

# Hybrid System Identification with Faulty Measurements and its Application to Activity Analysis

Necmiye Ozay

Mario Sznaier

**Abstract**—This paper addresses the problem of set membership identification of a class of discrete-time affine hybrid systems, switched affine models, in the presence of sensor failures. Given a finite collection of input/output measurements and a bound on the number of subsystems, the objective is to identify a suitable set of affine models along with a switching sequence that can explain the available experimental information. Contrary to existing work, here we allow for instantaneous failures in the measurement sensors at unknown times. These failures lead to corrupted input/output data, that if used in the identification process would result in substantial identification errors. The main result of the paper shows that, exploiting the fact that these failures are infrequent, combined with an algebraic-geometric argument, allows for recasting the problem into an optimization form where the objective is to simultaneously minimize the rank of a matrix and the number of nonzero rows of a second one. While in principle this is a challenging, non-convex problem, exploiting recent results on convex relaxations of rank and block-sparsity leads to an efficient, semi-definite optimization based identification algorithm. Finally, these results are illustrated using both simulations and a practical example that arises in computer vision where the aim is to analyze the activity of a person in the presence of sensor failures.

## I. INTRODUCTION AND MOTIVATION

This paper addresses the problem of robust identification of a class of discrete-time affine hybrid systems, switched affine models, in cases where the experimental data is collected using sensors subject to random instantaneous outages. Specifically, given a finite collection of input/output measurements, some of which do not contain information about the system to be identified, and a bound on the number of subsystems, our objective is two-fold: (i) identify the faulty measurements and, (ii) identify a suitable set of affine models along with a switching sequence that can explain the experimental data record once the faulty measurements are removed. This problem arises in many practical situations, typically involving data collected remotely and transmitted over a channel subject to outages. In particular, the application motivating this paper, activity recognition, is drawn from computer vision. The ability to recognize activities from video sequences is central to many applications, including detecting contextually abnormal activity (both for security and assisted living scenarios), and human

computer interfaces. Formally, this problem can be stated as: Given a video clip containing several activities, each one spanning an unknown time length, parse the sequence into subsequences, each corresponding to a single activity, classify each of these segments and determine whether or not the entire sequence contains abnormal activities (defined as those that are not contained in a given database). As recently shown in [13], [8], very high recognition rates can be achieved by postulating that the observed data is the output trajectory of an underlying switched affine system, where each submodel corresponds to a sub-activity. In this context, activity recognition reduces to a switched systems model (in)validation problem: determine whether or not each subsystem is contained in a database of models of known activities and associated uncertainty descriptions.

Clearly, a prerequisite for performing the above invalidation is the ability to robustly identify each of the subsystems as well as the switching sequence. Identification of switched affine systems has been an area of intense research in the past few years. An excellent tutorial and a summary of the state of the art as of 2007, can be found in [14]. Later developments include sparsification [12] and polynomial optimization based approaches [11], [4], [3] that extended the algebraic ideas originally proposed in [9] to handle process and measurement noise. However, these methods fail in scenarios of the type motivating this paper, characterized by the presence of corrupted or missing measurements, due for instance to occlusion, or interference or outages in the communication channel. Indeed, in these situations, existing algorithms would attempt to capture the faulty data by adding spurious sub-systems and switches to the overall model and switching sequence, leading to incorrect models. As a consequence, in the application of interest here, activity recognition, admissible data sequences could be labeled abnormal.

In principle, the problem above can be circumvented by restricting the identification algorithm to use only those measurements known to be reliable (note that data records with gaps pose no particular problems to the algebraic methods mentioned above). However, the main difficulty in pursuing this approach is that neither the underlying dynamics, the switching sequence nor the location of the faulty measurements are a-priori known<sup>1</sup>.

NO is with the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA. MS is with the Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115, USA. This work has been supported in part by NSF grants IIS-0713003, ECCS-0901433, AFOSR grant FA9550-09-0253 and the Alert DHS Center of Excellence under Award Number 2008-ST-061-ED0001. The work of NO was supported in part by the Boeing Corporation.

<sup>1</sup>Clearly if the underlying dynamics and switching sequence are known, faulty measurements can be identified through a model (in)validation step. Conversely, if the location of the faulty measurements is known, the underlying system can be identified by eliminating these from the data record. The switching sequence can then be identified via model (in)validation.

To address this difficulty, in this paper we will exploit the fact that the failures under consideration are sparse. When combined with an algebraic geometric identification algorithm, this observation allows for recasting the identification with faulty measurements problem into an optimization form where the objective is to simultaneously minimize the rank of a matrix and the number of nonzero rows of a second one, both generated from the experimental information. Finally, the use of recently developed convex relaxations for problems of this type, leads to a computationally efficient algorithm, based on the solution of a semi-definite optimization problem.

These results are illustrated using both simulations and real data involving an activity recognition example in the presence of sensor failures.

## II. PRELIMINARIES

For ease of reference, in this section we summarize the notation used in the paper and recall some results that will be used to recast the identification problem into a convex optimization form.

### A. Notation

By  $\mathbf{x}$  and  $\mathbf{M}$ , we, respectively, denote a vector in  $\mathbb{R}^n$  and a matrix in  $\mathbb{R}^{n \times m}$ .  $\|x\|_\infty \doteq \sup_i |x_i|$  is the  $\infty$ -norm of a vector.  $\|\mathbf{M}\|_{row,0}$  is the number of nonzero rows of the matrix  $\mathbf{M}$ .  $\mathbf{I}$  denotes the identity matrix of appropriate dimensions.  $\mathbf{M} \succeq \mathbf{N}$  means the matrix  $\mathbf{M} - \mathbf{N}$  is positive semidefinite. Veronese map of degree  $s$ ,  $\nu_s$ , is a mapping from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  with  $m = \binom{s+n-1}{s}$ , defined by

$$\nu_s([x_1, \dots, x_n]^T) = [\dots, \xi^s, \dots]^T$$

where  $\xi^s \doteq x_1^{s_1} x_2^{s_2} \dots x_n^{s_n}$ ,  $\sum s_i = s$ , e.g. all possible monomials of order  $s$ , in lexicographical order.

### B. The Problem of Moments

Given a sequence of scalars  $\{m_i\}_{i=1}^n$ , the *problem of moments* is to determine whether there exist a probability measure that has  $\{m_i\}$  as its first  $n$  moments (see references [15], [5], [1] for a historical review and details of the problem). In particular, in the sequel we are interested in probability measures that are supported on bounded symmetric intervals of the real line. In this case, the following theorem provides necessary and sufficient conditions for the existence of such a measure.

*Theorem 1:* Given a sequence  $\{m_i : i = 1, 2, \dots, n\}$ , there exists a probability measure supported on  $[-\epsilon, \epsilon]$  such that

$$m_i = \mathcal{E}_\mu(x^i) = \int_{-\epsilon}^{\epsilon} x^i \mu(dx)$$

if and only if

- when  $n = 2k + 1$  (odd case), the following holds

$$\epsilon \mathbf{M}(0, 2k) \succeq \mathbf{M}(1, 2k + 1) \quad (1)$$

$$\mathbf{M}(1, 2k + 1) \succeq -\epsilon \mathbf{M}(0, 2k) \quad (2)$$

- when  $n = 2k$  (even case), the following holds

$$\mathbf{M}(0, 2k) \succeq 0 \quad (3)$$

$$\epsilon^2 \mathbf{M}(0, 2k - 2) \succeq \mathbf{M}(2, 2k) \quad (4)$$

where  $\mathbf{M}(i, i + 2j)$  is the  $(j + 1)$  by  $(j + 1)$  Hankel matrix formed from the moments, that is:

$$\mathbf{M}(i, i + 2j) \doteq \begin{bmatrix} m_i & m_{i+1} & \dots & m_{i+j} \\ m_{i+1} & \ddots & \ddots & m_{i+j+1} \\ \vdots & \ddots & \ddots & \vdots \\ m_{i+j} & \dots & \dots & m_{i+2j} \end{bmatrix}, \quad (5)$$

and where  $m_0 = 1$ .

*Proof:* Direct application of Theorem III.2.3 and Theorem III.2.4 in [5].  $\blacksquare$

The problem of moments, especially its multivariate extensions, has been used in the optimization community to convert polynomial optimization problems into a hierarchy of convex semidefinite programming problems with increasing size ([6], [7]). In this paper, we apply similar ideas to feasibility problems involving a combination of polynomial and rank constraints. By exploiting the problem structure, we show that it suffices to use one dimensional distributions for which the moments can be precisely characterized by fixed sized Linear Matrix Inequalities of the form given in Theorem 1.

## III. PROBLEM SETUP

In this paper we consider the problem of set membership identification of single input single output, switched autoregressive exogenous (SARX) linear models of the form:

$$\begin{aligned} y_t &= \sum_{i=1}^{n_a} a_i(\sigma_t) y_{t-i} + \sum_{i=1}^{n_c} c_i(\sigma_t) u_{t-i} + \eta_t \\ \tilde{y}_t &= \begin{cases} y_t & \text{if } f_t = 0 \\ \theta_t & \text{otherwise} \end{cases} \end{aligned} \quad (6)$$

where  $u$ ,  $\tilde{y}$  and  $\eta$  denote the input, output and process noise respectively;  $f_t$  is an unknown sparse binary sequence that represents the reliability of measurements (i.e.  $f_t = 1$  for time instances when the measurement sensor fails),  $\theta_t$  denotes the faulty measurements, and  $\sigma_t \in \{1, \dots, s\}$  is the discrete state or mode of the system. Different values of  $\sigma_t$  correspond to  $s$  different hybrid submodels. Given the number of submodels  $s$  and input/output data over the interval  $[t_0, T]$ , our goal is to find a model of the form (6) that interpolates this experimental data within a given process noise level  $\|\eta_t\|_\infty \leq \epsilon$  and with as few faulty measurements as possible.

In the ideal case of noiseless measurements and sensor faults, Vidal et al. [16], [9] proposed an elegant algebraic geometric solution to the identification of SARX models. The case where the system is subject to bounded process and measurement noise, was recently addressed in [11], [4], [3], by exploiting a combination of results from the classical theory of moments and convex optimization. In this paper, we further extend this framework to accommodate (infrequent) faulty measurements, due for instance to temporary sensor faults.

#### IV. MAIN RESULTS

In this section we present the main result of the paper, a semi-definite programming based algorithm for identifying SARX models in the presence of measurement noise and sensor faults. Towards this goal, we first recall the main result in [11] that reduces the SARX identification problem to a constrained rank minimization. In the absence of sensor faults (i.e.  $\tilde{y}_t = y_t$ ), (6) can be rewritten as

$$\mathbf{b}(\sigma_t)^T \mathbf{r}_t = 0 \quad (7)$$

where  $\mathbf{r}_t = [-\tilde{y}_t + \eta, \tilde{y}_{t-1}, \dots, \tilde{y}_{t-n_a}, u_{t-1}, \dots, u_{t-n_c}]^T$  and  $\mathbf{b}(\sigma_t) = [1, a_1(\sigma_t), \dots, a_{n_a}(\sigma_t), c_1(\sigma_t), \dots, c_{n_c}(\sigma_t)]^T$ . Next, note that (7) holds for some  $\sigma_t$  iff

$$p_s(\mathbf{r}) = \prod_{i=1}^s (\mathbf{b}_i^T \mathbf{r}_t) = \mathbf{c}_s^T \nu_s(\mathbf{r}_t) = 0 \quad (8)$$

holds for all  $t$  independent of which of the  $s$  submodels is active at time  $t$ . In the above equality,  $\mathbf{b}_i \in \mathbb{R}^{n_a+n_c+1}$  is the parameter vector corresponding to the  $i^{\text{th}}$  submodel,  $\mathbf{r}_t$  is the regressor vector at time  $t$ , and  $\nu_s(\cdot)$  is the Veronese map of degree  $s^2$ . Rewriting  $\mathbf{r}_t$  explicitly as a combination of the known regressor vector  $\tilde{\mathbf{r}}_t$  of measurements and inputs, and unknown noise terms  $\eta_t$ ; and collecting all data into a matrix form leads to:

$$\mathbf{V}_s \mathbf{c}_s \doteq \begin{bmatrix} \nu_s(\tilde{\mathbf{r}}_{t_0}, \eta_{t_0})^T \\ \vdots \\ \nu_s(\tilde{\mathbf{r}}_T, \eta_T)^T \end{bmatrix} \mathbf{c}_s = \mathbf{0} \quad (9)$$

Thus, in this context, the identification problem reduces to finding an admissible noise sequence  $\{\eta_t\}$  that renders  $\mathbf{V}_s$  rank deficient, and a vector  $\mathbf{c}_s$  in its null space, leading to the following feasibility problem:

$$\begin{aligned} & \text{find} && \eta_t \\ & \text{subject to} && \text{rank}[\mathbf{V}_s(\tilde{\mathbf{r}}_t, \eta_t)] \leq h \\ & && \|\eta_t\|_\infty \leq \epsilon \end{aligned} \quad (10)$$

where  $h$  is the number of columns of  $\mathbf{V}_s(\tilde{\mathbf{r}}_t, \eta_t)$  minus one. The main result in [11] shows that this problem can be recast into a rank minimization problem affine in the optimization variables (via a moments argument) and relaxed to an SDP optimization. Unfortunately, the reasoning above breaks down in the presence of sensor failures. In this case, when a faulty measurement is lifted using the Veronese map, the rows of the embedded data matrix that contain the faulty measurement are typically far from the row space of the ‘‘clean’’ Veronese map, rendering problem (10) infeasible. As we show in the sequel, this difficulty can be circumvented by reducing the problem to a rank minimization with additional sparsity constraints. Briefly, the main idea is to introduce a (row sparse) error matrix  $\mathbf{E}$ , with the same dimensions as  $\mathbf{V}_s$  that, in a sense to be made precisely later, compensates the effect of the corrupted measurements. The first step towards this goal is to reformulate Problem (10) as a feasibility

<sup>2</sup>This is essentially a noisy version of the hybrid decoupling constraint proposed in [16].

problem in terms of the moments of an unknown probability distribution function as follows:

$$\begin{aligned} & \text{find} && \mathbf{m}^{(t)} \\ & \text{subject to} && \text{rank}[\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)})] \leq h \\ & && (1) - (2) \quad \forall \mathbf{m}^{(t)} \quad \text{if } s \text{ is odd} \\ & && (3) - (4) \quad \forall \mathbf{m}^{(t)} \quad \text{if } s \text{ is even} \end{aligned} \quad (11)$$

where  $\mathbf{m}^{(t)} = [m_1^{(t)}, \dots, m_s^{(t)}]$  is the moment sequence corresponding to  $\eta_t$  and  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)})$  is a matrix linear in the moments, obtained by replacing each  $k^{\text{th}}$  degree monomial  $\eta_t^k$  in  $\mathbf{V}_s(\tilde{\mathbf{r}}_t, \eta_t)$  with the corresponding  $k^{\text{th}}$  order moment  $m_k^{(t)}$ .

*Theorem 2:* Problem (10) is feasible if and only if problem (11) is feasible. Moreover, if  $\mathbf{c}$  belongs to the nullspace of  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)})$  in a solution of (11), then there exists a noise sequence  $\eta_t^*$  with  $\|\eta_t^*\|_\infty \leq \epsilon$  such that  $\mathbf{c}$  belongs to the nullspace of  $\mathbf{V}_s(\tilde{\mathbf{r}}_t, \eta_t^*)$ .

*Proof:* If (10) is feasible, then there exists some sequence  $\eta_t^*$  such that  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{*(t)})$  with  $\mathbf{m}^{*(t)} = [\eta_t^*, (\eta_t^*)^2, \dots, (\eta_t^*)^s]$  (i.e. all distributions have point support) has rank  $r_1 \leq h$  and  $\mathbf{m}^{*(t)}$  satisfies the LMI constraints (1)-(2) or (3)-(4). Hence, (11) feasible.

Consider now a feasible solution  $\mathbf{m}^{*(t)}$  of (11). From Theorem 1, feasibility of the LMIs guarantees the existence of  $T - t_0 + 1$  measures  $\mu^{*(t)}$ , each supported on  $[-\epsilon, \epsilon]$  and having  $\mathbf{m}^{(t)}$  as its moments. Let  $\mathbf{c}$  be in the nullspace of  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{*(t)})$  (i.e.  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{*(t)})\mathbf{c} = \mathbf{0}$ ). Thus, for each row of  $\mathbf{V}_s$ ,  $\mathcal{E}_{\mu^{*(t)}}[\nu_s(\tilde{\mathbf{r}}_t, \eta_t)^T] \mathbf{c} = \mathcal{E}_{\mu^{*(t)}}[\nu_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)})^T \mathbf{c}] = 0$ . Since  $\nu_s(\tilde{\mathbf{r}}_t, \eta_t)^T \mathbf{c}$  is a polynomial function of  $\eta_t$  (hence continuous) and  $\mu^{*(t)}$  is supported on  $[-\epsilon, \epsilon]$ , from direct application of the mean value theorem for integration it follows that there exist  $\eta_t^* \in [-\epsilon, \epsilon]$  for all  $t$  such that  $\nu_s(\tilde{\mathbf{r}}_t, \eta_t^*)^T \mathbf{c} = 0$ . Thus, whenever the nullspace of the solution of (10) is non-trivial, so is that of (11), which proves the theorem. ■

As indicated before, in the presence of faulty measurements, (11) is typically no longer feasible, since the corrupted measurements do not satisfy (7), or equivalently, there is no vector  $\mathbf{c} \neq 0$  such that  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)})\mathbf{c} = 0$  for all  $t$ . The effect of the corrupted rows can be eliminated by introducing an error matrix  $\mathbf{E}$  such that  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)}) + \mathbf{E}$  is rank deficient. Further, under the assumption that faults are infrequent,  $\mathbf{E}$  should be row sparse (that is, only a few rows, corresponding to the rows in  $\tilde{\mathbf{V}}$  affected by the corrupted measurements, should be non-zero). From the reasoning above, it follows that faulty measurements can be accommodated by considering the following optimization problem:

$$\begin{aligned} & \text{minimize}_{\mathbf{m}^{(t)}, \mathbf{E}} && \|\mathbf{E}\|_{\text{row}, 0} \\ & \text{subject to} && \text{rank}[\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)}) + \mathbf{E}] \leq h \\ & && (1) - (2) \quad \forall \mathbf{m}^{(t)} \quad \text{if } s \text{ is odd} \\ & && (3) - (4) \quad \forall \mathbf{m}^{(t)} \quad \text{if } s \text{ is even.} \end{aligned} \quad (12)$$

Note that if there are  $k$  faults,  $\mathbf{E}$  would have at most  $k(n_a+1)$  nonzero rows, since a single faulty measurement affects  $n_a+1$  rows of  $\tilde{\mathbf{V}}$ . Therefore, minimizing the nonzero rows of  $\mathbf{E}$  amounts to minimizing the number of faulty measurements.

*Theorem 3:* Let the optimum of (12) be  $e$  and let  $\{r_i\}_{i=1}^e$  be the indices of nonzero rows of  $\mathbf{E}^*$  in the optimal solution. If  $\mathbf{c}$  is a vector in the null space of  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)*}) + \mathbf{E}^*$ , then, for  $t \in [t_0, T] \setminus \{r_i\}_{i=1}^e$ , there exists a noise sequence  $\eta_t^*$  with  $\|\eta_t^*\|_\infty \leq \epsilon$  such that  $\mathbf{c}$  belongs to the nullspace of  $\mathbf{V}_s(\tilde{\mathbf{r}}_t, \eta_t^*)$ .

*Proof:* Since  $\mathbf{c}$  is in the nullspace of  $\mathbf{N} \doteq \tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)*}) + \mathbf{E}^*$ , it is also in the nullspace of the submatrix  $\mathbf{N}'$  that is formed by eliminating the rows of  $\mathbf{N}$  that correspond to  $\{r_i\}_{i=1}^e$ . Note that only  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)*})$  contributes to  $\mathbf{N}'$ , since the corresponding rows in  $\mathbf{E}^*$  are all zeros. Therefore,  $\mathbf{N}'\mathbf{c} = 0$  together with the moment constraints in (12) implies that, for all  $t \in [t_0, T] \setminus \{r_i\}_{i=1}^e$ , there exists measures  $\mu^{*(t)}$ , each supported on  $[-\epsilon, \epsilon]$  and having  $\mathbf{m}^{(t)}$  as their moments. Hence  $\mathcal{E}_{\mu^{*(t)}}[\nu_s(\tilde{\mathbf{r}}_t, \eta_t)^T] \mathbf{c} = 0$ . Similar to the proof of Theorem 2, we can invoke the mean value theorem for integration to conclude that there exist  $\eta_t^* \in [-\epsilon, \epsilon]$  for all  $t \in [t_0, T] \setminus \{r_i\}_{i=1}^e$  such that  $\nu_s(\tilde{\mathbf{r}}_t, \eta_t^*)^T \mathbf{c} = 0$ . ■

Intuitively, the result above states that, in the presence of outliers, it is possible to modify just a few rows of the Veronesi map (precisely those corrupted by the outliers) in such a way that the modified matrix and the ideal, uncorrupted Veronesi map have the same null space, for some admissible noise sequence  $\{\eta^*\}$ . It follows that, if there are enough uncorrupted rows to completely characterize this null space, then it can be identified by solving Problem (12). Once this null space has been identified, the coefficients of each submodel can be recovered by proceeding as in [16], [9]. Note that this approach is capable of handling faulty measurements of the input sequence as well, since these also result in corrupted rows in the embedded data matrix.

*Remark 1:* It is worth emphasizing that in the proposed approach the matrix  $\mathbf{E}$  enters linearly the constraints in (12) and its elements are decoupled from those of the moments sequence  $\mathbf{m}^{(t)}$ . For comparison, introducing a matrix  $\mathbf{E}$  in (10) leads to polynomial constraints involving both the elements of  $\eta_t$  and  $\mathbf{E}$  (originating from expanding the constraint  $\text{rank}[\mathbf{V}_s(\tilde{\mathbf{r}}_t, \eta_t^*) + \mathbf{E}] \leq h$ ), hence necessitating the consideration of the joint moments. Similarly, working directly with the sequence  $\theta_t$ , unknowns  $y_t$  and  $f_t$  in (6), requires considering the joint moments of the sequences  $\{\eta_t\}$ ,  $\{\theta_t\}$ ,  $\{y_t\}$  and  $\{f_t\}$ , substantially increasing the computational complexity.

#### A. Convex relaxations for rank and row sparsity

In principle, Problem (12) is generically NP-hard, due to both, the objective function and the rank constraint. However, as we show next, efficient convex relaxations can be obtained by combining recent results on rank minimization and block sparsification. The main idea is to first replace the objective function by

$$\text{rank}[\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}_t, \mathbf{m}^{(t)}) + \mathbf{E}] + \gamma \|\mathbf{E}\|_{\text{row},0}$$

where  $\gamma$  is a suitably chosen Lagrange multiplier. Next, note that enforcing row sparsity of  $\mathbf{E}$  is equivalent to enforcing sparsity of a vector formed by norms of its rows. Following

the arguments in [17], [10], we relax  $\|\mathbf{E}\|_{\text{row},0}$  to  $\sum_i \|\mathbf{E}_i\|_2$ , where  $\mathbf{E}_i$  denotes the  $i^{\text{th}}$  row of  $\mathbf{E}$ , and we relax rank to nuclear norm. Finally, using a semi-definite characterization of the nuclear norm [2], combined with a re-weighted heuristic, leads to the following algorithm, based on solving a sequence of convex optimization problems:

---

#### Algorithm 1: SARX Identification with outliers

---

initialize:

$$k = 0$$

$$W_y^{(0)} = I_{m \times m}$$

$$W_z^{(0)} = I_{n \times n}$$

$$w_i^{(0)} = \frac{1}{\sqrt{N_p}}$$

$\delta =$  small positive constant

REPEAT

Solve

$$\min_{\mathbf{m}, Y^{(k)}, Z^{(k)}, \mathbf{E}^{(k)}} \text{Tr} \left[ W_y^{(k)} Y^{(k)} \right] + \text{Tr} \left[ W_z^{(k)} Z^{(k)} \right] + \gamma^{(k)} \sum_{i=1}^{N_p} w_i^{(k)} \|\mathbf{E}_i^{(k)}\|_2$$

subject to

$$\begin{bmatrix} Y^{(k)} & \tilde{\mathbf{V}}_s(\mathbf{m}^{(k)}) + \mathbf{E}^{(k)} \\ [\tilde{\mathbf{V}}_s(\mathbf{m}^{(k)}) + \mathbf{E}^{(k)}]^T & Z^{(k)} \end{bmatrix} \succeq 0$$

and (1)-(2)  $\forall \mathbf{m}$  if  $s$  is odd,

or (3)-(4)  $\forall \mathbf{m}$  if  $s$  is even.

Let  $\tilde{\mathbf{V}}_s(\mathbf{m}^{(k)}) + \mathbf{E}^{(k)} = UDV^T$ .

Set  $\epsilon = D(m, m)$ .

Set  $W_y^{(k+1)} = (Y^{(k)} + \epsilon I)^{-1}$ .

Set  $W_z^{(k+1)} = (Z^{(k)} + \epsilon I)^{-1}$ .

Set  $w_i^{(k+1)} = \frac{1}{\|\mathbf{E}_i^{(k)}\|_2 + \delta}$ .

Set  $\gamma^{(k+1)} = \gamma^{(0)} \frac{\|W_y^{(k)}\|_2 + \|W_z^{(k)}\|_2}{2 \|[w_1^{(k)} \dots w_{N_p}^{(k)}]\|_2}$ .

Set  $k = k + 1$ .

UNTIL (a convergence criterion is reached)

RETURN  $\tilde{\mathbf{V}}_s(\mathbf{m}^{(k)}), \mathbf{E}^{(k)}, \mathbf{m}^{(k)}$

---

where  $\tilde{\mathbf{V}}_s(\tilde{\mathbf{r}}, \mathbf{m}^{(k)})$  is denoted by  $\tilde{\mathbf{V}}_s(\mathbf{m}^{(k)})$  for brevity and where  $\delta > 0$  is a small regularization constant. The initial value of the Lagrange multiplier,  $\gamma^{(0)}$  is a tuning parameter that controls the number of measurements that are discarded. A small value of  $\gamma^{(0)}$  will typically result in a large number of measurements labeled as faulty and dropped. On the other hand, a large value of  $\gamma^{(0)}$  will result in fewer discarded measurements and hence may preclude finding a rank deficient solution. Once a suitable value of  $\gamma^{(0)}$  is selected, the update rule for  $\gamma^{(k)}$  in the algorithm above attempts to keep the relative weights of the rank minimization and row sparsity enforcing terms approximately constant throughout the optimization.

## V. ILLUSTRATIVE EXAMPLES

In this section we illustrate the ability of the proposed method to correctly identify a SARX system in the presence of outliers using both an academic example and a non-trivial computer vision application.

### A. Academic Example

Consider a hybrid system that switches among the following two ARX subsystems

$$y_t = 0.2y_{t-1} + 0.24y_{t-2} + 2u_{t-1} + \eta_t \quad (\text{Submodel 1})$$

$$y_t = -1.4y_{t-1} - 0.53y_{t-2} + u_{t-1} + \eta_t \quad (\text{Submodel 2})$$

modeled as

$$y_t = p_1(\sigma_t)y_{t-1} + p_2(\sigma_t)y_{t-2} + p_3(\sigma_t)u_{t-1} + \eta_t$$

$$\tilde{y}_t = \begin{cases} y_t & \text{if } f_t = 0 \\ \theta_t & \text{otherwise} \end{cases} \quad (13)$$

where  $\sigma_t \in \{1, 2\}$  depending on which model is active at time  $t$ . The system experimental data consists of  $T = 100$  input/output measurements obtained by setting  $\sigma_t = 1$  for  $t \in [1, 25] \cup [51, 75]$  and  $\sigma_t = 2$  for  $t \in [26, 50] \cup [76, 100]$ , corrupted with noise bounded by  $\|\eta\|_\infty = 0.23$ . The measured input  $u(t)$  used to drive the system is randomly drawn from a zero mean unit variance Gaussian distribution. Sensor faults were simulated at  $t \in \{10, 50, 80\}$  by using values  $\theta_t$  sampled from a zero mean Gaussian distribution with standard deviation equal to five times the standard deviation of the sequence  $y(t)$ . The parameter values used for the simulation are shown in Table I together with the results obtained by our method and the original algebraic method of [16]. Figures 1 and 2 show the clustering of data into different submodels. As seen there, the proposed method outperforms the method in [16].

		True	Moments-based	GPCA
Submodel 1	$p_1$	0.2000	0.0725	-0.0876
	$p_2$	0.2400	0.2953	0.4065
	$p_3$	2.0000	1.9645	2.0644
Submodel 2	$p_1$	-1.4000	-1.2738	-1.3972
	$p_2$	-0.5300	-0.4269	-0.3278
	$p_3$	1.0000	1.0867	-10.3061

TABLE I  
ESTIMATED AND TRUE VALUES OF PARAMETERS

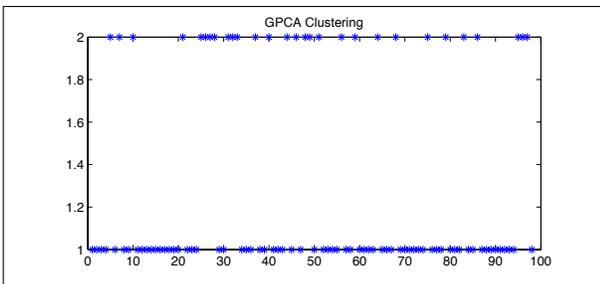


Fig. 1. Clustering via GPCA.

### B. Computer vision application: Activity analysis with sensor faults

Next, we illustrate the ability of the method to handle realistic scenarios by applying it to a non-trivial computer vision problem: human activity analysis. The goal here is to

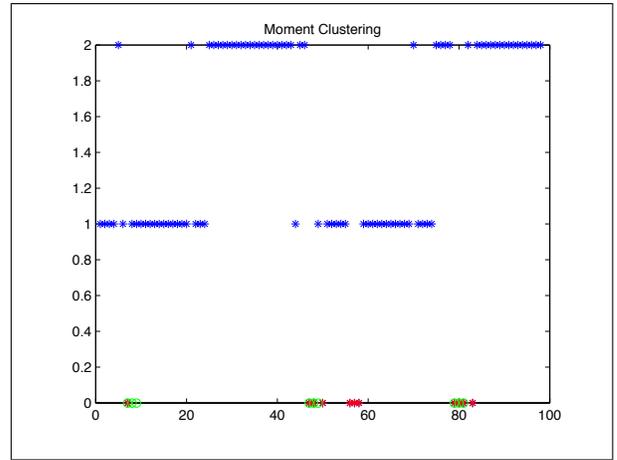


Fig. 2. Clustering via moments-based method. The outlying rows detected are denoted with red stars and the true outliers are denoted by green circles.

segment a video clip containing multiple activities into its constituent sub-activities and to find a model characterizing each of these, as a first step towards recognizing contextually abnormal situations. The data used in this particular example consists of 55 frames extracted from a video sequence of a person walking and bending in front of the camera. Three randomly chosen frames were corrupted with large amounts of noise to simulate instantaneous sensor failures (in this case interference in the wireless communication channel from the sensor to the base station). Figure 3 shows some sample frames from the sequence. Half way through the sequence the person bends down, then stands up and resumes walking. These frames were modeled as the output of an underlying switched affine system, with 2 submodels, each corresponding to a given activity. In particular, the horizontal<sup>3</sup> position of the center of mass was modeled as the output of a first order switched affine autoregressive system:

$$x_t = a(\sigma_t)x_{t-1} + d(\sigma_t) + \eta_t \quad (14)$$

where  $a(\sigma_t)$  and  $d(\sigma_t)$  are unknown parameters. We set  $\|\eta_t\|_\infty = 3$ , allowing  $\pm 3$  pixels noise.

For the measurements, we use a simple tracker based on background subtraction to estimate the location of the center of mass of the person in each frame. Sensor failures were captured with the following measurement equation:

$$\tilde{x}_t = \begin{cases} x_t & \text{if } f_t = 0 \\ \theta_t & \text{otherwise} \end{cases} \quad (15)$$

where  $f_t$  is an unknown (yet sparse) binary sequence with  $f_t = 0$  for healthy conditions and  $f_t = 1$  for times were the sensors fail, and where  $\theta_t$  is random Gaussian noise whose mean and standard deviation is chosen to be equal to those of  $x_t$ .

Figures 4 and 5 compare the results of applying the proposed identification method and GPCA in this scenario.

<sup>3</sup>It may seem more natural to use the vertical position. However, this would have resulted in 3 segments, corresponding to roughly no vertical motion, downward and upward motion, while there are only two different activities involved.

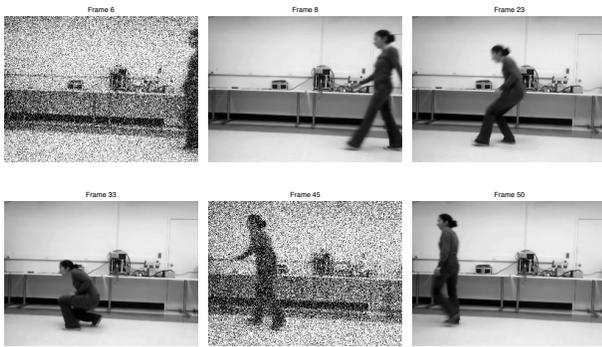


Fig. 3. Sample clean and corrupted frames from the video.

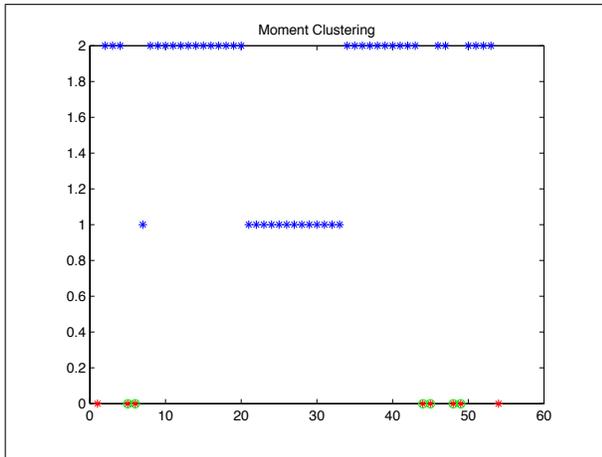


Fig. 4. Activity segmentation via the proposed method. The outlying rows detected are denoted with red stars and the true outliers are denoted with green circles.

As illustrated there, while the proposed method is able to identify the underlying subsystems, the switching sequence and the corrupted frames, leading to the correct segmentation, GPCA fails to do so. Although proposed method is computationally more expensive than GPCA (i.e. the former requires solving an SDP whereas the latter requires computing a singular value decomposition), it still leads to a tractable convex problem and performs significantly better.

## VI. CONCLUSIONS

In this paper we considered the problem of identifying a SARX model from noisy, potentially faulty measurements. This scenario is motivated by several application domains,

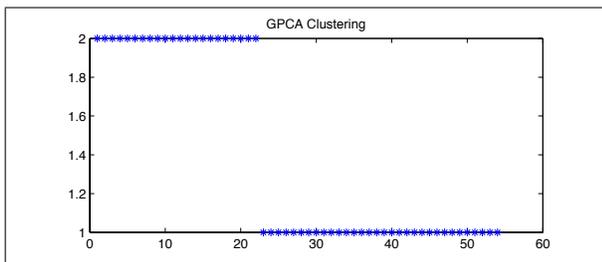


Fig. 5. Activity segmentation via GPCA.

including computer vision, where data is transmitted through channels subject to interference or outages. Under these conditions, existing algorithms will attempt to identify a model that also fits the corrupted data, by adding, if necessary, spurious subsystems, leading to an incorrect overall model. The main result of the paper shows that this situation can be avoided by recasting the problem into a rank minimization form, subject to additional block-sparsity constraints. While in principle this leads to challenging NP hard optimization problems, recent results on convex relaxations of rank and block-sparsity can be exploited to obtain a computationally tractable algorithm. These results were illustrated both with synthetic data and a non-trivial computer vision example: segmentation of a video clip containing multiple activities in the presence of corrupted frames. Research is currently underway seeking to extend these results to more general switched non-linear systems.

## REFERENCES

- [1] R.E. Curto and Fialkow L.A. Recursiveness, positivity, and truncated moment problems. *Houston J. Math.*, 17:603–635, 1991.
- [2] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.
- [3] C. Feng, C. Lagoa, N. Ozay, and M. Sznaier. Hybrid system identification: An sdp approach. In *Proc. 49<sup>th</sup> IEEE Conf. Dec. Control*, 2010.
- [4] C. Feng, C. Lagoa, and M. Sznaier. Hybrid system identification via sparse polynomial optimization. In *American Control Conference*.
- [5] M. G. Krein and A. A. Nudelman. *The Markov Moment Problem and Extremal Problems*, volume 50 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1977.
- [6] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optimization*, 11(3):796–817, 2001.
- [7] M. Laurent. *Emerging Applications of Algebraic Geometry*, chapter Sums of squares, moment matrices and optimization over polynomials, pages 157–270. Springer, 2009.
- [8] B. Li et al. Activity recognition using dynamic subspace angles. In *CVPR*, 2011.
- [9] 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.
- [10] M. McCoy and J. Tropp. Two Proposals for Robust PCA using Semidefinite Programming. *ArXiv e-prints*, December 2010.
- [11] N. Ozay, C. Lagoa, and M. Sznaier. Robust identification of switched affine systems via moments-based convex optimization. In *Proc. 48<sup>th</sup> IEEE Conf. Dec. Control*, pages 4686–4691, 2009.
- [12] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps. A sparsification approach to set membership identification of a class of affine hybrid systems. In *Proc. 47<sup>th</sup> IEEE Conf. Dec. Control*, pages 123–130, 2008.
- [13] 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 *CDC*, pages 7624–7630, 2010.
- [14] 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.
- [15] J.A. Shohat and J.D. Tamarkin. *The Problem of Moments*. Math. Surveys I. American Mathematical Society, Providence, RI, 1943.
- [16] R. Vidal, S. Soatto, Y. Ma, and S. Sastry. An algebraic geometric approach to the identification of linear hybrid systems. In *IEEE Conference on Decision and Control*, pages 167–172, December 2003.
- [17] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust pca via outlier pursuit. *CoRR*, abs/1010.4237, 2010.