Hybrid System Identification: An SDP Approach

C. Feng*, C. M. Lagoa*, N. Ozay** and M. Sznaier**

Abstract— **The problem of identifying discrete time affine hybrid systems with noisy measurements is addressed in this paper. Given a finite number of measurements of input/output and a bound on the measurement noise, the objective is to identify a switching sequence and a set of affine models that are compatible with the a priori information, while minimizing the number of affine models. While this problem has been successfully addressed in the literature if the input/output data is noise-free or corrupted by process noise, results for the case of measurement noise are limited, e.g., a randomized algorithm has been proposed in a previous paper [3]. In this paper, we develop a deterministic approach. Namely, by recasting the identification problem as polynomial optimization, we develop deterministic algorithms, in which the inherent sparse structure is exploited. A finite dimensional semi-definite problem is then given which is equivalent to the identification problem. Moreover, to address computational complexity issues, an equivalent rank minimization problem subject to deterministic LMI constraints is provided, as efficient convex relaxations for rank minimization are available in the literature. Numerical examples are provided, illustrating the effectiveness of the algorithms.**

I. INTRODUCTION

In recent years, considerable effort has been put in the problem of identification of hybrid systems. In general, a hybrid system is a system whose behavior is determined by switching dynamics. These systems arise in many different contexts, for examples, circuit network, biological systems, systems with interaction with logic devices and continuous processes. In addition, they can be used to approximate nonlinear dynamics. Thus, due to the potential application to a vast set of practical problems, the problem of identifying input/output hybrid models has attracted considerable attention, and several approaches have been developed.

For the identification problem of piecewise affine (PWA) systems, there are many results available in the literature. One may refer to a thorough review [12] for a summary of recent developments. In the case where measurements are noise-free, an algebraic procedure, known as Generalized Principal Component Analysis (GPCA), has been proposed in [8], [15] to efficiently solve the problem. The problem can be also formulated as a mixed linear integer optimization problem [14] or in terms of linear complementary inequalities [1], leading to generically NP-hard problems. More recently, a greedy algorithm has been proposed to identify

the system while minimizing the number of switches [10]. For robust identification of PWA systems subject to process noise, an efficient moment-based convex approach using convex relaxations on rank minimization has been proposed in [9]. A similar approach was also pursued in [11] to solve a different problem: segmenting a collection of noisy measurements into subspaces. However, to the best of our knowledge, for the case of measurement noise, the only result available is the approach proposed in [3] where a randomized algorithm is provided based on sparse polynomial optimization.

In this paper, we continue the line of research started in [3]. First, we provide an equivalent polynomial optimization problem, which inherently has a sparse structure and satisfies the so-called *running intersection property*. This sparse structure can be used to significantly reduce computational complexity, e.g. see [5], [6], [16]. The reasoning behind the proposed approach bears strong connection to hybrid decoupling used in GPCA. However, the approach in this paper preserves all system parameters as a part of optimizing variables, while GPCA eliminates the structure of the parameters (this is elaborated in Section IV). Moreover, it is shown that the sparse polynomial optimization problem can be solved via a *fixed size* semi-definite program (SDP).

Furthermore, for larger size problems, an equivalent rank minimization problem is provided. In formulating this rank minimization problem, inspired by the results in [9], we use similar tricks to isolate the system parameters from the unknown noise terms, and, hence, eliminate them from the decision variables in the formulated rank minimization problem. A major advantage of the approach in this paper is that the matrix in the objective function is symmetric and of much smaller size than the ones used in [9]. One may note that this feature notably reduces the computational burden in solving convex relaxations of rank minimization problems.

The remainder of the paper is organized as follows. Section II defines notation used and presents some background results related to sparse polynomial optimization. In Section III, we formally define the identification problem of PWA systems in the presence of measurement noise. In Section IV, we reformulate the identification problem as a polynomial optimization problem and show that it has an intrinsically sparse structure. A fixed size SDP is provided which is proven to be equivalent to the identification problem. In Section V, an equivalent rank minimization problem with fixed size LMI constraints is given and a drop rank algorithm is then provided based on a convex relaxation of rank minimization. Illustrative numerical examples are provided in Section VI. Section VII concludes the paper with some final remarks and directions for further research.

II. PRELIMINARIES

A. Notation

x^i abbreviation for $x_i^{i_1} \cdots x_d^{i_d}$ where d is the dimension of the vector x

$\mathbf{E}_\mu[p(x)]$ the mean value of $p(x)$ w.r.t the probability measure μ on the random variable x

$\{m_i\}_0^N$ the moment sequence where $m_i = \mathbf{E}_\mu x^i$ for some probability measure μ and $0 \le i_1 + \ldots + i_d \le N$

$M_N(m)$ the moment matrix in $\mathcal{R}^{\binom{N+d}{N} \times \binom{N+d}{N}}$ constructed by $\{m_i\}_0^{2N}$

$M \succeq 0$ the matrix M is positive semi-definite

$\|x\|_p$ the ℓ_p norm of the vector x, $p = 2$ or ∞

B. General Polynomial Optimization

Consider the following general constrained polynomial optimization problem:

$$p_K^* := \min_{x \in K} p_0(x) \qquad (P1)$$

where $K \subset \mathbb{R}^d$ is a compact semi-algebraic set with nonempty interior defined as

$$K \doteq \{x \colon p_i(x) \ge 0, i = 1, \ldots L\}$$

where $p_i(x)$ are polynomials with total degree d_i. This problem is usually not convex, and hence, hard to solve in general. Yet, let's consider a related problem in the probability measure space:

$$\tilde{p}_K^* := \min_{\mu \in \mathcal{P}(K)} \int p_0(x)\mu(dx) := \min_{\mu \in \mathcal{P}(K)} \mathbf{E}_\mu \left[p_0(x)\right] \quad (P2)$$

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 following result, taken from [5], establishes the equivalence between the two problems:

Theorem 1: Problems (P1) and (P2) are equivalent; that is:

- $\tilde{p}_K^* = p_K^*$.
- If x^* is a global minimizer of (P1), then the Dirac distribution $\mu^* = \delta_{x^*}$ with support on the point x^* is a global minimizer of (P2).
- For every optimal solution μ^* of (P2), $p_0(x) = p_K^*$-μ^* almost everywhere.

One direct consequence of this theorem is that, it is possible to develop a *convergent* sequence of LMI based convex relaxations to problem (P1), where the optimization variables are $m_i \doteq \mathbf{E}_\mu x^i$, the moments of the unknown distribution μ.

If in addition $p_0(x) - p_K^*$ has a Sum-of-Squares (SOS) representation on K, i.e.,

$$p_0(x) - p_K^* = t_0^2(x) + \sum_{i=1}^{L} p_i(x) t_i^2(x) \qquad (1)$$

for some polynomial $t_0(x)$ of degree at most N and some polynomials $t_i(x)$ of degree at most $N - d_i/2$, then, it

is possible to construct an *equivalent* LMI based convex optimization problem for P2. To this effect, let

$$p_N^* = \min_m \sum_\alpha p_{0,\alpha} m_\alpha \qquad (2)$$

$$\text{s.t.} \quad M_N(m) \succeq 0,$$
$$M_{N_i}(p_i m) \succeq 0, i = 1, \ldots, d,$$

where $p_{0,\alpha}$ is the coefficient of x^α in $p_0(x)$; N_i is the smallest integer that no less than $N - d_i/2$; and $M_N(m)$ is the so-called *moment matrix* and $M_{N_i}(p_i m)$ is the so-called *localizing matrix*, both of which are constructed from the truncated moment sequence $\{m_i\}_0^N$. For illustration and clarity of exposition, consider the case where $x \in R^2$, the moment matrix $M_N(m)$ is defined as

$$M_N(m) = \begin{bmatrix} M_{0,0}(m) & M_{0,1}(m) & \cdots & M_{0,N}(m) \\ M_{1,0}(m) & M_{1,1}(m) & \cdots & M_{1,N}(m) \\ \vdots & \vdots & \ddots & \vdots \\ M_{N,0}(m) & M_{N,1}(m) & \cdots & M_{N,N}(m) \end{bmatrix}$$

where

$$M_{j,k}(m) = \begin{bmatrix} m_{j+k,0} & m_{j+k-1,1} & \cdots & m_{j,k} \\ m_{j+k-1,1} & m_{j+k-2,2} & \cdots & m_{j-1,k+1} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k,j} & m_{k-1,j+1} & \cdots & m_{0,j+k} \end{bmatrix}.$$

Note that the size of $M_N(m)$ is $\binom{d+N}{N} \times \binom{d+N}{N}$. The localizing matrix $M_{N_i}(p_i m)$ is defined as

$$M_{N_i}(p_i m)(i,j) = \sum_\alpha p_{i,\alpha} m(\beta(i,j) + \alpha)$$

where $p_{i,\alpha}$ is the coefficient of x^α in $p_i(x)$, $m(i,j)$ is the entry (i,j) of $M_N(m)$ and $\beta(i,j)$ is the subscript of m_β. To illustrate, consider $g_1 = a - x_1 x_2$, then

$$M_1(g_1 m) = \begin{bmatrix} a - m_{11} & am_{10} - m_{22} & am_{01} - m_{12} \\ am_{10} - m_{22} & am_{20} - m_{31} & am_{11} - m_{22} \\ am_{01} - m_{12} & am_{11} - m_{22} & am_{02} - m_{13} \end{bmatrix}$$

By the end, according to Theorem 4.2 in [5], we have

Theorem 2 (General Polynomial Optimization with SOS): In problem (P1), if $p_0(x) - p_K^*$ has the representation form (1), then

$$p_N^* = p_K^*. \qquad (3)$$

C. Sparse Polynomial Optimization

The previous section describes how to build a finite SDP to solve a polynomial optimization problem given that a SOS representation is known to exist at a priori. However, in terms of complexity, it might become computationally intractable if the problem size is large, i.e., the dimension of x is large and/or the degree N is large. Note that the dimension of the moment matrix $M_N(m)$ is $\binom{d+N}{d}$, which grows polynomially in N or d but still very fast, as pointed out in [5], [7], [13]. On the other hand, however, many polynomial optimization problems encountered in practice have a sparse structure that can be exploited to decrease

computational complexity; i.e., the polynomial p_i only contains a small fraction of the overall variables; e.g., see [16]. If the set of indices of variables in each p_i satisfies the so-called *running intersection property*, the size of the LMIs can be significantly reduced, i.e., at most $\binom{N+\xi}{N}$ where ξ is the largest number of variables appearing in each polynomial, e.g., see [6]. We now state the definition of this property.

Definition 1 (Running Intersection Property): Let I_k, $k = 1, \ldots, \tilde{d}$, be the subsets of variables $X \doteq \{x_1, \ldots, x_d\}$ satisfying $\bigcup_{k=1}^{\tilde{d}} I_k = X$. If

  i) *each constraint polynomial $p_i(x)$ uses only variables in I_k for some k;*
  ii) *the objective polynomial can be written as $p_0 = p_{0,1} + \cdots + p_{0,l}$ where each $p_{0,i}$ uses only variables in I_k for some k,*

then running intersection property *is satisfied in Problem* (P1) *if the collection $\{I_1, \ldots, I_{\tilde{d}}\}$ obeys the following condition:*

$$I_{k+1} \cap \left(\cup_{j=1}^{k} I_j\right) \subseteq I_s \text{ for some } s \leq k, \quad (4)$$

for every $k = 1, \ldots, \tilde{d} - 1$.

Similar to the result stated in Theorem 2, for a sparse polynomial optimization problem that satisfies the *running intersection property* and has a sparse SOS representation on K, one can construct a finite SDP as well. For simplicity, we denote $M_N(m, I_k)$ the moment matrix for the reduced variables in the set I_k and denote $M_{N_i}(p_j m, I_k)$ the localizing matrix with the reduced variables in I_k. In the spirit of results in [5], [6], we have the following theorem.

Theorem 3 (Sparse Polynomial Optimization with SOS): *Assume that* (P1) *satisfies Running Intersection Property and let*

$$p_N^* = \min_{m} \quad \sum_{\alpha} p_{0,\alpha} m_\alpha \quad (5)$$
$$s.t. \quad M_N(m, I_k) \succeq 0, \quad k = 1, \ldots, \tilde{d}$$
$$M_{N_i}(p_i m, I_{k(i)}) \succeq 0, \quad i = 1, \ldots, d,$$

where $p_i(x)$ contains only variables in $I_{k(i)}$, and $p_0(x)$ has a sparse *SOS representation on K, i.e.,*

$$p_0(x) - p_K^* = \sum_{k=1}^{L} \left(t_{k,0}^2(x) + \sum_{i=1}^{d} p_i(x) t_{k,i}^2(x) \right) \quad (6)$$

Then,

$$p_N^* = p_K^*. \quad (7)$$

III. Set Membership Identification

In this section, we define the hybrid system identification problem. We consider the problem of identifying single-input-single-output switched linear systems of the form

$$y_k = -\sum_{i=1}^{n} a_i(\sigma_k)(y_{k-i} - e_{k-i}) + \sum_{i=1}^{m} b_i(\sigma_k) u_{k-i} + e_k \quad (8)$$

where u, y and e denote input, output and noise, respectively. The magnitude of noise is bounded by $\bar{e} > 0$. Moreover, a_i and b_i are the parameters of the system, where $\sigma_k \in \{1, 2, \ldots, s\}$ denotes which sub-system is active at time k.

Without any additional restrictions, the problem admits infinitely many solutions. For example, one can assign a trivial model corresponding to each measurement. Thus, one needs to add additional constraints or objectives to make the problem meaningful. In this paper, we aim at minimizing the number of sub-systems and/or minimizing the order of the linear models. For simplicity, let us assume that s, the number of switched sub-systems, is known and (n, m), the order of the linear models, is also known. This assumption does not imply loss of generality since one can always increase s and/or (n, m) one by one until a meaningful solution is found. Then, the problem of interest can be formally stated as follows.

Problem 1: Given input and corrupted output measurements u, y over the interval $[1, L]$, a bound $\bar{e}$ on the ℓ_∞ norm of measurement noise e (i.e. $|e_k| \leq \bar{e}$ for $k \in [1, L]$), the number of sub-models s and the order of sub-models (n and m), find a hybrid affine model of the form (8) that is consistent with all a priori information and the measurement data, or conclude that none exists.

Compared to the switched autoregressive exogenous (SARX) linear models considered in [1], [9], the model (8) assumes that one has measurement noise and does not contain unmodeled system dynamics.[1] In [3], the case with measurement noise is considered and a polynomial optimization problem is formulated for finding a compatible hybrid model, if any, using an algebraic procedure known as GPCA. In the next section, we use a related but different algebraic procedure, with two main advantages. First, the parameters of the linear models are explicitly included in the optimization variables, i.e., once the polynomial optimization problem is solved, the system parameters are known immediately. Hence, there is no need to use parameter recovering algorithms as in GPCA related approaches. Second, it is shown the polynomial optimization problem can be solved via a fixed size SDP program.

IV. Algebraic Reformulation

In this section, based on the so-called *hybrid decoupling constraint* introduced in [15], one can see that equation (8) is equivalent to the polynomial equation

$$p_{0,k}(e, a, b) \doteq \prod_{j=1}^{s} [(y_k + \sum_{i=1}^{n} a_i(j)(y_{k-i} - e_{k-i}) - \sum_{i=1}^{m} b_i(j) u_{k-i} - e_k)] = 0 \quad (9)$$

which holds for all k. Then, the identification problem is equivalent to find some admissible noise e and parameters $a_i(j)$ and $b_i(j)$, so that $p_k(e) = 0, k = 1, \ldots, L$.

To address this issue, we consider the following polynomial optimization problem.

---

[1]This formulation can be easily modified to include unmodeled dynamics.

Problem 2: Given the number of the sub-systems s and the order of linear models (n, m), find

$$\min_{e,a,b} \quad p_0(e, a, b) \qquad (10)$$

$$s.t. \qquad \|e\|_\infty \leq \bar{e},$$

where

$$p_0(e, a, b) = \sum_{k=1}^{L} p_{0,k}(e, a, b)^2.$$

The equivalence between the above problem and Problem 1 is established in the following theorem.

Theorem 4: Given the number of the sub-systems s and the order of the sub-models (n, m), if there exists at least one compatible hybrid affine model for Problem 1, then there exist noise e and parameters (a, b) so that the minimum of Problem 2 is zero. The converse is also true.

Proof: If there is a compatible hybrid model, then, given the true values of parameters and noise, one have $p_{0,k}(e, a, b) = 0$ for all k. Hence, $p_0 = 0$. Since p_0 is a SOS, $p_0 \geq 0$, hence, the minimum of (10) is zero. Conversely, if $p_0 = 0$ for some e, a, b, then $p_{0,k}(e, a, b) = 0$ for all k. Since $p_{0,k}$ is a product of s polynomials, then one of them is equal to zero. Denote the index of such a polynomial by $\sigma(k)$. Hence, a system with a, b being its parameters is a compatible model, and $\sigma(k)$ is the index of the active sub-model at time k. $\square$

Remark 1: As mentioned before, there is a connection between the formulation above and GPCA based approaches. There are, however, substantial differences between these two. In GPCA, a algebraic procedure is used to construct the so-called *Veronese* matrix. The related problem is to find the null space of the matrix and to *extract* system parameters from the null vector, e.g., see [15]. On the other hand, in our procedure, the system parameters are a part of the optimization variables. Hence, a necessary and sufficient condition is derived in Theorem 4; and once the optimization problem is solved, the parameters are determined immediately.

Although Problem 1 and Problem 2 are equivalent, at first glance it seems that the reformulation is at least equally difficult to solve, since (10) contains variables a, b and e where the dimension of e is equal to the number of measurements. However, if one carefully checks the structure of the polynomial p_0, it can be found that it is sparse. In fact, p_0 is a sum of squares of $p_{0,k}$ and for each k, $p_{0,k}$ is a polynomial of variables a, b and $e_{k-n}, \ldots, e_k$. Hence, problem (10) inherently satisfies the *running intersection property* defined in Definition 1, with

$$I_k = \{a, b, e_k, \ldots, e_{k+n}\}, \quad k = 1, 2, \ldots, L. \qquad (11)$$

Moreover, p_0 is a SOS already, implying that, if the minimum of (10) is zero, $p_0 - p_0^*$ can be represented in the form of (6); i.e., one just needs to take $t_{k,0} = p_{0,k}$ and $t_{k,i} = 0$. Consequently, Problem 2 can be solved via a fixed and finite sized SDP. This is summarized as follows.

Theorem 5: Given the number of the sub-systems s and the order of the sub-models (n, m), consider the following optimization problem

$$p_s^* = \min_m \quad \sum_\alpha p_{0,\alpha} m_\alpha \qquad (12)$$

$$s.t. \quad M_N(m, I_k) \succeq 0, \quad k = 1, \ldots, L$$

$$M_{N_i}(p_i m, I_{\beta(i)}) \succeq 0, \quad i = 1, \ldots, L + n,$$

where $I_{\beta(i)} = \{a, b, e_{\beta(i)}, \ldots, e_{\beta(i)+n}\}$ *is the partial variable set for* p_i, $\beta(i) = i$ *for* $i = 1, \ldots, L$ *and* $\beta(i) = L$ *for* $i = L + 1, L + n$; $N = 2s$ *and* $N_i = 2s - 1$ *for* $i = 1, \ldots, L + n$; $p_i = \bar{e}^2 - e_i^2, i = 1, \ldots, L + n$. *Then, if there exists at least one compatible hybrid affine model for Problem 1, $p_s^* = 0$ is the optimum of* (12). *Conversely, if $p_s^* = 0$, rank $M_N(m, I_k) = $ rank $M_{N_i}(m, I_k)$ for all k, and rank $M_N(m, I_k \cap I_j) = 1$ for all pairs (j, k) with $I_k \cap I_j \neq \emptyset$, then, there exists at least one compatible model.*

Proof: This is a direct consequence of Theorem 3, given the fact that running intersection property holds for the collection of the variable sets I_k defined in (11) and the fact that p_0 has a representation of (6) by taking $t_{k,0} = p_{0,k}$ and $t_{k,i} = 0$. $\square$

Remark 2: The rank condition rank $M_N(m, I_k) = $ rank $M_{N_i}(m, I_k)$ and rank $M_N(m, I_k \cap I_j) = 1$ is a sufficient condition to guarantee that the optimum of the SDP relaxation is the same to the one of the corresponding polynomial optimization problem, see e.g. [4], [6]. With this rank condition being satisfied, an algorithm is given in [4], which can always extract an optimal moment sequence corresponding to a probability measure with point support.

Remark 3: In the case where both input noise and output noise are considered, the results above still apply. One just need to add more variables for input noise terms in the polynomial optimization problem, which has a similar structure as shown above.

As illustrated above, by taking into account the sparsity, the optimization problem can be solved if its size is relatively small. One important observation is that the complexity is proportional to the number of measurements, i.e. if one fixes the structure of the hybrid system, the maximum size of LMIs in the SDP remains the same and the number of LMIs increases proportionally. Hence, if the hybrid system is relatively simple, the identification problem can be solved via solving (12) directly. However, if the hybrid system becomes more complex, say, the number of sub-systems increases or the order of the sub-models increases, the problem becomes numerically difficult to solve. This comes from the fact that the size of I_k is equal to $n_\nu = n + 1 + s(n + m)$, and the max size of the moment matrices is equal to $\binom{n_\nu + 2s}{2s}$.

To overcome this numerical difficulty, motivated by GPCA related algorithms, we formulate an equivalent rank minimization problem by using quadratic functions of rows of the Veronese matrix to eliminate the system parameters a, b in the optimization problem. This is described in the next section.

V. An Equivalent Rank Minimization Problem

It can be seen that the main computational difficulty comes from the fact that all system parameters are a part of the

decision variables in the polynomial optimization. Thus, if the structure of the hybrid system is too complex, one may need large computational power to solve (12).

In this section, we show that, by manipulating the rows of the noisy Veronese matrix, a rank minimization problem with finite LMIs can be formulated and it is equivalent to the original identification problem. Though rank minimization problem is in general NP-hard, efficient relaxations are available especially for symmetric positive semi-definite matrix, which is the case encountered.

We now review GPCA approach which provides the motivation for the results presented. That is, one can write (9) into matrix forms, in which the noise terms e_k and the parameters a, b are isolated. Note that the information of the parameters is included in the null vector of the Veronese matrix. A similar idea has been used in [3] in the proposed randomized algorithm, where the computational cost is reduced by fixing each part of the variables in the iterations.

First let's recall the construction of Veronese matrix. Consider the polynomial equations (9) from time $k = 1$ to L, by collecting all measurements, one can build the noisy Veronese matrix V_s and polynomial equations as

$$V_s(\mathbf{r}, e)x \doteq \begin{bmatrix} v_s(\mathbf{r}_1, e) \\ \vdots \\ v_s(\mathbf{r}_L, e) \end{bmatrix} x = 0. \tag{13}$$

where x is a vector with elements being polynomial functions of the system parameters. Then, the identification problem is equivalent to finding admissible noise e so that $V_s(\mathbf{r}, e)$ has a non-trivial null space, i.e. finding e and a non-zero vector x so that (13) holds. Once such a null space is found, it can be used to recover the parameters of the sub-systems; e.g. see [15]. This is summarized in the following proposition.

Proposition 1 (Theorem 1, 2 in [15]): Given the number of the sub-systems s, the order of sub-models (n, m), if there exists at least one compatible hybrid affine model for Problem 1, then there exists some admissible noise e such that $V_s(\mathbf{r}, e)$ is rank deficient; and the system parameters a, b can be extracted from its null space.

A simple example is illustrated on how to construct such a noisy Veronese matrix.

Example 1: For $s = 2$ and the order $(n, m) = (1, 1)$, equation 9 can be written that

$$\begin{aligned} (y_k + a_1(y_{k-1} - e_{k-1}) - b_1 u_k - e_k) \cdot \\ (y_k + a_2(y_{k-1} - e_{k-1}) - b_2 u_k - e_k) &= 0. \end{aligned}$$

Hence, the k-th row of the noisy Veronese matrix V_s can be written as

$$v_k(\mathbf{r}, e) = \begin{pmatrix} (y_k - e_k)^2 \\ (y_{k-1} - e_{k-1})(y_k - e_k) \\ -u_k(y_k - e_k) \\ -u_k(y_{k-1} - e_{k-1}) \\ (y_{k-1} - e_{k-1})^2 \\ u_k^2 \end{pmatrix}^T$$

Remark 4: The size of the Veronese matrix is $L \times \binom{n+m+s}{s}$. For each row of the matrix, the highest total degree of the polynomials (in e) is s, the number of sub-systems.

Given the noisy Veronese matrix constructed above, let's define a matrix

$$Q(m) \doteq \sum_{i=1}^{L} \mathbf{E}_\mu \{v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)\} \tag{14}$$

where m is the truncated moment sequence corresponding to probability measure μ. Hence, $Q(m)$ is a symmetric matrix linear in m. Now, we are ready to propose the equivalent rank minimization problem.

Problem 3: Given the number of the sub-systems s, the order of sub-models (n, m), find a rank deficient matrix $Q(m)$ defined in (14) subject to

$$\begin{aligned} M_N(m, I_k) \succeq 0, \quad k = 1, \dots, L, \\ M_{N_i}(p_i m, I_{\beta(i)}) \succeq 0, \quad i = 1, \dots, L + n. \end{aligned}$$

where $I_{\beta(i)} = \{e_{\beta(i)}, \dots, e_{\beta(i)+n}\}$ is the partial variable set for p_i, $\beta(i) = i$ for $i = 1, \dots, L$ and $\beta(i) = L$ for $i = L+1, \dots, L+n$; $N = s$ and $N_i = s-1$ for $i = 1, \dots, L+n$; and $p_i = \bar{e}^2 - e_i^2, i = 1, \dots, L + n$.

Then, we observe that this problem is indeed equivalent to the problem of finding an admissible noise sequence that results in a rank deficient Veronese matrix.

Theorem 6: If there exists an admissible noise sequence e such that the Veronese matrix $V_s(\mathbf{r}, e)$ is rank deficient, then Problem 3 has a feasible solution. Conversely, if Problem 3 has a feasible solution m, and if rank $M_N(m, I_k) =$ rank $M_{N_i}(m, I_k)$ for all k and rank $M_N(m, I_k \cap I_j) = 1$ for all pairs (j, k) with $I_k \cap I_j \neq \emptyset$, then, $V_s(\mathbf{r}, e)$ is rank deficient for some admissible noise e.

Proof: First note that there exists some unit vector x and admissible noise e such that

$$V_s(\mathbf{r}, e)x = 0$$

if and only if

$$\sum_{i=1}^{L} x^T v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)x = 0. \tag{15}$$

Since $v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)$ is positive semi-definite, it is equivalent to finding a unit vector x^* that the minimum of the following problem

$$\min_e \quad (x^*)^T \left[\sum_{i=1}^{L} v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e) \right] x^* \tag{16}$$

$$\text{s.t.} \quad e \in K$$

is zero, where K is defined as

$$K \doteq \{e : p_k \doteq \bar{e}^2 - e_k^2 \geq 0, k = 1, \dots, L + n\}.$$

Note that the objective function in (16) is a SOS in terms of e for any fixed x, hence, it can be represented in the form of (6). Moreover, the running intersection property is

satisfied for (16). Therefore, by Theorem 3, it follows that the minimum of the following problem

$$\min_{m} \quad (x^*)^T Q(m) x^* \tag{17}$$
$$\text{s.t.} \quad M_N(m, I_k) \succeq 0, k = 1, \ldots, L,$$
$$M_{N_i}(p_i m, I_{k(i)}) \succeq 0, i = 1, \ldots, L + n,$$

is zero, where $N = s$, $N_i = s - 1$ for all i, and Q is defined in (14). Since $v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)$ is positive semi-definite, $Q(m)$ is positive semi-definite as well.[2] Hence, we have $(x^*)^T Q(m) x^* = 0$ for some unit vector x^* if and only if the matrix $Q(m)$ is rank deficient.

Conversely, if the rank condition holds, one can always extract a moment sequence m^* corresponding to a probability measure with point support, such that $Q(m^*)$ is rank deficient. Hence, there is an admissible noise e that $\sum_{i=1}^{L} v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)$ is rank deficient. This implies the Veronese matrix is rank deficient for the same noise e, which concludes the proof. $\square$

Although rank minimization is NP-hard, efficient convex relaxations are available. In particular, good approximate solutions can be obtained by using a log-det heuristic that relaxes rank minimization to a sequence of convex problems, e.g., see [2], [9]. Furthermore, as stated in Theorem 6, it suffices to find a rank deficient solution. Thus, we use a modification of log-det heuristic that aims at dropping the rank by one, as illustrated in the following.

Algorithm 1 Drop Rank

Set $X \leftarrow Q(m)$, $X_0 \leftarrow I$, $k \leftarrow 0$.
repeat
  Solve

$$X_{k+1} \leftarrow \quad \arg\min \quad \text{Tr}(X_k + \delta I)^{-1} X$$
$$\text{s.t.} \quad m \in \mathcal{M}$$

  where $\mathcal{M}$ is the feasible set in Problem 3.
  Decompose the symmetric matrix $X_k = T^{-1} D T$.
  Set $\delta \leftarrow \min \text{diag}(D)$.
  Set $k \leftarrow k + 1$.
until a convergence criterion is reached.
return X_k

Remark 5: Given a rank deficient matrix $Q(m^*)$ is found with a moment sequence m^*, a null vector can be easily determined by computing its eigenvectors and eigenvalues. Moreover, by Theorem 6, there exists some admissible noise e that the Veronese matrix $V_s(\mathbf{r}, e)$ is rank deficient. One can extract such a noise value e from the moment sequence m^* using the algorithm introduced in [4]. Once the noise values are estimated, the problem can be converted to the noise free case by plugging the noise estimates in to the Veronese matrix and the system parameters can be computed using the procedure introduced in [8].

[2]Note that $Q(m)$ is p.s.d. since m satisfies the LMI constraints in (17). This comes directly from the fact that the finite moment condition is the dual formulation of the related SOS problem.

VI. NUMERICAL RESULTS

In this section, we present several numerical examples for illustrating the proposed algorithms. In the first example, a simple hybrid system with limited measurements is considered. The identification problem is then solved by solving the SDP problem (12) directly, according to Theorem 5. In the second example, a more complex hybrid model is considered. It is then shown that, the rank minimization relaxation algorithm in Algorithm 1 works efficiently on identifying the system parameters. A few comments is also given regarding how the noise bound may affect identification results.

A. Example I: via Sparse Polynomial Optimization

In this example, we consider a hybrid linear switching system with $s = 3$ and $(n, m) = (1, 0)$, where the sub-models are

$$y_k = -0.9(y_{k-1} - e_{k-1}) + u_{k-1} + e_k \quad \text{(Submodel 1)}$$
$$y_k = -0.5(y_{k-1} - e_{k-1}) + u_{k-1} + e_k \quad \text{(Submodel 2)}$$
$$y_k = +0.7(y_{k-1} - e_{k-1}) + u_{k-1} + e_k \quad \text{(Submodel 3)}$$

and the hybrid system is modeled as (8), i.e.,

$$y_k = -a_1(\sigma_k)(y_{k-1} - e_{k-1}) + u_{k-1} + e_k,$$

where $\sigma_k \in \{1, 2, 3\}$ depending on which sub-model is active at time k. In the simulation, we set $\sigma(k) = 1$ for $k = [1, 5]$, $\sigma(k) = 2$ for $k = [6, 10]$ and $k = [16, 20]$, $\sigma(k) = 3$ for $k = [11, 15]$. The experimental data was obtained with unit step input and with uniformly randomly generated noise bounded by $\bar{e}$. First we set the noise bound $\bar{e} = 0.1$ and then increase the bound by setting $\bar{e} = 0.3$. The identification problems are solved based on Theorem 5 via solving a SDP relaxation built from the equivalent sparse polynomial optimization problem.

The parameter values used in the simulation and their estimated values with respect to different noise bounds are illustrated in Table I.

TABLE I
ESTIMATED AND TRUE VALUES OF PARAMETERS

		True	$\bar{e} = 0.1$	$\bar{e} = 0.3$
Submodel 1	a_1	0.9000	0.9352	0.5825
Submodel 2	a_1	0.5000	0.4867	0.5825
Submodel 3	a_1	-0.7000	-0.6841	-0.6652

Moreover, the active sub-system at time k can be determined once the sub-models are identified. The results are shown in Figure I.
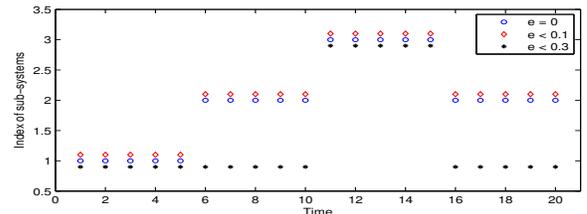

Fig. 1.　Identified and true active sub-systems v.s. time

Remark 6: As one can see from Table I and Figure I, if the noise bound is large, there are compatible hybrid systems with a smaller number of sub-systems. However, one should note that this is not a failure of our algorithm. In fact, the hybrid system obtained is still compatible with all the measurements and a priori information. This should not be surprising since when the measurements are limited, the difference between model dynamics could be covered by the noise, especially when the noise is large. Moreover, when the noise is large, one may get a compatible hybrid system with linear sub-models of lower orders.

B. Example II: via Rank Minimization

In this example, we consider a more complex hybrid system to show the computational efficiency of using Theorem 6 and Algorithm 1. It is assumed that there are three sub-systems with the order $(n, m) = (2, 1)$. That is,

$$y_k = -a_{1,i}(y_{k-1} - e_{k-1}) - a_{2,i}(y_{k-2} - e_{k-2}) + b_{1,i}u_{k-1} + e_k$$

where $i \in \{1, 2, 3\}$ depending on which sub-system is active at time k We take 120 measurement with 11 switches among these three sub-systems. The simulation is run for two different noise levels, i.e. $\bar{e} = 0.1$ and $\bar{e} = 0.3$. The input signal u is uniformly randomly generated between -1 and 1. In Table II, we present the values of the system parameters identified by the algorithm for the noise levels considered.

TABLE II

ESTIMATED AND TRUE VALUES OF PARAMETERS

		True	Noise $e < 0.1$	Noise $e < 0.3$
Submodel 1	a_1	0.9000	0.9183	0.9586
	a_2	0.1800	0.1736	0.2379
	b_1	0.2000	0.2310	0.2863
Submodel 2	a_1	0.5000	0.5155	0.5494
	a_2	0.0600	0.0564	0.1014
	b_1	1.0000	1.1310	1.0802
Submodel 3	a_1	-1.0000	-0.9716	-0.9618
	a_2	0.3000	0.3144	0.3287
	b_1	0.6000	0.6369	0.6991

Moreover, the active sub-system at time k can be determined once the sub-models are identified. The results are shown in Figure I.
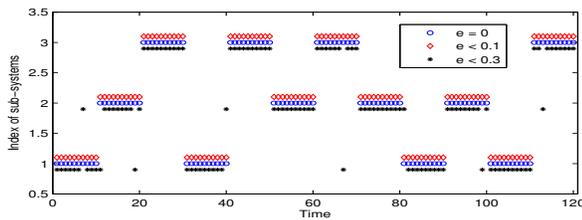

Fig. 2. Identified and true active sub-systems v.s. time

VII. CONCLUDING REMARKS

This paper addresses the identification problem of discrete-time affine hybrid systems with input/output data corrupted by measurement noise. The proposed approach first formulates the identification problem as a polynomial optimization problem. It is shown that the optimization problem obtained inherently has a sparse structure, which can be used to significantly reduce the size of the SDPS, and, hence, reduce computational cost. Moreover, since the objective function has a SOS representation, the relaxation is equivalent to the original problem (size of relaxation is determined exclusively by the number of sub-models). Furthermore, to address the computational difficulties when the hybrid system is complex, a drop rank approach is given to solve the original identification problem. Though rank minimization problem is in general NP-hard, efficient convex relaxations are available in the literature. Numerical examples are provided to illustrate the effectiveness of the proposed algorithm.

It is shown in the paper, by isolating system parameters from the decision variables in the optimization problem, the computational cost can be substantially reduced. However, the structure information of system parameters is then lost. Hence, ongoing work is aimed at developing approaches to separate the search of admissible noise and systems parameters while preserving the structure information.

REFERENCES

[1] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Trans. Automat. Contr.*, 50(10):1567–1580, 2005.
[2] M. Fazel, H. Hindi, and S. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proc. Amer. Contr. Conf.*, 2003.
[3] C. Feng, C. M. Lagoa, and M. Sznaier. Hybrid system identication via sparse polynomial optimization. In *Proc. Amer. Contr. Conf.*, 2010.
[4] D. Henrion and J. B. Lasserre. Detecting global optimality and extracting solutions in gloptipoly. Technical report, LAAS-CNRS, 2005.
[5] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2001.
[6] Jean B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM J. on Optimization*, 17(3):822–843, 2006.
[7] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. *Emerging applications of algebraic geometry*, 149:157–270, 2009.
[8] Yi Ma and René Vidal. Identification of deterministic switched arx systems via identification of algebraic varieties. In *Hybrid Systems: Computation and Control*, pages 449–465, 2005.
[9] N. Ozay, C. Lagoa, and M. Sznaier. Robust identification of switched affine systems via moments-based convex optimization. In *Proc. IEEE Conf. Dec. Contr.*, 2009.
[10] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps. A sparsification approach to set membership identification of a class of affine hybrid systems. In *Proc. IEEE Conf. Dec. Contr.*, pages 123–130, 2008.
[11] N. Ozay, M. Sznaier, C. M. Lagoa, and O. Camps. GPCA with denoising: a moment-based convex approach. submitted to IEEE Conference on Computer Vision and Pattern Recognition, 2010.
[12] S. Paoletti, A. Juloski, G. Ferrari-trecate, and R. Vidal. Identification of hybrid systems: a tutorial. *European Journal of Control*, 13(2):123–130, 2007.
[13] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2, Ser. B):293–320, 2003.
[14] J. Roll, A. Bemporad, and L. Ljung. Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1):37 – 50, 2004.
[15] R. Vidal, S. Soatto, Y. Ma, and S. Sastry. An algebraic geometric approach to the identification of a class of linear hybrid systems. In *Proc. IEEE Conf. Dec. Contr.*, pages 167–172, 2003.
[16] H. Waki, S. Kim, M. Kojima, and M. Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. on Optimization*, 17(1):218–242, 2006.