Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Identification of Markov Jump Autoregressive Processes from Large Noisy Data Sets ⋆

Sarah Hojjatinia and Constantino Lagoa ∗

∗ *School of Electrical Engineering and Computer Science, The Pennsylvania State University, USA.*

Abstract: This paper introduces a novel methodology for the identification of switching dynamics for switched autoregressive linear models. Switching behavior is assumed to follow a Markov model. The system's outputs are contaminated by possibly large values of measurement noise. Although the procedure provided can handle other noise distributions, for simplicity, it is assumed that the distribution is Normal with unknown variance. Given noisy input-output data, we aim at identifying switched system coefficients, parameters of the noise distribution, dynamics of switching and probability transition matrix of Markovian model. System dynamics are estimated using previous results which exploit algebraic constraints that system trajectories have to satisfy. Switching dynamics are computed with solving a maximum likelihood estimation problem. The efficiency of proposed approach is shown with several academic examples. Although the noise to output ratio can be high, the method is shown to be effective in the situations where a large number of measurements is available.

Keywords: Markov Switched ARX, Identification for Control

1. INTRODUCTION

While identification of linear time invariant systems is by now a well understood problem, identification of switched and hybrid systems is considerably less developed, even in the piecewise affine case. Existing methods exploit a number of algebraic, optimization-based technique to find subsystem dynamics and switching surfaces; see Paoletti et al. (2007). A common feature is the computational complexity entailed in dealing with noisy measurements: in this case algebraic procedures lead to nonconvex optimization problems, while optimization methods lead to mixed integer/linear programming; see Roll et al. (2004); Ansaripour et al. (2016).

Similarly, methods relying on probabilistic priors lead to combinatorial problems; see Juloski et al. (2005) . This can be avoided by using clustering-based methods as in Nakada et al. (2005); Ferrari-Trecate et al. (2003). However, these require "fair sampling" of each cluster, which constrains the data that can be used. In Ozay et al. (2012, 2015); Bako (2011), some sparsification based-techniques for identification of affine switched models have been developed that allow for several types of noise.

This paper develops effective methods for identifying switching dynamics from large noisy data sets, for a broad class of systems described by switching autoregressive models. These systems can be considered a generalization of piecewise linear models, and breach the gap between linear and nonlinear models, retaining many of the tractability properties of the former, while providing descriptions that more accurately capture the features of practical problems over broader scenarios.

In identifying the parameters of switched models, the dynamics of switching play an important role. The interest in Markovian jump systems, switched system with switching dynamics based on a Markov chain, has been growing since they have a broad range of application in different areas and real world problems such as economic systems, power systems, networked control systems, neuroscience, and health care; see Shi and Li (2015); Hojjatinia and Lagoa (2019); Hojjatinia et al. (2020a). An example for application of Markovian jump systems in health care is m-health interventions for increasing light physical activity; see Lagoa et al. (2017). More precisely, the availability of activity tracking devices allows gathering of large amount of data such as individual's physical activity which is a dynamic behavior, so it can be modeled as a dynamical system. Furthermore, its characteristics may remarkably change based on the time in a day, weekdays or weekends, location, etc; which, motivated the approach of modeling it as a Markovian jump system; see Conroy et al. (2019); Hojjatinia et al. (2020c).

In comparison to the large amount of literature on analysis and control of Markovian jump systems, the identification problem seems to have received very little attention. In Hojjatinia et al. (2017) a new method for the identification of parameters of Markovian jump system is provided. The probability transition matrix is estimated using a suitable convex optimization problem. However, due to computational complexity, the number of measurement that the proposed approach is able to handle is limited and only

⋆ This work was partially supported by National Institutes of Health (NIH) Grant R01 HL142732, and National Science Foundation (NSF) Grant #1808266.

process noise was considered. In this paper, we focus on cases involving a very large number of measurements, possibly affected by large values of noise. In this case, polynomial/moments based approaches become ineffective, and different methodologies need to be devised. The approach we propose builds upon the same premises as Hojjatinia et al. (2019, 2020b).

More precisely we start by assuming that the output measurements are corrupted by random Normal measurement noise with unknown variance. Then, we exploit the availability of a large number of measurements and the results in Hojjatinia et al. (2019) to determine high confidence estimates of the systems parameters and the variance of the measurement noise. Finally, by using a maximum likelihood approach, we estimate the probability transition matrix and dynamics of switching. The approach can be easily extended to other noise distributions, as long as the number of unknown parameters of the distribution is "low."

1.1 Paper Organization

The paper is structured as follows: after this introduction, problem statement is defined in Section 2. Identification of system coefficients and noise parameters are reviewed in Section 3. In Section 4, the method for identification of probability transition matrix is described. Numerical results are shown in Section 5. Finally, Section 6 concludes the paper highlighting some possible future research directions.

2. PROBLEM STATEMENT

A precise description of the problem addressed is provided in this section. Assumptions needed to solve the problem are also introduced.

2.1 System Model

We consider switched autoregressive (SAR) linear models of the form

$$x_k = \sum_{j=1}^{n_a} a_{j\delta_k}\ x_{k-j} + \sum_{j=1}^{n_c} c_{j\delta_k}\ u_{k-j} \tag{1}$$

where $x_k \in \mathbb{R}$ is the output at time k and $u_k \in \mathbb{R}$ is input at time k. The variable $\delta_k \in \{1, ..., n\}$ denotes the subsystem active at time k, where n is the total number of sub-systems. Furthermore, $a_{j\delta_k}$ and $c_{j\delta_k}$ denote unknown coefficients corresponding to mode δ_k. Time k takes values over the non-negative integers. The latent discrete state δ_k evolves according to a Markov chain with transition probability matrix P, whose ij entry is

$$P_{ij} = P\{\delta_{k+1} = j \mid \delta_k = i\} \tag{2}$$

Output is assumed to be contaminated by (possibly large) noise; i.e. observations are of the form:

$$y_k = x_k + \eta_k \tag{3}$$

where η_k, denotes measurement noise.

The following assumptions are made on the system model and noise.

Assumption 1. Throughout this paper it is assumed that:

- Model orders n_a and n_c are available.
- Number of sub-systems n is available.
- Switching sequence is based on a Markov process.
- Each subsystem is "visited" infinitely often. More precisely, the Markov chain is assumed to be irreducible so it is possible to go to any state from any state; see e.g., Grimmett et al. (2001).
- Measurement noise η_k has zero mean Normal distribution with unknown variance.
- Noise η_k is independent from η_l for $k \neq l$, and identically distributed.
- Input sequence u_k applied to the system is known and bounded.
- There exists a finite constant L so that $|x_k| \leq L$ for all positive integers k.

Note that, the approach in this paper can be extended for any noise distribution as long as the number of unknown distribution parameters is "small."

2.2 Problem Definition

The main objective of this paper is to develop algorithms to identify the parameters of SAR systems, noise parameters, and dynamics of switching from noisy observations. More precisely, we aim at solving the following problem:

Problem 1. Given Assumption 1, an input sequence u_k, $k = -n_c + 1, \ldots, N - 1$ and noisy output measurements y_k, $k = -n_a + 1, \ldots, N$, determine

(1) Coefficients of the SAR model $a_{i,j}$, $i = 1, 2, \ldots, n_a$, $j = 1, 2, \ldots, n$, $c_{i,j}$, $i = 1, 2, \ldots, n_c$, $j = 1, 2, \ldots, n$,
(2) Noise distribution parameters,
(3) Switching sequence δ_k , $k = 1, 2, \ldots, n$ which is based on a Markov process.

3. REVIEW: IDENTIFICATION OF SYSTEM COEFFICIENTS AND NOISE PARAMETERS

To identify the coefficients of SAR system with measurement noise from large amount of data, we adopt the approach developed in Hojjatinia et al. (2019). For the sake of completeness, we briefly summarize the approach in this section. We refer the reader to Hojjatinia et al. (2019) for more details.

First, we review earlier results on an algebraic reformulation of the SAR identification problem for the case where no noise is present. Details on the algebraic approach to switched system identification can be found in Vidal et al. (2003).

The equation (1) is equivalent to

$$b_{\delta_k}^T\ r_k = 0 \tag{4}$$

where

$$r_k = [x_k,\ x_{k-1},\ \cdots,\ x_{k-n_a},\ u_{k-1},\ \cdots,\ u_{k-n_c}]^T$$

is the known regressor at time k, and

$$b_{\delta_k} = [-1,\ a_{1\delta_k},\ \cdots,\ a_{n_a\delta_k},\ c_{1\delta_k},\ \cdots,\ c_{n_c\delta_k}]^T.$$

is the vector of unknown coefficients at time k. Hence, independently of which of the n submodels is active at time k, we have

$$\Upsilon_n(r_k) = \prod_{i=1}^{n} b_i^T r_k = c_n^T \nu_n(r_k) = 0, \tag{5}$$

where the vector of parameters corresponding to the i-th submodel is denoted by $b_i \in \mathbb{R}^{n_a+n_c+1}$, and $\nu_n(\cdot)$ is Veronese map of degree n Harris (2013)

$$\nu_n([x_1, \ \cdots, \ x_s]^T) = [\cdots, \ x_1^{n_1} x_2^{n_2} \cdots x_s^{n_s}, \ \cdots]^T$$

which contains all monomials of order n in lexicographical order, and c_n is a vector whose entries are polynomial functions of unknown parameters b_i (see Vidal et al. (2005) for explicit definition). The Veronese map above is also known as polynomial embedding in machine learning Vidal et al. (2005). Veronese matrix V_n is of the form

$$V_n(r) = \left[\nu_n(r_1)^\top, \ \cdots, \ \nu_n(r_N)^\top\right]^\top \tag{6}$$

where r, without subscript, denotes the set of all regressor vectors. Note that the number of rows of the Veronese matrix V_n is equal to the number of measurements available for the regressor N. Therefore, a reformulation of the previous results to address the problem of identification from very large data sets is as follows Hojjatinia et al. (2019).

For the noiseless case, a reformulation of the hybrid decoupling constraint shows identifying the coefficients of the sub-models is equivalent to finding the singular vector c_n associated with the minimum singular value of the matrix

$$\overline{\mathcal{M}}_N = \frac{1}{N}\sum_{k=1}^{N} \nu_n(r_k)\nu_n^T(r_k) \doteq \frac{1}{N}\sum_{k=1}^{N} M_k \tag{7}$$

where, matrices are of size $\binom{n+n_a+n_c}{n}$, and size does not depend on the number of measurements, which is especially important in the case of very large data sets.

Identifying the parameters of the SAR model is equivalent to finding a vector in the null space of the matrix $\overline{\mathcal{M}}_N$. Under mild conditions, the null space of the matrix above has dimension one if and only if the data is compatible with the assumed model Ozay et al. (2015). However, when output is corrupted by noise, x_k is not known and, therefore, this matrix cannot be computed. However, we can use available information on the statistics of the noise and the measurements collected to compute approximations of the matrix $\overline{\mathcal{M}}_N$ and, consequently, approximations of vectors in its null space.

We start by noting that although x_k are unknown, the following holds

$$\begin{aligned} E[x_k^h] &= E[(y_k - \eta_k)^h] = E[y_k^h] - \sum_{d=1}^{h}\binom{h}{d} E[x_k^{h-d}]\, E[\eta_k^d] \\ &= E[y_k^h] - \sum_{d=1}^{h}\binom{h}{d} E[x_k^{h-d}]\, m_d \\ &\forall k = 1, 2, \cdots, N. \end{aligned} \tag{8}$$

where $E(\cdot)$ denotes expectation and m_d is the d^{th} moment of noise.

Hence, assuming that distribution of the noise, and the input signal are given and fixed, there exists an affine function $M(\cdot)$ so that

$$\begin{aligned} M_k &= E\{M[mon_n(y_k, \ldots, y_{k-n_a}, u_{k-1}, \ldots, u_{k-n_b})]\} \\ &= M\{E[mon_n(y_k, \ldots, y_{k-n_a}, u_{k-1}, \ldots, u_{k-n_b})]\}. \end{aligned}$$

where $mon_n(\cdot)$ denote a function that returns a vector with all monomials up to order n of its argument Hojjatinia et al. (2019).

This can be exploited to identify the parameters of the SAR system. The only thing needed is an estimation of the matrix $\overline{\mathcal{M}}_N$ in (7). It turns out that this can be done using the available noisy measurements. More precisely, we can construct the matrix

$$\widehat{\overline{\mathcal{M}}}_N \doteq \frac{1}{N}\sum_{k=1}^{N} M[mon_n(y_k, \ldots, y_{k-n_a}, u_{k-1}, \ldots, u_{k-n_b})]$$

and it is shown in Hojjatinia et al. (2019) that this matrix converges to $\overline{\mathcal{M}}_N$ in (7) as $N \to \infty$ almost surely. Hence, for large number of measurements N, the null space of the matrix $\widehat{\overline{\mathcal{M}}}_N$ can be used to determine the coefficients of the subsystems.

The above assumes knowledge of the moments of the noise. However, this does not need to be the case. In this paper, measurement noise is assumed to have a Normal distribution with zero mean and unknown variance σ^2. Since the moments are known functions of the variance, $\overline{\mathcal{M}}_N$ is a known function of σ and estimation of variance can be performed by minimizing the minimum singular value of matrix above over the allowable values of σ. More precisely, the parameters of the submodels and the variance of the noise can be identified using the following algorithm: Let n_a, n_c, n, some parameters of the noise and $\sigma_{\max}$ be given.

Step 1. Compute matrix $\widehat{\overline{\mathcal{M}}}_N$ as a function of the unknown noise parameter σ.
Step 2. Find the value $\sigma^* \in [0, \ \sigma_{\max}]$ that minimizes the minimum singular value of $\widehat{\overline{\mathcal{M}}}_N$.
Step 3. Let c_n be associated singular vector.
Step 4. Determine the coefficients of the subsystems from the vector c_n.

In order to perform Step 3 in Algorithm, we adopt polynomial differention algorithm for mixtures of hyperplanes, introduced by Vidal (Vidal and Sastry, 2003, pp. 69–70). In practice for sufficiently large N, the above algorithm provides both a good estimate of the systems coefficients and noise parameters, especially if we take σ^* to be the smallest value of σ for which the minimum singular value of $\widehat{\overline{\mathcal{M}}}_N$ is below a given threshold ϵ. Previous work cannot address the identification of switching dynamics and estimating the probability transition matrix of Markov jump models, this problem is explicitly addressed in the following section.

4. IDENTIFICATION OF PROBABILITY TRANSITION MATRIX

In Section 3, the algorithms and procedure of identifying noise parameters and system coefficients have been presented. In this section the switching behavior and dynamics of switching are considered. This is done in two steps:

The first step is to identify switches that have the highest probability of occurrence. Then, in the second step, by considering these switches as a good estimate of switching sequence, we estimate the transition probabilities.

4.1 Maximum likelihood switch sequence

Assume that the noise variance and system coefficients have been identified. To do the first step in identification of switching dynamics, i.e., determine the switches with highest probability, we start by building the following sequence based on available data and identified coefficients and parameters. Considering equations (1) and (3):

$$x_k = \sum_{j=1}^{n_a} a_{j\delta_k}\ x_{k-j} + \sum_{j=1}^{n_c} c_{j\delta_k}\ u_{k-j}$$
$$y_k = x_k + \eta_k$$

Since $x_k = y_k - \eta_k$. we have

$$\begin{aligned}\eta_k - &\sum_{j=1}^{n_a} a_{j\delta_k}\ \eta_{k-j} = \\ &y_k - \sum_{j=1}^{n_a} a_{j\delta_k}\ y_{k-j} - \sum_{j=1}^{n_c} c_{j\delta_k}\ u_{k-j}\end{aligned} \tag{9}$$

Since we have identified the coefficients $a_{j\delta_k}$ and $c_{j\delta_k}$, and input output (u, y) are available, we are able to determine the realization of the random variable in the right hand side of equation (9) for the all possible values of the switching sequence δ_k. Define

$$z_k(\delta_k) = y_k - \sum_{j=1}^{n_a} a_{j\delta_k}\ y_{k-j} - \sum_{j=1}^{n_c} c_{j\delta_k}\ u_{k-j} \tag{10}$$

To identify the most probable realization of the switches, we can use the values of $z_k(\delta_k)$ for each possible active system at time k ($\delta_k = \{1, \cdots, n\}$), and determine the sequence δ_k, $k = 1, 2, \ldots, N$ of maximum likelihood. However, for a fixed switching sequence, $z_k(\delta_k)$ is a sequence of correlated random variables. Even though the measurement noise is iid, $z_k(\delta_k)$ depends on η_{k-l} , $l = 0, 1, \ldots, n_a$ leading to a sequence of correlated random variables. Therefore, determining the values of δ_k , $k = 1, 2, \ldots, N$ that lead to the highest likelihood is a complex combinatorial problem.

To circumvent this, we start by noting that $z_k(\delta_k)$ is independent of $z_l(\delta_l)$ if $l > k + n_a$. Therefore, if enough data is available, we can use only independent "snippets" of data of low enough length for which: i) maximum likelihood sequence can be easily computed and ii) given that they are independent, likelihood can be computed individually for each snippet.

Hence, in the identification procedure proposed in this paper, we only consider snippets of data of the length n_l that are separated in time by at least n_a sample periods. More precisely, we consider snippets of data of length n_l, compute its joint distribution as a function of the switching sequence, determine the maximum likelihood switches for this snippet, skip the next n_a data points, and repeat the process until we run out of data.

We now elaborate on this. Take snippets of data of length n_l, denoted as a vector Z_k defined as:

$$\begin{aligned}Z_k =&[z_k(\delta_k), \cdots,\ z_{k+n_l-1}(\delta_{k+n_l-1})]^T \\ &\forall\ k = (n_a + 1) + (n_a + n_l) \times l, \\ &\forall\ l = 0, 1, 2, \cdots, \mathrm{int}(\frac{N}{n_a + n_l}) - (n_a + n_l + 1)\end{aligned} \tag{11}$$

where $\mathrm{int}(\cdot)$ refers to integer part (round towards zero) of its argument. In this way, each snippet Z_k is independent from other snippets and each of these has a mutltivariable Normal distribution whose covariance matrix is a function of the switching sequence. As a reminder, an n_l dimension multivariate Normal distribution has density function

$$f(x) = \frac{1}{(2\,\pi)^{n_l/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x)^T|\Sigma|^{-1}(x)\}$$

where Σ is the covariance matrix of dimension $n_l \times n_l$.

Note that, at each time k, there are n^{n_l} possible switching sequences for the snippet Z_k, since $\delta_k \in \{1, \cdots, n\}$. Therefore, if n_l is small enough, we can compute the likelihood value for each of the n^{n_l} choices and take the most likely sequence of subsystems as the one leading to the highest likelihood. Hence, given the independence of Z_k, estimating the most likely switching sequence in the used snippets can be done by solving the following problem

$$\begin{aligned}\max_{Z_k,\delta_k} \quad & \sum_k \log[f(Z_k)] \\ \text{s.t.} \quad & Z_k = [z_k(\delta_k), \cdots,\ z_{k+n_l-1}(\delta_{k+n_l-1})]^T \\ & \delta_k \in \{1, \cdots, n\} \\ & \forall\ k = (n_a + 1) + (n_a + n_l) \times l, \\ & \forall\ l = 0, 1, 2, \cdots, \mathrm{int}(\frac{N}{n_a + n_l}) - (n_a + n_l + 1)\end{aligned} \tag{12}$$

whose optimization can be done separately for each term of the sum. Therefore, its complexity is exponential in n_l but linear in the number of snippets and can be efficiently solved if n_l is "not too large." Note that this optimization problem is a convex quadratic one that can be solved efficiently in many ways.

Remark 1. As previously mentioned, in the above formulation, we do not use all available data when computing high likelihood switchings. More precisely, we only use $n_l/(n_l + n_a)$ of the data. Hence, any choice of n_l is a compromise between computational complexity and fraction of the data used and the "right" choice should be done by taking into account how many measurements are available.

Solving Problem (12), allows us to determine how many times a specific "jump" occurs in this high likelihood sequence of switches. This can be done in the following way:

Step 1. Solve problem (12). Recall that this can be done by solving the problem separately for each k.

Step 2. For each k, let $n_{ij}^{(k)}$ be the number of times the transition from system i to system j occurs in the maximum likelihood switch sequence for snippet k.

Step 3. Compute the total number of transitions from system i to system j observed in all snippets

$$n_{ij} = \sum_k n_{ij}^{(k)}$$

Given this high likelihood estimate of how often a transition occurs in the snippets, we can estimate the probability transition matrix. This can be done by computing the frequency of transitions as

$$P_{ij} = \frac{n_{ij}}{\sum_{j=1}^{n} n_{ij}} \qquad i = 1, \cdots, n \quad j = 1, \cdots, n \tag{13}$$

To be able to have convergence of the estimates, the following assumption is made

Assumption 2. Let m_k be the number of subsystems compatible with the information at time k when there is no measurement noise. More precisely

$$m_k = \text{cardinality} \left\{ \delta : x_k = \sum_{j=1}^{n_a} a_{j\delta}\, x_{k-j} + \sum_{j=1}^{n_c} c_{j\delta}\, u_{k-j} \right\}$$

and define

$$N_k = \text{cardinality} \left\{ l : 0 \leq l \leq k \text{ and } m_l > 1 \right\}.$$

Then

$$\lim_{k\to\infty} \frac{N_k}{k} = 0$$

The assumption above implies that the number of ambiguous transitions is much smaller than the number of measurements and these ambiguities will have little impact on the outcome of the identification algorithm for large number of measurements. With this assumption, as the number of observations goes to infinity, the solution of this problem converges to the true probability transition matrix. More precisely, we have the following result

Theorem 1. Let $\hat{P}_N$ represent the estimated transition probability matrix of N measurements of switched autoregressive processes, obtained from (13), and P_{true} be the true transition probability matrix. Assume that the Markov model for switching is aperiodic and Assumptions 1 and 2 hold. Then,

$$\lim_{N\to\infty} \lim_{\sigma^2\to 0} \; ||\hat{P}_N - P_{true}|| \; \to 0 \tag{14}$$

where σ^2 is the variance of the measurement noise.

Sketch of proof: We start by noting that under Assumption 2, the results in Hojjatinia et al. (2019) imply that as $\sigma^2 \to 0$ and for N sufficiently large, the estimate of the coefficients of the subsystems provided by the algorithm in Section 3 converges to the true ones.

Hence, and again using the fact that $\sigma^2 \to 0$, the optimization problem (12) identifies the true active system except for the times k where m_k (defined in Assumption 2) is strictly greater than one. Therefore, for most k, the right transitions are identified.

Finally, the assumption on aperiodicity of the underlying Markov chain implies that the probability of a transition occurring in a snippet is equal to the probability of occurring anywhere in the sequence of measurements. Therefore, and by the Law of Large Numbers Grimmett et al. (2001), the frequency of a transition occurring in the snippets converges to its probability. In other words, as the variance of the noise converges to zero, the estimated probability transition matrix of the switching sequence converges to the true one.

4.2 An Example

To better illustrate the approach, we provide an example of how to do the maximum likelihood estimation of probabilities required for identification of probability transition matrix. Therefore, consider the problem of identifying switching dynamics for SAR system with $n = 2$ subsystems of the form

$$\begin{array}{ll} \text{subsystem 1}: & x_k = a_1\, x_{k-1} + b_1\, u_{k-1} \\ \text{subsystem 2}: & x_k = a_2\, x_{k-1} + b_2\, u_{k-1} \end{array} \tag{15}$$

and noisy measurements

$$y_k = x_k + \eta_k \tag{16}$$

where η_k has zero mean Normal distribution, and $n_a = 1$. We consider snippets of data of length $n_l = 2$, and skip $n_a = 1$ sample measurement in between the snippets of data, i.e. two snippets of data for subsystem 1 are like:

$$\begin{cases} \eta_k - a_1\eta_{k-1} = y_k - (a_1 y_{k-1} + b_1 u_{k-1}) \\ \eta_{k+1} - a_1\eta_k = y_{k+1} - (a_1 y_k + b_1 u_k) \end{cases}$$

$$\begin{cases} \eta_{k+3} - a_1\eta_{k+2} = y_{k+3} - (a_1 y_{k+2} + b_1 u_{k+2}) \\ \eta_{k+4} - a_1\eta_{k+3} = y_{k+4} - (a_1 y_{k+3} + b_1 u_{k+3}) \end{cases}$$

and we have skipped this one:

$$\eta_{k+2} - a_1\eta_{k+1} = y_{k+2} - (a_1 y_{k+1} + b_1 u_{k+1})$$

For this example

$$\begin{aligned} Z_k =& [z_k(\delta_k),\, z_{k+1}(\delta_{k+1})]^T \\ & \forall\, k = 2 + 3 \times l, \\ & \forall\, l = 0, 1, 2, \cdots, \text{int}(\frac{N}{3}) - 4 \end{aligned}$$

and,

$$z_k(\delta_k) \in \left\{ z_k(\delta_k = 1),\, z_k(\delta_k = 2) \right\}$$

Therefore, the set of possible active sequences can be:

$$\begin{aligned} Z_k =& \begin{cases} [z_k(\delta_k = 1),\, z_{k+1}(\delta_{k+1} = 1)]^T \\ [z_k(\delta_k = 1),\, z_{k+1}(\delta_{k+1} = 2)]^T \\ [z_k(\delta_k = 2),\, z_{k+1}(\delta_{k+1} = 1)]^T \\ [z_k(\delta_k = 2),\, z_{k+1}(\delta_{k+1} = 2)]^T \end{cases} \\ & \forall\, k = 2 + 3 \times l, \\ & \forall\, l = 0, 1, 2, \cdots, \text{int}(\frac{N}{3}) - 4 \end{aligned} \tag{17}$$

At each time k, there are $n^{n_l} = 4$ possible Z_k cases. For system shown in equation (15), Z_k cases are as follows:

$$\begin{cases} \eta_k - a_1\eta_{k-1} = y_k - (a_1 y_{k-1} + b_1 u_{k-1}) \\ \eta_{k+1} - a_1\eta_k = y_{k+1} - (a_1 y_k + b_1 u_k) \end{cases}$$

$$\begin{cases} \eta_k - a_1\eta_{k-1} = y_k - (a_1 y_{k-1} + b_1 u_{k-1}) \\ \eta_{k+1} - a_2\eta_k = y_{k+1} - (a_2 y_k + b_2 u_k) \end{cases}$$

$$\begin{cases} \eta_k - a_2\eta_{k-1} = y_k - (a_2 y_{k-1} + b_2 u_{k-1}) \\ \eta_{k+1} - a_1\eta_k = y_{k+1} - (a_1 y_k + b_1 u_k) \end{cases}$$

$$\begin{cases} \eta_k - a_2\eta_{k-1} = y_k - (a_2 y_{k-1} + b_2 u_{k-1}) \\ \eta_{k+1} - a_2\eta_k = y_{k+1} - (a_2 y_k + b_2 u_k) \end{cases}$$

The value of probability density function for the Multivariate Normal distribution at each of n^{n_l} sequences Z_k will be computed, and the one which has the maximum value of the likelihood will be considered as the set of active subsystems at that point.

5. NUMERICAL RESULTS

In this section, we will address the problem of identifying switching dynamics in Markovian jump systems. The values of true coefficients in this example has taken from the example in Hojjatinia et al. (2019), which are $a_1 = 0.3$, $b_1 = 1$, $a_2 = -0.5$, and $b_2 = -1$. Measurement noise is assumed to have zero-mean Normal distribution. So, AR system with $n = 2$ subsystems, $n_a = 1$, and $n_c = 1$ in this example are as follows:

$$\begin{aligned} \text{subsystem 1}: \quad & x_k = 0.3\, x_{k-1} + 1\, u_{k-1} \\ \text{subsystem 2}: \quad & x_k = -0.5\, x_{k-1} - 1\, u_{k-1} \\ \text{noisy output}: \quad & y_k = x_k + \eta_k \end{aligned} \tag{18}$$

Total number of $N = 10^6$ input-output data is available. Output is corrupted with random measurement noise which is Normal with zero mean and different values of variance. The proposed algorithm is coded and run in Python.

Noise to output ratio (γ) is defined as

$$\gamma = \frac{\max\, |\eta|}{\max\, |y|} \tag{19}$$

Simulation results for several experiments are shown in Table 1. For each experiment, a random probability transition matrix has been generated, which is shown in column 2 of the table. By using the algorithms mentioned in the paper, for each experiment probability transition matrix has been estimated from noisy measurements, which is shown in column 3 of the table. The normalized Frobenius norm between true and estimated values of probability transition matrix has been computed and shown in column 4 of the table ($\left\|\hat{P} - P_{true}\right\|_F / \|P_{true}\|_F$).

For each experiment noise to output ratio and variance of noise are shown in columns 5 and 6 of the table. As we see in this table, the value of entries in probability transition matrix are very close to the true values, even when the noise variance is high with noise magnitude in average around 30% of the signal magnitude.

For example, in experiment 6 the value of $\gamma = 0.5223$ shows that noise to output ratio of approximately 52%; even with this very large value of corruption with measurement noise, the proposed method works well and the normalized Frobenius norm between true and estimated values of probability transition matrix is only 0.1153. As expected and in the Table 1 is shown, for smaller values of noise to output ratio (γ), the estimated values for probability transition matrix are closer to the true probability transition matrix and normalized Frobenius norm of their difference has smaller values. However, with the prposed approach even for large values of noise to output ratio, the difference in Frobenius norm is still low.

Figure 1 demonstrates the convergence of probability transition matrix as number of measurements grows. This figure is based on a random experiment, for the values of coefficients in equation (18), fixed variance of noise $\sigma^2 = 0.03$ and noise to output ratio $\gamma = 0.15$. As we see in this figure the values of normalized Frobenius norm between true and estimated probability transition matrix decreases, when number of measurement increases. As we observe in Figure 1, the value of difference between true and estimated probability transition matrix decreases from 0.2419 at $k = 100$ to 0.0659 at $k = 10^6$. It shows even for the case of having 15% noise to output ratio, the approximated switching dynamics and transition probability matrix are close to the true ones.

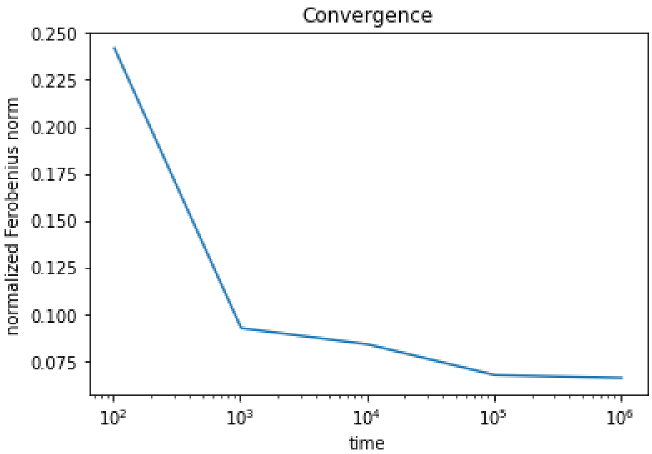

Fig. 1. Convergence of estimated probability transition matrix

6. CONCLUSION AND FUTURE WORK

In this paper we have proposed a methodology for identification of switching dynamics in Markovian jump SAR models. Given large noisy input-output data, by using previously developed procedures for identification of switched system from large noisy data sets, we estimate the parameters of the noise, and then, identify the coefficients of each submodel. Then, by using the novel procedure presented in this paper for estimation of probability transition matrix, we identify the switching dynamics and computed the probability transition matrix of Markov chain. Even for large values of measurement noise, numerical simulations show a low estimation error. The Frobenius norm between estimated and true probability transition matrix is small even in the case of large noise to output ratio. For future work, we can consider the problem of identifying switched ARX models and switching dynamics form large noisy data sets, but with the process noise. We will also test the effectiveness of the proposed approaches in "real" applications with emphasis on estimating individual response to treatments aimed at improving light physical activity.

REFERENCES

Ansaripour, A., Mata, A., Nourazari, S., Kumin, H., et al. (2016). Some explicit results for the distribution problem of stochastic linear programming. *Open Journal of Optimization*, 5(04), 140.

Bako, L. (2011). Identification of switched linear systems via sparse optimization. *Automatica*, 47(4), 668–677.

Conroy, D.E., Hojjatinia, S., Lagoa, C.M., Yang, C.H., Lanza, S.T., and Smyth, J.M. (2019). Personalized models of physical activity responses to text message micro-interventions: A proof-of-concept application of control systems engineering methods. *Psychology of Sport and Exercise*, 41, 172–180.

Table 1. Identifying probability transition matrix (PTM) for different value of noise variance, different system run and $N = 10^6$ measurements.

experiment #	**Value 1** True PTM	**Value 2** Estimated PTM	**Value 3** Normalized Ferobenius norm	**Value 4** γ	**Value 5** σ^2
1	$\begin{pmatrix} 0.1837 & 0.8163 \\ 0.3424 & 0.6576 \end{pmatrix}$	$\begin{pmatrix} 0.2116 & 0.7884 \\ 0.3472 & 0.6528 \end{pmatrix}$	0.035810	0.0888	0.01
2	$\begin{pmatrix} 0.4286 & 0.5714 \\ 0.1412 & 0.8588 \end{pmatrix}$	$\begin{pmatrix} 0.3897 & 0.6103 \\ 0.1776 & 0.8224 \end{pmatrix}$	0.066922	0.1439	0.03
3	$\begin{pmatrix} 0.1748 & 0.8252 \\ 0.4921 & 0.5079 \end{pmatrix}$	$\begin{pmatrix} 0.2439 & 0.7561 \\ 0.4802 & 0.5198 \end{pmatrix}$	0.090075	0.1967	0.05
4	$\begin{pmatrix} 0.5056 & 0.4944 \\ 0.6885 & 0.3115 \end{pmatrix}$	$\begin{pmatrix} 0.5087 & 0.4913 \\ 0.6412 & 0.3588 \end{pmatrix}$	0.064687	0.2273	0.07
5	$\begin{pmatrix} 0.2587 & 0.7413 \\ 0.4536 & 0.5464 \end{pmatrix}$	$\begin{pmatrix} 0.3200 & 0.6800 \\ 0.4474 & 0.5526 \end{pmatrix}$	0.082236	0.2650	0.09
6	$\begin{pmatrix} 0.3991 & 0.6009 \\ 0.1811 & 0.8189 \end{pmatrix}$	$\begin{pmatrix} 0.3661 & 0.6339 \\ 0.2651 & 0.7349 \end{pmatrix}$	0.115335	0.5223	0.27
7	$\begin{pmatrix} 0.5350 & 0.4650 \\ 0.6467 & 0.3533 \end{pmatrix}$	$\begin{pmatrix} 0.5180 & 0.4819 \\ 0.5788 & 0.4212 \end{pmatrix}$	0.096751	0.4603	0.29

Ferrari-Trecate, G., Muselli, M., Liberati, D., and Morari, M. (2003). A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2), 205–217.

Grimmett, G., Grimmett, G.R., Stirzaker, D., et al. (2001). *Probability and random processes.* Oxford university press.

Harris, J. (2013). *Algebraic geometry: a first course*, volume 133. Springer Science & Business Media.

Hojjatinia, S., Aliyari Shoorehdeli, M., Fatahi, Z., Hojjatinia, Z., and Haghparast, A. (2020a). Improving the izhikevich model based on rat basolateral amygdala and hippocampus neurons, and recognizing their possible firing patterns. *Basic and Clinical Neuroscience*, 11(1), 79–90.

Hojjatinia, S. and Lagoa, C.M. (2019). Comparison of different spike sorting subtechniques based on rat brain basolateral amygdala neuronal activity. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2251–2258.

Hojjatinia, S., Lagoa, C.M., and Dabbene, F. (2017). A method for identification of markovian jump arx processes. *IFAC-PapersOnLine*, 50(1), 14088 – 14093. 20th IFAC World Congress.

Hojjatinia, S., Lagoa, C.M., and Dabbene, F. (2019). Identification of switched autoregressive systems from large noisy data sets. In *2019 American Control Conference (ACC)*, 4313–4319. IEEE.

Hojjatinia, S., Lagoa, C.M., and Dabbene, F. (2020b). Identification of switched autoregressive exogenous systems from large noisy datasets. *International Journal of Robust and Nonlinear Control*.

Hojjatinia, S., Lagoa, C.M., West, A.B., Hojjatinia, S., Turrisi, T., Brunke-Reese, D., and Conroy, D.E. (2020c). Abstract p196: Personalized dynamical system models of individual text message effects on changes in physical activity. *Circulation*, 141(Suppl_1), AP196–AP196.

Juloski, A.L., Weiland, S., and Heemels, W. (2005). A bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control*, 50(10), 1520–1533.

Lagoa, C.M., Conroy, D.E., Hojjatinia, S., and Yang, C.H. (2017). Modeling subject response to interventions aimed at increasing physical activity: A control systems approach. *Poster Presented at 5th International Conference on Ambulatory Monitoring of Physical Activity and Movement.*

Nakada, H., Takaba, K., and Katayama, T. (2005). Identification of piecewise affine systems based on statistical clustering technique. *Automatica*, 41(5), 905–913.

Ozay, N., Lagoa, C., and Sznaier, M. (2015). Set membership identification of switched linear systems with known number of subsystems. *Automatica*, 51, 180 – 191.

Ozay, N., Sznaier, M., Lagoa, C.M., and Camps, O.I. (2012). A sparsification approach to set membership identification of switched affine systems. *IEEE Transactions on Automatic Control*, 57(3), 634–648.

Paoletti, S., Juloski, A.L., Ferrari-Trecate, G., and Vidal, R. (2007). Identification of hybrid systems a tutorial. *European journal of control*, 13(2-3), 242–260.

Roll, J., Bemporad, A., and Ljung, L. (2004). Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1), 37–50.

Shi, P. and Li, F. (2015). A survey on markovian jump systems: modeling and design. *International Journal of Control, Automation and Systems*, 13(1), 1–16.

Vidal, R., Soatto, S., Ma, Y., and Sastry, S. (2003). An algebraic geometric approach to the identification of a class of linear hybrid systems. In *Proc. 42nd IEEE Conf. Decision and Control*, volume 1, 167–172.

Vidal, R., Ma, Y., and Sastry, S. (2005). Generalized principal component analysis (gpca). *IEEE transactions on pattern analysis and machine intelligence*, 27(12), 1945–1959.

Vidal, R.E. and Sastry, S. (2003). *Generalized principal component analysis (gpca): an algebraic geometric approach to subspace clustering and motion segmentation.* Electronics Research Laboratory, College of Engineering, University of California.