

Convex Certificates for Model (In)validation of Switched Affine Systems with Unknown Switches

Necmiye Ozay

Mario Sznaier

Constantino Lagoa

Abstract

Checking validity of a model is a crucial step in the process of system identification. This is especially true when dealing with switched affine systems since, in this case, the problem of system identification from noisy data is known to be generically NP-Hard and can only be solved in practice by using heuristics and relaxations. Therefore, before the identified models can be used for instance for controller design, they should be systematically validated against additional experimental data. In this paper we address the problem of model (in)validation for multi-input multi-output switched affine systems in output error form with unknown switches. As a first step, we prove that necessary and sufficient invalidation certificates can be obtained by solving a sequence of convex optimization problems. In principle, these problems involve increasingly large matrices. However, as we show in the paper by exploiting recent results from semi-algebraic geometry, the proposed algorithm is guaranteed to stop after a finite number of steps that can be explicitly computed from the a-priori information. In addition, this algorithm exploits the sparse structure of the underlying optimization problem to substantially reduce the computational burden. 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 and can be found in many different contexts. Examples are as diverse as manufacturing processes, biological systems, communication networks, automobiles and chemical processes.

This work was supported in part by NSF grants CNS-1329422, ECCS-1201973, ECCS-0731224, CMMI-0838906 and ECCS-0901433; AFOSR grant FA9550-09-1-0253; and DHS grant 2008-ST-061-ED0001. N. Ozay is with the EECS Department, University of Michigan, Ann Arbor, MI, 48109. M. Sznaier is with the ECE Department, Northeastern University, Boston, MA, 02115. Constantino Lagoa is with the EE Department, Penn State University, University Park, PA 16802.

Switched affine systems form a specific class of hybrid systems where continuous states evolve according to a set of affine dynamics and the discrete state (i.e., the mode of the system) dictates which dynamics within this set is followed at a given time. As such, switched affine system models can also be used to approximate nonlinear dynamics. Therefore, during the past decade, a considerable research effort has been devoted to the problem of set membership identification of such models from experimental data; e.g., see the tutorial paper [23] for a summary of the main issues in this area and [11] for recent developments in the field. Due to the mixture of continuous and discrete state variables and the presence of noise, the identification problem is generically NP-Hard. Hence, the majority of existing identification algorithms are based on heuristics or relaxations ([10], [2], [21], [19], [8], [7], [1], [15], [22])¹. Therefore, before using such models, it is imperative that they are validated against additional experimental data. Note that this validation can never be achieved using a finite data record, since there is always the possibility that the model will be invalidated by data not yet seen. Thus, from a practical standpoint, the problem of interest is model invalidation, rather than validation: establish whether a given set of a-priori assumptions (model, noise and uncertainty levels) is consistent with the observed experimental data. Models that do not pass an (in)validation step are rejected, while those that do are accepted as correct, with the proviso that they could be invalidated by future data. In this context, it is desirable to obtain tight (ideally necessary and sufficient) invalidation certificates, to guarantee that valid models are not rejected. Further, these certificates can be used to sharpen existing bounds on model uncertainty, leading to less conservative controller designs. Model (in)validation of Linear Time Invariant (LTI) systems has been extensively studied in the past two decades and both time and frequency domain results are available in the literature (see for instance [25], [4], [27] and references therein). A related line of research is model (in)validation of Linear Parameter Varying (LPV) systems ([30], [3]) 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

¹Note that some of these identification methods are for switched ARX models and some for switched output-error models. Invalidation is an important problem for both classes of models but the focus of the current paper is output-error models as invalidation problem for switched ARX models is relatively easier as discussed later in the paper. Also note that the proposed invalidation method is agnostic to how the output-error model is obtained and can be used regardless of whether it is derived from first principles or is the result of applying any identification technique to experimental data.

addressed in [26] using sum of squares methods and barrier functions. However, to the best of our knowledge, similar results for discrete-time switched affine systems with unknown switches are not available in the literature,² 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 model to be (in)validated by the experimental data. More precisely, given a set of affine output-error models and experimental input/output data, certificates are provided for the existence or nonexistence of a switching sequence such that the resulting output of the switched model interpolates the given experimental data within a given noise bound. In order to develop such certificates, we start by showing that determining whether a given model is consistent with the observed experimental data can be accomplished without estimating the (unknown) switching sequence. Rather, this problem can be reduced to checking whether a suitably defined semialgebraic set is empty or not. Then, 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, infinite dimensional convex optimization problem. In principle, this would require the solving of a sequence of convex programs of increasing complexity. However, as shown in the paper, the process stops in a finite number of steps bounded above by a constant that depends only on the number of data points and the orders of the models involved. While this bound on the number of steps is usually very large, this result is one of the very few available guaranteeing that emptiness of a certain class of semi algebraic sets can be established by solving a convex optimization problem with fixed size. Moreover, if while solving intermediate problems of smaller size either a strictly 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 can be stopped. The former case provides an invalidation certificate, while the latter shows that the model and data observed so far are consistent. 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. Finally, we further extend these results to invalidation of

²While this paper was under review, the invalidation problem was also considered in [5] where a mixed integer programming approach is used to impose constraints on switching

models with parametric uncertainty and provide algorithms for computing uncertainty bounds based on additional experimental data.

In the last part 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 video streams from multiple sources. However, interesting (e.g. abnormal) activities are infrequent, typically appearing in just a few frames. Thus, in order to avoid errors due to human operator overload, it is important to automatically flag these frames for further inspection, while discarding the portions of the data guaranteed not to contain “contextually abnormal” events. As we show in Section VI-C, this problem can be recast into a piecewise-affine model invalidation form and solved using the framework developed in this paper.

A preliminary version of this work has appeared in a conference paper [20], parts of which were included in the book chapter [9] (see also [18]). The current paper, in addition to containing more in-depth discussions on the methodology, extends the earlier versions in three aspects: (i) we provide a global upper bound for convergence of the semidefinite programming hierarchy for the invalidation task, (ii) we extend the results to the case where the a-priori model contains parametric uncertainty and provided a reformulation to compute parameter bounds, and (iii) we include more examples.

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 results of this paper.

A. Notation and Definitions

The set of real numbers is denoted by \mathbf{R} . By \mathbf{x} and \mathbf{M} , we denote a vector in \mathbf{R}^n and a matrix in $\mathbf{R}^{n \times m}$, respectively. The j^{th} entry of a vector \mathbf{x} is denoted by $x^{(j)}$. $\|\mathbf{x}\|_\infty \doteq \sup_i |x^{(i)}|$ is the ∞ -norm of a vector. $\|\mathbf{M}\|_{max}$ is the entrywise infinity norm of the matrix \mathbf{M} . $\mathbf{M} \succeq \mathbf{N}$ indicates that the matrix $\mathbf{M} - \mathbf{N}$ is positive semidefinite. $\langle \mathbf{M}, \mathbf{N} \rangle \doteq \text{trace}(\mathbf{M}^T \mathbf{N})$, is the inner product of two matrices \mathbf{M} and \mathbf{N} . $\mathbf{R}[x_1, \dots, x_n]$ denotes the ring of polynomials in n variables over \mathbf{R} . When n is clear from context, $\mathbf{R}[\mathbf{x}]$ is used for short. The mean value of the function $p(\mathbf{x})$

w.r.t the Borel measure μ on the random variable \mathbf{x} is denoted by $\mathcal{E}_\mu[p(\mathbf{x})]$. Given a set I , $|I|$ denotes its cardinality. Finally, \mathbf{N}_n is the set of positive integers up to n , i.e. $\mathbf{N}_n \doteq \{1, \dots, n\}$.

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 $u_1, \dots, u_m \in \mathbf{R}[\mathbf{x}]$.

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

B. The problem of moments and polynomial optimization.

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

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

where $K \subset \mathbf{R}^n$ is a compact semialgebraic set defined by c polynomial inequalities of the form $g_k(\mathbf{x}) \geq 0$, $k = 1, \dots, c$. In the sequel, we briefly summarize some results relating polynomial optimization to the problem of moments that is used to recast problem (P1) (and hence model invalidation) into a (possibly infinite-dimensional) convex optimization form (see [12], [13], [16] for more details).

1) *The problem of moments:* Let K be a compact subset of \mathbf{R}^n . 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 Borel measure μ supported on K that has $\{m_\alpha\}$ as its α^{th} moments. That is:

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

where $\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$ (for a historical review and details of the problem, see [6], [28] and references therein). As shown in [12], [6], the existence of such a measure can be characterized by positive semidefiniteness of some infinite matrices, the so-called moment $\mathbf{M}(m_\alpha)$ and localization matrices $\mathbf{L}(g_k m_\alpha)$ where $g_k(\mathbf{x}) \geq 0$ are the polynomials defining K .

Next, we briefly discuss how to build truncated versions of \mathbf{M} and \mathbf{L} of a given sequence $\mathbf{m} \doteq \{m_\alpha\}$ 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

$\mathbf{0} = \alpha^{(1)} < \dots < \alpha^{(S_N)}$, where $S_N \doteq \binom{N+n}{n}$ is the number of moments in \mathbf{R}^n up to order N . The truncated version of \mathbf{M} is defined as follows:

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

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

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

2) *Moments-based polynomial optimization*: In this section, we recall some results from [12] 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:

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

where $\mathcal{P}(K)$ is the space of Borel measures on K with $\mu(K) = 1$. Although (P2) is an infinite dimensional problem, it is, in contrast to (P1), convex. The next result, taken from [12], establishes the relation between the two problems:

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

- $\tilde{p}_K^* = p_K^*$.
- If \mathbf{x}^* is a global minimizer of (P1), then the Dirac measure $\mu^* = \delta_{\mathbf{x}^*}$ with a point support at \mathbf{x}^* is a global minimizer of (P2).
- For every optimal solution μ^* of (P2), $p(x) = p_K^*$, except in a set of μ^* -measure zero.

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 Borel measure by using (2) and (3) to define the constraint set. To this effect, define

$$\begin{aligned} p_N^* &= \min_{\mathbf{m}} \sum_{\alpha} p_{\alpha} m_{\alpha} \\ \text{s.t.} \quad &\mathbf{M}_N(\mathbf{m}) \succeq 0, \\ &\mathbf{L}_N(g_k \mathbf{m}) \succeq 0, k = 1, \dots, c, \end{aligned} \quad (4)$$

where p_{α} is the coefficient of the monomial \mathbf{x}^{α} in the polynomial $p(\mathbf{x})$, i.e. $p(\mathbf{x}) = \sum_{\alpha} p_{\alpha} \mathbf{x}^{\alpha}$. Then, according to Theorem 4.2 in [12], we have:

Theorem 2. As N goes to infinity, p_N^* in (4) approaches to p_K^* from below.

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. Consider problem (P1) where the constraint set $K \in \mathbf{R}^n$ is defined by c polynomials g_k . Let $\{I_i\}_{i=0}^{c'}$ be a collection of subsets of variables in $X \doteq \{x_1, \dots, x_n\}$, satisfying $\bigcup_{i=0}^{c'} I_i = X$. Suppose each g_k contains variables only in I_i for some i . Also, assume the objective function p can be partitioned as $p(\mathbf{x}) = p_1(\mathbf{x}) + \dots + p_l(\mathbf{x})$ where each p_k contains variables only in I_i for some i . Then, the running intersection property is satisfied in Problem (P1) if the collection $\{I_i\}_{i=0}^{c'}$ satisfies

$$I_{i+1} \cap \bigcup_{j=0}^i I_j \subseteq I_r \text{ for some } r \leq i. \quad (5)$$

Similar to the convergence result stated in Theorem 2, for a polynomial optimization problem that satisfies the *running intersection property*, it is possible to construct a hierarchy of semidefinite programs of smaller size. Assume that (P1) satisfies the running intersection property with the family of subsets $\{I_i\}_{i=0}^{c'}$ and partition of the objective function $\{p_j\}_{j=1}^l$. Let

$$\begin{aligned} p_N^* &= \min_{\mathbf{m}} \sum_{j=1}^l \sum_{\alpha(j)} p_{j,\alpha(j)} m_{\alpha(j)} \\ \text{s.t.} \quad & \mathbf{M}_N(\mathbf{m}_{I_i}) \succeq 0, i = 0, \dots, c', \\ & \mathbf{L}_N(g_k \mathbf{m}_{I_{i(k)}}) \succeq 0, k = 1, \dots, c, \end{aligned} \quad (6)$$

where $p_{j,\alpha(j)}$ is the coefficient of the $\alpha(j)^{th}$ monomial in the polynomial p_j , $\mathbf{M}_N(\mathbf{m}_{I_i})$ denote the moment matrix and $\mathbf{L}_N(g_k \mathbf{m}_{I_{i(k)}})$ the localizing matrix for the subset of variables in $I_{i(k)}$. Next result is adapted from [13].

Theorem 3 (Sparse Polynomial Optimization). As N goes to infinity, p_N^* in (6) approaches to p_K^* from below.

For the case of generic polynomials and constraints, solving problem (P1) using the method of moments for a given order N requires considering moments and localization matrices containing $O(n^{2N})$ variables. On the other hand, if the running intersection property holds, using Theorem 3 it is possible to define $c + c' + 1$ smaller sized matrices each containing variables only in some I_i

(i.e. number of variables is $O(c'\kappa^{2N})$, where $\kappa \doteq \max_{0 \leq i \leq c'} |I_i|$). 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 SWITCHED OUTPUT ERROR MODELS

We now provide a formal definition of the (in)validation problem addressed in this paper and show that it is equivalent to finding if a suitable polynomial optimization problem has zero as its optimum. Computationally tractable (in)validation certificates are then obtained by exploiting results from polynomial optimization and semi algebraic geometry.

A. Problem Statement

The systems considered in this paper are the so-called multi-input, multi-output (MIMO) switched affine systems which have an output error model structure of the form:

$$\begin{aligned} \mathbf{y}_t &= \sum_{k=1}^{n_a} \mathbf{A}_k(\sigma_t) \mathbf{y}_{t-k} + \sum_{k=1}^{n_c} \mathbf{C}_k(\sigma_t) \mathbf{u}_{t-k} + \mathbf{f}(\sigma_t) \\ \tilde{\mathbf{y}}_t &= \mathbf{y}_t + \boldsymbol{\eta}_t \end{aligned} \quad (7)$$

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 . No dwell-time assumptions are made, hence the mode signal σ_t can switch arbitrarily fast among the s submodels G_i , each of which is associated with the set of its coefficient matrices $\{\mathbf{A}_1(i), \dots, \mathbf{A}_{n_a}(i), \mathbf{C}_1(i), \dots, \mathbf{C}_{n_c}(i), \mathbf{f}(i)\}$.

In this context, the model (in)validation problem addressed in this paper can be formally stated as:

Problem 1. *Given a nominal hybrid model of the form (7) with s submodels G_1, \dots, G_s , an a priori bound ϵ on the ℓ_∞ norm of the noise, and experimental data $\{\mathbf{u}_t, \tilde{\mathbf{y}}_t\}_{t=0}^T$, determine whether or not there exists a switching sequence $\sigma = \{\sigma_t\}_{t=0}^T$ such that the corresponding trajectory of (7), corrupted by admissible noise $\boldsymbol{\eta}$ can explain the observed data, or, equivalently whether or not the consistency set*

$$\mathcal{T} = \{(\boldsymbol{\eta}, \sigma) \mid \|\boldsymbol{\eta}_t\|_\infty \leq \epsilon, \sigma_t \in \mathbf{N}_s \text{ subject to (7)} \forall t \in [0, T]\}$$

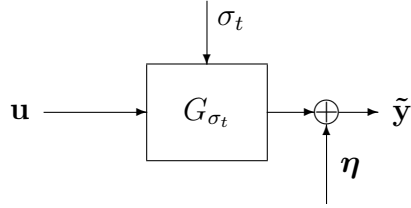


Fig. 1. Problem Setup. The coefficient matrices of the submodels G_i and a bound on the noise are known *a priori*. The experimental data consists of input/output measurements, \mathbf{u} and $\tilde{\mathbf{y}}$. The mode signal σ_t and noise sequence η are unknown.

is nonempty.

Remark 1. *It should be emphasized that the equations (7) are coupled through the noise terms. Indeed, if the i^{th} sub model is active at time t , (7) is equivalent to*

$$\begin{aligned} & \mathbf{A}_1(i)(\tilde{\mathbf{y}}_{t-1} - \boldsymbol{\eta}_{t-1}) + \dots + \mathbf{A}_{n_a}(i)(\tilde{\mathbf{y}}_{t-n_a} - \boldsymbol{\eta}_{t-n_a}) \\ & - (\tilde{\mathbf{y}}_t - \boldsymbol{\eta}_t) + \mathbf{C}_1(i)\mathbf{u}_{t-1} + \dots + \mathbf{C}_{n_c}(i)\mathbf{u}_{t-n_c} + \mathbf{f}(i) = \mathbf{0} \end{aligned} \quad (8)$$

which consists of n_y linear equations in $n_a + 1$ unknown noise vectors $\boldsymbol{\eta}_{t-n_a:t}$. Therefore, Problem 1 cannot be solved by just looking at feasibility of s sets of equations (each corresponding to one subsystem) one-at-a-time. Rather, to proceed along these lines, one will be forced to look at all possible combinations of subsystems over the relevant time horizon. On the other hand, for switched autoregressive models with exogenous inputs (ARX), where the noise terms are decoupled, a similar invalidation problem reduces to checking the feasibility of a set of linear inequalities and is relatively easier to solve.

B. A Convex Certificate for (In)validating MIMO Switched Output Error Models

In this section, we present the main result of the paper, convex necessary and sufficient invalidation certificates in the form of strict positivity of the solution to a finite-dimensional convex optimization problem. Towards this goal, we start by defining a second consistency set \mathcal{T}' , independent of the switching signal σ_t , whose non-emptiness is equivalent to the non-emptiness of the original consistency set \mathcal{T} . Note that if the i^{th} submodel is active at time t

(i.e., $\sigma_t = i$), the following should hold:

$$[h_{t,i}^{(1)}(\boldsymbol{\eta}_{t-n_a:t}) = 0] \wedge \dots \wedge [h_{t,i}^{(n_y)}(\boldsymbol{\eta}_{t-n_a:t}) = 0] \quad (9)$$

where \wedge denotes the logical AND operation and $h_{t,i}^{(j)}(\boldsymbol{\eta}_{t-n_a:t}) = 0$ denotes the equation corresponding to the j^{th} row of the system of equations (8). Algebraically, the expression above can be written as:

$$g_{t,i}(\boldsymbol{\eta}_{t-n_a:t}) \doteq \sum_{j=1}^{n_y} \left[h_{t,i}^{(j)}(\boldsymbol{\eta}_{t-n_a:t}) \right]^2 = 0. \quad (10)$$

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. (10) should hold true for some $i \in \{1, \dots, s\}$ and some admissible value of the noise $\boldsymbol{\eta}_{t-n_a:t}$. This condition can be expressed as

$$[g_{t,1}(\boldsymbol{\eta}_{t-n_a:t}) = 0] \vee \dots \vee [g_{t,s}(\boldsymbol{\eta}_{t-n_a:t}) = 0] \quad (11)$$

where \vee denotes the logical OR operation, or algebraically³

$$p_t(\boldsymbol{\eta}_{t-n_a:t}) \doteq \prod_{i=1}^s g_{t,i}(\boldsymbol{\eta}_{t-n_a:t}) = 0. \quad (12)$$

Hence, consistency between the *a priori* and *a posteriori* information on a switched affine system is equivalent to the existence of an admissible solution $\boldsymbol{\eta}_{t-n_a:t}$ to the set of polynomial equalities (12), leading to the following result. The proof is omitted since it is an immediate consequence of the definitions of the polynomials above and the bounds on the measurement noise.

Lemma 1. *Let*

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

where

$$f_{t,j}(\eta_t^{(j)}) \doteq \epsilon^2 - \left[\eta_t^{(j)} \right]^2 \quad (13)$$

Then, \mathcal{T} is empty if and only if \mathcal{T}' is empty.

The result above shows that the problem of model (in)validation is equivalent to checking whether a set defined by polynomial inequalities is empty or not. Hence, one could use the

³This idea is similar to the hybrid decoupling constraint proposed in [17].

Positivstellensatz, and derive sum of squares certificates for the emptiness of \mathcal{T}' , as proposed in [24] for general polynomial inequalities. However, such an approach does not take into account the structure of the problem at hand and can lead to very complex semidefinite optimization problems. A closer look at the polynomials used in the definition of \mathcal{T}' reveals an underlying sparse structure that can be exploited to substantially reduce the computational complexity entailed in checking its emptiness.

We now elaborate on this. Consider the following optimization problem

$$\begin{aligned} o^* = \min_{\boldsymbol{\eta}} \quad & \sum_{t=n_a}^T p_t(\boldsymbol{\eta}_{t-n_a:t}) \\ \text{s.t.} \quad & f_{t,j}(\eta_t^{(j)}) \geq 0 \quad \forall t \in [0, T], j \in \mathbb{N}_{n_y}. \end{aligned} \quad (14)$$

We now argue that

$$o^* \geq 0 \quad \text{and} \quad o^* = 0 \Leftrightarrow \mathcal{T}' \neq \emptyset.$$

The inequality $o^* \geq 0$ is a consequence of the fact that the objective function in (14) is a sum of squares polynomial.⁴ Moreover, given the definition of the set \mathcal{T}' and the optimization problem (14), o^* is equal to zero if and only if there exists a noise sequence that satisfies the polynomial constraints defining the set \mathcal{T}' or, in other words, if and only if \mathcal{T}' is nonempty. Equivalently, $o^* > 0$ if and only if $\mathcal{T}' = \emptyset$.

The main advantage in reformulating the (in)validation problem as an optimization problem such as (14) is the fact that one can now exploit the results on sparse polynomial optimization reviewed in Section II-B3. More precisely, the following result holds.

Lemma 2. *Problem (14) satisfies the running intersection property.*

Proof. Consider the collection $\{I_i\}_{i=0}^{T-n_a}$ of subsets of the variables $\boldsymbol{\eta}_{0:T}$ where $I_i = \{\boldsymbol{\eta}_{i:i+n_a}\}$ for all i . Note that $\bigcup_{i=0}^{T-n_a} I_i = \{\boldsymbol{\eta}_{0:T}\}$. Now consider the polynomials $f_{t,j}(\eta_t^{(j)})$ and $p(\boldsymbol{\eta}_{0:T}) \doteq \sum_{t=n_a}^T p_t(\boldsymbol{\eta}_{t-n_a:t})$ defining, respectively, the constraints and the objective function of the problem (14). For $t \leq n_a$, each $f_{t,j}$ contains variables only from I_0 ; and for $t > n_a$, each $f_{t,j}$ contains variables only from I_{t-n_a} . Similarly, each p_t contains variables only from I_{t-n_a} . Moreover,

$$I_{i+1} \cap \bigcup_{j=0}^i I_j = \{\boldsymbol{\eta}_{(i+1):(i+n_a+1)}\} \cap \{\boldsymbol{\eta}_{0:(i+n_a)}\} = \{\boldsymbol{\eta}_{(i+1):(i+n_a)}\} \subset I_i,$$

⁴Since it is formed by multiplication and addition of SoS polynomials in (10), and the cone of SoS polynomials is closed under these operations.

for all i , satisfying (5) with $r = i$. Therefore, the running intersection property in Definition 3 holds. \square

The result above, together with a moments approach to polynomial optimization, can be used to formulate a sequence of convex optimization problems whose solution converges to o^* and sparse structure is exploited to reduce computational complexity. Specifically, define the following convex optimization problem

$$\begin{aligned}
d_N^* = \min_{\mathbf{m}} \quad & \sum_{t=n_a}^T l_t(\mathbf{m}_{t-n_a:t}) \\
\text{s.t.} \quad & \mathbf{M}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \quad \forall t \in [n_a, T] \\
& \mathbf{L}_N(f_{t,j} \mathbf{m}_{t-n_a:t}) \succeq 0 \quad \forall t \in [n_a + 1, T], j \in \mathbf{N}_{n_y} \\
& \mathbf{L}_N(f_{t,j} \mathbf{m}_{0:n_a}) \succeq 0 \quad \forall t \in [0, n_a], j \in \mathbf{N}_{n_y}
\end{aligned} \tag{15}$$

where each l_t is the linear functional of moments defined as $l_t(\mathbf{m}_{t-n_a:t}) \doteq \mathcal{E} \{p_t(\boldsymbol{\eta}_{t-n_a:t})\}$, 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 $N \geq s$. Note that, in the problem above, we do not use a “global” moment matrix $\mathbf{M}_N(\mathbf{m}_{t-n_a:T})$. Instead, we use smaller $\mathbf{M}_N(\mathbf{m}_{t-n_a:t})$, leading to a substantial reduction in the number of variables used; see Remark 3.

Again, the fact that the objective function of (14) is a sum of squares polynomial implies that $d_N^* \geq 0$ for all N . Moreover, application of Theorem 2 implies that $d_N^* \uparrow o^*$ as $N \rightarrow \infty$. Therefore, one has the following necessary and sufficient conditions for model (in)validation.

Theorem 4. *The following statements are equivalent:*

- (i) *The consistency set \mathcal{T}' 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 zero⁵:*

$$\begin{aligned}
r^* = \min_{\mathbf{m}} \quad & \sum_{t=n_a}^T l_t(\mathbf{m}_{t-n_a:t}) \\
\text{s.t.} \quad & \mathbf{M}_s(\mathbf{m}_{t-n_a:t}) \succeq 0 \quad \forall t \in [n_a, T] \\
& \text{rank} [\mathbf{M}_s(\mathbf{m}_{t-n_a:t})] = 1 \quad \forall t \in [n_a, T] \\
& f_{t,j}(\mathbf{m}_{t-n_a:t}) \geq 0 \quad \forall t \in [n_a + 1, T], j \in \mathbf{N}_{n_y} \\
& f_{t,j}(\mathbf{m}_{0:n_a}) \geq 0 \quad \forall t \in [0, n_a], j \in \mathbf{N}_{n_y}
\end{aligned} \tag{16}$$

⁵With slight abuse of notation, in (16) $f_{t,j}$ is used to denote the linear functional of moments defined by $f_{t,j}(\mathbf{m}_{t-n_a:t}) \doteq \mathcal{E} \{f_{t,j}(\boldsymbol{\eta}_{t-n_a:t})\}$.

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) \Leftrightarrow (ii) Recall that $\mathcal{T}' = \emptyset \iff o^* > 0$. Since $d_N^* \uparrow o^*$ as $N \rightarrow \infty$, 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}' is empty.

(i) \Leftrightarrow (iii) To prove this equivalence, we show that r^* in (16) is equal to o^* in (14). Assume $\boldsymbol{\eta}^*$ is an optimizer of (14), then the moments of the Borel measure μ^* with point support at $\boldsymbol{\eta}^*$ is feasible for (16) with the same objective value which implies $r^* \leq o^*$. On the other hand, if \mathbf{m}^* is a minimizer of (16), the rank condition implies that there is a corresponding measure μ^* with point support, say at $\boldsymbol{\eta}^*$. Since this value of $\boldsymbol{\eta}^*$ is a feasible point of (14), $o^* \leq r^*$. Therefore, $r^* = o^*$ from which we conclude that $r^* > 0$ is equivalent to (i). \square

The theorem above states that if the consistency set is empty, then there exists a finite relaxation order N_o such that $d_{N_o}^* > 0$. Remarkably, as shown below, a general upper bound on the relaxation order can be explicitly computed from the *a-priori* information.

Theorem 5. *Assume that the consistency set \mathcal{T}' is empty. Then $d_N^* > 0$ for all*

$$N \geq s^{T-n_a+1} + 1 \doteq N_o \quad (17)$$

Proof. See Appendix A. \square

Remark 2. *The bound above is significant from a theoretical standpoint, since it is one of the very few results of this type available in the literature for constrained polynomial optimization problems involving real rather than discrete-valued variables. On the other hand, from a practical standpoint, this bound is usually conservative, leading to optimization problems that may be too large. In some cases, the consistency of the experimental data record with the model and a priori assumptions can be certificated at smaller relaxation orders by resorting to a variation of the so-called flat extension property for sparse polynomial optimization stated in [13]. In particular, if for some $N < N_o$, $\tilde{p}_N \leq 0$ and the solution \mathbf{m}^* of (15) satisfies $\text{rank} [\mathbf{M}_N(\mathbf{m}_{0:n_a}^*)] = 1$; $\text{rank} [\mathbf{M}_N(\mathbf{m}_{t-n_a:t-1}^*)] = 1$ and $\text{rank} [\mathbf{M}_N(\mathbf{m}_{t-n_a:t}^*)] = \text{rank} [\mathbf{M}_{N-2}(\mathbf{m}_{t-n_a:t}^*)] \quad \forall t \in [n_a, T]$, then this certifies that $o^* = d_N^* = 0$; hence \mathcal{T}' is not empty.*

Remark 3. It is important to highlight the complexity reduction achieved by employing the running intersection property established in Lemma 2 while forming the optimization problem (15). The conventional moment relaxation of order N in [12] would require $O((Tn_y + n_y)^{2N})$ optimization variables with a moment matrix of the size $\begin{pmatrix} N + Tn_y + n_y \\ Tn_y + n_y \end{pmatrix}$. On the other hand, (15) involves only $O((n_a n_y + n_y)^{2N})$ variables with $T - n_a + 1$ moment matrices of the size $\begin{pmatrix} N + n_a n_y + n_y \\ n_a n_y + n_y \end{pmatrix}$ 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

As mentioned in the previous section, the relaxation bound provided in Theorem 5 can be very conservative, and it is often possible to assert whether the data collected and the *a priori* conditions are consistent or not by using a value of N much lower than N_0 . Hence, in practice, one solves a sequence of problems with increasing N until either consistency with the model is established or one obtains a value of d_N^* strictly greater than zero and the model is deemed invalid. To do this, for each N , it is necessary to determine if either $d_N^* = 0$ or $d_N^* > 0$.

This leads to a potential numerical difficulty. All numerical solvers have finite precision and, in practice, they will almost always give a numerical solution to problem (15) that is non-zero. To address this problem, we propose to use the dual of (15) to obtain a more numerically robust certificate of strict positiveness of d_N^* . We now elaborate on this.

The constraints in problem (15) are Linear Matrix Inequality (LMI) constraints and, hence, it can be represented in the standard dual form of semidefinite programs; i.e., it can be reformulated as

$$\begin{aligned} d_N^* = \inf_{\mathbf{m}} \quad & \mathbf{b}^T \mathbf{m} + c_o \\ \text{s.t.} \quad & \sum_{\alpha \in \mathcal{I}} \mathbf{A}_\alpha m_\alpha \succeq \mathbf{C} \end{aligned} \quad (18)$$

where \mathcal{I} is the set of the multi-indexes of all moments that occur in (15) except the zeroth moment $m_0 = 1$ which is used to form the constant terms \mathbf{C} and c_o . Consider now the corresponding primal problem which is of the form

$$\begin{aligned} p_N^* = \max_{\mathbf{X}} \quad & \langle \mathbf{C}, \mathbf{X} \rangle + c_o \\ \text{s.t.} \quad & \langle \mathbf{A}_\alpha, \mathbf{X} \rangle = b_\alpha \quad \forall \alpha \in \mathcal{I} \\ & \mathbf{X} \succeq 0 \end{aligned} \quad (19)$$

where, without loss of generality, \mathbf{X} can be chosen to have the same block-diagonal structure of \mathbf{A}_α s and \mathbf{C} . Results on duality show that $p_N^* \leq d_N^*$ and, hence, $p_N^* > 0$ is a certificate for model invalidation. Moreover, if (19) is a feasible problem, then $p_N^* = d_N^*$ and

$$\mathcal{T}' = \emptyset \quad \Leftrightarrow \quad p_{N^*}^* \text{ for some } N^* \leq N_0.$$

Returning to numerical considerations, let \tilde{d}_N and \tilde{p}_N be the solutions to problems (15) and (19) respectively, obtained using any numerical solver. Then,

$$\tilde{p}_N \leq p_N^* = d_N^* \leq \tilde{d}_N$$

and, hence, one can always use $\tilde{p}_N > 0$ as a certificate for emptiness of the set \mathcal{T}' . In other words, strict positivity of \tilde{p}_N is an invalidation certificate.

V. HANDLING PARAMETRIC UNCERTAINTY

In this section, we discuss model invalidation in the presence of parametric uncertainty. To this effect, we consider the following model:

$$\begin{aligned} \mathbf{y}_t &= \sum_{k=1}^{n_a} [\mathbf{A}_k(\sigma_t) + \Delta_{\mathbf{A}_k(\sigma_t)}] \mathbf{y}_{t-k} + \sum_{k=1}^{n_c} [\mathbf{C}_k(\sigma_t) + \Delta_{\mathbf{C}_k(\sigma_t)}] \mathbf{u}_{t-k} \\ \tilde{\mathbf{y}}_t &= \mathbf{y}_t + \boldsymbol{\eta}_t \end{aligned} \quad (20)$$

where Δ_X (for $X \in \{\mathbf{A}_i, \mathbf{C}_j\}$ with $i \in [1, n_a], j \in [1, n_c]$) represent bounded parametric uncertainty (i.e. $\|\Delta_X\|_{max} \leq \gamma \quad \forall X$). Also, let Δ denote the the set of all uncertain parameters, i.e., all entries in Δ_X for all X . Following along lines similar to those in Section III-A, a consistency set $\mathcal{T}(\boldsymbol{\eta}, \Delta, \sigma)$ and its semialgebraic equivalent $\mathcal{T}'(\boldsymbol{\eta}, \Delta)$ can be defined. All the results of Sections III-B and IV (including the running intersection property) extend to the case with parametric uncertainty in a straight forward way. This only requires additional variables, due to Δ_X 's, in the optimization problems.

Next, we propose a modification of the optimization problem that allows to use input/output data from invalidation experiments to obtain bounds on the uncertain parameters. Let us denote an entry of an uncertainty matrices Δ_X by δ_l for $l \in \mathbb{N}_{|\Delta|}$. For any bound $\gamma > 0$ on the uncertainty, we define the polynomials $q_1(\delta_l) \doteq \gamma - \delta_l$ and $q_2(\delta_l) \doteq \delta_l + \gamma$. Now, consider the

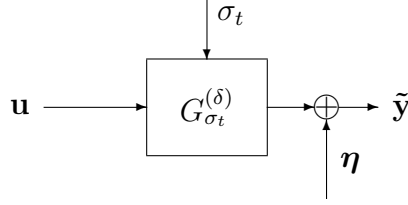


Fig. 2. Problem setup with parametric uncertainty. The coefficient matrices of the submodels G_i is subject to parametric uncertainty δ . The parametric form of $G_i^{(\delta)}$ and a bound on the noise and uncertainty are known *a priori*. The experimental data consists of input/output measurements, \mathbf{u} and $\tilde{\mathbf{y}}$. The mode signal σ_t , the noise sequence η , and the actual value of the uncertain parameters δ are unknown.

following polynomial optimization problem:

$$\begin{aligned}
 \gamma^* = \arg \min_{\gamma, \Delta, \eta} \quad & \gamma \\
 \text{s.t.} \quad & p_t(\boldsymbol{\eta}_{t-n_a:t}, \Delta) = 0 \quad \forall t \in [n_a, T], \\
 & q_1(\delta_l) \geq 0, q_2(\delta_l) \geq 0 \quad \forall l \in \mathbf{N}_{|\Delta|}, \\
 & f_{t,j}(\eta_t^{(j)}) \geq 0 \quad \forall t \in [0, T] \quad \forall j \in \mathbf{N}_{n_y}.
 \end{aligned} \tag{21}$$

The solution of (21), γ^* , is such that the for all uncertainties with $\|\Delta_X\|_{max} < \gamma^*$, the model is invalidated. Moreover, for any bound γ' larger than γ^* , one can find an uncertainty with $\|\Delta_X\|_{max} \leq \gamma'$ and noise sequence that would not invalidate the model.

Note that (21) is a polynomial optimization problem satisfying the running intersection property. Hence, we can resort to moments-based relaxations for sparse polynomial optimization to construct a series of convex programs to find the uncertainty bound γ^* . In particular, it is possible to find a lower bound for γ^* by solving the semidefinite program:

$$\begin{aligned}
 \tilde{\gamma}_N = \arg \min_{\mathbf{m}} \quad & \mathcal{E}(\gamma) \\
 \text{s.t.} \quad & \mathbf{M}_N(\mathbf{m}, I_t) \succeq 0 \quad \forall t \in [0, T - n_a + 1] \\
 & \mathbf{L}_N(p_t \mathbf{m}, I_{t-n_a}) \succeq 0 \quad \forall t \in [n_a, T] \\
 & \mathbf{L}_N(p_t \mathbf{m}, I_{t-n_a}) \preceq 0 \quad \forall t \in [n_a, T] \\
 & \mathbf{L}_N(f_{t,j} \mathbf{m}, I_{\beta(t)-n_a}) \succeq 0 \quad \forall t \in [0, T], \forall j \in \mathbf{N}_{n_y} \\
 & \mathbf{L}_N(q_1 \mathbf{m}, I_t) \succeq 0 \quad \forall t \in [0, T - n_a + 1] \\
 & \mathbf{L}_N(q_2 \mathbf{m}, I_t) \succeq 0 \quad \forall t \in [0, T - n_a + 1]
 \end{aligned} \tag{22}$$

where $I_t = \{\gamma, \Delta_X, \boldsymbol{\eta}_k : \forall X, \forall k \in [t, t + n_a]\}$ is a partial variable set with $\beta(t) = n_a$ for $t \leq n_a$ and $\beta(t) = t$ for $t > n_a$. Note that the solution, $\tilde{\gamma}_N$, of (22) gives a lower bound for γ^* (i.e. $\tilde{\gamma}_N \leq \gamma^*$). Hence the model is invalidated for all uncertainties with $\|\Delta_X\|_{max} \leq \tilde{\gamma}_N \forall X$. Moreover, it is possible to obtain an arbitrarily good approximation of γ^* by increasing the relaxation order N as stated in the next theorem.

Theorem 6. *As the relaxation order N increases in (22), $\tilde{\gamma}_N$ converges to γ^* . That is as $N \rightarrow \infty$, $\tilde{\gamma}_N \uparrow \gamma^*$.*

VI. 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 SparsePOP [31] and SEDUMI [29].

A. Academic Examples

We consider the following submodels:

$$y_t = 0.2y_{t-1} + 0.3y_{t-2} + 2u_{t-1} \quad (G_1)$$

$$y_t = -1.5y_{t-1} - 0.5y_{t-2} + 1.2u_{t-1} \quad (G_2)$$

$$y_t = 1.7y_{t-1} - 0.7y_{t-2} + 0.6u_{t-1} \quad (G_3)$$

and the measurement equation:

$$\tilde{y}_t = y_t + \eta_t. \quad (23)$$

We ran different sets of simulations involving 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, the numerical solution of the optimization problem (19), \tilde{p}_s^* , 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 $\tilde{p}_s^* < 0$.

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 with a uniform distribution and with uniform random noise $\|\eta_t\| \leq 0.5$. 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	\tilde{p}_s	Result
G_1, G_2, G_3	G_1, G_2	-1.1484×10^{-8}	not invalidated
G_1, G_2	G_1, G_2, G_3	0.9134	invalidated

TABLE I

INVALIDATION RESULTS FOR EXAMPLE 1.

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 with a uniform distribution and with uniform random noise $\|\eta_t\| \leq \epsilon$. 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.*

Submodels	A priori ϵ	Actual ϵ	\tilde{p}_s	Result
G_1, G_2, G_3	0.8	1	0.0040	invalidated
G_1, G_2, G_3	1.2	1	-2.8520×10^{-7}	not invalidated

TABLE II

INVALIDATION RESULTS FOR EXAMPLE 2.

Example 3. (Submodel perturbation) *For this set of experiments, we generated input/output data by perturbing the coefficients of a priori submodels $\{G_1, G_2, G_3\}$, and again using a random switching sequence σ_t with a uniform distribution and uniformly sampled random noise $\|\eta_t\| \leq 0.5$. We denote the submodel whose coefficient values are perturbed by e percent of their original values as $G_i + E_{e\%}$. This case is more challenging than the one considered in Example*

I 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	\tilde{p}_s	Result
G_1, G_2, G_3	$G_1 + E_{1\%}, G_2 + E_{1\%}, G_3 + E_{1\%}$	0.5714	invalidated
G_1, G_2	$G_1 + E_{5\%}, G_2 + E_{5\%}$	0.8454	invalidated
G_1, G_2	$G_1 + E_{2\%}, G_2 + E_{2\%}$	0.2242	invalidated
G_1, G_2	$G_1 + E_{2\%}$	5.2776×10^{-4}	invalidated

TABLE III

INVALIDATION RESULTS FOR EXAMPLE 3.

To give a sense of computational complexity and to demonstrate the advantages of exploiting the problem structure, we also report the time required to solve the problem on a 2.2GHz Intel Core Duo Laptop with 2GB of memory running Windows. Table IV summarizes mean computation times when the number of submodels used as part of the a priori information increases. Table V summarizes how the mean computation time scales with increasing time horizon over which data is collected when the a priori information contains three submodels. Computation times were averaged over six runs with random noise and input values. Note that it is not possible to solve any of these problems without exploiting the sparse problem structure on the same platform due to insufficient memory. Although even when exploiting the structure, we hit the memory bound for three submodels and $T = 600$ data points, on-going research indicates that memory efficiency could be further improved by employing first order optimization techniques that, as opposed to interior point methods, do not require storing the Hessian.

s	2	3	4	5
time (sec)	10.5	41.5	147	602

TABLE IV

MEAN COMPUTATION TIME AS THE NUMBER OF SYSTEMS s INCREASES.

B. Uncertainty Estimation

Consider the following two models with parametric model uncertainty in their input channels:

$$y_t = 0.2y_{t-1} + 0.24y_{t-2} + (2 + \delta_1)u_{t-1}, \quad (G_1^{(\delta)})$$

T	96	150	300	600
time (sec)	41.5	62	158	N/A

TABLE V

MEAN COMPUTATION TIME AS THE NUMBER OF INPUT/OUTPUT DATA POINTS, T , INCREASES (N/A STANDS FOR OUT OF MEMORY ERROR).

$$y_t = -1.4y_{t-1} - 0.53y_{t-2} + (1 + \delta_2)u_{t-1}, \quad (G_2^{(\delta)})$$

and the measurement equation same as before:

$$\tilde{y}_t = y_t + \eta_t. \quad (24)$$

The input/output data was generated from this model with mode signal $\sigma_t = 1$ for $t \in [0, 24] \cup [49, 72]$ and $\sigma_t = 2$ for $t \in [25, 48] \cup [73, 96]$, and noise, η , was independently uniformly randomly sampled from the interval $[-0.5, 0.5]$, (i.e., $\epsilon = 0.5$). We ran the experiment with different values of δ_1 and δ_2 (unknown to the algorithm), and estimated the lower bound on uncertainty as described in Section V. Table VI summarizes the results. Note that in this case, the optimal uncertainty bound γ^* is not known since it corresponds to the “best-case” noise sequence estimate that would lead to the least amount of uncertainty required to have the model interpolate the data within the given noise level. Whereas, in our experiments we use an arbitrary noise sequence consistent with the *a priori* information. The estimated γ_N values (where a second order relaxation, $N = 2$, is used in the experiments) are lower bounds for γ^* and the size of the true uncertainty is an upper bound, that is $\gamma_N \leq \gamma^* \leq \max(|\delta_1|, |\delta_2|)$. As can be seen from the results, the gap between the lower and upper bounds is relatively tight.

δ_1	δ_2	$\max(\delta_1 , \delta_2)$	γ_2
1	1	1	0.9304
-1.8	2	2	1.9422
5	-5	5	4.9584

TABLE VI

ESTIMATING THE SIZE OF UNCERTAINTY USING MODEL INVALIDATION. δ_1 AND δ_2 ARE THE TRUE VALUES OF THE UNCERTAINTY. γ_2 IS THE ESTIMATED LOWER BOUND ON THE MINIMUM UNCERTAINTY SIZE, γ^* .

C. A Practical Example: Activity Monitoring

We now present an application of the approach developed in this paper to a non-trivial problem arising in computer vision: activity monitoring. The objective here is to decide if a person's behavior being captured by a camera is "normal" or "not normal." In other words, one wants to determine if the behavior observed is a combination of known normal "elementary behaviors" or not. To apply the results in this paper, each of the elementary behaviors is modeled as the response of an ARX system. Then, the problem of classifying a behavior as normal/not normal is equivalent to model validation/invalidation of a hybrid system whose submodels are those corresponding to elementary behaviors.

In this example, two elementary behaviors were considered: i) walk and ii) wait. The model for "walk" was identified from the behavior of the center of mass of the person in the video whose sample frames are shown in Fig. 3. Linear programming was used to obtain the autoregressive model with the least ℓ_∞ norm of the process noise resulting in the following model

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

where (x_t, y_t) is the normalized coordinate of the center of the person in the t^{th} frame. The model used for "wait" was simply

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

The image coordinate system was normalized so that the measurements satisfy $(x, y) \in [0, 1] \times [0, 1]$ with a measurement noise bounded by $\|\boldsymbol{\eta}_t\| \leq 0.04$. This noise level together with the submodels (A_1) and (A_2) for "normal" activities constitute the *a priori* information.

To test the proposed approach, four video sequences were used. In all of them the position of the center of mass of the person in the video was determined using the silhouettes obtained by background subtraction. The resulting trajectories of the center of mass were then used for model (in)validation. In the first sequence, the person walks, waits and walks again; i.e., the overall activity is normal. In the second sequence, the person runs and our method found a certificate for invalidity by determining that $\tilde{p}_s > 0$. In the third sequence, the person walks and then starts jumping. In the last sequence, the person passes in front of the camera by jumping.



Fig. 3. Training sequence used in identification of the submodel (A_1) for walking.

In these last two sequences, the proposed algorithm determined that the model was invalid and, hence, the activity was flagged as "not normal."

Sample frames from the sequences are shown in Fig. 4 and the results are summarized in Table VII. The primal optimal objective values, \tilde{p}_s , are also reported in Table VII. Since, in this example, image coordinates are scaled to $[0, 1] \times [0, 1]$, the values \tilde{p}_s obtained have a magnitude much smaller than the ones in previous examples. However, note that the positive range of \tilde{p}_s is just a matter of scaling. It is the sign of \tilde{p}_s that is important. The positivity of \tilde{p}_s (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	\tilde{p}_s	Result
walk, wait	walk, wait	-2.3303×10^{-8}	not invalidated
walk, wait	run	2.3707×10^{-5}	invalidated
walk, wait	walk, jump	5.0293×10^{-7}	invalidated
walk, wait	jump	1.5998×10^{-6}	invalidated

TABLE VII

INVALIDATION RESULTS FOR ACTIVITY MONITORING.

VII. CONCLUSIONS

This paper addressed the problem of model (in)validation of uncertain switched affine systems in a set membership framework. Given *a priori* information consisting of a nominal model, worst-case, deterministic bounds on the measurement noise and model uncertainty, we developed necessary and sufficient certificates for consistency of this information with observed experimental data. As shown in the paper, these certificates can be obtained by solving a convex optimization problem whose size can be bounded from the *a-priori* data and does not require estimating

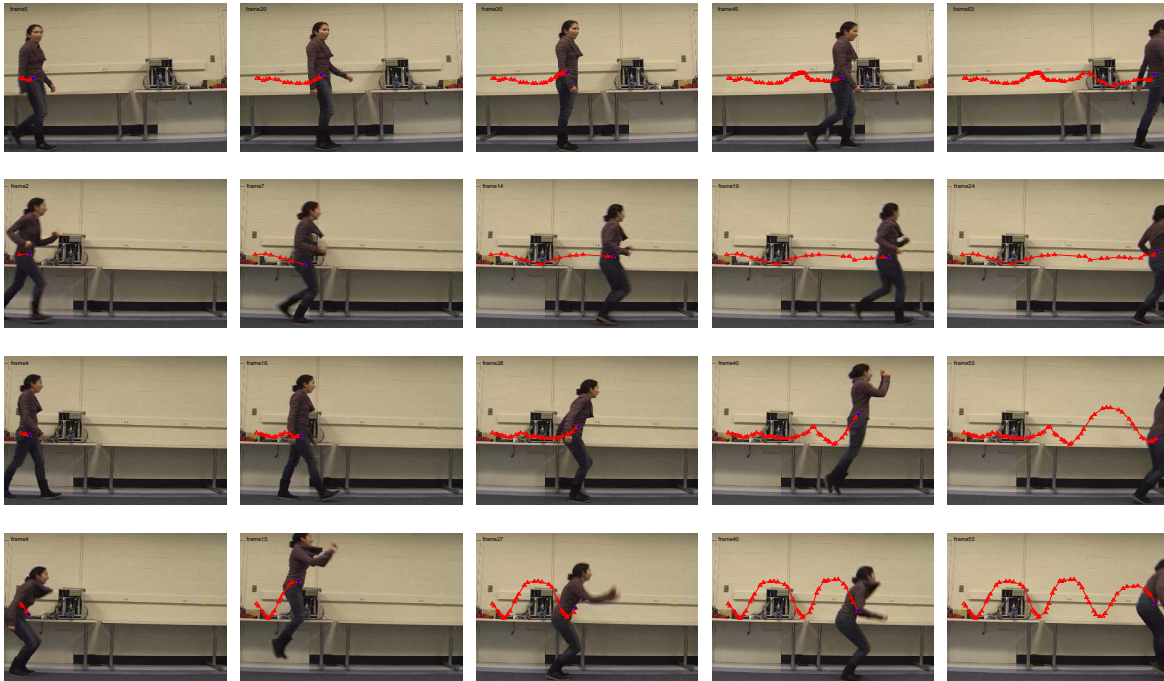


Fig. 4. First row: Walk, wait, walk sequence (not invalidated). Second row: Running sequence (invalidated). Third row: Walk, jump sequence (invalidated). Fourth row: Jumping sequence (invalidated).

the (unknown) switching sequence. We also provided a computationally attractive alternative, based on solving a sequence of convex optimization problems of increasing size until either a positive solution is found or the so-called flat extension property holds. 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 practical problem arising in video-analytics: detecting contextually abnormal activity. An interesting fact borne out by our extensive experiments is that in cases where the experimental data invalidates the *a priori* assumptions, the relaxation obtained using moments of order s (the number of subsystems) already provides an invalidation certificate, suggesting the existence of tighter bounds than those provided by Theorem 5. Research is currently under way seeking to prove this fact.

REFERENCES

- [1] L. Bako. Identification of switched linear systems via sparse optimization. *Automatica*, 47(4):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, Oct. 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] J. Chen. Frequency-domain tests for validation of linear fractional uncertain models. *IEEE Transactions on Automatic Control*, 42(6):748–760, June 1997.
- [5] Y. Cheng, Y. Wang, M. Sznaier, N. Ozay, and C. Lagoa. A convex optimization approach to model (in) validation of switched arx systems with unknown switches. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 6284–6290. IEEE, 2012.
- [6] R.E. Curto and L.A. Fialkow. Truncated k-moment problems in several variables. *Journal of Operator Theory*, 54(1):189–226, 2005.
- [7] C. Feng, C. M. Lagoa, N. Ozay, and M. Sznaier. Hybrid system identification: An sdp approach. In *Proc. 49th IEEE Conf. Dec. Control*, 2010.
- [8] C. Feng, C. M. Lagoa, and M. Sznaier. Hybrid system identification via sparse polynomial optimization. In *American Control Conference*, 2010.
- [9] C. Feng, N. Ozay, C.M. Lagoa, and M. Sznaier. Identification and model (in) validation of switched arx systems: A moment-based approach. In *Linear Parameter-Varying System Identification: New Developments and Trends*, pages 347–379. World Scientific, 2011.
- [10] 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.
- [11] A. Garulli, S. Paoletti, and A. Vicino. A survey on switched and piecewise affine system identification. In *Proceedings of the IFAC Symposium on System Identification*, July 2012.
- [12] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optimization*, 11(3):796–817, 2001.
- [13] J.B. Lasserre. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM J. Optimization*, 17(3):822–843, 2006.
- [14] J.B. Lasserre, M. Laurent, and P. Rostalski. Semidefinite characterization and computation of zero-dimensional real radical ideals. *Foundations of Computational Mathematics*, 8:607–647, 2008.
- [15] F. Lauer, G. Bloch, and R. Vidal. A continuous optimization framework for hybrid system identification. *Automatica*, 47(3):608 – 613, 2011.
- [16] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In M. Putinar and S. Sullivant, editors, *Emerging Applications of Algebraic Geometry*, pages 157–270. Springer, 2009.
- [17] 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.
- [18] N. Ozay. *Convex relaxations for robust identification of hybrid models*. PhD thesis, Northeastern University, 2010.
- [19] N. Ozay, C. M. 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, December 2009.
- [20] N. Ozay, M. Sznaier, and C. M. Lagoa. Model (in) validation of switched ARX systems with unknown switches and its application to activity monitoring. In *Proc. 48th IEEE Conf. Dec. Control*, pages 7624 –7630, December 2010.

- [21] N. Ozay, M. Sznaier, C. M. 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, December 2008.
- [22] N. Ozay, M. Sznaier, C. M. Lagoa, and O.I. Camps. A sparsification approach to set membership identification of switched affine systems. *IEEE Transactions on Automatic Control*, 57(3):634–648, march 2012.
- [23] 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.
- [24] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming Ser. B*, 96(2):293–320, 2003.
- [25] 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.
- [26] S. Prajna. Barrier certificates for nonlinear model validation. *Automatica*, 42(1):117–126, 2006.
- [27] R. Sánchez Peña and M. Sznaier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.
- [28] K. Schmudgen. The k-moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289(1):203–206, 1991.
- [29] J. F. Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones, 1999.
- [30] M. Sznaier and M.C. Mazzaro. An lmi approach to control-oriented identification and model (in) validation of lpv systems. *IEEE Transactions on Automatic Control*, 48(9):1619–1624, sept. 2003.
- [31] H. Waki, S. Kim, M. Kojima, M. Muramatsu, and H. Sugimoto. Algorithm 883: Sparsepop—a sparse semidefinite programming relaxation of polynomial optimization problems. *ACM Transactions on Mathematical Software*, 35(2):15, 2008.

APPENDIX

A. Proof of Theorem 5

For notational simplicity denote $q_{t,i}(\boldsymbol{\eta}) \doteq g_{t,i}(\boldsymbol{\eta}_{t-n_a:t})$ in (10), and $r_k(\boldsymbol{\eta}) \doteq f_{t,j}(\eta_t^{(j)})$ in (13), with index $k \in \mathbf{N}_{n_r}$ where $n_r \doteq (T+1)n_y$. Then, problem (14) can be rewritten as

$$\begin{aligned} o^* &= \min_{\boldsymbol{\eta}} q(\boldsymbol{\eta}) = \sum_{t=n_a}^T \prod_{i=1}^s q_{t,i}(\boldsymbol{\eta}) \\ \text{s.t. } & r_k(\boldsymbol{\eta}) \geq 0 \quad \forall k \in \mathbf{N}_{n_r}, \end{aligned} \quad (25)$$

where the polynomials $q_{t,i}$ are convex, quadratic and a sum of squares, and r_k are concave and also quadratic. In order to prove the theorem, we need to introduce the following preliminary result showing that if the he moment based relaxation of Problem (25) of order $N = s^{T-n_a+1} + 1$ has optimum zero then a related optimization problem also has optimal value 0. To this effect, define the multi-index $\mathbf{i} \doteq [i_{n_a}, i_{n_a+1}, \dots, i_T] \in \mathbf{N}_s^{T-n_a+1}$ and the associated quadratic SoS polynomials $q_{\mathbf{i}}(\boldsymbol{\eta}) \doteq q_{n_a, i_{n_a}}(\boldsymbol{\eta}) + q_{n_a+1, i_{n_a+1}}(\boldsymbol{\eta}) + \dots + q_{T, i_T}(\boldsymbol{\eta})$. Finally, let $q_{big}(\boldsymbol{\eta}) \doteq \prod_{\mathbf{i}} q_{\mathbf{i}}(\boldsymbol{\eta})$.

Lemma 3. *Consider the following optimization problem:*

$$\begin{aligned} o_{big}^* &= \min_{\boldsymbol{\eta}} q_{big}(\boldsymbol{\eta}) \\ \text{s.t. } & r_k(\boldsymbol{\eta}) \geq 0 \quad \forall k \in \mathbf{N}_{n_r}, \end{aligned} \quad (26)$$

Let d_N^* and $d_{big,N}^*$ denote the optimum of the moment based relaxations of order $N = s^{T-n_a+1} + 1$ of Problems (25) and (26), respectively. Then $d_N^* = 0 \Rightarrow d_{big,N}^* = 0$.

Proof. By construction $q_{big}(\boldsymbol{\eta}) = 0$ if and only if $q(\boldsymbol{\eta}) = 0$, and q_{big} can be represented as

$$q_{big}(\boldsymbol{\eta}) = \sum_{t=n_a}^T q_{aux,t}(\boldsymbol{\eta}) \prod_{i=1}^s q_{t,i}(\boldsymbol{\eta})$$

for some SoS polynomials $q_{aux,t}(\boldsymbol{\eta})$. To see this recall that q_{big} is given by

$$q_{big}(\boldsymbol{\eta}) \doteq \prod_{[i_{n_a}, \dots, i_T] \in \{1,2,\dots,s\}^{T-n_a+1}} (q_{n_a, i_{n_a}}(\boldsymbol{\eta}) + q_{n_a+1, i_{n_a+1}}(\boldsymbol{\eta}) + \dots + q_{T, i_T}(\boldsymbol{\eta})).$$

and, hence, one can express it as

$$q_{big}(\boldsymbol{\eta}) = \prod_{[i_{n_a}, \dots, i_{T-1}] \in \{1,2,\dots,s\}^{T-n_a}} \prod_{i_T \in \{1,2,\dots,s\}} (q_{n_a, i_{n_a}}(\boldsymbol{\eta}) + q_{n_a+1, i_{n_a+1}}(\boldsymbol{\eta}) + \dots + q_{T, i_T}(\boldsymbol{\eta})).$$

Next consider the inner product. In this case, one has

$$\begin{aligned} & \prod_{i_T \in \{1,2,\dots,s\}} (q_{n_a, i_{n_a}}(\boldsymbol{\eta}) + q_{n_a+1, i_{n_a+1}}(\boldsymbol{\eta}) + \dots + q_{T, i_T}(\boldsymbol{\eta})) \\ &= h_{i_{n_a}, \dots, i_{T-1}}(\boldsymbol{\eta}) (q_{n_a, i_{n_a}}(\boldsymbol{\eta}) + q_{n_a+1, i_{n_a+1}}(\boldsymbol{\eta}) + \dots + q_{T-1, i_{T-1}}(\boldsymbol{\eta})) + \prod_{i_T=1}^s q_{T, i_T}(\boldsymbol{\eta}). \end{aligned}$$

Replacing this leads to

$$\begin{aligned} q_{big}(\boldsymbol{\eta}) &= \prod_{i_T \in \{1,2,\dots,s\}} \left[h_{i_{n_a}, \dots, i_{T-1}}(\boldsymbol{\eta}) (q_{n_a, i_{n_a}}(\boldsymbol{\eta}) + q_{n_a+1, i_{n_a+1}}(\boldsymbol{\eta}) + \dots + q_{T-1, i_{T-1}}(\boldsymbol{\eta})) + \prod_{i_T=1}^s q_{T, i_T}(\boldsymbol{\eta}) \right] \\ &= \tilde{h}_T(\boldsymbol{\eta}) \left[\prod_{i_T \in \{1,2,\dots,s\}} (q_{n_a, i_{n_a}}(\boldsymbol{\eta}) + q_{n_a+1, i_{n_a+1}}(\boldsymbol{\eta}) + \dots + q_{T-1, i_{T-1}}(\boldsymbol{\eta})) \right] + q_{aux,T}(\boldsymbol{\eta}) \prod_{i_T=1}^s q_{T, i_T}(\boldsymbol{\eta}) \end{aligned}$$

where

$$\tilde{h}_T(\boldsymbol{\eta}) = \prod_{i_T \in \{1,2,\dots,s\}} h_{i_{n_a}, \dots, i_{T-1}}(\boldsymbol{\eta}).$$

Note that all polynomials in the expressions above are obtained by multiplying and summing SoS polynomials. Hence, all of them, and in particular, $q_{aux,T}$, are SoS polynomials. Now, repeat the reasoning for the part in between brackets to obtain the term corresponding to $t = T - 1$. Repeating the reasoning until one obtains the term corresponding to $t = n_a$ results in the required result.

Since $d_N^* = 0$, there exists a moment matrix M_N such that

$$\mathcal{E}_{M_N}(q(\boldsymbol{\eta})) = \sum_{t=n_a}^T \mathcal{E}_{M_N} \left(\prod_{i=1}^s q_{t,i}(\boldsymbol{\eta}) \right) = 0$$

where, by a slight abuse of notation we used \mathcal{E} to denote computing the expected value of $q(\boldsymbol{\eta})$ using the moment sequence in M_N . Since the polynomials $q_{t,i}(\boldsymbol{\eta})$ are SoS, Lemma 3.7 in [14] implies that

$$\mathcal{E}_{M_N} \left(\prod_{i=1}^s q_{t,i}(\boldsymbol{\eta}) \right) = 0 \text{ for all } t = n_a, n_a + 1, \dots, T.$$

Next, recall (Lemma 5.7 in [16]) that the null space of a truncated moments matrix of order N , albeit not a polynomial ideal, has ‘‘truncated ideal like’’ properties when dealing with polynomials whose product has degree up to $N - 1$. Hence the equality above implies

$$\mathcal{E}_{M_N} \left(q_{aux,t}(\boldsymbol{\eta}) \prod_{i=1}^s q_{t,i}(\boldsymbol{\eta}) \right) = 0 \text{ for all } t = n_a, n_a + 1, \dots, T$$

which, in turn, implies that $\mathcal{E}_{M_N}(q_{big}(\boldsymbol{\eta})) = 0$ from where it follows that $d_{big,N}^* = 0$. \square

Proof of Theorem 5: The fact that $o^* = 0 \Rightarrow d_N^* = 0$ is a consequence of the fact that the objective function is a sum of squares. To prove the converse, assume by contradiction that $d_N^* = 0$ but $o^* > 0$. Hence, from Lemma 3 we have that $d_{big,N}^* = 0$ and $o_{big}^* > 0$, which in turn implies that, for all \mathbf{i} , $\gamma_{\mathbf{i}} > 0$, where

$$\gamma_{\mathbf{i}} \doteq \min_{\boldsymbol{\eta}} q_{\mathbf{i}}(\boldsymbol{\eta}) \text{ s.t. } r_k(\boldsymbol{\eta}) \geq 0 \quad \forall k \in \mathbf{N}_{n_r}, \quad (27)$$

Since this is a convex quadratic problem, from Lemma 3.10 in [16] we have that

$$q_{\mathbf{i}}(\boldsymbol{\eta}) = \gamma_{\mathbf{i}} + s_{0,\mathbf{i}}(\boldsymbol{\eta}) + \sum_{k=1}^{n_r} \lambda_{k,\mathbf{i}} r_k(\boldsymbol{\eta})$$

where $s_{0,\mathbf{i}}$ is a quadratic SoS polynomial and where $\lambda_{k,\mathbf{i}} \geq 0$ are scalars. Recursively substituting this expression in the definition of q_{big} leads to an expansion of the form:

$$\begin{aligned} q_{big}(\boldsymbol{\eta}) &= \prod_{\mathbf{i} \in \mathbf{N}_s^{T-n_a+1}} q_{\mathbf{i}}(\boldsymbol{\eta}) = [\gamma_{\mathbf{i}_1} + s_{0,\mathbf{i}_1}(\boldsymbol{\eta}) + \sum_{k=1}^{n_r} \lambda_{k,\mathbf{i}_1} r_k(\boldsymbol{\eta})] \prod_{\mathbf{i} \in \mathbf{N}_s^{T-n_a+1} \setminus \{\mathbf{i}_1\}} q_{\mathbf{i}}(\boldsymbol{\eta}) \\ &= \gamma_{\mathbf{i}_1} \prod_{\mathbf{i} \in \mathbf{N}_s^{T-n_a+1} \setminus \{\mathbf{i}_1\}} q_{\mathbf{i}}(\boldsymbol{\eta}) + u_{0,1}(\boldsymbol{\eta}) + \sum_{k=1}^{n_r} u_{k,1} r_k(\boldsymbol{\eta}) \\ &= \gamma_{\mathbf{i}_1} \gamma_{\mathbf{i}_2} \prod_{\mathbf{i} \in \mathbf{N}_s^{T-n_a+1} \setminus \{\mathbf{i}_1, \mathbf{i}_2\}} q_{\mathbf{i}}(\boldsymbol{\eta}) + u_{0,2}(\boldsymbol{\eta}) + \sum_{k=1}^{n_r} u_{k,2} r_k(\boldsymbol{\eta}) \\ &\quad \vdots \\ &= \gamma + s_0(\boldsymbol{\eta}) + \sum_{k=1}^{n_r} s_k(\boldsymbol{\eta}) r_k(\boldsymbol{\eta}) \end{aligned} \quad (28)$$

where $\gamma \doteq \prod_{\mathbf{i} \in \mathbf{N}_s^{T-n_a+1}} \gamma_{\mathbf{i}} > 0$, and where s_0 and s_k are sum of squares polynomials of degree up to $2N$ and $2(N-1)$ respectively. This expansion together with Theorem 4.2 in [12] imply that $d_{big,N}^* \geq \gamma > 0$, contradicting the hypothesis.