2010 American Control Conference
Marriott Waterfront, Baltimore, MD, USA
June 30-July 02, 2010

Hybrid System Identification via Sparse Polynomial Optimization

Chao Feng*, Constantino M. Lagoa* and Mario Sznaier**

Abstract— **In this paper, the problem of identifying discrete time affine hybrid systems with measurement noise is considered. Given a finite collection of measurements and a bound on the noise, the objective is to identify a hybrid system with the smallest number of sub-systems that is compatible with the a priori information. While this problem has been addressed in the literature if the input/output data is noise-free or corrupted by process noise, it remains open for the case of measurement noise. To handle this case, we propose a new approach based on recasting the problem into a polynomial optimization form and exploiting its inherent sparse structure to obtain computationally tractable problems. Combining these ideas with a randomized Hit and Run type approach leads to further computational complexity reduction, allowing for solving realistically sized problems. Numerical examples are provided, illustrating the effectiveness of the algorithm and its potential to handle large size problems.**

I. Introduction

A hybrid system is a system whose behavior is determined by switching dynamics. These systems arise in many different contexts, for example, circuit network, biological systems, systems with interaction with logic devices and continuous processes, and in addition, they can be used to approximately model nonlinear dynamics. Due to the potential application to a vast set of practical problems, the problem of identifying input/output hybrid systems models from experimental data has attracted considerable attention.

There are many results available in the literature concerning the identification of piecewise affine switched systems (PWAS); see the thorough review [9] for a summary of recent developments. For example, in [12], [6],the identification problem can be solved efficiently by an algebraic procedure, known as Generalized Principal Component Analysis (GPCA), for-noise free data. The problem can be also formulated as a mixed linear integer optimization problem as shown in [1], [11]. More recently, a greedy algorithm has been proposed to identify the system while minimizing the number of switches among sub-systems [8]. Finally, [7] proposed a convex optimization algorithm based on moment conditions for robust identification of PWAS subject to process noise. While these methods are effective in many situations, to the best of our knowledge, the identification of input/output piecewise affine systems subject to both process and measurement noise has not been addressed.

This work was supported by the National Science Foundation under grants CMMI-0838906, ECCS-0731224 and ECCS-0501166.

EE Department, Pennsylvania State University, University Park, PA 16802, USA (e-mail: feng@psu.edu, lagoa@engr.psu.edu).

**ECE Department, Northeastern University, Boston, MA 02115, USA (m.sznaier@neu.edu)

Motivated by this fact and by the approach introduced in [7], in this paper we propose a randomized algorithm aim at solving the PWAS identification problem when data is corrupted by measurements noise. It is based on the results on sparse polynomial optimization [13], [5] and randomized techniques [2], [3]. The starting point of the proposed method is the algebraic procedure introduced in [12], [6] for noiseless data. As shown there, the parameters of the system can be recovered from the null space of a so-called Veronese matrix built from the input/output data. In the case under consideration here, the entries of this matrix depend polynomially on the unknown noise. Thus, the *a priori* information and the experimental data are consistent if and only if there exists an admissible noise sequence that renders this matrix rank deficient. As we show in the paper, the problem of finding such a sequence and the corresponding null space from a "noisy" Veronese matrix can be reformulated as a polynomial optimization problem. In principle, these problems can be solved by constructing a hierarchy of semi-definite program (SDP) relaxations [4], [10]. However, the dimension of these SDP relaxations and the entailed computational complexity is not small, limiting the size of problems that can be handled. To circumvent this difficulty, we exploit the fact that the reformulated problem has an inherently sparse structure, and hence recent results can be exploited to substantially reduce its computational complexity, allowing for solving medium size problems. Finally, to handle larger problems, we propose an algorithm that combines sparse polynomial optimization and randomized Hit-and-Run ideas.

The remainder of the paper is organized as follows. Section II defines notation used and presents background results related to sparse polynomial optimization. In Section III, we formally define the problem and review the algebraic approach to identification of (noiseless) PWAS systems. In Section IV, we reformulate the identification problem as a polynomial optimization problem and show that it has an intrinsically sparse structure. We further provide an analysis of the computational complexity of the method and develop a Hit-and-Run type algorithm aimed at decreasing the computational burden. Illustrative numerical examples are provided in Section V. Section VI concludes the paper with remarks, future work and some open questions for future research.

II. Preliminaries

For the reader's convenience, in this section we define the notation used and briefly summarize some results on polynomial optimization that are used in the proposed algorithm. For a more detailed exposition on (sparse) polynomial optimization, the reader is referred to [4], [13], [5].

A. Notation

x^i	abbreviation for $x_i^{i_1}\cdots x_d^{i_d}$
$\mathbf{E}_\mu[p(x)]$	the mean value of $p(x)$ w.r.t the probability measure μ on the random variable x
$\{m_i\}_{i=0}^N$	a moment sequence where $m_\mathbf{i} = \mathbf{E}_\mu x^\mathbf{i}$
$M_N(m)$	the moment matrix constructed by $\{m_i\}_{i=0}^{2N}$
$M \succeq 0$	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:

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

where $K \subset \mathbb{R}^\mathbb{N}$ is a compact set. This problem is usually non-convex, hence hard to solve. Next, consider the related problem:

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

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 [4], establishes the relation 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 $\mu^* = \delta_{x^*}$ 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 in the case of semi-algebraic constraint sets, e.g $K = \{x\colon p_i(x) \geq 0, i = 1,\dots d\}$, where p_i are polynomials with total degree d_{p_i}, 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. To this effect, let

$$\begin{aligned} p_N^* = \min_m \quad & \sum_\alpha p_{0,\alpha} m_\alpha \\ \text{s.t.} \quad & M_N(m) \succeq 0, \\ & M_{N_i}(p_i m) \succeq 0, i = 1,\dots,d, \end{aligned} \tag{1}$$

where $p_{0,\alpha}$ is the coefficient of x^α in $p_0(x)$; $M_N(m)$ 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 moments m_i. For instance, in the case where $x \in R^2$, the moment matrix $M_N(m)$ consists of the block matrix $\{M_{j,k}\}_{0\leq j,k\leq N}$ defined by

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

and

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

The localizing matrix $M_{N_i}(g_i m)$ is defined as

$$M_{N_i}(g_i m)(i,j) = \sum_\alpha g_{i,\alpha} m_{\{\beta(i,j)+\alpha\}}$$

where $g_{i,\alpha}$ is the coefficient of x^α in $g_i(x)$, $m(i,j)$ is the entry (i,j) of $M_N(m)$, $\beta(i,j)$ is the subscript of m_β, and N_i is the smallest integer that no less than $N - d_{p_i}$. Then, according to Theorem 4.2 in [4], we have

Theorem 2 (General Polynomial Optimization):

$$p_N^* \uparrow p_K^*. \tag{2}$$

as N increases to infinity,

C. Sparse Polynomial Optimization

While a hierarchy of SDP relaxations can be built to asymptotically solve a general polynomial optimization problem, in terms of computational complexity, it can be applied only to relatively small size problems. Indeed, according to [4], [10], the largest size of the LMIs in the SDP is $\binom{d+N}{d}$. On the other hand, many polynomial optimization problems encountered in practice (such as the one considered in this paper) 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 indeterminate variables. If the sets of indices of variables in each p_i and in the objective p_0 satisfy the so-called *running intersection property*, the size of the LMIs can be significantly reduced, i.e. $\binom{\xi+N}{\xi}$ where ξ is the largest number of variables appearing in the polynomials. We now briefly state the definition of this property.

Definition 1 (Running Intersection Property): Let I_k be the set of indices of the variables that appear in p_k, $k = 1,\dots,d$, and write $p_0 = p_{0,1} + \cdots + p_{0,l}$ where each $p_{0,i}$ uses only variables $\{X_i | i \in I_k\}$ for some k. Then, the running intersection property *is satisfied if the collection $\{I_1,\dots,I_d\}$ obeys the following condition:*

$$I_{k+1} \cap \bigcup_{j=1}^{k} I_j \subseteq I_s \text{ for some } s \leq k, \tag{3}$$

for every $k = 1,\dots,d-1$.

Similar to the result stated in Theorem 2, for a sparse polynomial optimization problem that satisfies the *running intersection property*, the convergence property holds as well. For simplicity, let us denote $M_N(m, I_k)$ the moment matrix for the reduced variables in the set I_k, and denote $M_{N_i}(g_j m, I_k)$ the localizing matrix with the reduced variables in I_k. We now restate Theorem 3.6 in [5] that formalize this approach.

Theorem 3 (Sparse Polynomial Optimization): Let

$$\begin{aligned} p_N^* = \min_m \quad & \sum_\alpha p_\alpha m_\alpha \\ \text{s.t.} \quad & M_N(m, I_k) \succeq 0, \quad k = 1, \dots, d \\ & M_{N_i}(g_j m, I_k) \succeq 0, \quad k = 1, \dots, d. \end{aligned} \tag{4}$$

Then, as N increases to infinity,

$$p_N^* \uparrow p_K^*. \tag{5}$$

III. Set Membership Identification of Linear Hybrid Systems

In this section, we formally define the hybrid system identification problem and formulate it as a constrained polynomial optimization. More precisely, we consider the problem of identifying single-input-single-output switched linear systems of the form

$$y_k = \sum_{i=1}^{n} a_i(\sigma_k)(y_{k-i} - e_{k-i}) + \sum_{i=1}^{m} b_i(\sigma_k) u_{k-i} + e_k \tag{6}$$

where u, y and e are 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 $\sigma_k \in \{1, 2, \dots, s\}$ denotes which sub-system is active at time k.

Without additional restrictions, the problem admits infinitely many solutions, i.e., one can assign a trivial model corresponding to each measurement. Thus, one needs to add additional objectives to make the problem meaningful. In this paper, we aim at minimizing the number of sub-systems, denoted by s. For simplicity, assume that s is known. This assumption does not imply loss of generality since one can always increase s one by one until a meaningful solution is found. Then, the problem of interest can be 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 (6) that is consistent with all a priori information given.

Compared to the switched autoregressive exogenous (SARX) linear models considered in [1], [7], the model (6) assumes that one has measurement noise and does not contain unmodeled system dynamics.[1] We are now ready to formulate the identification problem as finding the null space of a a so-called "noisy" Veronese matrix.

A. Algebraic Reformulation

In the algebraic procedure, known as GPCA introduced in [12], [6], if the noise $e = 0$, equation (6) can be equivalently represented as

$$\mathbf{b}(\sigma_k)^T \mathbf{r}_k = 0$$

where $\mathbf{r}_k = [-y_k, y_{k-1}, \dots, y_{k-n}, u_{k-1}, \dots, u_{k-m}$, and $\mathbf{b}(\sigma_k) = [1, a_1(\sigma_k), \dots, a_n(\sigma_k), \dots, b_1(\sigma_k), \dots, b_m(\sigma_k)]$.

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

Based on the so-called *hybrid decoupling constraint*, one can write this as a polynomial equation

$$p_s(\mathbf{r}) = \prod_{i=1}^{s} (\mathbf{b}_i^T \mathbf{r}_k) = \mathbf{c}_s^T v_s(\mathbf{r}_k) = 0, \tag{7}$$

which holds at any time k. Therefore, by collecting all measurements, we can build the so-called *Veronese* matrix V_s and polynomial equations

$$V_s \mathbf{c}_s \doteq \begin{bmatrix} v_s(\mathbf{r}_1)^T \\ \vdots \\ v_s(\mathbf{r}_L)^T \end{bmatrix} \mathbf{c}_s = 0. \tag{8}$$

Hence, the problem is converted into the related problem of finding the null space of the Veronese matrix V_s.

In the case where measurement noise needs to be considered, one can preform a similar reasoning and derive the following polynomial equation.

$$p_s(\mathbf{r}, e) = \prod_{i=1}^{s} (\mathbf{b}_i^T \tilde{\mathbf{r}}_k) = \mathbf{c_s}^T v_s(\mathbf{r}_k) = 0, \tag{9}$$

where

$$\tilde{r}_k = [y_k - e_k, \dots, y_{k-n} - e_{k-n}, u_{k-1}, \dots, u_{k-m}].$$

Then, a "noisy" Veronese matrix $V_s(\mathbf{r}, e)$ can be constructed and 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 that

$$V_s(\mathbf{r}, e) x = 0. \tag{10}$$

Once such a null space is found, it can be used to recover the parameters of the sub-systems; e.g. see [12].

Now we present a simple example to illustrate this procedure and highlight the difficulty of the problem.

Example 1: For $s = 2$ and the order $(n, m) = (1, 1)$, equation (7) can be written that

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

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 1: Note that given the noisy measurements u and y, the Veronese matrix above is a matrix polynomial function of e. Moreover, there is a special structure on how the noise "appears" in the matrix, i.e., e_k appears not only at the k-th row of $V(e)$, but also at the rows $k+1, k+2, \dots, k+n$. Hence, there are elements in the matrix being multivariate polynomials. Note in SARX linear models, each e_k only appears in a single row, which guarantees that every element

in the Veronese matrix is either a constant or a univariate polynomial, motivating the proposed moments-based convex optimization algorithm in [7]. However, this algorithm no longer applies when measurement noise is considered.

IV. Main Results

In this section, we provide the details of our approach to hybrid systems identification. The exposition is divided in two parts. First, we show that an admissible sequence e and a vector x that satisfy (10) can be found by solving a polynomial optimization problem. The sparse structure of the optimization problem is then exploited to obtain computationally tractable relaxations that work well for moderate size problems. Finally, to further decrease computational complexity and to solve larger size problems, we proposed a Hit-and-Run type randomized algorithm.

A. An Equivalent Sparse Polynomial Optimization Problem

Finding the null space of the noisy Veronese matrix is equivalent to minimizing its minimal singular value and determining the corresponding eigenvector. Thus, we reformulate it as a general polynomial optimization problem by introducing an additional unitary vector.

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

$$\begin{aligned} \min_{e,x} \quad & \|V_s(\mathbf{r}, e)x\|_2^2 \\ s.t. \quad & \bar{e}^2 - e_k^2 \geq 0, k = 1, \ldots, L, \\ & \|x\|_2^2 - 1 \geq 0. \end{aligned} \tag{11}$$

It is easy to see this problem equivalent to Problem 1.

Proposition 1: Given the number of the sub-systems s and the order of the sub-models (n, m), if there exists at least one compatible model for Problem 1, then there exist e and x so that the minimum of Problem 2 is zero.

Although these two problems are equivalent, at first glance it seems that the reformulation is at least equally difficult to solve, since it contains variables x and e where the dimension of e is equal to the number of measurements. However, if one carefully checks the structure of the polynomial $\|V_s(\mathbf{r}, e)x\|_2^2$, it can be found that the polynomial is sparse. More specifically, the polynomial can be rewritten as follows.

$$\begin{aligned} p_0 &\doteq \|V_s(\mathbf{r}, e)x\|_2^2 \\ &= \sum_{i=1}^{L} (x^T v_s(\mathbf{r}_i))^2 \\ &= p_{0,1}(x, e_1, \ldots, e_n) + p_{0,2}(x, e_2, \ldots, e_{n+1}) \\ &\quad + \ldots + p_{0,L}(x, e_L, \ldots, e_{l+n}) \end{aligned} \tag{12}$$

where $p_{0,i}$s are polynomials. Hence, the problem (11) satisfies the *running intersection property* defined in Definition 1. Thus, one can construct convergent SDP relaxations of significantly smaller size to solve Problem 2. [2]

[2]These observations still hold true when input noise is considered, where one just need to add more variables for the input noise sequence.

The following simple example illustrates how the sparse structure substantially reduces computational complexity.

Example 2: Consider two sub-systems ($s = 2$) with order $n = 1$ and $m = 1$, and 20 measurements ($L = 20$). Then, in Problem 2, one would have to build one LMI with size $\Theta_1 = \binom{20+6+4}{4} = 27,405$ and 21 LMIs of size $\Theta_2 = \binom{20+6+2}{2} = 378$ for $N = 4$ in the SDP relaxation. On the other hand, by taking into account the sparsity, one could build a SDP with 20 LMIs of size $\Theta_3 = \binom{2+6+4}{4} = 495$, and 21 LMIs of size $\Theta_4 = \binom{2+6+2}{2} = 45$, for the same relaxation order.

Another important observation is that the complexity is proportional to the data length. Hence, if the hybrid system is relative simple, the identification problem can be solved via solving Problem 2 directly. However, if the hybrid system becomes complicate, say, the number of sub-systems increases and/or the order of the sub-models increases, the problem becomes numerically difficult to solve.

To overcome this difficulty, we propose an iterative algorithm. More precisely, we start from a random unit vector x_0, and solve (11) for $x = x_0$ to find e that minimize the objective function. Then, we fix e and let x be the singular vector of $V_s(\mathbf{r}, e)$ corresponding to its minimal singular value. We repeat this iteration until x and e converge. This iteration problem is formally defined as follows.

Problem 3: Given an initial unit vector x_0, find e^* and x^* such that

$$\begin{aligned} e^*(x_0) = \arg \quad \min_{e} \quad & \|V_s(\mathbf{r}, e)x^*\|_2^2 \\ \textit{s.t.} \quad & \|e\|_\infty \leq \bar{e}, \end{aligned} \tag{13}$$

and

$$\begin{aligned} x^*(x_0) = \arg \quad \min_{x} \quad & \|V_s(\mathbf{r}, e^*)x\|_2^2 \\ \textit{s.t.} \quad & \|x\|_2^2 = 1. \end{aligned} \tag{14}$$

It is easy to see that the computational cost is significantly reduced by fixing a part of the unknown variables. On the negative side, however, one may encounter a convergence problem. That is, counterexamples can be constructed where the procedure above converges to some local minimum, depending on the initial vector x_0. Therefore, we propose a randomized algorithm that circumvents this problem, which is described in the next section.

B. A Hit-and-Run Algorithm

In this section, we propose a randomized algorithm to circumvent the local convergence problem.

First of all, let us state the following proposition that helps in the understanding the algorithm.

Proposition 2: Let

$$\gamma = \|V_s(\mathbf{r}, e^*(x_0))x^*(x_0)\|_2^2$$

where $e^*(x_0)$ and $x^*(x_0)$ are the solutions of Problem 3 for some initial unit vector x_0. Then, the minimum of Problem 2 is γ if and only if the minimum of the problem

$$\begin{aligned} \min_{e,\beta,d} \quad & \|V_s(\mathbf{r}, e)(x^* + \beta d)\|_2^2 - \gamma\|x^* + \beta d\|_2^2 \\ \textit{s.t.} \quad & \|e\|_\infty \leq \bar{e}. \end{aligned} \tag{15}$$

is zero.

Proof: We prove it by contradiction. For the "if" part, assume that the minimum of Problem 2 is not γ. Then there exist some noise e and some unit vector $\tilde{x}$ that

$$\|V_s(\mathbf{r}, e)\tilde{x}\|_2^2 < \gamma.$$

Let $d = \tilde{x} - x^*$ and $\beta = 1$, then,

$$\|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 < \gamma.$$

Hence,

$$\begin{aligned}&\|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 - \gamma\|x + \beta d\|_2^2 \\ &\quad = \|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 - \gamma < 0\end{aligned}$$

which is a contradiction.

For the "only if" part, assume the minimum of problem (15) is less than zero for some direction d and some scalar β. Note this minimum is always non-positve, since the objective is always zero if one take $\beta = 0$. Then, let the unit vector

$$\tilde{x} = \frac{x^* + \beta d}{\|x^* + \beta d\|_2},$$

one has

$$\begin{aligned}&\|V_s(\mathbf{r}, e)(x^* + \beta d)\|_2^2 - \gamma\|x^* + \beta d\|_2^2 \\ &\quad = \|x^* + \beta d\|_2^2 \cdot (\|V_s(\mathbf{r}, e)\tilde{x}\|_2^2 - \gamma) < 0\end{aligned}$$

Hence, γ cannot be the minimum of Problem 2, which is a contradiction. □

Thus, solving problem (15) is equivalent to solving Problem 2, which is also a sparse polynomial optimization problem satisfying running intersection property. Now we are ready to state the randomized algorithms, which is summarized in Algorithm 1.

C. Computational Cost Analysis

In Step 1, since x is fixed, the objective polynomial in problem (11) becomes

$$p_0 = p_{0,1}(e_1, \ldots, e_{n+1}) + \ldots + p_{0,L}(e_L, \ldots, e_{L+n}).$$

Hence, for a fixed relaxation order N, the size of the moment matrices is $\binom{n+1+N}{N}$ and the size of localizing matrices is $\binom{n+N-1}{N-2}$, in the constructed SDP. The number of LMIs is $2L+1$. Note also that since the total degree of p_0 is equal to $2s$, it usually does not require a high relaxation order to solve the optimization problem at or very close to the optimality.

Step 2 is a singular value decomposition for a constant matrix. The size of the matrix $V_s(\mathbf{r}, e)^T V_s(\mathbf{r}, e)$ is $\binom{n+m+s}{s}$, which is not related to the data length of measurement data.

In Step 4, to construct a SDP relaxation for a fixed relaxation order N, the size of the moment matrices is $\binom{n+2+N}{N}$ and the size of localizing matrices is $\binom{n+N}{N-2}$. The number of LMIs is $2L + 1$.

By the analysis above, we observe that this algorithm can be applied to complicate problems encountered in practice. However, it is unclear how many H&R iterations are needed to find a singular value close enough to zero, given a solution exists. In our numerical examples, we found the algorithm usually terminates and gives a good estimate in less than 100 iterations.

Algorithm 1 Hit-and-Run Algorithm

Step 0: Generate a random vector x_0 that $\|x_0\| = 1$
Set $x^* \leftarrow x_0$
repeat
repeat
Step 1: Set $x \leftarrow x^*$, solve

$$e^* \leftarrow \arg\min_e \quad \|V_s(\mathbf{r}, e)x\|_2^2 \quad s.t. \quad \|e\|_\infty \le \bar{e}$$

Step 2: Set $e \leftarrow e^*$, solve

$$x^* \leftarrow \arg\min_x \quad \|V_s(\mathbf{r}, e)x\|_2^2 \quad s.t. \quad \|x\|_2^2 = 1$$

until

$$\min_e \ \|V_s(\mathbf{r}, e)x^*\|_2^2 \ = \ \min_x \ \|V_s(\mathbf{r}, e^*)x\|_2^2$$
$$s.t. \quad \|e\|_\infty \le \bar{e} \qquad s.t. \quad \|x\|_2^2 = 1$$

Step 3: Set $\gamma \leftarrow \|V_s(\mathbf{r}, e^*)x^*\|_2^2$
repeat
Step 4: (H&R) Pick a random direction d, solve

$$(e, \beta) \leftarrow \arg\min \quad \|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 - \gamma\|x + \beta d\|_2^2$$
$$\text{s.t.} \quad \|e\|_\infty \le \bar{e}$$

until a convergence criterion is reached
until
return x^*

V. Numerical Results

In this section, we present some numerical results for the proposed randomized algorithm. In the first part, the CPU time needed for solving the sub-problems in Algorithm 1 is given. In the second part, we use an academic example to illustrate the effectiveness of the proposed algorithm.

A. CPU Time of Each Iteration for Different Size Problems

In this example, we run the code in Matlab2007R on a computer with 1.8GHz CPU and 1 GB RAM. Moreover, we used the package provided by Waki et al. [14] to solve the sparse polynomial optimization problem. The CPU time used for formulating the optimization problem and solving the SDP relaxations is listed in the table followed.

N	L	s	Conversion Time (sec)	SeDuMi Time (sec)
2	20	2	2.21	0.19
4	20	2	2.48	1.02
8	20	2	4.51	20.89
2	40	2	8.28	0.31
4	40	2	9.33	2.18
8	40	2	13.29	49.64
2	80	2	112.12	0.60

N: relaxation order; L: number of measurements; s: number of sub-systems.

Remark 2: As shown in the table, if one fixes the relaxation order N, the CPU time for solving the optimization problems is proportional to the number of measurements L. On the other hand, the conversion time (that is, the time used to formulate the sparse optimization problems into standard SeDuMi forms) is long when L is large. However, one should note this formulation can be done symbolically once the structure of the hybrid system is fixed. Thus it can be expected that the algorithm can be applied to handle problems of much larger size, if the coding is optimized,

B. Illustrative Example

In this section, we consider an illustrative example to demonstrate the effectiveness of the proposed algorithm. It is assumed that there are three sub-systems with the order $(n, m) = (2, 1)$. That is,

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

where $i = 1, 2, 3$. We take 90 measurement with 5 switches. The simulation is run for two different noise levels, i.e. $\bar{e} = 0.05$ and $\bar{e} = 0.2$ (note that the GPCA approach usually cannot handle noisy data well, especially for large noise; see [7]). The input signal u is random ranging from -1 to 1. In Table I, we present the values of the system parameters identified by the algorithm for the noise levels considered.

TABLE I

ESTIMATED AND TRUE VALUES OF PARAMETERS

		True	Noise $e < 0.05$	Noise $e < 0.2$
Sub-model 1	a_1	1.8000	1.7642	1.9320
	a_2	0.8100	0.8261	0.7034
	b_1	1.5000	1.5134	1.5820
Sub-model 2	a_1	0.5000	0.4742	0.3394
	a_2	0.0600	0.0436	0.0240
	b_1	1.0000	0.9821	1.3740
Sub-model 3	a_1	-1.4000	-1.3627	-1.1360
	a_2	0.7400	1.7642	0.6722
	b_1	1.0000	1.0461	1.3836

Moreover, the measurements are classified to one of the the sub-models, indicating which one is active at time k. This is shown in Figure I.

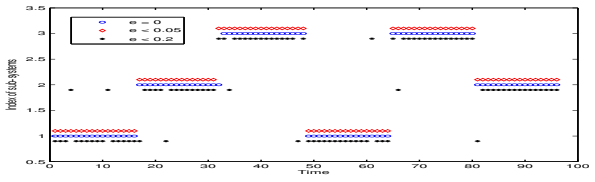

Fig. 1. Identified and true active sub-systems v.s. time

It can be seen that, in the case of small noise, the parameters of the sub-systems can be closely recovered and almost all (except one) measurements are correctly classified. For the case of large noise, one could conclude that the measurements are "misclassified" at several time instants. However, it should be noted that, given the model identified above, the classification obtained is indeed compatible with the measured data and a priori information on the magnitude of the noise. In fact, if the dynamic of sub-models are close and the noise is large, one may get a hybrid system with less sub-models. This should not be regarded as a failure of the algorithm, as the system identified is compatible with all the given information.

VI. Concluding Remarks

This paper addresses the identification problem of discrete-time affine hybrid systems with input/output data corrupted by measurement noise. The approach suggested in the paper formulates the identification problem as a polynomial optimization problem. The inherent sparse structure is then exploited to significantly reduce the computational cost needed to solve its SDP relaxations. Furthermore, a randomized algorithm is proposed (by using ideas from Hit-and-Run) to speed up the computation, in order to be able to handle much more larger sized problems. Numerical examples are provided to illustrate the effectiveness of the proposed algorithm and the potential to attack large size problems.

Hence, ongoing work is aimed at optimizing the code, as discussed in Section V-A. On the other hand, we are also working on finding a bound on the expected number of H&R iterations needed for a given precision.

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] Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556, 2004.

[3] F. Dabbene, P. Shcherbakov, and B. T. Polyak. A randomized cutting plane scheme with geometric convergence: probabilistic analysis and SDP applications. In *Proc. IEEE Conf. Dec. Contr.*, pages 3044–3049, 2008.

[4] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2001.

[5] Jean B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM J. on Optimization*, 17(3):822–843, 2006.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2, Ser. B):293–320, 2003.

[11] J. Roll, A. Bemporad, and L. Ljung. Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1):37 – 50, 2004.

[12] 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.

[13] 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.

[14] H. Waki, S. Kim, M. Kojima, M. Muramatsu, and H. Sugimoto. SparsePOP-a sparse semidefinite programming relaxation of polynomial optimization problems. *ACM Trans. Math. Softw.*, 35(2):1–13, 2008.