Model (In)validation of Switched ARX Systems with Unknown Switches and its Application to Activity Monitoring

Necmiye Ozay   Mario Sznaier   Constantino Lagoa

Abstract— Identification of switched linear systems has received considerable attention during the past few years. Since the problem is generically NP-Hard, the majority of existing algorithms are based on heuristics or relaxations. Therefore, it is crucial to check the validity of the identified models against additional experimental data. This paper addresses the problem of model (in)validation for multi-input multi-output switched affine autoregressive exogenous systems with unknown switches. Our main result provides necessary and sufficient conditions for a given model to be (in)validated by the experimental data. In principle, checking these conditions requires solving a sequence of convex optimization problems involving increasingly large matrices. However, as we show in the paper, if in the process of solving these problems either a positive solution is found or the so-called flat extension property holds, then the process terminates with a certificate that either the model has been invalidated or that the experimental data is indeed consistent with the model and a–priori information. By using duality, the proposed approach exploits the inherently sparse structure of the optimization problem to substantially reduce its computational complexity. The effectiveness of the proposed method is illustrated using both academic examples and a non-trivial problem arising in computer vision: activity monitoring.

I. INTRODUCTION AND MOTIVATION

Hybrid systems, dynamical systems where continuous and discrete states interact, are ubiquitous in many different contexts (e.g. biological systems, systems incorporating logical and continuous elements, manufacturing, automotive, etc.). In addition, these systems can be used to approximate nonlinear dynamics. Thus, during the past few years, a considerable research effort has been devoted to the problem of identifying hybrid systems, leading to several methods (see the excellent tutorial paper [13] for a summary of the main issues and recent developments in the field). Since the identification problem is generically NP-Hard, the majority of existing identification algorithms are based on heuristics or relaxations ([6], [1], [12], [11], [5]). Hence, a crucial step before using the resulting models, is to check their validity against additional experimental data. Model (in)validation of Linear Time Invariant (LTI) systems has been extensively addressed in the past two decades and both time and frequency domain results are available in the literature (see for instance [15], [3], [17] and references therein). A related line of research is model (in)validation of Linear Parameter Varying (LPV) systems ([20], [2]) where it is assumed that parameter values are measurable during the experiment and used as part of *a posteriori* data during the (in)validation step. Finally, (in)validation of continuous-time nonlinear models was addressed in [16] using sum of squares methods and barrier functions. However, to the best of our knowledge, similar results for discrete-time switched linear systems with unknown switches has not been considered, with the main difficulty here being the combination of noisy measurements and an unknown mode signal.

The main result of this paper is a necessary and sufficient condition for a multi-input multi-output switched affine autoregressive exogenous model to be (in)validated by the experimental data. Specifically, given a nominal model and experimental input/output data, we provide certificates for the existence/nonexistence of a switching sequence such that the resulting output sequence interpolates the given experimental data within a given noise bound. The starting point to obtain such certificates is to recast the (in)validation problem as one of checking whether a semialgebraic set is empty. By using a combination of recent results on moments-based sparse polynomial optimization and duality we show that emptiness of this set is equivalent to strict positivity of the solution of a related, convex optimization problem. In principle, checking this condition requires solving a sequence of convex optimization problems involving increasingly large matrices. However, as we show in the paper, if in the process of solving these problems either a positive solution is found or the rank of certain matrices formed using the solution to the dual problem ceases to increase (the so-called flat extension property), then the process terminates with either an invalidation or a consistency certificate. A salient feature of the proposed approach is its ability to exploit the inherently sparse structure of the optimization problem to substantially reduce its computational complexity.

In the second portion of the paper, these results are illustrated both with academic examples 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 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 that play a key role in establishing the main result of this paper.

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 $\mathbf{x} \in \mathbf{R}^n$

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

$\mathbf{R}[x_1, \ldots, x_n]$	the ring of polynomials in n variables over $\mathbf{R}$. When n is clear from context, $\mathbf{R}[\mathbf{x}]$ is used.
$\mathbf{N}_n$	positive integers up to n, i.e.

$$\mathbf{N}_n \doteq \{1, \ldots, n\}$$

$\wedge$ ($\vee$)	logical AND (OR)
$< \mathbf{M}, \mathbf{N} >$	trace($\mathbf{M}^T \mathbf{N}$)
$\mathbf{M} \succeq \mathbf{N}$	the matrix $\mathbf{M} - \mathbf{N}$ is positive semidefinite.

Definition 1: A polynomial $p \in \mathbf{R}[\mathbf{x}]$ 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 $u_1, \ldots, u_m \in \mathbf{R}[\mathbf{x}]$.

Definition 2: A set $K \subset \mathbf{R}^n$ is said to be semialgebraic if its defined by a finite number of polynomial equations and inequalities

B. The problem of moments and polynomial optimization.

In this paper, we will reduce the (in)validation problem to a polynomial optimization over a semialgebraic set, that is, a problem of the form:

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

where $K \subset \mathbf{R}^n$ is a compact semialgebraic set defined by d polynomial inequalities of the form $g_k(\mathbf{x}) \geq 0$, $k = 1, \ldots, d$. In the sequel, we briefly summarize some results relating polynomial optimization to the problem of moments that will be used to recast problem (P1) (and hence model invalidation) into a (possibly infinite-dimensional) convex optimization form (see [7], [8], [9] for more details).

1) The problem of moments: Let K be a compact subset of $\mathbf{R}^n$. Given a multisequence of scalars $\{m_{\boldsymbol{\alpha}}\}$, indexed by a multi-index $\boldsymbol{\alpha} \in \mathbf{N}^n$, the *K-moment problem* is to determine whether there exist a probability measure μ supported on K that has $\{m_{\boldsymbol{\alpha}}\}$ as its $\boldsymbol{\alpha}^{th}$ moments. That is:

$$m_{\boldsymbol{\alpha}} = \mathbf{E}_\mu(\mathbf{x}^{\boldsymbol{\alpha}}) \doteq \int_{\mathbf{K}} \mathbf{x}^{\boldsymbol{\alpha}} \mu(dx) \qquad (1)$$

where $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ (for a historical review and details of the problem, see [4], [18] and references therein). As shown in [7], [4], the existence of such a measure can be characterized by positive semidefiniteness of some infinite matrices, the so-called moment $\mathbf{M}(m_{\boldsymbol{\alpha}})$ and localization

matrices $\mathbf{L}(g_k m_{\boldsymbol{\alpha}})$ where $g_k(x) \geq 0$ are the polynomials defining K.

Next, we briefly discuss how to build truncated versions of $\mathbf{M}$ and $\mathbf{L}$ of a given sequence $\mathbf{m} \doteq \{m_{\boldsymbol{\alpha}}\}$ that contains all the moments up to order $2N$. Although the order of the subsequence is immaterial, for the sake of clarity of presentation, we arrange the moments according to a graded reverse lexicographic order (grevlex) of the corresponding monomials so that we have $\mathbf{0} = \boldsymbol{\alpha}^{(1)} < \ldots < \boldsymbol{\alpha}^{(S_N)}$, where $S_N \doteq \begin{pmatrix} N + n \\ n \end{pmatrix}$ is the number of moments in $\mathbf{R}^n$ up to order N. The truncated version of $\mathbf{M}$ is defined as follows:

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

Let $g_k(x) = \sum_{\boldsymbol{\beta}} g_{k, \boldsymbol{\beta}^{(l)}} x^{\boldsymbol{\beta}^{(l)}}$ be one of the defining polynomials of K with coefficients $g_{k, \boldsymbol{\beta}^{(l)}}$ and degree δ_k, then the corresponding truncated localization matrix is defined as:

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

2) Moments-based polynomial optimization: In this section, we recall some results from [7] that establish a connection between polynomial optimization and the problem of moments. In general, problem (P1) is non-convex, hence hard to solve. Instead, we consider a related problem:

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

where $\mathcal{P}(K)$ is the space of finite Borel signed measures on K. Although (P2) is an infinite dimensional problem, it is, in contrast to (P1), convex. The next result, taken from [7], 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), $p(x) = p_K^*$, μ^* almost everywhere.

One direct consequence of this theorem is that the problem of finding p_K^* in problem (P1) can be reduced to a sequence of Linear Matrix Inequalities (LMI) optimization problems in the moments of the unknown distribution by using (2) and (3) to define the constraint set. To this effect, define

$$p_N^* = \min_{\mathbf{m}} \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} m_{\boldsymbol{\alpha}}$$
$$\text{s.t.}$$
$$\mathbf{M}_N(\mathbf{m}) \succeq 0, \qquad (4)$$
$$\mathbf{L}_N(g_k \mathbf{m}) \succeq 0, k = 1, \ldots, d,$$

Theorem 2: As $N \to \infty$, $p_N^* \uparrow p_K^*$.

3) Exploiting the sparse structure: The next property will play a key role in reducing the computational complexity of problem (P1) by exploiting its structure.

Definition 3: Let $K \in \mathbf{R}^n$ be a semialgebraic set defined by d polynomials g_k. Let $I_k \subset \{1, \ldots, n\}$ be the set of indices of variables such that each g_k contains variables from some I_k and assume that the objective function p can be

partitioned as $p = p_1 + \ldots + p_l$ where each p_j contains only variables from some I_k. If there exists a reordering $I_{k'}$ of I_k such that for every $k' = 1, \ldots, d-1$:

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

then the *running intersection property* is satisfied.

For the case of generic polynomials and constraints, solving problem (P1) using the method of moments requires considering moments and localization matrices containing $O(n^{2N})$ variables. On the other hand, if the running intersection property holds, it can be shown [8], [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). In many practical applications, including the one considered in this paper, $\kappa \ll n$. Hence, exploiting the sparse structure substantially reduces the number of variables in the optimization (and hence the computational complexity), while still providing convergent relaxations.

III. (IN)VALIDATING MIMO SARX MODELS

In this section we formally state the problem under consideration and show that it can be reduced to a polynomial optimization over a semialgebraic set. In turn, this allows for exploiting the results briefly discussed in section II-B to obtain computationally tractable (in)validation certificates.

A. Problem Statement

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

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

where $\mathbf{u}_t \in \mathbf{R}^{n_u}$ is the input, $\tilde{\mathbf{y}}_t \in \mathbf{R}^{n_y}$ is the measured output corrupted by the noise $\boldsymbol{\eta}_t \in \mathbf{R}^{n_y}$, and $\sigma_t \in \mathbf{N}_s$ is the discrete mode signal indicating which of the s submodels is active at time t. We do not make any dwell-time assumptions, hence the mode signal σ_t can switch arbitrarily among the s submodels G_i, each of which is associated with the set of its coefficient matrices $\{\mathbf{A}_1(i), \ldots, \mathbf{A}_{n_a}(i), \mathbf{C}_1(i), \ldots, \mathbf{C}_{n_c}(i), \mathbf{f}(i)\}$.

The model (in)validation problem for the setup described above and shown in Figure 1 can be formally stated as follows:

Problem 1: Given a nominal hybrid model of the form (6) together with its s submodels $G_1, \ldots, G_s$, an *a priori* bound ϵ on noise, and experimental data $\{\mathbf{u}_t, \tilde{\mathbf{y}}_t\}_{t=t_0}^{T}$, determine whether or not the *a priori* information and the *a posteriori* experimental data are consistent, i.e. whether the consistency set

$$\begin{aligned}
\mathcal{T}(\boldsymbol{\eta}, \sigma) &= \{\|\boldsymbol{\eta}_t\|_\infty \leq \epsilon, \sigma_t \in \mathbf{N}_s \\
&\quad \text{subject to } (6) \ \forall t \in [t_0, T]\}
\end{aligned}$$

is nonempty.

Since $\mathcal{T}(\boldsymbol{\eta}, \sigma)$ contains all possible noise and mode signal sequences that can explain the observed data, clearly establishing that $\mathcal{T}(\boldsymbol{\eta}, \sigma) = \emptyset$ is equivalent to invalidating the model.
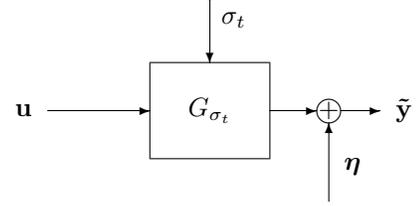

Fig. 1. Problem Setup. The coefficient matrices of the submodels G_i and a bound on the noise are known *a priori*. The experimental data consists of input/output measurements, $\mathbf{u}$ and $\tilde{\mathbf{y}}$. The mode signal σ_t and noise sequence η are unknown.

B. A Convex Certificate for (In)validating MIMO SARX Models

Next, we present the main result of the paper, showing that a necessary and sufficient condition for the model to be invalidated by the experimental data is strict positivity of the solution to a related convex optimization problem. We begin by constructing a semialgebraic set $\mathcal{T}'(\boldsymbol{\eta})$ and showing its equivalence, in a sense to become clear later, to the consistency set $\mathcal{T}(\boldsymbol{\eta}, \sigma)$. To this effect, assume that the ith submodel is active at time t. Rearranging the terms in Eq. (6) yields:

$$\begin{aligned}
&\mathbf{A}_1(i)(\tilde{\mathbf{y}}_{t-1} - \boldsymbol{\eta}_{t-1}) + \ldots + \mathbf{A}_{n_a}(i)(\tilde{\mathbf{y}}_{t-n_a} - \boldsymbol{\eta}_{t-n_a}) \\
&- (\tilde{\mathbf{y}}_t - \boldsymbol{\eta}_t) + \mathbf{C}_1(i)\mathbf{u}_{t-1} + \ldots + \mathbf{C}_{n_c}(i)\mathbf{u}_{t-n_c} + \mathbf{f}(i) = \mathbf{0}
\end{aligned} \tag{7}$$

which consists of n_y linear equations in $n_a + 1$ unknown noise vectors $\boldsymbol{\eta}_{t:t-n_a}$. Denote the equation corresponding to j^{th} output by $h_{t,i}^{(j)}(\boldsymbol{\eta}_{t:t-n_a})$. Then Eq. (7) is equivalent to

$$[h_{t,i}^{(1)}(\boldsymbol{\eta}_{t:t-n_a}) = 0] \wedge \ldots \wedge [h_{t,i}^{(n_y)}(\boldsymbol{\eta}_{t:t-n_a}) = 0] \tag{8}$$

or algebraically

$$g_{t,i}(\boldsymbol{\eta}_{t:t-n_a}) \doteq \sum_{j=1}^{n_y} \left[h_{t,i}^{(j)}(\boldsymbol{\eta}_{t:t-n_a}) \right]^2 = 0. \tag{9}$$

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, in order for the set of submodels given as part of *a priori* information not to be invalidated by the experimental data, Eq. (9) should hold true for some $i \in \{1, \ldots, s\}$[1]. This condition can be expressed as

$$[g_{t,1}(\boldsymbol{\eta}_{t:t-n_a}) = 0] \vee \ldots \vee [g_{t,s}(\boldsymbol{\eta}_{t:t-n_a}) = 0] \tag{10}$$

[1]This idea is similar to the hybrid decoupling constraint proposed in [10]

or algebraically

$$p_t(\boldsymbol{\eta}_{t:t-n_a}) \doteq \prod_{i=1}^{s} g_{t,i}(\boldsymbol{\eta}_{t:t-n_a}) = 0. \qquad (11)$$

Next we use the equalities above to define a semialgebraic set and establish its relationship with the consistency set $\mathcal{T}$:

Proposition 1: Let

$$\mathcal{T}'(\boldsymbol{\eta}) \quad \doteq \quad \left\{\boldsymbol{\eta} \mid f_{t,j}(\eta_t^{(j)}) \geq 0 \,\, \forall t \in [0,T], j \in \mathbf{N}_{n_y} \text{and} \right.$$
$$\left. p_t(\boldsymbol{\eta}_{t:t-n_a}) = 0 \,\, \forall t \in [n_a, T] \right\}.$$

where

$$f_{t,j}(\eta_t^{(j)}) \doteq \epsilon^2 - \left[\eta_t^{(j)}\right]^2$$

Then, $\mathcal{T}(\boldsymbol{\eta}, \sigma)$ is empty if and only if $\mathcal{T}'(\boldsymbol{\eta})$ is empty.

Proof: Follows from the construction of $\mathcal{T}'(\boldsymbol{\eta})$ and the fact that $|\eta_t^{(j)}| \leq \epsilon \iff f_{t,j}(\eta_t^{(j)}) \geq 0$ ∎

At this point, one can use the *Positivstellenstaz* and derive sum of squares certificates for the emptiness of $\mathcal{T}'(\boldsymbol{\eta})$, as proposed in [14]. However, as we show next, adopting a dual approach based on the theory of moments allows for exploiting the inherently sparse structure of the problem. In order to pursue this approach, start by considering the following optimization problem:

$$o^* = \min_{\boldsymbol{\eta}} \quad \sum_{t=n_a}^{T} p_t(\boldsymbol{\eta}_{t:t-n_a})$$
$$\text{s.t.}$$
$$f_{t,j}(\eta_t^{(i)}) \geq 0 \,\, \forall t \in [0,T], j \in \mathbf{N}_{n_y}. \qquad (12)$$

Note that $o^* \geq 0$, since the objective function in (12) is a sum of squares polynomial[2]. Further, if $\mathcal{T}'(\boldsymbol{\eta})$ is non-empty, then there exist a noise sequence $\boldsymbol{\eta}^*$ for which (12) attains its minimum $o^* = 0$. Equivalently, $o^* > 0 \iff \mathcal{T}'(\boldsymbol{\eta}) = \emptyset$.

Proposition 2: Problem (12) above satisfies the running intersection property.

Proof: Consider the $T - n_a + 1$ subsets I_k of the variables $\boldsymbol{\eta}_{0:T}$ where each I_k contains only the variables $\boldsymbol{\eta}_{k:k+n_a}$. One can associate each $f_{t,j}$ with I_0 for $t \leq n_a$ and with I_{t-n_a} for $t > n_a$. The collection I_{kj} formed by repeating each I_k, n_y times, satisfies (5) hence the running intersection property in Definition 3 holds. ∎

Next, we use results from the theory of moments to obtain a convex problem where the objective and constraints have affine (rather than polynomial) dependence on the data and exploits sparsity. Consider the related moments-based optimization:

$$d_N^* = \min_{\mathbf{m}} \sum_{t=n_a}^{T} l_t(\mathbf{m}_{t-n_a:t})$$
$$\text{s.t.}$$
$$\mathbf{M}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \,\, \forall t \in [n_a, T]$$
$$\mathbf{L}_N(f_{t,j}\mathbf{m}_{t-n_a:t}) \succeq 0 \,\, \forall t \in [n_a+1, T], j \in \mathbf{N}_{n_y}$$
$$\mathbf{L}_N(f_{t,j}\mathbf{m}_{0:n_a}) \succeq 0 \,\, \forall t \in [0, n_a], j \in \mathbf{N}_{n_y}$$
$$(13)$$

where each l_t is the linear functional of moments defined as $l_t(\mathbf{m}_{t-n_a:t}) \doteq \mathbf{E}\left\{p_t(\boldsymbol{\eta}_{t:t-n_a})\right\}$, $\mathbf{E}$ denotes expectation

[2]since it is formed by multiplication and addition of SOS polynomials in (9), and the cone of SOS polynomials is closed under these operations.

and where $\mathbf{M}_N$ and $\mathbf{L}_N$ are the moments and localization matrices associated with a truncated moments sequence containing terms up to order $2N$ with $N \geq s$. Clearly $d_N^* \geq 0$, since the objective function in (12) is a sum of squares polynomial, and, from Theorem 2, $d_N^* \uparrow o^*$ as $N \to \infty$. These observations lead to the following necessary and sufficient conditions for (in)validation:

Proposition 3: The following statements are equivalent:

(i) The consistency set $\mathcal{T}'(\boldsymbol{\eta})$ is empty
(ii) There exists some finite N_o such that $d_{N_o}^* > 0$
(iii) The solution r^* to the following optimization problem is strictly greater than zero:

$$r^* = \min_{\mathbf{m}} \sum_{t=n_a}^{T} l_t(\mathbf{m}_{t-n_a:t})$$
$$\text{s.t.}$$
$$\mathbf{M}_s(\mathbf{m}_{t-n_a:t}) \succeq 0 \,\, \forall t \in [n_a, T]$$
$$\text{rank}\left[\mathbf{M}_s(\mathbf{m}_{t-n_a:t})\right] = 1 \,\, \forall t \in [n_a, T]$$
$$\mathbf{L}_s(f_{t,j}\mathbf{m}_{t-n_a:t}) \succeq 0 \,\, \forall t \in [n_a+1, T], j \in \mathbf{N}_{n_y}$$
$$\mathbf{L}_s(f_{t,j}\mathbf{m}_{0:n_a}) \succeq 0 \,\, \forall t \in [0, n_a], j \in \mathbf{N}_{n_y}$$
$$(14)$$

where each of the $T - n_a + 1$ moments sequences $\mathbf{m}_{t-n_a:t}$, $t \in [n_a, T]$, contains moments up to order $2s$ (i.e. two times number of submodels).

Proof: (i)⟺(ii) Recall that $\mathcal{T}'(\boldsymbol{\eta}) = \emptyset \iff o^* > 0$. Since $d_N^* \uparrow o^*$ as $N \to \infty$, if $o^* > 0$, there exist N_o such that $d_{N_o}^* > 0$. On the other hand, if $d_{N_o}^* > 0$ then $o^* > 0$ since $d_{N_o}^* < o^*$. Hence, $\mathcal{T}'(\boldsymbol{\eta})$ is empty.

(i)⟺(iii) To prove this equivalence, we show that r^* in (14) is equal to o^* in (12). Assume η^* is an optimizer of (12), then the moments of the distribution μ^* with point support at η^* is feasible for (14) with the same objective value which implies $r^* \leq o^*$. On the other hand, if $\mathbf{m}^*$ is a minimizer of (14), the rank condition implies that there is a corresponding measure μ^* with point support, say at η^*. Since this value of η^* is a feasible point of (12), $o^* \leq r^*$. Therefore, $r^* = o^*$ from which we conclude that $r^* > 0$ is equivalent to (i). ∎

Note that by forming a single block diagonal matrix containing all LMI constraints in (13), it can be transformed into the standard dual form of semidefinite programs. That is:

$$d_N^* = \inf_{\mathbf{m}} \quad \mathbf{b}^T\mathbf{m} + c_o$$
$$\text{s.t.} \qquad (15)$$
$$\sum_{\boldsymbol{\alpha} \in \mathcal{I}} \mathbf{A}_{\boldsymbol{\alpha}} m_{\boldsymbol{\alpha}} \succeq \mathbf{C}$$

where $\mathcal{I}$ is the set of the multi-indexes of all moments that occur in (13) except the zeroth moment $m_{\mathbf{0}} = 1$ which is used to form the constant terms $\mathbf{C}$ and c_o.

Remark 1: It is important to highlight the complexity reduction achieved by employing the running intersection property while forming the optimization problem (13). The conventional moment relaxation of order N in [7] would require $O((Tn_y+n_y)^{2N})$ variables with a moment matrix of size $\binom{N+Tn_y+n_y}{Tn_y+n_y}$. On the other hand, (13) involves only $O((n_an_y+n_y)^{2N})$ variables with $T - n_a + 1$ moment matrices of size $\binom{N+n_an_y+n_y}{n_an_y+n_y}$ where, in general, the

length of the experimental data T is substantially larger than the order of the regressor n_a (i.e. $n_a \ll T$).

IV. NUMERICAL CONSIDERATIONS

From Proposition 3 it follows that (in)validation/validation certificates can be obtained, for instance, by solving a sequence of problems with increasing N until either a solution with $d^* > 0$ is found (in which case the data invalidates the model) or condition (iii) fails, in which case the data record is consistent with the model and *a-priori* information. However, a potential numerical difficulty here stems from the fact that in practice $d^* = 0$ is never reached since there is no closed form solution to the problem and numerical solvers terminate when the duality gap drops below a certain level. Thus, from a practical standpoint, more useful conditions can be obtained by examining the primal problem, which can be written as

$$
\begin{aligned}
p_N^* = \quad &\max_{\mathbf{X}} \quad < \mathbf{C}, \mathbf{X} > + c_o \\
&\text{s.t.} \\
&< \mathbf{A}_\alpha, \mathbf{X} > = b_\alpha \quad \forall \alpha \in \mathcal{I} \\
&\mathbf{X} \succeq 0
\end{aligned}
\tag{16}
$$

where, without loss of generality, $\mathbf{X}$ can be chosen to have the same block-diagonal structure that $\mathbf{A}_\alpha$s and $\mathbf{C}$. Note that, from duality, $p_N^* \le d_N^*$, and hence $p_N^* > 0$ provides an invalidation certificate. In fact if problem (16) is feasible, strong duality holds [8]. Thus $p_N^* > 0$ is strictly equivalent to $d_N^* > 0$, leading to the following result:

Proposition 4: The consistency set $\mathcal{T}'(\boldsymbol{\eta})$ is empty if and only if $p_{N_o}^* > 0$ for some finite N_o.

Proof: Follows directly from strong duality and Proposition 3. ∎

Let $\tilde{d}_N$ and $\tilde{p}_N$ be the numerical solutions to problems (13) and (16) respectively, obtained using a standard SDP solver. As noted before, in practice, one always has $\tilde{d}_N > 0$. Thus it can be hard to ascertain whether conditions (ii) or (iii) in Proposition 3 fail. On the other hand, $\tilde{p} > 0$ is an invalidation certificate.

Remark 2: A certificate that the existing data record is consistent with the model and *a priori* assumptions can be obtained by resorting to a variation of the so-called flat extension property for sparse polynomial optimization stated in [8]. In particular, if for some N, $\tilde{p} \le 0$ and the dual solution satisfies $\text{rank}\left[\mathbf{M}_N(\mathbf{m}_{0:n_a}^*)\right] = 1$; $\text{rank}\left[\mathbf{M}_N(\mathbf{m}_{t-n_a:t-1}^*)\right] = 1$ and $\text{rank}\left[\mathbf{M}_N(\mathbf{m}_{t-n_a:t}^*)\right] = \text{rank}\left[\mathbf{M}_{N-2}(\mathbf{m}_{t-n_a:t}^*)\right]$ $\forall t \in [n_a, T]$ where $\mathbf{m}^*$ denotes an optimal solution of (13), then this certifies that $o^* = d_N^* = 0$; hence $\mathcal{T}'(\boldsymbol{\eta})$ is not empty.

Remark 3: It is also worth pointing out that all of the results above (including the running intersection property) hold true in the presence of bounded parametric model uncertainty. For instance, if $a_i(k, l)$ is an entry in one of the coefficient matrices of the ith submodel, it is possible to use $a_i(k, l) + \delta_i(k, l)$ with $|\delta_i(k, l)| \le \delta^*$ in the *a priori* model. This only requires additional variables in the optimization problem.

V. ILLUSTRATIVE EXAMPLES

In this section we illustrate the effectiveness of the proposed method both using academic examples and a computer vision application. In all cases, we used the moments relaxation corresponding to $N = s$ and the resulting SDP problem was solved using SEDUMI [19].

A. Academic Examples

We consider the ARX submodels:

$$y_t = 0.2y_{t-1} + 0.24y_{t-2} + 2u_{t-1} \tag{G_1}$$

$$y_t = -1.4y_{t-1} - 0.53y_{t-2} + u_{t-1} \tag{G_2}$$

$$y_t = 1.7y_{t-1} - 0.72y_{t-2} + 0.5u_{t-1} \tag{G_3}$$

and the measurement equation:

$$\tilde{y}_t = y_t + \eta_t. \tag{17}$$

We ran different sets of simulations investigating different sources of model mismatch. In all cases, we collected input/output data $\{u_t, \tilde{y}_t\}$ for $t \in [0, 96]$ and tried to (in)validate the *a priori* model. In all experiments, when we used data inconsistent with the *a priori* information, $\tilde{p}^*$ turned out to be positive. Hence, we correctly invalidated the model in each of such cases. On the other hand, whenever the *a priori* information was consistent with *a posteriori* data, we had $\tilde{p}^* < 0$.

Example 1: (Submodel mismatch) For the first set of experiments, we generated input/output data using different subsets of $\{G_1, G_2, G_3\}$ with a random switching sequence σ_t and with uniform random noise $||\boldsymbol{\eta}_t|| \le 0.5$. The noise used in the experiments was within the *a priori* noise bound. We used both correct and incorrect *a priori* submodel sets. Hence, the model should be invalid when a submodel that is not contained in the *a priori* submodel set is used in the actual experiment. The results are summarized in Table I.

A priori	Actual	Result
G_1, G_2, G_3	G_1, G_2, G_3	not invalidated
G_1, G_2, G_3	G_1, G_2	not invalidated
G_1, G_2	G_1, G_2, G_3	**invalidated**
G_1, G_2	G_2	not invalidated
G_1, G_2, G_3	G_1	not invalidated
G_1, G_2	G_2, G_3	**invalidated**

TABLE I

INVALIDATION RESULTS FOR EXAMPLE 1. THE VALUES OF $\tilde{p}$ WERE RESPECTIVELY $-3.8441e - 008$, $-8.2932e - 009$, 0.8585, $-5.4026e - 008$, $-1.5490e - 007$ AND 0.7566.

Example 2: (Noise bound mismatch) For this example, we generated input/output data using different subsets of $\{G_1, G_2, G_3\}$ with a random switching sequence σ_t and with uniform random noise $||\boldsymbol{\eta}_t|| \le \epsilon$. The *a priori* submodel set and the actual submodel set used in the experiment were the same. The source of invalidation was the actual noise level exceeding the *a priori* bound. The results of this set of experiments are summarized in Table II.

Example 3: (Submodel perturbation) For this set of experiments, we generated input/output data by perturbing

Submodels	A priori ϵ	Actual ϵ	Result
G_1, G_2, G_3	0.5	1	**invalidated**
G_1, G_2, G_3	0.8	1	**invalidated**
G_1, G_2, G_3	1.2	1	not invalidated
G_1, G_2	0.5	1	**invalidated**
G_1, G_2	0.8	1	**invalidated**
G_1, G_2	1.2	1	not invalidated

TABLE II

INVALIDATION RESULTS FOR EXAMPLE 2. THE VALUES OF $\tilde{p}$ WERE RESPECTIVELY 0.0724, 0.0035, $-5.1810e - 007$, 0.0737, 0.0034 AND $-1.4930e - 007$.

the coefficients of *a priori* submodels $\{G_1, G_2, G_3\}$, and again using a random switching sequence σ_t and uniformly sampled random noise $||\boldsymbol{\eta}_t|| \le 0.5$. We denote the submodel whose coefficient values are perturbed by e percent of their original values as $G_i + \Delta_e$. This case is more challenging than the one considered in Example 1 since the dynamics are similar. Nevertheless, we could invalidate in each of our trials. The results for this example are summarized in Table III.

A priori	Actual	Result
G_1, G_2, G_3	$G_1 + \Delta_5, G_2 + \Delta_5, G_3 + \Delta_5$	**invalidated**
G_1, G_2, G_3	$G_1 + \Delta_2, G_2 + \Delta_2, G_3 + \Delta_2$	**invalidated**
G_1, G_2, G_3	$G_1 + \Delta_1, G_2 + \Delta_1, G_3 + \Delta_1$	**invalidated**
G_1, G_2	$G_1 + \Delta_5, G_2 + \Delta_5$	**invalidated**
G_1, G_2	$G_1 + \Delta_2, G_2 + \Delta_2$	**invalidated**
G_1, G_2	$G_1 + \Delta_1, G_2 + \Delta_1$	**invalidated**
G_1, G_2	$G_1 + \Delta_2$	**invalidated**

TABLE III

INVALIDATION RESULTS FOR EXAMPLE 3. THE VALUES OF $\tilde{p}$ WERE ,RESPECTIVELY, 0.8963, 0.0997, 0.0080, 0.0308, $2.8638e - 004$, $2.9069e - 006$ AND $6.2061e - 006$.

B. A Practical Example: Activity Monitoring

In this section we illustrate the application of the proposed model invalidation framework to a non-trivial problem arising in computer vision: activity monitoring. Here we start with a set of dynamic models associated with "normal" behavior. If a person passing in front of the camera exhibits a combination of the normal activities (i.e. his/her behavior can be modeled with a hybrid system that has the normal activity dynamics as submodels), then the activity is considered not interesting. On the other hand, if he/she does something different than what has been encoded in the initial set of normal dynamics, this is an indication of an interesting event. In such cases the model should be invalidated.

We used a training video of walk shown in Fig. 2 to identify an autoregressive model for the dynamics of the center of mass of a person walking. By minimizing the ℓ_∞ norm of process noise via linear programming, we obtained the following model for "walk":

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.4747 & 0.0628 \\ -0.3424 & 1.2250 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 0.5230 & -0.1144 \\ 0.3574 & -0.2513 \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} \quad (A_1)$$

where (x_t, y_t) is the normalized coordinate of the center of the person in the tth frame. Another activity that we considered normal is "waiting" which can simply be modeled as:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix}. \quad (A_2)$$

We normalized the image coordinate system so that $(x, y) \in [0, 1] \times [0, 1]$ and set the measurement noise level to $||\boldsymbol{\eta}_t|| \le 0.04$. This bound together with the submodels (A_1) and (A_2) for "normal" activities constitute the *a priori* information.

As for test purposes, we used three different video sequences. Background subtraction was used to locate the person; and the center of mass was estimated and tracked using the silhouettes. Then the center of mass trajectories were used for model (in)validation. In the first sequence, the person walks, waits and walks again; so the overall activity is normal. In the second sequence, the person runs for which our method found the certificate for invalidity by finding $\tilde{p} > 0$. In the third sequence, the person walks and then starts jumping. Again, our method flagged this abnormal activity by verifying the invalidity of the model. Sample frames from the sequences are shown in Fig. 3 and the results are summarized in Table IV. The primal optimal objective values, $\tilde{p}$, are also reported in Table IV. Since image coordinates are scaled to $[0, 1] \times [0, 1]$ in this example, higher order polynomials of values within this range results in low values of $\tilde{p}$. However, note that the positive range of $\tilde{p}$ is just a matter of scaling. It is the sign of $\tilde{p}$ that is important. The positivity of $\tilde{p}$ (as well as the existence of any primal feasible point with positive objective value) is a certificate for the invalidity of the model.

A priori	Actual	Result
walk, wait	walk, wait	not invalidated
walk, wait	run	**invalidated**
walk, wait	walk, jump	**invalidated**

TABLE IV

INVALIDATION RESULTS FOR ACTIVITY MONITORING. THE VALUES OF $\tilde{p}$ WERE, RESPECTIVELY, $-2.3303e - 008$, $2.3707e - 005$, AND $5.0293e - 007$.

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. In principle, computing these certificates entails solving a sequence of convex optimization problems involving increasingly large matrices. However, as we show here, if in the process of solving these problems either a positive solution is found or the so-called flat extension property holds, then the process
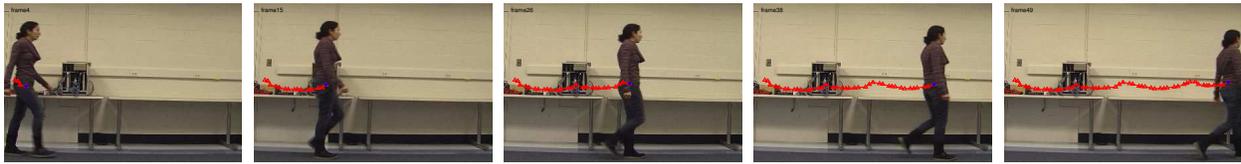
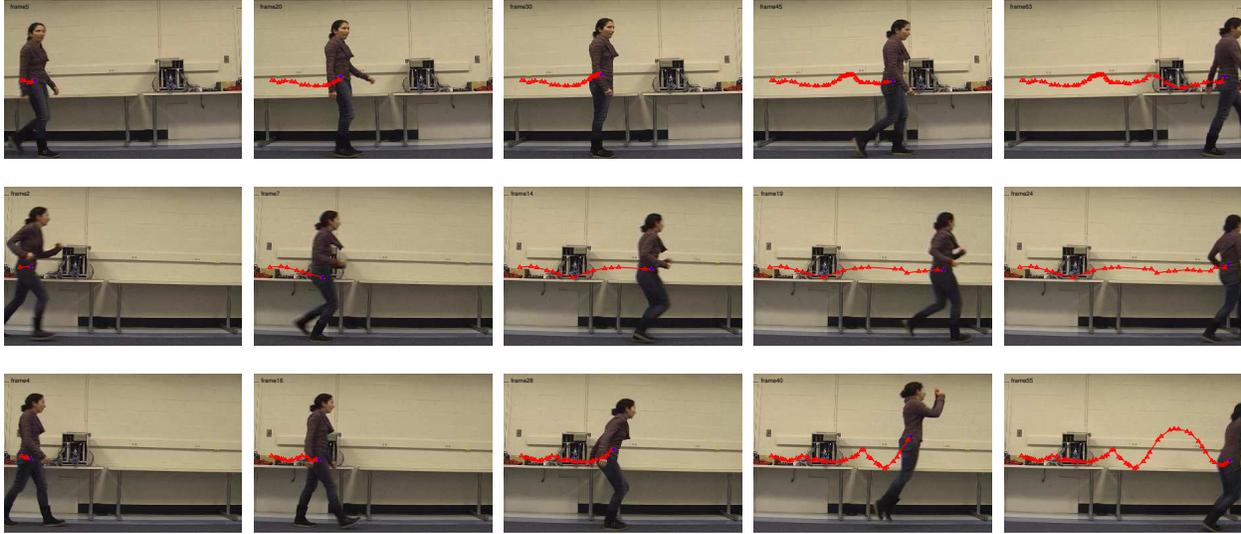Fig. 2. Training sequence used in identification of the submodel (A_1) for walking.

Fig. 3. Top: Walk, wait, walk sequence (not invalidated). Middle: Running sequence (invalidated). Bottom: Walk, jump sequence (invalidated).

terminates with a certificate that either the model has been invalidated (first case) or that the experimental data is indeed consistent with the model and *a–priori* information. By using duality, the proposed approach exploits the inherently sparse structure of the optimization problem to substantially reduce its computational complexity. The effectiveness of the proposed method was illustrated using both academic examples and a non-trivial problem arising in computer vision: activity monitoring. Research currently under way seeks to extend these results to the (in)validation of hybrid models in state space form.

REFERENCES

[1] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic Control*, 50(10):1567–1580, 2005.

[2] 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.

[3] J. Chen. Frequency-domain tests for validation of linear fractional uncertain models. *IEEE Transactions on Automatic Control*, 42(6):748–760, June 1997.

[4] R.E. Curto and Fialkow L.A. Truncated k-moment problems in several variables. *Journal of Operator Theory*, 54(1):189–226, 2005.

[5] C. Feng, C. Lagoa, and M. Sznaier. Hybrid system identification via sparse polynomial optimization. In *American Control Conference*, 2010. (to appear).

[6] 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.

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

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

[9] M. Laurent. *Emerging Applications of Algebraic Geometry*, chapter Sums of squares, moment matrices and optimization over polynomials, pages 157–270. Springer, 2009.

[10] Y. Ma and R. Vidal. A closed form solution to the identification of hybrid arx models via the identification of algebraic varieties. In *Hybrid Systems Computation and Control*, pages 449–465, March 2005.

[11] 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.

[12] 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.

[13] 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.

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

[15] K. Poolla, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal. A time domain approach to model validation. *IEEE Transactions on Automatic Control*, 39(5):951–959, 1994.

[16] S. Prajna. Barrier certificates for nonlinear model validation. *Automatica*, 42(1):117–126, 2006.

[17] R. Sánchez Peña and M. Sznaier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.

[18] K. Schmudgen. The k-moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289(1):203–206, 1991.

[19] Jos F. Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones, 1999.

[20] 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.

[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.