Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Subspace-based identification of discrete time-delay system

Qiang LIU[†1], Jia-chen MA[1,2]
([1]School of Astronautics, Harbin Institute of Technology, Harbin 150001, China)
([2]School of Information and Electrical Engineering, Harbin Institute of Technology, Weihai 264209, China)
[†]E-mail: lqianghit@163.com

Received Oct. 23, 2015; Revision accepted Mar. 14, 2016; Crosschecked May 6, 2016

Abstract: We investigate the identification problems of a class of linear stochastic time-delay systems with unknown delayed states in this study. A time-delay system is expressed as a delay differential equation with a single delay in the state vector. We first derive an equivalent linear time-invariant (LTI) system for the time-delay system using a state augmentation technique. Then a conventional subspace identification method is used to estimate augmented system matrices and Kalman state sequences up to a similarity transformation. To obtain a state-space model for the time-delay system, an alternate convex search (ACS) algorithm is presented to find a similarity transformation that takes the identified augmented system back to a form so that the time-delay system can be recovered. Finally, we reconstruct the Kalman state sequences based on the similarity transformation. The time-delay system matrices under the same state-space basis can be recovered from the Kalman state sequences and input-output data by solving two least squares problems. Numerical examples are to show the effectiveness of the proposed method.

Key words: Identification problems, Time-delay systems, Subspace identification method, Alternate convex search, Least squares
http://dx.doi.org/10.1631/FITEE.1500358 **CLC number:** TP242.6

1 Introduction

Time delays often appear in various engineering systems, such as chemical processes, mechanical systems, transmission lines, and economic systems (Kolmanovskii and Myshkis, 1999; Niculescu, 2001; Richard, 2003; Yang *et al.*, 2015; Bayrak and Tatlicioglu, 2016). The existence of time delays usually causes performance degradation and is frequently a source of instability. Thus, the identification of linear systems with unknown time delays is a prerequisite for system analysis and control design.

As for linear systems with state delays, problems of controllability, observability, controller design (Gao and Chen, 2007), and filter design have received considerable attention. However, system identification of linear time-delay systems is a less developed area (Drakunov *et al.*, 2006; Yang and Gao, 2014). Nakagiri and Yamamoto (1995) and Lunel (2001) illustrated the complexity and intractability of the identification of linear time-delay systems. Identifiability analysis is developed for a linear dynamic system with a delayed state, control input, and measurement output described by functional differential equations (Lunel, 2001; Belkoura and Orlov, 2002; Orlov *et al.*, 2002). The time-delay system is identifiable in principle if and only if this system is weakly controllable and the identifiability of the system parameters and delays can then be enforced by any sufficiently non-smooth input signal (Belkoura and Orlov, 2002). Once the identifiability of the system is verified, an adaptive identification algorithm should be developed to estimate the unknown parameters and time delays. An on-line adaptive identifier was developed by Orlov *et al.* (2003), which enables the simultaneous identification of the system parameters and delays. The structure of the identifier is inspired by the Lyapunov-Razumikhin redesign techniques and has a series-parallel configuration (Orlov *et al.*,

ORCID: Qiang LIU, http://orcid.org/0000-0003-2464-8007

2003). By choosing an appropriate Lyapunov-Razumikhin function and an adaptive identification law, the error dynamic system is guaranteed to be stable, which implies that the identifier parameters converge to the corresponding model parameters. By using the Lyapunov direct method, parallel and series-parallel model estimators for time-delay system identification have been proposed (Park *et al.*, 2013; Hachicha *et al.*, 2014). The stability condition of these two model estimators was given in two linear matrix inequalities. The previously mentioned parameter identifiers were developed for continuous time-delay systems, which all require that the model structure be known and system states be measurable. This restricts their application in practical industrial processes. Advanced techniques for handling system identification problems for discrete time-delay systems are still lacking in the literature.

Subspace identification methods, which have undergone tremendous development in the last 20 years in both theory and practice, offer an attractive alternative to the prediction error method due to simple and general parameterization for multi-input multi-output (MIMO) systems. These methods are computationally tractable and robust because mathematical tools, such as QR factorization and singular value decomposition (SVD), are used. Another advantage of subspace methods is that no prior information about the system is needed. Common and popular approaches in subspace identification include canonical variate analysis, the multi-variable output error state-space algorithm, and the subspace state-space system identification (N4SID) algorithm, and they all fall into a unifying framework in which these algorithms can be interpreted as an SVD of a weighted matrix (Qin, 2006; Lima and Barros, 2015). These algorithms are initially developed for discrete time systems operated in an open loop. Many closed-loop subspace identification methods have also been proposed recently. Wang and Qin (2006) proposed the use of parity space and principal component analysis for closed-loop identification with colored input excitation. In Huang *et al.* (2005), parity space and orthogonal projection were employed to deal with the correlation problem between the input and the unmeasured disturbance under feedback control. In fact, subspace methods can be used as an initial step in many other theoretical frameworks, such as data-driven fault diagnosis (Ding *et al.*, 2009) and data-driven control loop performance monitoring (Huang and Kadali, 2008). Subspace methods can also be used to provide a proper initial estimate for iterative identification algorithms (Lyzell *et al.*, 2009).

In this study, we investigate the identification problems for discrete stochastic time-delay systems with a single delay in the state vector. To handle the time-delay property, the state augmentation technique will be used to transform the time-delay system into an equivalent linear time-invariant (LTI) system. Then subspace methods can be used to provide an initial consistent estimate for the augmented system model. To recover the time-delay system matrices from the estimated augmented model, an alternate convex search (ACS) algorithm is presented so that the time-delay system matrices under the same state-space basis can be recovered from the Kalman state sequences and input-output data by solving two least squares problems.

2 Problem formulation

Consider the linear discrete stochastic time-delay system described by the following delay differential equations:

$$\begin{cases} \boldsymbol{x}(k+1)=\boldsymbol{A}\boldsymbol{x}(k)+\boldsymbol{A}_{\mathrm{d}}\boldsymbol{x}(k-d)+\boldsymbol{B}\boldsymbol{u}(k)+\boldsymbol{w}(k), \\ \boldsymbol{y}(k)=\boldsymbol{C}\boldsymbol{x}(k)+\boldsymbol{D}\boldsymbol{u}(k)+\boldsymbol{v}(k), \end{cases} \quad (1)$$

where $\boldsymbol{u}(k)\in\mathbb{R}^l$, $\boldsymbol{y}(k)\in\mathbb{R}^m$, $\boldsymbol{x}(k)\in\mathbb{R}^n$, $\boldsymbol{w}(k)\in\mathbb{R}^n$, and $\boldsymbol{v}(k)\in\mathbb{R}^m$ are system input, output, state, process noise, and measurement noise, respectively. System matrices $\boldsymbol{A}$, $\boldsymbol{A}_{\mathrm{d}}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ are with appropriate dimensions under the same state-space basis, and d is the unknown constant delay in the state vector.

To derive the identification algorithm of the time-delay system (1), we introduce the following assumptions:

(A1) The time-delay system under consideration is stable and identifiable.

(A2) $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are white noise sequences with zero mean, and their covariance matrices are $\boldsymbol{Q}$ and $\boldsymbol{R}$, respectively.

(A3) The input $\boldsymbol{u}(k)$ and noise sequences $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are uncorrelated. That is, this system is operated in an open loop.

(A4) The input signal is persistently exciting, which enforces system identifiability (Orlov *et al.*, 2002; Qin, 2006).

It is well known that most existing identification methods are confined to dynamic systems without time delays. Linear time-delay systems as described by Eq. (1) cannot directly fall into the well-established frameworks of conventional identification methods. Therefore, to apply identification algorithms such as subspace methods to the time-delay system (1), methods such as the state augmentation technique can be used to transform this system into an equivalent LTI system. Then estimates of augmented system matrices and Kalman state sequences up to a similarity transformation can be obtained using a subspace state-space system identification (N4SID) algorithm. Due to similarity transformation equivalence, system matrices ($\boldsymbol{A}$, $\boldsymbol{A}_\mathrm{d}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$) are mixed up in the estimated augmented system matrices, which takes the identified augmented system back to a form that facilitates the recovery of the time-delay system (1) in this paper. A further important issue is the development of a method guaranteeing that system matrices are recovered under the same state-space basis.

In summation, the linear time-delay system identification problem is: given a set of input and output measurements, estimate system matrices ($\boldsymbol{A}$, $\boldsymbol{A}_\mathrm{d}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$) up to a similarity transformation, with constant time delay d and noise covariance matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$.

3 Identification of LTI systems

3.1 Derivation of an equivalent LTI system

For the linear time-delay system (1), the state augmentation technique is used to associate this system with an equivalent LTI system. Denoting an augmented state vector $\boldsymbol{z}(k)=[\boldsymbol{x}^\mathrm{T}(k)\ \ \boldsymbol{x}^\mathrm{T}(k-1)\ \ \ldots\ \ \boldsymbol{x}^\mathrm{T}(k-d)]^\mathrm{T}$ and employing identical equations $z_j(k+1)=z_{j-1}(k)$, j=2, 3, …, d+1, a higher dimensional LTI system model can be constructed as

$$\begin{cases}\boldsymbol{z}(k+1)=\boldsymbol{A}_l\boldsymbol{z}(k)+\boldsymbol{B}_l\boldsymbol{u}(k)+\boldsymbol{E}_l\boldsymbol{w}(k),\\ \boldsymbol{y}(k)=\boldsymbol{C}_l\boldsymbol{z}(k)+\boldsymbol{D}_l\boldsymbol{u}(k)+\boldsymbol{v}(k),\end{cases}\tag{2}$$

where $\boldsymbol{z}(k)\in\mathbb{R}^{(d+1)n}$, and the system matrices ($\boldsymbol{A}_l$, $\boldsymbol{B}_l$, $\boldsymbol{C}_l$, $\boldsymbol{D}_l$, $\boldsymbol{E}_l$) are given as

$$\begin{cases}\boldsymbol{A}_l=\begin{bmatrix}\boldsymbol{A} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{A}_\mathrm{d}\\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0}\\ \vdots & \vdots & & \vdots & \vdots\\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\in\mathbb{R}^{(d+1)n\times(d+1)n},\\ \boldsymbol{B}_l=[\boldsymbol{B}^\mathrm{T}\ \boldsymbol{0}\ldots\boldsymbol{0}\ \boldsymbol{0}]^\mathrm{T}\in\mathbb{R}^{(d+1)n\times l},\\ \boldsymbol{C}_l=[\boldsymbol{C}\ \boldsymbol{0}\ldots\boldsymbol{0}\ \boldsymbol{0}]^\mathrm{T}\in\mathbb{R}^{m\times(d+1)n},\\ \boldsymbol{D}_l=\boldsymbol{D}\in\mathbb{R}^{m\times l},\\ \boldsymbol{E}_l=[\boldsymbol{I}\ \boldsymbol{0}\ldots\boldsymbol{0}\ \boldsymbol{0}]^\mathrm{T}\in\mathbb{R}^{(d+1)n\times n}.\end{cases}\tag{3}$$

Therefore, the primary time-delay system identification problem can be reformulated as identifying model parameters of the augmented system described by Eq. (2). This can be realized by applying the N4SID algorithm.

3.2 Identification of the augmented system model

Before the application of identification algorithms, the identifiability of the augmented system model (2) should be verified. As stated in Ljung (1987), the state-space model is identifiable if the augmented system is controllable and observable. Here, we have the following proposition (the proof is given in Appendix A):

Proposition 1 ($\boldsymbol{A}_l$, $\boldsymbol{B}_l$) is controllable if and only if for any $\lambda\in\mathbb{C}$, $\mathrm{rank}[\lambda^d(\lambda\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}_\mathrm{d}\boldsymbol{B}]=n$; ($\boldsymbol{C}_l$, $\boldsymbol{A}_l$) is observable if and only if for any $\lambda\in\mathbb{C}$, $\mathrm{rank}[(\lambda^d(\lambda\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}_\mathrm{d})^\mathrm{T}(\lambda^d\boldsymbol{C})^\mathrm{T}]^\mathrm{T}=n$.

Remark 1 It is easy to show that Proposition 1 is consistent with the Popov-Belevitch-Hautus (PBH) rank test (Kailath, 1980) for an LTI system under d=0, which means that there is no state delay in system (1).

If the augmented system is controllable and observable, the identifiability of the system can be enforced by an arbitrary persistently exciting input signal.

Proposition 2 Consider the augmented system given by the tuple ($\boldsymbol{A}_l$, $\boldsymbol{B}_l$, $\boldsymbol{C}_l$, $\boldsymbol{D}_l$), and assume that the input signal is persistently exciting. Then the elements of state vector $\boldsymbol{z}(k)$ are linearly independent if and only if the system is controllable.

The proof of Proposition 2 is given in Appendix B.

We briefly review the N4SID algorithm here. The input and output block Hankel matrices are defined as follows (Huang and Kadali, 2008):

$$\boldsymbol{U}_p \triangleq \boldsymbol{U}_{0|i-1} \triangleq \begin{pmatrix} \boldsymbol{u}_0 & \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_{j-1} \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_j \\ \vdots & \vdots & & \vdots \\ \boldsymbol{u}_{i-1} & \boldsymbol{u}_i & \cdots & \boldsymbol{u}_{i+j-2} \end{pmatrix}, \quad (4)$$

$$\boldsymbol{U}_f \triangleq \boldsymbol{U}_{i|2i-1} \triangleq \begin{pmatrix} \boldsymbol{u}_i & \boldsymbol{u}_{i+1} & \cdots & \boldsymbol{u}_{i+j-1} \\ \boldsymbol{u}_{i+1} & \boldsymbol{u}_{i+2} & \cdots & \boldsymbol{u}_{i+j} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{u}_{2i-1} & \boldsymbol{u}_{2i} & \cdots & \boldsymbol{u}_{2i+j-2} \end{pmatrix}, \quad (5)$$

where $\boldsymbol{U}_p \in \mathbb{R}^{il\times j}$, $\boldsymbol{U}_f \in \mathbb{R}^{il\times j}$, and i and j are user-defined parameters. Matrices $\boldsymbol{Y}_p$, $\boldsymbol{Y}_f$, $\boldsymbol{W}_f$, and $\boldsymbol{V}_f$ can be defined in a similar way. The parameter i is closely associated with the order of the system to be identified; thus, it is typically selected to be large enough so that $i>(d+1)n$. Also, j should be chosen to be very large to reduce noise sensitivity. Denote $\boldsymbol{M}_p = \left[\boldsymbol{Y}_p^{\mathrm{T}} \ \boldsymbol{U}_p^{\mathrm{T}}\right]^{\mathrm{T}}$ and $\boldsymbol{M}_f = \left[\boldsymbol{Y}_f^{\mathrm{T}} \ \boldsymbol{U}_f^{\mathrm{T}}\right]^{\mathrm{T}}$. Based on the above data structure, the following subspace matrix equation can be derived:

$$\boldsymbol{Y}_f = \boldsymbol{\Gamma}_i \boldsymbol{Z}_f + \boldsymbol{H}_i^d \boldsymbol{U}_f + \boldsymbol{H}_i^w \boldsymbol{W}_f + \boldsymbol{V}_f, \quad (6)$$

where

$$\boldsymbol{\Gamma}_i = [\boldsymbol{C}_l \ \ \boldsymbol{C}_l\boldsymbol{A}_l \ \ \boldsymbol{C}_l\boldsymbol{A}_l^2 \ \ldots \ \boldsymbol{C}_l\boldsymbol{A}_l^i]^{\mathrm{T}},$$

$$\boldsymbol{H}_i^d = \begin{bmatrix} \boldsymbol{D}_l & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}_l\boldsymbol{B}_l & \boldsymbol{D}_l & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}_l\boldsymbol{A}_l\boldsymbol{B}_l & \boldsymbol{C}_l\boldsymbol{B}_l & \boldsymbol{D}_l & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{C}_l\boldsymbol{A}_l^{i-2}\boldsymbol{B}_l & \boldsymbol{C}_l\boldsymbol{A}_l^{i-3}\boldsymbol{B}_l & \boldsymbol{C}_l\boldsymbol{A}_l^{i-4}\boldsymbol{B}_l & \cdots & \boldsymbol{D}_l \end{bmatrix},$$

$\boldsymbol{H}_i^w \triangleq \boldsymbol{H}_i^d |_{\boldsymbol{D}_l\to\boldsymbol{0},\boldsymbol{B}_l\to\boldsymbol{E}_l}$, and the state sequence is

$$\boldsymbol{Z}_f=[\boldsymbol{Z}(i)\ \boldsymbol{Z}(i+1)\ \ldots\ \boldsymbol{Z}(i+j-1)]\in\mathbb{R}^{(d+1)n\times j}.$$

Step 1: Perform an oblique projection of Eq. (6) along $\boldsymbol{U}_f$ onto the row space of $\boldsymbol{M}_p$. Based on Assumption (A3) and the property of oblique projection, $\boldsymbol{W}_f /_{\boldsymbol{U}_f} \boldsymbol{M}_p = \boldsymbol{0}$, $\boldsymbol{V}_f /_{\boldsymbol{U}_f} \boldsymbol{M}_p = \boldsymbol{0}$, and the following equation is derived:

$$\boldsymbol{Y}_f /_{\boldsymbol{U}_f} \boldsymbol{M}_p = \boldsymbol{\Gamma}_i \boldsymbol{Z}_f /_{\boldsymbol{U}_f} \boldsymbol{M}_p, \quad (7)$$

where $\boldsymbol{A}/_{\boldsymbol{B}}\boldsymbol{C}$ denotes the oblique projection of the row space of $\boldsymbol{A}$ along the row space of $\boldsymbol{B}$ onto the row space of $\boldsymbol{C}$. The oblique projection is performed using LQ decomposition, and details of this algorithm can be found in Qin *et al.* (2007).

Step 2: Compute the SVD of $\boldsymbol{Y}_f /_{\boldsymbol{U}_f} \boldsymbol{M}_p$. The order of the augmented system (2) can be decided by the number of significant singular values. The Kalman state sequence ($\hat{\boldsymbol{Z}}_f$), which is the estimate of $\boldsymbol{Z}_f$, can be constructed with the right singular vectors. With the available state sequences, augmented system matrices are obtained by solving the following least squares problem:

$$\underset{\boldsymbol{A}_l,\boldsymbol{B}_l,\boldsymbol{C}_l,\boldsymbol{D}_l}{\arg\min} \left\| \begin{pmatrix} \hat{\boldsymbol{Z}}(k+1) \\ \boldsymbol{Y}_{i|i} \end{pmatrix} - \begin{pmatrix} \boldsymbol{A}_l & \boldsymbol{B}_l \\ \boldsymbol{C}_l & \boldsymbol{D}_l \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{Z}}(k) \\ \boldsymbol{U}_{i|i} \end{pmatrix} \right\|_{\mathrm{F}}^2, \quad (8)$$

where $\|\cdot\|_{\mathrm{F}}^2$ is the Frobenious norm; e.g., for a certain matrix $\boldsymbol{M}$, there is $\|\boldsymbol{M}\|_{\mathrm{F}}^2=\mathrm{tr}(\boldsymbol{M}\boldsymbol{M}^{\mathrm{T}})$.

Remark 2 As presented in step 2, the order of the augmented system, $(d+1)\times n$, is obtained. If the order of time-delay system (n) is known a priori, e.g., from first principles, the time-delay d is then calculated easily. However, if not previously known, different pairs of n and d should be tried. The correct values are obtained when the system model (1) whose structure is decided by a certain pair of n and d gives a minimum output prediction error.

Remark 3 Note that the N4SID algorithm used here is intended to provide an initial consistent model estimate for the augmented system operated in an open loop. This can also be extended to a closed-loop operating system; however, subspace identification methods presented in Huang *et al.* (2005) and Wang and Qin (2006) should be used instead of N4SID.

3.3 Recovery of time-delay system matrices

It is well known that the system matrices $\left(\hat{\boldsymbol{A}}_l, \hat{\boldsymbol{B}}_l, \hat{\boldsymbol{C}}_l, \hat{\boldsymbol{D}}_l\right)$ obtained by the N4SID algorithm in Section 3.2 are closely related to the system matrices ($\boldsymbol{A}_l$, $\boldsymbol{B}_l$, $\boldsymbol{C}_l$, $\boldsymbol{D}_l$) defined in Eq. (3) by a non-singular matrix $\boldsymbol{T}$. This similarity transformation equivalence can be expressed by the following bilinear equations:

$$\begin{cases} \boldsymbol{T}\hat{\boldsymbol{A}}_l = \boldsymbol{A}_l\boldsymbol{T}\boldsymbol{T}\hat{\boldsymbol{B}}_l = \boldsymbol{B}_l, \\ \hat{\boldsymbol{C}}_l = \boldsymbol{C}_l\boldsymbol{T}\hat{\boldsymbol{D}}_l = \boldsymbol{D}_l. \end{cases} \tag{9}$$

Due to the arbitrariness of $\boldsymbol{T}$, the time-delay system matrices ($\boldsymbol{A}$, $\boldsymbol{A}_\text{d}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$) are mixed up in the estimated augmented system matrices. Therefore, a method should be developed to transform the identified augmented system back into a form so that the time-delay system can be recovered. An ACS algorithm based on Eq. (9) is presented in this subsection.

We define ($\boldsymbol{A}_l(\boldsymbol{\theta})$, $\boldsymbol{B}_l(\boldsymbol{\theta})$, $\boldsymbol{C}_l(\boldsymbol{\theta})$, $\boldsymbol{D}_l(\boldsymbol{\theta})$), which have the same structure as defined in Eq. (3), leaving $\boldsymbol{A}$, $\boldsymbol{A}_\text{d}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ unknown. Vector $\boldsymbol{\theta}$ contains all these unknown parameters. The objectives of this algorithm are to find $\boldsymbol{\theta}$ and a non-singular matrix $\boldsymbol{T}$ that transforms the estimated system matrices $\left(\hat{\boldsymbol{A}}_l, \hat{\boldsymbol{B}}_l, \hat{\boldsymbol{C}}_l, \hat{\boldsymbol{D}}_l\right)$ into ($\boldsymbol{A}_l(\boldsymbol{\theta})$, $\boldsymbol{B}_l(\boldsymbol{\theta})$, $\boldsymbol{C}_l(\boldsymbol{\theta})$, $\boldsymbol{D}_l(\boldsymbol{\theta})$). It is worth noting that the left-upper corner and the right-upper corner of $\boldsymbol{A}_l(\boldsymbol{\theta})$ have the same dynamics as $\boldsymbol{A}$ and $\boldsymbol{A}_\text{d}$, respectively. To approximate the solution, the idea is to solve

$$\begin{cases} (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{T}}) = \arg\min\limits_{\boldsymbol{\theta},\boldsymbol{T}} F(\boldsymbol{\theta},\boldsymbol{T}), \\ F(\boldsymbol{\theta},\boldsymbol{T}) = \| \boldsymbol{T}\hat{\boldsymbol{A}}_l - \boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{T} \|_\text{F}^2 + \| \boldsymbol{T}\hat{\boldsymbol{B}}_l - \boldsymbol{B}_l(\boldsymbol{\theta}) \|_\text{F}^2 \\ \qquad\qquad + \| \hat{\boldsymbol{C}}_l - \boldsymbol{C}_l(\boldsymbol{\theta})\boldsymbol{T} \|_\text{F}^2 . \end{cases} \tag{10}$$

Clearly, this is a biconvex optimization problem, and the solution of this problem can be achieved by solving two convex optimization subproblems that are activated in cycles. On the basis of this, our ACS algorithm is described as follows:

Step 1: Initialize the unknown parameter $\boldsymbol{\theta}$.

Step 2: For fixed $\boldsymbol{\theta}$, Eq. (9) is linear for $\boldsymbol{T}$. The estimate of $\boldsymbol{T}$ can be found by solving a least squares problem

$$\arg\min_{\boldsymbol{T}} \| \boldsymbol{A}_{\hat{\boldsymbol{\theta}}}\text{vec}(\boldsymbol{T}) - \boldsymbol{b}_{\hat{\boldsymbol{\theta}}} \|_\text{F}^2, \tag{11}$$

where

$$\boldsymbol{A}_{\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \boldsymbol{I}_n \otimes \boldsymbol{A}_l(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{A}}_l^\text{T} \otimes \boldsymbol{I}_n \\ \hat{\boldsymbol{B}}_l^\text{T} \otimes \boldsymbol{I}_n \\ \boldsymbol{I}_n \otimes \boldsymbol{C}_l(\hat{\boldsymbol{\theta}}) \end{pmatrix}, \quad \boldsymbol{b}_{\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \boldsymbol{0} \\ \text{vec}\left(\boldsymbol{B}_l(\hat{\boldsymbol{\theta}})\right) \\ \text{vec}(\hat{\boldsymbol{C}}_l) \end{pmatrix},$$

where '⊗' is the Kronecker product, the 'vec' operator builds a vector from a matrix by stacking its columns on top of one another, and $\boldsymbol{I}_n$ denotes the n-dimensional identity matrix.

Step 3: For the fixed $\boldsymbol{T}$ obtained in step 2, Eq. (9) is linear for $\boldsymbol{\theta}$. An estimate for $\boldsymbol{\theta}$ can be found by solving a least squares problem:

$$\arg\min_{\boldsymbol{\theta}} \left(\left\| \boldsymbol{A}_{\hat{\boldsymbol{T}}} \begin{pmatrix} \text{vec}(\boldsymbol{A}_l(\boldsymbol{\theta})) \\ \text{vec}(\boldsymbol{B}_l(\boldsymbol{\theta})) \\ \text{vec}(\boldsymbol{C}_l(\boldsymbol{\theta})) \end{pmatrix} - \boldsymbol{b}_{\hat{\boldsymbol{T}}} \right\|_\text{F}^2 \right), \tag{12}$$

subject to

$$\boldsymbol{A}_\text{eq} \begin{pmatrix} \text{vec}(\boldsymbol{A}_l(\boldsymbol{\theta})) \\ \text{vec}(\boldsymbol{B}_l(\boldsymbol{\theta})) \\ \text{vec}(\boldsymbol{C}_l(\boldsymbol{\theta})) \end{pmatrix} = \boldsymbol{b}_\text{eq}, \tag{13}$$

where

$$\boldsymbol{A}_{\hat{\boldsymbol{T}}} = \begin{pmatrix} \hat{\boldsymbol{T}}^\text{T} \otimes \boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \hat{\boldsymbol{T}}^\text{T} \otimes \boldsymbol{I}_m \end{pmatrix}, \quad \boldsymbol{b}_{\hat{\boldsymbol{T}}} = \begin{pmatrix} \text{vec}(\hat{\boldsymbol{T}}\hat{\boldsymbol{A}}_l) \\ \text{vec}(\hat{\boldsymbol{T}}\hat{\boldsymbol{B}}_l) \\ \text{vec}(\hat{\boldsymbol{C}}_l) \end{pmatrix},$$

$\boldsymbol{b}_\text{eq}$ is a vector which consists of 0 or 1, and $\boldsymbol{A}_\text{eq}$ is a matrix which decides the arguments whose values are restricted. All of the known arguments in $\boldsymbol{A}_l(\boldsymbol{\theta})$, $\boldsymbol{B}_l(\boldsymbol{\theta})$, and $\boldsymbol{C}_l(\boldsymbol{\theta})$ are included in equality constraint (13).

Step 4: Calculate the value of error

$$\begin{aligned} e = &\| \hat{\boldsymbol{T}}\hat{\boldsymbol{A}}_l - \boldsymbol{A}_l(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{T}} \|_\text{F} + \| \hat{\boldsymbol{T}}\hat{\boldsymbol{B}}_l - \boldsymbol{B}_l(\hat{\boldsymbol{\theta}}) \|_\text{F} \\ &+ \| \hat{\boldsymbol{C}}_l - \boldsymbol{C}_l(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{T}} \|_\text{F} . \end{aligned} \tag{14}$$

If $e<\varepsilon$ (ε is an arbitrarily small positive constant), then one solution $\hat{\boldsymbol{\theta}}$ of the unknown arguments is obtained and $\hat{\boldsymbol{T}}$ is the corresponding transformation matrix;

else, return to step 2.

Remark 4 The cost functions in Eqs. (11) and (12) are derived directly from cost function in Eq. (10) based on the definition of the Frobenious norm and the well-known property of the Kronecker product vec($\boldsymbol{EXF}$)=($\boldsymbol{F}^{\mathrm{T}}\otimes\boldsymbol{E}$)vec($\boldsymbol{X}$).

Remark 5 Note that this is a biconvex optimization problem. The convergence property of this algorithm is not guaranteed unless the initial values of $\boldsymbol{\theta}$ are chosen well. Related discussion of the choice of starting points for a similar iterative algorithm can be found in Xie and Ljung (2002), Gorski *et al.* (2007), and Prot and Mercère (2011).

Following steps 1–4, estimates of original time-delay system matrices $\hat{\boldsymbol{A}}, \hat{\boldsymbol{A}}_{\mathrm{d}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}$, and $\hat{\boldsymbol{D}}$ are derived. If the initial values $\boldsymbol{\theta}$ are chosen well, the local convergence of our algorithm can be proved. Details of the proof can be seen in Appendix C.

Once we have the solution, estimates of time-delay system matrices $\hat{\boldsymbol{A}}, \hat{\boldsymbol{A}}_{\mathrm{d}}, \hat{\boldsymbol{B}}$, and $\hat{\boldsymbol{C}}$ can be extracted from $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{D}}=\boldsymbol{D}_l$. It is easy to verify that $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{A}}_{\mathrm{d}}$ have the same eigenvalues as $\boldsymbol{A}$ and $\boldsymbol{A}_{\mathrm{d}}$, respectively, which shows the effectiveness of our algorithm. Unfortunately, the estimated system matrices are not under the same state-space basis, which can be easily inferred from

$$\begin{cases}\boldsymbol{P}\hat{\boldsymbol{A}}=\boldsymbol{A}\boldsymbol{P},\\ \boldsymbol{P}\hat{\boldsymbol{B}}=\boldsymbol{B}\Rightarrow\boldsymbol{P}, \text{ but } \boldsymbol{P}\hat{\boldsymbol{A}}_{\mathrm{d}}\neq\boldsymbol{A}_{\mathrm{d}}\boldsymbol{P},\\ \hat{\boldsymbol{C}}=\boldsymbol{C}\boldsymbol{P}.\end{cases} \tag{15}$$

To solve this problem, we reconstruct the Kalman state sequences $\hat{\boldsymbol{Z}}_f$ obtained in Section 3.2.

$$\hat{\boldsymbol{Z}}_f{}'=\hat{\boldsymbol{T}}\hat{\boldsymbol{Z}}_f \text{ and } \hat{\boldsymbol{Z}}'=\hat{\boldsymbol{Z}}_f{}'(1:n,:), \tag{16}$$

where $\hat{\boldsymbol{Z}}_f{}'(1:n,:)$ is the extraction of the matrix elements from the first row to the nth row.

Then the time-delay system matrices ($\boldsymbol{A}$, $\boldsymbol{A}_{\mathrm{d}}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$) are estimated instead, by solving the following two least squares problems:

$$\underset{\boldsymbol{A},\boldsymbol{A}_{\mathrm{d}},\boldsymbol{B}}{\arg\min}\|\boldsymbol{Z}'(k+1)-\boldsymbol{A}\boldsymbol{Z}'(k)-\boldsymbol{A}_{\mathrm{d}}\boldsymbol{Z}'(k-d)-\boldsymbol{B}\boldsymbol{U}_{i|i}\|_{\mathrm{F}}^2, \tag{17}$$

$$\underset{\boldsymbol{C},\boldsymbol{D}}{\arg\min}\|\boldsymbol{Y}_{i|i}-\boldsymbol{C}\boldsymbol{Z}'(k)-\boldsymbol{D}\boldsymbol{U}_{i|i}\|_{\mathrm{F}}^2. \tag{18}$$

4 Simulation

In this section, a simulation study is presented to demonstrate the performance of our method. To comply with the standard practice in the system identification literature, Monte Carlo simulations are performed. Three different performance criteria are adopted in this study. A Bode plot from the Monte Carlo simulations (Fig. 1) is used to represent the bias error, while a scatter plot of estimated poles (Figs. 2–4) is used to represent the variance error of the estimation (Huang and Kadali, 2008). The step response curve is also used to compare the input-output characteristics between the actual system and the estimated system.

Consider the following discrete-time system with a single delay in state vector:

$$\begin{cases}\boldsymbol{x}(k+1)=\begin{bmatrix}0.5 & 0.2\\ 0 & 0.1\end{bmatrix}\boldsymbol{x}(k)+\begin{bmatrix}0.1 & 0.2\\ 0 & 0.2\end{bmatrix}\boldsymbol{x}(k-2)\\ \qquad +\begin{bmatrix}0.4\\ 0.2\end{bmatrix}\boldsymbol{u}(k),\\ \boldsymbol{y}(k)=[2\ \ 1]\boldsymbol{x}(k)+0.9\boldsymbol{u}(k)+\boldsymbol{v}(k).\end{cases} \tag{19}$$

It is easy to prove that the system under consideration is stable and that the augmented system is controllable and observable. The measurement noise $\boldsymbol{v}(k)$ is Gaussian white with standard deviation 0.02^2. The excitation signal $\boldsymbol{u}(k)$ is a Gaussian white noise sequence with variance 1. Each simulation generates 900 data points. We generate 50 simulated datasets, each with the same excitation signal $\boldsymbol{u}(k)$ but with a different noise sequence $\boldsymbol{v}(k)$.

5 Conclusions

Identification problems of linear stochastic time-delay systems in discrete time have been investigated in this paper. To bypass the difficulties introduced into system identification by the time-delay property, the state augmentation technique has been employed to associate the time-delay system with an equivalent LTI system. Then the N4SID algorithm was used to provide an initial consistent estimate for the augmented system matrices. An ACS

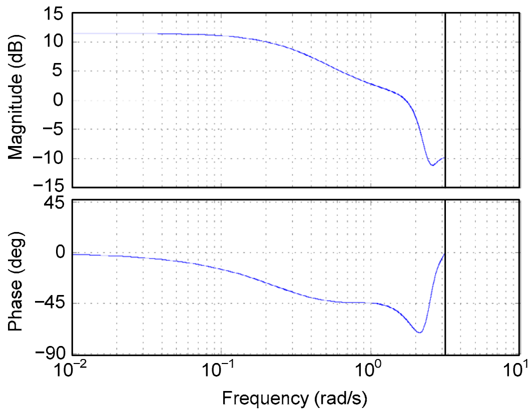

Fig. 1 Bode plot of the estimated augmented system

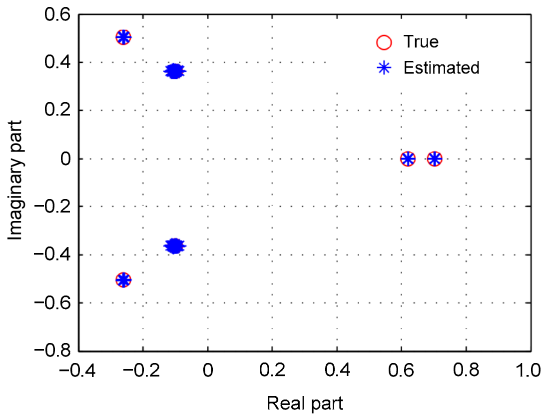

Fig. 2 Pole locations of the estimated augmented system

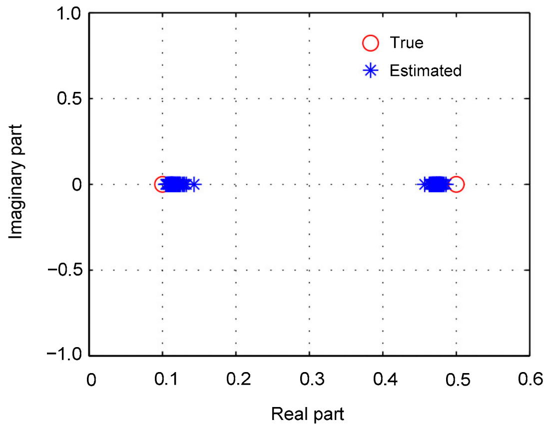

Fig. 3 Eigenvalue locations of the estimated A

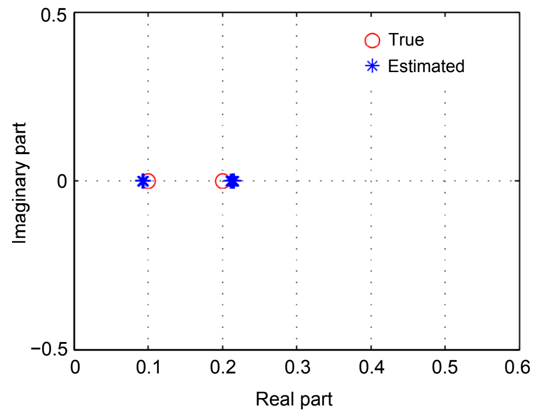

Fig. 4 Eigenvalue locations of the estimated A_d

algorithm was also presented for recovering the time-delay system matrices from the estimated augmented matrices, and the local convergence of this algorithm has been proved. Finally, the time-delay system matrices up to a similarity transformation were derived by solving two least squares problems based on reconstructed Kalman state sequences and input-output data. A numerical example was provided to show the effectiveness of the developed method.

References

Bayrak, A., Tatlicioglu, E., 2016. A novel online adaptive time delay identification technique. *Int. J. Syst. Sci.*, **47**(7):1574-1585. http://dx.doi.org/10.1080/00207721.2014.941958

Belkoura, L., Orlov, Y., 2002. Identifiability analysis of linear delay-differential systems. *IMA J. Math. Contr. Inform.*, **19**(1-2):73-81. http://dx.doi.org/10.1093/imamci/19.1_and_2.73

Ding, S.X., Zhang, P., Naik, A., *et al.*, 2009. Subspace method aided data-driven design of fault detection and isolation systems. *J. Process Contr.*, **19**(9):1496-1510. http://dx.doi.org/10.1016/j.jprocont.2009.07.005

Drakunov, S.V., Perruquetti, W., Richard, J.P., *et al.*, 2006. Delay identification in time-delay systems using variable structure observers. *Ann. Rev. Contr.*, **30**(2):143-158. http://dx.doi.org/10.1016/j.arcontrol.2006.08.001

Gao, H., Chen, T., 2007. New results on stability of discrete-time systems with time-varying state delay. *IEEE Trans. Autom. Contr.*, **52**(2):328-334. http://dx.doi.org/10.1109/TAC.2006.890320

Gorski, J., Pfeuffer, F., Klamroth, K., 2007. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Math. Methods Oper. Res.*, **66**(3):373-407. http://dx.doi.org/10.1007/s00186-007-0161-1

Hachicha, S., Kharrat, M., Chaari, A., 2014. N4SID and MOESP algorithms to highlight the ill-conditioning into subspace identification. *Int. J. Autom. Comput.*, **11**(1): 30-38. http://dx.doi.org/10.1007/s11633-014-0763-z

Huang, B., Kadali, R., 2008. Dynamic Modeling, Predictive Control and Performance Monitoring: a Data-Driven Subspace Approach. Springer, London, UK. http://dx.doi.org/10.1007/978-1-84800-233-3

Huang, B., Ding, S.X., Qin, S.J., 2005. Closed-loop subspace identification: an orthogonal projection approach. *J. Process Contr.*, **15**(1):53-66. http://dx.doi.org/10.1016/j.jprocont.2004.04.007

Kailath, T., 1980. Linear Systems. Prentice-Hall, New Jersey, USA.

Kolmanovskii, V., Myshkis, A., 1999. Introduction to the Theory and Applications of Functional Differential Equations. Springer, the Netherlands. http://dx.doi.org/10.1007/978-94-017-1965-0

Kudva, P., Narendra, K.S., 1973. An Identification Procedure for Discrete Multivariable Systems. Technical Report No. AD0768992, Yale University, USA.

Lima, R.B.C., Barros, P.R., 2015. Identification of time-delay systems: a state-space realization approach. Proc. 9th IFAC Symp. on Advanced Control of Chemical Processes, p.254-259. http://dx.doi.org/10.1016/j.ifacol.2015.08.190

Ljung, L., 1987. System Identification: Theory for the User. PTR Prentice Hall, New Jersey, USA.

Lunel, S.M.V., 2001. Parameter identifiability of differential delay equations. *Int. J. Adapt. Contr. Signal Process.*, **15**(6):655-678. http://dx.doi.org/10.1002/acs.690

Lyzell, C., Enqvist, M., Ljung, L., 2009. Handling Certain Structure Information in Subspace Identification. Report, Linköping University Electronic Press, Sweden.

Nakagiri, S., Yamamoto, M., 1995. Unique identification of coefficient matrices, time delays and initial functions of functional differential equations. *J. Math. Syst. Estimat. Contr.*, **5**(3):323-344.

Niculescu, S.I., 2001. Delay Effects on Stability: a Robust Control Approach. Springer-Verlag, London, UK. http://dx.doi.org/10.1007/1-84628-553-4

Orlov, Y., Belkoura, L., Richard, J.P., *et al.*, 2002. On identifiability of linear time-delay systems. *IEEE Trans. Autom. Contr.*, **47**(8):1319-1324. http://dx.doi.org/10.1109/TAC.2002.801202

Orlov, Y., Belkoura, L., Richard, J.P., *et al.*, 2003. Adaptive identification of linear time-delay systems. *Int. J. Robust Nonl. Contr.*, **13**(9):857-872. http://dx.doi.org/10.1002/rnc.850

Park, J.H., Han, S., Kwon, B., 2013. On-line model parameter estimations for time-delay systems. *IEICE Trans. Inform. Syst.*, **96**(8):1867-1870.

Pourboghrat, F., Chyung, D.H., 1989. Parameter identification of linear delay systems. *Int. J. Contr.*, **49**(2):595-627. http://dx.doi.org/10.1080/00207178908559656

Prot, O., Mercère, G., 2011. Initialization of gradient-based optimization algorithms for the identification of structured state-space models. Proc. 18th IFAC World Congress, p.10782-10787.

Qin, P., Kanae, S., Yang, Z.J., *et al.*, 2007. Identification of lifted models for general dual-rate sampled-data systems based on input-output data. Proc. Asian Modelling and Simulation, p.7-12.

Qin, S.J., 2006. An overview of subspace identification. *Comput. Chem. Eng.*, **30**(10-12):1502-1513. http://dx.doi.org/10.1016/j.compchemeng.2006.05.045

Richard, J.P., 2003. Time-delay systems: an overview of some recent advances and open problems. *Automatica*, **39**(10): 1667-1694. http://dx.doi.org/10.1016/S0005-1098(03)00167-5

Wang, J., Qin, S.J., 2006. Closed-loop subspace identification using the parity space. *Automatica*, **42**(2):315-320. http://dx.doi.org/10.1016/j.automatica.2005.09.012

Xie, L., Ljung, L., 2002. Estimate physical parameters by black box modeling. Proc. 21st Chinese Control Conf., p.673-677.

Yang, X., Gao, H., 2014. Multiple model approach to linear parameter varying time-delay system identification with EM algorithm. *J. Franklin Inst.*, **351**(12):5565-5581. http://dx.doi.org/10.1016/j.jfranklin.2014.09.015

Yang, X., Lu, Y., Yan, Z., 2015. Robust global identification of linear parameter varying systems with generalised expectation-maximisation algorithm. *IET Contr. Theory Appl.*, **9**(7):1103-1110. http://dx.doi.org/10.1049/iet-cta.2014.0694

Appendix A: Proof of Proposition 1

Proof The PBH rank test (Kailath, 1980) is used to prove Proposition 1.

1. According to the PBH rank test, the augmented system is controllable if and only if for any $\lambda\in\mathbb{C}$, rank$[\lambda\boldsymbol{I}-\boldsymbol{A}_l\boldsymbol{B}_l]=(d+1)n$. That is,

$$\text{rank}\begin{bmatrix} \lambda\boldsymbol{I}-\boldsymbol{A} & \boldsymbol{0} & \cdots & \boldsymbol{0} & -\boldsymbol{A}_\text{d} & \boldsymbol{B} \\ -\boldsymbol{I} & \lambda\boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -\boldsymbol{I} & \lambda\boldsymbol{I} & \boldsymbol{0} \end{bmatrix}=(d+1)n.$$

By exploring the primary matrix transformation, the equation above can be transformed into

$$\text{rank}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \lambda^d(\lambda\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}_\text{d} & \boldsymbol{B} \\ -\boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}=(d+1)n.$$

The above equation depends on rank$[\lambda^d(\lambda\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}_\text{d}\boldsymbol{B}]=n$. So, the first part of Proposition 1 is proved.

2. According to the PBH rank test, the augmented system is observable if and only if for any

$\lambda\in\mathbb{C}$, $\text{rank}[\boldsymbol{C}_l^{\mathrm{T}}(\lambda\boldsymbol{I}-\boldsymbol{A}_l)^{\mathrm{T}}]^{\mathrm{T}}=(d+1)n$. That is,

$$\text{rank}\begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{0} \\ \lambda\boldsymbol{I}-\boldsymbol{A} & \boldsymbol{0} & \ldots & \boldsymbol{0} & -\boldsymbol{A}_{\mathrm{d}} \\ -\boldsymbol{I} & \lambda\boldsymbol{I} & \ldots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I} & \ldots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & -\boldsymbol{I} & \lambda\boldsymbol{I} \end{bmatrix}=(d+1)n.$$

This equation can be transformed into the following equivalent form by conducting a primary matrix transformation:

$$\text{rank}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{0} & \lambda^d\boldsymbol{C} \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{0} & \lambda^d(\lambda\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}_{\mathrm{d}} \\ -\boldsymbol{I} & \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{0} \\ 0 & -\boldsymbol{I} & \ldots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & -\boldsymbol{I} & \boldsymbol{0} \end{bmatrix}=(d+1)n.$$

It is easy to show that the above equation depends on $\text{rank}[(\lambda^d(\lambda\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}_{\mathrm{d}})^{\mathrm{T}}(\lambda^d\boldsymbol{C})^{\mathrm{T}}]^{\mathrm{T}}=n$. So, the proof of Proposition 1 is completed.

Appendix B: Proof of Proposition 2

Proof Based on the augmented system (2), we obtain

$$\boldsymbol{z}(k)=(p\boldsymbol{I}-\boldsymbol{A}_l(\boldsymbol{\theta}))^{-1}\boldsymbol{B}_l(\boldsymbol{\theta})\boldsymbol{u}(k),$$

where p is the shift operator such that $p[\boldsymbol{z}(k)]\triangleq\boldsymbol{z}(k+1)$.

We define $f(\lambda)$ as the characteristic polynomial of $\boldsymbol{A}_l$. Then

$$f(\lambda)=\lambda^N+\alpha_N\lambda^{N-1}+\alpha_{N-1}\lambda^{N-2}+\ldots+\alpha_2\lambda+\alpha_1,$$

where $N=(d+1)n$. It is well known (Kudva and Narendra, 1973) that

$$(\lambda\boldsymbol{I}-\boldsymbol{A}_l(\boldsymbol{\theta}))^{-1}=\frac{1}{f(\lambda)}\sum_{i=1}^{N}f_i(\lambda)\boldsymbol{A}_l(\boldsymbol{\theta})^{i-1},$$

where the auxiliary polynomials are defined as

$$f_N(\lambda)=1,\ f_0(\lambda)=f(\lambda),$$

and

$$f_{i-1}(\lambda)=\lambda f_i(\lambda)+\alpha_i,\ i\in\{1, 2, \ldots, N\}.$$

Therefore, $\boldsymbol{z}(k)$ can be written as

$$\boldsymbol{z}(k)=\frac{1}{f(p)}\sum_{i=1}^{N}f_i(p)\boldsymbol{A}_l(\boldsymbol{\theta})^{i-1}\boldsymbol{B}_l(\boldsymbol{\theta})\boldsymbol{u}(k).$$

We first show the sufficiency of Proposition 2. Suppose that the input signal is persistently exciting and the augmented system is controllable, but the elements of the state vector $\boldsymbol{z}(k)$ are linearly dependent. That is to say, there exists a nonzero vector $\boldsymbol{\alpha}\in\mathbb{R}^N$ such that $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{z}(k)=0$ (Pourboghrat and Chyung, 1989). Then

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{z}(k)=\boldsymbol{\alpha}^{\mathrm{T}}f^{-1}(p)\sum_{i=1}^{N}f_i(p)\boldsymbol{A}_l(\boldsymbol{\theta})^{i-1}\boldsymbol{B}_l(\boldsymbol{\theta})\boldsymbol{u}(k)=0.$$

Using the definition of $f_i(p)$, the equation above can be further written as

$$\begin{aligned}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{z}(k)=&\boldsymbol{\alpha}^{\mathrm{T}}f^{-1}(p)[\boldsymbol{B}_l(\boldsymbol{\theta})p^{N-1}+(\alpha_N\boldsymbol{B}_l(\boldsymbol{\theta})+\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta}))p^{N-2}\\&+(\alpha_{N-1}\boldsymbol{B}_l(\boldsymbol{\theta})+\alpha_N\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})+\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta}))p^{N-3}+\ldots\\&+(\alpha_2\boldsymbol{B}_l(\boldsymbol{\theta})+\alpha_3\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})+\ldots+\boldsymbol{A}_l(\boldsymbol{\theta})^{N-1}\boldsymbol{B}_l(\boldsymbol{\theta}))]\boldsymbol{u}(k)=0.\end{aligned}$$

It is easy to see that if the input signal is persistently exciting, then the equation above implies

$$\begin{aligned}&\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{B}_l(\boldsymbol{\theta})=0,\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})=0,\ \ldots,\\&\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_l(\boldsymbol{\theta})^{N-2}\boldsymbol{B}_l(\boldsymbol{\theta})=0,\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_l(\boldsymbol{\theta})^{N-1}\boldsymbol{B}_l(\boldsymbol{\theta})=0.\end{aligned}$$

That is,

$$\boldsymbol{\alpha}^{\mathrm{T}}[\boldsymbol{B}_l(\boldsymbol{\theta})\ \boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})\ \ldots\ \boldsymbol{A}_l(\boldsymbol{\theta})^{N-2}\boldsymbol{B}_l(\boldsymbol{\theta})\ \boldsymbol{A}_l(\boldsymbol{\theta})^{N-1}\boldsymbol{B}_l(\boldsymbol{\theta})]=0.$$

Since $\boldsymbol{\alpha}$ is an arbitrary nonzero vector, the equation above implies

$$\text{rank}[\boldsymbol{B}_l(\boldsymbol{\theta})\ \boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})\ \ldots\ \boldsymbol{A}_l(\boldsymbol{\theta})^{N-2}\boldsymbol{B}_l(\boldsymbol{\theta})\ \boldsymbol{A}_l(\boldsymbol{\theta})^{N-1}\boldsymbol{B}_l(\boldsymbol{\theta})]<N.$$

This means the augmented system is uncontrollable, which is a contradiction. Therefore, if the input signal is persistently exciting and the augmented system is controllable, the elements of state vector $\boldsymbol{z}(k)$ are linearly independent.

Next, let us show the necessity of Proposition 2. Suppose the input signal is persistently exciting and the elements of the state vector $\boldsymbol{z}(k)$ are linearly independent, but the augmented system is uncontrollable. Therefore, there exists a nonzero vector $\boldsymbol{\alpha}\in\mathbb{R}^N$ such that

$$\boldsymbol{\alpha}^{\mathrm{T}}[\boldsymbol{B}_l(\boldsymbol{\theta})\ \boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})\ \ldots\ \boldsymbol{A}_l(\boldsymbol{\theta})^{N-2}\boldsymbol{B}_l(\boldsymbol{\theta})\ \boldsymbol{A}_l(\boldsymbol{\theta})^{N-1}\boldsymbol{B}_l(\boldsymbol{\theta})]=0,$$

which means

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{B}_l(\boldsymbol{\theta})=0,\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})=0,\ \ldots,$$
$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_l(\boldsymbol{\theta})^{N-2}\boldsymbol{B}_l(\boldsymbol{\theta})=0,\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_l(\boldsymbol{\theta})^{N-1}\boldsymbol{B}_l(\boldsymbol{\theta})=0.$$

Then it is easy to show that

$$\begin{aligned}&\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{z}(k)=\boldsymbol{\alpha}^{\mathrm{T}}f^{-1}(p)[\boldsymbol{B}_l(\boldsymbol{\theta})p^{N-1}+(\alpha_N\boldsymbol{B}_l(\boldsymbol{\theta})+\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta}))p^{N-2}\\&+(\alpha_{N-1}\boldsymbol{B}_l(\boldsymbol{\theta})+\alpha_N\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})+\boldsymbol{A}_l(\boldsymbol{\theta})^2\boldsymbol{B}_l(\boldsymbol{\theta}))p^{N-3}+\ldots\\&+(\alpha_2\boldsymbol{B}_l(\boldsymbol{\theta})+\alpha_3\boldsymbol{A}_l(\boldsymbol{\theta})\boldsymbol{B}_l(\boldsymbol{\theta})+\ldots+\boldsymbol{A}_l(\boldsymbol{\theta})^{N-1}\boldsymbol{B}_l(\boldsymbol{\theta}))]\boldsymbol{u}(k)=0.\end{aligned}$$

Since $\boldsymbol{\alpha}$ is an arbitrary nonzero vector, the equation above implies that the elements of the state vector are linearly dependent, which is a contradiction to our assumption. Therefore, if the input signal is persistently exciting and the elements of the state vector are linearly independent, the augmented system is controllable.

Appendix C: Proof of the convergence of our ACS algorithm

Proof As shown in the algorithm procedures, solution of the biconvex optimization problem is achieved by solving two convex subproblems that are activated alternately. Therefore, the first step of the proof is to verify the solvability and uniqueness of solutions of the two convex subproblems. Since the optimization problems in steps 2 and 3 are two least squares problems, they are solvable. Moreover, uniqueness of the solutions can be guaranteed under the assumption that the coefficient matrices of unknown parameter vectors in every iteration are of full column rank. That is, $\boldsymbol{A}_{\hat{\boldsymbol{\theta}}}$ and $[\boldsymbol{A}_{\hat{\boldsymbol{T}}};\boldsymbol{A}_{\mathrm{eq}}]$ are of full column rank. Based on the properties of the determinant and the Kronecker product, it is easy to prove that the assumption that the second matrix is of full column rank is equivalent to that the transformation matrix $\boldsymbol{T}$ is non-singular. It must be remembered that all derivations are based on the existence of bilinear Eq. (9), which is enforced by the identifiability of the augmented system. Since the cost function $F(\boldsymbol{\theta}, \boldsymbol{T})$ is continuous and differentiable with respect to unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{T}$, it is almost certain that for a partial optimum rank$[\lambda^d(\lambda\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}_{\mathrm{d}}\boldsymbol{B}]=n$ there exists a non-empty set S of starting points in the domain of attraction of $F(\boldsymbol{\zeta}^*)$, such that $\boldsymbol{\zeta}^*$ is the output of our ACS algorithm, because the bilinear Eq. (9) holds. Otherwise, the identifiability assumption is not verified any more. Therefore, if the starting point $\boldsymbol{\zeta}_0=(\boldsymbol{\theta}_0, \boldsymbol{T}_0)$ is chosen well, the full column rank assumption of the two coefficient matrices in every iteration is satisfied. Then the sequence $\{\boldsymbol{\zeta}_i\}_{i=1, 2, \ldots}$ generated by our ACS algorithm converging to partial optimum $\boldsymbol{\zeta}^*=(\boldsymbol{\theta}^*, \boldsymbol{T}^*)$ can be proved by the procedures presented in Gorski *et al.* (2007).