ℓ_∞ Worst-Case Optimal Estimators for Switched ARX Systems.

Y. Cheng Y. Wang M. Sznaier

Abstract— **This paper considers the problem of worst-case estimation for switched piecewise linear models, in cases where the mode-variable is not directly observable. Our main result shows that worst case point wise optimal estimators can be designed by solving a constrained polynomial optimization problem. In turn, this problem can be relaxed to a sequence of convex optimizations by exploiting recent results on moments-based semi-algebraic optimization. Theoretical results are provided showing that this approach is guaranteed to find the optimal filter in a finite number of steps, bounded above by a constant that depends only on the number of data points available and the parameters of the model. Finally, we briefly show how to extend these results to accommodate parametric uncertainty.**

I. Introduction and motivation

Switched linear systems are ubiquitous in many applications ranging from manufacturing processes, communication systems and biology to reconfigurable control. Thus, a large research effort has been devoted in the past decade to the problems of identification, control and estimation of systems represented by switched autoregressive models.

In the context of estimation, existing results can be roughly classified into three categories. The first class of algorithms involves situations where the mode variable is directly accessible. In this case, the problem can be solved by considering for instance a gain switched Luenberger observer [2], obtained by solving a set of linear matrix inequalities. In the case of ℓ_∞ bounded noise, worst case optimal filters can be synthesized by solving on-line a linear programming problem with $\mathcal{O}(r)$ variables, where r is the memory of the filter [11].

In the second category of algorithms, the assumption on the knowledge of the discrete state is relaxed. Rather than assuming exact knowledge of the discrete state, the problem is formulated in a probabilistic context, where either a probability vector for different modes is provided or it is assumed that mode transitions occur according to a model described as a first-order Markov chain [9][5]. However, in many practical situations these probabilistic descriptions are not directly available.

The third family of algorithms considers a very general setup where the discrete state is not directly accessible and no priors exist on the probability of transitions. Rather, in these cases an estimator is constructed to estimate the states from the knowledge of the inputs and outputs [10][3][4][1]. In [10] and [3], an multi-model observer is designed, which consists of a location observer to decide the active submodel and a continuous observer to estimate the continuous states. These type of methods require observability of the switching system which is nontrivial to check. Alternatively, several methods completely avoid estimating the mode variable. For instance [4], proposes an asymptotic observer constructed directly from the measured data using an algebraic approach. However, this approach is only available for systems sharing common dynamics, where the switching is restricted to the measurement equations. In [1], both the unknown discrete mode and the continuous state vector are jointly estimated by minimizing a receding-horizon quadratic cost function that penalizes weighted ℓ_2 norms of the estimation errors in the state, measurement and process noise over the set $\mathcal{G}$ of switching patterns compatible with the experimental data observed so far. A potential difficulty here stems from the computational complexity entailed in minimizing the objective function over all patterns in $\mathcal{G}$, and in propagating this set.

In this paper, we consider the problem of finding worst case optimal estimators for switched ARX models in cases where the mode is not directly accessible and the measurements are corrupted by unknown but bounded noise. This scenario captures the features present in many applications where no a priori information is available about the noise, except its support set. As in [1], we propose a receding horizon strategy. However, the proposed algorithm does not require explicitly propagating the admissible switching set or minimizing over all its elements. Rather, this is accomplished implicitly by recasting the problem into a constrained polynomial optimization form where admissible switching sequences are characterized as those satisfying a set of polynomial constraints. As we show in the paper by appealing to Information Based Complexity ideas, worst case point wise optimal estimators can be obtained by solving a semi-algebraic optimization problem. In turn, the use of moments-based polynomial optimization tools allows to relax this problem to a sequence of convex optimization problems, each of which furnishes a suboptimal estimator with guaranteed worst case bounds. In addition, we provide theoretical results showing that an optimal solution can be obtained by considering a relaxation whose size can be determined a-priori and depends only on the memory of the estimator and the ARX model. Finally, in the last part of the paper, we briefly indicate how to extend these results to accommodate model uncertainty.

This work was supported in part by NSF grant ECCS–0901433; AFOSR grant FA9559–12–1–0271; and DHS grant 2008-ST-061-ED0001. Y. Cheng, Y. Wang and M. Sznaier are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115.

II. Preliminaries

For ease of reference, in this section we summarize the notation used in the paper and recall some results on polynomial optimization and information based complexity that play a key role in establishing the main result of this paper.

A. Notation

$\mathbf{y}\,(\mathbf{M})$	a vector in $\mathbf{R}^n$ (a matrix in $\mathbf{R}^{n\times m}$)
$\|\|\mathbf{y}\|\|_\infty$	∞ norm of the vector $\mathbf{y}$: $\|\|\mathbf{y}\|\|_\infty \doteq \max_i \|y_i\|$
ℓ_∞	Banach space of vector sequences equipped with the norm $\|\mathbf{y}\|_{\ell_\infty} \doteq \sup_i \|\mathbf{y}_i\|_\infty$
$\mathbf{N}_{n_s}$	set of positive integers up to n_s
$\mathbf{M}\succeq\mathbf{N}$	$\mathbf{M}-\mathbf{N}$ is positive semidefinite

B. Background on Information Based Complexity

In this section, we recall some results from Information Based Complexity (IBC) required to establish (worst-case) optimality of the proposed filter. Here we just consider the case of bounded operators in ℓ_∞. A general treatment can be found for instance in the book [13].

Let K be a set in ℓ_∞ and consider two linear operators S_y, S_z: $\ell_\infty \to \ell_\infty$. In this context, the estimation problem can be stated as, given an element $f_0 \in K$, find an estimate $\hat{z}$ of $z \doteq S_z f_0$ using noisy experimental information $y = S_y f_0 + \eta$, where the noise η is only known to belong to some bounded set $\mathcal{N} \subset \ell_\infty$. Given an estimation algorithm $\hat{z} = \mathcal{A}(y)$ (not necessarily linear), it is of interest to compute its worst case approximation error. For a given measurement y, define the consistency set:

$$\mathcal{T}(y) \doteq \{f \in K : \; y = S_y f + \eta \text{ for some } \eta \in \mathcal{N}\} \tag{1}$$

that is, the set of all possible elements in K that could have generated the observed data y. Since all the elements f that could have generated y belong to $\mathcal{T}(y)$, it follows that the local error $\epsilon(y,\mathcal{A})$ is given by

$$\epsilon(\mathcal{A},y) \doteq \sup_{f\in\mathcal{T}(y)} ||S_z f - \mathcal{A}(y)||_\infty \tag{2}$$

Definition 1: An algorithm $\mathcal{A}_o(\cdot)$ is said to be locally (or point wise) optimal if $\epsilon(\mathcal{A}_o,y) = \inf_{\mathcal{A}}\epsilon(\mathcal{A},y)$, that is, if, for every measurement y, it produces the best (in the worst-case sense) estimate of all possible algorithms.

The minimum local error $\epsilon(\mathcal{A}_0,y)$ is sometimes referred to as the local radius of information, $r(y)$. It provides a lower bound on achievable performance, since no estimation algorithm can have smaller worst case error when operating on the same measured data y.

C. Moments based polynomial optimization

Next, we recall results from polynomial optimization the classical theory of moments which will play a key role in developing tractable estimation algorithms. Let $K \subset \mathcal{R}^n$ be a compact semi-algebraic set defined by a collection of polynomial inequalities of the form $g_k(\mathbf{x}) \geq 0$, $k = 1,\cdots,d$, that is,

$$K = \{\mathbf{x} \,|\, \mathbf{x} \in \mathcal{R}^n \; g_k(\mathbf{x}) \geq 0, \; k = 1,\cdots,d\} \tag{3}$$

and consider the problem of minimizing a multivariate polynomial $p = \sum_\alpha p_{\boldsymbol\alpha} x^{\boldsymbol\alpha}$ over the set K. As shown in [6], this problem is equivalent to $\min_\mu \mathcal{E}(p)$ where $\mathcal{E}$ denotes expectation and μ denotes the set of all Borel measures supported in K, or equivalently

$$\begin{array}{ll} p^* & = \min_{\mathbf{m}} \sum_\alpha p_{\boldsymbol\alpha} m_{\boldsymbol\alpha} \\ & \text{subject to } \exists\mu, \text{ supported in } K \text{ such that} \\ & m_{\boldsymbol\alpha} = \mathcal{E}_\mu(x^{\boldsymbol\alpha}) \end{array} \tag{4}$$

Existence of such a representing measure μ is equivalent to positive semidefiniteness of the (infinite) moment $\mathbf{M}(\mathbf{m})$ and localizing $\mathbf{L}(g_k\mathbf{m})$ matrices [6], Thus, an equivalent convex, albeit infinite dimensional reformulation of (4) is given by:

$$\begin{array}{rll} p^* = & \min_{\mathbf{m}} & \sum_\alpha p_{\boldsymbol\alpha} m_{\boldsymbol\alpha} \\ & \text{s.t.} & \mathbf{M}(\mathbf{m}) \succeq 0, \\ & & \mathbf{L}(g_k\mathbf{m}) \succeq 0, k = 1,\ldots,d, \end{array} \tag{5}$$

A truncated version of this problem involving moments of order up to $2N$ is given by:

$$\begin{array}{rll} p_N^* = & \min_{\mathbf{m}} & \sum_\alpha p_{\boldsymbol\alpha} m_{\boldsymbol\alpha} \\ & \text{s.t.} & \mathbf{M}_N(\mathbf{m}) \succeq 0, \\ & & \mathbf{L}_N(g_k\mathbf{m}) \succeq 0, k = 1,\ldots,d, \end{array} \tag{6}$$

where

$$\begin{aligned} \mathbf{M}_N(\mathbf{m})(i,j) &= m_{\alpha^{(i)}+\alpha^{(j)}}, \; \forall i,\, j = 1,\cdots,S_N \\ \mathbf{L}_N(g_k\mathbf{m})(i,j) &= \sum_\beta g_{k,\beta^l} m_{\beta^{(l)}+\alpha^{(i)}+\alpha^{(j)}}, \\ & \quad \forall i,\, j = 1,\cdots,S_{N-\lfloor\frac{\delta_k}{2}\rfloor} \end{aligned} \tag{7}$$

where $S_N = \begin{pmatrix} N+n \\ n \end{pmatrix}$(e.g. the number of moments in $\mathcal{R}^n$ up to order N). The main result of [6] shows that $p_N^* \uparrow p^*$, monotonically, thus providing a hierarchy of convergent relaxations.

1) Exploiting the sparse structure: In many cases of practical interest, both the polynomial objective and the constraints that define the set K exhibit a sparse structure that can be used to reduce the computational complexity entailed in solving the (truncated) problems (6). The following property plays a key role in exploiting this structure:

Definition 2: [7] Assume that the polynomial p can be partitioned into $p = p_1 + \cdots + p_d$ such that each p_k and that the constraints g_k that define the set K contain only variables indexed by elements of some subset $I_k \subset \{1,\cdots,n\}$. If there exists a reordering $I_{k'}$ of I_k such that for every $k' = 1,\cdots,d-1$:

$$I_{k'+1} \cap \cup_{j=1}^{k'} I_j \subset I_s \quad \text{for some } s \leq k' \tag{8}$$

then the *running intersection property* is satisfied.

It can be shown that for problems that satisfy the *running intersection property*, it is possible to construct a hierarchy of semidefinite programs of smaller size. Specifically, partition the objective function $\{p_k\}_{k=1}^d$ according to the sets $\{I_k\}$ and consider the problem:

$$\begin{array}{rll} p_N^* = & \min_{\mathbf{m}} & \sum_{k=1}^d \sum_{\alpha(j)} p_{k,\boldsymbol\alpha(j)} m_{\boldsymbol\alpha(j)} \\ & \text{s.t.} & \mathbf{M}_N(\mathbf{m}_{I_k}) \succeq 0, k = 1,\ldots,d, \\ & & \mathbf{L}_N(g_k\mathbf{m}_{I_k}) \succeq 0, k = 1,\ldots,d, \end{array} \tag{9}$$

where $p_{k,\boldsymbol{\alpha}(j)}$ is the coefficient of the $\boldsymbol{\alpha}(j)^{th}$ monomial in the polynomial p_k, $\mathbf{M}_N(\mathbf{m}_{I_k})$ denote the moment matrix and $\mathbf{L}_N(g_k\mathbf{m}_{I_k})$ the localizing matrix for the subset of variables in I_k. Then, as shown in [7] $p_N^* \uparrow p^*$. It is worth emphasizing that for the case of generic polynomials and constraints, an N^{th} order relaxation requires considering moments and localizing matrices containing $O(n^{2N})$ variables. On the other hand, if the running intersection property holds, it is possible to define d sets of smaller sized matrices each containing variables only 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.

D. Problem Statement

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

$$\begin{array}{rcl} \mathbf{y}_t &=& \sum_{k=1}^{n_a} \mathbf{A}_k(\sigma_t)\mathbf{y}_{t-k} + \sum_{k=1}^{n_c} \mathbf{C}_k(\sigma_t)\mathbf{u}_{t-k} \\ \hat{\mathbf{y}}_s &=& \mathbf{y}_s + \boldsymbol{\eta}_s,\ s = t, t-1, ..t-n_a \end{array} \tag{10}$$

where $\mathbf{u}_t \in \mathbf{R}^{n_u}$, $\mathbf{y}_t, \hat{\mathbf{y}}_t \in \mathbf{R}^{n_y}$ and $\sigma_t \in \mathbf{N}_{n_s}$ denote the input, output, its noisy measurements, and the discrete mode signal, respectively. No assumptions are made in terms of dwell time, thus the system can switch arbitrarily fast among the n_s submodels G_i, each associated with a set of its coefficient matrices $\{\mathbf{A}_1(i), \ldots, \mathbf{A}_{n_a}(i), \mathbf{C}_1(i), \ldots, \mathbf{C}_{n_c}(i)\}$. The goal is to estimate a (scalar) linear combination of values of $\mathbf{y}$ using the most recent r noisy measurements $\hat{\mathbf{y}}$, where r (the memory of the estimator), is a design parameter. Formally, this problem can be stated as follows:

Problem 1: Given a nominal switched ARX system of the form (10), an a-priori bound ϵ on the ℓ_∞ norm of the noise, e.g. $\|\boldsymbol{\eta}\|_\infty \leq \epsilon$, and a linear functional

$$\begin{array}{rcl} z_t &\doteq& \sum_{i=0}^{n_z} \mathbf{h}_i^T \mathbf{y}_{t-i} \doteq \text{Trace}\left(\mathbf{H}^T\mathbf{Y}_t\right); \\ \mathbf{Y}_t &\doteq& \begin{bmatrix}\mathbf{y}_t & \ldots & \mathbf{y}_{t-n_z}\end{bmatrix}, \\ \mathbf{H}_t &\doteq& \begin{bmatrix}\mathbf{h}_0 & \ldots & \mathbf{h}_{n_z}\end{bmatrix} \end{array} \tag{11}$$

where $\mathbf{h}_i$ are given vectors, find a (worst case) locally optimal estimate $\hat{z}_t$ of z_t using the most recent $r \geq \max\{n_a+1, n_z+1\}$ noisy measurements of $\mathbf{y}$, that is find $\hat{z} = \mathcal{A}(\{\mathbf{u}_t, \hat{\mathbf{y}}_t\}_{t-r+1}^t)$ that minimizes the worst case estimation error $\max_{z_t} ||\hat{z}_t - z_t||_\infty$.

III. Bounded Complexity Locally Optimal Estimators

In this section we present the main result of this paper: a convex optimization based algorithm for finding bounded complexity ℓ_∞ (point wise) optimal estimators. The main idea is to first recast this problem into a linear optimization form over a semi-algebraic set, which in turn can be relaxed to a sequence of convex optimization problems by exploiting the results outlined in section II-C.

A. A Semi-Algebraic Optimization Reformulation

Given a sequence of input/output data $\{\hat{\mathbf{y}}_i, \mathbf{u}_i\}$ $i = t-r+1, \ldots, t$, define the consistency set $\mathcal{T}(\hat{\mathbf{y}}, \mathbf{u})$:

$$\mathcal{T}(\hat{\mathbf{y}}, \mathbf{u}) \doteq \big\{\{\mathbf{y}_i\}_{t-r+1}^t \text{ such that (10) holds for some } \|\boldsymbol{\eta}\|_\infty \leq \epsilon \text{ and } \{\sigma_i\}_{t-r+1+n_a}^t \in \mathbf{N}_{n_s}.\ \big\} \tag{12}$$

that is the set of all possible values of $\mathbf{y}$ that are consistent with the observed data. As we show next, $\mathcal{T}(\hat{\mathbf{y}}, \mathbf{u})$ has a semi-algebraic representation. Define a set of $(r-n_a)n_s$ indicator variables $x_{i,j} \in \{0,1\}$, $i = 1, \ldots, n_s$, $j = t-(r-n_a)+1, \ldots, t$. Clearly, (10) holds if and only if the following set of equations is feasible:

$$\begin{array}{l} x_{i,s}\|\mathbf{y}_s - \sum_{k=1}^{n_a} \mathbf{A}_k(i)\mathbf{y}_{s-k} - \sum_{k=1}^{n_c} \mathbf{C}_k(i)\mathbf{u}_{s-k}\|_2^2 = 0 \\ x_{i,s}^2 = x_{i,s},\ \sum_i x_{i,s} = 1,\ i = 1, \ldots, n_s, \\ \forall\ s \in [t-r+1+n_a, t] \end{array} \tag{13}$$

where we used the fact that $x \in \{0,1\} \iff x = x^2$. Thus, it follows that $\mathcal{T}(\mathbf{y}, \mathbf{u})$ is equivalent to:

$$\mathcal{T}'(\hat{\mathbf{y}}, \mathbf{u}) \doteq \big\{\{\mathbf{y}_i\}_{t-r+1}^t \text{: subject to eq. (13), } \|\eta_s\|_\infty \leq \epsilon \text{ and } \mathbf{y}_s = \hat{\mathbf{y}}_s - \boldsymbol{\eta}_s \text{ for all } s \in [t-r+1+n_a, t].\ \big\} \tag{14}$$

Next, we present the proposed estimator, based on solving two optimization problems. Given $\{\hat{\mathbf{y}}_t, \mathbf{u}_t\}_{t-r+1}^t$, define $\hat{z}_t^+$, $\hat{z}_t^-$ as the solutions to the following optimization problems:

$$\hat{z}_t^+ = \max_{\mathbf{y} \in \mathcal{T}'(\hat{\mathbf{y}}, \mathbf{u})} \text{Trace}\left(\mathbf{H}^T\mathbf{Y}_t\right) \tag{15}$$

$$\hat{z}_t^- = \min_{\mathbf{y} \in \mathcal{T}'(\hat{\mathbf{y}}, \mathbf{u})} \text{Trace}\left(\mathbf{H}^T\mathbf{Y}_t\right) \tag{16}$$

and define the (central) estimator

$$\hat{z}_t \doteq \frac{\hat{z}_t^+ + \hat{z}_t^-}{2} \tag{17}$$

Lemma 1: $\hat{z}_t$ is a point wise optimal worst-case estimator of z_t.

Proof: This result is followed immediately from Theorem 2.4 in [12] by noting that $\hat{z}_t$ is the Chebyshev center of the set

$$\mathcal{HT}' \doteq \left\{ \begin{array}{l} z \colon z \doteq \text{Trace}\left(\mathbf{H}^T\mathbf{Y}\right) \\ \text{for some } \{\mathbf{y}\}_{t-r+1}^t \in \mathcal{T}'(\hat{\mathbf{y}}, \mathbf{u})\} \end{array} \right.$$

■

Remark 1: By directly substituting $\hat{\mathbf{y}}_k = \mathbf{y}_k + \boldsymbol{\eta}_k$ into (13) it follows that the optimization problems (15) and (16) can be rewritten as:

$$\begin{array}{ll} \hat{z}_t^+\ (\hat{z}_t^-) = & \max_{\boldsymbol{\eta}_k}(\min_{\boldsymbol{\eta}_k})\text{Trace}\left(\mathbf{H}^T\boldsymbol{\Xi}\right) + f_o(\hat{\mathbf{y}}) \\ \text{subject to:} & x_{i,s}p_{i,s}(\boldsymbol{\eta}) = 0,\ s \in [t-r+1+n_a, t] \\ & \|\boldsymbol{\eta}\|_\infty \leq \epsilon \\ & x_{i,s} = x_{i,s}^2;\ \sum_i x_{i,s} = 1 \end{array} \tag{18}$$

where $\boldsymbol{\Xi} \doteq \begin{bmatrix}\boldsymbol{\eta}_t & \ldots & \boldsymbol{\eta}_{t-n_z}\end{bmatrix}$, $p_{i.s}(\boldsymbol{\eta})$ is a (quadratic) SoS polynomial, and where $f_o(\hat{\mathbf{y}})$ is a function of the measured data only.

B. A Moments-Based Convex Relaxation

Problems (18) are computationally challenging due to the non-convex polynomial constraints. However, as shown next the use of the moment-based techniques allows for obtaining a (monotonic) sequence of bounds from which convergent suboptimal estimators can be obtained. To this effect, rewrite $p_{i,s}$ in (18) explicitly as $p_{i,s} = \mathbf{c}_s^T \boldsymbol{\eta}^\alpha$, where the coefficients $\mathbf{c}_s$ are a function of the parameters of the SARX model and the measured data $\{\mathbf{u}, \hat{\mathbf{y}}\}$, and where $\boldsymbol{\eta}^\alpha$ is a vector containing the monomials of the unknown noise $\boldsymbol{\eta}_t$ in a suitable ordering. From the results outlined in section II-C, it follows that a sequence of convex relaxations of problems (18) is given by:

$$\hat{z}^+_{t,N}(\hat{z}^-_{t,N}) = \max_{\mathbf{m}}(\min_{\mathbf{m}})\text{Trace}\left(\mathbf{H}^T\mathbf{P}_N\mathbf{M}_N\right) + f_o(\hat{\mathbf{y}}) \quad (19)$$

subject to

$$\begin{array}{c}\mathbf{M}_N(\mathbf{m}) \succeq 0 \\ \mathbf{L}_N(\mathbf{m}) \succeq 0\end{array} \quad (20)$$

where $\mathbf{m}$ denotes the moment sequence of $\boldsymbol{\eta}_t, \cdots, \boldsymbol{\eta}_{t-r+1}$ up to order $2N$ and where the matrix $\mathbf{P}_N$ selects the elements of $\mathbf{M}$ corresponding to the first order moments of $\boldsymbol{\eta}^t_{t-n_z}$ and $\mathbf{L}_N$ is the moments localizing matrix corresponding to the constraints (13) and $\|\boldsymbol{\eta}\|_\infty \leq \epsilon$.

Lemma 2: The estimator $\hat{z}_N \doteq \frac{\hat{z}^+_{t,N}+\hat{z}^-_{t,N}}{2}$ satisfies $\max_{z\in\mathcal{HT}'} |z - \hat{z}_N| \leq e_N \doteq \frac{\hat{z}^+_{t,N}-\hat{z}^-_{t,N}}{2}$. Moreover e_N is a non-increasing sequence and $\lim_{N\to\infty} e_N = e$, the optimal worst case error.

Proof: From the results in section II-C it follows that $\hat{z}^+_{t,N} \geq \hat{z}^+_t$ and $\hat{z}^-_{t,N} \leq \hat{z}^-_t$. Hence, given any $z \in \mathcal{HT}'$, we have that:

$$\begin{array}{rcl} z - \hat{z}_N & \leq & \hat{z}^+_{t,N} - \hat{z}_N = \frac{\hat{z}^+_{t,N}-\hat{z}^-_{t,N}}{2} \\ z - \hat{z}_N & \geq & \hat{z}^-_{t,N} - \hat{z}_N = -\frac{\hat{z}^+_{t,N}-\hat{z}^-_{t,N}}{2} \end{array}$$

Thus $|z - \hat{z}_N| \leq e_N$ for all $z \in \mathcal{HT}$. The facts that e_k is non-increasing, convergent follow from the results in section II-C, showing that $\hat{z}^+_{t,N}$, $\hat{z}^-_{t,N}$ are convergent non-increasing/non-decreasing sequences respectively. ■

Theorem 1: The optimal value in problem (19) is achieved for some $N \leq N_o \doteq r - n_a + 1$.

Proof: The proof, omitted due to space constraints, is based upon showing that the objective function in (18) admits an expansion of the form:

$$\text{Trace}\left(\mathbf{H}^T\boldsymbol{\Xi}\right) - z^* \doteq u_o + \sum u_i f_i$$

where z^* denotes the optimal value, f_i denote the constraints, and u_o, u_i are sum of squares polynomials such that degree of $u_o \leq 2N_o$ and degree of $u_i f_i \leq 2N_o$. The desired result follows then from Putinar's Positivstellensatz. ■

C. Computational Complexity Considerations

It can be easily seen that the objective and constraints in Problem (19) can be partitioned into $r - n_a$ subsets I_k, each containing the noise variables for $n_a + 1$ consecutive time instants and the n_s indicator variables associated with the index k. Thus, it follows that it satisfies the running intersection property and hence it can be solved by considering $r - n_a$ smaller moments matrices of size $\binom{N+n_a n_y+n_y+n_s}{n_a n_y+n_y+n_s}$ rather than a single one of size $\binom{N+rn_y+(r-n_a)n_s}{rn_y+(r-n_a)n_s}$. While this results in a substantial reduction on the number of variables that need to be considered (from $\mathcal{O}((rn_y + (r - n_a)n_s)^{2N})$ to $\mathcal{O}((n_a n_y + n_y + n_s)^{2N})$, the computational burden can limit the "memory" of the estimator (the design parameter r) that can be used in practice. To circumvent this difficulty, in the sequel we present a suboptimal filter that allows for using a shorter memory $r_0 < r$ by propagating past bounds on the estimation error in a receding horizon fashion, leading to the following algorithm:

Algorithm 1: Receding Horizon Estimator

Choose the window length r_0, $r_0 < r$. Set the number of iterations to $r - r_0 + 1$;
Set $z_s^+ = +\infty,\ z_s^- = -\infty,\ s = 1, \ldots, r$
For $i = r_0 + 1, \ldots, r$, **repeat**

1- Solve Problem (19) with $r = r_0$ subject to the additional constraints
$\text{Trace}(\mathbf{H}^T\mathbf{Y}_{s-n_z:s}) \leq z_s^+,$
$\text{Trace}(\mathbf{H}^T\mathbf{Y}_{s-n_z:s}) \geq z_s^-,\ s = i - r_0 + 1, \ldots, i$

2- Update z_i^+ and z_i^- with the results from [1-];

end for.

D. Consistency Considerations

Note that the optimization problems (15) and (16) are uncoupled, in the sense that the worst case noise and switching sequences could potentially be different. Hence the estimator (17) is not interpolatory, in the sense that it does not necessarily belong to the consistency set $\mathcal{T}'(\hat{\mathbf{y}}, \mathbf{u})$. This raises the issue of whether the search for optimal estimators should be restricted to those inside the consistency set, for instance by enforcing the constraint that the optimizing sequences in (15) and (16) are the same. As we show next with a simple example, as long as one is interested in minimizing the worst case estimation error, the answer to this question is negative: estimators outside the consistency set can have smaller worst case error than those inside it[1]. To this effect consider a system that switches between two first order models:

$$y_{k+1} = y_k \quad \text{(sys1)}$$

and

$$y_{k+1} = -0.5 y_k \quad \text{(sys2)}$$

with noise bound $|\epsilon| \leq 1$. Assume that two noisy measurements are available:

$$\begin{array}{rcl}\hat{y}_1 & = & y_1 + \epsilon_1 = 2 \\ \hat{y}_2 & = & y_2 + \epsilon_2 = 0\end{array} \quad (21)$$

[1]This is related to the fact that, unless one is working in a Hilbert space, the Chebyshev center of a set can be outside it, even if the set is convex.

The goal is to find a worst-case optimal estimator for y_2. Optimizing $y^+ - y^-$ (the diameter of information) subject to the constraint that both y^+ and y^- have to have been generated by the same switching sequence leads to $y^+_{\text{sys1}} = y^-_{\text{sys1}} = 1$ for the case where the measurements are both generated by the first system and $y^+_{\text{sys2}} = -0.5$, $y^-_{\text{sys2}} = -1$ if both are generated by the second one. The corresponding central estimators are given by $y_{c,\text{sys1}} = 1$ and $y_{c,\text{sys2}} = -0.75$. Note that the worst case estimation error of $y_{c,\text{sys1}}$ is $1 - (-1) = 2$, achieved if the true signal was generated by the second system. Similarly, the worst case estimation error of $y_{c,\text{sys2}}$ is $1-(-0.75) = 1.75$, achieved when the true signal was generated by the first system. On the other hand, optimizing y^+ and y^- separately leads to $y^+ = 1$, $y^- = -1$ and $y_c = 0$ with a worst case error of 1, regardless of whether the true system is the first or the second one. Note that the estimate $y_c = 0$ is outside the consistency set, since this set is $\{1\} \cup [-1.5, -0.5]$.

E. Handling Parametric Uncertainty

In this section we briefly outline how to extend the basic estimator presented in section III-A to account for parametric uncertainty. Assume that some coefficients in (10) are only known to belong to some semi-algebraic set, e.g. $\mathbf{A}_i \in \mathcal{A}_i$, $\mathbf{C}_j \in \mathcal{C}_j$, for some[2] i, j. In this case, treating $\mathbf{A}_i, \mathbf{C}_j$ as unknowns in (13) and adding the constraints $\mathbf{A}_i \in \mathcal{A}_i$, $\mathbf{C}_j \in \mathcal{C}_j$ to the description (14) leads to a semi-algebraic optimization problem that can be solved using the techniques outlined above.

IV. Examples

In this section, we illustrate with several examples.

A. Example with no model uncertainty

Consider first the case of a system that switches arbitrarily fast between the following two subsystems:

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

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

The goal is to estimate $z_t = y_t$ from the noisy measurements

$$\hat{y}_t = y_t + \eta_t,\ \|\eta\|_\infty \le 0.5$$

Figures 1 and 2 shows the performance of the proposed estimator with memory $r_0 = 11$, over a time horizon $T = 30$. As illustrated there, the proposed estimator substantially reduces the uncertainty in the estimation, and performance improves as more measurements are collected.

[2]simple examples are structured additive uncertainty $\mathbf{M} + \mathbf{\Delta}$ and multiplicative uncertainty $(1 + \delta)\mathbf{M}$.

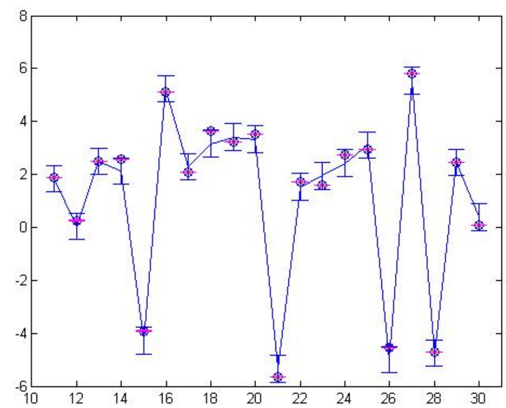

Fig. 1. Estimation results in the absence of model uncertainty (blue line: $[\hat{y}_t - 0.5, \hat{y}_t + 0.5]$, magenta box: suboptimal bounds by Algorithm 1, red star: y_t, and black circle: our estimate)

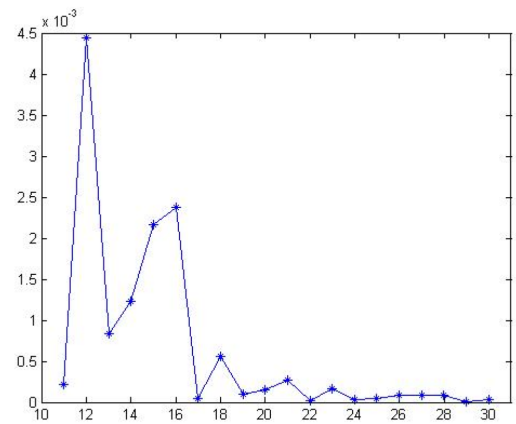

Fig. 2. Estimation error for Example 1

B. Example with Parametric Uncertainty

The next example illustrates the ability of the proposed estimators to handle parametric uncertainty. In this case, we assumed that the coefficients of the nominal model are subject to time-varying multiplicative parametric uncertainty leading to descriptions of the form

$$\begin{array}{rcl} \mathbf{y}_t & = & \sum_{k=1}^{n_a}(1+\delta_{A_k,t})\mathbf{A}_k(\sigma_t)\mathbf{y}_{t-k} \\ & & + \sum_{k=1}^{n_c}(1+\delta_{C_k,t})\mathbf{C}_k(\sigma_t)\mathbf{u}_{t-k} \\ \hat{\mathbf{y}}_s & = & \mathbf{y}_s + \boldsymbol{\eta}_s,\ s = t, t-1, ..t - n_a \end{array} \qquad (22)$$

where the only information available about $\delta_{A_k,t}$ and $\delta_{C_k,t}$ is a bound of the form:

$$||\delta_{A_k,t}||_\infty \le \epsilon_{A_k} \quad \text{and} \quad ||\delta_{C_k,t}||_\infty \le \epsilon_{C_k} \qquad (23)$$

This uncertainty can be handled by the proposed algorithm by adding the new variables $\delta_{A_k,t}$ and $\delta_{C_k,t}$ to the problem, and suitably modifying the moments sequence $\mathbf{m}$ and corresponding matrix $\mathbf{M}(\mathbf{m})$ to include the new terms, and adding the constraints corresponding to (23) to the localizing matrix $\mathbf{L}(\mathbf{m})$.

Figures 3 and 4 show the results of two experiments with 5% and 10% uncertainty, respectively (e.g. $\epsilon_{A_k} = \epsilon_{C_k} = 0.05$ and $\epsilon_{A_k} = \epsilon_{C_k} = 0.10$). Finally, Figure 5 shows the actual estimation error for both cases, as well as the one obtained when the system is perfectly known. As expected, model uncertainty results in larger estimation errors.

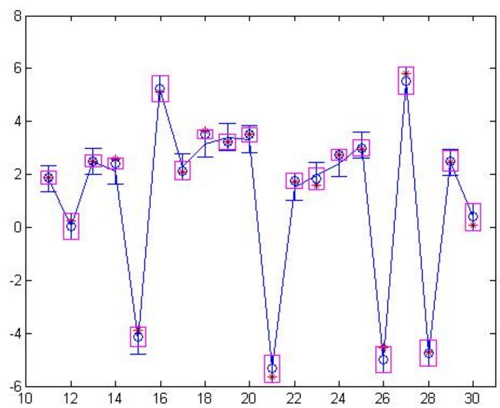

Fig. 3. Estimation with 5% parametric uncertainty.

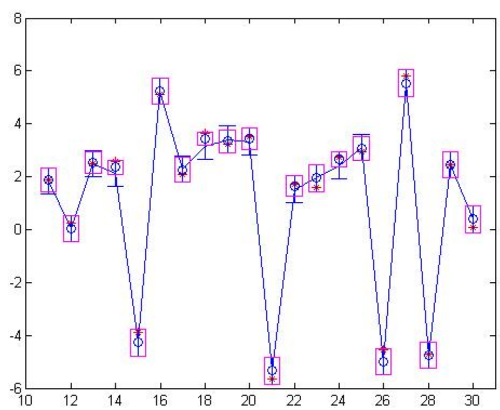

Fig. 4. Estimation with 10% parametric uncertainty.

V. Conclusions

This paper considers the problem of worst case estimation for switching systems in cases where the mode variable is not accessible. By exploiting a combination of results from information based complexity and polynomial optimization, we show that point wise optimal estimators can be obtained by solving a (convergent) sequence of convex relaxations. Each one of these relaxations provides a suboptimal estimator with guaranteed worst case estimation error, and this error converges monotonically to the optimal as the size of the relaxations increases. Moreover, as shown in the paper, the size of the relaxation needed to exactly solve the problem is bounded by a constant that depends only on the memory of the ARX model and the estimator. Finally, we address

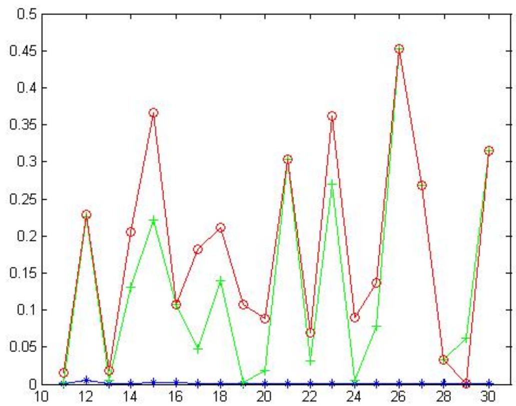

Fig. 5. Absolute value of the actual estimation error for different uncertainty levels (blue star: $\epsilon_{A_k} = \epsilon_{C_k} = 0$, green cross: $\epsilon_{A_k} = \epsilon_{C_k} = 0.05$, red circle: $\epsilon_{A_k} = \epsilon_{C_k} = 0.10$)

the issue of computational complexity by exploiting the inherently sparse structure of the problem and by considering a restricted memory estimator in a receding horizon context. A point worth noting is that consistent numerical experience shows that the bounds obtained in Theorem 1 are conservative and that indeed optimal solutions are obtained with substantially smaller size relaxations. However, no formal proof of this fact is currently available.

References

[1] A. Alessandri, M. Baglietto, and G. Battistelli. Receding-horizon estimation for switching discrete-time linear systems. *Automatic Control, IEEE Transactions on*, 50(11):1736–1748, 2005.

[2] A. Alessandri and P. Coletta. Design of luenberger observers for a class of hybrid linear systems. *Hybrid systems: computation and control*, pages 7–18, 2001.

[3] M. Babaali and M. Egerstedt. Asymptotic observers for discrete-time switched linear systems. In *Proceedings of the 16th IFAC World Congress*, volume 16, 2005.

[4] M. Babaali, M. Egerstedt, and E. W. Kamen. A direct algebraic approach to observer design under switching measurement equations. *Automatic Control, IEEE Transactions on*, 49(11):2044–2049, 2004.

[5] A. Doucet, A. Logothetis, and V. Krishnamurthy. Stochastic sampling algorithms for state estimation of jump markov linear systems. *Automatic Control, IEEE Transactions on*, 45(2):188–202, 2000.

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

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

[8] 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.

[9] A. B. Memon and E. I. Verriesty. Suboptimal multi-mode state estimation and mode detection. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 2034–2039. IEEE, 2012.

[10] J. Ragot, D. Maquin, E. A. Domlan, et al. Switching time estimation of piecewise linear systems. application to diagnosis. In *5th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, Safeprocess' 2003*, 2003.

[11] M. Sznaier, B. Yimaz, and F. Blanchini. Bounded complexity ℓ_∞ filters for switched systems. In *American Control Conference*, 2011.

[12] J. F. Traub and H. Woźniakowski. *A general theory of optimal algorithms*. Academic Press New York, 1980.

[13] J.F. Traub, G.W. Wasilkowski, and H. Wozniakowski. *Information-Based Complexity*. Academic Press, New York, 1988.

[14] V. A Yakubovich. S-procedure in nonlinear control theory. *Vestnik Leningrad Univ*, (4):73–93, 1977.