

A Convex Optimization Approach to Model (In)validation of Switched ARX Systems with Unknown Switches

Y. Cheng

Y. Wang

M. Sznaier

N. Ozay

C. Lagoa

Abstract—This paper considers the problem of (in)validating switched affine models from noisy experimental data, in cases where the mode-variable is not directly observable. This problem, the dual of identification, is a crucial step when designing controllers using models identified from experimental data. Our main results are convex certificates, obtained by exploiting a combination of sparsification and polynomial optimization tools, for a given model to either be consistent with the observed data or be invalidated by it. These results are illustrated using both academic examples and a non-trivial application: detecting abnormal activities using video data.

I. INTRODUCTION AND MOTIVATION

The problem of identifying switched affine models has been the object of considerable attention during the past few years. Since it has been shown that this problem is generically NP-hard, a large portion of these efforts has been directed towards developing computationally efficient algorithms, leading to a number of techniques based either on heuristics or relaxations (see for instance [9], [18], [2], [17], [15], [8], [13], [1] and references therein). Thus, a key step before using the resulting models, is to (in)validate them using additional experimental data. In the case where the mode variable can be directly measured, the problem is closely related to that of validating Linear Parameter Varying (LPV) models and can be solved using techniques similar to those proposed in [19], [3]. However, in many practical situations, the mode variable is not directly available. Examples of this situation include, amongst others, fault detection problems and general data segmentation problems arising in the context of systems biology and video-analytics. The case where the discrete state must be inferred from the measured, noisy data, is considerably less developed. In [16], Ozay *et al.* proposed necessary and sufficient conditions for a switched affine autoregressive exogenous model to be (in)validated by the experimental data, based upon the idea of recasting the problem into one of checking emptiness of a semialgebraic set. In turn, as shown there, this condition can be checked by solving a sequence of convex optimization problems involving increasingly large matrices until either a positive solution is found or the rank of certain matrices formed using the solution to the dual problem ceases to increase. The former case provides an invalidation certificate,

while the latter indicates that the observed experimental data is consistent with the model under consideration. While this approach has been shown to be very effective with relatively small sized problems, its computational complexity increases with the problem size. In addition, in its present form, it cannot accommodate constraints on the trajectory of the mode variable. Equivalently, the underlying graph is assumed to be complete.

To circumvent these difficulties, in this paper we present an alternative framework for model (in)validation of switched affine models. Contrary to previous work based on exploiting the so called hybrid decoupling constraint, the present approach relies on introducing a set of sparse “indicator” variables s_i . In this context, the problem reduces to seeking solutions to a set of linear inequalities such that $s_i \in \{0, 1\}$, or showing that no such solutions exist. Further, topological constraints in the underlying graph (e.g. sets of admissible transitions) translate into simple linear constraints in these indicator variables.

Since the problem above is NP-hard, in the first part of the paper, we develop a linear programming based relaxation based upon using a (weighted) ℓ_1 norm as surrogate for cardinality [6], [20], [4]. This relaxation is computationally efficient and can handle large scale problems. However, in some cases it may fail to establish whether or not the model is (in)validated by the experimental data¹. In order to handle these cases, in the second part of the paper we develop an alternative convex relaxation, based on the solution of a constrained polynomial optimization problem using moments-based techniques. The main result of this section shows that this relaxation is guaranteed to provide a necessary and sufficient (in)validation certificate, using moment matrices of order at most $T + 1$, where T denotes the number of experimental data points.

These results are illustrated both with academic examples, where we compare the performance of the relaxations proposed here and the one developed in [16], and a non-trivial problem arising in computer vision: activity monitoring. Typically, a visual surveillance system captures high volume data streams from multiple cameras. However, interesting (e.g. abnormal) activities are rare. Thus, it is important to be able to automatically eliminate the normal behavior and trigger an appropriate response when something potentially interesting or abnormal occurs. As we show in Section V-B, this problem can be recast into a piecewise affine model

¹Due to its special structure, the problem does not satisfy, in general, the restricted isometry property. Thus, there is no guarantee that minimizing the ℓ_1 norm will indeed find the sparsest solution.

This work was supported in part by NSF grant ECCS-0901433; AFOSR grant FA9559-12-1-0271; and DHS grant 2008-ST-061-ED0001. Y. Cheng, Y. Wang and M. Sznaier are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115. N. Ozay is with the Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA 91125, Constantino Lagoa is with the Department of Electrical Engineering, Penn State University, University Park, PA 16802.

invalidation form and solved using the framework developed in this paper.

II. PRELIMINARIES

For ease of reference, in this section we summarize the notation used in the paper and recall some results on sparse polynomial optimization needed to establish its main results.

A. Notation and Definitions

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

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

$\|\mathbf{x}\|_1$ ℓ_1 -norm of the vector $\mathbf{x} \in \mathbf{R}^n$

$$\|\mathbf{x}\|_1 \doteq \sum_i |x_i|$$

$\|\mathbf{x}\|_0$ ℓ_0 quasinorm of \mathbf{x} (e.g. number of non-zero elements)

$\mathbf{M} \succeq \mathbf{N}$ the matrix $\mathbf{M} - \mathbf{N}$ is positive semidefinite.

Definition 1: A polynomial $p(\mathbf{x})$ is said to be a *sum of squares polynomial* (SoS), if it can be written as $p = \sum_{j=1}^m u_j^2$ for some polynomials u_1, \dots, u_m .

B. The Problem of Moments

Next, we recall results from the classical theory of moments that will play a key role in obtaining convex (in)validation certificates. Let $K \subset \mathbf{R}^n$ be a compact semi-algebraic set defined by a collection of polynomial inequalities of the form $g_k(\mathbf{x}) \geq 0$. Given a multisequence of scalars $\{m_\alpha\}$, indexed by a multi-index $\alpha \in \mathbf{N}^n$, the *K-moment problem* is to determine whether there exist a probability measure μ supported on K that has $\{m_\alpha\}$ as its α^{th} moments. That is:

$$m_\alpha = \mathbf{E}_\mu(\mathbf{x}^\alpha) \doteq \int_K \mathbf{x}^\alpha \mu(dx) \quad (1)$$

where $\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$. As shown in [11], [5], existence of such a measure is equivalent to positive semidefiniteness of the (infinite) moment $\mathbf{M}(m_\alpha)$ and localization $\mathbf{L}(g_k m_\alpha)$ matrices. Truncated versions of these matrices, involving moments of order up to $2N$ are given by:

$$\mathbf{M}_N(\mathbf{m})(i, j) = m_{\alpha^{(i)} + \alpha^{(j)}} \text{ for all } i, j \leq S_N. \quad (2)$$

where $S_N \doteq \binom{N+n}{n}$ (e.g. the number of moments in \mathbf{R}^n up to order N), and where the moments have been arranged according to a grevlex order of the corresponding monomials, so that $\mathbf{0} = \alpha^{(1)} < \dots < \alpha^{(S_N)}$. Similarly, if the polynomials defining the set K have the form $g_k(\mathbf{x}) = \sum_\beta x^\beta$, with degree δ_k , the corresponding truncated localization matrix is given by:

$$\mathbf{L}_N(g_k \mathbf{m})(i, j) = \sum_\beta g_{k, \beta^{(i)}} m_{\beta^{(i)} + \alpha^{(i)} + \alpha^{(j)}} \text{ for all } i, j \leq S_{N - \lfloor \frac{\delta_k}{2} \rfloor} \quad (3)$$

C. The Running Intersection Property and Sparse Polynomial Optimization

Definition 2: Consider a polynomial optimization problem of the form:

$$p_K^* := \min_{x \in K} p(x) \quad (4)$$

where $K \subset \mathbf{R}^n$ is defined by d polynomial inequalities of the form $g_k(\mathbf{x}) \geq 0$, $k = 1, \dots, d$. Assume that the polynomial p can be partitioned into $p = p_1 + \dots + p_l$ such that each p_j and each g_k contains only variables indexed by elements of some subset $I_k \subset \{1, \dots, n\}$. If there exists a reordering $I_{k'}$ of I_k such that for every $k' = 1, \dots, d-1$:

$$I_{k'} \cap \bigcup_{j=1}^{k'} I_j \subseteq I_s \text{ for some } s \leq k' \quad (5)$$

then the *running intersection property* is satisfied.

The significance of this property is that it allows for substantial computational complexity reduction. Specifically, in the case of generic polynomials, a moments relaxation of order N entails considering moments and localization matrices containing $O(n^{2N})$ variables. On the other hand, if the running intersection property holds, it can be shown [12], [21] that it is possible to define d sets of smaller sized matrices each containing only variables in I_k (i.e. number of variables is $O(\kappa^{2N})$, where κ is the maximum cardinality of I_k). Since typically $\kappa \ll n$, this leads to substantial reduction in the number of variables in the optimization (and hence its computational complexity).

D. Problem Statement

In the sequel, we consider multi-input, multi-output (MIMO) switched affine autoregressive exogenous (SARX) models of the form:

$$\begin{aligned} \xi_t &= \sum_{k=1}^{n_a} \mathbf{A}_k(\sigma_t) \xi_{t-k} \\ &\quad + \sum_{k=1}^{n_c} \mathbf{C}_k(\sigma_t) \mathbf{u}_{t-k} + \mathbf{f}(\sigma_t) \\ \mathbf{y}_t &= \xi_t + \eta_t \end{aligned} \quad (6)$$

where $\mathbf{u}_t \in \mathbf{R}^{n_u}$, $\mathbf{y}_t \in \mathbf{R}^{n_y}$ and $\sigma_t \in \mathbf{N}_s \doteq \{1, \dots, n_s\}$ denote the input, the measured output corrupted by the noise $\eta_t \in \mathbf{R}^{n_y}$, and the discrete state, respectively. Since in this paper we do not make any dwell-time assumptions, the mode signal σ_t can switch arbitrarily amongst subsets of the n_s subsystems G_i . In this context, the model invalidation problem of interest, illustrated in Fig. 1, can be formally stated as follows:

Problem 1: Given a nominal hybrid model of the form (6) together with its n_s submodels G_1, \dots, G_{n_s} , an *a priori* bound ϵ on noise, and experimental data $\{\mathbf{u}_t, \mathbf{y}_t\}_{t=0}^{T-1}$, determine whether or not the *a priori* information and the *a posteriori* experimental data are consistent, i.e. whether the consistency set

$$\mathcal{T}(\eta, \sigma) = \left\{ \|\eta_t\|_\infty \leq \epsilon, \sigma_t \in \mathbf{N}_s \text{ subject to (6)} \forall t \in [0, T-1] \right\}$$

is nonempty.

Remark 1: The formulation above does not place constraints on the admissible transitions of the mode variable

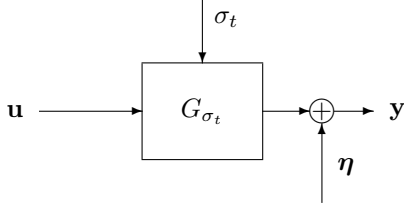


Fig. 1. Problem Setup. The coefficient matrices of the submodels G_i and a bound on the noise are known *a priori*. The experimental data consists of input/output measurements, \mathbf{u} and \mathbf{y} . The mode signal σ_t and noise sequence $\boldsymbol{\eta}$ are unknown.

σ (equivalently, the underlying graph is assumed to be complete). In section IV we will briefly indicate how to incorporate constraints on the set of admissible transitions.

III. CONVEX (IN)VALIDATION CERTIFICATES

In this section we present the main result of this paper: convex necessary and sufficient conditions for a MIMO SARX model to be (in)validated by experimental data. The first step towards this goal is to show that the problem can be reduced to a mixed integer linear program (MILP). While in principle this problem is NP-hard, as shown here, convex (in)validation certificates can be obtained by appealing to a combination of sparsification and polynomial optimization arguments.

A. A MILP reformulation

For notational simplicity, re-write equation (6) as

$$\begin{aligned} & \mathbf{A}_1(i)(\mathbf{y}_{t-1} - \boldsymbol{\eta}_{t-1}) + \dots + \mathbf{A}_{n_a}(i)(\mathbf{y}_{t+n_a} - \boldsymbol{\eta}_{t+n_a}) \\ & - (\mathbf{y}_t - \boldsymbol{\eta}_t) + \mathbf{C}_1(i)\mathbf{u}_{t-1} + \dots + \mathbf{C}_{n_c}(i)\mathbf{u}_{t-n_c} + \mathbf{f}(i) \doteq \\ & \mathbf{g}_{i,t} - \mathbf{h}_i\boldsymbol{\eta}_{t:t-n_a} = \mathbf{0} \end{aligned} \quad (7)$$

where we have defined $\mathbf{h}_i \doteq [-\mathbf{I} \quad \mathbf{A}_1(i) \quad \dots \quad \mathbf{A}_{n_a}(i)]$ and $\boldsymbol{\eta}_{t:t-n_a}^T \doteq [\boldsymbol{\eta}_t^T \quad \dots \quad \boldsymbol{\eta}_{t-n_a}^T]^T$. Note that since the mode signal σ_t is unmeasurable, the actual subsystem G_i that is active at any given time t is not known. However, regardless of the value of σ_t , the set of submodels given as part of *a priori* information is not invalidated by the experimental data if and only if Eq. (7) holds true for some $i \in \{1, \dots, n_s\}$ for all t . Clearly, this condition is equivalent to the existence of a set of “indicator” variables $s_{i,t}$ and admissible noise sequence $\boldsymbol{\eta}_t$ such that

$$\begin{aligned} & s_{i,t}(\mathbf{g}_{i,t} - \mathbf{h}_i\boldsymbol{\eta}_{t:t-n_a}) = \mathbf{0} \quad \forall t \in [0, T-1] \\ & \text{subject to} \\ & \sum_{i=1}^{n_s} s_{i,t} = 1 \\ & s_{i,t} \in \{0, 1\} \text{ and } \|\boldsymbol{\eta}_t\|_\infty \leq \epsilon \end{aligned} \quad (8)$$

Equivalently, by defining the auxiliary variables $\boldsymbol{\eta}_{i,t:t-n_a} \doteq s_{i,t}\boldsymbol{\eta}_{t:t-n_a}$, it can be easily shown that the condition above

is equivalent to feasibility of:

$$\begin{aligned} & s_{i,t}\mathbf{g}_{i,t} - \mathbf{h}_i\boldsymbol{\eta}_{i,t:t-n_a} = \mathbf{0} \\ & \text{subject to} \\ & s_{i,t} \in \{0, 1\}, \sum_i s_{i,t} = 1, \quad \forall t \in [0, T-1] \\ & \|\boldsymbol{\eta}_{i,t:t-n_a}\|_\infty \leq s_{i,t}\epsilon, \text{ and} \\ & \sum_{i=1}^{n_s} \boldsymbol{\eta}_{i,t:t-n_a} = \boldsymbol{\eta}_{t:t-n_a} \end{aligned} \quad (9)$$

Note that the problem above is a MILP feasibility problem that can be solved with existing tools. However, while this approach works well for small to medium sized problems, its poor scaling properties make it impractical for larger sized ones. As we show next, this difficulty can be circumvented by exploiting recent results in sparsification and polynomial optimization to derive convex certificates that can be computed in polynomial time.

B. Sparsification Based Certificates

In this section we derive sparsification based sufficient conditions for the model (6) to be (in)validated by the experimental data. The starting point is the following result relating feasibility of (9) to the cardinality of the solution to a linear programming (LP) problem.

Lemma 1: Eq. (9) is feasible if and only if there exist a solution $\mathbf{s} \doteq \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_T\}$, $\mathbf{s}_j \doteq [s_{1,j} \quad \dots \quad s_{n_s,j}]$, with $\|\mathbf{s}\|_0 = T$ to the following set of (linear) inequalities:

$$\begin{aligned} & s_{i,t}\mathbf{g}_{i,t} - \mathbf{h}_i\boldsymbol{\eta}_{i,t:t-n_a} = \mathbf{0} \quad \forall i, t \\ & 0 \leq s_{i,t} \leq 1, \quad \forall i, t \\ & \sum_i s_{i,t} = 1, \quad \forall t \\ & \|\boldsymbol{\eta}_{i,t:t-n_a}\|_\infty \leq s_{i,t}\epsilon, \quad \text{and} \\ & \sum_{i=1}^{n_s} \boldsymbol{\eta}_{i,t:t-n_a} = \boldsymbol{\eta}_{t:t-n_a} \end{aligned} \quad (10)$$

Proof: Follows from the fact that feasibility of (9) is equivalent to the existence, at each time instant t , of some $i \in [1, \dots, n_s]$ such that (8) holds with $s_{i,t} = 1, s_{j,t} = 0, j \neq i$. Hence (10) admits a solution with cardinality T . On the other hand, since $\sum_i s_{i,t} = 1$, it follows that if (10) admits a solution with cardinality T , this solution has exactly one $s_{i,t} = 1$ at each time t . ■

From the result above, it follows that validation/(in)validation certificates can be obtained by finding sparse solutions to the set of inequalities (10). In turn, using the standard ℓ_1 norm based relaxation of cardinality, combined with the re-weighted heuristic proposed in [4], [7] leads to the Algorithm outlined below:

Algorithm 1: Sparsification Based (In)Validation Certificates

- 1.- Solve (iteratively)

$$\begin{aligned} & \text{minimize}_{\mathbf{s}, \boldsymbol{\eta}} \quad \sum_{i,t} w_{i,t}^{(k)} s_{i,t} \\ & \text{subject to (10)} \end{aligned} \quad (11)$$

where $w_{i,t}^{(k+1)} = (s_{i,t}^{(k)} + \delta)^{-1}$, $s_{i,t}^{(k)}$ denotes the optimal solution at the k^{th} iteration, with $w^{(1)} = [1, 1, \dots, 1]^T$, and where δ is a (small) regularization constant.

- 2.- If the problem above is infeasible, then the experimental data invalidates the model (6). On the other

hand, if the problem is feasible and its optimal solution \mathbf{s}_o satisfies $s_{i,j} \in \{0, 1\}$, then the experimental data is consistent with the model (6) and the *a priori* information.

The algorithm above provides a computationally efficient way obtaining convex (in)validation certificates –infeasibility of (11)– or establishing that indeed the experimental data is compatible with the assumed priors (when the solution satisfies $s_{i,j} \in \{0, 1\}$). Note however that these conditions are only sufficient, since they can't elucidate the situation when the relaxation above provides a feasible solution with non-integer elements. Such a situation can arise both in the case of valid models (when due to noise some data points can be explained by more than one model), or invalid ones (if one of the models can be obtained as a linear combination of the others). In principle, one can attempt to address this case proceeding in the spirit of MILP, by setting the variables with non-integer values to all possible combinations in $\{0, 1\}$, and, for each combination, solving (11) with respect to the rest of the variables, in an attempt to find an integer solution or establish that such a solution is infeasible. However, this approach becomes quickly intractable when the number of non-integer elements in \mathbf{s}_o is not small. To circumvent this difficulty, next we use results from the theory of moments to obtain a convex necessary and sufficient (in)validation certificate.

Theorem 1: The following statements are equivalent:

- (i) Problem (8) is feasible.
- (ii) $p^* = 0$, where p^* denotes the optimal value of the objective function in the following constrained polynomial optimization problem:

$$p^* = \min_{\mathbf{s}, \boldsymbol{\eta}} \sum_{i,t} s_{i,t}^2 \|\mathbf{g}_{i,t} - \mathbf{h}_i \boldsymbol{\eta}_{t:t-n_a}\|^2$$

subject to

$$s_{i,t} = s_{i,t}^n$$

$$\sum_i s_{i,t}^2 = 1 \quad \forall t$$

$$\|\boldsymbol{\eta}_{t:t-n_a}\|_\infty \leq \epsilon$$
(12)

- (iii) $p_m^* = 0$, where p_m^* denotes the optimal value of the following convex optimization problem:

$$p_m^* = \min_{\mathbf{m}} \sum_{i,t} l_{i,t}$$

s.t.

$$\mathbf{M}_{T+1}(\mathbf{m}) \succeq 0$$

$$\mathbf{L}_{T+1}(\mathbf{m}) \succeq 0$$
(13)

where each $l_{i,t}$ is the linear functional of moments defined as $l_{i,t}(\mathbf{m}_{t-n_a:t}) \doteq \mathbf{E} \{s_{i,t}^2 \|\mathbf{g}_{i,t} - \mathbf{h}_i \boldsymbol{\eta}_{t:t-n_a}\|^2\}$, \mathbf{E} denotes expectation and where \mathbf{M}_{T+1} and \mathbf{L}_{T+1} are the moments and localization matrices associated with a truncated moments sequence containing terms up to order $2(T+1)$, as outlined in section II-B.

Proof: Only an outline is given, due to space constraints. (i) \iff (ii) follows from the fact that since the objective in (12) is a sum of squares, $p^* \geq 0$. Further, $p^* = 0 \iff$ there exist sequences $\{\boldsymbol{\eta}\}$ and $\{i_t\}$ such that $\|\boldsymbol{\eta}\| \leq 0$,

$s_{i_t} \neq 0$ and $(\mathbf{g}_{t,i_t} - \mathbf{h}_{i_t} \boldsymbol{\eta}_{t:t-n_a}) = \mathbf{0}$. Next, let

$$p(\mathbf{s}, \boldsymbol{\eta}) \doteq \sum_{i,t} s_{i,t}^2 \|\mathbf{g}_{i,t} - \mathbf{h}_i \boldsymbol{\eta}_{t:t-n_a}\|^2$$
(14)

The main idea in establishing that (ii) \iff (iii) is to first use Lemma 3.10 in [14] to show that

$$p - p^* = p_o + \sum_{it} \beta_{it} (\epsilon^2 - \boldsymbol{\eta}_{it}^2) + (\sum_{\mathcal{I}} s_{\mathcal{I}}^2 - 1) p^*$$
(15)

for some $2(T+1)$ order SoS polynomial $p_o(\boldsymbol{\eta}_i, s_{i,t})$ and some $2T$ order SoS polynomials $\beta_{it}(s_{i,t})$, (where, given a multi-index \mathcal{I} , we defined $s_{\mathcal{I}} \doteq s_{i_1,1} \dots s_{i_T,T}$). The proof follows now from Putinar Positivstellensatz (see for instance section 3.6.2 in [14]) and Theorem 4.2 (b) in [11]. ■

Remark 2: It can be easily shown that problem (12) exhibits the running intersection property. Thus, when solving problem (13) one only needs to consider smaller moment matrices, each involving the moments of the variables $s_{1,t}, \dots, s_{n_a,t}, \boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1} \dots \boldsymbol{\eta}_{t-n_a}$.

From Theorem 1 and the fact that $p^* \geq p_m^*$ (see for instance [11]), it follows that necessary and sufficient convex (in)validation certificates can be obtained by solving a (finite) sequence of convex optimization problems, as outlined in Algorithm 2 below.

Algorithm 2: Moments Based (In)Validation Certificates

- 1.- Set $N = 2$
 - 2.- Solve $p_m^* = \min_{\mathbf{m}} \sum_{i,t} l_{i,t}$ subject to:

$$\left. \begin{array}{l} \mathbf{M}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \\ \mathbf{L}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \end{array} \right\} \forall t \in [n_a, T]$$
 - 3.- If $p_m^* > 0$ or $p_m^* = 0$ and $\text{rank}[\mathbf{M}_N(\mathbf{m}_{t-n_a:t})] = \text{rank}[\mathbf{M}_{N-1}(\mathbf{m}_{t-n_a:t})]$ or $N = T + 1$ then stop. else set $N = N + 1$ and go to step 2.
 - 4.- If $p_m^* > 0$ the model is invalid, otherwise the data record is consistent with the *a-priori* assumptions.
-

Remark 3: It is worth emphasizing that, while Theorem 1 provides a guaranteed stopping condition for the algorithm above, in practice this condition has proved to be conservative. Indeed, consistent numerical experience supports the conjecture that only second order moment matrices need to be considered. However, at the present time no formal proof of this fact is available.

IV. INCORPORATING STRUCTURAL CONSTRAINTS

In many situations of practical interest, *a-priori* constraints are available on the admissible transitions of the mode variable. Incorporating these constraints can rule out models that may otherwise be deemed valid. For instance, in systems biology applications, only certain transitions can take place

from a given metabolic stage. Similarly, when dealing with human activities, only some sequence of activities are “normal”: for instance, depending on the context, a sequence such as $\{run, walk, stop\}$ could be normal, while $\{walk, stop, run\}$ may indicate an abnormal event. As we briefly discuss below, one of the advantages of the proposed framework is that structural constraints on the underlying graph can be easily accommodated. Specifically, assume that the transition from $\sigma_t = i$ to $\sigma_{t+\tau} = j$ is not admissible (that is, if at time “ t ” the state of the system is “ i ”, then it cannot be “ j ” τ stages later). Clearly, such constraints are easily incorporated in our framework by simply adding constraints of the form:

$$s_{i,t} + s_{j,t+\tau} \leq 1 \quad (16)$$

Similarly, situations where the state of the system at time t constrains its state at time $t + \tau$ to some set $\mathcal{I}_{t+\tau}$ simply reduce to:

$$\sum_{i \in \mathcal{I}_t} s_{i,t} = \sum_{j \in \mathcal{I}_{t+\tau}} s_{j,t+\tau} \quad (17)$$

(that is, at time $t + \tau$ the system can be in a state in $\mathcal{I}_{t+\tau}$ iff at time t it was in \mathcal{I}_t)

A point worth noting is that these additional constraints have relatively little effect on the overall computational complexity when using the sparsification based approach proposed in section III-B. On the other hand, they directly impact the running intersection property noted in Remark 2, since they couple the variables in the set \mathcal{I}_t to those in the set $\mathcal{I}_{t+\tau}$. Thus, when solving the optimization problem (13), a single larger moment matrix corresponding to all the variables in $\mathcal{I}_t \cup \mathcal{I}_{t+\tau}$ must be considered, rather than two smaller ones.

V. EXAMPLES

In this section we illustrate the effectiveness of the proposed method both using academic examples and a computer vision application.

A. Academic Examples

The goal of these examples is to compare the performance of the sparsification and moments based relaxations introduced in this paper against the method in [16]. In all cases, the resulting convex optimization problems were solved using cvx [10].

A.1 Example without constraints on the switching sequence

In this example we considered data generated by a system that switched randomly between the following three subsystems:

$$\xi_t = 0.2\xi_{t-1} + 0.24\xi_{t-2} + 2u_{t-1} \quad (G_1)$$

$$\xi_t = -1.4\xi_{t-1} - 0.53\xi_{t-2} + u_{t-1} \quad (G_2)$$

$$\xi_t = 1.7\xi_{t-1} - 0.72\xi_{t-2} + 0.5u_{t-1} \quad (G_3)$$

and the measurement equation:

$$y_t = \xi_t + \eta_t$$

with uniform random noise $\|\eta_t\|_\infty \leq \epsilon_t = 0.5$. The results obtained from two different noise realizations are summarized in Table I. The second and third columns in the rows labeled “results using moments” and “results using [16]” show both the order of the relaxation used and the value of the objective function in the corresponding polynomial optimization problem: a value $p > 0$ certifies that the model is invalid, while a value $p^* \approx 0$, together with a flat-extension, indicates that the data observed so far is consistent with the *a-priori* assumptions.

TABLE I
INVALIDATION RESULTS FOR EXAMPLE A.1 (PART 1)

Actual	G_1, G_2, G_3	G_1, G_2, G_3
A Priori Information	G_1, G_2, G_3	G_1, G_2, G_3
Results using Sparsification	feasible $s_{i,t} \in \{0, 1\}$	feasible $s_{i,t} \notin \{0, 1\}$
Interpretation	not invalidated	no decision
Time (sec.)	3.6808	4.4611
Results using Moments	2 $-3.0399e-07$	2 $-2.4735e-07$
Interpretation	not invalidated	not invalidated
Time (sec.)	2 6.8146	2 6.4891
Results from [16]	3 $-2.4832e-07$	3 $-1.0710e-07$
Interpretation	not invalidated	not invalidated
Time (sec.)	3 76.4373	3 66.6724

Next, we tested the different approaches by attempting to validate data obtained switching amongst a subset of $\{G_i\}$ against the hypothesis that it was generated using a *different* subset. The results are summarized in Table II. As shown in Table I and II, as expected, the sparsification based approach runs considerably faster than the moments based approaches. However, as illustrated there, in some cases it fails to provide a conclusive certificate. Regarding the moments based approaches, in the examples shown, both approaches have comparable performance. Note that (when exploiting the running intersection property), the approach in [16] requires only $n_y n_a$ variables, compared to $n_y n_a + n_s$ for Algorithm 2. Thus, for the same order relaxation, the former approach has substantially lower computational complexity. On the other hand, empirical results show Algorithm 2 requires considering moment relaxations of order 2, versus order n_s for [16]. Thus, it is preferable for scenarios where n_a or n_s are large.

TABLE II
INVALIDATION RESULTS FOR EXAMPLE A.1 (PART 2)

Actual	G_1, G_2, G_3	G_2, G_3
A Priori Information	G_1, G_2	G_1, G_2
Results using Sparsification	infeasible	infeasible
Interpretation	invalidated	invalidated
Time (sec)	0.1949	0.1980
Results using Moments	2 7.3123	2 14.2226
Interpretation	invalidated	invalidated
Time (sec.)	2 2.7642	2 2.4306
Results using [16]	2 6.8922	2 13.3500
Interpretation	invalidated	invalidated
Time (sec.)	2 5.2854	2 4.0118

A.2 Example with constraints on the switching sequence

Next, we use a simple example to illustrate the ability of the proposed methods to handle constraints on the admissible transitions. The data was generated using the same ARX submodels in A.1, and the measurement equation:

$$y_t = \xi_t + \eta_t$$

with $\|\eta_t\|_\infty \leq 0.5$. We further impose the constraint that the underlying graph has the topology shown in Fig. 2, that is one-step transitions from G_1 to G_2 are not admissible. Following the idea outlined in Section IV, this translates into a constraint of the form $s_{1,t} + s_{2,t+1} \leq 1, \forall t$. The results obtained running both Algorithm 1 and 2 are summarized in Table III. For completeness, we also report the results of running these algorithms, as well as the one in [16], without imposing the additional structural constraint. As shown in the table, if this constraint is not enforced, then the model is deemed to be valid.

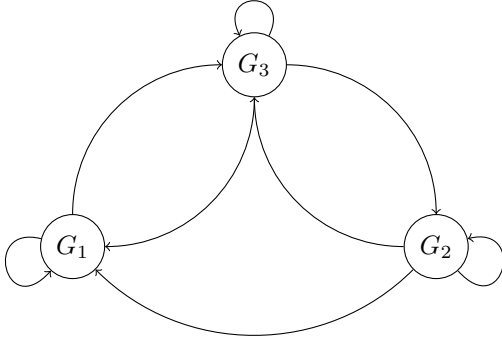


Fig. 2. A structurally constrained transition graph. One-step transitions from G_1 to G_2 are not allowed.

B. Contextually Abnormal Activity Recognition

In this section we apply the proposed model invalidation framework to a non-trivial problem arising in computer vision: activity monitoring. Following the framework proposed in [16], we will model a sequence of activities as the output of switched ARX system, where each submodel corresponds to a given activity. As shown in [16], contextually abnormal activities can be detected by simply invalidating the data record against the set of trajectories that can be obtained by switching among a set of known, normal activities, whose associated models can be obtained using identification tools. In this section, we further develop this framework and show that using the algorithms proposed here allows for not only recognizing abnormal activities, but also abnormal *activity sequences*. The model obtained in [16] for the activities $\{walk, wait, run\}$, given below for ease of reference, are:

$$\begin{aligned} \begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} 0.4747 & 0.0628 \\ -0.3424 & 1.2250 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 0.5230 & -0.1144 \\ 0.3574 & -0.2513 \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} \end{aligned} \quad (\text{walk})$$

TABLE III
INVALIDATION UNDER STRUCTURAL CONSTRAINTS

Actual	G_1, G_2, G_3 (with transitions from G_1 to G_2)
A Priori Information	G_1, G_2, G_3 (transitions from G_1 to G_2 not allowed)
Results using Sparsification without Constraints	feasible $s_{i,t} \notin \{0, 1\}$
Interpretation	no decision
Time (sec.)	3.712371
Results using Sparsification with Constraints	infeasible
Interpretation	invalidated
Time (sec.)	0.2695
Results using Moments without Constraints	2 $-4.0371e-07$
Interpretation	not invalidated
Time (sec.)	2 3.1370
Results using Moments with Constraints	2 0.1751
Interpretation	invalidated
Time (sec.)	2 64.8279
Results using [16]	3 $-8.3433e-08$
Interpretation	not invalidated
Time (sec.)	3 27.3684

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} \quad (\text{wait})$$

$$\begin{aligned} \begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} 0.6058 & 0.0003 \\ 0.2597 & 0.8589 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 0.3608 & 0.1853 \\ -0.2381 & 0.1006 \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} \end{aligned} \quad (\text{run})$$

where (x_t, y_t) denote the coordinates, at time t , of the center of mass of the person. Sample frames from video sequences involving combinations of these activities are shown in Fig. 3. As in [16], background subtraction was used to locate the person and (x_t, y_t) were estimated using the silhouettes. Table IV summarizes the results of applying Algorithms 1 and 2 to the resulting time series. For the purpose of this example, we assumed that transitions from *run* to *walk* were not admissible. Hence, a sequence of the form $\{walk \rightarrow run\}$ should be labeled normal, while $\{run \rightarrow walk\}$ should be flagged as contextually abnormal. As shown there, both the sparsification and moment-based algorithms correctly distinguish between these two situations.

VI. CONCLUSIONS

In this paper we considered the model (in)validation problem for switched ARX systems with unknown switches. Given a nominal model, a bound on the measurement noise and experimental input output data, we provided a necessary and sufficient condition that certifies the existence/nonexistence of admissible noise and switching sequences such that the resulting output sequence interpolates the given experimental data within the noise bound. The main result of this paper shows that these certificates can be computed by solving convex optimization problems. Moreover, these problems can be easily modified to incorporate



Fig. 3. Detecting contextually abnormal activities in video sequences. For the purpose of this example we assumed that transitions from running to walking are not allowed. Hence the top sequence should be labeled abnormal, while the bottom one (walking \rightarrow running) should not be flagged.

TABLE IV
INVALIDATION RESULTS FOR EXAMPLE B

Actual	walk, run (with transitions from run to walk)	walk, run (without transitions from run to walk)
A Priori Information	walk, run (transitions from run to walk not allowed)	walk, run (transitions from run to walk not allowed)
Results using Sparsification with Constraints	infeasible	feasible $s_{i,t} \in \{0, 1\}$
Interpretation	invalidated	not invalidated
Time (sec.)	2.1836	1.0577
Results using Moments with Constraints	2 0.001801	2 $-3.3581e-08$
Interpretation	invalidated	not invalidated
Time (sec.)	2 1.2968e+03	2 0.8407e+03

constraints on the topology of the underlying graph. In many practical cases, a validation/(in)validation certificate can be obtained by simply solving a sequence of LP problems. Cases where this approach fails to ascertain the validity of the model can be addressed, at the price of additional computational complexity, by using a moments-based reformulation containing moments of order up to $2(T + 1)$, where T denotes the number of experimental data points. The effectiveness of the proposed method was illustrated using both academic examples and a non-trivial problem arising in computer vision: activity monitoring. An interesting feature borne out by consistent numerical experience is the fact that, when using the moments-based approach, flat extensions (and hence exact solutions) were achieved for relaxations involving only moments of order up to 4 (e.g. $N = 2$). However, no formal proof of this fact is currently available.

REFERENCES

- [1] L. Bako. Identification of switched linear systems via sparse optimization. *Automatica*, 47:668–677, 2011.
- [2] 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.
- [3] F. Bianchi and R. Sánchez Peña. Robust identification/invalidation in an lpv framework. *International Journal of Robust and Nonlinear Control*, 20(3):301–312, Mar. 2009.
- [4] E. J. Candes, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.
- [5] R.E. Curto and Fialkow L.A. Truncated k -moment problems in several variables. *Journal of Operator Theory*, 54(1):189–226, 2005.
- [6] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, 2006.
- [7] 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.
- [8] C. Feng, C. Lagoa, and M. Sznaier. Hybrid system identification via sparse polynomial optimization. In *American Control Conference*, 2010.
- [9] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. A clustering technique for the identification of piecewise affine and hybrid systems. *Automatica*, 39:205–207, 2003.
- [10] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. <http://cvxr.com/cvx/>, April 2011. (web page and software).
- [11] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optimization*, 11(3):796–817, 2001.
- [12] J.B. Lasserre. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM J. Optimization*, 17(3):822–843, 2006.
- [13] F. Lauer, G. Bloch, and R. Vidal. A continuous optimization framework for hybrid system identification. *Automatica*, 47:608–613, 2011.
- [14] M. Laurent. *Emerging Applications of Algebraic Geometry*, chapter Sums of squares, moment matrices and optimization over polynomials, pages 157–270. Springer, 2009.
- [15] N. Ozay, C. Lagoa, and M. Sznaier. Robust identification of switched affine systems via moments-based convex optimization. In *Proc. 48th IEEE Conf. Dec. Control*, pages 4686–4691, 2009.
- [16] N. Ozay, M. Sznaier, and C. Lagoa. Model (in)validation of switched arx systems with unknown switches and its application to activity monitoring. In *Proc. 49th IEEE Conf. Dec. Control*, pages 7624–7630, 2010.
- [17] 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. 47th IEEE Conf. Dec. Control*, pages 123–130, 2008.
- [18] 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.
- [19] M. Sznaier and M.C. Mazzaro. An lmi approach to control-oriented identification and model (in) validation of lpv systems. *Automatic Control, IEEE Transactions on*, 48(9):1619–1624, sept. 2003.
- [20] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.
- [21] 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.