Constrained subspace method for the identification of structured state-space models (cosmos)

Yu, Chengpu; Ljung, Lennart; Wills, Adrian; Verhaegen, Michel

DOI
10.1109/TAC.2019.2957703

Publication date
2020

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Automatic Control

Constrained Subspace Method for the Identification of Structured State-Space Models (COSMOS)

Chengpu Yu, Lennart Ljung, Adrian Wills, and Michel Verhaegen

***Abstract*—In this paper, a unified identification framework called constrained subspace method for structured state-space models (COSMOS) is presented, where the structure is defined by a user specified linear or polynomial parametrization. The new approach operates directly from the input and output data, which differs from the traditional two-step method that first obtains a state-space realization followed by the system-parameter estimation. The new identification framework relies on a subspace inspired linear regression problem which may not yield a consistent estimate in the presence of process noise. To alleviate this problem, the linear regression formulation is imposed by structured and low-rank constraints in terms of a finite set of system Markov parameters and the user specified model parameters. The non-convex nature of the constrained optimization problem is dealt with by transforming the problem into a difference-of-convex optimization problem, which is then handled by the sequential convex programming strategy. Numerical simulation examples show that the proposed identification method is more robust than the classical prediction-error method (PEM) initialized by random initial values in converging to local minima, but at the cost of heavier computational burden.**

***Index Terms*—Subspace identification, Markov-parameter estimation, Hankel matrix factorization**

I. Introduction

Structured state-space (or gray-box) models are popular for describing practical physical models in terms of system parameters having physical interpretation or in terms of network structures [1]–[3]. The identification of structured state-space models using observed input and output data is a fundamental identification problem which has been intensively investigated in the literature [4]–[6].

In the literature, there are two classes of methods for the identification of structured state-space models. The *first class* is to identify the parameterized state-space model directly from observed input and output data using the traditional prediction-error method (PEM) [1]. The PEM has the best possible asymptotic accuracy; however, when the associated optimization problem has many minima, the PEM is sensitive to the selection of the initial parameter estimate. Different from the black-box model, the gray-box usually starts from some physical insight. However when that insight is only restricted to a particular model parametrization with no or limited a priori initial estimates of these parameters, the chances for ending up into local minima increases rapidly. The *second class* is a two-step estimation framework, where the system matrices of a state-space model are identified (up to a similarity transformation) using the input and output data through the subspace identification method, followed by the estimation of system parameters from estimated system matrices [7]. Since the system-parameter estimation problem in the second step does not involve any input and output data, the associated optimization problem has a smaller scale than that of the PEM. Recently, many identification methods [8]–[10] have been developed under this two-step estimation framework.

For the system-parameter estimation in the two-step estimation framework, there are generally two kinds of estimation methods. One is to simultaneously estimate the system parameters and the similarity transformation provided the estimated system matrices (up to a similarity transformation) [8], [11]. This turns out to be a bilinear estimation problem with a nonsingularity constraint on the similarity transformation, for which the nonsingularity constraint is difficult to handle in practice [8]. To avoid estimating the similarity transformation, the system parameters embedded in a structured state-space model can be estimated by the model-matching principle [9], [10], where the difference-of-convex programming (DCP) scheme [10] turns out to be an effective way to handle the non-convex parameter-estimation problem. It is noted that non-convex optimization problems originated from the system identification subject are usually difficult to tackle, and many convexification schemes have been provided for approximating the global solution, in response to the recent special issue of the IEEE Transactions on Automatic Control on "relaxation methods in identification and estimation problem" [12].

Analogous to the PEM, this paper investigates the identification of structured (parameterized) state-space models directly from the input and output data. A unified identification framework is developed for structured state-space models that inherits the features of both the PEM and the subspace method, which is called constrained subspace method for structured state-space models (COSMOS). The COSMOS is developed based on the fact that the output prediction error obtained by a k-step-ahead predictor can be represented as the convolution of a sequence of finite-length Markov parameters and the system input, which enables the accurate identification of finite-length Markov parameters directly from the input and output data.

C. Yu is with School of Automation, Beijing Institute of Technology, Beijing 100081, China (yuchengpu@bit.edu.cn)

L. Ljung is with Division of Automatic Control, Department of Electrical Engineering, Linkoping University, Sweden (lennart.ljung@liu.se)

A. Wills is with the School of Electrical Engineering and Computer Science, University of Newcastle, Australia (adrian.wills@newcastle.edu.au)

M. Verhaegen is with the Delft Center for Systems and Control, Delft University, Delft 2628CD, Netherlands (m.verhaegen@tudelft.nl)

The research leading to these results has received funding from the National Natural Science Foundation of China (NSFC) under Grants No. 61873301, No. 61933002 and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339681.

This idea enables to incorporate the model-matching technique of our previous work [10] to estimate the parameters embedded in structured system matrices and the Markov parameters directly from the input and output data. The novelties of the proposed COSMOS framework against the relevant works are summarized below.

First, the least-squares estimation framework [13] is adopted for estimating the Markov parameters directly from the input and output data, and sufficient conditions for the consistent estimation have been provided. However, when the system input is not a white noise or there exists the process noise, the obtained least-squares solution of the finite-length Markov parameters may not be consistent (see Remark 2). To alleviate this problem, in addition to the lower-triangular block Toeplitz structure of the convolution matrix constructed by Markov parameters adopted in [14], [15], the low-rank property of the Hankel matrix constructed by Markov parameters is also exploited for the Markov-parameter estimation, which results in a rank-constrained least-squares optimization problem. Analogous to the Ho-Kalman realization of a state-space model, the low-rank factorization of the block Hankel matrix involved in the rank-constrained optimization problem provides a foundation for the identification of structured system matrices.

Second, different from DCP-based identification method [10], [16] that is developed under the two-step estimation framework, the COSMOS provides a novel identification framework that estimates the system parameters directly from the observed input and output data. A rank constrained optimization solution is provided for the identification of parameterized state-space models directly from the input and output data [17], which requires to estimate the system parameters as well as the state sequence simultaneously, resulting in a large number of variables to be determined when the data length is very long. In contrast, the system-parameter estimation in the COSMOS is inherently based on the factorization of a Hankel matrix constructed from a finite number of Markov parameters, so that the number of variables to be estimated in the proposed method is independent of the data length, leading to a more tractable identification framework. In addition, different from [10], [16], [17], the COSMOS framework in this paper is applicable for handling polynomially parameterized state-space models such as the system matrices in Kronecker-product forms.

The paper is organized as follows. Section II formulates the identification problem of linearly or polynomially parameterized state-space models. Section III presents a rank-constrained optimization method for the estimation of finite-length Markov parameters and analyzes the estimation consistency. Section IV reviews the estimation of structured matrices using the structured and low-rank matrix factorization method, and its effectiveness is illustrated through a low-rank matrix completion example. Section V provides three simulation examples to show the effectiveness of the proposed method for the identification of finite-length Markov parameters and system parameters, and conclusions are made in Section VI.

For the sake of brevity, the following notations are defined. Vectors and matrices are respectively represented by the lower-case letter a and the upper-case letter A. The trace, transpose and Moore-Penrose pseudo inverse of the matrix A are denoted as $\text{tr}[A]$, A^T and $A^\dagger$, respectively. The Frobenius norm of matrix A is represented as $\|A\|_F$. The singular values of matrix A are denoted by $\sigma_i(A)$, and the nuclear norm of matrix $A \in \mathbb{R}^{n\times n}$ is defined as $\|A\|_* = \sum_{i=1}^n \sigma_i(A)$. The rank operator for the matrix A is denoted as $\text{rank}(A)$. The identity matrix of appropriate size is denoted as I. For the time sequence $x(k)$, a collection of samples from the time k to $k+l$ is denoted as $x(k:k+l) = [x^T(k)\ x^T(k+1)\cdots x^T(k+l)]^T$. The expectation of the stochastic vector $x(k)$ is denoted as $E[x(k)]$, and the operator $\bar{E}\{x(k)\}$ is defined as $\bar{E}[x(k)] = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N E[x(k)]$ with an implied assumption that the limit exists with probability one. The block Hankel matrix of the sequence $x(k)$ (or a sequence of matrices) is defined as

$$\mathcal{H}_{s,h}[x(k:k+h+s)] = \begin{bmatrix} x(k) & x(k+1) & \cdots & x(k+h) \\ x(k+1) & x(k+2) & \ddots & x(k+h+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(k+s) & x(k+s+1) & \cdots & x(k+s+h) \end{bmatrix}.$$

Given a positive integer s, the extended observability matrix of (A, C) is defined as

$$\mathcal{O}_s[A,C] = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix},$$

and the extended controllability matrix of (A, B) is defined as

$$\mathcal{C}_s[A,B] = \begin{bmatrix} B & AB & \cdots & A^sB \end{bmatrix}.$$

The lower-triangular convolution matrix for the state-space model (A, B, C, D) is defined as

$$\mathcal{T}_s[CA^{s-1}B, \cdots, CB, D] = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}B & \cdots & CB & D \end{bmatrix}.$$

II. Problem formulation

The identification of discrete-time state-space models with structured system matrices is considered. Let the discrete-time state-space model be given as

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + w(k) \\ y(k) &= Cx(k) + Du(k) + v(k) \end{aligned} \tag{1}$$

where $x(k) \in \mathbb{R}^n, u(k) \in \mathbb{R}^m, y(k) \in \mathbb{R}^p$ are respectively the system state, input and output; $w(k) \in \mathbb{R}^n$ and $v(k) \in \mathbb{R}^p$ are respectively the process and measurement noises.

When the system matrices (A, B, C, D) have no specific structures, the identification problem becomes a black-box model identification problem for which the system matrices (A, B, C, D) can only be identified up to a similarity transformation using the subspace identification method [2], [18]. However, when the system matrices have structural constraints, the standard subspace method cannot handle the identification

problem. As a matter of fact, the structured state-space system identification problem is inherently a challenging non-convex optimization problem [7], [8], [10].

The structural constraints of the system matrices usually originate from the physical mechanisms of practical systems or the topologies of networked systems. Denote by $\mathcal{S}$ the parameterized matrix set that can be obtained from prior knowledge of the gray-box model to be identified. It follows that the true system matrices (A^*, B^*, C^*, D^*) belong to this set:

$$(A^*, B^*, C^*, D^*) \in \mathcal{S}. \tag{2}$$

Remark 1. *An example of the set $\mathcal{S}$ is the affinely parameterized structure considered in [10], [17]:*

$$\begin{aligned}\mathcal{S} = \Bigg\{ & A(\theta), B(\theta), C(\theta), D(\theta) : A(\theta) = A_0 + \sum_{i=1}^{l} A_i\theta_i, \\ & B(\theta) = B_0 + \sum_{i=1}^{l} B_i\theta_i,\ C(\theta) = C_0 + \sum_{i=1}^{l} C_i\theta_i, \\ & D(\theta) = D_0 + \sum_{i=1}^{l} D_i\theta_i,\ \theta \in \mathbb{R}^l \Bigg\}.\end{aligned} \tag{3}$$

where $\{A_i, B_i, C_i, D_i\}_{i=1}^{l}$ are known matrix bases. The above affinely parameterized structure can be used to represent various interconnection patterns of networked systems, such as 1D line pattern [19], [20], circular pattern [21] and general topologies [22].

In addition, for the two-dimensional state-space model considered in [23], the system matrices of its equivalent one-dimensional model are of Kronecker product forms. The corresponding structured matrix set can be represented as:

$$\begin{aligned}\mathcal{S} = \{ & A, B, C, D : A = A_1 \otimes A_2,\ B = B_1 \otimes B_2, \\ & C = C_1 \otimes C_2,\ D = D_l \otimes D_2\}\end{aligned} \tag{4}$$

where $A_i \in \mathbb{R}^{n_i \times n_i}, B_i \in \mathbb{R}^{n_i \times m_i}, C_i \in \mathbb{R}^{p_i \times n_i}, D_i \in \mathbb{R}^{p_i \times m_i}$ for $i = 1, 2$ such that $n_1 n_2 = n, m_1 m_2 = m, p_1 p_2 = p$. In this example, the system parameters includes all the entries of $\{A_1, A_2, B_1, B_2, C_1, C_2, D_1, D_2\}$.

Given the prior knowledge of the system structure, i.e., $(A, B, C, D) \in \mathcal{S}$, *the problem of interest* is to identify the structured state-space model (1) using the observed input and output data pairs $\{u(k), y(k)\}$. It is remarked that there might exist many state-space models of the same structure that can exactly depict the input-output mapping, and we aim to find one such structured state-space model in this paper. Although the problem formulation above is similar to that in our previous work [10], the contributions of this paper differs in the following two aspects:

1) In [10], the structured system matrices are identified by two successive steps: Markov-parameter estimation and structured Hankel matrix factorization, whereas in the proposed method these two steps are integrated in a single optimization framework so that the influence of Markov-parameter estimation error to the Hankel matrix factorization can be alleviated.
2) The developed identification approach allows more general matrix structures, not only affinely parameterized matrices as concerned in [10].

The following standard assumptions are made throughout the paper.

A1. The process noise $w(k)$ and the measurement noise $v(k)$ are zero-mean white Gaussian noises with the covariance matrix

$$E \begin{bmatrix} w(k_1) \\ v(k_1) \end{bmatrix} \begin{bmatrix} w(k_2) \\ v(k_2) \end{bmatrix}^T = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{k_1,k_2}$$

where δ_{k_1,k_2} is the Kronecker delta function. In addition, the noises $w(k)$ and $v(k)$ are uncorrelated with the input signal $u(k)$.

A2. The state-space model (1) is minimal in the sense that (A, C) is observable and $(A, [B, Q^{1/2}])$ is controllable.

A3. The input signal $u(k)$ is quasi-stationary and persistently exciting [1].

By Assumptions A1-A2, the state-space model (1) has an innovation form [24]:

$$\begin{aligned}\hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + Ke(k) \\ y(k) &= C\hat{x}(k) + Du(k) + e(k)\end{aligned} \tag{5}$$

where $K \in \mathbb{R}^{n \times p}$ is the Kalman gain, $e(k) \in \mathbb{R}^p$ is a zero-mean innovation signal, and $\hat{x}(k)$ is the one-step forward predictor of the state $x(k)$. The innovation signal $e(k)$ is a white noise that is uncorrelated with all the inputs and the past states, i.e.,

$$\begin{aligned}\bar{E}\{e(k)\hat{x}^T(k_0)\} &= 0 \text{ for all } k_0 \le k \\ \bar{E}\{e(k)u^T(k_0)\} &= 0 \text{ for all } k_0 \in \mathbb{N}.\end{aligned} \tag{6}$$

Since the original state-space model and its innovation form have the same system matrices (A, B, C, D), the identification of the system matrices from the innovation model will be considered in the sequel.

In the following, we start with the Markov-parameter estimation problem. This is then extended by a rank constraint on the block Hankel matrix constructed by the Markov parameters. Finally, an integrated identification framework (COSMOS) for the concerned structured state-space model is presented that operates directly on input and output data.

III. Direct estimation of Markov parameters from input-output data

The Markov-parameter sequence of the state-space model (1) is normally of infinite length, which can be calculated by expanding transfer functions [25] or state-space realizations [2]. In this way, the estimation error of the transfer functions or state-space models will influence the computation of finite-length Markov parameters. To overcome this problem, it is essential to estimate the finite-length Markov parameters from the input and output data directly.

There are several Markov-parameter estimation methods using directly the input and output data. The Markov-parameter estimation for a Box-Jenkin's model was recasted as a rank constrained infinite-dimensional optimization problem [26], and an approximate solution is then derived by solving a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2957703, IEEE Transactions on Automatic Control

nuclear-norm regularized optimization problem. The Kernel-based regularization methods [27], [28] can yield robust estimation for the Markov parameter sequence, where the estimation accuracy relies the Kernel selection. The above two kinds of methods usually yield biased estimation of the finite-length Markov parameters.

Based on the data equation of the state-space model, a structured least-squares solution for the Markov-parameter estimation was proposed in [13], [15], which can yield accurate estimation of Markov parameters. In the sequel, this structured least-squares solution will be reviewed and sufficient conditions for the consistent estimation of the Markov parameters will be given. Furthermore, by exploiting the low-rank property of the block Hankel matrix constructed by Markov parameters, a rank-constrained least-squares solution is provided for the Markov parameter estimation, which lays a foundation for the development of the identification method in Section IV.

A. Markov-parameter identification by least-squares optimization

For the state-space model in (1), the data equation can be written as

$$Y_k = O_s X_k + T_u U_k + T_e E_k \tag{7}$$

where U_k, Y_k, E_k are respectively the block Hankel matrices constructed from $u(k), y(k), e(k)$:

$$\begin{aligned} U_k &= \mathcal{H}_{s,h}[u(k:k+s+h)] \\ Y_k &= \mathcal{H}_{s,h}[y(k:k+s+h)] \\ E_k &= \mathcal{H}_{s,h}[e(k:k+s+h)], \end{aligned} \tag{8}$$

the extended observability matrix O_s and the convolution matrices $\{T_u, T_e\}$ are respectively defined as

$$\begin{aligned} O_s &= \mathcal{O}_s[A, C] \\ T_u &= \mathcal{T}_s[CA^{s-1}B \ \cdots \ CB \ \ D] \\ T_e &= \mathcal{T}_s[CA^{s-1}K \ \cdots \ CK \ \ I], \end{aligned} \tag{9}$$

the state sequence X_k is defined as

$$X_k = \left[\hat{x}(k) \ \ \hat{x}(k+1) \ \cdots \ \hat{x}(k+h)\right]. \tag{10}$$

Throughout the paper, it is stipulated that the dimension parameter $h \gg s$ such that Y_k, U_k are fat matrices.

In the data equation (7), due to the unknown term $O_s X_k$, the Markov parameters in the convolution matrix T_u are difficult to estimate. By exploiting the low rank property of the unknown matrix $O_s X_k$, the Markov-parameter estimation problem can be formulated as a low-rank minimization problem [29]. Due to the NP-hard property of the low-rank minimization, the nuclear norm was used instead of rank operator which results in an approximate estimation of Markov parameters.

By regarding the state sequence X_k in (7) as an unknown input sequence, the corresponding Markov-parameter estimation turns out to be a blind identification problem. Inspired from the blind subspace identification approach [30], if the basis of the state sequence X_k is available, the Markov-parameter estimation can be handled by the least-squares method. Following this route, the basis (row space) of the state sequence X_k will be provided, and then the Markov parameter estimation.

The basis of the state sequence X_k can be derived based on the N4SID method [18] and the PO-MOESP method [2], i.e., the current state can be linearly represented by the past input-output observations or the future input-output observations. Consider the following data equation constructed from the past input and output data that has a similar form with equation (7):

$$Y_{k-s-1} = O_s X_{k-s-1} + T_u U_{k-s-1} + T_e E_{k-s-1}. \tag{11}$$

When the extended observability matrix O_s has full column rank, the state sequence X_k can be represented as [13]:

$$\begin{aligned} X_k = &\underbrace{A^{s+1}O_s^\dagger}_{\Gamma_s} Y_{k-s-1} + \underbrace{[C_u - A^{s+1}O_s^\dagger T_u]}_{\Upsilon_s} U_{k-s-1} \\ &+ \underbrace{[C_e - A^{s+1}O_s^\dagger T_e]}_{\Xi_s} E_{k-s-1} \end{aligned} \tag{12}$$

where $C_u = [A^s B \cdots AB \ \ B]$ and $C_e = [A^s K \cdots AK \ \ K]$. The above equation indicates that the state sequence X_k can be linearly represented in terms of the past input and output data. As a consequence, the data equation (7) can be rewritten as

$$\begin{aligned} Y_k = &\underbrace{O_s\Gamma_s}_{\Phi_s} Y_{k-s-1} + \underbrace{O_s\Upsilon_s}_{\Psi_s} U_{k-s-1} + T_u U_k \\ &+ \underbrace{O_s\Xi_s}_{\Pi_s} E_{k-s-1} + T_e E_k. \end{aligned} \tag{13}$$

The least-squares estimation for the convolution matrix T_u that includes the Markov parameters as its block entries can then be formulated as [13]:

$$\min_{\Phi_s, \Psi_s, T_u} \quad \|Y_k - \Phi_s Y_{k-s-1} - \Psi_s U_{k-s-1} - T_u U_k\|_F^2 \tag{14}$$

It can be observed from the above optimization problem that the optimal solution of the convolution matrix T_u can be obtained by the oblique projection [18] of the row space of Y_k along the row space of $\left[U_{k-s-1}^T \ \ Y_{k-s-1}^T\right]^T$ on the row space of U_k. The least-squares estimate for the matrix T_u in (14) will be analyzed in the following lemma.

Lemma 1. *Suppose that Assumptions A1-A3 are satisfied and the dimension parameter s is larger than the observability index of (A, C). Let*

$$\begin{aligned} &\bar{E}\left\{\left[\begin{array}{c} y(k-s-1:k-1) \\ \hline u(k-s-1:k-1) \\ \hline u(k:k+s) \end{array}\right]\left[\begin{array}{c} y(k-s-1:k-1) \\ \hline u(k-s-1:k-1) \\ \hline u(k:k+s) \end{array}\right]^T\right\} \\ &= \left[\begin{array}{cc} R_{11} & R_{12} \\ R_{21} & R_{22} \end{array}\right] \end{aligned} \tag{15}$$

and

$$\bar{E}\left\{e(k-s-1:k-1)e^T(k-s-1:k-1)\right\} = R_{ee}. \tag{16}$$

Denote by T_u^h the least-squares estimate of T_u in (14) *that depends on the dimension parameter h defined in* (8)*. As the*

dimension parameter h (or the data length) tends to infinity, the following equality holds with probability one:

$$\lim_{h\to\infty} T_u^h = T_u^* + [\Pi_s^* R_{ee} T_e^{*,T} \;\; 0] \left[R_{11} - R_{12}R_{22}^{-1}R_{21}\right]^{-1} R_{12}R_{22}^{-1} \tag{17}$$

where T_u^*, Π_s^* and T_e^* represent the true values of T_u, Π_s and T_e, respectively.

Proof: First, by Assumptions A1-A3, it can be established that the following matrix has full rank [31], i.e.,

$$\bar{E}\left\{\begin{bmatrix} y(k-s-1:k-1) \\ \hline u(k-s-1:k-1) \\ \hline u(k:k+s) \end{bmatrix}\begin{bmatrix} y(k-s-1:k-1) \\ \hline u(k-s-1:k-1) \\ \hline u(k:k+s) \end{bmatrix}^T\right\} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} > 0.$$

Because of the white noise property of $e(k)$, we have

$$\bar{E}\left\{e(k:k+s)\left[e^T(k-s-1:k-1) \;\; y^T(k-s-1:k-1) \;\; u^T(k-s-1:k-1) \;\; u^T(k:k+s)\right]\right\} = 0. \tag{18}$$

Then, by the first-order optimality condition of the objective function in (14), it can be obtained that

$$\begin{bmatrix}\Phi_s^h & \Psi_s^h & T_u^h\end{bmatrix}\begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}\begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}^T = Y_k \begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}^T, \tag{19}$$

with Φ_s^h, Ψ_s^h and T_s^h being respectively the estimates of Φ_s, Ψ_s and T_s. Let $T_u^*, \Phi_s^*, \Psi_s^*, \Pi_s^*, T_e^*$ be respectively the true values of the matrix variables $T_u, \Phi_s, \Psi_s, \Pi_s, T_e$ satisfying that

$$Y_k = \Phi_s^* Y_{k-s-1} + \Psi_s^* U_{k-s-1} + T_u^* U_k + \Pi_s^* E_{k-s-1} + T_e^* E_k.$$

Substituting the above expression of Y_k into (19) yields that

$$\begin{aligned}&\begin{bmatrix}\Phi_s^h & \Psi_s^h & T_u^h\end{bmatrix}\begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}\begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}^T \\ &= \begin{bmatrix}\Phi_s^* & \Psi_s^* & T_u^*\end{bmatrix}\begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}\begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}^T \\ &\quad + \left[\Pi_s^* E_{k-s-1} + T_e^* E_k\right]\begin{bmatrix} Y_{k-s-1} \\ U_{k-s-1} \\ U_k \end{bmatrix}^T.\end{aligned} \tag{20}$$

By dividing both hand sides of the above equation by $h+1$ and taking the limits as $h \to \infty$, the following equality holds with probability one:

$$\begin{aligned}[\Phi_s^\infty \;\; \Psi_s^\infty \;\; T_u^\infty]\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} &= [\Phi_s^* \;\; \Psi_s^* \;\; T_u^*]\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \\ &\quad + \begin{bmatrix} \Pi_s^* R_{ee} T_e^{*,T} & 0 & 0 \end{bmatrix}.\end{aligned}$$

It then follows that

$$\begin{aligned}[\Phi_s^\infty \;\; \Psi_s^\infty \;\; T_u^\infty] &= [\Phi_s^* \;\; \Psi_s^* \;\; T_u^*] \\ &\quad + \left[\begin{array}{c|c|c} \Pi_s^* R_{ee} T_e^{*,T} & 0 & 0 \end{array}\right]\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}^{-1}.\end{aligned} \tag{21}$$

Due to the positive definite property of the matrix $\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$, by the inverse lemma of a 2×2 block matrix [32], the estimate of T_u can be expressed as

$$T_u^\infty = T_u^* + [\Pi_s^* R_{ee} T_e^{*,T} \;\; 0] \left[R_{11} - R_{12}R_{22}^{-1}R_{21}\right]^{-1} R_{12}R_{22}^{-1}.$$

This completes the proof of the lemma. ■

Remark 2. *The least-squares solution of the convolution matrix T_u is given in* (17)*, from which several insights can be obtained.*

First, it can be observed from equation (17) *that the estimation error is linearly proportional to Π_s^* which can be explicitly written as*

$$\Pi_s^* = O_s \Xi_s^* = O_s C_e - O_s A^{s+1} O_s^\dagger T_e.$$

It can be found that the last block of Π_s^* is proportional to A^s, which tends to zero with probability one as $s \to \infty$; therefore, the last block row of T_u^h tends to last block row of T_u^*. This result is in accord with the consistency of the Markov-parameter estimation in [13].

Second, it can be seen from equation (17) *that the correlation between U_k and $[Y_{k-s-1}^T \;\; U_{k-s-1}^T]^T$ is another factor that prevents the consistent estimation of T_u. If the input signal $u(k)$ is a zero-mean white noise, we can obtain that $R_{21} = \bar{E}\{u(k:k+s)[y^T(k-s-1:k-1) \; u^T(k-s-1:k-1)]\} = 0$; thus, the convolution matrix T_u can be consistently estimated.*

Third, according to equation (21) *in the proof of Lemma 1, we can see that the existence of the innovation noise prevents the consistent estimation of the convolution matrix T_u. If we set the innovation noise $e(k)$ to be zero, it can still be proven that the matrix T_u can be consistently estimated.*

Remark 3. *The least-squares solution given in* (14) *is analogous to the classical prediction-error framework in [1, Chapter 3.2], where the $\Phi_s Y_{k-s-1} + \Psi_s U_{k-s-1}$ can be viewed as the $(s+1)$-step-ahead predictor for the output Y_k, and $T_u U_k$ is the corresponding prediction error (or residual). Therefore, the least-squares solution for T_u in* (14) *can be interpreted as a combination of the prediction-error method and the subspace identification method.*

B. Low-rank regularized estimation of Markov parameters

The least-squares estimation framework (14) enables the estimation of finite-length Markov parameters using directly the input and output data. When only a finite number of data are available, the estimation accuracy will be influenced by the noise perturbation; therefore, in order to improve the estimation accuracy, the structural properties of the convolution matrix T_u was exploited and a structured least-squares solution was provided in [13]–[15]. In fact, the block Hankel matrix constructed from Markov parameters has a rank equal to the system order; therefore, this low-rank property will be adopted in this paper to improve the estimation accuracy of Markov parameters.

For the identification of the structured state-space model (1), the system order is assumed to be known. Then, a rank-constrained least-squares optimization problem is provided for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2957703, IEEE Transactions on Automatic Control

the Markov-parameter estimation:

$$\begin{aligned}\min_{\Phi_s,\Psi_s,T_u,H_u,M_i} \quad & \|Y_k - \Phi_s Y_{k-s-1} - \Psi_s U_{k-s-1} - T_u U_k\|_F^2 \\ s.t. \quad & T_u = \mathcal{T}_s\left[M_0 \ \cdots \ M_{s-1}\right] \\ & H_u = \mathcal{H}_{l_1,l_2}\left[M_0 \ \cdots \ M_{s-1}\right] \\ & \text{rank}(H_u) = n,\end{aligned} \tag{22}$$

where the dimension parameters l_1, l_2 and s are chosen such that $l_1 + l_2 = s - 1$ and $l_1, l_2 \geq n$. Due to the NP-hard property of the above rank-constrained optimization problem, the nuclear norm is usually adopted as a convex relaxation of the rank operator [33]. By solving the associated nuclear-norm optimization problem, the Markov parameters as well as the system order can be approximately estimated. In this paper, a difference-of-convex programming algorithm will be proposed to deal with the above rank-constrained optimization problem.

Apart from the block Toeplitz structure of the matrix T_u, the low rank constraint of the block Hankel matrix H_u enables more accurate estimation of Markov parameters, especially when only a short batch of noisy output observations are available. This will be demonstrated by a simulation example in Section V.

IV. Estimation of structured system matrices

The objective of the concerned identification problem is to identify the system parameters rather than the Markov parameters. To this end, the parameter estimation will be investigated based on the Markov-parameter estimation framework in (22). The block Hankel matrix H_u in (22) can be represented as the product of an observability matrix and a controllability matrix with shifting properties [2]. By adopting the framework for the structured factorization of a block Hankel matrix [10], the structural constraints of the system matrices can be explicitly imposed. In this section, the Hankel-matrix factorization framework provided in [10] is reviewed briefly, the structural constraints of the system matrices will be incorporated to the rank minimization problem in (22), and an iterative optimization method will be developed for the estimation of system parameters.

A. Hankel matrix factorization with structural constraints

The block Hankel matrix H_u in the optimization problem (22) possesses a *structured* and *low-rank* factorization.

The *low-rank* property can be seen from the following expression:

$$H_u = O_{l_1} C_{l_2} \tag{23}$$

where $O_{l_1} = \mathcal{O}_{l_1}[A, C]$ and $C_{l_2} = \mathcal{C}_{l_2}[A, B]$. It is not difficult to see that rank $(H_u) = n$.

The *structural* property refers to the parametrization of the system matrices embedded in the Hankel matrix. The matrices C and B are respectively the first block entries of O_{l_1} and C_{l_2}, i.e.,

$$C = O_{l_1}(1:p,:), \quad B = C_{l_2}(:,1:m). \tag{24}$$

As a consequence, the structural constraints on C and B can be explicitly imposed. The structural constraint on the matrix A is less straightforward. It can however be expressed by exploiting the shifting property of O_{l_1} or C_{l_2}. Denote

$$\bar{O}_{l_1} = \begin{bmatrix} CA \\ \vdots \\ CA^{l_1+1} \end{bmatrix} \text{ and } \bar{C}_{l_2} = \begin{bmatrix} AB & \cdots & A^{l_2+1}B \end{bmatrix}.$$

The shifting property can be represented as

$$\bar{O}_{l_1} = O_{l_1} A, \quad \bar{C}_{l_2} = A C_{l_2} \tag{25}$$

where the structural constraint of A can then be imposed.

To sum up, the structured and low-rank factorization of the Hankel matrix H_u is captured by the equations (23), (24) and (25). Equation (24) appears to be linear constraints, whereas equations (23) and (25) are inherently bilinear constraints. The following lemma shows that the bilinear equations in (23) and (25) can be compactly represented by a single rank constraint.

Lemma 2. *The bilinear equations in* (23) *and* (25) *hold if and only if there exists a matrix variable* $\bar{A}$ *such that*

$$\textit{rank} \begin{bmatrix} H_u & O_{l_1} & \bar{O}_{l_1} \\ C_{l_2} & I & A \\ \bar{C}_{l_2} & A & \bar{A} \end{bmatrix} = n. \tag{26}$$

Proof: Sufficiency: The second block column (row) of the matrix in (26) has column rank n since it contains an identity block entry. Then, the rank constraint (26) implies that the first and the third block columns (rows) should be linearly dependent of the second column (row). As a result, the bilinear equations in (23) and (25) can be derived.

Necessity: By substituting the expressions of $\bar{O}_{l_1}$ and $\bar{C}_{l_2}$ into (26), when $\bar{A} = A^2$, it is easy to verify that all the block columns (rows) can be linearly represented by the second block column (row). Therefore, the rank equation in (26) holds. ∎

For the identification of a parameterized state-space model, the system order is known, i.e., rank$[H_u] = n$. The structural constraints of the system matrices can be incorporated with the proposed Markov-parameter estimation framework (22). By Lemma 2, the structured state-space identification problem can be addressed by solving the following rank constrained optimization problem:

$$\begin{aligned}\min_{\Theta} \quad & \|Y_k - \Phi_s Y_{k-s-1} - \Psi_s U_{k-s-1} - T_u U_k\|_F^2 \\ s.t. \quad & T_u = \mathcal{T}_s\left[M_0 \ \cdots \ M_{s-1}\right] \\ & H_u = \mathcal{H}_{l_1,l_2}\left[M_0 \ \cdots \ M_{s-1}\right] \\ & \text{rank} \begin{bmatrix} H_u & O_{l_1} & \bar{O}_{l_1} \\ C_{l_2} & I & A \\ \bar{C}_{l_2} & A & \bar{A} \end{bmatrix} = n \\ & C = O_{l_1}(1:p,:), \quad B = C_{l_2}(:,1:m) \\ & \{A, B, C\} \in \mathcal{S}\end{aligned} \tag{27}$$

where $\Theta = \{\Phi_s, \Psi_s, T_u, H_u, M_i, O_{l_1}, \bar{O}_{l_1}, C_{l_2}, \bar{C}_{l_2}, A, B, C, \bar{A}\}$ denotes the variable set, and $\mathcal{S}$ represents the set of structured system matrices as illustrated in Remark 1. Apart from the rank constraint and possible polynomial parametrization of the matrix set $\mathcal{S}$, the above optimization problem is convex. Its numerical solution will be investigated in the next section.

Remark 4. *The optimization problem provides an approach to identify structured system matrices directly from the input and output observations, instead of sequentially estimating the Markov parameters and then the structured system matrices as done in [10], [23]; thus, it provides a unified framework for the identification of structured state-space models.*

B. Iterative optimization method

In this section, numerical solutions for the rank-constrained optimization problem (27) will be investigated. To simplify the notation, we denote

$$\begin{aligned} f(\Theta) :=& \|Y_k - \Phi_s Y_{k-s-1} - \Psi_s U_{k-s-1} - T_u U_k\|_F^2 \\ s.t.\quad & T_u = \mathcal{T}_s\left[M_0 \ \cdots \ M_{s-1}\right] \\ & H_u = \mathcal{H}_{l_1,l_2}\left[M_0 \ \cdots \ M_{s-1}\right] \\ & C = O_{l_1}(1:p,:), \quad B = C_{l_2}(:,1:m) \\ & \{A,B,C\} \in \mathcal{S} \end{aligned}$$

and

$$H(\Theta) = \begin{bmatrix} H_u & O_{l_1} & \bar{O}_{l_1} \\ C_{l_2} & I & A \\ \bar{C}_{l_2} & A & \bar{A} \end{bmatrix}.$$

The matrix $H(\Theta)$ is linearly represented in terms of Θ. When the structured system matrix set $\mathcal{S}$ can be linearly parameterized, it is easy to see that $f(\Theta)$ is a convex function with respect to Θ. Otherwise, when the matrix set $\mathcal{S}$ is polynomially parameterized, the structural constraints can be formulated as rank constraints using the technique in Subsection IV-A. This is illustrated in the following example.

Example 1. *Suppose that the matrix A in the LTI system* (1) *has a third-order polynomial parametrization as follows:*

$$A(\theta) = \sum_{i=1}^{l}\sum_{j=1}^{l}\sum_{k=1}^{l} A_{i,j,k}\theta_i\theta_j\theta_k, \tag{28}$$

where the matrix coefficients $A_{i,j,k}$ are known. The above polynomial parametrization can be equivalently represented as

$$\begin{aligned} A(\theta) &= \sum_{i=1}^{l}\sum_{j=1}^{l}\sum_{k=1}^{l} A_{i,j,k}\Omega_{i,j,k} \\ \text{vec}(\Omega) &= \theta\otimes\theta\otimes\theta, \end{aligned} \tag{29}$$

where Ω represents a tensor and $\text{vec}(\Omega)$ *denotes the vectorization of Ω. The involved tensor notations and operators follow those in [32, Chapter 12.4]. The tensor Ω in* (29) *is a symmetric and rank-1 tensor, which can be equivalently represented as*

$$\begin{aligned} &\Omega_{i,j,k} = \Omega_{i,k,j} = \Omega_{j,i,k} = \cdots = \Omega_{k,j,i} \quad i,j,k \in \{1,\cdots,l\} \\ &\text{rank}\begin{bmatrix} \Omega(:,1,1) & \cdots & \Omega(:,1,l) & \cdots & \Omega(:,l,l) \end{bmatrix} = 1 \end{aligned} \tag{30}$$

where the linear constraints in the first equation represent the symmetrical structure of Ω, and the second equation is the rank-one constraint.

From the above example, we can see that a polynomially parameterized matrix can be equivalently represented as the combination of symmetrical constraints and rank-1 constraints.

To simplify the notation, we treat only one rank constraint and the function $f(\Theta)$ is assumed to be convex in the sequel; however, this does not affect the implementation of the proposed method for the identification of polynomially parameterized state-space models, like the identification of the Kronecker structured system matrices in Subsection V-C. The optimization problem in (27) can be written as

$$\begin{aligned} \min_{\Theta}\quad & f(\Theta) \\ s.t.\quad & \text{rank}[H(\Theta)] = n. \end{aligned} \tag{31}$$

To solve the above rank-constrained optimization problem, the difference-of-convex (DC) programming method proposed in our previous work [10] is adopted. Define the Ky Fan n-norm of the matrix $H(\Theta)$ as

$$g_n(H(\Theta)) = \sum_{i=1}^{n}\sigma_i(H(\Theta)). \tag{32}$$

Replacing the rank constraint $\text{rank}[H(\Theta)] = n$ by a difference-of-convex equality constraint $\|H(\Theta)\|_* - g_n(H(\Theta)) = 0$ yields that

$$\begin{aligned} \min_{\Theta}\quad & f(\Theta) \\ s.t.\quad & \|H(\Theta)\|_* - g_n(H(\Theta)) = 0. \end{aligned} \tag{33}$$

Due to the non-convexity of the above optimization problem, an approximate solution is obtained by solving the following penalized optimization problem:

$$\min_{\Theta}\quad f(\Theta) + \rho\left[\|H(\Theta)\|_* - g_n(H(\Theta))\right], \tag{34}$$

where ρ is a positive penalty coefficient. The term $\rho\left[\|H(\Theta)\|_* - g_n(H(\Theta))\right]$ imposes a penalty for violating the constraint $\|H(\Theta)\|_* - g_n(H(\Theta)) = 0$. Due to the inequality $\|H(\Theta)\|_* - g_n(H(\Theta)) \geq 0$, the objective function is always nonnegative. When the penalty parameter ρ tends to infinity, the global minimum of (34) will approach that of (33).

To solve the optimization problem (34), the sequential convex programming approach [34] will be adopted, for which it is crucial to linearize the concave part of the objective function in (34). Given the estimate $\hat{\Theta}^j$ and the SVD decomposition

$$H\left(\hat{\Theta}^j\right) = \begin{bmatrix} U_1^j & U_2^j \end{bmatrix}\begin{bmatrix} \Sigma_1^j & \\ & \Sigma_2^j \end{bmatrix}\begin{bmatrix} V_1^j \\ V_2^j \end{bmatrix} \tag{35}$$

with U_1^j and V_1^j being respectively the left and right singular vectors associated with the largest n singular values, the sequential convex programming approach boils down to iteratively solving the following convex optimization problem

$$\Theta^{j+1} := \arg\min_{\Theta}\quad f(\Theta) + \rho\left(\|H(\Theta)\|_* - \text{tr}\left[U_1^{j,T}H(\Theta)V_1^j\right]\right). \tag{36}$$

Through the above iterative optimization, it can be established that the objective function decreases as the iteration index k increases [34]; thus, it is a descent algorithm.

In order to make the optimal solution to (36) better approximate that to (33), the penalty coefficient ρ in (36) needs to be increased along with the iteration index j. The corresponding convex-concave procedure is detailed in **Algorithm 1**, where the relative error tolerance ϵ is a small value and the maximum

value of the penalty coefficient ρ_{max} is set for the convergence sake. In other words, when the maximum value of the penalty coefficient is reached, it becomes an optimization problem of the form (34) for which the objective function is guaranteed to converge.

Algorithm 1 Convex-concave procedure for (33)

1) Set the initial values: $U_1^0 = 0$ and $V_1^0 = 0$.
2) Repeat
   2-1): Obtain the estimate $\hat{\Theta}^{j+1}$ by solving (36).
   2-2): Compute U_1^{j+1} and V_1^{j+1} by the SVD in (35).
   2-3): $\rho := \min\{\mu\rho, \rho_{max}\}$ where $\mu \geq 1$ and $\rho_{max} > 0$.
3) Until the relative decrease of the objective function is smaller than ϵ.

It is remarked that the estimate $\hat{\Theta}^1$ is generated by solving a nuclear-norm regularized convex optimization problem, which usually yields a good initial parameter estimate for the remaining iterations.

C. Attractive convergence property of Algorithm 1

In the proposed convex-concave procedure, the rank constraint is to be relaxed sequentially, which enables the proposed method to find a good initial parameter estimate as well as getting around local minima. In order to show the capability of the proposed method in avoiding local minima, the matrix completion problem without noise perturbation will be simulated. The performance comparison of the proposed method, the nuclear norm method [35], the log-det method [36] and the manifold optimization method [37] will be demonstrated.

In the simulation, a 10×10 matrix with rank 5 is randomly generated, which is denoted by Ω. The considered matrix completion problem is to select a number of entries as unknown parameters that are to be estimated such that the reconstructed matrix has a low rank (the rank is 5 in this case). The number of parameters ranges from 1 to 50. The unknown parameters are randomly selected from the 10×10 matrix.

Denote by $\mathcal{D}$ the index set of the known entries in the matrix Ω. The nuclear norm method [35] is to solve the following optimization problem

$$\begin{aligned} \min_{\Psi \in \mathbb{R}^{10\times 10}} \quad & \|\Psi\|_* \\ s.t. \quad & \Psi(i,j) = \Omega(i,j), \quad (i,j) \in \mathcal{D}. \end{aligned}$$

The log-det method [36] is to iteratively solve the following optimization problem

$$\begin{aligned} & [X^{k+1}, Z^{k+1}, \Psi^{k+1}] = \\ & \arg\min_{X,Z,\Psi} \quad \mathrm{tr}\left[\left(\begin{bmatrix} X^k & \\ & Z^k \end{bmatrix} + \sigma I\right)^{-1} \begin{bmatrix} X & \\ & Z \end{bmatrix}\right] \\ & s.t. \quad \begin{bmatrix} X & \Psi \\ \Psi^T & Z \end{bmatrix} \geq 0 \\ & \qquad \Psi(i,j) = \Omega(i,j), \quad (i,j) \in \mathcal{D}, \end{aligned}$$

where the parameter σ is set to $\sigma = 10^{-6}$ as suggested in the simulation example of [36]. The initial parameter estimate is obtained by setting $X^0 = 0$ and $Z^0 = 0$. The stopping criterion is set to

$$\frac{\left\|\Psi^{k+1} - \Psi^k\right\|_F}{\left\|\Psi^k\right\|_F} \leq 10^{-10},$$

and the maximum number of iterations is set to 100.

The manifold optimization method [37] is to solve the following rank-constrained optimization problem

$$\begin{aligned} \min_{\Psi} \quad & \|\Psi(\mathcal{D}) - \Omega(\mathcal{D})\|_F^2 \\ s.t. \quad & \mathrm{rank}(\Psi) = 5. \end{aligned}$$

The initial parameter is obtained by setting the unknown parameters to be zero and computing the best rank-5 approximation of Ω in the sense of Frobenius norm. The stopping criterion is set to be the same as the log-det method.

For the proposed method - Algorithm 1, the penalty coefficient ρ is set to be a constant $\rho = 0.1$, and the stopping threshold is set to $\epsilon = 10^{-10}$. Also, the maximum number of iterations is set to 100.

For a fixed number of unknown parameters, the above four different methods individually perform 100 Monte-Carlo trials, and the success rate is defined as the total number of trials whose relative estimation errors are smaller than 10^{-6}, i.e.,

$$\frac{\|\hat{\Psi} - \Omega\|_F}{\|\Omega\|_F} \leq 10^{-6}.$$

Fig. 1 shows the success rates of the four different methods, where it can be seen that the success rate of the proposed method is nearly 100% when the number of parameters is less than 15, and it is significantly larger than other methods when the number of parameters is larger than 15. In particular, when the number of parameters is 24, the success rates of the proposed method, the manifold optimization method, the log-det method and the nuclear norm method are respectively 52%, 15%, 0, 0. The comparison results in Fig. 1 can be explained as follows.

- The performance of the nuclear norm method is not as good as the log-det method and the proposed method, since the nuclear norm method provides an initial parameter estimate for both the log-det method and the proposed method.
- The better performance of the proposed method against the log-det method might be caused by the fact that the exact rank information is utilized in the proposed method. It can be observed from Fig. 1 that, when the number of unknown parameters is 20, the success rates of the proposed method and the log-det method are respectively 77% and 0.
- The manifold optimization method is inherently a gradient-type method on the embedded manifold of fixed-rank matrices, which is easily getting stuck into local minima when the number of parameters becomes larger. This explains the better performance of the proposed method against the manifold optimization method.
- When the number of parameters is larger than 10, the manifold optimization method and the proposed method perform better than the log-det method and the nuclear norm method, since the manifold optimization method and the proposed method have made use of the rank information.

From the above performance analysis, it can be seen that the proposed method and the manifold optimization method is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2957703, IEEE Transactions on Automatic Control

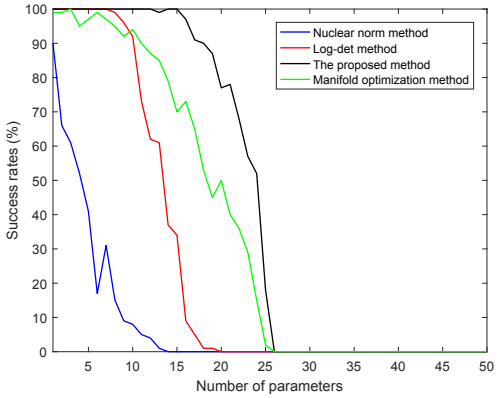

Fig. 1. Success rates for the nuclear norm method, the log-det method, the manifold optimization method and the proposed method.

TABLE I
COMPARISON OF COMPUTATION TIMES.

Method	Times (s)	No Iter.	Time/Iter. (s)
Nuclear norm method	0.3437	1	0.3437
Log-det method	1.2290	3	0.4097
The proposed method	1.3818	3	0.4606
Manifold optimization method	6×10^{-4}	12	5×10^{-5}

more suitable for dealing with rank-constrained optimization problems, while the nuclear norm method and the log-det method are more appropriate for handling rank minimization problems. Since the rank constraint is crucial for the identification problem concerned in this paper, the proposed method and the manifold optimization method (or the gradient-type optimization method on a specific subspace) will be performed and compared for handling the concerned identification problem in next section.

The numerical simulations are run on a laptop with a 2.9 GHz processor and a 8.0 GB RAM. One Monte-Carlo trial is performed by four different methods with 10 unknown parameters, and the corresponding computation times are shown in Table I. It can be observed that:

- the computation time of the proposed method per iteration is slightly higher than the nuclear norm method and the log-det method, because the log-det method is inherently a re-weighted nuclear norm method [36] and the proposed method needs to run an extra SVD decomposition at each iteration;
- the computation time of the manifold optimization method is much less than the other three methods that need to solve a nuclear norm optimization problem in each iteration.

V. NUMERICAL SIMULATIONS

In this section, three simulation examples are provided to validate the proposed COSMOS identification method: the first one is to show the effectiveness of the rank-constrained optimization approach proposed in Section III on estimating Markov parameters using a short batch of input and output data; the second one is to apply the COSMOS on identifying a linearly parameterized state-space model; the third one is to show the performance of the COSMOS on identifying a state-space model with the system matrices having Kronecker-product forms, as described in Remark 1.

A. Estimation of finite-length Markov parameters

In this simulation example, the proposed rank-constrained optimization method (22) is simulated to show the identification performance for the finite-length Markov parameters, where the rank-constrained optimization problem is solved using the convex-concave procedure in Algorithm 1.

The simulated innovation model is described by the following matrices:

$$A = \begin{bmatrix} 1.5610 & -0.6414 \\ 1.0000 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, K = \begin{bmatrix} 0.1477 \\ 0.0388 \end{bmatrix}$$
$$C = \begin{bmatrix} 0.0715 & 0.0072 \end{bmatrix}, D = 0.0201.$$

The input signal $u(k)$ and the innovation signal $e(k)$ are generated as white noise sequences. The data length is set to 300. The Markov-parameter sequence of length 15 will be estimated. To demonstrate the identification performance against different noise levels, the signal-to-noise ratio is defined as

$$\text{SNR} = 10 \log \frac{\text{var}[y(k) - e(k)]}{\text{var}[e(k)]}$$

and the relative estimation error is defined as

$$\text{Relative error} = \frac{\sum_{i=0}^{14} \|\hat{M}_i - M_i^*\|_F}{\sum_{i=0}^{14} \|M_i^*\|_F}, \tag{37}$$

where $\hat{M}_i$ and M_i^* are respectively the estimated and the true Markov parameters.

For the comparison purpose, the proposed rank-constrained least-squares (RCL) method (22), the structured least-squares (SLS) estimation method [15] and the "TC" kernel (TCK) based FIR estimation method [27] will be simulated.

For the rank-constrained optimization in (22), the dimension parameters are set to $s = 6, l = 6$. It is handled using the convex-concave procedure in Algorithm 1, where the involved parameters are set to

$$\rho = 0.01,\ \mu = 1.02,\ \rho_{max} = 10,\ \epsilon = 10^{-10},$$

and the maximum number of iterations is set to 100, i.e., the implementation of Algorithm 1 stops if the stopping criterion is not reached at the 100-th iteration.

The TCK method is implemented using the Matlab commands as follows [38, Subsection 3.2]:

```
aropt=arxRegulOptions;
aropt.RegulKernel='TC';
[L,R]=arxRegul(data,[0 14 0],aropt);
aopt=arxOptions;
aopt.Regularization.Lambda=L;
aopt.Regularizatoin.R=R;
mest=arx(data,[0 14 0],aopt);
```

To evaluate the identification performance for three different identification methods, 30 Monte-Carlo trials are carried out at each SNR. Fig. 2 shows the mean relative errors at different SNRs and Fig. 3 provides box plots for the three identification methods, where we can observe that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2957703, IEEE Transactions on Automatic Control

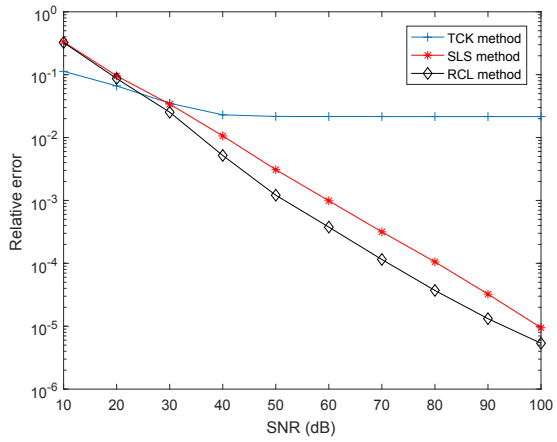

Fig. 2. Example A: mean relative errors of the "TC" Kernel (TCK) method, the structured least-squares (SLS) method and the rank-constrained least-squares (RCL) method at different SNRs.

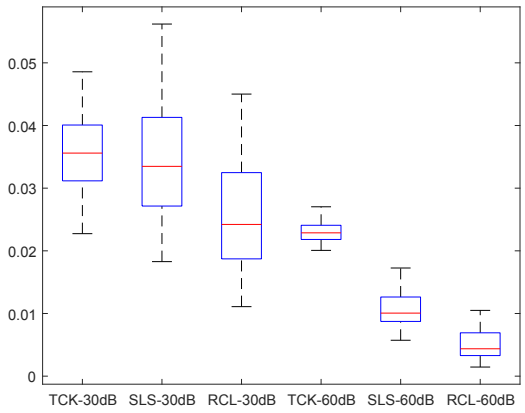

Fig. 3. Example 1: box-plots of the TCK method, the SLS method and the RCL method through 50 Monte-Carlo trials.

1) The TCK method yields an approximate estimate even if the noise level tends to zero, which is caused by the improper tuning of the regularization parameters. The relative error is about 0.02 when the SNR is larger than 40 dB.
2) The relative-error curves of the SLS method and the RCL method decay along with the increase of SNR, indicating that the finite-length Markov parameters can be accurately estimated in the absence of noise.
3) The proposed rank-constrained least-squares method performs better than the structured least-squares method. This is because the low-rank property of the Hankel matrix is exploited in the proposed RCL method.

The above three algorithms are run in the Matlab environment as described in Subsection IV-C. The computation times are shown in Table II, where we can see that the SLS and TCK have similar computation times that are much less than the proposed RCL method. This is because the proposed RCL method needs to solve a nuclear norm regularized optimization problem in each iteration.

TABLE II
COMPUTATION TIMES OF THE TCK, SLS AND RCL METHODS.

Method	Times (s)	No Iter.	Time/Iter. (s)
TCK	0.4672	1	0.4672
SLS	0.5402	1	0.5402
RCL	6.9855	9	0.7762

B. Identification of a linearly parameterized state-space model

This simulation example aims to show the effectiveness of the COSMOS on identifying linearly parameterized matrices in the presence of measurement noise. The state-space model to be identified in this example is a three-compartment model which is determined by the following structured system matrices

$$A(\theta) = \begin{bmatrix} -\theta_1 & \theta_2 & 0 \\ \theta_1 & -(\theta_2+\theta_3) & \theta_4 \\ 0 & \theta_3 & -\theta_4 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$
$$K = \begin{bmatrix} 0.2 & 0.1 & 0.5 \end{bmatrix}^T,\ C = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix},\ D = 0.$$

Different from the previous example which is to identify the finite-length Markov parameters, the structured state-space model identification problem aims to identify the parameter vector θ embedded in the structured system matrices. It has been shown in [39] that the parameter vector θ in the above compartment model is identifiable.

In this simulation example, the parameter vector θ is set to $\theta = [0.10 \quad 0.32 \quad 0.21 \quad 0.45]$ and the identification performance of the COSMOS against the measurement-noise level is demonstrated. The input signal $u(k)$ and the innovation noise $e(k)$ are generated as white noises. The data length is set to 50. The implementation settings of the COSMOS are the same as those in Example A. For the comparison purpose, the classical prediction-error method (PEM) [1] is simulated as well. Since the PEM is a gradient descent method, its performance is sensitive to the initial estimate of the parameter vector. Without any prior knowledge of the system parameters, the initial estimate of θ is randomly generated. The stopping criterion for the COSMOS and the PEM is that the relative parameter-estimation error is smaller than 10^{-10}, and the maximum number of iterations for the COSMOS and the PEM is set to 100.

To measure the identification performance, the following two criteria are defined.

Normalized estimation error (NEE). Denote by $\hat{\theta}^r$ and θ^* the estimate of θ at the $r-$th Monte-Carlo trial and the true parameter vector, respectively. The NEE criterion is defined as

$$\text{NEE}^r = \frac{\|\hat{\theta}^r - \theta^*\|}{\|\theta^*\|}.$$

This criterion is directly used to show the identification accuracy of the system parameters.

Output relative error (ORE). According to equation (27), we denote by $\Phi_s^r, \Psi_s^r, T_s^r$ respectively the estimates of Φ_s, Ψ_s, T_s at the r-th Monte-Carlo trial. The OLE criterion is defined as

$$\text{ORE}^r = \frac{\|Y_k - \Phi_s^r Y_{k-s-1} - \Psi_s^r U_{k-s-1} - T_s^r U_k\|_F}{\|Y_k\|_F}.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2957703, IEEE Transactions on Automatic Control

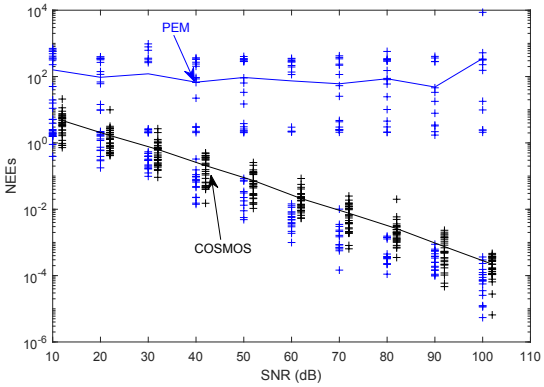

Fig. 4. Example B: performance of identifying the linearly parameterized state-space model in terms of the NEE criterion. The blue crosses represent the NEE values obtained by the PEM, while the black crosses are provided by the COSMOS. The blue curve denotes the mean NEE values of the COSMOS at different SNRs, while the black curve corresponds to mean NEE values of the PEM. Note that the NEE values of these two methods are computed at the same SNRs; however, they are slightly separated for better comparison.

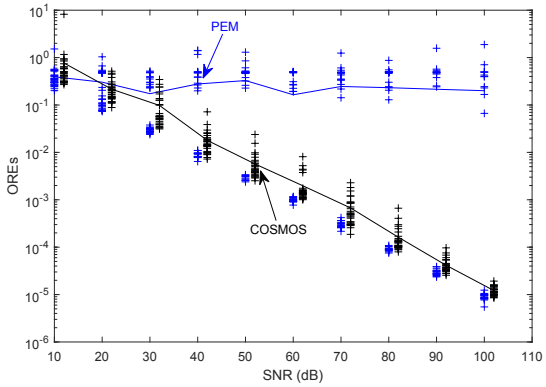

Fig. 5. Example B: performance of identifying the linearly parameterized state-space model in terms of the ORE criterion. The blue crosses represent the ORE values obtained by the PEM, while the black crosses are provided by the COSMOS. The blue curve denotes the mean ORE values of the COSMOS at different SNRs, while the black curve corresponds to mean ORE values of the PEM. Note that the ORE values of these two methods are computed at the same SNRs; however, they are slightly separated for better comparison.

Different from the NEE criterion, the ORE is used to show the estimation accuracy of system dynamics rather than system parameters, and it works well even if the system parameters are unidentifiable.

Fig. 4 and Fig. 5 show the performance of the COSMOS and the PEM, respectively, in terms of the NEE and the ORE criteria. At each SNR, the NEE and ORE values are computed by running 30 Monte-Carlo trials. It can be observed that the NEE and ORE values of the COSMOS are closely distributed around their corresponding mean values. The mean NEE and ORE values of the COSMOS decay along with the increase of the SNR, indicating that the COSMOS can obtain accurate parameter estimation in the absence of measurement noise. However, for the PEM, there are two separate clusters of NEE and ORE values at each SNR, indicating that the PEM with random initialization sometimes gets stuck into local minima.

In the simulation, the PEM takes about 5.26s for 100 iterations, while the COSMOS takes about 94.53s for 100 iterations. This is caused by the fact that the COSMOS needs to solve a nuclear norm regularized optimization problem in each iteration.

C. Identification of a state-space model with Kronecker structured system matrices

Different from the simulation example in Subsection V-B that concerns the identification of linearly parameterized state-space models, this simulation example aims to deal with the identification of a state-space model with its system matrices having Kronecker-product forms [23]. The PEM is simulated for comparison where the Kronecker structured state-space model is defined using the command `idgrey` in the Matlab environment.

The structured state-space model to be identified is determined by the following system matrices:

$$\begin{aligned} A(\theta) &= \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix} \otimes \begin{bmatrix} \theta_5 & \theta_6 \\ \theta_7 & \theta_8 \end{bmatrix}, \\ B &= \begin{bmatrix} -1.35 & -0.32 \end{bmatrix}^T \otimes \begin{bmatrix} -0.12 & -0.81 \end{bmatrix}^T \\ C &= \begin{bmatrix} 1.55 & 0.62 \end{bmatrix} \otimes \begin{bmatrix} 0.28 & -1.13 \end{bmatrix}. \end{aligned} \tag{38}$$

The system matrices K and D are set to be zero. In the simulation, the vector θ is set to

$$\begin{aligned} \theta = \big[& -0.49 \quad -0.13 \quad 0.72 \quad -0.33 \quad -0.47 \\ & 0.40 \quad -0.09 \quad 0.72 \big]^T. \end{aligned}$$

The input signal $u(k)$ and the innovation noise $e(k)$ are generated as white noises. The data length is set to 50. For the implementation of the COSMOS, there are two penalty coefficients which are set to $\rho_1 = 0.1$ and $\rho_2 = 0.1$: one corresponds to the rank constraint of the matrix $H(\Theta)$ and the other corresponds to the rank-one constraint of the reshuffled matrix of the Kronecker product form. The other simulation settings are set to be the same as those in Example A. The stopping criterion for the COSMOS and the PEM is that the relative parameter-estimation error is smaller than 10^{-10}, and the maximum number of iterations for the COSMOS and the PEM is set to 100.

Due to the fact that the Kronecker-structured state-space model cannot be uniquely determined from the input and output data, only the output relative error (ORE) criterion is adopted to evaluate the identification performance here.

Fig. 6 provides a scatter plot for the identification performance of the PEM and the COSMOS in terms of the ORE criterion. At each SNR value, 30 Monte-Carlo trials are performed and the ORE values are computed. It can be observed that the COSMOS performs much better than the PEM in terms of ORE values when the SNR is larger than 20 dB. The averaging ORE values of the COSMOS decay along with the increase of the SNR, indicating that the COSMOS can yield accurate output prediction in the absence of measurement

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2957703, IEEE Transactions on Automatic Control

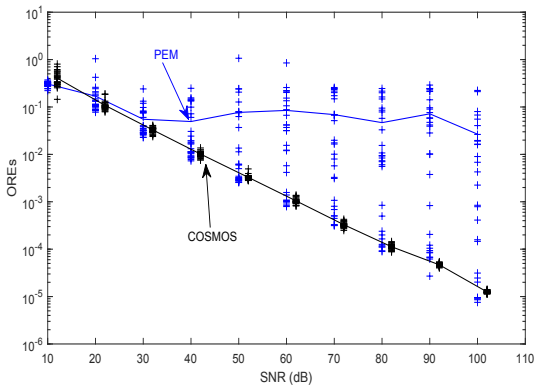

Fig. 6. Example C: performance of identifying the Kronecker-structured state-space model in terms of the ORE criterion. The blue crosses represent the ORE values obtained by the PEM, while the black crosses are provided by the COSMOS. The blue curve denotes the mean ORE values of the COSMOS at different SNRs, while the black curve corresponds to mean ORE values of the PEM. Note that the ORE values of these two methods are computed at the same SNRs; however, they are slightly separated for better comparison.

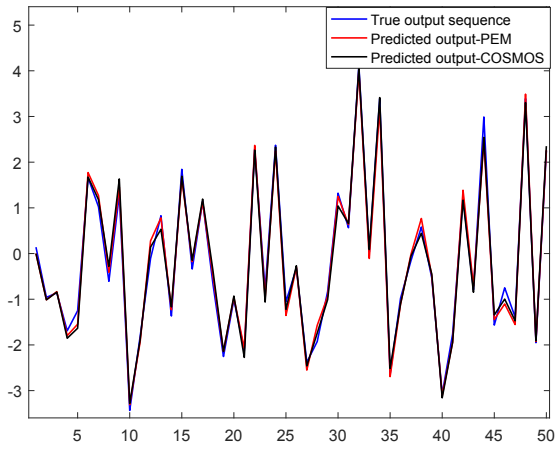

Fig. 7. Example C: predicted output data of a Monte-Carlo trial at SNR=20 dB.

noise. In addition, the ORE values of the COSMOS are more concentrated around their corresponding mean values than the PEM, implying that the COSMOS can yield more reliable identification results than the PEM with random initializations.

For a Monte-Carlo trial at SNR=20dB, the predicted output data against the true output observation is plotted in Fig. 7. In the simulation, the PEM takes about 5.72s for 100 iterations, while the COSMOS takes about 103.70s for 100 iterations. This is caused by the fact that the COSMOS needs to solve a nuclear norm regularized optimization problem in each iteration.

VI. Conclusions

In this paper, the identification of structured state-space models using directly the input and output data has been considered, and a new framework called COSMOS has been developed by combining the spirits of the PEM and the subspace method. The proposed COSMOS can simultaneously estimate the finite-length Markov parameters and the embedded system parameters. This qualifies the COSMOS as a complement to the PEM and the subspace method. In addition, the COSMOS can provide accurate estimation of linearly or polynomially parameterized system matrices, turning out to be more robust than the PEM initialized at random initial values in converging to local minima.

Since the proposed COSMOS framework relies only on finite-length of Markov parameters, it can be used for dealing with large-scale gray-box system models, such the identification work of QUARKS [40]. In addition, the proposed COSMOS framework allows the system matrices to be polynomially parameterized, it can be extended for dealing with high-dimensional system identification problems [23]. However, the COSMOS needs to solve a nuclear-norm regularized optimization problem in each iteration, resulting in heavier computational burden than the PEM. Therefore, investigation will be made on improving the computational efficiency.

References

[1] L. Ljung, *System Identification: Theory for the User*. Pearson Education, 1999.
[2] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.
[3] J. Sun, Q. Yang, X. Liu, and J. Chen, "Event-triggered consensus for linear continuous-time multi-agent systems based on a predictor," *Information Sciences*, vol. 459, pp. 278 – 289, 2018.
[4] C. Yu, J. Chen, and M. Verhaegen, "Subspace identification of individual systems in a large-scale heterogeneous network," *Automatica*, vol. 109, p. 108517, 2019.
[5] Y. Liu, D. Feng, and S. Yang, "An efficient hierarchical identification method for general dual-rate sampled-data systems ," *Automatica*, vol. 50, no. 3, pp. 962–970, 2014.
[6] D. Jie, D. Feng, X. P. Liu, and G. Liu, "Hierarchical least squares identification for linear siso systems with dual-rate sampled-data," *IEEE Transactions on Automatic Control*, vol. 56, no. 11, pp. 2677–2683, 2011.
[7] P. Parrilo and L. Ljung, "Initialization of physical parameter estimates," in *Proc. 13th IFAC Symposium on System Identification*, (Rotterdam, The Netherlands), pp. 1524 – 1529, Aug. 2003.
[8] G. Mercere, O. Prot, and J. Ramos, "Identification of parameterized gray-box state-space systems: From a black-box linear time-invariant representation to a structured one," *Automatic Control, IEEE Transactions on*, vol. 59, pp. 2873–2885, Nov 2014.
[9] D. Vizer, G. Mercere, O. Prot, and E. Laroche, "h_∞-norm-based optimization for the identification of gray-box lti state-space model parameters," *Systems & Control Letters*, vol. 92, pp. 34–41, 2016.
[10] C. Yu, L. Ljung, and M. Verhaegen, "Identification of structured state-space models," *Automatica*, vol. 90, pp. 54–61, 2018.
[11] C. Yu, M. Verhaegen, S. Kovalsky, and R. Basri, "Identification of structured lti mimo state-space models," in *IEEE Conference on Decision and Control*, 2016.
[12] D. Regruto, F. Dabbene, and D. E. Rivera, "Guest editorial: Special issue on relaxation methods in identification and estimation problems," *Automatic Control IEEE Transactions on*, vol. 59, no. 11, pp. 2869–2870, 2014.
[13] K. Peternell, W. Scherrer, and M. Deistler, "Statistical analysis of novel subspace identification methods," *Signal Processing*, vol. 52, no. 2, pp. 161–177, 1996.
[14] A. Chiuso, "The role of vector autoregressive modeling in predictor-based subspace identification.," *Automatica*, vol. 43, no. 6, pp. 1034–1048, 2007.
[15] G. Mercere, I. Markovsky, and J. A. Ramos, "Innovation-based subspace identification in open- and closed-loop," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 2951–2956, 2016.
[16] C. Yu, L. Ljung, and M. Verhaegen, "Gray box identification using difference of convex programming," pp. 9462 – 9467, 2017. 20th IFAC World Congress.

[17] A. Wills, C. Yu, L. Ljung, and M. Verhaegen, "Affinely parameterized state-space models: Ways to maximize the likelihood functions," in *18th IFAC Symposium on System Identification*, (Stockholm, Sweden), 2018.

[18] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: TheoryImplementationApplications*. Springer Science & Business Media, 1996.

[19] A. Haber and M. Verhaegen, "Subspace identification of large-scale interconnected systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 10, pp. 2754–2759, 2014.

[20] C. Yu, M. Verhaegen, and A. Hansson, "Subspace identification of local systems in one-dimensional homogeneous networks," *IEEE Transactions on Automatic Control*, vol. 63, no. 4, pp. 1126–1131, 2018.

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

[22] C. Yu and M. Verhaegen, "Subspace identification of distributed clusters of homogeneous systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 463–468, 2017.

[23] B. Sinquin and M. Verhaegen, "K4sid: Large-scale subspace identification with kronecker modeling," *IEEE Transactions on Automatic Control*, 2018.

[24] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. No. BOOK, Prentice Hall, 2000.

[25] N. F. Al-Muthairi and S. Bingulac, "An algorithm for calculating markov parameters from transfer function matrices," *Computers & Electrical Engineering*, vol. 21, no. 2, pp. 77–87, 1995.

[26] H. Hjalmarsson, J. S. Welsh, and C. R. Rojas, "Identification of box-jenkins models using structured arx models and nuclear norm relaxation," in *16th IFAC symposium on system identification*, no. 16, pp. 322–327, Brussels Belgium, 2012.

[27] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[28] A. Chiuso, "Regularization and bayesian learning in dynamical systems: Past, present and future," *Annual Reviews in Control*, vol. 41, pp. 24–38, 2016.

[29] M. Verhaegen and A. Hansson, "N2sid: Nuclear norm subspace identification of innovation models," *Automatica*, vol. 72, pp. 57–63, 2016.

[30] D. Scobee, L. Ratliff, R. Dong, H. Ohlsson, M. Verhaegen, and S. S. Sastry, "Nuclear norm minimization for blind subspace identification (n2bsid)," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, pp. 2127–2132, IEEE, 2015.

[31] M. Jansson and B. Wahlberg, "On consistency of subspace methods for system identification," *Automatica*, vol. 34, no. 12, pp. 1507–1519, 1998.

[32] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3. JHU Press, 2012.

[33] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *Siam Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2012.

[34] T. Lipp and S. Boyd, "Variations and extension of the convexcconcave procedure," *Optimizationa & Engineering*, vol. 17, no. 2, pp. 263–287, 2016.

[35] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *Siam Review*, vol. 52, no. 3, pp. 471–501, 2010.

[36] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *American Control Conference*, 2003.

[37] N. Boumal, B. Mishra, P. A. Absil, and R. Sepulchre, "Manopt, a matlab toolbox for optimization on manifolds," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1455–1459, 2013.

[38] L. Ljung, R. Singh, and T. Chen, "Regularization features in the system identification toolbox," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 745–750, 2015.

[39] R. Bellman and K. J. Astrom, "On structural identifiability," *Mathematical Biosciences*, vol. 7, no. 3, pp. 329–339, 1970.

[40] B. Sinquin and M. Verhaegen, "Quarks: Identification of large-scale kronecker vector-autoregressive models," *IEEE Transactions on Automatic Control*, 2018.

PLACE PHOTO HERE

Chengpu Yu Chengpu Yu received the B. E. and M. E. degrees in electrical engineering from the University of Electronic Science and Technology of China, in 2006 and 2009, respectively, and the Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 2014.

He was with the Internet of Things lab at Nanyang Technological University as a research associate from 2013 to 2014, and with Delft Center for Systems and Control as a PostDoc from 2014 to 2017. Since 2018, he has been with Beijing Institute of Technology as a full professor. His research interests include system identification, distributed optimization and optical imaging.

PLACE PHOTO HERE

Lennart Ljung Lennart Ljung received his PhD in Automatic Control from Lund Institute of Technology in 1974. Since 1976 he is Professor of the chair of Automatic Control In Linkoping, Sweden. He has held visiting positions at IPU (Moscow), Stanford, MIT, Berkeley and Newcastle University (NSW) and has written several books on System Identification and Estimation. He is an IEEE Fellow, an IFAC Fellow and an IFAC Advisor. He is as a member of the Royal Swedish Academy of Sciences (KVA), a member of the Royal Swedish Academy of Engineering Sciences (IVA), an Honorary Member of the Hungarian Academy of Engineering, an Honorary Professor of the Chinese Academy of Mathematics and Systems Science, and a Foreign Member of the US National Academy of Engineering (NAE) as well as a member of the Academia European.

He has received honorary doctorates from the Baltic State Technical University in St Petersburg, from Uppsala University, Sweden, from the Technical University of Troyes, France, from the Catholic University of Leuven, Belgium and from Helsinki University of Technology, Finland. In 2003 he received the Hendrik W. Bode Lecture Prize from the IEEE Control Systems Society, and in 2007 the IEEE Control Systems Award. He received the Quazza Medal in 2002 and the Nichols Medal in 2017, both from IFAC.

PLACE PHOTO HERE

Adrian Wills Adrian Wills was born in Orange, N.S.W. Australia and received his B.E. (Elec.) and Ph.D. degrees from The University of Newcastle, Australia (Callaghan Campus) in May 1999 and May 2003, respectively. He is now with the University of Newcastle, Australia as an associate professor, and the focus of his research has been in the area of system identification.

PLACE PHOTO HERE

Michel Verhaegen Michel Verhaegen received an engineering degree in aeronautics from the Delft University of Technology, The Netherlands, in 1982, and the doctoral degree in applied sciences from the Catholic University Leuven, Belgium, in 1985. From 1985 to 1994 he has been a research fellow of the U.S. National Research Council (NRC) and the Dutch Academy of Arts and Sciences. In the period 1994-1999 he was an Associate Professor of the control laboratory of the Delft University of Technology and became a full professor at the faculty of Applied Physics of the University of Twente in the Netherlands in 1999. From 2001 on he moved back to the University of Delft and joined the Delft Center for Systems and Control. His main research directions include system identification, distributed and fault tolerant control and data driven controller design methodologies. Application areas include control for high resolution imaging and wind energy. In 2014 he was a recipient of an advanced ERC grant.