(3), 2027–2035 (2015)
3. H.B. Chen, Y.S. Xiao et al., Hierarchical gradient parameter estimation algorithm for Hammerstein nonlinear systems using the key term separation principle. *Appl. Math. Comput.* **247**, 1202–1210 (2014)
4. M. Dehghan, M. Hajarian, Iterative algorithms for the generalized centro-symmetric and central anti-symmetric solutions of general coupled matrix equations. *Eng. Comput.* **29**(5), 528–560 (2012)
5. F. Ding, *System Identification—New Theory and Methods* (Science Press, Beijing, 2013)
6. F. Ding, *System Identification—Performances Analysis for Identification Methods* (Science Press, Beijing, 2014)
7. F. Ding, K.P. Deng, X.M. Liu, Decomposition based Newton iterative identification method for a Hammerstein nonlinear FIR system with ARMA noise. *Circuits Syst. Signal Process.* **33**(9), 2881–2893 (2014)
8. F. Ding, J. Ding, Least squares parameter estimation with irregularly missing data. *Int. J. Adapt. Control Signal Process.* **24**(7), 540–553 (2010)
9. J. Ding, C.X. Fan, J.X. Lin, Auxiliary model based parameter estimation for dual-rate output error systems with colored noise. *Appl. Math. Model.* **37**(6), 4051–4058 (2013)
10. J. Ding, J.X. Lin, Modified subspace identification for periodically non-uniformly sampled systems by using the lifting technique. *Circuits Syst. Signal Process.* **33**(5), 1439–1449 (2014)
11. F. Ding, X.M. Liu, Y. Gu, An auxiliary model based least squares algorithm for a dual-rate state space system with time-delay using the data filtering. *J. Frankl. Inst. Eng. Appl. Math.* **353** (2015). doi:[10.1016/j.jfranklin.2015.10.025](https://doi.org/10.1016/j.jfranklin.2015.10.025)
12. F. Ding, G. Liu, X.P. Liu, Parameter estimation with scarce measurements. *Automatica* **47**(8), 1646–1655 (2011)
13. F. Ding, X.H. Wang, Q.J. Chen, Y.S. Xiao, Recursive least squares parameter estimation for a class of output nonlinear systems based on the model decomposition. *Circuits Syst. Signal Process.* **35** (2016). doi:[10.1007/s00034-015-0190-6](https://doi.org/10.1007/s00034-015-0190-6)
14. F. Ding, Y.J. Wang, J. Ding, Recursive least squares parameter estimation algorithms for systems with colored noise using the filtering technique and the auxiliary model. *Digit. Signal Process.* **37**, 100–108 (2015)

15. C.L. Fan, H.J. Li, X. Ren, The order recurrence quantification analysis of the characteristics of two-phase flow pattern based on multi-scale decomposition. *Trans. Inst. Meas. Control* **37**(6), 793–804 (2015)
16. Y.Q. Hei, W.T. Li, W.H. Fu, X.H. Li, Efficient parallel artificial bee colony algorithm for cooperative spectrum sensing optimization. *Circuits Syst. Signal Process.* **34**(11), 3611–3629 (2015)
17. Y.B. Hu, Iterative and recursive least squares estimation algorithms for moving average systems. *Simul. Model. Pract. Theory* **34**, 12–19 (2013)
18. Y.B. Hu, B.L. Liu, Q. Zhou, C. Yang, Recursive extended least squares parameter estimation for Wiener nonlinear systems with moving average noises. *Circuits Syst. Signal Process.* **33**(2), 655–664 (2014)
19. J. Huang, Y. Shi, H.N. Huang, Z. Li, 1-2-1-infinity filtering for multirate nonlinear sampled-data systems using T-S fuzzy models. *Digit. Signal Process.* **23**(1), 418–426 (2013)
20. Y. Ji, X.M. Liu, Unified synchronization criteria for hybrid switching-impulsive dynamical networks. *Circuits Syst. Signal Process.* **34**(5), 1499–1517 (2015)
21. Y. Ji, X.M. Liu et al., New criteria for the robust impulsive synchronization of uncertain chaotic delayed nonlinear systems. *Nonlinear Dyn.* **79**(1), 1–9 (2015)
22. Q.B. Jin, Z. Wang, X.P. Liu, Auxiliary model-based interval-varying multi-innovation least squares identification for multivariable OE-like systems with scarce measurements. *J. Process Control* **35**, 154–168 (2015)
23. Q.B. Jin, Z. Wang, J. Wang, Least squares based iterative identification for multivariable integrating and unstable processes in closed loop. *Appl. Math. Comput.* **242**, 10–19 (2014)
24. Q.B. Jin, Z. Wang, Q. Wang, R.G. Yang, Optimal input design for identifying parameters and orders of MIMO systems with initial values. *Appl. Math. Comput.* **224**, 735–742 (2013)
25. Q.B. Jin, Z. Wang, R.G. Yang, J. Wang, An effective direct closed loop identification method for linear multivariable systems with colored noise. *J. Process Control* **24**(5), 485–492 (2014)
26. J.H. Li, Parameter estimation for Hammerstein CARARMA systems based on the Newton iteration. *Appl. Math. Lett.* **26**(1), 91–96 (2013)
27. W.L. Li, Y.M. Jia, J.P. Du, J. Zhang, Robust state estimation for jump Markov linear systems with missing measurements. *J. Frankl. Inst.* **350**(6), 1476–1487 (2013)
28. J. Li, J.D. Zhu, Z.H. Feng, Y.J. Zhao, D.H. Li, Passive multipath time delay estimation using MCMC methods. *Circuits Syst. Signal Process.* **34**(12), 3897–3913 (2015)
29. X.G. Liu, J. Lu, Least squares based iterative identification for a class of multirate systems. *Automatica* **46**(3), 549–554 (2010)
30. H. Raghavan, A.K. Tangirala, R.B. Gopaluni, S.L. Shah, Identification of chemical processes with irregular output sampling. *Control Eng. Pract.* **14**(4), 467–480 (2006)
31. C.F. So, S.H. Leung, Maximum likelihood whitening pre-filtered total least squares for resolving closely spaced signals. *Circuits Syst. Signal Process.* **34**(8), 2739–2747 (2015)
32. Y.J. Wang, F. Ding, Iterative estimation for a nonlinear IIR filter with moving average noise by means of the data filtering technique. *IMA J. Math. Control Inf.* (2016). doi:[10.1093/imamci/dnv067](https://doi.org/10.1093/imamci/dnv067)
33. X.H. Wang, F. Ding, Convergence of the recursive identification algorithms for multivariate pseudo-linear regressive systems. *Int. J. Adapt. Control Signal Process.* **30** (2015). doi:[10.1002/acs.2642](https://doi.org/10.1002/acs.2642)
34. X.H. Wang, F. Ding, Recursive parameter and state estimation for an input nonlinear state space system using the hierarchical identification principle. *Signal Process.* **117**, 208–218 (2015)
35. D.Q. Wang, Y.P. Gao, Recursive maximum likelihood identification method for a multivariable controlled autoregressive moving average system. *IMA J. Math. Control Inf.* (2015). doi:[10.1093/imamci/dnv021](https://doi.org/10.1093/imamci/dnv021)
36. D.Q. Wang, H.B. Liu et al., Highly efficient identification methods for dual-rate Hammerstein systems. *IEEE Trans. Control Syst. Technol.* **23**(5), 1952–1960 (2015)
37. C. Wang, T. Tang, Several gradient-based iterative estimation algorithms for a class of nonlinear systems using the filtering technique. *Nonlinear Dyn.* **77**(3), 769–780 (2014)
38. D.Q. Wang, W. Zhang, Improved least squares identification algorithm for multivariable Hammerstein systems. *J. Frankl. Inst. Eng. Appl. Math.* **352**(11), 5292–5370 (2015)
39. L. Xu, The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. *Signal Process.* **120**, 660–667 (2016)
40. L. Xu, A proportional differential control method for a time-delay system using the Taylor expansion approximation. *Appl. Math. Comput.* **236**, 391–399 (2014)
41. L. Xu, Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. *J. Comput. Appl. Math.* **288**, 33–43 (2015)

42. L. Xu, L. Chen, W.L. Xiong, Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. *Nonlinear Dyn.* **79**(3), 2155–2163 (2015)
43. W.G. Zhang, Decomposition based least squares iterative estimation for output error moving average systems. *Eng. Comput.* **31**(4), 709–725 (2014)