Subspace identification of 1D spatially-varying systems using Sequentially Semi-Separable matrices

Sinquin, Baptiste; Verhaegen, Michel

DOI
10.1109/ACC.2016.7524891

Publication date
2016

Document Version
Accepted author manuscript

Published in
Proceedings of the 2016 American Control Conference (ACC 2016)

Accepted Author Manuscript. Link to published article (IEEE): http://dx.doi.org/10.1109/ACC.2016.7524891

Subspace Identification of 1D Spatially-Varying Systems using Sequentially Semi-Separable matrices

Baptiste Sinquin[1], Michel Verhaegen[1]

***Abstract*— We consider the problem of identifying 1D spatially-varying systems that exhibit no temporal dynamics. The spatial dynamics are modeled via a mixed-causal, anti-causal state space model. The methodology is developed for identifying the input-output map of e.g a 1D flexible beam described by the Euler-Bernoulli beam equation and equipped with a large number of actuators and sensors. It is shown that the static input-output map between the lifted inputs and outputs possess a so-called Sequentially Semi-Separable (SSS) matrix structure. This structure is of key importance to derive algorithms with linear computational complexity for controller synthesis of large-scale systems. A nuclear norm subspace identification method of the N2SID class is developed for estimating these state space models from input-output data. To enable the method to deal with a large number of repeated experiments a dedicated Alternating Direction Method of Multipliers (ADMM) algorithm is derived. It is shown in this paper that a nuclear norm relaxation on the SSS structure can be imposed which improves the estimates of the system matrices.**

***Index Terms*— nuclear norm subspace identification, spatially distributed systems, sequentially semi-separable matrices.**

I. Introduction

Research on scalable algorithms for system identification has received increased attention over the past years and is driven by projects such as smart grids for power networks, large scale adaptive optics [1] and also coordinated transportation or biology [2] to name but a few. Efficient modelling of large systems with thousands of actuators and sensors with a compact representation may be a key in starting control synthesis. This was e.g highlighted in [3] where efficient control design methods with linear computational complexity in the system size dimension were derived when the original state space system matrices where belonging to the so-called Sequentially Semi-Separable (SSS) matrices.

In general, Partial Differential Equations (PDE) model both spatial and temporal dynamics. However, for systems that are used in a frequency band far below their first resonance frequency, the temporal dynamics can be neglected, e.g deformable mirrors in adaptive optics. A spatio-temporal discretization of a 1D thin plate [4] can be described by the following string interconnected state space model:

$$x_j(k+1) = A_j x_j(k) + A_{\ell,j} x_{j-1}(k) + A_{r,j} x_{j+1}(k) + B_j u_j(k)$$

where (j,k) are indices respectively for space and time. The quasi-static assumptions implies $x_j(k+1) = x_j(k)$, hence the aforementioned system representation is re-casted into a descriptor form:

$$\begin{bmatrix} -A_{r,j} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_{j+1}(k) \\ x_j(k) \end{bmatrix} = \begin{bmatrix} A_j - I & A_{\ell,j} \\ I & 0 \end{bmatrix} \begin{bmatrix} x_j(k) \\ x_{j-1}(k) \end{bmatrix} + \begin{bmatrix} B_j \\ 0 \end{bmatrix} u_j(k) \tag{1}$$

in which $A_{r,j}$ is not necessarily full rank. The Kronecker-Weierstrass canonical form [5] that decouples causal and anti-causal directions from (1) give rise to a mixed causal, anti-causal state space model:

$$\begin{cases} \begin{bmatrix} x^c_{j+1,k} \\ x^a_{j-1,k} \end{bmatrix} = \begin{bmatrix} R_j & 0 \\ 0 & W_j \end{bmatrix} \begin{bmatrix} x^c_{j,k} \\ x^a_{j,k} \end{bmatrix} + \begin{bmatrix} Q_j \\ V_j \end{bmatrix} u_{j,k} \\ y_{j,k} = \begin{bmatrix} P_j & U_j \end{bmatrix} \begin{bmatrix} x^c_{j,k} \\ x^a_{j,k} \end{bmatrix} + D_j u_{j,k} + e_{j,k} \end{cases} \tag{2}$$

with $x^c_{j,k} \in \mathbb{R}^{n^c_j}$, $x^a_{j,k} \in \mathbb{R}^{n^a_j}$ and the measurement noise $e_j \in \mathbb{R}^{p_j}$. j denotes a spatial index within the range of indices $[1, N_s]$ whereas k represents the index of a given experiment within $[1, N_t]$. $u_{j,k} \in \mathbb{R}^{m_j}$ and $y_{j,k} \in \mathbb{R}^{p_j}$ are respectively the input and measurement from the k-th experiment taken at spatial position j. The map between the lifted inputs $u_{j,k}$ and outputs $y_{j,k}$ at time instant k is a SSS matrix. In adaptive optics (AO) this map is called the influence matrix and considered as dense in the so-called Matrix-Vector Mutliplication approach. The methods presented in this paper impose a SSS structure on this matrix. This structure opens the way for development of algorithms for control of extremely large AO systems with linear computational complexity. In this paper we focus as an important initial step on the identification of the generators of the SSS input/output map. SSS matrices were introduced in [6] in which their efficiency for distributed and fast computations is highlighted. The spatially-varying dynamics of system (2) in the context of a 1D flexible beam stem e.g from the varying material properties but also from the dimensions

[1] Both authors are with the Delft Center for Systems and Control, Delft University of Technology, the Netherlands.
Corresponding author: b.sinquin@tudelft.nl

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

of the inputs $u_{j,k}$ and outputs $y_{j,k}$ that can vary with the position j. This arises e.g when actuators and sensors are not collocated.

The model (2) describes a mixed causal, anti-causal linear time-varying (LTV) system. This class of systems was analyzed in the context of SSS matrices in [7]. The identification of causal LTV in the old MOESP framework was considered in [8], the identification of mixed causal, anti-causal LTI state space models in that same framework was considered in [9]. In this paper we generalize the methods of [8] to the case of mixed causal, anti-causal varying systems and treat the problem in the new N2SID framework [10]. *It leads to three major improvements over existing works. First, it enables to handle both the mixed causal, anti-causal and time-varying properties. Second, it avoids the use of the Kronecker decomposition in separating the causal and anti-causal part. The avoidance of the calculation of the generalized eigenvalues contributes to numerical robustness. A third improvement that contributes to more robustness is in the estimation of the system matrices without the need of projections and the possibility to take the constraints of the SSS structure into account.* The derived N2SID variant can be used as an initial estimate to the parametric identification methods of SSS models as proposed in [11] following the Extended Kalman filtering methodology or in [12] using output-error identification method.

The paper is organized as follows. Section II introduces structural notions and formulates the data equation, Section III recasts the non-unique rank minimization problem into a unique convex problem with structure information on the SSS matrix. Section IV uses the shift-invariance property of the SSS matrix to estimate the generators up to a similarity transformation. Dealing with large datasets is made possible through the use of ADMM whose implementation is described in Section V. The method is illustrated with numerical experiments in Section VI.

Notations. $X^\dagger$ and X^T represent respectively the Moore pseudo-inverse and transpose of X. The Frobenius norm of the matrix X is denoted as $\|X\|_F$. $rank(X)$ is equal to the number of non-zero singular values of X whereas the nuclear norm $\|X\|_\star$ is defined as the sum of the singular values. For X and Y matrices the inequality $X \prec Y$ ($\preceq$) means that $Y-X$ is (semi-)positive-definite. The inner product $\langle X, Y\rangle$ is equal to $Trace(X^TY)$. $\lfloor x \rfloor$ denotes the floor part of the real number x. The standard Matlab notations are used for both vectors and matrices: $X(:,i)$ denotes the i-th column of X, $X(i,:)$ the i-th line. $X(:)$ is the matrix X reshaped columnwise into a vector. $x_{a:b}$ denotes the sequence $(x_a, \ldots, x_b)$. The block-diagonal matrix $\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$ is written with $blkdiag(X, Y)$.

II. Problem formulation

The discrete spatial dynamics of the 1D system under consideration (2) are defined as a function of the spatial index j that takes integer values in the interval $[1, N_s - s + 1]$. Here s is an integer defining the size of the matrices to be processed in the subspace identification method. On the interval $[1, N_s - s + 1]$ we consider a window of size s as depicted in Figure 1, on which we define the following extended observability matrix with $s > n_j^c + n_{j+s-1}^a$:

$$\begin{aligned} \mathcal{O}_{j:j+s-1} &:= (\mathcal{O}^c_{j:j+s-1} | \mathcal{O}^a_{j:j+s-1}) \\ &:= \begin{bmatrix} P_j & U_j W_{j+1} W_{j+2} \ldots W_{j+s-1} \\ P_{j+1} R_j & U_{j+1} W_{j+2} \ldots W_{j+s-1} \\ \vdots & \vdots \\ P_{j+s-1} R_{j+s-2} \ldots R_j & U_{j+s-1} \end{bmatrix} \end{aligned}$$

A definition for uniform observability is mentioned in

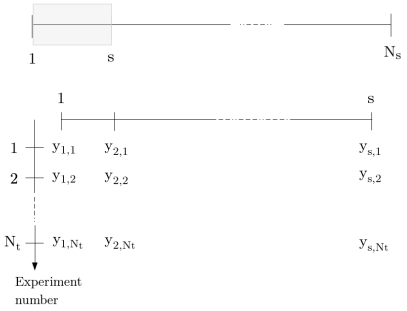

Fig. 1. Schematic representation of a spatial local zone. Above: general view. Under: concatenation of experiments for a zone starting at index 1.

[15] for LTV systems. We extend it here to the mixed causal, anti-causal case.

Definition 1: *The pairs (P_j, R_j) and (U_{j+s-1}, W_{j+s-1}) are uniformly observable if there exists a strictly positive integer l and positive constants b_1, b_2 such that the following inequality on the Gramian holds:*

$$0 \prec b_1 I \preceq \mathcal{O}_{j:j+l-1} \mathcal{O}^T_{j:j+l-1} \preceq b_2 I$$

If we denote the smallest integer l for which this holds, Sylvester inequality is used to prove that:

$$rank(\mathcal{O}_{j:j+l-1}) = n_j^c + n_{j+l-1}^a$$

The controllability matrix is defined with:

$$\begin{aligned} C_{j:j+s-1} &:= \begin{bmatrix} C^c_{j:j+s-1} \\ C^a_{j:j+s-1} \end{bmatrix} \\ &= \begin{bmatrix} Q_{j+s-1} & \ldots & R_{j+s-1} R_{j+s-2} \ldots Q_j \\ W_j W_{j+1} \ldots V_{j+s-1} & \ldots & V_j \end{bmatrix} \end{aligned}$$

Controllability and observability are dual notions, and uniform controllability is defined similarly as in Definition 1.

Definition 3: *Let us consider the mixed causal, anti-causal state space representation* (2)*. The set of SSS generators is the set of matrices* $\{P_j, R_j, Q_j, D_j, U_j, W_j, V_j\}$.

Definition 4: *A bounded set of SSS generators is said to be uniform if it is uniformly observable and uniformly controllable.*

We are now ready to formulate the problem.

Consider:

the state space representation (2) for j defined in the interval $[1, N_s]$, the experiment index k in the range $[1, N_t]$, and measurements of inputs $u_{j,k}$ and outputs $y_{j,k}$, then:

Estimate, for all $j \in [2, N_s - 1]$**:**

- the causal and anti-causal orders (n_j^c, n_j^a) corresponding to a uniform SSS realization.
- the generators $P_j, R_j, Q_j, D_j, U_j, W_j, V_j$, up to a similarity transformation.

For a local spatial zone $j \in [1, s]$ we have the following relationship:

$$\begin{bmatrix} y_{1,1} \\ \vdots \\ y_{s,1} \end{bmatrix} = \mathcal{O}_{1:s} \begin{bmatrix} x_{1,1}^c \\ x_{s,1}^a \end{bmatrix} + \mathcal{T}_{1:s} \begin{bmatrix} u_{1,1} \\ \vdots \\ u_{s,1} \end{bmatrix} + \begin{bmatrix} e_{1,1} \\ \vdots \\ e_{s,1} \end{bmatrix} \quad (3)$$

where $\mathcal{T}_{1:s}$ is a *SSS matrix* defined in (4). $N_t - 1$ other independent experiments are performed and input-output data is compactly written in the data equation:

$$Y_{1:s,N_t} = \mathcal{O}_{1:s} X_{1:s,N_t} + \mathcal{T}_{1:s} U_{1:s,N_t} + E_{1:s,N_t} \quad (5)$$

where

$$Y_{1:s,N_t} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,N_t} \\ \vdots & \vdots & & \vdots \\ y_{s,1} & y_{s,2} & \cdots & y_{s,N_t} \end{bmatrix}$$

and

$$X_{1:s,N_t} = \begin{bmatrix} x_{1,1}^c & x_{1,2}^c & \cdots & x_{1,N_t}^c \\ x_{s,1}^a & x_{s,2}^a & \cdots & x_{s,N_t}^a \end{bmatrix}$$

The matrices $U_{1:s,N_t}$ and $E_{1:s,N_t}$ are built similarly to $Y_{1:s,N_t}$ from the data $u_{j,k}$ and $e_{j,k}$. When $s > n_j^c + n_{j+s-1}^a$, from Sylvester's inequality it comes that the matrix $\mathcal{O}_{1:s} X_{1:s,N_t}$ is of low rank. Hence identifying the SSS matrix $\mathcal{T}_{1:s}$ in the N2SID framework [10] boils down to the following optimization problem:

$$\min_{\hat{Y}_{1:s,N_t}, \mathcal{T}_{1:s,N_t}} rank(\hat{Y}_{1:s,N_t} - \mathcal{T}_{1:s} U_{1:s,N_t}) + \lambda \|\hat{Y}_{1:s,N_t} - Y_{1:s,N_t}\|^2 \quad (6)$$

with λ a regularization parameters that establishes the trade-off between the two cost functions. Uniqueness of the rank minimization problem shall be studied here. For the N2SID method the Toeplitz structure was sufficient to guarantee a unique solution to the rank minimization problem. For (6) more care is required.

III. Towards a convex and structured optimization problem

For sake of clarity we drop in this section the indices s and N_t mentioning respectively the spatial position and the number of experiments. The causal and anti-causal orders are n^c and n^a.

A. A non-unique solution to the rank minimization problem

Lemma 1: *We assume the input is persistently exciting such that the compound matrix* $\begin{bmatrix} X \\ U \end{bmatrix}$ *has full row rank. Then in the noise-free case* $E = 0$*, the solution* $\mathcal{T}$ *of:*

$$\min_{\mathcal{T}} \quad rank(Y - \mathcal{T}U) \quad (7)$$

is not unique.

Proof: Let Θ_1 and Θ_2 be two solutions of (7). The difference of both solutions $\Delta := \Theta_1 - \Theta_2$ is introduced.

$$\begin{aligned} Y - \Theta_1 U &= Y - (\Theta_2 + \Delta)U \\ &= \begin{bmatrix} \mathcal{O} & \Delta \end{bmatrix} \begin{bmatrix} X \\ -U \end{bmatrix} \end{aligned}$$

Using Sylvester's inequality and the persistence of excitation of the input show that:

$$rank\left(\begin{bmatrix} \mathcal{O} & \Delta \end{bmatrix} \begin{bmatrix} X \\ -U \end{bmatrix} \right) = rank(\begin{bmatrix} \mathcal{O} & \Delta \end{bmatrix})$$

For $\Delta = 0$, the rank equals $n^a + n^c$ and the rank is minimal. Now we can find a non-zero Δ such that the above rank does not increase:

$$rank(\begin{bmatrix} \mathcal{O} & \Delta \end{bmatrix}) = n^c + n^a \quad (8)$$

One example of such a Δ is:

$$\Delta = \begin{bmatrix} P_1 \\ P_2 R_1 \\ \vdots \\ P_s R_{s-1} \dots R_1 \end{bmatrix} \begin{bmatrix} 1_{n^c \times 1} & 0 & \dots & 0 \end{bmatrix} \quad (9)$$

■

B. A biased, although unique, solution with the nuclear norm relaxation

The nuclear norm convex relaxation in subspace identification used in [13] and [10] has open up the way for improvements upon classical methods. Let us relax (7) into the following nuclear norm optimization:

$$\min_{\mathcal{T}} \quad \|Y - \mathcal{T}U\|_\star \quad (10)$$

It is well known that the nuclear norm relaxation introduces a bias as it minimizes the whole vector of singular

$$\mathcal{T}_{1:s} = \begin{bmatrix} D_1 & U_1V_2 & U_1W_2V_3 & \dots & U_1W_2\dots W_{s-1}V_s \\ P_2Q_1 & D_2 & U_2V_3 & \dots & U_2W_3\dots W_{s-1}V_s \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & U_{s-1}V_s \\ P_sR_{s-1}\dots R_2Q_1 & P_sR_{s-1}\dots R_3Q_2 & \dots & P_sQ_{s-1} & D_s \end{bmatrix} \tag{4}$$

values instead of only the last $s-n$ values that don't contribute to the observable subspace.

Lemma 2: *Let the conditions of Lemma 1 hold and denote T_0 the true values of the SSS matrix, and $\mathcal{T}$ as in* (4)*. Then the estimate $\hat{\mathcal{T}}$ solving* (10) *is unique but biased as follows:*

$$\hat{\mathcal{T}} = T_0 + \Gamma$$

for $\Gamma \neq 0$ when X is not orthogonal to U.

Proof: Following the same line as in the previous proof,

$$\|Y - \hat{\mathcal{T}}U\|_\star = \| \begin{bmatrix} -\Gamma & \mathcal{O} \end{bmatrix} \begin{bmatrix} U \\ X \end{bmatrix} \|_\star \tag{11}$$

The following RQ factorization is introduced:

$$\begin{bmatrix} U \\ X \end{bmatrix} = \begin{bmatrix} R_u & 0 \\ R_{ux} & R_x \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}$$

Inserting this equation in (11) yields:

$$\| \begin{bmatrix} -\Gamma & \mathcal{O} \end{bmatrix} \begin{bmatrix} U \\ X \end{bmatrix} \|_\star = \| \begin{bmatrix} -\Gamma R_u + \mathcal{O}R_{ux} & \mathcal{O}R_x \end{bmatrix} \|_\star \tag{12}$$

$\begin{bmatrix} U \\ X \end{bmatrix}$ is full row rank, hence the optimal solution to the above problem is $\Gamma = \mathcal{O}R_{ux}R_u^{-1}$. The underbound of the above cost function is obtained by:

$$\|\mathcal{O}R_x\|_\star \leq \| \begin{bmatrix} -\Gamma R_u + \mathcal{O}R_{ux} & \mathcal{O}R_x \end{bmatrix} \|_\star \tag{13}$$

We here proved the solution to the nuclear norm problem (10) is not biased if $R_{ux} = 0$. Determining the bias Γ with the Frobenius norm solution is only valid if no additional structure is enforced as it would be the case when dealing with only-causal varying systems. In other words,

$$\|Y - \hat{\mathcal{T}}U\|_\star \leq \|Y - T_0U\|_\star$$

To prove that the above biased solution is unique, we now consider a deviation γ to the computed solution $\hat{\Gamma} = \mathcal{O}R_{ux}R_u^{-1}$:

$$\Gamma = \hat{\Gamma} + \gamma \tag{14}$$

$$\| \begin{bmatrix} -\Gamma R_u + \mathcal{O}R_{ux} & \mathcal{O}R_x \end{bmatrix} \|_\star = \| \begin{bmatrix} \gamma R_u & -\mathcal{O}R_x \end{bmatrix} \|_\star \tag{15}$$

which yields $\gamma = 0$ when minimizing with respect with γ. Hence the solution to the nuclear norm problem (10) is unique. ■

C. A structured SSS matrix

The block-terms in the matrix $\mathcal{T}_{1:s}$, such as P_2Q_1, which are called with some abuse of terminology the Markov parameters, are not constant along the diagonals. However, $s-1$ Hankel submatrices appear in both lower and upper parts of $\mathcal{T}_{1:s}$ and can be factorized into a product of observability and controllability matrices. For example, by taking the following partition of $\mathcal{T}_{1:s}$:

$$\forall j \in [2, s], \quad \mathcal{T}_{1:s}(\sum_{i=1}^{j-1} p_i + 1 : \sum_{i=1}^{s} p_i, 1 : \sum_{i=1}^{j-1} m_i)$$

with block-columns in reverse order, the generalized Hankel matrix $\mathcal{H}_j^c$ is formed:

$$\begin{aligned} \mathcal{H}_j^c &:= \begin{bmatrix} P_jQ_{j-1} & \dots & P_jR_{j-1}\dots Q_1 \\ P_{j+1}R_jQ_{j-1} & \dots & P_{j+1}R_j\dots Q_1 \\ \vdots & & \vdots \\ P_s\dots Q_{j-1} & \dots & P_s\dots Q_1 \end{bmatrix} \\ &= \begin{bmatrix} P_j \\ P_{j+1}R_j \\ \vdots \\ P_s\dots R_j \end{bmatrix} \begin{bmatrix} Q_{j-1} & \dots & R_{j-1}\dots Q_1 \end{bmatrix} \\ &= \mathcal{O}_{j:s}^c\mathcal{C}_{1:j-1}^c \end{aligned}$$

Each SSS realization is uniform, hence with Sylvester's inequality:

$$\text{rank}(\mathcal{H}_j^c) = n_j^c$$

The upper-side contains as well the generalized Hankel matrices $\mathcal{H}_j^a$ of rank n_j^a:

$$\forall j \in [1, s-1], \quad \mathcal{H}_j^a := \mathcal{O}_{1:j}^a\mathcal{C}_{j+1:s}^a$$

The structure is highlighted in Figure 2.

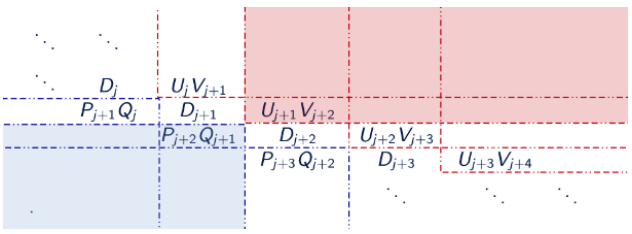

Fig. 2. Schematic representation of the low rank matrices inside a SSS matrix.

However, both n_j^c and n_j^a are unknown and an upper bound $r \in \mathbb{N}$ of these ought to be chosen such that:

$$\max(n) + 1 < r < s - \max(n) + 1 \tag{16}$$

where $\max(n) = \max(\{n_j^c\}_{j\in[2,s]}, \{n_j^a\}_{j\in[1,s-1]})$. The additional low rank constraints on the matrices $\mathcal{H}_j^c$ and $\mathcal{H}_j^a$, that is a consequence of the SSS structure, are integrated by convex relaxation in the cost function (6) as follows:

$$\begin{aligned}\min_{\hat{Y}_{1:s,N},\mathcal{T}_{1:s}} \quad & \|\hat{Y}_{1:s,N} - \mathcal{T}_{1:s}U_{1:s,N}\|_\star + \lambda\|\hat{Y}_{1:s,N} - Y_{1:s,N}\|_F^2 \\ & +\mu\sum_{i=r}^{s-r+2}\left(\|\mathcal{H}_i^c\|_\star + \|\mathcal{H}_{i-1}^a\|_\star\right)\end{aligned} \tag{17}$$

Here λ and μ are regularization parameters.

IV. Matrices estimation

Analysing the SSS matrix $\mathcal{T}_{1:s}$ reveals $7s-8$ generators as it has already been studied in [7]. The procedure to estimate these generators and hence the 7 system matrices to define the mixed causal, anti-causal system (2) is here not rewritten because of space limitation. The quadruplets (P_1, R_1, W_1, V_1) and (R_s, Q_s, U_s, W_s) cannot be determined from $\mathcal{T}_{1:s}$. However only a spatial window of length s has been studied so far. Solving (17) on a minimum of $\lfloor \frac{N_s}{s} \rfloor + 1$ overlapping windows enable to estimate a total of $7N_s - 8$ generators .

V. Alternating Direction Method of Multipliers

A solver is proposed to cope with the

$$n_x = \sum_{i=1}^{s} p_i\left(N_t + \sum_{i=1}^{s} m_i\right)$$

unknowns of problem (17). The Alternating Direction Method of Multipliers (ADMM) is used following the line of [13] and [10]. It is of major importance to relieve the computational burden implied by both the increased number of variables compared to N2SID and the additional low rank constraints. We define the vector of unknowns x first. For all $i \in [1, \sum_{i=1}^{s} p_i]$:

$$\begin{aligned} v_i &= \mathcal{T}_{1:s}(i,:) \qquad \hat{y}_i = \hat{Y}_{1:s,N_t}(i,:) \\ x_i &= \begin{bmatrix}\hat{y}_i & v_i\end{bmatrix} \qquad x = \begin{bmatrix}x_{1:\sum_{i=1}^{s} p_i}\end{bmatrix}\end{aligned} \tag{18}$$

The nuclear norm is not differentiable which is why the ADMM algorithm in [13] introduces a consensus variable to deal with this specificity. The following lemma makes the use of this approach still possible.

Lemma 3: *Let $(Q_1, \ldots, Q_q) \in \mathbb{R}^{m_1\times n_1}\times\ldots\times\mathbb{R}^{m_q\times n_q}$. Let P be the $q-$block diagonal matrix built from the sequence $(Q_1, \ldots, Q_q)$. Then: $\|P\|_\star = \sum_{i=1}^{q}\|Q_i\|_\star$*

The term $\hat{Y}_{1:s,N_t} - \mathcal{T}_{1:s}U_{1:s,N_t}$ is linear in its unknowns (18), the linear operator associated is denoted with $\mathcal{A}_1(x)$. The linear operator $\mathcal{H}_i^c(x_{i:\sum_{i=1}^{s} p_i})$ maps the vector x into the Hankel matrix $\mathcal{H}_i^c$. The linear operator $\mathcal{B}_{\mathcal{H}^c}(.)$ defined as:

$$\mathcal{B}_{\mathcal{H}^c}(x) := blkdiag(\mathcal{H}_t^c(x_{t:\sum_{i=1}^{s} p_i}), \ldots, \mathcal{H}_{s-t+1}^c(x_{s-t+1:\sum_{i=1}^{s} p_i}))$$

maps the vector x into a block diagonal matrix built from the causal Hankel operators on which there is a low rank property enforced in (17). The operator $\mathcal{B}_{\mathcal{H}^a}(.)$ is defined similarly. The linear operator $\mathcal{A}_2(x)$ handles the SSS Hankel structure, both causal and anti-causal:

$$\mathcal{A}_2(x) := blkdiag(\mathcal{B}_{\mathcal{H}^c}(x), \mathcal{B}_{\mathcal{H}^a}(x))$$

Finally we introduce: $\mathcal{A}(x) := blkdiag(\mathcal{A}_1(x), \mathcal{A}_i(x))$

The Gramian matrix M is defined with $\mathcal{A}_{adj}(\mathcal{A}(x)) = Mx$ where $\mathcal{A}_{adj}$ denotes the adjoint of $\mathcal{A}$. Let us introduce a lemma to deal with the high-dimensionality of M and be able to compute it distributively.

Lemma 4: *Let $x \in \mathbb{R}^{n_x}$ and $\mathcal{A}(x)$ a linear operator with block diagonals $\mathcal{A}_1(x)$ and $\mathcal{A}_2(x)$. Then,*

$$\langle\mathcal{A}(x), \mathcal{A}(x)\rangle = \langle x, \sum_{i=1}^{2} M_i x\rangle$$

where M_i for $i = 1, 2$ are such that $\mathcal{A}_{adj,i}(\mathcal{A}_i(x)) = M_i x$. *Proof:* Using the definition of the inner product. ∎

Let us then analyze the Gramian associated to $\mathcal{A}_1(.)$.

Lemma 5:

- *M_1 is block-diagonal, with $\sum_{i=1}^{s} p_i$ blocks $M_1[k]$ of size $\mathbb{R}^{(N_t+\sum_{i=1}^{s} m_i)\times(N_t+\sum_{i=1}^{s} m_i)}$.*
- *Moreover, $M_1[k]$ is given by:*

$$M_1[k] = \begin{bmatrix} I_{N_t} & -U_{1:s,N_t}^T \\ -U_{1:s,N_t} & U_{1:s,N_t}U_{1:s,N_t}^T \end{bmatrix}$$

Proof: The linear operator $\mathcal{A}_1(x)$ is decomposed with:

$$\mathcal{A}_1(x) = \begin{bmatrix}\hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_s\end{bmatrix} - \begin{bmatrix}v_1 \\ v_2 \\ \vdots \\ v_s\end{bmatrix} U_{1:s,N_t}$$

Hence,

$$\begin{aligned}\langle\mathcal{A}_1(x), \mathcal{A}_1(x)\rangle &= \textstyle\sum_{j=1}^{s}\|\hat{y}_j - v_j U_{1:s,N_t}\|_2^2 \\ &= \textstyle\sum_{j=1}^{s}\langle\begin{bmatrix}\hat{y}_j^T \\ v_j^T\end{bmatrix}, \begin{bmatrix} I_{N_t} & -U_{1:s,N_t}^T \\ -U_{1:s,N_t} & U_{1:s,N_t}U_{1:s,N_t}^T\end{bmatrix}\begin{bmatrix}\hat{y}_j^T \\ v_j^T\end{bmatrix}\rangle \\ &= \textstyle\sum_{j=1}^{s}\langle x_j^T, M_1[j]\, x_j^T\rangle\end{aligned}$$

∎

Lemma 6:

- *M_2 is block-diagonal, with $\sum_{i=1}^{s} p_i$ blocks $M_2[k]$ of size $\mathbb{R}^{(N_t+\sum_{i=1}^{s} m_i)\times(N_t+\sum_{i=1}^{s} m_i)}$.*
- *$M_2[k]$ is diagonal:*

$$M_2[k] = \begin{bmatrix} 0_{N_t} & \\ & diag(P_2(k,:)) \end{bmatrix}$$

where $P_2 \in \mathbb{R}^{\sum_{i=1}^{s} p_i \times \sum_{i=1}^{s} m_i}$ *is a matrix such that the element* $P_2(i,j)$ *is equal to the number of times the parameter* $v_i(j)$ *is included in a Hankel operator in* (17).

Proof: $\mathcal{B}_{\mathcal{H}^c}(x)$ is block-diagonal by definition. Lemma 4 indicates that M_2 is the sum of all Gramian matrices associated to each block. Therefore, for one given causal Hankel operator indexed with j:

$$\langle \mathcal{H}_j^c(x_{j:\sum_{i=1}^{s} p_i}), \mathcal{H}_j^c(x_{j:\sum_{i=1}^{s} p_i})\rangle = \langle x, (P_j^c)^T P_j^c x\rangle$$

where P_j^c is a diagonal matrix that selects the entries from $\mathcal{H}_j^c$. The above equation is derived for the anti-causal part as well, which yields:

$$M_2 = \sum_{j=t}^{s-t+1} (P_j^c)^T P_j^c + (P_j^a)^T P_j^a$$

Hence it comes that M_2 is diagonal and contains in the diagonal the number of times a given variable appears in a Hankel operator. ■

Let us consider a 5×5 SSS matrix for illustration. SISO systems of order 1, with $s = 5, r = 3$ are considered. The matrix P_2 is then:

$$P_2 = \begin{bmatrix} 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 2 & 2 \\ 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 \end{bmatrix}$$

The conclusion is that the Gramian matrix can be computed very efficiently by taking into account the structure of the problem. The inverse of a block diagonal matrix is the matrix with inverted diagonal blocks, therefore it is not necessary to form M but rather work with $\sum_{i=1}^{s} p_i$ block matrices.

VI. Numerical experiments

A sequence of $N_s = 30$ SSS generators is randomly generated. Although the methodology applies to mixed causal-anti causal MIMO systems, we consider here SISO with a *causal only* part of order equal to 2 in order to have a comparison with the MOESP method [8]. The upper bounds r and s are respectively chosen equal to 4 and 15, and this choice is not unique. Within the bounds mentioned in (16), it should be such that the number of low rank constraints on the SSS structure is maximized, and hence r is optimally equal to $n + 2$. Experiments have been carried out on Matlab R2015b on a processor Intel Xeon E5-1620 with 8GB memory.

A. Dealing with measurement noise

The first set of experiments analyses how the algorithm behaves in presence of measurement noise. The set of identification data contains $N_t = 100$ independent time experiments. For each of them, the spatial input is zero-mean white Gaussian noise with unit variance. The Signal to Noise Ratio (SNR) ranges from -5dB to 40dB with a step of 5. 100 runs are performed for each SNR. The trade-off parameter λ weighting the fit between the measured and predicted output is generated with $logspace(1, 3, 6)$. Equation (17) is considered first with $\mu = 0$, then logarithmically spaced in the range $\left[10^{-2}, 5\right]$ with 5 values. The quality criteria is the Normalized Root Mean Square Error (NRMSE). For all regularization parameters tested, the system with the smallest NRMSE between the true SSS matrix and the estimated one is selected. Figure 3 highlights a smaller residual error when adding information on the SSS structure. For high SNR, the estimates obtained from (17) don't converge to the true estimates contrary to MOESP. The reason is the bias introduced by the nuclear norm, which has been theoretically analyzed in Lemma 2.

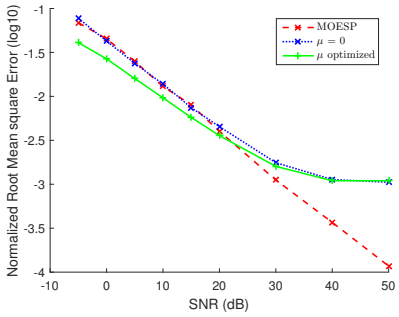

Fig. 3. Normalized Root Mean Square Error between the real values of the SSS entries and the estimated ones as a function of the noise. Mean over 100 different noise conditions.

B. Large datasets

Noise conditions of 5dB are considered in order to analyse how the residual error evolves when increasing the data length available for identification. The same $7N_s$ generators and input data are considered. The data batch is truncated according to the data length wanted. Here again considering the low rank properties of the SSS matrix achieves better results than without as can be seen in Figure 4.

VII. Conclusion

In summary we have formulated the problem of identifying 1D distributed systems as the identification of mixed causal anti-causal *and* spatially varying systems. For this, a new solution is developed with the N2SID framework. The convex relaxation proposed using the nuclear norm leads to a unique solution. The latter however includes a bias that has been reduced by imposing a SSS structure. Finally the memory requirements are relieved by analysing the structure of the Gramian matrix

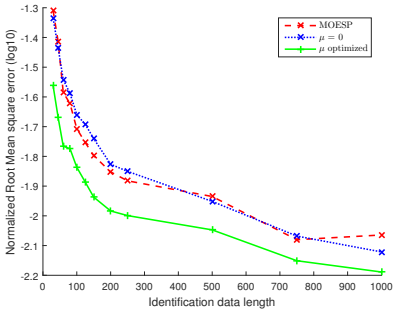

Fig. 4. Normalized Root Mean Square Error between the real values of the SSS entries and the estimated ones as a function of the length of the identification data.

M, and more specifically using the pattern of additional low rank constraints on the SSS matrix.

REFERENCES

[1] R. Gilmozzi, J. Spyromilio, "The European Extremely large Telescope (E-ELT)", *The Messenger 127, 11-19*, 2007.

[2] R. E. Kearney, L. W. Hunter, "System identification of human joint dynamics", *Biomedical Engineering, 18(1):5587*, 1990.

[3] J. K. Rice, M. Verhaegen, "Distributed control in multiple dimensions: a structure preserving computational technique", *IEEE Trans. Autom. Control, Vol. 56, No. 3, pp. 516-530*, 2011.

[4] L. Meirovitch, "Principles and techniques of vibrations", *Prentice Hall*, 1997.

[5] F. R. Gantmacher, "The Theory of Matrices, Vol.2", *Chelsea Publishing Company, New York*, 1960.

[6] S. Chandrasekaran, P. Dewilde, M. Gu, T. Pals, A.J. van der Veen, "Fast stable solvers for sequentially semi-separable linear systems of equations", *Technical report, Lawrence Livermore National Laboratory*, 2003.

[7] P. Dewilde, A-J van der Veen, "Time-Varying Systems and Computations", *Kluwer Academic Publishers, Dordrecht, the Netherlands*, 1998.

[8] M. Verhaegen, X. Yu, "A Class of Subspace Model Identification Algorithms to Identify Periodically and Arbitrarily Time-varying Systems", *Automatica, Vol. 31, No. 2, pp. 201-216*, 1995.

[9] M. Verhaegen, "A subspace model identification solution to the identification of mixed causal, anti-causal LTI systems", *SIAM J. Matrix Anal. Appl., Vol.17, No. 2, pp. 332-347*, 1996.

[10] M. Verhaegen, A. Hansson, "N2SID: Nuclear Norm Subsystem Identification", *arXiv:1501.04495 [cs.SY]*, 2015.

[11] J. K. Rice, M. Verhaegen, "Efficient System Identification of Heterogeneous Distributed Systems via a Structure Exploiting Extended Kalman Filter", *IEEE Trans. Autom. Control, Vol 56, No. 7*, 2011.

[12] P. Torres, J.W. van Wingerden, M. Verhaegen, "Output-Error Identifi- cation of Large Scale 1-D Spatially Varying Interconnected Systems", *IEEE Trans. Autom. Control, Vol. 60, No. 1*, 2015.

[13] Z. Liu, A. Hansson, L. Vandenberghe, "Nuclear norm system identi- fication with missing inputs and outputs", *Syst. Control Lett., vol. 62, no. 8, pp. 605-612*, 2013.

[14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers", *Found. Trends Mach. Learning, 3, pp. 1122*, 2010.

[15] A. H. Jazwinski, "Stochastic Processes and Filtering Theory", *Academic Press, New York*, 1970.