

A Least Squares Approach to the Subspace Identification Problem

L. Bako^{1,2}, G. Mercère³ and S. Lecoeuche^{1,2}

Abstract—In this paper, we propose a new method for the identification of linear Multiple Inputs-Multiple Outputs (MIMO) systems. By introducing a particular user-defined matrix that does not change the rank of the extended observability matrix when multiplying this latter matrix on the left, the subspace identification problem is recasted into a simple least squares problem with all regressors available. Therefore, the Singular Value Decomposition algorithm which is a customary tool in subspace identification can be avoided, thus making our method appealing for recursive implementation. The technique is such that the state coordinates basis of the estimated matrices is completely determined by the aforementioned user-defined matrix, that is, given such a matrix, the state basis of the identified matrices does not change with respect to the realization of input-output data.

I. INTRODUCTION

The identification of linear dynamical Multiple Input-Multiple Output (MIMO) systems achieved in the last two decades a remarkable development from the so-called subspace methods [1], [2], [3]. The main appealing feature of these methods over the more traditional error prediction methods [4], is that they directly provide minimal and not necessarily canonical state space models from input-output (I/O) data.

However, the application of subspace identification methods to the estimation of certain types of systems such as composite systems with linear constituent submodels [5], switched linear systems [6], [7] or recursive identification [8], [9], sometimes comes with some technical difficulties. One issue is related to the multiplicity of possible bases for representing the system equations in the state space. For example, in the case of recursive subspace identification, most of the existing contributions [8], [9] do not guarantee that the state basis remains fixed during the whole recursive identification procedure. This is also a crucial problem when dealing for example with the identification of multi-models or switched systems. Indeed, in these cases, all the different submodels of the system must be obtained in the same basis [6]. Failing to this requirement may have the consequence of altering the I/O behavior of the system.

A second issue is related to the SVD algorithm that is generally required in subspace identification. In fact, subspace methods involve an SVD step in which one decides arbitrarily on the basis of the range space of the extended

observability matrix to be estimated. An important problem is that the SVD is computationally heavy and technically hard to update recursively, thus making its use in online identification a fastidious task [8]. Moreover, updating an SVD neither solves the problem of coordinate basis deviation during the estimation. Therefore, an attempt to reconstruct the I/O behavior from the estimated matrices may result in a shifted behavior.

In this paper, we focus on developing a new, simple, efficient and SVD-free identification method for linear dynamical state space models, that could overcome the aforementioned problems. Since a state model holds only up to a similar transformation, one can indeed choose in advance the basis of the model to be identified. This is carried out by introducing a user-defined matrix Λ_f (notation of the paper), that preserves the rank properties of the extended observability matrix when multiplying this latter matrix on the left. We then show that, under the assumption that the considered system is observable, if one draws Λ_f randomly from a uniform distribution for example, we can obtain a consistent realization of the system. Being based on a conversion of the subspace estimation problem into a Least Squares problem, the developed method can be regarded as an interesting solution to the problem of subspace tracking frequently encountered in signal array processing, recursive system identification and many other applications [10].

The outline of the paper is as follows. In Section II we formulate the subspace identification problem. A relevant sufficiency of excitation concept is defined and used to derive conditions that guarantee consistency of the subspace identification. In Section III, we present a new subspace-based identification algorithm. Some illustrative simulation results shown in Section IV demonstrate the applicability of our method.

II. PROBLEM STATEMENT

We consider a Linear Time-Invariant (LTI) system described by a discrete-time model of the form

$$\begin{cases} x(t+1) &= Ax(t) + Bu(t) + w(t) \\ y(t) &= Cx(t) + Du(t) + v(t), \end{cases} \quad (1)$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_u}$, $y(t) \in \mathbb{R}^{n_y}$ are respectively the state, the input and output vector and $w(t) \in \mathbb{R}^n$ and $v(t) \in \mathbb{R}^{n_y}$ symbolize the process noise and measurement noise. Here, these noises are assumed to be zero-mean white noise processes. A, B, C, D are the system matrices relatively to a certain basis of the state space.

The identification problem can then be formulated as follows: given I/O data $\{u(t), y(t)\}_{t=1}^N$ generated by a system

¹ Ecole des Mines de Douai, Département Informatique et Automatique, 59508, Douai, France

² Laboratoire d'Automatique, Génie Informatique et Signal, UMR CNRS 8146, Université des Sciences et Technologies de Lille, 59655, Villeneuve d'Ascq, France

³ Université de Poitiers, Laboratoire d'Automatique et d'Informatique Industrielle, 86022, Poitiers, France

of the form (1), estimate the matrices (A, B, C, D) in any state coordinates.

To begin with the identification procedure, let us define

$$u_f(t) = [u(t)^\top \cdots u(t+f-1)^\top]^\top \in \mathbb{R}^{fn_u}, \quad (2)$$

with $f > n$. In a similar manner as $u_f(t)$, we define $y_f(t) \in \mathbb{R}^{fn_y}$, $w_f(t) \in \mathbb{R}^{fn}$ and $v_f(t) \in \mathbb{R}^{fn_y}$. Finally, in accordance with these signal vectors, we define the matrices

$$\Gamma_f = [(C)^\top \quad (CA)^\top \quad \cdots \quad (CA^{f-1})^\top]^\top \in \mathbb{R}^{fn_y \times n},$$

$$H_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ CA^{f-2}B & \cdots & CB & D \end{bmatrix} \in \mathbb{R}^{fn_y \times fn_u},$$

$$G_f = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ CA^{f-2} & \cdots & C & 0 \end{bmatrix} \in \mathbb{R}^{fn_y \times fn}.$$

From the recurrent equation (1), one can easily obtain, by successive substitutions over a time horizon $[t, t+f-1]$,

$$y_f(t) = \Gamma_f x(t) + H_f u_f(t) + e_f(t), \quad t \geq 1 \quad (3)$$

with $e_f(t) = G_f w_f(t) + v_f(t)$.

Let N be an integer such that $n < f \ll N$ and define

$$X_{t,N} = [x(t) \quad \cdots \quad x(t+N-1)] \in \mathbb{R}^{n \times N},$$

$$U_{t,f,N} = [u_f(t) \quad \cdots \quad u_f(t+N-1)] \in \mathbb{R}^{fn_u \times N}.$$

Similarly to $U_{t,f,N}$, we define also $Y_{t,f,N} \in \mathbb{R}^{fn_y \times N}$, $E_{t,f,N} \in \mathbb{R}^{fn_y \times N}$. Then, based on Eq. (3), we can write the following block data equation for $t = f+1$.

$$Y_{f+1,f,N} = \Gamma_f X_{f+1,N} + H_f U_{f+1,f,N} + E_{f+1,f,N}. \quad (4)$$

In the literature of subspace identification, data matrices of the form $Y_{1,f,N}$ and $Y_{f+1,f,N}$ are sometimes respectively referred to as *past* and *future* data. In order to alleviate the notations we will, throughout the paper, denote whenever possible the future data more simply as $Y = Y_{f+1,f,N}$, $U = U_{f+1,f,N}$, $E = E_{f+1,f,N}$ and $X = X_{f+1,N}$. In this way, Eq. (4) can be written simply as

$$Y = \Gamma_f X + H_f U + E. \quad (5)$$

Based on this embedded data equation, subspace methods for solving the problem of identifying system (1) proceed by first extracting either the state sequence X or the extended observability matrix Γ_f , making use of geometric projection techniques and rank factorization algorithms such as the SVD [2], [11]. A second step then consists in computing the system matrices in an arbitrary state coordinates basis.

However, the use of these methods in for instance a recursive identification context may be subject to two important issues: 1) the SVD factorization in addition of being computationally cumbersome, is technically difficult to update, 2) the state space basis in which the matrices are obtained

may change during the identification procedure or from one mode to another when dealing with multi-modal systems.

The method developed in the present paper mainly intends to overcome the aforementioned difficulties. But, in a more general framework, the proposed method turns out to be an interesting Least Squares solution to the well known problem of subspace tracking.

Consistent estimation of the parameter matrices in (5) that we can re-write as

$$Y = [\Gamma_f \quad H_f] \begin{bmatrix} X \\ U \end{bmatrix} + E, \quad (6)$$

requires the measured data to have some properties of richness. Therefore, we start by introducing such conditions that will be useful throughout the paper. We first make the following definition of sufficiency of excitation of the exogenous input $\{u(t)\}$.

Definition 1 (sufficiently exciting input sequence): The process $\{u(t)\}$ is said to be sufficiently exciting of order at least l if there exist an integer N_0 and a time index t_0 such that $\text{rank}(U_{t,l,N_0}) = ln_u$ for all $t \leq t_0$. We then say that $\{u(t)\}$ is SE(l).

Note that Definition 1 differs from the classical definition of persistency of excitation [4] that concerns infinite sequences of signals. Here, a finite horizon is considered instead as the number of data available for identification is generally finite in practice.

A consequence of Definition 1 is as follows.

Proposition 1: Assume that the system (1) is reachable and let the noise terms $w(t)$ and $v(t)$ be identically zero in (1). Then the following holds.

If $\{u(t)\}$ is SE($n+f$) in the sense of Definition 1 with $N \geq N_0 + n$ and $f+1 \leq t_0$, then

$$\text{rank} \left(\begin{bmatrix} X_{f+1,N} \\ U_{f+1,f,N} \end{bmatrix} \right) = n + fn_u.$$

Proof: Assume that there are $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^{fn_u}$ verifying

$$\alpha^\top X_{f+1,N} + \beta^\top U_{f+1,f,N} = 0. \quad (7)$$

We then need to show that α and β are both necessarily equal to zero. Since $N \geq N_0 + n$, for any t such that $f+1 \leq t \leq f+1+N-N_0$,

$$\alpha^\top X_{t,N_0} + \beta^\top U_{t,f,N_0} = 0. \quad (8)$$

On the other hand, we can write from the system equations in (1),

$$x(t) = A^n x(t-n) + \Delta_n u_n(t-n), \quad t > n, \quad (9)$$

where $\Delta_n = [A^{n-1}B \quad \cdots \quad AB \quad B] \in \mathbb{R}^{n \times nn_u}$. By making use of the Cayley-Hamilton theorem, we get after some calculations

$$x(t) = -(a^\top \otimes I_n) x_n(t-n) + \Delta_n (\mathcal{K} \otimes I_{n_u}) u_n(t-n), \quad (10)$$

where $x_n(t-n)$ and $u_n(t-n)$ are defined as in (2), \otimes refers to the Kronecker product, $a = [a_1 \quad \cdots \quad a_n]^\top$ is a vector formed with the coefficients a_j , $j = 1, \dots, n$ of the characteristic polynomial of A , $\text{ch}_A(z) = \det(zI - A) =$

$a_1 + a_2 z + \dots + a_n z^{n-1} + z^n$. Here, the matrix \mathcal{K} is defined as

$$\mathcal{K} = \begin{bmatrix} a_2 & a_3 & \cdots & a_{n-1} & a_n & 1 \\ a_3 & a_4 & \cdots & a_n & 1 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ a_n & 1 & \cdots & 0 & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

For all $t > n$, Eq. (10) reads as

$$x(t) + a_n x(t-1) + \dots + a_1 x(t-n) - \Delta_n(\mathcal{K} \otimes I_{n_u}) u_n(t-n) = 0. \quad (11)$$

Define $\tau = f + n + 1 > n$. Then, Eq. (11) implies

$$X_{\tau, N_0} + a_n X_{\tau-1, N_0} + \dots + a_1 X_{\tau-n, N_0} - \Delta_n(\mathcal{K} \otimes I_{n_u}) U_{\tau-n, n, N_0} = 0. \quad (12)$$

Multiplying the equation (12) on the left by α^\top yields

$$\alpha^\top X_{\tau, N_0} + a_n \alpha^\top X_{\tau-1, N_0} + \dots + a_1 \alpha^\top X_{\tau-n, N_0} - \bar{\alpha}^\top U_{\tau-n, n, N_0} = 0, \quad (13)$$

with $\bar{\alpha}^\top = \alpha^\top \Delta_n(\mathcal{K} \otimes I_{n_u})$. Combining now the foregoing equation with (8) results in

$$-\beta^\top U_{\tau, f, N_0} - a_n \beta^\top U_{\tau-1, f, N_0} - \dots - a_1 \beta^\top U_{\tau-n, f, N_0} - \bar{\alpha}^\top U_{\tau-n, n, N_0} = 0. \quad (14)$$

Note that all the matrices of the form U_{k, f, N_0} , $\tau - n \leq k \leq \tau$ can be expressed as a combination of the rows of $U_{\tau-n, f+n, N_0}$. In this way, we have $U_{k, f, N_0} = P_k U_{\tau-n, f+n, N_0}$ for $\tau - n \leq k \leq \tau$, and $U_{k, n, N_0} = Q_k U_{\tau-n, f+n, N_0}$, with

$$P_{\tau-j} = \begin{bmatrix} \mathbf{0}_{f n_u \times (n-j) n_u} & I_{f n_u} & \mathbf{0}_{f n_u \times j n_u} \end{bmatrix} \in \mathbb{R}^{f n_u \times (f+n) n_u}$$

$$Q_{\tau-n} = \begin{bmatrix} I_{n n_u} & \mathbf{0}_{n n_u \times f n_u} \end{bmatrix} \in \mathbb{R}^{n n_u \times (f+n) n_u},$$

for $j = 0, \dots, n$. Using these notations, we get

$$\left(\beta^\top P_\tau + a_n \beta^\top P_{\tau-1} + \dots + a_1 \beta^\top P_{\tau-n} + \bar{\alpha}^\top Q_{\tau-n} \right) U_{\tau-n, f+n, N_0} = 0 \quad (15)$$

Since $\{u(t)\}$ is sufficiently of order at least $n + f$ the matrix $U_{\tau-n, f+n, N_0} = U_{f+1, f+n, N_0}$ is full row rank. Consequently, the term of (15) that is in the parentheses is null. If we let β^\top and $\bar{\alpha}^\top$ be partitioned as

$$\bar{\alpha}^\top = [\bar{\alpha}_1^\top \quad \cdots \quad \bar{\alpha}_n^\top], \quad \bar{\alpha}_j^\top \in \mathbb{R}^{1 \times n_u},$$

$$\beta^\top = [\beta_1^\top \quad \cdots \quad \beta_f^\top], \quad \beta_j^\top \in \mathbb{R}^{1 \times n_u},$$

then, by straightforward calculations, we arrive at

$$\begin{bmatrix} 1 & a_n & \cdots & \cdots & a_1 & & & & & & \\ & 1 & a_n & \cdots & \cdots & a_1 & \mathbf{0} & & & & \\ & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \ddots & & & & \\ & & & & 1 & a_n & \cdots & \cdots & a_1 & & \end{bmatrix} \begin{bmatrix} \beta_1^\top \\ \vdots \\ \beta_f^\top \end{bmatrix} = 0,$$

and

$$\begin{bmatrix} a_1 & & & & & & & & & & \\ a_2 & a_1 & \mathbf{0} & & & & & & & & \\ \vdots & \vdots & \ddots & & & & & & & & \\ a_n & a_{n-1} & \cdots & \cdots & a_1 & & & & & & \end{bmatrix} \begin{bmatrix} \beta_1^\top \\ \vdots \\ \beta_n^\top \end{bmatrix} + \begin{bmatrix} \bar{\alpha}_1^\top \\ \vdots \\ \bar{\alpha}_n^\top \end{bmatrix} = 0.$$

It follows immediately that $\beta = 0$ and $\bar{\alpha} = \alpha^\top \Delta_n(\mathcal{K} \otimes I_{n_u}) = 0$. Since the system (1) is reachable, we have $\text{rank}(\Delta_n) = n$ and so, from the definition of \mathcal{K} , $\Delta_n(\mathcal{K} \otimes I_{n_u})$ is clearly full row rank n . Consequently, we have also $\alpha = 0$ and hence, the rows of $[X_{f+1, N}^\top \quad U_{f+1, f, N}^\top]^\top$ are linearly independent. ■

For future reference, we shall also state the following proposition.

Proposition 2: Assume that the system (1) is reachable and observable and let $w(t)$ and $v(t)$ be identically null in (1). Then the following statements are equivalent.

1) $\text{rank} \left(\begin{bmatrix} X \\ U \end{bmatrix} \right) = n + f n_u.$

2) $\text{rank}(X \Pi_U^\perp) = n$, where

$$\Pi_U^\perp = I_N - U^\top (U U^\top)^{-1} U, \quad (16)$$

I_N being the identity matrix of order N .

3) $\text{rank} \left(\begin{bmatrix} Y \\ U \end{bmatrix} \right) = n + f n_u.$

4) $\text{rank}(Y \Pi_U^\perp) = n.$

Proof: 1) \iff 2): Notice that

$$\begin{bmatrix} X \\ U \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix}^\top = \begin{bmatrix} X X^\top & X U^\top \\ U X^\top & U U^\top \end{bmatrix}.$$

Then, by using the identity [12]

$$\begin{bmatrix} I_n & -X U^\top (U U^\top)^{-1} \\ 0 & I_{f n_u} \end{bmatrix} \begin{bmatrix} X X^\top & X U^\top \\ U X^\top & U U^\top \end{bmatrix} \times \begin{bmatrix} I_n & 0 \\ -(U U^\top)^{-1} U X^\top & I_{f n_u} \end{bmatrix} = \begin{bmatrix} X \Pi_U^\perp X^\top & 0 \\ 0 & U U^\top \end{bmatrix},$$

it is clear that

$$\begin{aligned} \text{rank} \left(\begin{bmatrix} X \\ U \end{bmatrix} \right) &= \text{rank} \left(\begin{bmatrix} X \Pi_U^\perp X^\top & 0 \\ 0 & U U^\top \end{bmatrix} \right) \\ &= f n_u + \text{rank}(X \Pi_U^\perp X^\top) \\ &= f n_u + \text{rank}(X \Pi_U^\perp), \end{aligned}$$

from which it can be concluded that

$$\text{rank} \left(\begin{bmatrix} X \\ U \end{bmatrix} \right) = n + f n_u \iff \text{rank}(X \Pi_U^\perp) = n.$$

1) \iff 3): The result follows directly from the relation

$$\begin{bmatrix} Y \\ U \end{bmatrix} = \underbrace{\begin{bmatrix} \Gamma_f & H_f \\ 0 & I_{f n_u} \end{bmatrix}} \begin{bmatrix} X \\ U \end{bmatrix} \quad (17)$$

since the underbraced matrix is full column rank when the system is observable, i.e., when $\text{rank}(\Gamma_f) = n$.

3) \iff 4): A proof of this statement is similar to the one derived for 1) \iff 2). ■

III. A NEW SUBSPACE IDENTIFICATION METHOD

We develop in this section a new method for the identification of linear state space models. In contrast to most of the existing subspace techniques, our method does not require any SVD. Therefore, it can be naturally and straightforwardly extended to recursive identification [8], [13], [9]. Moreover, it allows to set up very simply the state basis of the matrices to be identified.

To begin with, assume the order n of the system (1) to be available. Consider a known matrix $\Lambda_f \in \mathbb{R}^{n \times f n_y}$ that satisfies¹

$$\text{rank}(\Lambda_f \Gamma_f) = n. \quad (18)$$

If we multiply Eq. (5) on the left by Λ_f , we get

$$\Lambda_f Y = (\Lambda_f \Gamma_f) X + \Lambda_f H_f U + \Lambda_f E. \quad (19)$$

Since $T = \Lambda_f \Gamma_f \in \mathbb{R}^{n \times n}$ is nonsingular, one realizes a state coordinates basis change by setting $X \leftarrow \bar{X} = TX$. Then, the state sequence can be obtained in the new basis as

$$\bar{X} = \Lambda_f Y - \Lambda_f H_f U - \Lambda_f E. \quad (20)$$

As a consequence of this change of the coordinates system, the matrices (A, B, C, D) change into $(\bar{A}, \bar{B}, \bar{C}, \bar{D}) = (TAT^{-1}, TB, CT^{-1}, D)$ and Γ_f becomes $\bar{\Gamma}_f = \Gamma_f T^{-1}$ but H_f remains unchanged as it does not depend on the state basis.

For the sake of clarity, we shall first consider the noise-free case, i.e., the case where E in (5) is identically null. Then, we shall demonstrate the applicability of the method on noisy data.

A. Deterministic case

In the absence of noise in the data, the state equation (20) reduces to

$$\bar{X} = \Lambda_f Y - \Lambda_f H_f U. \quad (21)$$

It is obvious from this relation that the choice of Λ_f determines completely the state basis since it directly defines a linear combination of the I/O data that gives the state.

To obtain $\bar{\Gamma}_f$, we can report Eq. (21) in Eq. (5). This results immediately in

$$\begin{aligned} Y &= \bar{\Gamma}_f \bar{X} + H_f U \\ &= \bar{\Gamma}_f \Lambda_f Y + (I_{f n_y} - \bar{\Gamma}_f \Lambda_f) H_f U \\ &= \begin{bmatrix} \bar{\Gamma}_f & \Omega_f \end{bmatrix} \begin{bmatrix} \Lambda_f Y \\ U \end{bmatrix}, \end{aligned} \quad (22)$$

where $\Omega_f = (I_{f n_y} - \bar{\Gamma}_f \Lambda_f) H_f$. This shows that the subspace identification problem can be transformed into an ordinary Least Squares problem by eliminating the unknown state. As shown by Propositions 1 and 2, when the input process is

sufficiently exciting of order at least $n + f$, and the model (1) is minimal, it holds that

$$\text{rank} \left(\begin{bmatrix} Y \\ U \end{bmatrix} \right) = n + f n_u.$$

By following a similar procedure as in the proof of Proposition 2 (see Eq. (17)), one can easily establish that the following holds also

$$\text{rank} \left(\begin{bmatrix} \Lambda_f Y \\ U \end{bmatrix} \right) = n + f n_u \text{ and } \text{rank}(\Lambda_f Y \Pi_U^\perp) = n.$$

Thus, we can estimate directly the matrices $\bar{\Gamma}_f$ and Ω_f from Eq. (22) and then deduce the system matrices.

Remark 1: If $\Lambda_f \in \mathbb{R}^{n \times f n_y}$ obeys the rank property (18), then by denoting $\text{im}(\cdot)$ and $\text{ker}(\cdot)$ respectively the image and the kernel operators for matrices, we have:

- 1) $\text{im}(\Gamma_f) = \text{im}(\bar{\Gamma}_f) = \text{ker}(I_{f n_y} - \bar{\Gamma}_f \Lambda_f)$.
- 2) $\text{rank}(I_{f n_y} - \bar{\Gamma}_f \Lambda_f) = f n_y - n$.
- 3) The matrix $I_{f n_y} - \bar{\Gamma}_f \Lambda_f$ is a projection matrix that projects the ambient vector space $\mathbb{R}^{f n_y}$ onto $\text{im}(I_{f n_y} - \bar{\Gamma}_f \Lambda_f) = \text{ker}(\Lambda_f)$ along $\text{im}(\Gamma_f)$. In this way, Eq. (22) can be viewed as a projection of Eq. (5) obtained by applying $(I_{f n_y} - \bar{\Gamma}_f \Lambda_f)$ to it.

Thanks to this remark, we know that $I_{f n_y} - \bar{\Gamma}_f \Lambda_f$ is not full column rank and so H_f cannot be retrieved directly from the estimate $\hat{\Omega}_f$ of Ω_f in (22). However, by appropriately exploiting the structure of H_f , it is still possible to recover the matrices B and D from $\hat{\Omega}_f$ and $\hat{\bar{\Gamma}}_f$ similarly as in [1, p. 54].

Another way to proceed is to remove the term Ω_f from (22) by multiplying this latter equation on the right by Π_U^\perp before computing $\bar{\Gamma}_f$. We then get

$$\begin{aligned} \bar{\Gamma}_f &= Y \Pi_U^\perp (\Lambda_f Y \Pi_U^\perp)^\dagger, \\ &= \Sigma_{yu} \Lambda_f^\top (\Lambda_f \Sigma_{yu} \Lambda_f^\top)^{-1}, \end{aligned} \quad (23)$$

where $\Sigma_{yu} = \frac{1}{N} Y \Pi_U^\perp Y^\top$. Here, Σ_{yu} is a correlation matrix while Π_U^\perp defined in (16) is a matrix of projection onto the orthogonal complement of the rows space of U . The symbol \dagger refers here to the Moore-Penrose inverse. In practice, the orthogonal projection Π_U^\perp may be implemented by resorting to methods that are known to be numerically robust such as the QR factorization [2]. Once $\bar{\Gamma}_f$ is available, an extraction of the matrices \bar{A} and \bar{C} can be simply achieved by exploiting the so-called A-invariance property of $\bar{\Gamma}_f$ [1].

Now, given the matrices \bar{A} and \bar{C} , it remains to determine the matrices \bar{B} and \bar{D} . This can be achieved by solving a linear regression problem. For more details on the concrete procedure, we refer to, for example the papers [2], [8].

B. Stochastic case

Consider now the more realistic case where the data are subject to noise. Then, the combination of Eq. (5) and Eq. (20) gives

$$Y = \bar{\Gamma}_f \Lambda_f Y + (I_{f n_y} - \bar{\Gamma}_f \Lambda_f) H_f U + (I_{f n_y} - \bar{\Gamma}_f \Lambda_f) E.$$

¹Discussion on the concrete determination of such a matrix Λ_f is deferred to Subsection III-C.

As previously, the term $\Omega_f = (I_{fn_y} - \bar{\Gamma}_f \Lambda_f) H_f$ is removed by post-multiplying the equation by Π_U^\perp . We get

$$Y \Pi_U^\perp = \bar{\Gamma}_f \Lambda_f Y \Pi_U^\perp + (I_{fn_y} - \bar{\Gamma}_f \Lambda_f) E \Pi_U^\perp. \quad (24)$$

To deal with the noisy data, we adopt the well known method of instrumental variable (IV) [4]. The basic idea of this method is to choose a matrix of instruments $Z \in \mathbb{R}^{n_z \times N}$, $n_z \geq n$ (formed for example with past input and output [3]) to zero out the effect of the noise while preserving the information conveyed by the data. By multiplying Eq. (24) on the right by Z^\top and dividing by the number of samples N , we obtain

$$\begin{aligned} \frac{1}{N} Y \Pi_U^\perp Z^\top &= \bar{\Gamma}_f \frac{1}{N} \Lambda_f Y \Pi_U^\perp Z^\top + (I_{fn_y} - \bar{\Gamma}_f \Lambda_f) \frac{1}{N} E Z^\top \\ &\quad - (I_{fn_y} - \bar{\Gamma}_f \Lambda_f) \frac{1}{N} E U^\top \left(\frac{1}{N} U U^\top \right)^{-1} \frac{1}{N} U Z^\top. \end{aligned} \quad (25)$$

Under the assumption that the sequence $\{u(t)\}$ is ergodic and independent of $\{w(t)\}$ and $\{v(t)\}$, the last term of (25) vanishes when $N \rightarrow \infty$. Thus, Eq. (25) reduces to

$$\begin{aligned} \lim_{N \rightarrow \infty} \left(\frac{1}{N} Y \Pi_U^\perp Z^\top \right) &= \bar{\Gamma}_f \lim_{N \rightarrow \infty} \left(\frac{1}{N} \Lambda_f Y \Pi_U^\perp Z^\top \right) \\ &\quad + (I_{fn_y} - \bar{\Gamma}_f \Lambda_f) \lim_{N \rightarrow \infty} \left(\frac{1}{N} E Z^\top \right). \end{aligned} \quad (26)$$

In view of the previous equation, we require the instruments matrix Z to verify the following two conditions

$$\begin{cases} \lim_{N \rightarrow \infty} \left(\frac{1}{N} E Z^\top \right) = 0, \\ \text{rank} \left(\lim_{N \rightarrow \infty} \frac{1}{N} \Lambda_f Y \Pi_U^\perp Z^\top \right) = n. \end{cases} \quad (27)$$

When Z fulfills the requirements (27), the following asymptotic relation holds

$$\lim_{N \rightarrow \infty} \frac{1}{N} Y \Pi_U^\perp Z^\top = \bar{\Gamma}_f \lim_{N \rightarrow \infty} \left(\frac{1}{N} \Lambda_f Y \Pi_U^\perp Z^\top \right). \quad (28)$$

From this equation, we can estimate $\bar{\Gamma}_f$, then compute as previously \bar{A} and \bar{C} and subsequently obtain \bar{B} and \bar{D} by linear regression as indicated in Subsection III-A. Turning now to the question of how to choose the instruments matrix Z , we will follow [14] where the concatenation of the the past inputs and past outputs,

$$Z = \begin{bmatrix} U_{1,f,N}^\top & Y_{1,f,N}^\top \end{bmatrix}^\top \quad (29)$$

is shown to be a valid set of instruments.

C. Setting the matrix Λ_f

The method described above for the identification of state space models relies on the possibility to find a matrix $\Lambda_f \in \mathbb{R}^{n \times fn_y}$ that verifies the rank condition (18). This naturally raises the question of whether such a matrix always exists and more importantly, how to determine it while Γ_f is unknown. The question of the existence is readily answered under the assumption that the system is observable since then, $\Lambda_f = \Gamma_f^\top$ (for example) verifies (18). But, as we do not know Γ_f , we cannot set Λ_f to be equal to Γ_f^\top . However, as stated by Propositions 1 and 2, if $\{u(t)\}$ is SE($n + f$),

then $\text{rank}(\Lambda_f \Gamma_f) = n$ if and only if $\text{rank}(\Lambda_f Y \Pi_U^\perp) = n$. In view of this remark, Λ_f can be selected just such that $\text{rank}(\Lambda_f Y \Pi_U^\perp) = n$ because contrarily to Γ_f , $Y \Pi_U^\perp$ is known. But this solution may require an SVD factorization.

Therefore, an alternative solution is to generate Λ_f at random as suggested by the following proposition.

Proposition 3: Assume that the system (1) is observable. Let the above matrix Λ_f be drawn randomly from a uniform distribution. Then, it holds with probability one that

$$\text{rank}(\Lambda_f \Gamma_f) = n.$$

Proof: Let $\lambda = \text{vec}(\Lambda_f) \in \mathbb{R}^{fn_y}$, where $\text{vec}(\cdot)$ is the vectorization operator. Denote $\Lambda_f = \Lambda_f(\lambda)$, that is, let us view Λ_f as a function of λ and consider the set $\mathcal{S} = \left\{ \lambda \in \mathbb{R}^{fn_y} / P(\lambda) = \det(\Lambda_f(\lambda) \Gamma_f) = 0 \right\}$ of all λ such that $\text{rank}(\Lambda_f(\lambda) \Gamma_f) < n$. Consider the probability measure Pr associated to the uniform distribution and defined on a σ -algebra \mathcal{R} (containing \mathcal{S}) over \mathbb{R}^{fn_y} . Clearly, for a given Γ_f , the polynomial P is not identically null since we have for example $P(\text{vec}(\Gamma_f^\top)) \neq 0$. Then, the hypersurface \mathcal{S} is a subset of the probability space $(\mathbb{R}^{fn_y}, \mathcal{R}, \text{Pr})$ that is of dimension strictly less than fn_y . From the measure theory [15], a subset such as \mathcal{S} is known to be a null set. Thus, the property $\text{rank}(\Lambda_f(\lambda) \Gamma_f) = n$ holds almost everywhere.² ■

Proposition 3 states in other words that if we draw randomly the matrix Λ_f from a uniform distribution for example, the rank property (18) holds with probability one. The system matrices are then correctly estimated by the above method with probability one.

Remark 2 (On the estimation of the order): In case the order n of the system (1) is unknown, it would be also interesting to estimate it without having to resort to the SVD. To this purpose, let Λ_f be selected in $\mathbb{R}^{fn_y \times fn_y}$ such that³ for any $r \leq n$, it holds that $\text{rank}(\Lambda_f^{1:r} \Gamma_f) = r$, where $\Lambda_f^{1:r} = \Lambda_f(1 : r, :)$. Then from the embedded data equation $Y = \Gamma_f X + H_f U$, one can write

$$\begin{bmatrix} \Lambda_f^1 \\ \Lambda_f^2 \end{bmatrix} Y \Pi_U^\perp = \begin{bmatrix} \Lambda_f^1 \\ \Lambda_f^2 \end{bmatrix} \Gamma_f X \Pi_U^\perp,$$

where $\Lambda_f^1 = \Lambda_f(1 : n, :)$ and $\Lambda_f^2 = \Lambda_f(n + 1 : fn_y, :)$. Taking the square of the first term of this equation yields

$$\Sigma = \begin{bmatrix} \Lambda_f^1 \Sigma_{yu} (\Lambda_f^1)^\top & \Lambda_f^1 \Sigma_{yu} (\Lambda_f^2)^\top \\ \Lambda_f^2 \Sigma_{yu} (\Lambda_f^1)^\top & \Lambda_f^2 \Sigma_{yu} (\Lambda_f^2)^\top \end{bmatrix},$$

with $\Sigma_{yu} = Y \Pi_U^\perp Y^\top$. The order is then $n = \max \{ r : \text{rank}(\Sigma(1 : r, 1 : r)) = r \}$ and can be determined by following a similar procedure as in [7].

IV. SIMULATION EXAMPLE

We shall in this section, evaluate the performance of our method on a numerical example. To this purpose, we consider a linear system of order three with two inputs and two outputs. The system matrices are given by

²(18) is satisfied everywhere in \mathbb{R}^{fn_y} except on a set of measure null.

³In fact, as in Proposition 3, this is the case almost surely when Λ_f is generated randomly.

$$A = \begin{bmatrix} -0.0398 & -0.4909 & -0.6115 \\ 0.4784 & 0.4808 & -0.4169 \\ -0.6214 & 0.4020 & -0.2488 \end{bmatrix}, B = \begin{bmatrix} 0.7433 & 0 \\ 0 & -1.3193 \\ -0.8510 & -0.0181 \end{bmatrix},$$

$$C = \begin{bmatrix} -0.0028 & 2.3968 & 0 \\ -0.0857 & 0.0264 & -0.6553 \end{bmatrix}, D = \begin{bmatrix} -1.7384 & 0.0425 \\ 0.7127 & 0 \end{bmatrix}.$$

To identify this system, the excitation input is chosen as a zero-mean white Gaussian noise with unit variance. The state noise $w(t)$ and output noise $v(t)$ are both set as white noises in a reasonable proportion of 25 dB respectively of the state and the output. The user-defined parameter f is set to be 5. We then carry out a Monte-Carlo simulation with 1000 different realizations of noise and input. Each of these simulations generates 1000 I/O data samples. With this setting, we perform two sets of simulations: one with an instrumental variable in the form of (29) and the second without any instrumental variable. The results can be judged through the following graphs.

In Figure 1, we present the poles of the actual system and the poles of the estimated model over all the 1000 simulations. One can notice that when no instrumental variable is introduced, the estimator seems to be biased when the data are corrupted by noise.

To further show the performance of our method in a statistical sense, we represent in Figure 2 the histograms of the relative errors $\|M - \hat{M}\|/\|M\|$ between the first ten Markov parameters (which are invariant under state basis change) of the true system together with the Markov parameters of the estimated model.

All the simulation results presented clearly illustrate the potential of our method as a serious alternative to the classical subspace methods for the identification of state space models.

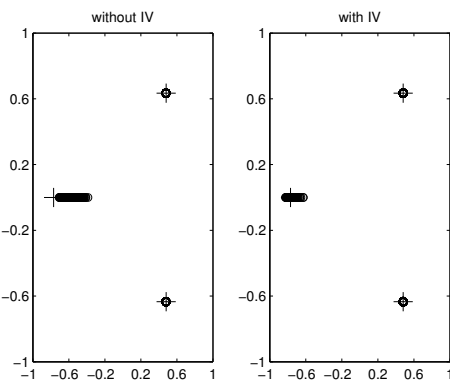


Fig. 1. Poles of the system superposed to that of its estimate. The big crosses refer here to the poles of the true system.

V. CONCLUSION

We have proposed a new method for the identification of MIMO state space models. Contrarily to most of the existing subspace identification schemes, our method is SVD-free and therefore, lends itself to a straightforward extension to recursive identification of multivariable systems. Moreover the state space basis of the system matrices to be estimated

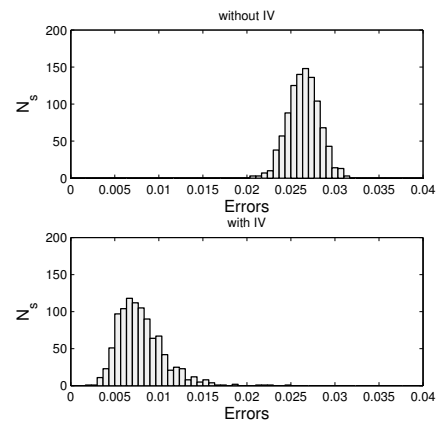


Fig. 2. Histograms of the relative errors between the Markov parameters of the system on the one hand and that of the model estimated by our method on the other hand. N_s stands for number of simulations.

is fixed in advance through a conversion of the subspace problem into a least squares one. This is an important feature for example in the framework of multi-modal systems identification since in such situation, all the different submodels need to be identified in the same state coordinates basis. In future work we will carry out a comparison study of our algorithms with the standard subspace identification methods such as MOESP and N4SID in terms of complexity and performance. We will also consider to extend our method to the identification of multivariable switched linear systems described by state space models.

REFERENCES

- [1] P. Van Overschee and B. De Moor, *Subspace identification for linear systems. theory, implementation, applications*. Kluwer Academic Publishers, 1996.
- [2] T. Katayama, *Subspace methods for system identification*. Springer-Verlag, 2005.
- [3] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, 2007.
- [4] L. Ljung, *System Identification. Theory for the user*. 2nd ed., PTR Prentice Hall, Upper Saddle River, USA., 1999.
- [5] V. Verdult, L. Ljung, and M. Verhaegen, "Identification of composite local linear state-space models using a projected gradient search," *International Journal of Control*, vol. 75, pp. 1385–1398, 2002.
- [6] V. Verdult and M. Verhaegen, "Subspace identification of piecewise linear systems," in *Conference on Decision and Control*, 2004.
- [7] L. Bako, G. Mercère, and S. Lecoeuche, "Online subspace identification of switching systems with possibly varying orders," in *European Control Conference*, 2007.
- [8] M. Lovera, T. Gustafsson, and M. Verhaegen, "Recursive subspace identification of linear and non-linear wiener state-space models," *Automatica*, vol. 36, pp. 1639–1650, 2000.
- [9] H. Oku and H. Kimura, "Recursive 4SID algorithms using gradient type subspace tracking," *Automatica*, vol. 38, pp. 1035–1043, 2002.
- [10] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, pp. 67–94, 1996.
- [11] S. J. Qin, "An overview of subspace identification," *Computers and Chemical Engineering*, vol. 30, pp. 1502–1513, 2006.
- [12] T. Kailath and A. H. Sayed, *Fast Reliable Algorithms for Matrices with Structure*. SIAM, Philadelphia, PA, 1999.
- [13] G. Mercère, L. Bako, and S. Lecoeuche, "Propagator-based methods for recursive subspace model identification," *Signal Processing*, vol. 2008, pp. 468–491, 88.
- [14] C. T. Chou and M. Verhaegen, "Subspace algorithms for the identification multivariable dynamic error-in-variables models," *Automatica*, vol. 33, pp. 1857–1869, 1997.
- [15] P. R. Halmos, *Measure Theory*. New York: Springer-Verlag, 1974.