

# Identification of Switched Autoregressive Systems from Large Noisy Data Sets

Sarah Hojjatinia<sup>1</sup>, Constantino M. Lagoa<sup>2</sup>, and Fabrizio Dabbene<sup>3</sup>

**Abstract**—The paper introduces a novel methodology for the identification of coefficients of switched autoregressive linear models. We consider the case when the system's outputs are contaminated by possibly large values of measurement noise. It is assumed that only partial information on the probability distribution of the noise is available. Given input-output data, we aim at identifying switched system coefficients and parameters of the distribution of the noise which are compatible with the collected data. System dynamics are estimated through expected values computation and by exploiting the strong law of large numbers. We demonstrate the efficiency of the proposed approach with several academic examples. The method is shown to be extremely effective in the situations where a large number of measurements is available; cases in which previous approaches based on polynomial or mixed-integer optimization cannot be applied due to very large computational burden.

## I. INTRODUCTION

The interest in the study of hybrid systems has been persistently growing in the last years, due to their capability of describing real-world processes in which continuous and discrete time dynamics coexist and interact. Besides classical automotive and chemical processes, emerging applications include computer vision, biological systems, and communication networks.

Moreover, hybrid systems can be used to efficiently approximate nonlinear dynamics, with broad application, ranging from civil structures to robotics and systems biology, that entail extracting information from high volume data streams [13], [18]. In the case of high dimensional data, nonlinear order reduction or low dimensional sparse representations techniques [9], [5], [16], are very effective in handling static data, but most do not exploit dynamical information of the data.

In the literature, several results have been obtained for the analysis and control of hybrid systems, formally characterizing important properties such as stability or reachability, and proposing different control designs [10]. In parallel, researchers rapidly realized that first-principle models may be hard to derive especially with the increase of diverse application fields. This sparked interest on the problem of

identifying hybrid (switched) models starting from experimental data; see for instance the tutorial paper [14] and the survey [3].

It should be immediately pointed out that this identification problem is not a simple one, since the simultaneous presence of continuous and discrete state variables gives it a combinatorial nature. The situation becomes further complicated in the presence of unknown-but-bounded noise. In this case the problem is in general NP-hard. Several approaches have been proposed to address this difficulty, see e.g. [8]. The paper [15] reformulates the problem as a mixed-integer program. These techniques proved to be very effective in situations involving relatively small noise levels or moderate dimensions, but they do not appear to scale well, and their performance deteriorates as the noise level or problem size increase.

Of particular interest are recent approaches based on convex optimization: in [1] some relaxation based on sparsity are proposed, while [12] develops a moment based approach to identify the switched autoregressive exogenous system, and [6] adapts it toward Markovian jump systems identification. These methods are surely more robust, and represent the choice of reference for medium-size problems and medium values of noise, and have found applications in several contexts, ranging from segmentation problems arising in computer vision to biomedical systems.

However, the methods still rely on the solution of rather large optimization problems. Even if the convex nature of these problems allows to limit the complexity growth, there are several situations for which their application becomes critical. For instance, identification problems cases that involve quite high noise levels and/or large number of measurements.

An enlightening example, which serves as a practical motivation for our developments, arises in healthcare applications: the availability of *activity tracking devices* allows to gather a large amount of information of the physical activity of an individual. Physical activity is a dynamic behavior, which in principle can be modeled as a dynamical system [7]. Moreover, its characteristics may significantly change depending on the time of the day, position, etc. This motivated the approach of modeling it as a switching system [2].

In this paper, we focus on cases involving a very large number of sample points, possibly affected by large levels of noise. In this situation, polynomial/moments based approaches become ineffective, and different methodologies need to be devised. The approach we propose builds upon

<sup>1</sup>Sarah Hojjatinia is with the School of Electrical Engineering and Computer Science, The Pennsylvania State University, University Park, PA, USA, [szh199@psu.edu](mailto:szh199@psu.edu)

<sup>2</sup>Constantino M. Lagoa is with the School of Electrical Engineering and Computer Science, The Pennsylvania State University, University Park, PA, USA, [lagoa@psu.edu](mailto:lagoa@psu.edu)

<sup>3</sup>Fabrizio Dabbene is with CNR-IEIT, Politecnico di Torino, 10129 Torino, Italy, [fabrizio.dabbene@ieit.cnr.it](mailto:fabrizio.dabbene@ieit.cnr.it)

This work was partially supported by National Institutes of Health (NIH) Grant R01 HL142732, National Science Foundation (NSF) Grant #1808266 and the International Bilateral Joint CNR-JST Lab COOPS.

the same premises as [12]: the starting point is the algebraic procedure due to Ma and Vidal [11], where it has been shown for noiseless processes, it is possible to identify the different subsystems in a switching system by recurring to a Generalized Principal Component Analysis (GPCA). In particular, we infer the parameters of each subsystem from the null space of a matrix  $V_n(r)$  constructed from the input-output data  $r$  via a nonlinear embedding (the Veronese map).

The approach was extended to the cases in presence of process noise in [12], showing how the entries of this matrix depend polynomially on the unknown noise terms. Then, the problem was formulated in an unknown-but-bounded setting, looking for an admissible noise sequence rendering the matrix  $V_n(r)$  rank deficient. This problem was then relaxed using polynomial optimization methods.

In this work, we follow the same line of reasoning, but then take a somewhat different route. First, we consider random noise, and we assume that *some* information on the noise is available. Then, instead of relaxing the problem, we exploit the availability of a large number of measurements to make recourse to law-of-large-numbers type of reasoning. This allows us to devise an algorithm characterized by an extremely low complexity in terms of required operations. The ensuing optimization problem involves only the computation of the singular vector associated with the minimum singular value of a matrix that can be efficiently computed and whose size does not depend on the number of measurements.

#### A. Paper Organization

The paper is structured as follows: after this introduction, there is a brief notation section. Section II includes the problem statement. In Section III, algebraic reformulation of switched autoregressive (SAR) linear system identification problem for noiseless data is reviewed. The problem of identifying SAR system in the presence of noise is surveyed in Section IV. In Section V, the algorithm for estimating unknown noise parameters is described. Numerical results are shown in Section VI. Finally, Section VII concludes the paper highlighting some possible future research directions.

#### B. Notation

Given a scalar random variable  $X$ , we denote by  $m_d$  its  $d^{th}$  moment, which may be computed according to the following integral

$$m_d = E[x^d] = \int_{-\infty}^{\infty} x^d f(x) dx \quad (1)$$

where  $E[\cdot]$  refers to expectation, and  $f(x)$  is the probability density function of  $X$ . Additionally, the variance of  $X$  is indicated by  $s^2$ . For instance, if  $X$  has a normal distribution with zero mean and variance  $s^2$ , i.e.  $f(x) = \frac{1}{s\sqrt{2\pi}} e^{-x^2/2s^2}$ , its moments are given by

$$m_d = E[x^d] = \begin{cases} 0 & \text{if } d \text{ is odd} \\ s^d (d-1)!! & \text{if } d \text{ is even} \end{cases} \quad (2)$$

where  $!!$  denotes double factorial ( $n!!$  is the product of all numbers from  $n$  to 1 that have the same parity as  $n$ ).

## II. PROBLEM STATEMENT

In this section, a complete description of the problem is addressed. In addition, the required assumptions are defined to solve the problem.

#### A. System Model

We consider SAR systems of the form

$$x_k = \sum_{j=1}^{n_a} a_{j\sigma(k)} x_{k-j} + \sum_{j=1}^{n_c} c_{j\sigma(k)} u_{k-j} \quad (3)$$

where  $x_k \in \mathbb{R}$  is the output at time  $k$  and  $u_k \in \mathbb{R}$  is input at time  $k$ . The variable  $\sigma(k) \in \{1, \dots, n\}$  denotes the sub-system active at time  $k$ , where  $n$  is the total number of sub-systems. Furthermore,  $a_{j\sigma(k)}$  and  $c_{j\sigma(k)}$  denote unknown coefficients corresponding to mode  $\sigma(k)$ . Time  $k$  takes values over the non-negative integers.

In practice, output is always contaminated by noise; i.e. we assume that we observe

$$y_k = x_k + \eta_k \quad (4)$$

where  $\eta_k$ , denotes measurement noise.

The following assumptions are made on the system model and noise.

*Assumption 1:* Throughout this paper it is assumed that:

- Upper bounds on  $n_a$  and  $n_c$  are available.
- Upper bound on the number of subsystems  $n$  is available.
- Noise  $\eta_k$  at time  $k$  is independent from  $\eta_l$  for  $k \neq l$ , and identically distributed with probability density  $f(\eta|\theta)$ ; where  $\theta$  is a (low dimensional) vector of unknown parameters.
- Input sequence  $u_k$  applied to the system is known and bounded.
- There exists a finite constant  $L$  so that  $|x_k| \leq L$  for all positive integers  $k$ .

#### B. Problem Definition

The main objective of this paper is to develop algorithms that are able to identify the coefficients of a SAR model from noisy observations. More precisely, we aim at solving the following problem:

*Problem 1:* Given Assumption 1, an input sequence  $u_k$ ,  $k = -n_c + 1, \dots, N - 1$  and noisy output measurements  $y_k$ ,  $k = -n_a + 1, \dots, N$ , determine coefficients of the SAR model  $a_{i,j}$ ,  $i = 1, 2, \dots, n_a$ ,  $j = 1, 2, \dots, n$ ,  $c_{i,j}$ ,  $i = 1, 2, \dots, n_c$ ,  $j = 1, 2, \dots, n$ , and the noise distribution parameters  $\theta$ .

## III. NOISELESS CASE: A REVIEW

As a motivation for the approach presented in this paper, we review and slightly reformulate earlier results on an algebraic reformulation of the SAR identification problem for the case where no noise is present. We refer the reader to [19] for details on this algebraic approach to switched system identification.

### A. Hybrid Decoupling Constraint

We start by noting that equation (3) is equivalent to

$$b_{\sigma(k)}^T r_k = 0 \quad (5)$$

where we introduced the (known) regressor at time  $k$

$$r_k = [x_k, x_{k-1}, \dots, x_{k-n_a}, u_{k-1}, \dots, u_{k-n_c}]^T$$

and the vector of unknown coefficients at time  $k$

$$b_{\sigma(k)} = [-1, a_{1\sigma(k)}, \dots, a_{n_a\sigma(k)}, c_{1\sigma(k)}, \dots, c_{n_c\sigma(k)}]^T.$$

Hence, independently of which of the  $n$  submodels is active at time  $k$ , we have

$$P_n(r_k) = \prod_{i=1}^n b_i^T r_k = c_n^T \nu_n(r_k) = 0, \quad (6)$$

where the vector of parameters corresponding to the  $i$ -th submodel is denoted by  $b_i \in \mathbb{R}^{n_a+n_c+1}$ , and  $\nu_n()$  is Veronese map of degree  $n$  [4]

$$\nu_n([x_1, \dots, x_s]^T) = [\dots, x_1^{n_1} x_2^{n_2} \dots x_s^{n_s}, \dots]^T$$

which contains all monomials of order  $n$  in lexicographical order, and  $c_n$  is a vector whose entries are polynomial functions of unknown parameters  $b_i$  (see [20] for explicit definition). The Veronese map above is also known as polynomial embedding in machine learning [20].

Equation (6) holds for all  $k$ , and these equalities can be expressed in matrix form

$$V_n(r) c_n = [\nu_n(r_1)^T, \dots, \nu_n(r_N)^T]^T c_n = 0 \quad (7)$$

where  $r$ , without the subscript, denotes the set of all regressor vectors. Clearly, we are able to identify  $c_n$  (and hence the system's parameters) if and only if  $V_n(r)$  is rank deficient. In that case, the vector  $c_n$  can be found by computing the nullspace of  $V_n$ .

### B. A Reformulation of the Hybrid Decoupling Constraint

Note that the number of rows of the Veronese matrix  $V_n$  is equal to the number of measurements available for the regressor; i.e., in the notation of our paper, the number of rows is  $N$ . Therefore, a reformulation of the results in the previous section is needed to be able to address the problem of identification from very large data sets.

As mentioned in the previous section, in the absence of noise, the SAR system identification is equivalent to finding a vector  $c_n$  satisfying

$$c_n^T \nu_n(r_k) = 0 \text{ for all } k = 1, 2, \dots, N.$$

This is in turn equivalent to finding  $c_n$  so that

$$\frac{1}{N} \sum_{k=1}^N c_n^T \nu_n(r_k) \nu_n^T(r_k) c_n = 0$$

As a result, for the noiseless case, identifying the coefficients of the sub-models is equivalent to finding the singular vector  $c_n$  associated with the minimum singular value of the matrix

$$\overline{\mathcal{M}}_N = \frac{1}{N} \sum_{k=1}^N \nu_n(r_k) \nu_n^T(r_k) \doteq \frac{1}{N} \sum_{k=1}^N M_k \quad (8)$$

Note that, by using this equivalent condition, we only need to consider matrices of size  $\binom{n+n_a+n_c}{n}$ . In other words, the size of this matrix *does not depend* on the number of measurements. This is especially important when considering very large data sets.

## IV. SAR SYSTEM IDENTIFICATION IN THE PRESENCE OF NOISE

Now, we address the case where the measurements of output of the switched autoregressive system are corrupted by noise. As a first step, we consider the case where the distribution of the noise is known, so its moments  $m_d$  are available.

As seen in the previous section, identifying the parameters of the SAR model is equivalent to finding a vector in the null space of the matrix  $\overline{\mathcal{M}}_N$ . Under mild conditions, the null space of the matrix above has dimension one if and only if the data is compatible with the assumed model. However, if noise is present,  $x_k$  is not known; therefore, this matrix cannot be computed. In this section, we use available information on the statistics of the noise to compute approximations of the matrix  $\overline{\mathcal{M}}_N$ , consequently approximations of vectors in its null space.

### A. On the Powers of $x_k$

Since we do not have access to the values of the output  $x_k$  to estimate the values of the quantities in equation (8), we need to relate the powers of  $x_k$  to the measurements and available information of the noise.

Note that  $x_k$  is a (unknown) deterministic quantity. Therefore

$$x_k^h = E[x_k^h] \quad (9)$$

Since  $x_k = y_k - \eta_k$  we have

$$x_k^h = E[x_k^h] = E[(y_k - \eta_k)^h] \quad \forall k = 1, 2, \dots, N \quad (10)$$

Assume, for simplicity the distribution of the noise is symmetric with respect to the origin. As a result, all odd moments are zero (in particular, the noise is zero mean, i.e.  $m_1 = 0$ ). This assumption is made to simplify the calculations below and the approach can be immediately extended to the non-symmetric case.

We concentrate on computing the expected value of powers of  $x_k$  recursively and in a closed form. First, we give an example of how to compute the expected value of powers of  $x_k$  for powers  $h = 1, 2$ . For  $h = 1$ , we have

$$x_k = E[x_k] = E[y_k - \eta_k] = E[y_k] - m_1 = E[y_k] \quad (11)$$

while, for  $h = 2$ , we can write

$$x_k^2 = E[x_k^2] = E[(y_k - \eta_k)^2] = E[y_k^2] - 2E[y_k \eta_k] + E[\eta_k^2]. \quad (12)$$

Note that  $E[y_k^2]$  can be estimated from collected data, and  $E[\eta_k^2]$  is equal to second moment of noise ( $m_2$ ), which is assumed to be known. To estimate the value of  $E[y_k \eta_k]$ , consider the following

$$E[y_k \eta_k] = E[(x_k + \eta_k) \eta_k] = E[x_k \eta_k] + E[\eta_k^2]. \quad (13)$$

The quantities  $x_k$  and  $\eta_k$  are mutually independent and, therefore,  $E[x_k \eta_k] = E[x_k]E[\eta_k]$ , with  $E[\eta_k] = m_1 = 0$ . As a consequence, we have

$$E[y_k \eta_k] = E[\eta_k^2] \quad (14)$$

and finally the value of equation (12) is

$$\begin{aligned} E[x_k^2] &= E[y_k^2] - 2E[\eta_k^2] + E[\eta_k^2] = E[y_k^2] - E[\eta_k^2] \\ &= E[y_k^2] - m_2 \end{aligned} \quad (15)$$

The reasoning above can be generalized to any power of  $x_k$ . More precisely, we have the following result whose proof is an immediate consequence of the derivations so far.

**Lemma 1:** The expected value of the powers of  $x_k$  satisfies

$$\begin{aligned} E[x_k^h] &= E[(y_k - \eta_k)^h] = E[y_k^h] - \sum_{d=1}^h \binom{h}{d} E[x_k^{h-d}] E[\eta_k^d] \\ &= E[y_k^h] - \sum_{d=1}^h \binom{h}{d} E[x_k^{h-d}] m_d \\ \forall k &= 1, 2, \dots, N. \end{aligned} \quad (16)$$

### B. On the Structure of $M_k$

We derive some of the properties of the matrices  $M_k$  now. An immediate consequence of the results in the previous section is the following.

**Lemma 2:** Assume that the noise distribution, some parameters of the noise, and the input signal are given and fixed. Let  $\text{mon}_n(\cdot)$  denote a function that returns a vector with all monomials up to order  $n$  of its argument. Then there exists an affine function  $M(\cdot)$  so that

$$\begin{aligned} M_k &= E\{M[\text{mon}_n(y_k, \dots, y_{k-n_a})]\} \\ &= M\{E[\text{mon}_n(y_k, \dots, y_{k-n_a})]\}. \end{aligned}$$

It should be noted that the random variables  $y_k$  and  $y_l$  are mutually independent for  $k \neq l$ , so the function above can further be represented as a multilinear function of the moments of  $y_k$ .

### C. An Example of Construction of $M_k$

To better illustrate the approach used in this paper, we provide an example of how to construct the matrix  $M_k$  required for identification. To this end, consider the problem of identifying a SAR system with  $n = 2$  subsystems of the form

$$\begin{aligned} \text{subsystem 1 : } & x_k = a_1 x_{k-1} + b_1 u_{k-1} \\ \text{subsystem 2 : } & x_k = a_2 x_{k-1} + b_2 u_{k-1} \end{aligned} \quad (17)$$

from noisy measurements

$$y_k = x_k + \eta_k \quad (18)$$

where  $\eta_k$  has a symmetric distribution. We can rewrite the system as in equation (6). In particular, the vector  $c_2$  as a function of the parameters of the subsystems, assumes the form

$$c_2 = [1, -(a_1 + a_2), -(b_1 + b_2), a_1 a_2, a_1 b_2 + b_1 a_2, b_1 b_2]^T.$$

The regressor vector  $r_k$  at time  $k$

$$r_k = [x_k \quad x_{k-1} \quad u_{k-1}]^T$$

gives rise to the following Veronese vector

$$\nu_n(r_k) = r_k \otimes r_k = \begin{bmatrix} x_k^2 \\ x_k x_{k-1} \\ x_k u_{k-1} \\ x_{k-1}^2 \\ x_{k-1} u_{k-1} \\ u_{k-1}^2 \end{bmatrix} \quad (19)$$

whose size is  $l \times 1$ , with  $l = \binom{n+n_a+n_c}{n} = \binom{2+1+1}{2} = 6$ . From  $r_k$  and  $\nu_n(r_k)$ , we can compute matrix  $M_k$ , which is given in Figure 1. Then, as we have the values of noisy output  $y_k$ , we compute expected value of powers of  $x_k$  in terms of expected value of powers of  $y_k$  and moments of measurement noise. Following the results of Lemma 1, we obtain the second matrix in Figure 1. For system of equation (17),  $M_k$  is given by the two expression in Figure 1.

### D. Identification Algorithm

As mentioned before, to identify the parameters of the SAR system, we need to be able to estimate the matrix  $\overline{M}_N$  in equation (8). It turns out that it can be done by using the available noisy measurements. More precisely, we have the following result.

**Theorem 1:** Let  $M(\cdot)$  and  $\text{mon}_n(\cdot)$  be the functions defined in Lemma 2. Define

$$\widehat{\overline{M}}_N \doteq \frac{1}{N} \sum_{k=1}^N M[\text{mon}_n(y_k, \dots, y_{k-n_a})]$$

Then, as  $N \rightarrow \infty$ ,

$$\widehat{\overline{M}}_N - \overline{M}_N \rightarrow 0 \text{ a.s.}$$

**Sketch of proof:** See Appendix.

As a result, the empirical average computed using the noisy measurements (where expected values of monomials are replaced by the measured monomial values) converges to the desired matrix in equation (8). Therefore we propose the following algorithm for identification of a SAR system.

**Algorithm 1:** Let  $n_a, n_c, n$  and some parameters of the noise be given.

Step 1. Compute matrix

$$\widehat{\overline{M}}_N = \frac{1}{N} \sum_{k=1}^N M[\text{mon}_n(y_k, \dots, y_{k-n_a})]$$

$$M_k = \nu_n(r_k) \nu_n^T(r_k) = \begin{pmatrix} x_k^4 & x_k^3 x_{k-1} & x_k^3 u_{k-1} & x_k^2 x_{k-1}^2 & x_k^2 x_{k-1} u_{k-1} & x_k^2 u_{k-1}^2 \\ * & x_k^2 x_{k-1}^2 & x_k^2 x_{k-1} u_{k-1} & x_k x_{k-1}^3 & x_k x_{k-1}^2 u_{k-1} & x_k x_{k-1} u_{k-1}^2 \\ * & * & x_k^2 u_{k-1}^2 & x_k x_{k-1}^2 u_{k-1} & x_k x_{k-1} u_{k-1}^2 & x_k u_{k-1}^3 \\ * & * & * & x_k^4 & x_k^3 u_{k-1} & x_k^2 u_{k-1}^2 \\ * & * & * & * & x_k^2 u_{k-1}^2 & x_k u_{k-1}^3 \\ * & * & * & * & * & u_{k-1}^4 \end{pmatrix}$$

$$M_k = \begin{pmatrix} E[y_k^4] - 6m_2(E[y_k^2] - m_2) - m_4 & (E[y_k^3] - 3m_2 E[y_k]) E[y_{k-1}] & (E[y_k^3] - 3m_2 E[y_k]) u_{k-1} & & & \\ * & (E[y_k^2] - m_2)(E[y_{k-1}^2] - m_2) & (E[y_k^2] - m_2) E[y_{k-1}] u_{k-1} & & & \\ * & * & (E[y_k^2] - m_2) u_{k-1}^2 & \dots & & \\ * & * & * & & & \\ * & * & * & & & \\ * & * & * & & & \end{pmatrix}$$

$$\dots \begin{pmatrix} (E[y_k^2] - m_2)(E[y_{k-1}^2] - m_2) & (E[y_k^2] - m_2) E[y_{k-1}] u_{k-1} & (E[y_k^2] - m_2) u_{k-1}^2 \\ (E[y_{k-1}^3] - 3m_2 E[y_{k-1}]) E[y_k] & (E[y_{k-1}^2] - m_2) E[y_k] u_{k-1} & E[y_k] E[y_{k-1}] u_{k-1}^2 \\ (E[y_{k-1}^2] - m_2) E[y_k] u_{k-1} & E[y_k] E[y_{k-1}] u_{k-1}^2 & E[y_k] u_{k-1}^3 \\ E[y_{k-1}^4] - 6m_2(E[y_{k-1}^2] - m_2) - m_4 & (E[y_{k-1}^3] - 3m_2 E[y_{k-1}]) u_{k-1} & (E[y_{k-1}^2] - m_2) u_{k-1}^2 \\ * & (E[y_{k-1}^2] - m_2) u_{k-1}^2 & E[y_{k-1}] u_{k-1}^3 \\ * & * & u_{k-1}^4 \end{pmatrix}$$

Fig. 1. Example of construction of  $M_k$

- Step 2. Let  $c_n$  be the singular vector associated with the minimum singular value of  $\widehat{\mathcal{M}}_N$ .
- Step 3. Determine the coefficients of the subsystems from the vector  $c_n$ .

In order to perform Step 3 in Algorithm 1, we adopt polynomial differentiation algorithm for mixtures of hyperplanes, introduced by Vidal [21, pp. 69–70].

#### V. ESTIMATING UNKNOWN NOISE PARAMETERS

We now address the case where the distribution of the noise is not completely known. As mentioned in Assumption 1, the distribution of the noise is known except for a few parameters  $\theta$ . For simplicity of exposition, let's consider the case where the noise has a normal distribution with zero mean and unknown variance  $s^2$ . The reasoning extends to any noise distribution with a small number of unknown parameters.

In such a case, the objective is to simultaneously estimate system parameters and the variance of noise. We start by noting that computing  $\widehat{\mathcal{M}}_N$  using the true value of the variance results in a rank deficient matrix. Moreover, given collected data  $y_k$  and  $u_k$ , the matrix  $\widehat{\mathcal{M}}_N$  is a continuous function of the moments of noise and, hence, a *known* continuous function of the standard deviation  $s$ . Given previous convergence results, the true value of  $s$  will make  $\widehat{\mathcal{M}}_N$  to have a very small minimum singular value (especially for large values of  $N$ ). For this reason, estimation of  $s$  can be performed by minimizing the minimum singular value of matrix above over the allowable values of  $s$ . More precisely, we propose the following algorithm

*Algorithm 2:* Let  $n_a$ ,  $n_c$ ,  $n$ , some parameters of the noise and  $s_{\max}$  be given.

Step 1. Compute matrix

$$\widehat{\mathcal{M}}_N = \frac{1}{N} \sum_{k=1}^N M[\text{mon}(y_k, \dots, y_{k-n_a})]$$

as a function of the noise parameter  $s$ .

- Step 2. Find the value  $s^* \in [0, s_{\max}]$  that minimizes the minimum singular value of  $\widehat{\mathcal{M}}_N$ .
- Step 3. Let  $c_n$  be associated singular vector.
- Step 4. Determine the coefficients of the subsystems from the vector  $c_n$ .

Note that the nonconvex optimization of Step 2 can be solved via an easily implementable line-search. However, the solution  $s^*$  might not be unique; i.e., there might exist several values of  $s$  that lead to a minimum singular value very close to zero. In practice, our experience has been that, for sufficiently large  $N$ , the above algorithm provides both a good estimate of the systems coefficients, and noise parameters; especially if we take  $s^*$  to be the smallest value of  $s$  for which the minimum singular value of  $\widehat{\mathcal{M}}_N$  is below a given threshold  $\epsilon$ .

#### VI. NUMERICAL RESULTS

In the following examples, we address the problem of identifying a two-modes switched system of the form of equation (17), whose true coefficients are  $a_1 = 0.3$ ,  $b_1 = 1$ ,  $a_2 = -0.5$ , and  $b_2 = -1$ . Measurement noise is assumed to be zero-mean with Normal distribution. In the numerical examples presented,  $N = 10^6$  input-output data is given. True and identified coefficients for different variances of noise, are presented in Table I. Variance of noise and noise to output ratio for each experiment are also shown in this table. The provided noise to output ratio ( $\gamma$ ) is defined as

$$\gamma = \frac{\max |\eta|}{\max |y|} \quad (20)$$

Results are as expected even for high values of noise in comparison to output. As it is illustrated in Table I, the identified parameters are very close to true values which demonstrates the convergence of proposed algorithm even for small signal to noise ratio. Moreover, the algorithm requires

TABLE I  
IDENTIFYING POLYNOMIAL COEFFICIENTS FOR DIFFERENT VALUES OF NOISE VARIANCE AND DIFFERENT SYSTEM RUN.

system #	Value 1	Value 2 $-(a_1 + a_2)$	Value 3 $-(b_1 + b_2)$	Value 4 $a_1 a_2$	Value 5 $a_1 b_2 + b_1 a_2$	Value 6 $b_1 b_2$	Value 7 $\gamma$	Value 8 $s^2$	Value 9 estimation of $s^2$
true parameters	1	0.2	0	-0.15	-0.8	-1	-	-	-
identification 1	1	0.2002	0.0001	-0.1503	-0.7989	-0.9996	0.2410	0.1	0.1000
identification 2	1	0.2002	0.0011	-0.1510	-0.7974	-1.0004	0.5187	0.5	0.4980
identification 3	1	0.1977	0.0046	-0.1548	-0.7997	-0.9966	0.6494	1	1.0010
identification 4	1	0.2120	0.0003	-0.1485	-0.8006	-1.0017	0.8516	2	1.9950

a very small computational effort. For the case of  $10^6$  measurements and using an off-the-shelf core i5 laptop with 8 Gigs of RAM, the running time is between 7 to 8 seconds, which shows the effectiveness of approach for very large data sets.

The second norm of error between true coefficients of system and estimated coefficients,  $\|c_n - \hat{c}_n\|_2$ , as a function of number of measurements,  $N$ , is depicted in Figure 2 for different values of noise variance. As it can be seen from Figure 2, the error decreases as the number of measurements increases. Rate of convergence is fast, despite the fact that, in some of the experiments, a large amount of noise is used. It should be noted that these results are for one experiment, and given that this is a realization of a random process, error is not always decreasing. For all values of noise variance, error will eventually decrease and the estimated values of coefficients converge to the true values.

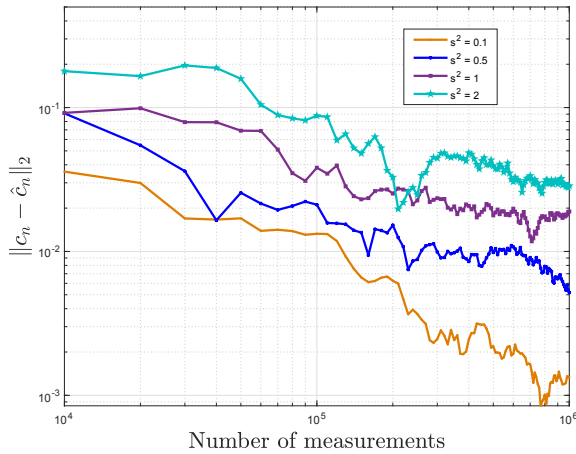


Fig. 2. Estimation error of system coefficients

Now that we have identified the coefficients of polynomial, it is time to identify each subsystems' coefficients. For the above mentioned example, Table II shows the values of subsystems coefficients for different experiments related to different values of noise variance. As we see in this table the value of coefficients are very close to the true values, even when the noise variance is high with noise magnitude in average around 85% of the signal magnitude.

The estimation of noise variance based on the structure of matrix  $M_k$  is shown in Table I as well. The estimates of noise variance are very close to the true values of variance.

TABLE II  
IDENTIFYING SUBMODELS' COEFFICIENTS FOR DIFFERENT VALUES OF NOISE VARIANCE.

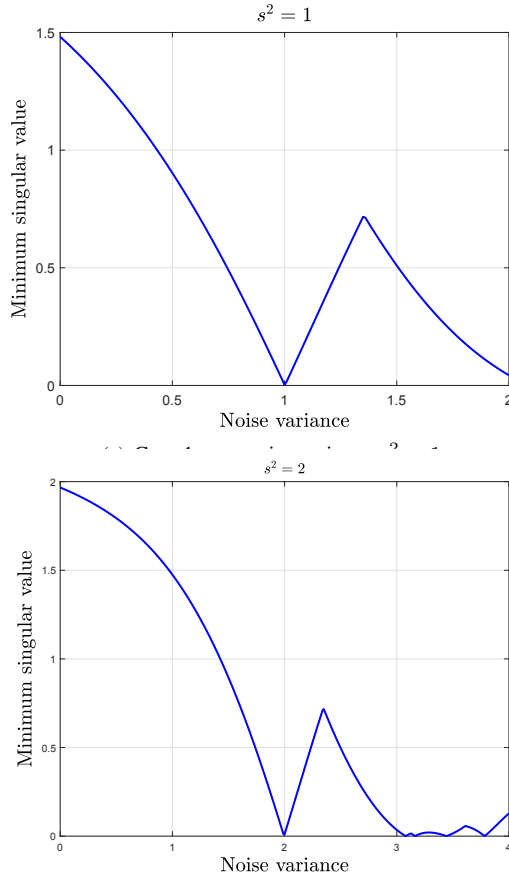
coeff- icients	true values	variance $s^2 = 0.1$	variance $s^2 = 0.5$	variance $s^2 = 1$	variance $s^2 = 2$
$a_1$	0.3	0.3002	0.2981	0.3006	0.2938
$b_1$	1	0.9988	1.0007	0.9412	1.0031
$a_2$	-0.5	-0.4996	-0.5000	-0.5006	-0.5059
$b_2$	-1	-0.9999	-0.9991	-1.0004	-1.0011

By knowing the structure of matrix  $M_k$ , the dependence of every entry on the moments of noise, and the relation in between these moments and the unknown variance (see Section I-B), we are able to estimate the noise parameter (in this case, noise variance). This illustrates the capability of the proposed algorithm to estimate both system and noise parameters even for large values of noise.

Two examples of the process of estimating the unknown variance of noise are shown in Fig. 3; where Fig. 3(a) is for the case of given data contaminated with noise of variance 1, and Fig. 3(b) is for data with measurement noise of variance 2. By taking  $s^*$  as the smallest local minimum, the estimated variance for both cases in Fig. 3(a) and Fig. 3(b) is very close to the true values.

## VII. CONCLUSION AND FUTURE WORK

In this paper we have proposed a methodology to identify the coefficients of switched autoregressive processes and unknown noise parameters, starting from partial information of the noise and given input-output data of switched system. The approach is shown to be particularly efficient in the case of large amount of data, situation that makes it possible to exploit law-of-large-numbers type of results. The approach requires the computation of singular value decomposition of a specially constructed input-output Veronese matrix. The ensuing singular vector is then related to the switched system parameters to be identified. We prove that the estimated parameters converge to the true ones as the number of measurements grows. Numerical simulations show a low estimation error, even in the case of large measurement noise. Also, in cases that noise distribution is not completely known, simulation results show very close estimation of unknown parameters of noise to the true values. In future work, we will consider the problem of identifying switched systems with process noise from large amount of noisy data. Moreover, we will address the problem of identifying



(b) Case 2: true noise variance =  $s^2 = 2$ .

Fig. 3. Estimation of noise variance

switching dynamics in switched processes form large noisy data sets.

## REFERENCES

- [1] Laurent Bako. Identification of switched linear systems via sparse optimization. *Automatica*, 47(4):668–677, April 2011.
- [2] David E Conroy, Sarah Hojjatinia, Constantino M Lagoa, Chih-Hsiang Yang, Stephanie T Lanza, and Joshua M Smyth. Personalized models of physical activity responses to text message micro-interventions: A proof-of-concept application of control systems engineering methods. *Psychology of Sport and Exercise*, 41:172–180, 2019.
- [3] Andrea Garulli, Simone Paoletti, and Antonio Vicino. A survey on switched and piecewise affine system identification. *IFAC Proceedings Volumes*, 45(16):344–355, 2012.
- [4] Joe Harris. *Algebraic geometry: a first course*, volume 133. Springer Science & Business Media, 2013.
- [5] Sarah Hojjatinia, Korkut Bekiroglu, and Constantino M Lagoa. Parsimonious voltaerra system identification. In *2018 Annual American Control Conference (ACC)*, pages 1933–1938. IEEE, 2018.
- [6] Sarah Hojjatinia, Constantino M. Lagoa, and Fabrizio Dabbene. A method for identification of markovian jump arx processes. *IFAC-PapersOnLine*, 50(1):14088 – 14093, 2017. 20th IFAC World Congress.
- [7] Constantino M. Lagoa, David E. Conroy, Sarah Hojjatinia, and Chih-Hsiang Yang. Modeling subject response to interventions aimed at increasing physical activity: A control systems approach. *Poster Presented at 5th International Conference on Ambulatory Monitoring of Physical Activity and Movement*, 2017.
- [8] Fabien Lauer, Gérard Bloch, and René Vidal. A continuous optimization framework for hybrid system identification. *Automatica*, 47(3):608–613, 2011.

- [9] Ruei-Sung Lin, Che-Bin Liu, Ming-Hsuan Yang, Narendra Ahuja, and Stephen Levinson. Learning nonlinear manifolds from time series. In *European Conference on Computer Vision*, pages 245–256. Springer, 2006.
- [10] Jan Lunze and Françoise Lamnabhi-Lagarigue. *Handbook of hybrid systems control: theory, tools, applications*. Cambridge University Press, 2009.
- [11] Yi Ma and René Vidal. *Identification of Deterministic Switched ARX Systems via Identification of Algebraic Varieties*, pages 449–465. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [12] Necmiye Ozay, Constantino Lagoa, and Mario Sznaiier. Set membership identification of switched linear systems with known number of subsystems. *Automatica*, 51:180 – 191, 2015.
- [13] Necmiye Ozay, Mario Sznaiier, and Constantino Lagoa. Convex certificates for model (in) validation of switched affine systems with unknown switches. *IEEE Transactions on Automatic Control*, 59(11):2921–2932, 2014.
- [14] Simone Paoletti, Aleksandar Lj Juloski, Giancarlo Ferrari-Trecate, and René Vidal. Identification of hybrid systems a tutorial. *European journal of control*, 13(2-3):242–260, 2007.
- [15] Jacob Roll, Alberto Bemporad, and Lennart Ljung. Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1):37–50, 2004.
- [16] Lawrence K Saul and Sam T Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *Journal of machine learning research*, 4(Jun):119–155, 2003.
- [17] Pranab K. Sen and Julio d. M. Singer. *Large sample methods in statistics: an introduction with applications*. Chapman & Hall, New York, 1993.
- [18] Mario Sznaiier, Octavia Camps, Necmiye Ozay, and Constantino Lagoa. Surviving the upcoming data deluge: A systems and control perspective. In *53rd IEEE Conference on Decision and Control*, pages 1488–1498. IEEE, 2014.
- [19] R. Vidal, S. Soatto, Yi Ma, and S. Sastry. An algebraic geometric approach to the identification of a class of linear hybrid systems. In *Proc. 42nd IEEE Conf. Decision and Control*, volume 1, pages 167–172, December 2003.
- [20] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1945–1959, 2005.
- [21] Rene Esteban Vidal and Shankar Sastry. *Generalized principal component analysis (gpca): an algebraic geometric approach to subspace clustering and motion segmentation*. Electronics Research Laboratory, College of Engineering, University of California, 2003.

## APPENDIX

**Sketch of Proof of Theorem 1:** For simplicity of presentation, let

$$\widehat{M}_k \doteq M[\text{mon}_n(y_k, \dots, y_{k-n_a})]$$

We first note that, given the assumptions made on the noise,  $u_k$  and  $x_k$ , the entries of  $\widehat{M}_k$  have a variance uniformly bounded for all  $k$ . Moreover

$$k > l + n_a \Rightarrow \widehat{M}_k \text{ and } \widehat{M}_l \text{ are independent.}$$

Hence, by Kolmogorov’s Strong Law of Large Numbers [17] we have

$$\frac{1}{L} \sum_{l=1}^L \widehat{M}_{k+l(n_a+1)} - \frac{1}{L} \sum_{l=1}^L E[\widehat{M}_{k+l(n_a+1)}] \rightarrow 0 \text{ a.s.}$$

as  $L \rightarrow \infty$ . Since

$$E[\widehat{M}_k] = M_k \text{ for all positive integer } k$$

and applying the results above for  $k = 1, 2, \dots, n_a + 1$ , we conclude that

$$\frac{1}{N} \sum_{j=1}^N \widehat{M}_j - \frac{1}{N} \sum_{j=1}^N M_j \rightarrow 0 \text{ a.s.}$$

as  $N \rightarrow \infty$ .