Subspace identification of distributed clusters of homogeneous systems

Yu, Chengpu; Verhaegen, Michel

DOI
10.1109/TAC.2016.2555961

Publication date
2017

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Automatic Control

This manuscript has been accepted by IEEE Transactions on Automatic Control with the digital object identifier (DOI) 10.1109/TAC.2016.2555961.

Subspace Identification of Distributed Clusters of Homogeneous Systems

Chengpu Yu, Michel Verhaegen

***Abstract*—This note studies the identification of a network comprised of interconnected clusters of LTI systems. Each cluster consists of homogeneous dynamical systems, and its interconnections with the rest of the network are unmeasurable. A subspace identification method is proposed for identifying a single cluster using only local input and output data. With the topology of the concerned cluster being available, all the LTI systems within the cluster are decoupled by taking a transformation on the state, input and output data. To deal with the unmeasurable interconnections between the concerned cluster and the rest of the network, the Markov parameters of the decoupled LTI systems are identified first by solving a nuclear-norm regularized convex optimization, following the state-space realization of a single LTI system within the cluster by solving another nuclear-norm regularized optimization problem. The effectiveness of the proposed identification method is demonstrated by a simulation example.**

***Index Terms*—General network topology, decomposable systems, nuclear norm optimization.**

I. Introduction

The emergence of networked systems has stimulated a surge in research on distributed control and estimation problems. Despite many contributions on the control and estimation part, the identification of networked systems is far less developed. This note studies the distributed identification of large-scale interconnected system networks which can be carried out using either parametric identification methods or subspace identification methods.

To date, there have been several parametric approaches for the distributed identification of networked systems. The instrumental variable technique was used in [1] to identify distributed systems with identical subsystems. It requires all system inputs and outputs to form the instrumental variables, making this approach not scalable. In [2], using local system observations, the classic prediction-error method for closed-loop identification was employed to estimate a particular module in the network with some specific interconnection structure. In this method, the interconnection signals between the systems in a network were assumed to be measurable. This condition is easily violated when dealing with network approximations of systems governed by PDEs. The identification of a 1D heterogenous networked system, where the interconnection variables between the local systems are unmeasurable, was discussed in [3]. This approach exploits the parametrization of the SSS (Sequentially Semi-Separable) system matrices to derive an iterative extended Kalman filtering solution, whereby each iteration has linear computational complexity; however, the overall procedure is non-convex. A recent contribution to address the identification problem with missing interconnection measurements of the networked systems was considered in [4]. Though an efficient ADMM solution was presented, this approach is also non-convex in nature.

For a state-space represented network approximation of a continuous-time physical phenomenon (e.g., heat, wave or wind) or a continuous deformable membrane governed by PDEs [5], [6], the associated inner states are usually unmeasurable, which imposes a great challenge on the local system identification problem. To compensate for the missing information, the unknown interconnection signals were approximated by linear combinations of the local inputs and outputs [7]. The search for the related neighbors makes the problem non-convex. As an alternative, a nuclear-norm optimization approach was developed in [8] by exploiting the fact that the transfer function of the local dynamic is of low-order while that of the global local dynamic is of high-order; however, this approach is unable to identify the interconnections between neighboring systems.

In this note, we focus on the identification of a network consisting of distributed clusters of homogeneous dynamical systems. The concerned networked system belongs to the so-called α-heterogeneous system [9] or decomposable system [10]. It is remarked that the finite network approximation of a deformable membrane in adaptive-optics applications is an example of the network consisting of distributed clusters, for which the dynamics of boundary systems are different from those of the inner systems. Although several state-space model identification algorithms have been investigated for decomposable systems in [11], [12], the realization of the system matrices requires the solution of a Blinear Matrix Inequality (BMI) problem. As a consequence, the convex nature of the subspace identification is destroyed.

The goal of this note is to derive a *convex* optimization approach for the subspace identification problem of distributed clusters of homogeneous systems, with each cluster being interconnected by an arbitrary but known bidirected topology. Here we build upon our recent work in [13] where the problem of identifying the local system dynamics in a 1D distributed network was considered. Different from the local identification methods in [7], [8], the unknown interconnection signals to the concerned cluster are not approximated by linear combinations of local (active) inputs and outputs. Instead, by exploiting the spatial and temporal low-rank properties of the unknown

C. Yu and M. Verhaegen are with the Delft Center for Systems and Control, Delft University, Delft 2628CD, Netherlands (c.yu-4@tudelft.nl, m.verhaegen@tudelft.nl)

The work of C. Yu and M. Verhaegen is sponsored by the European Research Council, Advanced Grant Agreement No. 339681.

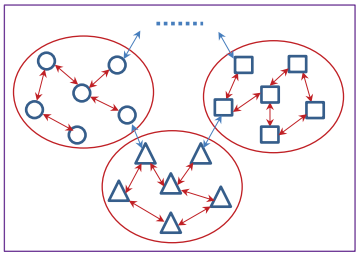

Fig. 1. Diagram of a network comprised of distributed clusters. The set of interconnected systems in an ellipse is called "local cluster", while the "network" refers to the union of all the distributed local clusters.

inputs that are revealed in the data equation of a state-space model, a nuclear-norm regularized optimization approach is developed. Numerical simulations show that the proposed approach can effectively handle the adverse effects caused by unmeasurable interconnection signals.

The rest of the note is organized as follows. Section II formulates the distributed identification problem. Section III presents an identification framework for distributed clusters homogeneous systems. In Section IV, a simulation example is provided to demonstrate the performance of the proposed identification algorithm, followed by the conclusions in Section V.

The following notations are adopted throughout the note. The lowercase (uppercase) x (X) is used to denote a vector (matrix). $X(m:n, p:q)$ is the submatrix of X with rows indexed from m to n and columns indexed from p to q, this notation is in accordance with Matlab style. $\text{diag}(x)$ denotes a diagonal matrix with its diagonal entries equal to the components of x. The superscripts ^{T} and $^{-1}$ are transpose and inverse operators, respectively. I is an identity matrix of appropriate dimension. $\otimes$ stands for the Kronecker product. $\|X\|_F$ and $\|X\|_*$ denote the Frobenius norm and the nuclear norm of the matrix X, respectively. $\mathbb{R}$ and $\mathbb{B}$ stand for the real and binary number sets, respectively.

II. Problem formulation

We consider the identification of a local cluster of homogeneous systems. As shown in Fig. 1, a local cluster refers to the set of systems in an ellipse. Suppose that the considered local cluster consists of N systems which are connected in a topology with its adjacency matrix being denoted by $P \in \mathbb{B}^{N\times N}$, and there are M systems in other clusters of the network that connect with the considered local cluster with the corresponding interconnection matrix denoted by $R \in \mathbb{B}^{N\times M}$.

The dynamics of the local cluster is described by the following state-space form:

$$\begin{aligned} x(k+1) &= (I\otimes A_a + P\otimes A_b)\,x(k) + (R\otimes A_b)\,v(k) \\ &\quad + (I\otimes B)\,u(k) \\ y(k) &= (I\otimes C_a + P\otimes C_b)\,x(k) + (R\otimes C_b)\,o(k) + w(k), \end{aligned} \tag{1}$$

where $x(k) = \left[x_1^T(k)\ \cdots\ x_N^T(k)\right]^T$, $u(k) = \left[u_1^T(k)\ \cdots\ u_N^T(k)\right]^T$, $y(k) = \left[y_1^T(k)\ \cdots\ y_N^T(k)\right]^T$ and $w(k) = \left[x_1^T(k)\ \cdots\ x_N^T(k)\right]^T$ are respectively the local-cluster state, input, output and measurement noise; $x_i(k) \in \mathbb{R}^n, u_i(k) \in \mathbb{R}^m, y_i(k) \in \mathbb{R}^p$ and $w_i(k) \in \mathbb{R}^p$, for $i = 1, \cdots, N$, are respectively the state, input, output and measurement noise of the i-th system operating within the local cluster; $A_a, A_b \in \mathbb{R}^{n\times n}$, $C_a, C_b \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{n\times m}$ are system matrices of a single system within the local cluster; $v(k) = \left[v_1^T(k)\ \cdots\ v_M^T(k)\right]^T$ and $o(k) = \left[o_1^T(k)\ \cdots\ o_M^T(k)\right]^T$ are the stacked state and output that come from the neighboring clusters of the network; $v_i(k) \in \mathbb{R}^n$ and $o_i(k) \in \mathbb{R}^p$, for $i = 1, \cdots, M$, are state and output of the i-th system from the neighboring clusters that connects to the concerned local cluster.

As shown in the state-space model (1), the unknown states and outputs of the neighboring clusters influence the dynamics of the concerned local cluster, causing the corresponding identification problem to be challenging. In many practical systems such as deformable mirror of adaptive optics, the controlled inputs of individual systems act locally; therefore, system inputs in the considered model in (1) are assumed to be decoupled. In addition, the matrices C_a and C_b are considered to be fat matrices; otherwise, the system states might be directly observed from their outputs, and the corresponding local system identification may become trivial.

For the state-space model of a local cluster in (1), the following assumptions are made.

A1. The local-cluster model in (1) is minimal and stable.
A2. The input of the local cluster, $u(k)$, is persistently exciting of any finite order [14], [15].
A3. The measurement noise of the local cluster, $w(k)$, is a white noise sequence and is uncorrelated with the input $u(k)$.
A4. The topology of the local cluster is bidirectional, i.e., the associated adjacency matrix P is symmetric.

Assumptions A1-A3 are standard assumptions for system identification problems. In Assumption A4, the considered network topology is assumed to be bidirectional, which is common in the network approximations of PDE systems [5], [6]. The bidirectional-topology assumption makes the concerned identification problem more complicated than those with directed topologies in [16], [17]; however, it implies that the adjacency matrix P is diagonalizable, which will be used for the subspace identification method in this note.

The *problem of interest* is stated as follows. Given the input-output data $\{u(k), y(k)\}_{k=1}^T$ and the eigenvalue decomposition $P = U\Lambda U^T$ with Λ being a real diagonal matrix $\Lambda = \text{Diag}\,(\lambda_1, \cdots, \lambda_N)$, the goal is to estimate the Markov parameters $(C_a+\lambda_i C_b)(A_a+\lambda_i A_b)^j B$ for $i = 1, \cdots, N, j = 0, 1, \cdots$ and subsequently use these estimated Markov parameters to estimate the individual system matrices A_a, A_b, B, C_a, C_b up to a similarity transformation, i.e. the estimated system matrices satisfies that $\hat{A}_a = QA_aQ^{-1}, \hat{A}_b = QA_bQ^{-1}, \hat{B} = QB, \hat{C}_a = C_aQ^{-1}, \hat{C}_b = C_bQ^{-1}$ with $Q \in \mathbb{R}^{n\times n}$ being a nonsingular ambiguity matrix.

It is remarked that, compared with the 1D distributed-system identification problem in [3], [13], the system matrices in (1) may not be sparse and banded. In this note, an approximate

solution that uses finite length data sequence is considered. This is a direct consequence of the use of the nuclear-norm optimization. The analysis of the consistency of the estimated models is out of scope in this note, though the numerical simulation section highlights that accurate estimates can be obtained.

III. Subspace identification of local clusters of homogeneous systems

A. Identification of Markov parameters

Under Assumption A4, the adjacency matrix P is symmetric, so it is always diagonalizable with its eigenvectors being mutually orthogonal and its eigenvalues being real [18]. Then, according to the properties of decomposable systems in [12], the system matrices in (1) can be transformed into block diagonal matrices by a transformation of the state, input and output data as summarized in the following lemma.

Lemma 1. *[12] Let $P = U\Lambda U^T$ with U an orthogonal matrix and Λ a real diagonal matrix. The local cluster in* (1) *can be equivalently transformed into:*

$$\begin{aligned}\hat{x}(k+1) &= \underbrace{(I\otimes A_a + \Lambda\otimes A_b)}_{\mathbf{A}}\hat{x}(k) + \underbrace{\left(U^TR\otimes A_b\right)}_{\mathbf{R}}v(k)\\ &+ \underbrace{(I\otimes B)}_{\mathbf{B}}\hat{u}(k),\\ \hat{y}(k) &= \underbrace{(I\otimes C_a + \Lambda\otimes C_b)}_{\mathbf{C}}\hat{x}(k) + \underbrace{\left(U^TR\otimes C_b\right)o(k) + \hat{w}(k)}_{\hat{e}(k)}.\end{aligned} \tag{2}$$

where $\hat{x}(k) = (U^T\otimes I)x(k)$, $\hat{u}(k) = (U^T\otimes I)u(k)$, $\hat{y}(k) = (U^T\otimes I)y(k)$ and $\hat{w}(k) = (U^T\otimes I)w(k)$. The following system matrices are block diagonal: $\mathbf{A} = I\otimes A_a + \Lambda\otimes A_b$, $\mathbf{B} = I\otimes B$, $\mathbf{C} = I\otimes C_a + \Lambda\otimes C_b$; however, the matrix $\mathbf{R}$ may be fully filled. The unknown outputs from neighboring clusters are absorbed into the measurement noise, which yields a combined noise denoted by $\hat{e}(k)$.

It is noted that the state-space model with block-diagonal system matrices in (2) enables us to explicitly present its data equation with finer structures.

Given T pairs of input and output data, the data equation of the state-space model in (2) can be written as

$$\hat{Y}_{s,r} = \mathcal{O}_s\hat{X}_r + \mathbf{T}_{u,s}\hat{U}_{s,r} + \mathbf{T}_{v,s}\hat{V}_{s,r} + \hat{E}_{s,r}, \tag{3}$$

where

$$\hat{Y}_{s,r} = \begin{bmatrix}\hat{y}(1) & \hat{y}(2) & \cdots & \hat{y}(r)\\ \hat{y}(2) & \hat{y}(3) & \iddots & \hat{y}(r+1)\\ \vdots & \iddots & \iddots & \vdots\\ \hat{y}(s) & \hat{y}(s+1) & \cdots & \hat{y}(T).\end{bmatrix} \in \mathbb{R}^{sNn\times sNn}$$

with the subscripts s and r representing respectively the numbers of vertical and horizontal blocks, satisfying that $s + r = T + 1$; $\hat{U}_{s,r}$, $\hat{U}_{s,r}$ and $\hat{W}_{s,r}$ are defined similarly as $\hat{Y}_{s,r}$; $\hat{X}_r = \begin{bmatrix}\hat{x}(1) & \cdots & \hat{x}(r)\end{bmatrix}$;

$$\mathcal{O}_s = \begin{bmatrix}\mathbf{C}\\ \mathbf{CA}\\ \vdots\\ \mathbf{CA}^{s-1}\end{bmatrix} \in \mathbb{R}^{sNn\times Nn}$$

is the extended observability matrix;

$$\mathbf{T}_{u,s} = \begin{bmatrix}0 & & & \\ \mathbf{CB} & 0 & & \\ \vdots & \ddots & \ddots & \\ \mathbf{CA}^{s-2}\mathbf{B} & \cdots & \mathbf{CB} & 0\end{bmatrix} \in \mathbb{R}^{sNp\times sNm},$$

and

$$\mathbf{T}_{v,s} = \begin{bmatrix}0 & & & \\ \mathbf{CR} & 0 & & \\ \vdots & \ddots & \ddots & \\ \mathbf{CA}^{s-2}\mathbf{R} & \cdots & \mathbf{CR} & 0\end{bmatrix} \in \mathbb{R}^{sNp\times sMn}.$$

The term $\mathcal{O}_s\hat{X}_r + \mathbf{T}_{v,s}\hat{V}_{s,r}$ on the right-hand side of the data equation in (3) is unknown. However, it has a low rank under some conditions.

Lemma 2. *The sum $\mathcal{O}_s\hat{X}_r + \mathbf{T}_{v,s}\hat{V}_{s,r}$ in the data equation* (3) *has a rank satisfying*

$$\textit{rank}\left(\mathcal{O}_s\hat{X}_r + \mathbf{T}_{v,s}\hat{V}_{s,r}\right) \le Nn + (s-1)Mn. \tag{4}$$

Proof: The above results can be straightforwardly obtained by considering the following rank properties:

$$\begin{aligned}&\text{rank}\left(\mathcal{O}_s\hat{X}_r + \mathbf{T}_{v,s}\hat{V}_{s,r}\right) \le \text{rank}\left(\mathcal{O}_s\hat{X}_r\right) + \text{rank}\left(\mathbf{T}_{v,s}\hat{V}_{s,r}\right),\\ &\text{rank}\left(\mathcal{O}_s\hat{X}_r\right) \le \text{rank}\left(\mathcal{O}_s\right) \le Nn,\\ &\text{rank}\left(\mathbf{T}_{v,s}\hat{V}_{s,r}\right) \le \text{rank}\left(\mathbf{T}_{v,s}\right) \le (s-1)Mn.\end{aligned}$$

■

From the above lemma, we can see that the sum $\mathcal{O}_s\hat{X}_r + \mathbf{T}_{v,s}\hat{V}_{s,r}$ is of low rank when the dimension parameters s and M satisfy

$$sNp > Nn + (s-1)Mn. \tag{5}$$

It is obvious that, when $N > M$, we can always find a positive integer s such that the above inequality holds. On the other hand, for a fixed dimension parameter s, when N is much larger than M, the associated low rank property will be more significant. In many practical networked systems, it is usually possible to find a local cluster for which the number of systems inside the cluster is larger than the number of interconnected systems outside the cluster. Examples are 1D networked systems with each system connecting with its two-neighboring systems and 2D networked systems with each system connecting with its four-neighboring systems.

Inspired by the idea of the N2SID method [19] and its extension in [13], by taking into account the low rank property of the sum $\mathcal{O}_s\hat{X}_r + \mathbf{T}_{v,s}\hat{V}_{s,r}$, the Markov parameters inside the

block Toeplitz matrix $\mathbf{T}_{u,s}$ can be estimated by the nuclear-norm optimization as follows:

$$\min_{\tilde{Y}_{s,r}\in\mathcal{H}_{s,r},\mathbf{T}_{u,s}\in\mathcal{T}_s}\sum_{k=1}^{T}\|\hat{y}(k)-\tilde{y}(k)\|_F^2+\alpha\|\tilde{Y}_{s,r}-\mathbf{T}_{u,s}\hat{U}_{s,r}\|_*, \tag{6}$$

where α is a regularization parameter to balance the least-squares term and the nuclear-norm regularization term; $\mathcal{H}_{s,r}$ denotes the set of block Hankel matrices having the form of $\hat{Y}_{s,r}$ and $\mathcal{T}_s$ represents the set of block Toeplitz matrices having the same structure as $\mathbf{T}_{u,s}$; $\tilde{Y}_{s,r}$ is a block Hankel matrix constructed by $\{\tilde{y}(k)\}_{k=1}^{T}$, which are variables used to represent noise-free measurements. The least-squares term is used to tackle the measurement-noise effect, while the nuclear-norm regularization term is adopted to deal with the unmeasurable states from neighboring clusters in the network.

It is noteworthy that the Markov parameters, as block entries of $\mathbf{T}_{u,s}$, are further block diagonal matrices. More specifically, $\mathbf{CA}^i\mathbf{B}$ can be explicitly expressed as

$$\begin{aligned}\mathbf{CA}^i\mathbf{B}=&\text{BDiag}\left[(C_a+\lambda_1C_b)(A_a+\lambda_1A_b)^iB,\cdots,\right.\\&\left.(C_a+\lambda_NC_b)(A_a+\lambda_NA_b)^iB\right],\end{aligned} \tag{7}$$

where BDiag$[\cdot]$ denotes a block-diagonal matrix.

Instead of directly estimating the coarse Markov parameters $\{\mathbf{CA}^i\mathbf{B}\}_{i=0}^{s-2}$ in (6), we regard the finer Markov parameters $\{(C_a+\lambda_jC_b)(A_a+\lambda_jA_b)^iB\}_{i=0,j=1}^{s-2,N}$ as the variables to be determined. Since such a finer parametrization of $\mathbf{T}_{s,r}$ does not destroy the convexity of (6), the finer Markov parameters $\{(C_a+\lambda_jC_b)(A_a+\lambda_jA_b)^iB\}_{i=0,j=1}^{s-2,N}$ can be reliably estimated. In addition, by considering the finer Markov parameters to be variables, the total number of variables involved the optimization problem in (6) can be reduced. The parametrization of $\tilde{Y}_{s,r}$ requires TNp variables, and the parametrization of $\mathbf{T}_{u,s}$ requires $(s-1)Npm$ variables. It is obvious that the number of decision variables in (6) can be very large once any or a combination of the dimension parameters s, T or N has a large value. To cope with this high-dimension optimization problem, the ADMM algorithm in [20] will be adopted for numerical simulations.

B. Estimation of system matrices

After having obtained the Markov parameters $\{(C_a+\lambda_jC_b)(A_a+\lambda_jA_b)^iB\}_{i=0,j=1}^{s-2,N}$, this subsection will be devoted to the estimation of system matrices $\{C_a, C_b, A_a, A_b, B\}$. Since the pattern matrix P is accessible beforehand, the values of $\{\lambda_i\}_{i=1}^{N}$ are available.

For notational simplicity without sacrificing the essence of the method, we shall present the estimation of system matrices using the Markov parameters up to the sixth moment, i.e. $\{(C_a+\lambda_jC_b)(A_a+\lambda_jA_b)^iB\}_{i=0,j=1}^{6,N}$.

Denote $C_j=C_a+\lambda_jC_b$ and $A_j=A_a+\lambda_jA_b$. Based on the known Markov parameters $\{C_jA_j^iB\}_{i=1}^{6}$ for $j\in\{1,\cdots,N\}$, we define the following block Hankel matrix:

$$\begin{aligned}H_j&=\begin{bmatrix}C_jB & C_jA_jB & C_jA_j^2B & C_jA_j^3B\\ C_jA_jB & C_jA_j^2B & C_jA_j^3B & C_jA_j^4B\\ C_jA_j^2B & C_jA_j^3B & C_jA_j^4B & C_jA_j^5B\\ C_jA_j^3B & C_jA_j^4B & C_jA_j^5B & C_jA_j^6B\end{bmatrix}\\&=\begin{bmatrix}C_j\\C_jA_j\\C_jA_j^2\\C_jA_j^3\end{bmatrix}\begin{bmatrix}B & A_jB & A_j^2B & A_j^3B\end{bmatrix}.\end{aligned} \tag{8}$$

In developing the estimation method for the system matrices, use will be made of the following assumption.

A5. The extended observability matrix $\begin{bmatrix}C_j\\C_jA_j\\C_jA_j^2\\C_jA_j^3\end{bmatrix}$ and the extended controllability matrix $\begin{bmatrix}B & A_jB & A_j^2B & A_j^3B\end{bmatrix}$, for $j=1,2,\cdots,N$, are of full column and row ranks, respectively

Under Assumption A5 and when $n<\min\{4p,4m\}$, it is easy to see that rank $(H_j)=n$, and the extended controllability matrix $\begin{bmatrix}B & A_jB & A_j^2B & A_j^3B\end{bmatrix}$ has the same row subspace as H_i.

Taking the SVD decomposition of H_i yields that

$$H_j=\begin{bmatrix}U_s^j & U_n^j\end{bmatrix}\begin{bmatrix}\Sigma_j & \\ & O\end{bmatrix}\begin{bmatrix}V_s^{j,T}\\V_n^{j,T}\end{bmatrix}, \tag{9}$$

where $\Sigma_j\in\mathbb{R}^{n\times n}$ is a nonsingular diagonal matrix, $U_s^j\in\mathbb{R}^{4p\times n}$, $U_n^j\in\mathbb{R}^{4p\times(4p-n)}$, $V_s^j\in\mathbb{R}^{4m\times n}$ and $V_n^j\in\mathbb{R}^{4m\times(4m-n)}$ are partial orthogonal matrices. Since H_j is known, the matrices on the right-hand side of (9) are considered to be known as well in the sequel.

By equations (8) and (9), we can derive that

$$\begin{bmatrix}B & A_jB & A_j^2B & A_j^3B\end{bmatrix}V_n^j=0,\quad j=1,\cdots,N. \tag{10}$$

Inspired by the subspace-based blind identification method in [21], we shall parameterize the row space of H_j for $j=1,\cdots,N$. Define a parametric matrix $\Phi\in\mathbb{R}^{n\times 15m}$ as

$$\begin{aligned}\Phi=\big[\ &B \quad A_aB \quad A_bB \quad A_a^2B \quad A_aA_bB\\ &A_bA_aB \quad A_b^2B \quad A_a^3B \quad A_a^2A_bB \quad A_aA_bA_aB\\ &A_aA_b^2B \quad A_bA_a^2B \quad A_bA_aA_bB \quad A_b^2A_aB \quad A_b^3B\big].\end{aligned}$$

For each $j\in\{1,\cdots,N\}$, there exists a constant matrix $\Theta_j\in\mathbb{R}^{15m\times 4m}$ satisfying

$$\begin{bmatrix}B & A_jB & A_j^2B & A_j^3B\end{bmatrix}=\Phi\Theta_j. \tag{11}$$

By the relation between Φ and $\begin{bmatrix}B & A_jB & A_j^2B & A_j^3B\end{bmatrix}$, the constant matrix Θ_j can be easily determined; hence Θ_j is considered to be known in the sequel.

Substituting (11) into (10) yields that

$$\Phi\Theta_jV_n^j=0 \text{ for } j=1,\cdots,N.$$

Then, by stacking the above equations for all $j \in \{1, \cdots, N\}$, we can obtain that

$$\Phi \underbrace{\begin{bmatrix} \Theta_1 & \Theta_2 & \cdots & \Theta_N \end{bmatrix}}_{\mathbf{\Theta}} \underbrace{\begin{bmatrix} V_n^1 & & & \\ & V_n^2 & & \\ & & \ddots & \\ & & & V_n^N \end{bmatrix}}_{\mathbf{V}_n} = 0. \tag{12}$$

In the above equation, the matrices $\mathbf{\Theta}$ and $\mathbf{V}_n$ are known, while Φ is the parametric matrix to be estimated.

Next, we shall analyze the properties of the solution to equation (12). Although the matrix $\mathbf{\Theta}$ may be a fat matrix, it is generally rank deficient.

Lemma 3. *For the coefficient matrix $\mathbf{\Theta} \in \mathbb{R}^{15m \times 4Nm}$ in (12), it has a rank satisfying*

$$\textit{rank}\,(\mathbf{\Theta}) \leq 10m, \tag{13}$$

where the equality holds if the adjacency matrix P has more than 4 different eigenvalues.

Proof: By re-ordering the block columns of $\mathbf{\Theta}$ that is determined by equation (11), we can obtain the following matrix

$$\left[\begin{array}{c|ccc|ccc|c} I \cdots I & & & & & & & \\ \hline & I & \cdots & I & & & & \\ & \lambda_1 I & \cdots & \lambda_N I & & & & \\ \hline & & & & I & \cdots & I & \\ & & & & \lambda_1 I & \cdots & \lambda_N I & \\ & & & & \lambda_1 I & \cdots & \lambda_N I & \\ & & & & \lambda_1^2 I & \cdots & \lambda_N^2 I & \\ \hline & & & & & & & \cdots \end{array}\right], \tag{14}$$

where the identity matrix I has size $m \times m$. From the structure of the above matrix, it is easy to see that rank $(\mathbf{\Theta}) \leq (1+2+3+4)m = 10m$. According to the properties of the Vandermonde matrix [18], we can find that rank $(\mathbf{\Theta}) = 10m$ as long as there are at least 4 different elements in $\{\lambda_1, \cdots, \lambda_N\}$, namely the adjacency matrix P has more than 4 different eigenvalues. ∎

The above lemma indicates that the matrix $\mathbf{\Theta}$ is always rank deficient, regardless of the topology of the local cluster.

By Lemma 3, the matrix $\mathbf{\Theta V}_n \in \mathbb{R}^{15m \times (4m-n)N}$ in equation (12) has the rank property satisfying

$$\text{rank}\,(\mathbf{\Theta V}_n) \leq \text{rank}\,(\mathbf{\Theta}) \leq 10m.$$

Therefore, when $n + 10m < 15m$, the parametric matrix Φ cannot be determined from equation (12) up to an $n \times n$ non-singular ambiguity matrix. However, by taking an insight into the structure of Φ, we can construct a matrix $\Gamma(\Phi) \in \mathbb{R}^{3n \times 7m}$ from the block entries of Φ such that it is of low rank:

$$\begin{aligned} \Gamma(\Phi) &= \begin{bmatrix} B & A_aB & A_bB & \cdots & A_b^2B \\ A_aB & A_a^2B & A_aA_bB & \cdots & A_aA_b^2B \\ A_bB & A_bA_aB & A_b^2B & \cdots & A_b^3B \end{bmatrix} \\ &= \begin{bmatrix} I \\ A_a \\ A_b \end{bmatrix} \begin{bmatrix} B & A_aB & A_bB & \cdots & A_b^2B \end{bmatrix}. \end{aligned} \tag{15}$$

Then, by combining the equation in (12) and the low rank property of $\Gamma(\Phi)$, we propose a nuclear-norm regularized optimization formulation as follows:

$$\begin{aligned} \min_{\Phi} \quad & \|\Phi\mathbf{\Theta}\mathbf{V}_n\|_F^2 + \beta\|\Gamma(\Phi)\|_* \\ s.t. \quad & \Phi M_l = M_r \end{aligned} \tag{16}$$

where β is a regularization parameter to make a tradeoff between the least-squares term and the nuclear-norm term, $M_l \in \mathbb{R}^{15m \times n}$ and $M_r \in \mathbb{R}^{n \times n}$ are constant and known matrices, and the equality constraint is provided to avoid the trivial solution of Φ. In principle, M_l has to be chosen such that it has full column rank and each of its columns is non-orthogonal to the row space of Φ, while M_r can be chosen as any non-singular matrix. In numerical simulations, the matrices M_l and M_r are randomly generated so that they satisfy the above mentioned properties with probability one [22].

Based on the estimate of Φ obtained from (16), we shall estimate the matrices A_a, A_b and B. Let the SVD decomposition of $\Gamma(\Phi)$ be given as follows:

$$\Gamma(\Phi) = \begin{bmatrix} U_s & U_n \end{bmatrix} \begin{bmatrix} \Sigma_s & \\ & \Sigma_n \end{bmatrix} \begin{bmatrix} V_s^T \\ V_n^T \end{bmatrix}, \tag{17}$$

where $U_s \in \mathbb{R}^{3n \times n}$, $U_n \in \mathbb{R}^{3n \times 2n}$, $V_s \in \mathbb{R}^{7m \times n}$ and $V_n \in \mathbb{R}^{7m \times (7m-n)}$ are partial orthogonal matrices, $\Sigma_s \in \mathbb{R}^{n \times n}$ and Σ_n are diagonal matrices with the nonzero entries of Σ_s being larger than those of Σ_n. Then, the estimates of A_a, A_b, B are respectively set to

$$\begin{aligned} &\hat{A}_a = U_s(n+1:2n,:), \;\; \hat{A}_b = U_s(2n+1:3n,:), \\ &\hat{B} = V_s^T(:, 1:m). \end{aligned} \tag{18}$$

Denote $M_j^i = (C_a + \lambda_j C_b)(A_a + \lambda_j A_b)^i B$ for $j = 1, \cdots, N$ and $i = 0, \cdots, 6$. The matrices C_a and C_b can be estimated by solving the following least-squares minimization problem:

$$\min_{C_a, C_b} \sum_{i=0,j=1}^{6,N} \left\| M_j^i - \begin{bmatrix} C_a & C_b \end{bmatrix} \begin{bmatrix} (A_a + \lambda_j A_b)^i B \\ \lambda_j (A_a + \lambda_j A_b)^i B \end{bmatrix} \right\|_F^2. \tag{19}$$

C. Summary of the identification approach

The subspace identification of a local cluster of homogeneous systems is carried out by sequentially estimating the Markov parameters and the system matrices. To ease the reference, the identification approach is summarized in Algorithm 1.

Algorithm 1: Identification of local clusters of homogeneous systems	
1)	Transform the local-cluster model into (2), as shown in Lemma 1;
2)	Estimate parameters $\{C_a + \lambda_j C_b)(A_a + \lambda_j A_b)^i B\}_{i=0,j=1}^{s-2,N}$ by solving the nuclear-norm optimization problem in (6);
3)	Estimate the parametric matrix Φ by solving (16);
4)	Take the SVD decomposition of $\Gamma(\Phi)$ shown in (17), and extract the estimates of A_a, A_b and B, as shown in (18).
5)	Estimate C_a and C_b by solving (19).

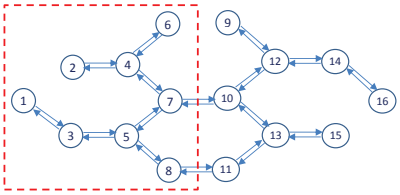

Fig. 2. The networked system considered in a simulation example.

IV. Numerical simulations

In this section, a numerical simulation example is provided to demonstrate the effectiveness of the proposed identification method. In the simulation, the system inputs and measurement noises are generated randomly, and 2000 pairs of input-output data are adopted for system identification. The regularization parameters in (6) and (16) are empirically set to $\alpha = 0.1, \beta = 0.01$. The high-dimension optimization problem in (6) is solved using the ADMM algorithm [20], while the small-scale convex optimization problem in (16) is solved using the CVX toolbox [23].

To assess the performance of Algorithm 1, the normalized impulse-response-fitting criterion is defined. For the matrix triplet $\{C_a, A_a, B\}$, the associated impulse-response-fitting criterion is defined as

$$\frac{1}{K}\sum_{j=1}^{K}\frac{\sum_{i=0}^{10}\|C_aA_a^iB-\hat{C}_a^j\left(\hat{A}_a^j\right)^i\hat{B}^j\|_F}{\sum_{i=0}^{10}\|C_aA_a^iB\|_F}, \qquad (20)$$

where K is the number of Monte-Carlo trials which is set to $K = 50$ and $\{\hat{C}_a^j, \hat{A}_a^j, \hat{B}^j\}$ are the j-th estimates of $\{C_a, A_a, B\}$, respectively. The normalized impulse-response-fitting criteria for the matrix triplets $\{C_b, A_a, B\}$, $\{C_a, A_b, B\}$ and $\{C_b, A_b, B\}$ are defined similarly.

In this simulation, we consider the identification of a local cluster of the network as shown in Fig. 2. The corresponding system matrices are defined as:

$$A_a = \begin{bmatrix} 0.3695 & -0.2017 \\ -0.1817 & 0.3209 \end{bmatrix}, A_b = \begin{bmatrix} 0.1699 & -0.1078 \\ 0.2842 & 0.1815 \end{bmatrix},$$
$$C_a = \begin{bmatrix} 0.6002 & 0.0163 \end{bmatrix}, C_b = \begin{bmatrix} 0.9817 & -0.3902 \end{bmatrix},$$
$$B = \begin{bmatrix} -0.5630 \\ -0.9674 \end{bmatrix}.$$

Fig. 3 shows the identification performance of the proposed identification algorithm under different noise levels. It can be observed that the associated estimation errors of impulse responses decrease along with the increase of the signal-to-noise ratio (SNR) when SNR$\leq$40 dB; however, they change slightly when SNR$\geq$50 dB, indicating that the estimates of system matrices are slightly biased. This might be caused by two factors: (a) the provided nuclear-norm optimization problem is a relaxed version of the low-rank optimization problem; (b) the unknown system outputs from neighboring clusters are considered to be measurement noises. Furthermore, it can be found in Fig. 3 that, if the unknown interconnection signals are neglected, i.e. the nuclear norm in (6) is replaced by the Frobenius norm, the associated estimates of system matrices

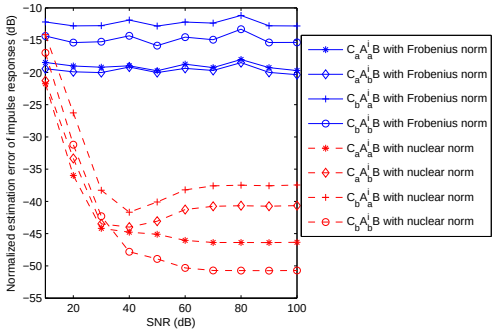

Fig. 3. Identification performance of Algorithm 1. Solid curves are plotted by neglecting the unknown interconnection signals (using the Frobenius-norm regularization), while dashed curves are plotted by our proposed algorithm (using the nuclear-norm regularization).

are very biased, which cannot be mitigated by increasing the SNR. From the above simulation results, we can see that our proposed identification is effective in handling the unknown interconnection signals.

V. Conclusion

In this note, we have developed a subspace identification algorithm for distributed clusters of homogeneous systems. The merit of the presented identification algorithm is that it is carried out by solving convex optimization problems; thus, it can yield more reliable identification results with relation to the general nonlinear optimization solutions. To implement the proposed algorithm, we identify the associated Markov parameters by solving a nuclear-norm regularized optimization problem, followed by the estimation of individual system matrices by solving another nuclear-norm regularized optimization problem. Numerical simulations have been provided to show the effectiveness of the proposed identification algorithm. In the future, by making use of the merit of low-rank optimization, the local identification of distributed heterogeneous systems will be investigated.

References

[1] M. Ali, A. Ali, S. Chughtai, and H. Werner, "Consistent identification of spatially interconnected systems," in *American Control Conference (ACC), 2011*, pp. 3583–3588, June 2011.

[2] P. M. V. den Hof, A. Dankers, P. S. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methodsbasic methods for consistent module estimates," *Automatica*, vol. 49, no. 10, pp. 2994 – 3006, 2013.

[3] J. Rice and M. Verhaegen, "Efficient system identification of heterogeneous distributed systems via a structure exploiting extended kalman filter," *Automatic Control, IEEE Transactions on*, vol. 56, pp. 1713–1718, July 2011.

[4] A. Hansson and M. Verhaegen, "Distributed system identification with admm," *arXiv:1403.4804*, 2014.

[5] U. Khan, J. M. Moura, *et al.*, "Distributing the kalman filter for large-scale systems," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4919–4935, 2008.

[6] T. Ruppel, *Modeling and Control of Deformable Membrane Mirrors*. INTECH Open Access Publisher, 2012.

[7] A. Haber and M. Verhaegen, "Subspace identification of large-scale interconnected systems," *Automatic Control, IEEE Transactions on*, vol. 59, pp. 2754–2759, Oct 2014.

[8] N. Matni and A. Rantzer, "Low-rank and low-order decompositions for local system identification," *arXiv:1403.7175*, 2014.

[9] P. Massioni, "Distributed control for alpha-heterogeneous dynamically coupled systems," *Systems & Control Letters*, vol. 72, pp. 30–35, 2014.

[10] P. Massioni and M. Verhaegen, "Distributed control for identical dynamically coupled systems: A decomposition approach," *Automatic Control, IEEE Transactions on*, vol. 54, pp. 124–135, Jan 2009.

[11] P. Massioni and M. Verhaegen, "Subspace identification of circulant systems," *Automatica*, vol. 44, no. 11, pp. 2825 – 2833, 2008.

[12] P. Massioni and M. Verhaegen, "Subspace identification of distributed, decomposable systems," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pp. 3364–3369, Dec 2009.

[13] C. Yu, M. Verhaegen, and A. Hansson, "Subspace identification of local 1d homogeneous systems," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 892–896, 2015.

[14] L. Ljung, *System Identification: Theory for the user (Second Edition)*. Prentice Hall PTR, 1999.

[15] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, 2007.

[16] P. Torres, J. van Wingerden, and M. Verhaegen, "PO-MOESP subspace identification of directed acyclic graphs with unknown topology," *Automatica*, vol. 53, pp. 60–71, 2015.

[17] P. Torres, J.-W. van Wingerden, and M. Verhaegen, "Hierarchical subspace identification of directed acyclic graphs," *International Journal of Control*, vol. 88, no. 1, pp. 123–137, 2015.

[18] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

[19] M. Verhaegen and A. Hansson, "Nuclear norm subspace identification (n2sid) for short data batches," *CoRR abs/1401.4273*, 2014.

[20] Z. Liu, A. Hansson, and L. Vandenberghe, "Nuclear norm system identification with missing inputs and outputs," *Systems & Control Letters*, vol. 62, no. 8, pp. 605–612, 2013.

[21] C. Yu and M. Verhaegen, "Blind multivariable arma subspace identification," *Automatica*, vol. 66, pp. 3–14, 2016.

[22] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis, Birkhäuser, 2013.

[23] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming.," *Web page and software available at http://stanford. edu/ boyd/cvx*, 2009.