Identification of Structured LTI MIMO State-Space Models

Chengpu Yu, Michel Verhaegen, Shahar Kovalsky and Ronen Basri

Abstract— **The identification of structured state-space model has been intensively studied for a long time but still has not been adequately addressed. The main challenge is that the involved estimation problem is a non-convex (or bilinear) optimization problem. This paper is devoted to developing an identification method which aims to find the global optimal solution under mild computational burden. Key to the developed identification algorithm is to transform a bilinear estimation to a rank constrained optimization problem and further a difference of convex programming (DCP) problem. The initial condition for the DCP problem is obtained by solving its convex part of the optimization problem which happens to be a nuclear norm regularized optimization problem. Since the nuclear norm regularized optimization is the closest convex form of the low-rank constrained estimation problem, the obtained initial condition is always of high quality which provides the DCP problem a good starting point. The DCP problem is then solved by the sequential convex programming method. Finally, numerical examples are included to show the effectiveness of the developed identification algorithm.**

I. INTRODUCTION

This paper studies the identification of structured state-space systems. In the literature, there are two kinds of parameterized state-space models [1]. One is the black-box model for which the associated system matrices are fully parameterized. The other is the gray-box model (also called structured model in this paper), where only parts of system matrices are parameterized as non-zero parameters. For the identification of black-box LTI system models, the subspace identification methods are commonly adopted as they generally yield reliable identification results [2]. Other identification methods for black-box systems, such as the regularized Gauss-Newton method and the gradient projection method, are also widely applied in practice [2, Chapter 7]. Although the gray-box model generally contains less parameters than the black-box model, it is usually difficult to obtain reliable identification results. The main reason is that the associated identification problem for the gray-box model is always non-convex and it may have many local optimal points. In this regard, it is difficult to find a global optimal solution under mild computational burden. Generally, the gray-box model is solved by either the output-error method or the prediction-error method for which the

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)

S. Kovalsky and R. Basri are with Weizmann Institute of Science, Department of Computer Science and Applied Mathematics (shahar.kovalsky@weizmann.ac.il, ronen.basri@weizmann.ac.il)

involved nonlinear optimization problem is always tackled by gradient-based algorithms [2], [3].

For the structured system model, apart from the direct identification method which is to estimate parameters directly from the system inputs and outputs, the indirect identification framework has also been intensively investigated recently, e.g., see [4], [5], [6]. This indirect identification framework contains two steps. First, the system matrices of the concerned state-space model are identified up to a similarity transformation using classical subspace identification methods, such as MOESP and N4SID [2]. Second, the system parameters are determined according to the system parametrization pattern and the obtained system matrices in the first step.

The second step of the indirect identification for a structured state-space system turns out to be a nonlinear estimation problem which has a smaller scale than that in the direct identification method. More specifically, the involved nonlinear estimation problem is bilinear, i.e. the system parameters and the similarity transformation matrix are coupled together. To solve such a kind of bilinear estimation problem, the alternating minimization algorithm is adopted in [7], which tries to minimize the objective function with respect to either the model parameters or the similarity transformation matrix. The main shortage of the alternating minimization algorithm is that it is likely to get stuck in local optimal points under randomly generated initial conditions. To cope with the initialization problem, the bilinear estimation problem is reformulated as a sum-of-squares which is then solved by semi-definite programming [8]. Due to the complexity of the sum-of-squares decomposition, this method can only be applied to a small-scale structured system. Since many variables are involved in this identification method, the estimate of the similarity transformation tends to be ill-conditioned.

Recently, in order to robustly solve the bilinear estimation problem for the identification of a structured system, a null-space-based technique is provided in [5]. It first stacks all the variables together, no matter independent or coupled, and then compute a subspace which the augmented variable vector lies in. By representing all the variables using the linear combination of the subspace basis, the dimension of the associated non-convex optimization problem can be reduced, hence decreasing computational burden and improving identification accuracy. In practical simulations, the estimate of the similarity transformation matrix may be ill conditioned or even singular when the adopted algorithm get stuck to local (not global) optimal points. To overcome this problem, the condition number of the similarity transforma-

tion matrix is constrained when dealing with the bilinear estimation problem [6], and a combination of BFGS method and spectral bundle algorithm is implemented to solve such a highly nonlinear optimization problem.

In this paper, we present a new identification method, which is a combination of the over-parametrization technique and low-rank constrained optimization. The main idea of this method is that the considered nonlinear estimation problem can be equivalently transformed to a linear estimation problem with a low rank constraint. The relaxation from a nonlinear estimation problem to a linear estimation problem introduces a number of redundant variables, while the low rank constraint reduces this redundancy. The low-rank constrained optimization problem is generally NP hard. However, for the investigated identification problem, the associated rank constrained optimization problem can be casted to a difference of convex programming (DCP) problem, which can be solved using many existing methods [9], [10]. The initial point for solving the DCP problem can be obtained by solving the convex part of the DCP problem. Interestingly, the convex part of the DCP problem is exactly a nuclear norm regularized optimization problem, which can be commonly employed as a heuristic way to solve low-rank constrained optimization problems. By solving the nuclear-norm regularized optimization problem, a good candidate of the initial point can usually be found. Then, by iteratively solving the DCP problem, the identification performance can be gradually improved, which usually yields a global optimal solution in practice. Compared with the methods in [5], [8], the reduction of the variable redundancy using the low-rank constraint is the key contribution of the present paper.

The rest of the paper is organized as follows. Section II provides the problem formulation and a discussion on system identifiability. Section III presents a new identification method for gray-box models. Section IV gives simulation examples to validate the proposed identification method. Finally, concluding remarks are given in Section V.

II. PROBLEM STATEMENT AND SYSTEM IDENTIFIABILITY

In this paper, we consider the identification of parameterized LTI state-space system model. Let $\theta \in \mathbb{R}^q$ be a parameter vector. The concerned state-space is described as

$$
\begin{aligned}
x(k+1) &= A(\theta)x(k) + B(\theta)u(k) \\
y(k) &= C(\theta)x(k) + w(k),
\end{aligned} \tag{1}
$$

where $u(k) \in \mathbb{R}^m, x(k) \in \mathbb{R}^n, y(k) \in \mathbb{R}^p$ and $w(k) \in \mathbb{R}^p$ are system input, state, output and measurement noise, respectively. The system matrices $A(\theta), B(\theta)$ and $C(\theta)$ are assumed to be affine functions with respect to $\theta = [\theta_1, \cdots, \theta_q]^T$:

$$
\begin{aligned}
A(\theta) &= A_0 + A_1\theta_1 + \cdots + A_q\theta_q, \\
B(\theta) &= B_0 + B_1\theta_1 + \cdots + B_q\theta_q, \\
C(\theta) &= C_0 + C_1\theta_1 + \cdots + C_q\theta_q.
\end{aligned} \tag{2}
$$

In practice, the parameter vector θ in (1) often has physical interpretation, such as the Newton's or Kirchoff's laws

[2, Example 7.1] and structure of a compartmental model [11]. Thus, the estimation of these parameters is practically meaningful.

For the system model in (1), the system input is assumed to be persistently exciting and the system model is minimal (observable and controllable). When the measurement noise $w(k)$ is uncorrelated with the system input $u(k)$, the associated system matrices can be consistently estimated up to a similarity transformation using the subspace identification method [1], [2]. Suppose that we have obtained a correct estimate of system matrices, up to a non-singular similarity transformation, denoted as $\hat{A}, \hat{B}$ and $\hat{C}$. Then, the following equation holds

$$
\begin{aligned}
\hat{A} &= T^{-1}A(\theta)T \\
\hat{B} &= T^{-1}B(\theta) \\
\hat{C} &= C(\theta)T,
\end{aligned} \tag{3}
$$

where $T \in \mathbb{R}^{n \times n}$ is a non-singular ambiguity matrix. In the above equation, both θ and T are unknown. Before proceeding to solve the above nonlinear equation, we shall discuss the system identifiability of (1) or the solution uniqueness of (3).

In order to check the identifiability of the state-space model in (1), a direct way is to provide the corresponding system transfer function in terms of θ and check whether the involved parameters can be uniquely determined from the system impulse response [11], [12]. There are generically two factors affecting the system identifiability: the number of involved parameters and the parametrization pattern of the model. For the state-space model in (1), if there are more than $n(p+m)$ parameters, the system is unidentifiable [1], [2]. Thus, a necessary condition is that the number of parameters should be less than $n(p + m)$. As an extreme case, if all the system matrices are fully parameterized, then they are unidentifiable. However, less than $n(p+m)$ does not necessarily result in the identifiability of the system model (1). In fact, the identifiability also depends on the structure of the concerned system model. It is shown in [11] that the state-space model described by the following parameterized system matrices is unidentifiable even though the number of parameters is strictly less than $n(m + p)$:

$$
A(\theta) = \begin{bmatrix} -\theta_1 & \theta_3 & 0 \\ \theta_1 & -(\theta_2 + \theta_3) & \theta_4 \\ 0 & \theta_3 & -\theta_4 \end{bmatrix}, B(\theta) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},
$$
$$
C(\theta) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}.
$$

It has been shown in [11] that the following system structures are identifiable: the diagonal structure, companion structure, cascaded compartmental structure with only system input and output taking place at the last compartment. It is also straightforward that any system model derived from the above identifiable structure under some similarity transformation is also identifiable.

Since this paper mainly focuses on developing a robust identification method for structured state-space system models, the model structure in (1) is assumed to be identifiable. The

above assumption indicates that the parameter vector θ can be uniquely determined from (3), as well as the ambiguity matrix T. In addition, by the assumption that the concerned system is minimal, we can see that T is a nonsingular matrix. Although θ and T are uniquely determined by (3), it is usually difficult to find the exact solution of the nonlinear estimation problem.

Equation (3) can be reformulated as

$$
\begin{aligned}
T\hat{A} &= A(\theta)T \\
T\hat{B} &= B(\theta) \\
\hat{C} &= C(\theta)T.
\end{aligned}
\tag{4}
$$

A basic approach to estimate T and θ is to solve the following optimization problem [7]:

$$
\begin{aligned}
&\arg\min_{\theta,T} \quad h(\theta,T) \\
h(\theta,T) &= \|T\hat{A} - A(\theta)T\|_F^2 + \|T\hat{B} - B(\theta)\|_F^2 \\
&\quad + \|\hat{C} - C(\theta)T\|_F^2.
\end{aligned}
\tag{5}
$$

Since the above estimation problem is non-convex, it is difficult to obtain a global optimal solution under mild computational burden. In order to obtain a solution close to the global optimal point, selection of the initial condition is crucial. Although a semi-definite programming method is provided in [8] to find a proper initial point, the involved the sums-of-squares decomposition of (5) is nontrivial and computationally expensive. Also, an increased number of decision variables are involved, which makes the estimation problem more under-determined. In this paper, we shall develop an effective and robust identification algorithm for structured state-space models. The associated algorithm relies on the combination of the over-parametrization technique and the low-rank constraint. The over-parametrization technique suggests to introduce more variables to transform a non-linear estimation problem into a linear estimation problem, as done in [8]; however, the low-rank constraint is employed to exploit the relations among all involved variables and to reduce their associated redundancy.

III. GRAY-BOX SYSTEM IDENTIFICATION

In this subsection, we aim to develop an algorithm to estimate θ and T from equation (4). Denote by $\otimes$ and $\text{vec}(\cdot)$ respectively the Kronecker product and the vectorization operators. The vectorized form of equation (4) is shown as follows:

$$
\underbrace{\begin{bmatrix} \hat{A}^T \otimes I - I \otimes A(\theta) \\ \hat{B}^T \otimes I \\ I \otimes C(\theta) \end{bmatrix}}_{M(\theta)} \text{vec}(T) = \underbrace{\begin{bmatrix} 0 \\ \text{vec}(B(\theta)) \\ \text{vec}(\hat{C}) \end{bmatrix}}_{N(\theta)},
\tag{6}
$$

where $M(\theta) \in \mathbb{R}^{(n^2+mn+pn)\times n^2}$ and $N(\theta) \in \mathbb{R}^{(n^2+mn+pn)}$ are introduced for notational simplicity. Since $A(\theta), B(\theta)$ and $C(\theta)$ are affine functions in terms of θ, $M(\theta)$ and $N(\theta)$ in the above equation are affine as well. Denote the affine

expressions of $M(\theta)$ and $N(\theta)$ as follows:

$$
\begin{aligned}
M(\theta) &= M_0 + M_1\theta_1 + \cdots + M_q\theta_q, \\
N(\theta) &= N_0 + N_1\theta_1 + \cdots + N_q\theta_q,
\end{aligned}
\tag{7}
$$

where $\{M_i\}_{i=0}^q$ and $\{N_i\}_{i=0}^q$ are constant coefficient matrices of $M(\theta)$ and $N(\theta)$, respectively. Let $\tau = \text{vec}(T)$. Then equation (6) can be equivalently expressed as

$$
M_0\tau + M_1\tau\theta_1 + \cdots + M_q\tau\theta_q = N_0 + N_1\theta_1 + \cdots + N_q\theta_q.
\tag{8}
$$

From the above equation, we can find that the associated estimation problem is bilinear. More specifically, the variables of θ and τ are coupled together. To cope with this, inspired by the over-parametrization technique, we introduce the following redundant variables:

$$
\begin{aligned}
\vartheta_1 &= \tau\theta_1, \\
&\vdots \\
\vartheta_q &= \tau\theta_q,
\end{aligned}
\tag{9}
$$

where $\vartheta_i \in \mathbb{R}^{n^2}$ for $i = 1, \cdots, q$. Then equation (8) is formed as

$$
M_0\tau + M_1\vartheta_1 + \cdots + M_q\vartheta_q = N_0 + N_1\theta_1 + \cdots + N_q\theta_q,
\tag{10}
$$

where $\{\vartheta_i\}_{i=1}^q$ and $\{\theta_i\}_{i=1}^q$ are unknown variables. The above equation is linear with respect to the unknown variables, but there are more variables than equations; thus, the associated estimation problem is under-determined. Constraining or regularizing the variables are typically used to mitigate this under-determinedness. Here, by exploring the inherent relations among the variables in (9), we can find that the following composed matrix is a rank one matrix:

$$
H(\vartheta,\theta) = \begin{bmatrix} \tau & \vartheta_1 & \cdots & \vartheta_q \\ 1 & \theta_1 & \cdots & \theta_q \end{bmatrix},
\tag{11}
$$

where $\vartheta = \begin{bmatrix} \vartheta_0 & \cdots & \vartheta_q \end{bmatrix} \in \mathbb{R}^{n^2 \times q}$.

As will be shown in the following proposition, equation (8) is equivalent to equation (10) with a rank constraint.

Proposition 1: Assume that there exists a unique solution pair (θ, τ) for equation (8). The variables (ϑ, θ) can be uniquely determined from equation (10) subject to the following rank constraint:

$$
\text{rank}\,[H(\vartheta,\theta)] = \text{rank}\begin{bmatrix} \tau & \vartheta_1 & \cdots & \vartheta_q \\ 1 & \theta_1 & \cdots & \theta_q \end{bmatrix} = 1.
\tag{12}
$$

Furthermore, (θ, τ) can be uniquely determined from (ϑ, θ).

Proof: To prove the above proposition, it suffices to prove that the variables (θ, τ) and (ϑ, θ) are in one-to-one mapping under the rank constraint. It is obvious that the variables (ϑ, θ) can be uniquely determined from (θ, τ). Next, we will show that (θ, τ) can be determined from (ϑ, θ) under the rank constraint as well.

Since $H(\vartheta, \theta)$ is a rank one matrix, its SVD decomposition can be represented as $H(\vartheta, \theta) = uv^T$ where $u \in \mathbb{R}^{(n^2+1)}$ and $v \in \mathbb{R}^{(q+1)}$. It follows that $uv^T = \begin{bmatrix} \tau & \vartheta_1 & \cdots & \vartheta_q \\ 1 & \theta_1 & \cdots & \theta_q \end{bmatrix} = \begin{bmatrix} \tau \\ 1 \end{bmatrix} \begin{bmatrix} 1 & \theta^T \end{bmatrix}$ with $\tau \in \mathbb{R}^{n^2}$.

It then follows that $u = \alpha \begin{bmatrix} \tau \\ 1 \end{bmatrix}$ and $v = \frac{1}{\alpha} \begin{bmatrix} 1 \\ \theta \end{bmatrix}$ with α a scalar ambiguity. Obviously, θ and τ can be respectively determined from v and u. Therefore, (θ, τ) can be determined from (ϑ, θ). ∎

By the above proposition, instead of dealing with the bilinear estimation of (8), we shall consider the following equivalent estimation problem:

$$\min_{\vartheta, \theta} \quad \| \sum_{i=0}^{q} M_i \vartheta_i - N_0 - \sum_{i=1}^{q} N_i \theta_i \|_2^2 \qquad (13)$$
$$s.t. \quad \text{rank}\,[H(\vartheta, \theta)] = 1.$$

The above rank constrained optimization problem is NP hard, which is difficult to find a global optimal solution under mild computational burden. One heuristic way for the above optimization problem is to solve the following nuclear norm regularized optimization problem:

$$\min_{\vartheta, \theta} \quad \| \sum_{i=0}^{q} M_i \vartheta_i - N_0 - \sum_{i=1}^{q} N_i \theta_i \|_2^2 + \lambda \| H(\vartheta, \theta) \|_*, \quad (14)$$

where λ is a regularization parameter. Since the above optimization problem is convex, we can reliably obtain an approximate estimate of (ϑ, θ).

Denote by $\sigma_i (H(\vartheta, \theta))$ the i-th largest singular value of $H(\vartheta, \theta)$ for $i = 1, \cdots, q+1$. Let

$$f_\kappa (H(\vartheta, \theta)) = \sum_{i=1}^{\kappa} \sigma_i (H(\vartheta, \theta)) \text{ for } \kappa = 1, \cdots, q+1.$$

It is remarked that $f_\kappa (\cdot)$ is a Ky Fan κ-norm [13]. Inspired by the truncated nuclear norm method in [14], the rank constraint in (13) can be replaced by the following constraint:

$$\sum_{i=2}^{q+1} \sigma_i (H(\vartheta, \theta)) = f_{q+1} (H(\vartheta, \theta)) - f_1 (H(\vartheta, \theta)) = 0.$$

As a consequence, we try to solve the following unconstrained optimization problem:

$$\min_{\vartheta, \theta} \quad \| \sum_{i=0}^{q} M_i \vartheta_i - N_0 - \sum_{i=1}^{q} N_i \theta_i \|_2^2 \qquad (15)$$
$$+ \lambda \left[f_{q+1} (H(\vartheta, \theta)) - f_1 (H(\vartheta, \theta)) \right].$$

According to the definition of $f_\kappa (\cdot)$, we observe that

$$f_{q+1} (H(\vartheta, \theta)) = \| H(\vartheta, \theta) \|_*$$
$$f_1 (H(\vartheta, \theta)) = \| H(\vartheta, \theta) \|_2.$$

Then equation (15) is therefore equivalent to

$$\min_{\vartheta, \theta} \quad \| \sum_{i=0}^{q} M_i \vartheta_i - N_0 - \sum_{i=1}^{q} N_i \theta_i \|_2^2 \qquad (16)$$
$$+ \lambda \left(\| H(\vartheta, \theta) \|_* - \| H(\vartheta, \theta) \|_2 \right).$$

Due to the convex properties of the nuclear norm and the spectral norm, it is easy to see that the above optimization is a difference of convex programming (DCP) problem, which can be efficiently solved by sequential convex programming method [9] or the difference convex algorithm (DCA) [10].

Since these two algorithms have similar convergence properties [15], we shall apply the sequential convex programming method to solve the optimization problem in (16).

Denote by $(\hat{\vartheta}^k, \hat{\theta}^k)$ the estimate of (ϑ, θ) at the k-th iteration. Let $\hat{u}^k$ and $\hat{v}^k$ be the left and right singular vectors associated with the largest singular value of $H(\hat{\vartheta}^k, \hat{\theta}^k)$, respectively. Then, linearizing the concave term of the objective function in (16) and leaving out the constant terms yields

$$\min_{\vartheta, \theta} \quad \| \sum_{i=0}^{q} M_i \vartheta_i - N_0 - \sum_{i=1}^{q} N_i \theta_i \|_2^2 \qquad (17)$$
$$+ \lambda \left(\| H(\vartheta, \theta) \|_* - (\hat{u}^k)^T H(\vartheta, \theta) \hat{v}^k \right).$$

Since (17) is convex, the estimate $(\hat{\vartheta}^{k+1}, \hat{\theta}^{k+1})$ can be reliably obtained using existing convex optimization tools. It is noteworthy that iterative minimization of (17) is the sequential convex programming method.

Since the DCP is still a non-convex optimization problem, the obtained solution using the adopted sequential convex programming method depends on the initial condition. To cope with this, the initial condition is obtained by solving the nuclear norm regularized optimization problem in (14). In fact, the optimization problem in (14) is the convex part of (16). Since the nuclear norm is the convex envelope of the low-rank constraint on the unit ball of the operator norm [16], the solution to (14) is usually a good candidate for the starting point of the sequential convex programming method. Numerous simulation results show that based on such an initial condition, the sequential convex programming method usually leads to a global optimal solution.

By the identifiability assumptions on the model structure, the matrix T has to be nonsingular at the global optimal point. However, getting stuck in local optimal points may cause the estimate of matrix T to be ill-conditioned and even singular [8]. Due to the continuity of the equation in (3), there always exists a region around the true solution such that the global optimal solution can be obtained once the initial point of the DCP programming approach falls in this region. In other words, if the initial solution is very close the global optimal point, the estimate of T will be automatically nonsingular.

IV. NUMERICAL SIMULATION

In this section, two examples are simulated to show the performance of the proposed identification method. The first example comes from [11], which is to identify compartmental structures of a networked system. The second one is to estimate a state-space represented printer belt derive model with its parameters having physical iterpretation [6]. The system orders of these two examples are set to three.

To evaluate the performance of the developed identification method, we use the root normalized mean square error criterion which is defined as

$$\text{rNMSE} = \sqrt{ \frac{1}{K} \sum_{i=1}^{K} \frac{\| \hat{\theta}^i - \theta^* \|^2}{\| \theta^* \|^2} }, \qquad (18)$$

where K is the number of Monte-Carlo trials which is set to 50 in the following simulations, $\hat{\theta}^i$ and θ^* are the i-th estimated and the true parameters, respectively.

In the simulations, both the system input and measurement noise are white noises which are generated independently. Therefore, the persistent excitation property for the system input and the consistent estimation conditions are satisfied with probability one. To estimate the system matrices, we employ the MOESP subspace identification method [2], [17]. In addition, the convex optimization problems involved are solved using the CVX toolbox. The regularization parameter λ in (14) is empirically set to $\lambda = 10^{-3}$. For each example, we show the identification performance at different noise levels. Moreover, we choose 400 input-output data for the system identification in each Monte-Carlo trial. The stopping criterion for the proposed iterative identification method is set to

$$\frac{\|\hat{\theta}^{i+1} - \hat{\theta}^i\|}{\|\hat{\theta}^i\|} \leq 10^{-6}.$$

Example 1: In this example, the system model comes from a networked system with compartmental structures [11]. The involved system matrices are parameterized as follows:

$$A(\theta) = \begin{bmatrix} -\theta_1 & \theta_3 & \\ \theta_1 & -\theta_2 - \theta_3 & \theta_4 \\ & \theta_2 & -\theta_4 \end{bmatrix}, B(\theta) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$
$$C(\theta) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}. \tag{19}$$

Compared with the unidentifiable example in Section II, the matrix C in this example has a different value, and the concerned system model here has been proven to be identifiable [11]. The true value of the parameter vector is set to

$$\theta = \begin{bmatrix} -0.394 & -0.893 & 0.325 & 0.383 \end{bmatrix}.$$

For the ease of reference, solving the optimization problem in (14) to estimate the system parameters is called the **nu**clear-**n**orm (NUN) method. The obtained estimate of the system parameters is the initial point for the *DCP method* presented in Section III. For the comparison purpose, based on initial point obtained by the *NUN method*, we carry out the **a**lternating **mi**nimization (AMI) method [7] to iteratively estimate the system parameters. For the purpose of fair comparison, both the *DCP method* and the *AMI method* run for 100 iterations.

Fig. 1 shows the identification performance of three different algorithms. We can find that the rNMSE curve corresponding to the *NUN method* changes slightly along with the increase of SNR. More specifically, the estimates obtained by the *NUN method* fluctuate around the point

$$\hat{\theta}^0 = \begin{bmatrix} -0.091 & -0.304 & 0.156 & -0.018 \end{bmatrix}.$$

However, those rNMSE curves of the *DCP method* and the *AMI method* decays along with the increase of SNR. From the comparison between the *DCP* and *AMI methods*, we can
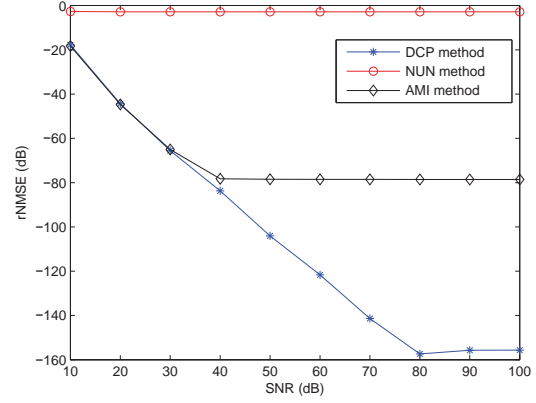

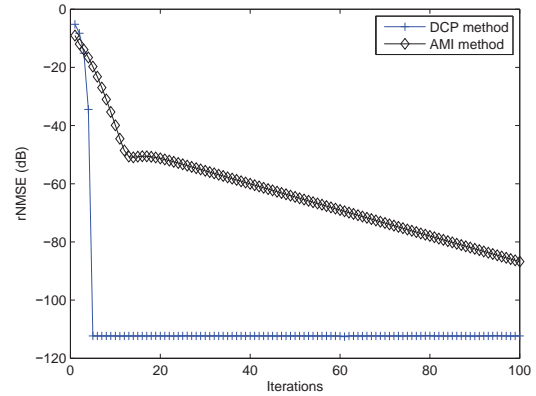Fig. 1. Example 1: identification performance against the SNR

Fig. 2. Example 1: identification performance against the number of iterations for a Monte-Carlo trial at SNR=50 dB.

see that the performance of the *AMI method* is worse than that of the *DCP method*.

Fig. 2 shows the identification performances of the *DCP* and *AMI methods* on a Monte-Carlo trial at SNR=50 dB, where we can find that the *AMI method* has not met the final stopping criterion up to 100 iterations. This explains the worse performance of the *AMI method* shown in Fig. 1. On the other hand, we can see that the *DCP method* can meet the stopping criterion within 5 iterations in this simulation example, indicating that the *DCP method* converges faster.

It is remarked that the same initial point is provided for the *DCP* and *AMI methods* at the zeroth iteration. However, the rNMSE curves in Fig. 2 start from the first iteration. This explains why their corresponding rNMSE values at the first iteration are slightly different.

Example 2: The example is extracted from [6], which aims to identify the physical parameters governing a printer belt drive system. The associated system matrices are param-

eterized as follows:

$$A(\theta) = \begin{bmatrix} 0 & -1 & 0.15 \\ 0.2 & 0 & 0 \\ \theta_1 & \theta_2 & \theta_3 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ \theta_4 \end{bmatrix}, C = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}^T. \tag{20}$$

The true value of θ is set to $\theta = \begin{bmatrix} -0.537 & 0.567 & -0.363 & 0.156 \end{bmatrix}$. All other simulation settings are the same as the previous example.

The identification performances of three different methods are shown in Fig. 3. The obtained estimates of the system parameters by the *NUN method* fluctuate around the point

$$\hat{\theta}^0 = \begin{bmatrix} -0.143 & 0.006 & -0.340 & 0.157 \end{bmatrix}.$$

Analogous to the previous example, the rNMSE curves of the *DCP* and *AMI methods* decay along with the increase of SNR. From Fig. 4, we can find that the *DCP method* can meet the stopping criterion within 20 iterations while the *AMI method* is not able to meet the stopping criterion up to 100 iterations.
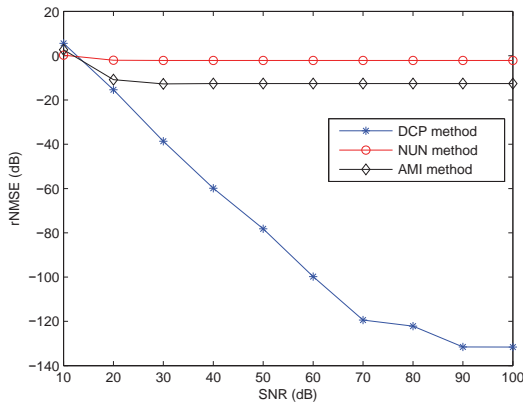

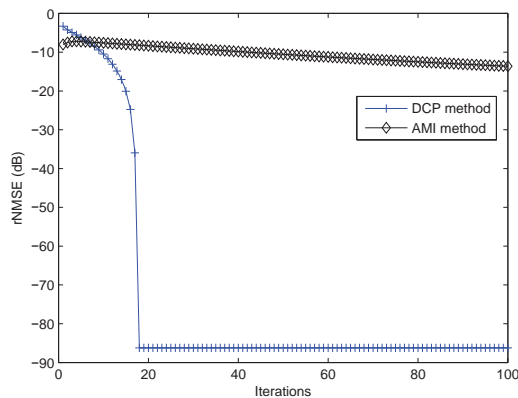Fig. 3.   Example 2: identification performance against the SNR.

Fig. 4.   Example 2: identification performance against the number of iterations for a Monte-Carlo trial at SNR=50 dB.

V. CONCLUSION

In this paper, the identification of structured state-space model has been studied and a new identification method has been devised. Key to the proposed identification method is that the concerned bilinear estimation is transformed to a rank constrained estimation problem and further a DCP problem, for which the initial condition can be robustly estimated by solving its convex optimization part. Simulation results show that the developed method usually converges to the global optimal solution. While the current approach is dedicated to the identification of structured systems, it can be straightforwardly generalized to solve many other bilinear estimation problems, such as the output feedback controller design, blind image deconvolution, source signal separation and so on.

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*. Pearson Education, 1998.
[2] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, 2007.
[3] C. Lyzell, "Structural reformulations in system identification," *Linköping Studies in Science and Technology. Dissertations*, no. 1475, 2012.
[4] C. Lyzell, M. Enqvist, and L. Ljung, "Handling certain structure information in subspace identification," 2009.
[5] O. Prot, G. Mercère, and J. Ramos, "A null-space-based technique for the estimation of linear-time invariant structured state-space representations," in *Proceedings of the IFAC Symposium on System Identification*, 2012.
[6] 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.
[7] L.-L. Xie and L. Ljung, "Estimate physical parameters by black box modeling," in *Proceedings of the 21st Chinese Control Conference*, pp. 673–677, 2002.
[8] L. Ljung and P. Parrilo, "Initialization of physical parameter estimates," 2003.
[9] S. Boyd, "Sequential convex programming," *Lecture Notes, Stanford University*, 2008.
[10] P. D. Tao *et al.*, "The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems," *Annals of Operations Research*, vol. 133, no. 1-4, pp. 23–46, 2005.
[11] R. Bellman and K. J. Åström, "On structural identifiability," *Mathematical biosciences*, vol. 7, no. 3, pp. 329–339, 1970.
[12] J. Norton, E. Walter, and L. Pronzato, *Identification of Parametric Models: from Experimental Data*. Communications and Control Engineering, Springer London, 2010.
[13] R. Bhatia, *Matrix Analysis*. Graduate Texts in Mathematics, Springer New York, 2013.
[14] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 9, pp. 2117–2130, 2013.
[15] G. R. Lanckriet and B. K. Sriperumbudur, "On the convergence of the concave-convex procedure," in *Advances in neural information processing systems*, pp. 1759–1767, 2009.
[16] 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.
[17] M. Verhaegen, V. Verdult, and N. Bergboer, "Filtering and system identification: an introduction to using matlab software," *Delft University of Technology*, vol. 68, 2007.