

# Robust Identification of Switched Affine Systems via Moments-Based Convex Optimization

Necmiye Ozay\*

Constantino Lagoa\*\*

Mario Sznaier\*

**Abstract**—This paper addresses the problem of robust identification of a class of discrete-time affine hybrid systems, switched affine models, in a set membership framework. Given a finite collection of noisy input/output data and a bound on the number of subsystems, the objective is to identify a suitable set of affine models along with a switching sequence that can explain the available experimental information. Our method builds upon an algebraic procedure proposed by Vidal *et al.* for noise free measurements. In the presence of norm bounded noise, this algebraic procedure leads to a very challenging nonconvex polynomial optimization problem. Our main result shows that this problem can be reduced to minimizing the rank of a matrix whose entries are affine in the optimization variables, subject to a convex constraint imposing that these variables are the moments of an (unknown) probability distribution function with finite support. Appealing to well known convex relaxations of rank leads to an overall semi-definite optimization problem that can be efficiently solved. These results are illustrated with two examples showing substantially improved identification performance in the presence of noise.

## I. INTRODUCTION AND MOTIVATION

In the past few years, considerable attention has been devoted to the problem of identifying hybrid systems, leading to several methods (see the excellent tutorial paper [9] for a summary of the main issues and recent developments). While successful in many situations, a common feature of these methods is the computational complexity entailed in dealing with noise: in this case algebraic procedures [13], [7] lead to challenging nonlinear/nonconvex optimization problems, while optimization methods lead to generically NP-hard problems, either necessitating the use of relaxations [1] or restricted to small size problems [10]. A sparsification-based convex relaxation was proposed in [8]. While this approach works well most of the time, it is not hard to construct counterexamples where it fails, due to its greedy nature, to find the minimum number of subsystems.

Motivated by these difficulties, in this paper we propose a convex optimization-based approach to the problem of identifying hybrid systems from noisy input/output data and some minimal *a-priori* information (bounds on the order and number of subsystems, and on the norm of the noise). The starting point is the algebraic procedure due to Vidal *et al.* [13], [7]. In the case of noiseless measurements, the (unknown) parameters of each subsystem are recovered from the null space of a matrix  $\mathbf{V}(\mathbf{r})$  constructed from the

input/output data  $\mathbf{r}$  via a nonlinear embedding (the Veronese map). In the case of noisy data, the entries of this matrix depend polynomially on the unknown noise terms. Thus, finding a model in the consistency set (e.g. a model that interpolates the data within the noise level) is equivalent to finding an admissible noise sequence  $\eta$  that renders the matrix  $\mathbf{V}(\mathbf{r})$  rank deficient, and a vector  $\mathbf{c}$  in its null space. However, this is not trivial, given the polynomial dependence noted above. The main result of this paper shows that the problem of jointly finding  $\eta$  and  $\mathbf{c}$  is equivalent to minimizing the rank of a matrix whose entries are affine in the optimization variables, subject to a convex constraint imposing that these variables are the moments of a suitable probability distribution function. This result is achieved by using first an idea similar to that of [6] relating polynomial optimization and the problem of moments, to eliminate the polynomial dependence on the optimization variables, albeit at the price of introducing infinitely many constraints. The structure of the problem can then be exploited to decouple it into several *finite dimensional* smaller ones, each involving only the moments of a one-dimensional distribution. Combining these ideas with a convex relaxation, similar to the log-det heuristic of [4], that aims at dropping the rank of  $\mathbf{V}$  by one and estimating a vector in its nullspace, allows for recasting the original problem into a semidefinite optimization form that can be solved efficiently.

The paper is organized as follows. Section II presents some background results related to polynomial optimization. In section III, we formally state the problem under consideration and review the original algebraic method of [13] for noiseless systems. The main results are presented in section IV. Section V illustrates these results with two examples, one academic and one practical. Finally, section VI concludes the paper with some remarks and directions for future research.

## II. PRELIMINARIES

For ease of reference, we summarize next the notation used in the paper and recall some results required to recast the identification problem into a convex optimization form.

### A. Notation

$\mathbf{x}, \mathbf{M}$  a vector in  $\mathbb{R}^n$  (matrix in  $\mathbb{R}^{n \times m}$ )  
 $\|\mathbf{x}\|_\infty$   $\infty$ -norm of the vector  $\mathbf{x} \in \mathbb{R}^n$

$$\|\mathbf{x}\|_\infty \doteq \sup_i |x_i|$$

\*Department of Electrical Engineering, Penn State University, University Park, PA 16802. lagoa@engr.psu.edu.

\*\*ECE Department, Northeastern University, Boston, MA 02115. {ozay.n, m.sznaier}@neu.edu

This work was supported in part by NSF grants ECS-0648054, IIS-0713003, and AFOSR grant FA9550-09-1-0253.

$\text{Tr}\{\mathbf{M}\}$  trace of the matrix  $\mathbf{M}$   
 $\mathbf{I}$  identity matrix  
 $\mathbf{M} \succeq \mathbf{N}$  the matrix  $\mathbf{M} - \mathbf{N}$  is positive semidefinite.

$\nu_s : \mathbb{R}^n \rightarrow \mathbb{R}^m$  where  $m = \binom{s+n-1}{s}$ .  
 Veronese map of degree  $s$ :

$$\nu_s([x_1, \dots, x_n]^T) = [\dots, \xi^s, \dots]^T$$

where  $\xi^s \doteq x_1^{s_1} x_2^{s_2} \dots x_n^{s_n}$ ,  $\sum s_i = s$ , e.g. all possible monomials of order  $s$ , in lexicographical order.

### B. The Problem of Moments

Given a sequence of scalars  $\{m_i\}_{i=1}^n$ , the *problem of moments* is to determine whether there exist a probability measure that has  $\{m_i\}$  as its first  $n$  moments (see references [11], [5], [3] for a historical review and details of the problem). In particular, in the sequel we are interested in probability measures that are supported on bounded symmetric intervals of the real line. In this case, the following theorem provides necessary and sufficient conditions for the existence of such a measure.

*Theorem 1:* Given a sequence  $\{m_i : i = 1, 2, \dots, n\}$ , there exists a probability measure supported on  $[-\epsilon, \epsilon]$  such that

$$m_i = \mathbf{E}_\mu(x^i) = \int_{-\epsilon}^{\epsilon} x^i \mu(dx)$$

if and only if

- when  $n = 2k + 1$  (odd case), the following holds

$$\epsilon \mathbf{M}(0, 2k) \succeq \mathbf{M}(1, 2k + 1) \quad (1)$$

$$\mathbf{M}(1, 2k + 1) \succeq -\epsilon \mathbf{M}(0, 2k) \quad (2)$$

- when  $n = 2k$  (even case), the following holds

$$\mathbf{M}(0, 2k) \succeq 0 \quad (3)$$

$$\epsilon^2 \mathbf{M}(0, 2k - 2) \succeq \mathbf{M}(2, 2k) \quad (4)$$

where  $\mathbf{M}(i, i + 2j)$  is the  $(j + 1)$  by  $(j + 1)$  Hankel matrix formed from the moments, that is:

$$\mathbf{M}(i, i + 2j) \doteq \begin{bmatrix} m_i & m_{i+1} & \dots & m_{i+j} \\ m_{i+1} & \ddots & \ddots & m_{i+j+1} \\ \vdots & \ddots & \ddots & \vdots \\ m_{i+j} & \dots & \dots & m_{i+2j} \end{bmatrix}, \quad (5)$$

and where  $m_0 = 1$ .

*Proof:* Direct application of Theorem III.2.3 and Theorem III.2.4 in [5].  $\blacksquare$

### C. Polynomial Optimization via Moments

This section reviews some results from [6] that relate polynomial optimization to the problem of moments. Specifically, consider the problem of minimizing a real valued polynomial:

$$p_K^* := \min_{x \in K} p(x) \quad (\text{P1})$$

where  $K \subset \mathbb{R}^n$  is a compact set defined by polynomial inequalities. This problem is usually non-convex, hence hard to solve. Next, we consider a related problem:

$$\tilde{p}_K^* := \min_{\mu \in \mathcal{P}(K)} \int p(x) \mu(dx) := \min_{\mu \in \mathcal{P}(K)} \mathbf{E}_\mu [p(x)] \quad (\text{P2})$$

where  $\mathcal{P}(K)$  is the space of finite Borel probability measures on  $K$ . Although (P2) is an infinite dimensional problem, it is, in contrast to (P1), convex. The next result, taken from [6], establishes the relation between the two problems:

*Theorem 2:* Problems (P1) and (P2) are equivalent; that is:

- $\tilde{p}_K^* = p_K^*$ .
- If  $x^*$  is a global minimizer of (P1), then  $\mu^* = \delta_{x^*}$  is a global minimizer of (P2).
- For every optimal solution  $\mu^*$  of (P2),  $p(x) = p_K^*$ ,  $\mu^*$  almost everywhere.

*Proof:* See Proposition 2.1 in [6].  $\blacksquare$

One direct consequence of this theorem is that when combined with Theorem 1, it is possible to convert the infinite dimensional problem (P2) in measures (or equivalently, the polynomial optimization problem (P1)) to a finite dimensional Linear Matrix Inequalities (LMI) optimization problem in the moments, as long as the probability measures are supported on compact intervals.

## III. SET MEMBERSHIP IDENTIFICATION OF HYBRID LINEAR ARX MODELS

In this paper we consider the problem of set membership identification of single input single output, switched autoregressive exogenous (SARX) linear models of the form:

$$y_t = \sum_{i=1}^{n_a} a_i(\sigma_t) y_{t-i} + \sum_{i=1}^{n_c} c_i(\sigma_t) u_{t-i} + \eta_t \quad (6)$$

where  $u$ ,  $y$  and  $\eta$  denote the input, output and noise, respectively, and where  $\sigma_t \in \{1, \dots, s\}$  is the discrete state or mode of the system. Different values of  $\sigma_t$  correspond to  $s$  different hybrid submodels. Our goal is to, given input/output data over the interval  $[t_0, T]$ , find a model of the form (6) that interpolates this experimental data within a given noise level  $\|\eta_t\|_\infty \leq \epsilon$ . Clearly, as stated this problem admits infinitely many solutions. For instance, one can always find a trivial hybrid model with  $T - t_0 + 1$  submodels or one model with a large order that perfectly fits the data. This situation can be avoided by adding additional constraints or performance criteria to the problem. For instance, in [8], the problem was regularized by finding a solution with minimum number of switches. In this paper, we are interested in finding a solution where  $s$ , the number of subsystems, is given a priori<sup>1</sup>. Practical situations where this problem is relevant arise for instance in segmentation problems in computer vision and medical image processing, where it is desired to segment an image or video clip into a given number of (not necessarily

<sup>1</sup>Note that by performing a line search one can then find a solution with the minimum number of subsystems, a problem solved in [1] via a greedy algorithm based on randomization and thermal relaxations.

contiguous) regions, for instance corresponding to healthy versus diseased tissue, or a fixed number of activities. Thus, the problem of interest here can formally be stated as follows:

*Problem 1:* Given input/output data over the interval  $[1, T]$ , a bound on the  $\ell_\infty$  norm of the noise (i.e.  $\|\eta\|_\infty \leq \epsilon$ ), and the number of submodels  $s$ , find a hybrid linear model of the form (6) that is consistent with the a priori information and experimental data.

In the noise free case (i.e.  $\eta_t = 0 \forall t$ ), the problem above can be elegantly solved using an algebraic procedure, Generalized Principal Component Analysis (GPCA), proposed by Vidal *et al.* [13], [7]. Note that in this case an equivalent representation of (6) is:

$$\mathbf{b}(\sigma_t)^T \mathbf{r}_t = 0 \quad (7)$$

where  $\mathbf{r}_t = [-y_t, y_{t-1}, \dots, y_{t-n_a}, u_{t-1}, \dots, u_{t-n_c}]^T$  and  $\mathbf{b}(\sigma_t) = [1, a_1(\sigma_t), \dots, a_{n_a}(\sigma_t), c_1(\sigma_t), \dots, c_{n_c}(\sigma_t)]^T$ , denote the regressor and (unknown) coefficients vectors at time  $t$ , respectively.

The idea behind the algebraic method is based on a polynomial constraint, the so-called *hybrid decoupling constraint*, that decouples the identification of model parameters from the identification of the discrete state and switching sequence. That is, (7) holds for some  $\sigma_t$  iff

$$p_s(\mathbf{r}) = \prod_{i=1}^s (\mathbf{b}_i^T \mathbf{r}_t) = \mathbf{c}_s^T \nu_s(\mathbf{r}_t) = 0 \quad (8)$$

holds for all  $t$  independent of which of the  $s$  submodels is active at time  $t$ . In the above equality,  $\mathbf{b}_i \in \mathbb{R}^{n_a+n_c+1}$  is the parameter vector corresponding to the  $i^{\text{th}}$  submodel,  $\mathbf{r}_t$  is the known regressor vector at time  $t$ , and  $\nu_s(\cdot)$  is the Veronese map of degree  $s$ . Collecting all data into a matrix form leads to:

$$\mathbf{V}_s \mathbf{c}_s \doteq \begin{bmatrix} \nu_s(\mathbf{r}_{t_0})^T \\ \vdots \\ \nu_s(\mathbf{r}_T)^T \end{bmatrix} \mathbf{c}_s = \mathbf{0} \quad (9)$$

Hence, one can solve for a vector  $\mathbf{c}_s$  in the nullspace of  $\mathbf{V}_s$ . Finally,  $\mathbf{b}_i$ , the parameters of the models can be computed from  $\mathbf{c}_s$  via polynomial differentiation (see the Appendix).

#### IV. MAIN RESULTS

In the presence of noise, the approach outlined above breaks down, since conditions (8) and (9) no longer hold. Indeed, the noisy equivalent of (8) is given by:

$$p_s(\mathbf{r}, \eta) = \prod_{i=1}^s (\mathbf{b}_i^T \tilde{\mathbf{r}}_t) = \mathbf{c}_s^T \nu_s(\tilde{\mathbf{r}}_t) = 0 \quad (10)$$

where  $\tilde{\mathbf{r}}_t = [-y_t + \eta_t, y_{t-1}, \dots, y_{t-n_a}, u_{t-1}, \dots, u_{t-n_c}]^T$ . Proceeding as in the noiseless case, a ‘‘noisy’’ data matrix  $\mathbf{V}_s(\mathbf{r}, \eta)$  can be built. However, finding the coefficients of each subsystem entails now finding both an admissible noise sequence  $\eta^\circ$  and a vector  $\mathbf{c}^\circ$  in the nullspace of  $\mathbf{V}_s(\mathbf{r}, \eta^\circ)$  such that

$$\mathbf{V}_s(\mathbf{r}, \eta^\circ) \mathbf{c}^\circ = 0 \quad (11)$$

Since  $\mathbf{V}_s$  is a polynomial function of the unknown noise terms  $\eta(t)$ , this approach leads to a computationally very challenging nonlinear, nonconvex optimization problem. However, as we show in the sequel, by exploiting the method of the moments, (11) can be recast into a constrained rank minimization form which in turn can be relaxed to an efficient convex optimization.

##### A. A moments based convex relaxation

Consider the following rank minimization problem:

$$\begin{aligned} & \text{minimize}_{\eta_t} \quad \text{rank } \mathbf{V}_s(\mathbf{r}_t, \eta_t) \\ & \text{subject to} \quad \|\eta_t\|_\infty \leq \epsilon \end{aligned} \quad (12)$$

Clearly, Problem 1 is solvable if and only if (12) admits a rank deficient solution. Indeed, in the case where the order of each subsystem is precisely  $(n_a, n_c)$  and  $\mathbf{V}_s$  has generically full column rank (e.g. enough data points have been collected from each subsystem), then there exists a noise sequence  $\eta_t^\circ$  such that the right null space of  $\mathbf{V}_s(\mathbf{r}_t, \eta_t^\circ)$  has dimension 1. When some of the subsystems have order lower than  $(n_a, n_c)$  or the number of subsystems is overestimated, the dimension of the nullspace of  $\mathbf{V}_s$  is higher than 1 (see [7] for details). In that sense, minimizing the rank of  $\mathbf{V}_s$  amounts to searching for the simplest model that explains the data.

Since in this paper we are interested in finding just one model consistent with the a-priori information (bounds on the  $\|\eta_t\|_\infty$ , the number of subsystems and their order), we will simply search for rank deficient solutions to (12). As we show next, this problem admits a computationally tractable relaxation.

Exploiting Theorem 1 and using the facts that (i)  $\eta_t$  and  $\eta_{\bar{t}}$  are independent for  $t \neq \bar{t}$ , and (ii)  $\eta_t$  only appears in the  $t^{\text{th}}$  row of  $\mathbf{V}_s$ , leads to the following moments optimization problem:

$$\begin{aligned} & \text{minimize}_{\mathbf{m}^{(t)}} \quad \text{rank } \tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{(t)}) \\ & \text{subject to} \quad (1) - (2) \quad \forall \mathbf{m}^{(t)} \quad \text{if } s \text{ is odd} \\ & \quad \quad \quad (3) - (4) \quad \forall \mathbf{m}^{(t)} \quad \text{if } s \text{ is even} \end{aligned} \quad (13)$$

where  $\mathbf{m}^{(t)} = [m_1^{(t)}, \dots, m_s^{(t)}]$  is the moment sequence corresponding to  $\eta_t$  and  $\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{(t)})$  is a matrix linear in the moments, obtained by replacing each  $k^{\text{th}}$  degree monomial  $\eta_t^k$  in  $\mathbf{V}_s(\mathbf{r}_t, \eta_t)$  with the corresponding  $k^{\text{th}}$  order moment  $m_k^{(t)}$ .

*Example 1:* For instance when  $s = 2$  and  $(n_a, n_c) = (1, 1)$ , then  $\mathbf{r}_t = [-y_t, y_{t-1}, u_{t-1}]^T$  and the rows of  $\mathbf{V}_s(\mathbf{r}_t, \eta_t)$  depend on  $\mathbf{r}$  and  $\eta$  as follows:

$$\nu_2(\mathbf{r}_t, \eta_t)^T = \begin{bmatrix} y_t^2 - 2y_t\eta_t + \eta_t^2 \\ -y_t y_{t-1} + y_{t-1}\eta_t \\ -y_t u_{t-1} + u_{t-1}\eta_t \\ y_{t-1}^2 \\ y_{t-1} u_{t-1} \\ u_{t-1}^2 \end{bmatrix}^T. \quad (14)$$

The corresponding row of  $\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{(t)})$  is given by:

$$\mathbf{E}_{\mu} [\nu_2(\mathbf{r}_t, \eta_t)^T] = \begin{bmatrix} y_t^2 - 2y_t m_1^{(t)} + m_2^{(t)} \\ -y_t y_{t-1} + y_{t-1} m_1^{(t)} \\ -y_t u_{t-1} + u_{t-1} m_1^{(t)} \\ y_{t-1}^2 \\ y_{t-1} u_{t-1} \\ u_{t-1}^2 \end{bmatrix}^T \quad (15)$$

Thus,  $\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{(t)})$  is affine in the unknown moments.

Note that Theorem 2 cannot be directly applied to show the equivalence of (12) and (13) since rank is not a polynomial function. Nevertheless, as we show next, a result similar to Theorem 2 can be obtained:

*Theorem 3:* There exists a rank deficient solution to problem (12) if and only if there exists a rank deficient solution to problem (13). Moreover, if  $\mathbf{c}$  belongs to the nullspace of the solution of (13) then there exists a noise value  $\eta^*$  with  $\|\eta^*\|_{\infty} \leq \epsilon$  such that  $\mathbf{c}$  belongs to the nullspace of  $\mathbf{V}_s(\mathbf{r}_t, \eta^*)$ .

*Proof:* Assume that the minimum rank  $r_1$  in (12) is achieved by some sequence  $\eta_t^*$ . Then  $\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{*(t)})$  with  $\mathbf{m}^{*(t)} = [\eta_t^*, (\eta_t^*)^2, \dots, (\eta_t^*)^s]$  (i.e. all distributions have point support) has rank  $r_1$  and  $\mathbf{m}^{*(t)}$  satisfies the LMI constraints. Hence the minimum rank obtained by solving (13) is less than or equal to the minimum rank obtained by solving (12).

Consider now an optimal solution  $\mathbf{m}^{*(t)}$  of (13). Note that, from Theorem 1, this guarantees the existence of  $T$  measures  $\mu^{*(t)}$ , each supported on  $[-\epsilon, \epsilon]$ . Let  $\mathbf{c}$  be in the nullspace of  $\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{*(t)})$  (i.e.  $\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{*(t)})\mathbf{c} = \mathbf{0}$ ). Thus, for each row of  $\mathbf{V}_s$ ,  $\mathbf{E}_{\mu^{*(t)}} [\nu_s(\mathbf{r}_t, \eta_t)]\mathbf{c} = \mathbf{E}_{\mu^{*(t)}} [\nu_s(\mathbf{r}_t, \eta_t)\mathbf{c}] = \mathbf{0}$ . By noting that  $\nu_s(\mathbf{r}_t, \eta_t)\mathbf{c}$  is a polynomial function of  $\eta_t$  (hence continuous) and  $\mu^{*(t)}$  is supported on  $[-\epsilon, \epsilon]$ , we can invoke the mean value theorem for integration to conclude that there exist  $\eta_t^* \in [-\epsilon, \epsilon]$  for all  $t$  such that  $\nu_s(\mathbf{r}_t, \eta_t^*)\mathbf{c} = \mathbf{0}$ .

Thus, whenever the nullspace of the solution of (12) is non-trivial, so is that of (13), which proves the theorem. ■

Although rank minimization is an NP-Hard problem, efficient convex relaxations are available. In particular, good approximate solutions can be obtained by using a log-det heuristic [4] that relaxes rank minimization to a sequence of convex problems. Furthermore, since from a set membership point of view it suffices to find a rank deficient solution, we propose a modification of log-det heuristic that aims at dropping the rank by one. The algorithm, which is inspired by the adaptive step size defined for weighted  $\ell_1$  minimization in [2], is summarized next:

---

#### Algorithm 1: Drop Rank

---

$\mathbf{X} \in \mathbb{R}^{m \times n}$  and assuming wlog  $m \leq n$ , initialize:

$$k = 0$$

$$\mathbf{W}_y^{(0)} = \mathbf{I}_{m \times m}$$

$$\mathbf{W}_z^{(0)} = \mathbf{I}_{n \times n}$$

REPEAT

Solve

$$\min_{\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}, \mathbf{Z}^{(k)}} \text{Tr} \begin{bmatrix} \mathbf{W}_y^{(k)} \mathbf{Y}^{(k)} & 0 \\ 0 & \mathbf{W}_z^{(k)} \mathbf{Z}^{(k)} \end{bmatrix}$$

subject to  $\begin{bmatrix} \mathbf{Y}^{(k)} & \mathbf{X}^{(k)} \\ \mathbf{X}^{T(k)} & \mathbf{Z}^{(k)} \end{bmatrix} \succeq 0$   
 $\mathbf{X}^{(k)} \in \mathcal{C}$

Decompose  $\mathbf{X}^{(k)} = \mathbf{U}\mathbf{D}\mathbf{V}^T$  using SVD.

Set  $\epsilon = \mathbf{D}(m, m)$ .

Set  $\mathbf{W}_y^{(k+1)} = (\mathbf{Y}^{(k)} + \epsilon \mathbf{I})^{-1}$ .

Set  $\mathbf{W}_z^{(k+1)} = (\mathbf{Z}^{(k)} + \epsilon \mathbf{I})^{-1}$ .

Set  $k = k + 1$ .

UNTIL (a convergence criterion is reached)

RETURN  $\mathbf{X}^{(k)}$

---

Above, for the sake of notational simplicity, we used  $\mathbf{X} = \tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{(t)})$ ; and  $\mathbf{X}^{(k)} \in \mathcal{C}$  stands for convex constraints, that is,  $\mathbf{m}^{(t)}$  lies on a convex set  $\mathcal{C}$  defined by LMIs in (13).

Assuming a rank deficient  $\tilde{\mathbf{V}}_s(\mathbf{r}_t, \mathbf{m}^{(t)})$  is found, a vector  $\mathbf{c}$  in its nullspace can be found by simply performing a singular value decomposition. From Theorem 3, it follows that  $\mathbf{c}$  is also in the nullspace of  $\mathbf{V}_s(\mathbf{r}_t, \eta_t)$  (i.e.  $\mathbf{V}_s(\mathbf{r}_t, \eta_t)(\mathbf{c}) = \mathbf{0}$ ). Hence, for each row, we have  $\nu_s(\mathbf{r}_t, \eta_t^*)\mathbf{c} = \mathbf{0}$  which is a polynomial equation in one variable. One can solve for the noise values by finding the roots of this polynomial that lie in  $[-\epsilon, \epsilon]$  (which is guaranteed to exist again by Theorem 3). Once the noise values are estimated, the problem can be converted to the noise free case by plugging the noise estimates into  $\mathbf{V}_s(\mathbf{r}_t, \eta_t)$  and the system parameters can be computed using the procedure described in the Appendix.

*Remark 1:* When the number of the submodels  $s$  is unknown, it is possible to search for minimum number of submodels that explains the data. This can be accomplished with a simple iteration on  $s$ ; starting with  $s = 1$  and increasing  $s$  up until a rank deficient solution to the problem (13) is found.

## V. ILLUSTRATIVE EXAMPLES

In this section we use both an academic and a practical example to illustrate the effectiveness of the proposed method.

### A. Academic Example

Consider a hybrid system that switches among the following three ARX subsystems

$$y_t = 0.2y_{t-1} + 0.24y_{t-2} + 2u_{t-1} + \eta_t \quad (\text{Submodel 1})$$

$$y_t = -1.4y_{t-1} - 0.53y_{t-2} + u_{t-1} + \eta_t \quad (\text{Submodel 2})$$

$$y_t = 1.7y_{t-1} - 0.72y_{t-2} + 0.5u_{t-1} + \eta_t \quad (\text{Submodel 3})$$

modeled as

$$y_t = p_1(\sigma_t)y_{t-1} + p_2(\sigma_t)y_{t-2} + p_3(\sigma_t)u_{t-1} + \eta_t. \quad (16)$$

where  $\sigma_t \in \{1, 2, 3\}$  depending on which model is active at time  $t$ . Experimental data was obtained by running a simulation for  $T = 96$  time steps with  $\|\eta\|_{\infty} = 0.25$  where  $\sigma_t = 1$  for  $t = [1, 32]$ ,  $\sigma_t = 2$  for  $t = [33, 64]$  and  $\sigma_t = 3$  for  $t = [65, 96]$ . The parameter values used for the simulation

are shown in Table I together with the results obtained by our robust moments–based method and the original algebraic method of [13]. Figures 1 and 2 show the clustering of data into different submodels. As seen there, the proposed method outperforms the method in [13]. Figure 3 shows the absolute error given the identified model. The error values are quite large for the method in [13] whereas they mostly satisfy the prior bound of  $\|\eta\|_\infty = 0.25$  for the new method. Indeed, in the latter case, the error exceeds the bound only at a single time instant. This is due to the convex relaxation used to solve the rank–minimization problem.

		True	Moments Based	GPCA
Submodel 1	$p_1$	0.2000	0.1964	0.2248
	$p_2$	0.2400	0.2332	0.3764
	$p_3$	2.0000	1.9287	2.5907
Submodel 2	$p_1$	-1.4000	-1.2959	-0.4491
	$p_2$	-0.5300	-0.4469	0.9188
	$p_3$	1.0000	1.0315	1.5262
Submodel 3	$p_1$	1.7000	1.6505	1.7213
	$p_2$	-0.7200	-0.6713	-0.7103
	$p_3$	0.5000	0.5007	1.2194

TABLE I  
ESTIMATED AND TRUE VALUES OF PARAMETERS

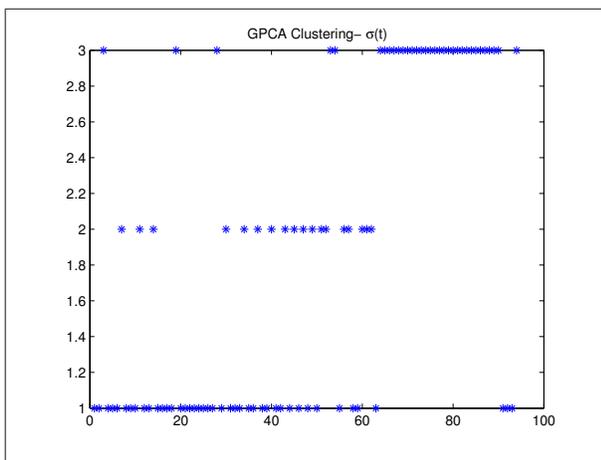


Fig. 1. Clustering via GPCA.

**B. A Practical Example: Human activity segmentation**

Next, we illustrate an application of the proposed method in a computer vision problem: human activity analysis. The input data, illustrated in Fig. 4, consists of 55 frames extracted from a video sequence of a person walking and bending in front of the camera. The goal here is to segment the two activities: walking and bending. In order to recast the problem into the identification of a piecewise affine system, we first used simple background subtraction to estimate the location of the center of mass of the person in each frame. The horizontal<sup>2</sup> position of the center of mass was then

<sup>2</sup>It may seem more natural to use the vertical position. However, this lead to 3 segments, corresponding to no vertical motion, downward and upward motion, while there are only two different activities involved.

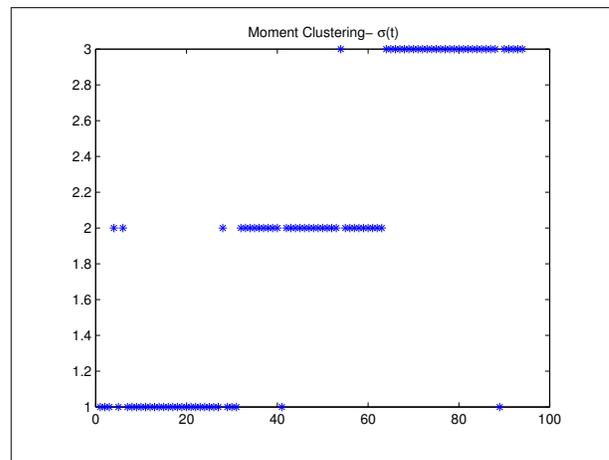


Fig. 2. Clustering via moments–based method.

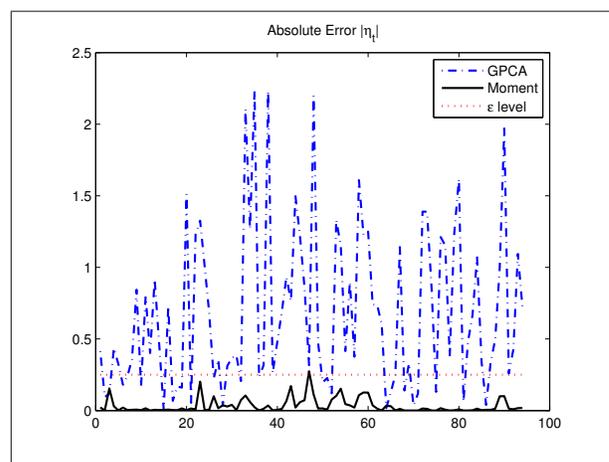


Fig. 3. Comparison of fitting errors for the moments–based and GPCA identification.

modeled as the output of a first order SAR system:

$$x_t = a(\sigma_t)x_{t-1} + f(\sigma_t) + \eta_t \tag{17}$$

where  $a(\sigma_t)$  and  $f(\sigma_t)$  are unknown parameters. We set  $\|\eta_t\|_\infty = 3$ , e.g.  $\pm 3$  pixels error in the position estimates.

Figure 6 shows that the proposed method is capable of segmenting the two activities, correctly labeling the activity at the initial and final portions of the video as the same. The single misclassification in the last frame of the sequence is due to an inaccurate estimate of the centroid of the person as she starts to leave the camera’s field of view. On the other hand, the classification with the original GPCA is less accurate as can be seen from Fig. 5.

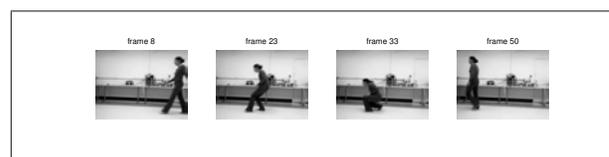


Fig. 4. Sample frames from the video.

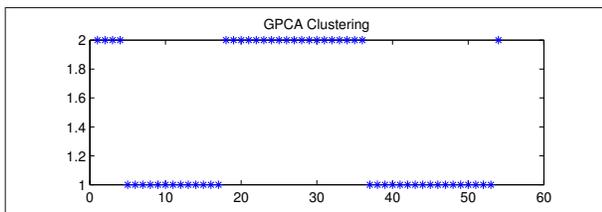


Fig. 5. Activity segmentation via GPCA.

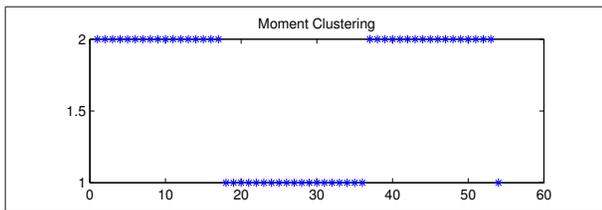


Fig. 6. Activity segmentation via moments-based method.

## VI. CONCLUSIONS

This paper considered the problem of identifying switched linear systems from input/output data and minimal *a priori* assumptions (bounds on the order and number of the subsystems, and on the magnitude of the noise). While this problem is known to be generically NP-hard, we showed that efficient convex relaxations can be obtained by recasting it into a moments optimization form. The effectiveness of this relaxation and its robustness to noise was illustrated using both an academic example and a non-trivial segmentation problems arising in computer vision. Research currently under way seeks to extend these results to the case of multi-output models, with the main difficulty here being the current lack of finite-dimensional characterizations of the moments of multivariate probability distribution functions.

## REFERENCES

- [1] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic Control*, 50(10):1567–1580, 2005.
- [2] E. J. Candes, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted  $\ell_1$  minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.
- [3] R.E. Curto and Fialkow L.A. Recursiveness, positivity, and truncated moment problems. *Houston J. Math*, 17:603–635, 1991.
- [4] M. Fazel, H. Hindi, and S. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *American Control Conference*, 2003.
- [5] M. G. Krein and A. A. Nudelman. *The Markov Moment Problem and Extremal Problems*, volume 50 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1977.
- [6] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optimization*, 11:796–817, 2001.
- [7] Y. Ma and R. Vidal. A closed form solution to the identification of hybrid arx models via the identification of algebraic varieties. In *Hybrid Systems Computation and Control*, pp. 449–465, March 2005.
- [8] N. Ozay, M. Sznajder, C. Lagoa, and O. Camps. A sparsification approach to set membership identification of a class of affine hybrid systems. In *Proc. 47<sup>th</sup> IEEE Conf. Dec. Control*, pp. 123–130, 2008.
- [9] S. Paoletti, A. Juloski, G. Ferrari-Trecate, and R. Vidal. Identification of hybrid systems: A tutorial. *European Journal of Control*, 13(2):242–260, 2007.
- [10] J. Roll, A. Bemporad, and L. Ljung. Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40:37–50, 2004.

- [11] J.A. Shohat and J.D. Tamarkin. *The Problem of Moments*. Math. Surveys I. American Mathematical Society, Providence, RI, 1943.
- [12] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca). *PAMI*, 27(12):1945–1959, December 2005.
- [13] R. Vidal, S. Soatto, and S. Sastry. An algebraic geometric approach to the identification of linear hybrid systems. In *IEEE Conference on Decision and Control*, pages 167–172, December 2003.

## APPENDIX

### A. Recovering the parameters of the model

Here we recall, for ease of reference, the polynomial differentiation based procedure proposed in [13] to recover the parameters of the model once  $\mathbf{c}_s$  is computed. The derivative of  $p_s(\mathbf{r})$  at a point  $\mathbf{r}$  is given by

$$Dp_s(\mathbf{r}) = \frac{\delta p_s(\mathbf{r})}{\delta \mathbf{r}} = \sum_{i=1}^s \prod_{j \neq i} (\mathbf{b}_j^T \mathbf{r}) \mathbf{b}_i \quad (18)$$

Since  $\mathbf{b}_i^T \mathbf{r} = 0$  when  $\mathbf{r}$  is generated by the  $i^{\text{th}}$  submodel (i.e.  $\sigma_t(\mathbf{r}) = i$ ), it follows from (18) that the parameter vector is given by:

$$\mathbf{b}_i = \frac{Dp_s(\mathbf{r})}{\mathbf{e}^T Dp_s(\mathbf{r})} \Big|_{\sigma_t(\mathbf{r})=i} \quad (19)$$

where  $\mathbf{e}^T = [1, 0, \dots, 0]$ .

Since, in general, the association of data points with submodels  $\sigma_t(\mathbf{r})$  is unknown, one can use the following heuristic function, suggested in [12], to choose one point from each submodel  $\{\mathbf{r}_{t_i}\}_{i=1}^s$ :

$$\mathbf{r}_{t_{i-1}} = \operatorname{argmin}_{\mathbf{r}_t: Dp_s(\mathbf{r}_t) \neq 0} \frac{\frac{|p_s(\mathbf{r}_t)|}{\|Dp_s(\mathbf{r}_t)\|} + \delta}{\left| (\mathbf{b}_1^T \mathbf{r}_t) \cdots (\mathbf{b}_s^T \mathbf{r}_t) \right| + \delta} \quad (20)$$

where  $\delta > 0$  is a small number to avoid division by zero.

Finally, given the parameter vectors  $\{\mathbf{b}_i\}_{i=1}^s$ , the mode signal can be computed as follows:

$$\sigma_t = \operatorname{argmin}_{i=1, \dots, s} (\mathbf{b}_i^T \mathbf{r}_t)^2. \quad (21)$$