Gray box identification using difference of convex programming

Yu, Chengpu; Ljung, Lennart; Verhaegen, Michel

DOI
10.1016/j.ifacol.2017.08.1469

Publication date
2017

Document Version
Final published version

Published in
IFAC-PapersOnLine

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 50-1 (2017) 9462–9467

Gray Box Identification Using Difference of Convex Programming

Chengpu Yu * **Lennart Ljung** ** **Michel Verhaegen** ***

* *Delft Center for Systems and Control, Delft University, Delft 2628CD, Netherlands (c.yu-4@tudelft.nl)*
** *Division of Automatic Control, Department of Electrical Engineering, Linkoping University, Sweden (ljung@isy.liu.se)*
*** *Delft Center for Systems and Control, Delft University, Delft 2628CD, Netherlands (m.verhaegen@tudelft.nl)*

Abstract: Gray-box identification is prevalent in modeling physical and networked systems. However, due to the non-convex nature of the gray-box identification problem, good initial parameter estimates are crucial for a successful application. In this paper, a new identification method is proposed by exploiting the low-rank and structured Hankel matrix of impulse response. This identification problem is recasted into a difference-of-convex programming problem, which is then solved by the sequential convex programming approach with the associated initialization obtained by nuclear-norm optimization. The presented method aims to achieve the maximum impulse-response fitting while not requiring additional (non-convex) conditions to secure non-singularity of the similarity transformation relating the given state-space matrices to the gray-box parameterized ones. This overcomes a persistent shortcoming in a number of recent contributions on this topic, and the new method can be applied for the structured state-space realization even if the involved system parameters are unidentifiable. The method can be used both for directly estimating the gray-box parameters and for providing initial parameter estimates for further iterative search in a conventional gray-box identification setup.

Keywords: Structured state-space model, convex-concave procedure.

1. INTRODUCTION

Nowadays, the control and identification of structured state-space system model have attracted great attention in the control community. There are two main sources of structured state-space models: the modeling of practical physical systems (Dorf and Bishop, 2011; Ljung, 1999; Verhaegen and Verdult, 2007) and the description of networked systems (Bellman and Åström, 1970; Van den Hof, 1998). When modeling physical systems, the non-zero entries of the system matrices always have physical meanings. Identification of the physical parameters can provide us a better understanding of the inner physical mechanism of the investigated object. On the other hand, a network connected system often can be represented as a structured system with the structure straightforwardly determined by the interconnections among the involved subsystems. Identification of such kind of structured system models provides the foundation for the model-based network control.

One simple example of gray-box model is the DC Servomotor (Ljung, 1999, Example 4.1) with time constant τ and steady state gain to angular velocity β. The continuous-time state-space expression is

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ 0 & -\frac{1}{\tau} \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \frac{\beta}{\tau} \end{bmatrix} u(t) \\ y(t) &= [\,1 \;\; 0\,]\, x(t). \end{aligned} \tag{1}$$

By reparameterizing $\theta_1 = \frac{-1}{\tau}, \theta_2 = \frac{\beta}{\tau}$, we have

$$\begin{aligned} \dot{x}(t) &= \left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \theta_1 \right) x(t) + \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \theta_2 \right) u(t) \\ y(t) &= [\,1 \;\; 0\,]\, x(t). \end{aligned}$$

Although the system matrices above are parameterized linearly in only two variables, it might be cumbersome to identify these two parameters using the sampled input-output (IO) data, unless we have good initial estimates.

In the literature, there are two kinds of methods to identify structured state-space models. *One* is the traditional gray-box set-up, to identify the parameterized state-space models directly from the IO data using the prediction-error method (Ljung, 1999; Verhaegen and Verdult, 2007). Since the involved identification problem is always a non-convex optimization problem, the conventional nonlinear optimization methods, such as regularized Gauss-Newton method (Ljung, 1999, Section 10.2), and the gradient project method (Verhaegen and Verdult, 2007, Chapter 7), are sensitive to the initial parameter estimate. This traditional setup thus requires reasonable knowledge of the parameters and structures to be identified. Since the gray box situation starts from some physical insight, this knowl-

⋆ This work is sponsored by the European Research Council, Advanced Grant Agreement No. 339681.

10.1016/j.ifacol.2017.08.1469

edge may be sufficient in some cases, but too demanding in other. Resorting to testing random initial parameters may not be feasible for large problems.

The other approach to structured state-space models is to first estimate an unstructured, black-box model using, e.g., subspace identification methods, followed by the recovery of the physical parameters embedded in the structured model. Using the classical subspace identification methods, such as MOESP and N4SID (Ljung, 1999; Verhaegen and Verdult, 2007), the system matrices in the first step can be consistently estimated under some mild conditions. The parameter recovery in the second step turns out to be a small-scale bilinear optimization problem.

To solve the bilinear optimization problem involved with the gray-box identification, an alternating minimization algorithm was developed in (Xie and Ljung, 2002) and a null-space based method was proposed in (Prot et al., 2012). In order to prevent the singular similarity transformation, a non-smooth optimization approach was presented in (Mercere et al., 2014). Furthermore, in order to avoid estimating the similarity transformation, an H_∞-norm-based identification algorithm was proposed in (Bergamasco and Lovera, 2013; Vizer et al., 2016). The above mentioned algorithms are sensitive to initial conditions. To cope with this problem, the bilinear optimization problem was reformulated into a sum-of-squares of polynomials which is then solved by the semi-definite programming method (Ljung and Parrilo, 2003); however, this technique is limited to solving small-scale problems having only a few unknown variables.

In this paper, a difference-of-convex programming (DCP) based method is developed for the identification of structured state-space models. This approach estimates the system parameters by the structured factorization of a block Hankel matrix of system impulse response, which is inspired by the Ho-Kalman decomposition method (Verhaegen and Verdult, 2007). More explicitly, the proposed method boils down to solving a structured low-rank matrix factorization problem. In this paper, this non-convex optimization problem is transformed into a difference-of-convex (DC) optimization problem which is then tackled by the sequential convex programming technique (Boyd, 2008).

The advantages of the proposed method against many recently developed methods are as follows. Different from the identification method in (Mercere et al., 2014), the proposed algorithm framework avoids the non-singularity constraint on the similarity transformation and can be applied to the realization of non-identifiable gray-box models. Unlike the model-matching H_∞ method (Vizer et al., 2016) which requires to solve an infinite-dimensional optimization problem, the proposed identification method is finite-dimensional so that it is more computational amenable.

2. PROBLEM STATEMENT

We consider a parameterized state-space model as follows

$$\begin{aligned} \dot{x}(t) &= A(\theta)x(t) + B(\theta)u(t) \\ y(kT) &= C(\theta)x(kT) + w(kT), \end{aligned} \tag{2}$$

where $u(t) \in \mathbb{R}^m, x(t) \in \mathbb{R}^n, y(t) \in \mathbb{R}^p$ and $w(k) \in \mathbb{R}^p$ are system input, state, output, and measurement noise, respectively; $\theta \in \mathbb{R}^l$ is the parameter vector; t and k represent continuous and discrete time indices, respectively; T is the sampling period.

The parameter vector θ in (2) typically represents unknown values of physical coefficients. Here, we assume that the structured system matrices are affine with respect to θ, i.e.

$$A(\theta) = A_0 + \sum_{i=1}^{l} A_i\theta_i, \quad B(\theta) = B_0 + \sum_{i=1}^{l} B_i\theta_i,$$
$$C(\theta) = C_0 + \sum_{i=1}^{l} C_i\theta_i,$$

where the coefficient matrices A_i, B_i and C_i are known. Besides the structures of the system matrices, the system order of (2) is known as a priori knowledge as well. The DC Servomotor example given in the introduction exactly fits the above structure.

Denote the corresponding true continuous-time transfer function by:

$$G(s,\theta) = C(\theta)\left(sI - A(\theta)\right)^{-1} B(\theta). \tag{3}$$

Although state-evolution equation in (2) is continuous, we can only obtain sampled IO data in practice with sampling period T. Denoting the discrete-time system, obtained by the sampling period T with the system input $u(t)$ being piecewise constant between the sampling instants kT, as

$$H(q,\theta) = C(\theta)\left(qI - A_T(\theta)\right)^{-1} B_T(\theta), \tag{4}$$

where

$$A_T = e^{A(\theta)T}, \quad B_T = \int_{\tau=0}^{T} e^{A(\theta)\tau} B(\theta) d\tau.$$

Given the sampled IO data $\{u(kT), y(kT)\}$ for $k = 0, 1, \cdots$ that are generated from model (2) for a certain value θ^*, the concerned gray-box identification problem is to estimate the parameter vector θ^* from the measured IO data.

3. GRAY-BOX APPROACH

The estimation of the parameter vector θ^* using the sampled IO data $\{u(kT), y(kT)\}$ is typically a *gray-box identification problem*. The traditional identification method is the prediction-error method (Ljung, 1999) in which the predicted or simulated outputs $\hat{y}(kT|\theta)$ are computed using the discrete-time model $H(q,\theta)$ for any θ. The corresponding prediction error criterion is written as

$$\begin{aligned} \min_{\theta} \quad & \frac{1}{N}\sum_{k=0}^{N-1} \|y(kT) - \hat{y}(kT|\theta)\|^2 \\ s.t. \quad & \hat{x}(kT+T) = A_T(\theta)\hat{x}(kT) + B_T(\theta)u(kT) \\ & \hat{y}(kT|\theta) = C(\theta)\hat{x}(kT) \quad \text{for } k = 0, \cdots, N-1. \end{aligned} \tag{5}$$

This general method has the best possible asymptotic accuracy, but the main drawback is that the optimization problem is (highly) non-convex and may have many local minima. The gradient-based optimization algorithms such as Gauss-Newton method (Ljung, 1999, Section 10.2), and gradient projection method (Verhaegen and Verdult, 2007, Chapter 7) can be used to solve (5). However,

the performance mainly relies on the selection of initial parameter estimate. The gray-box structure information may be sufficient to provide such initial estimates that are in the domain of attraction of the global minimum but otherwise one may have to resort to random initial parameters. It is shown in Ljung and Parrilo (2003) that the chances to reach the global minimum of (5) from random starting points may be very slim for problems of realistic sizes.

4. BLACK-BOX + ALGEBRAIC APPROACH

Besides the gray-box approach, there exist other routes to estimate the parameter vector θ^* from the sampled IO data. Even though the gray-box approach may end up in local minima, it is still possible to find the true system from data by a black-box approach. Subspace approaches like N4SID and MOESP (Ljung, 1999; Verhaegen and Verdult, 2007) can, under mild conditions, obtain the true discrete-time system $H(q,\theta^*)$ as the length of the IO data tends to infinity. That discrete-time system can be easily transformed to continuous-time using the zero-order hold interpolation approach (Franklin et al., 1998). As a result, the continuous-time transfer function $G(s,\theta^*)$ will be known, but in an unknown state-space basis:

$$G(s,\theta^*) = C^*(sI - A^*)^{-1}B^*. \tag{6}$$

The identification problem has now been transformed to an algebraic problem:
Given the values of A^*, B^*, C^*, determine the parameter vector θ satisfying

$$C^*(sI - A^*)^{-1}B^* = C(\theta)(sI - A(\theta))^{-1}B(\theta). \tag{7}$$

The estimate of θ obtained in this way can then be used as initial estimate in the minimization of (5). This approach was discussed in (Xie and Ljung, 2002; Ljung and Parrilo, 2003; Mercere et al., 2014).

5. SOLVING THE ALGEBRAIC PROBLEM

To solve the algebraic problem in (7), two routes are provided here: one is the similarity transformation of the state-space realization and the other is the structured low-rank factorization of the block Hankel matrix constructed by the system impulse response.

5.1 Using Similarity Transformation

Equation (7) means that there exists a similarity transformation Q such that

$$QA^* = A(\theta)Q, \quad QB^* = B(\theta), \quad C^* = C(\theta)Q. \tag{8}$$

From that we can form the criterion

$$\begin{aligned} V(Q,\theta) = &\|QA^* - A(\theta)Q\|_F^2 + \|QB^* - B(\theta)\|_F^2 \\ &+ \|C^* - C(\theta)Q\|_F^2 \end{aligned} \tag{9}$$

The optimization problem in (9) is a bilinear estimation problem and an alternating minimization method was proposed in (Xie and Ljung, 2002). In Ljung and Parrilo (2003), the optimization problem in (9) was minimized by a convex sum-of-squares method in case $A(\theta), B(\theta), C(\theta)$ are affine in θ; however, this method is limited to solving small-scale problems having only rather few unknown variables. Recently, to solve the bilinear optimization problem (9), a gradient projection method was given in (Verhaegen and Verdult, 2007, Chapter 7.5.4), a null-space-based optimization method was developed in Prot et al. (2012) and a difference-of-convex based method was proposed in Yu et al. (2015).

Note that a solution pair $(Q^\star, \theta^\star)$ to equation (8) might not be meaningful when $Q^\star$ is singular. In fact, equations (7) and (8) are equivalent if and only if Q is nonsingular (Kailath, 1980). To deal with this problem, a condition-number constraint on Q was considered in (Mercere et al., 2014), which turns out to be a non-smooth and highly non-convex optimization problem.

5.2 Using the Hankel Matrix of Impulse Response

In this section, aiming at dealing with the possible drawback of minimizing equation (9), a new identification approach is developed in this subsection. Firstly, given the impulse response of the concerned continuous system, a block Hankel matrix is constructed. Secondly, by exploiting the low-rank property of the block Hankel matrix and the shift properties of its row and column subspaces, the gray-box identification problem is formulated as a structured low-rank matrix factorization problem. Finally, the low-rank matrix factorization problem is numerically solved using the difference-of-convex programming method.

After obtaining a full-parameterized state-space realization $G(s,\theta^*) = C^*(sI - A^*)^{-1}B^*$, we can obtain the associated impulse response sequence denoted by

$$M_i(\theta^*) = C(\theta^*)A^i(\theta^*)B(\theta^*) = C^*(A^*)^iB^*$$

for $i = 0, 1, \cdots$. Let $H_{v,h}(\theta^*)$ be a block Hankel matrix constructed by Markov parameters

$$H_{v,h}(\theta^*) = \begin{bmatrix} M_0(\theta^*) & M_1(\theta^*) & \cdots & M_{h-1}(\theta^*) \\ M_1(\theta^*) & M_2(\theta^*) & \cdots & M_h(\theta^*) \\ \vdots & \vdots & \iddots & \vdots \\ M_{v-1}(\theta^*) & M_v(\theta^*) & \cdots & M_{v+h-2}(\theta^*) \end{bmatrix}, \tag{10}$$

where the subscripts v, h, satisfying $v, h \geq n$, denote the number of block rows and number of block columns, respectively. Given the block Hankel matrix $H_{v,h}(\theta^*)$, the concerned gray-box identification problem is formulated as

$$\begin{aligned} &\min_{\theta} \quad \|H_{v,h}(\theta^*) - H_{v,h}(\theta)\|_F^2 \\ &s.t. \quad H_{v,h}(\theta) = \begin{bmatrix} C(\theta)B(\theta) & \cdots & C(\theta)A^{h-1}(\theta)B(\theta) \\ \vdots & \iddots & \vdots \\ C(\theta)A^{v-1}(\theta)B(\theta) & \cdots & C(\theta)A^{v+h-2}(\theta)B(\theta) \end{bmatrix}. \end{aligned} \tag{11}$$

In the above equation, the block Hankel matrix $H_{v,h}(\theta)$ has a low-rank factorization as

$$\begin{aligned} H_{v,h}(\theta) = &\underbrace{\begin{bmatrix} C(\theta) \\ C(\theta)A(\theta) \\ \vdots \\ C(\theta)A^{v-1}(\theta) \end{bmatrix}}_{\mathcal{O}_v(\theta)} \\ &\times \underbrace{\begin{bmatrix} B(\theta) & A(\theta)B(\theta) & \cdots & A^{h-1}(\theta)B(\theta) \end{bmatrix}}_{\mathcal{C}_h(\theta)}, \end{aligned} \tag{12}$$

where $\mathcal{O}_v(\theta)$ and $\mathcal{C}_h(\theta)$ denote the extended observability and controllability matrix, respectively.

Denote $Y = H_{v,h}(\theta^*)$. By exploiting the shift properties embedded in extended observability and controllability matrices, the optimization problem (11) can be recasted into a structured low-rank matrix factorization problem:

$$\begin{aligned}\min_{\theta,\mathcal{O}_v,\mathcal{C}_h,X} \quad & \|Y-X\|_F^2 \\ s.t. \quad & X = \mathcal{O}_v\mathcal{C}_h \\ & \mathcal{O}_v\left(1:p,:\right) = C(\theta) \\ & \mathcal{C}_h\left(:,1:m\right) = B(\theta) \\ & \mathcal{O}_v\left(1:(v-1)p,:\right)A(\theta) = \mathcal{O}_v\left(p+1:vp,:\right) \\ & A(\theta)\mathcal{C}_h\left(:,1:(h-1)m\right) = \mathcal{C}_h\left(:,m+1:hm\right).\end{aligned} \tag{13}$$

In the above optimization problem, the first and the last two constraints in the above equation are bilinear. To solve this problem, the DCP-based identification framework (Yu et al., 2015) will be adopted, which contains the following three steps: (i) the bilinear optimization problem is transformed into a rank constrained optimization problem; (ii) the rank constrained problem is recasted into a DCP problem; (iii) the DCP problem is then solved using the sequential convex programming technique.

Step 1: The first constraint, $X = \mathcal{O}_v\mathcal{C}_h$, in (13) can be equivalently written as a rank constraint.

Lemma 1. (Doelman and Verhaegen, 2016) The bilinear equation $X = \mathcal{O}_v\mathcal{C}_h$ is equivalent to the rank constraint

$$\operatorname{rank}\begin{bmatrix} X & \mathcal{O}_v \\ \mathcal{C}_h & I_n \end{bmatrix} = n. \tag{14}$$

Analogously, the equivalent rank constraints for the last two constraints of (13) will be derived below. To simplify the notation, we denote $\bar{\mathcal{O}}_v = \mathcal{O}_v(1:(v-1)p,:)$, $\underline{\mathcal{O}}_v = \mathcal{O}_v(p+1:vp,:)$ and $\bar{\mathcal{C}}_h = \mathcal{C}_h(:,1:(h-1)m)$, $\underline{\mathcal{C}}_h = \mathcal{C}_h(:,m+1:hm)$. The last two constraints can be represented as

$$\begin{aligned}\bar{\mathcal{O}}_v A_0 + \sum_{i=1}^{q}\left(\bar{\mathcal{O}}_v\theta_i\right)A_i &= \underline{\mathcal{O}}_v \\ A_0\bar{\mathcal{C}}_h + \sum_{i=1}^{q} A_i\left(\bar{\mathcal{C}}_h\theta_i\right) &= \underline{\mathcal{C}}_h.\end{aligned} \tag{15}$$

An equivalent form of the combination of the fourth and fifth constraints is given in the following lemma.

Lemma 2. The constraint equation (15) is equivalent to

$$\begin{aligned}&\bar{\mathcal{O}}_v A_0 + \sum_{i=1}^{q}\Gamma_i A_i = \underline{\mathcal{O}}_v \\ &A_0\bar{\mathcal{C}}_h + \sum_{i=1}^{q} A_i\Upsilon_i = \underline{\mathcal{C}}_h \\ &\operatorname{rank}\begin{bmatrix} 1 & \theta_1 & \cdots & \theta_q \\ \operatorname{vec}(\bar{\mathcal{O}}_v) & \operatorname{vec}(\Gamma_1) & \cdots & \operatorname{vec}(\Gamma_q) \\ \operatorname{vec}(\bar{\mathcal{C}}_h) & \operatorname{vec}(\Upsilon_1) & \cdots & \operatorname{vec}(\Upsilon_q) \end{bmatrix} = 1.\end{aligned} \tag{16}$$

Proof of the above lemma can be found in Proposition 1 of Yu et al. (2015).

By Lemmas 1 and 2, the bilinear optimization problem in (13) can be equivalently formulated as a rank-constrained optimization problem as follows:

$$\begin{aligned}\min_{\theta,\mathcal{O}_v,\mathcal{C}_h,X,\Gamma,\Upsilon} \quad & \|Y-X\|_F^2 \\ s.t. \quad & \operatorname{rank}\begin{bmatrix} X & \mathcal{O}_v \\ \mathcal{C}_h & I_n \end{bmatrix} = n \\ & \mathcal{O}_v\left(1:p,:\right) = C(\theta) \\ & \mathcal{C}_h\left(:,1:m\right) = B(\theta) \\ & \bar{\mathcal{O}}_v A_0 + \sum_{i=1}^{q}\Gamma_i A_i = \underline{\mathcal{O}}_v \\ & \bar{\mathcal{O}}_v = \mathcal{O}_v(1:(v-1)p,:),\ \ \underline{\mathcal{O}}_v = \mathcal{O}_v(p+1:vp,:) \\ & A_0\bar{\mathcal{C}}_h + \sum_{i=1}^{q} A_i\Upsilon_i = \underline{\mathcal{C}}_h \\ & \bar{\mathcal{C}}_h = \mathcal{C}_h(:,1:(h-1)m),\ \ \underline{\mathcal{C}}_h = \mathcal{C}_h(:,m+1:hm) \\ & \operatorname{rank}\begin{bmatrix} 1 & \theta_1 & \cdots & \theta_q \\ \operatorname{vec}(\bar{\mathcal{O}}_v) & \operatorname{vec}(\Gamma_1) & \cdots & \operatorname{vec}(\Gamma_q) \\ \operatorname{vec}(\bar{\mathcal{C}}_h) & \operatorname{vec}(\Upsilon_1) & \cdots & \operatorname{vec}(\Upsilon_q) \end{bmatrix} = 1.\end{aligned} \tag{17}$$

The above optimization contains two rank constraints. To deal with the above rank constrained optimization, we shall further formulate it as a difference of convex optimization problem.

Step 2: For notational simplicity, we denote

$$T = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_q \\ \operatorname{vec}(\bar{\mathcal{O}}_v) & \operatorname{vec}(\Gamma_1) & \cdots & \operatorname{vec}(\Gamma_q) \\ \operatorname{vec}(\bar{\mathcal{C}}_h) & \operatorname{vec}(\Upsilon_1) & \cdots & \operatorname{vec}(\Upsilon_q) \end{bmatrix}.$$

Let $\sigma_i(T)$ be the i-th largest singular value of T for $i = 1, \cdots, q+1$. Define $f_\kappa(T) = \sum_{i=1}^{\kappa}\sigma_i(T)$, where $f_\kappa(\cdot)$ is a convex function (Qi and Womersley, 1996).

Inspired by the truncated nuclear norm method in (Hu et al., 2013; Yu et al., 2015), the rank constraint $\operatorname{rank}(T) = 1$ can be replaced by

$$f_{q+1}(T) - f_1(T) = \|T\|_* - f_1(T) = 0. \tag{18}$$

The above equation means that all the singular values of T except the largest one are zero.

Using the above strategy, instead of directly solving the rank constrained optimization problem in (17), we try to solve the following regularized optimization problem:

$$\begin{aligned}\min_{\theta,\mathcal{O}_v,\mathcal{C}_h,X,\Gamma,\Upsilon} \quad & \|Y-X\|_F^2 + \lambda_1\left(\|\Gamma\|_* - f_n(\Gamma)\right) \\ & + \lambda_2\left(\|T\|_* - f_1(T)\right) \\ s.t. \quad & \Gamma = \begin{bmatrix} X & \mathcal{O}_v \\ \mathcal{C}_h & I_n \end{bmatrix} \\ & \mathcal{O}_v\left(1:p,:\right) = C(\theta) \\ & \mathcal{C}_h\left(:,1:m\right) = B(\theta) \\ & \bar{\mathcal{O}}_v A_0 + \sum_{i=1}^{q}\Gamma_i A_i = \underline{\mathcal{O}}_v \\ & \bar{\mathcal{O}}_v = \mathcal{O}_v(1:(v-1)p,:),\ \ \underline{\mathcal{O}}_v = \mathcal{O}_v(p+1:vp,:) \\ & A_0\bar{\mathcal{C}}_h + \sum_{i=1}^{q} A_i\Upsilon_i = \underline{\mathcal{C}}_h \\ & \bar{\mathcal{C}}_h = \mathcal{C}_h(:,1:(h-1)m),\ \ \underline{\mathcal{C}}_h = \mathcal{C}_h(:,m+1:hm) \\ & T = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_q \\ \operatorname{vec}(\bar{\mathcal{O}}_v) & \operatorname{vec}(\Gamma_1) & \cdots & \operatorname{vec}(\Gamma_q) \\ \operatorname{vec}(\bar{\mathcal{C}}_h) & \operatorname{vec}(\Upsilon_1) & \cdots & \operatorname{vec}(\Upsilon_q) \end{bmatrix},\end{aligned} \tag{19}$$

where λ_1, λ_2 are non-negative regularization parameters. It is remarked that all the constraints in (19) are linear functions with respect to the unknown variables and the objective function is a difference-of-convex function.

Although the formulations (17) and (19) may not be strictly equivalent, they have the same global optimum.

Step 3: We shall develop a sequential convex programming method to solve the DC optimization problem in (19). In order to develop a sequential convex programming method, it is essential to linearize the concave terms in the objective function of (19). Let Γ^j be the estimate of Γ at the j-th iteration and its SVD decomposition be given as

$$\Gamma^j = \begin{bmatrix} U_1^j & U_2^j \end{bmatrix} \begin{bmatrix} S_1^j & \\ & S_2^j \end{bmatrix} \begin{bmatrix} V_1^{j,T} \\ V_2^{j,T} \end{bmatrix}, \tag{20}$$

where U_1^j and V_1^j are respectively the left and right singular vectors corresponding to the largest n singular values. It can be established that (Qi and Womersley, 1996)

$$U_1^j V_1^{j,T} \in \partial f_n\left(\Gamma^j\right). \tag{21}$$

Then, the linearization of $f_n(\Gamma)$ at the point $\Gamma = \Gamma^j$ is

$$f_n(\Gamma) \approx f_n(\Gamma^j) + \mathrm{tr}\left(U_1^{j,T}\left(\Gamma - \Gamma^j\right) V_1^j\right). \tag{22}$$

Let T^j be the estimate of T at the j-th iteration and its SVD decomposition be given as

$$T^j = \begin{bmatrix} L_1^j & L_2^j \end{bmatrix} \begin{bmatrix} \Sigma_1^j & \\ & \Sigma_2^j \end{bmatrix} \begin{bmatrix} R_1^{j,T} \\ R_2^{j,T} \end{bmatrix}, \tag{23}$$

where L_1^j and R_1^j are respectively the left and right singular vectors corresponding to the largest singular value. Then, the linearization of $f_1(T)$ at the point $T = T^j$ is

$$f_1(T) \approx f_1(T^j) + \mathrm{tr}\left(L_1^{j,T}\left(T - T^j\right) R_1^j\right). \tag{24}$$

Based on the linearizations in (22) and (24), the convex optimization problem to be solved at the $(j+1)$-th iteration is as follows:

$$\begin{aligned}
\min_{\theta,\mathcal{O}_v,\mathcal{C}_h,X,\Gamma,\Upsilon} \quad & \|Y - X\|_F^2 + \lambda_1\left(\|\Gamma\|_* - \mathrm{tr}\left(U_1^{j,T}\Gamma V_1^j\right)\right) \\
& + \lambda_2\left(\|T\|_* - \mathrm{tr}\left(L_1^{j,T} T R_1^j\right)\right) \\
s.t. \quad & \Gamma = \begin{bmatrix} X & \mathcal{O}_v \\ \mathcal{C}_h & I_n \end{bmatrix} \\
& \mathcal{O}_v\left(1:p,:\right) = C(\theta) \\
& \mathcal{C}_h\left(:,1:m\right) = B(\theta) \\
& \bar{\mathcal{O}}_v A_0 + \sum_{i=1}^{q} \Gamma_i A_i = \underline{\mathcal{O}}_v \\
& \bar{\mathcal{O}}_v = \mathcal{O}_v(1:(v-1)p,:),\ \ \underline{\mathcal{O}}_v = \mathcal{O}_v(p+1:vp,:) \\
& A_0\bar{\mathcal{C}}_h + \sum_{i=1}^{q} A_i \Upsilon_i = \underline{\mathcal{C}}_h \\
& \bar{\mathcal{C}}_h = \mathcal{C}_h(:,1:(h-1)m),\ \ \underline{\mathcal{C}}_h = \mathcal{C}_h(:,m+1:hm) \\
& T = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_q \\ \mathrm{vec}(\bar{\mathcal{O}}_v) & \mathrm{vec}(\Gamma_1) & \cdots & \mathrm{vec}(\Gamma_q) \\ \mathrm{vec}(\bar{\mathcal{C}}_h) & \mathrm{vec}(\Upsilon_1) & \cdots & \mathrm{vec}(\Upsilon_q) \end{bmatrix}.
\end{aligned} \tag{25}$$

To ease the reference, the above sequential convex programming procedure is summarized in Algorithm 1.

Since the difference-of-convex optimization problem in (19) is still non-convex, the performance of the provided sequential convex programming procedure relies on the initial conditions. However, by setting $T^0 = 0$ and $\Gamma^0 = 0$, we can find that the optimization problem in (19) is a nuclear-norm relaxation of the rank-constrained optimization problem in (17). Due to the fact that the nuclear norm is the convex envelope of the low-rank constraint on the unit spectral norm ball (Recht et al., 2010), the associated nuclear-norm optimization usually yields a good candidate for the starting point of the sequential convex programming procedure.

Algorithm 1 Sequential convex programming method for (19)

1) Set $\Gamma^0 = 0$ and $T^0 = 0$.
2) Repeat
   2-1): Compute respectively the left and right singular vectors of Γ^j and T^j as shown in (20) and (23).
   2-2): Obtain the estimates Γ^{j+1} and T^{j+1} by solving (25).
3) until $\frac{\|\theta^{j+1}-\theta^j\|_2}{\|\theta^j\|_2} \le \varepsilon$ with ε a small value.

6. NUMERICAL SIMULATIONS

The performance of the proposed identification method - Algorithm 1- is demonstrated in this section. For comparison purposes, the prediction-error method (PEM) (Ljung, 1999; Ljung and Parrilo, 2003) and the difference-of-convex programming (DCP) method (Yu et al., 2015) are simulated. The implementation details of these three methods are given below.

(1) Algorithm 1 is simulated by empirically setting the regularization parameters in (25) to $\lambda_1 = 10^{-4}$ and $\lambda_2 = 10^{-5}$. The tolerance of relative error is set to $\varepsilon = 10^{-4}$.
(2) PEM is simulated by firstly configuring the structure object using the Matlab command `idgrey`, see Ljung (2013) and then implementing the identification method using the Matlab command `pem`. The initial conditions are randomly generated following the standard Gaussian distribution.
(3) DCP method is simulated by setting the regularization parameter λ in equation (17) of Yu et al. (2015) to $\lambda = 10^{-4}$.

In the simulations, the maximum number of iterations for the these three methods is set to 100.

The simulation is conducted following the way in Ljung and Parrilo (2003). The state-space model of (2) is randomly generated by the Matlab command `rss`, and the system parameters to be estimated are randomly picked from the generated models. When simulating the DCP method and Algorithm 1, the system matrices A^*, B^* and C^* in (7) are assumed to be known.

To ensure the identifiability of the system parameters, the number of unknown parameters cannot be larger than $(p+m)n$; however, system parameters less than $(p+m)n$ may not always be identifiable (Ljung, 1999; Verhaegen and Verdult, 2007). Therefore, we use the *impulse-response fitting* to measure the identification performance. In the simulation, we choose the system order $n = 5$ and the input/output dimension $m = p = 1$. For each fixed number of free parameters, we carry out 100 Monte-Carlo trials by randomly generating the system model and randomly picking a fixed number of free parameters. The success rate is obtained by counting the number of successful trials using the criterion $\mathrm{IRF}^r \le 10^{-3}$. Denote by θ^r the estimate of θ at the r-th Monte-Carlo trial. The impulse-response fitting (IRF) of the r-th Monte-Carlo trial is defined as

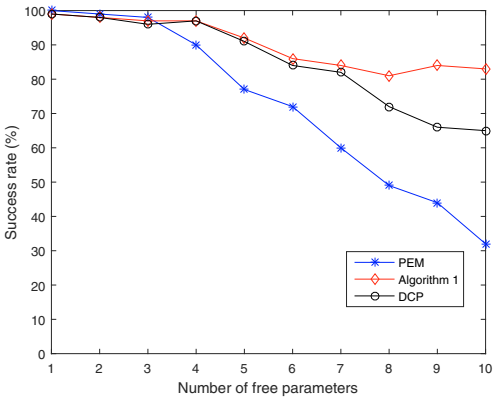

Fig. 1. Example 1: identification performance of PEM, DCP and Algorithm 1.

$$\mathrm{IRF}^r = \frac{\sum_{i=0}^{v+h-2} \|C(\theta^r)A^i(\theta^r)B(\theta^r) - C^*(A^*)^iB^*\|_F}{\sum_{i=0}^{v+h-2} \|C^*(A^*)^iB^*\|_F}, \tag{26}$$

where the dimension parameters v and h are defined in (10).

The identification performance of PEM, DCP and Algorithm 1 is shown in Fig. 1 from which we can draw the following conclusions.

(1) When the number of parameters is larger than 3, DCP and Algorithm 1 perform much better than PEM. This is because DCP and Algorithm 1 can find good initializations by nuclear-norm regularized optimization. In contrast, when the number of parameters are less than or equal to 3, PEM has a slightly better performance than DCP and Algorithm 1. This might be relevant to the selection of the regularization parameters of DCP and Algorithm 1.
(2) When the number of parameters is larger than 6, the success rate of Algorithm 1 is higher than that of DCP up to 20%. This might be caused by the fact that DCP does not consider the non-singularity constraint of the similarity transformation, while Algorithm 1 implicitly guarantees the non-singularity of the similarity transformation. However, when the number of parameters is less than or equal to 6, DCP and Algorithm 1 have similar performance. This might be because, when the number of free parameters becomes smaller, singular similarity transformations are less likely to occur.

7. CONCLUSIONS

In this paper, we have proposed a new gray-box identification method by exploiting the low-rank and structured factorization of the Hankel matrix of impulse response. This method uses the system impulse-response fitting as the objective function while avoiding the explicit non-singularity constraint on similarity transformation; thus, it can be applied to the state-space realizations of non-identifiable gray-box models. Compared with the classical prediction-error method initialized at random parameter values, the proposed method can yield better performance since it can find a good initialization by nuclear-norm based optimization.

REFERENCES

Bellman, R. and Åström, K.J. (1970). On structural identifiability. *Mathematical biosciences*, 7(3), 329–339.

Bergamasco, M. and Lovera, M. (2013). State space model identification: from unstructured to structured models with an h_∞ approach. *IFAC Proceedings Volumes*, 46(2), 202–207.

Boyd, S. (2008). Sequential convex programming. *Lecture Notes, Stanford University*.

Doelman, R. and Verhaegen, M. (2016). Sequential convex relaxation for convex optimization with bilinear matrix equalities. In *2016 European Control Conference*.

Dorf, R.C. and Bishop, R.H. (2011). *Modern control systems*. Pearson.

Franklin, G.F., Powell, J.D., and Workman, M.L. (1998). *Digital control of dynamic systems*, volume 3. Addison-wesley Menlo Park.

Hu, Y., Zhang, D., Ye, J., Li, X., and He, X. (2013). Fast and accurate matrix completion via truncated nuclear norm regularization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(9), 2117–2130.

Kailath, T. (1980). *Linear systems*, volume 1. Prentice-Hall Englewood Cliffs, NJ.

Ljung, L. (1999). *System Identification: Theory for the User*. Pearson Education.

Ljung, L. (2013). *The System Identification Toolbox: The Manual*. The MathWorks Inc. 1st edition 1986, Edition 8.3 2013, Natick, MA, USA.

Ljung, L. and Parrilo, P. (2003). Initialization of physical parameter estimates.

Mercere, G., Prot, O., and Ramos, J. (2014). 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*, 59(11), 2873–2885. doi:10.1109/TAC.2014.2351853.

Prot, O., Mercere, G., and Ramos, J. (2012). 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*.

Qi, L. and Womersley, R.S. (1996). On extreme singular values of matrix valued functions. *Journal of Convex Analysis*, 3, 153–166.

Recht, B., Fazel, M., and Parrilo, P.A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3), 471–501.

Van den Hof, J. (1998). Structural identifiability of linear compartmental systems. *Automatic Control, IEEE Transactions on*, 43(6), 800–818.

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

Vizer, D., Mercere, G., Prot, O., and Laroche, E. (2016). h_∞-norm-based optimization for the identification of gray-box lti state-space model parameters. *Systems & Control Letters*, 92, 34 – 41.

Xie, L.L. and Ljung, L. (2002). Estimate physical parameters by black box modeling. In *Proceedings of the 21st Chinese Control Conference*, 673–677.

Yu, C., Verhaegen, M., Kovalsky, S., and Basri, R. (2015). Identification of structured lti mimo state-space models. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, 2737–2742. IEEE.