Automatica 47 (2011) 668–677

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Identification of switched linear systems via sparse optimization☆

Laurent Bako*

Univ Lille Nord de France, F-59000 Lille, France
EMDouai, IA, F-59500 Douai, France

ARTICLE INFO

Article history:
Available online 21 February 2011

Keywords:
Switched linear systems
System identification
Sparse optimization
Hybrid systems

ABSTRACT

The work presented in this paper is concerned with the identification of switched linear systems from input-output data. The main challenge with this problem is that the data are available only as a mixture of observations generated by a finite set of different interacting linear subsystems so that one does not know a priori which subsystem has generated which data. To overcome this difficulty, we present here a sparse optimization approach inspired by very recent developments from the community of compressed sensing. We formally pose the problem of identifying each submodel as a combinatorial ℓ_0 optimization problem. This is indeed an NP-hard problem which can interestingly, as shown by the recent literature, be relaxed into a (convex) ℓ_1-norm minimization problem. We present sufficient conditions for this relaxation to be exact. The whole identification procedure allows us to extract the parameter vectors (associated with the different subsystems) one after another without any prior clustering of the data according to their respective generating-submodels. Some simulation results are included to support the potentialities of the proposed method.

1. Introduction

We consider in this paper the problem of identifying a switched linear system from a collection of input–output data. A switched linear system corresponds to the behavior that results from the interaction (mainly switching) between a finite set of linear dynamical subsystems. Examples of such systems arise in many different engineering fields, typically genetic regulatory networks study, air traffic management, nonlinear systems control, manufacturing processes modeling, computer vision, etc. (see e.g., Liberzon, 2003; Sun & Ge, 2005, for more examples). Mathematically speaking, a switched linear system can be viewed as a relation (model) of the form

$$y(t) = \theta_{\lambda_t}^\top x(t) + e(t), \tag{1}$$

relating a vector $x(t) \in \mathbb{R}^n$ called the regressor vector and a signal $y(t) \in \mathbb{R}$ designated as the output of the system being modeled at time t. Here, $\lambda_t \in \{1, \ldots, s\}$ is the discrete state or the discrete mode, i.e., the index of the active submodel at time t and $\theta_{\lambda_t} \in \mathbb{R}^n$ is the associated parameter vector (PV). The sequence $\{e(t)\}$ of errors refers to potential mismatch or noise; it is assumed to be bounded.

In a general situation, the vector $x(t)$ appearing in (1) need not be structured but when dealing with the input–output behavior of switched dynamical systems, it sometimes takes the form

$$x(t) = \left[y(t-1) \ \cdots \ y(t-n_a) \ u(t-1)^\top \ \cdots \ u(t-n_b)^\top\right]^\top, \tag{2}$$

where $u(t) \in \mathbb{R}^{n_u}$ and $y(t) \in \mathbb{R}$ are respectively the input and output of the considered system, n_a and n_b are its orders. The dimension of $x(t)$ is therefore $n = n_a + n_b n_u$ and the model (1) is designated as a Switched Auto-Regressive eXogenous (SARX) model.

1.1. The switched system identification problem

Given observations $\{x(t), y(t)\}_{t=1}^N$ generated by a switched linear model of the form (1), with $x(t)$ defined as in (2), we are interested here in estimating the parameter vectors $\{\theta_j\}_{j=1}^s$.

We start by recalling from Vidal, Chiuso, and Soatto (2002) that the problem of inferring a switched model such as (1) from a set of finite measurements, admits multiple solutions so that the identification problem is not well-posed. If the structural indices n_a and n_b are not fixed, then one can find for example a trivial switched linear model consisting of one single submodel with large orders that fits all the finite dataset. Even if finite and fixed values are assigned to n_a and n_b, there are still infinitely many switched models that explain the data. For example, it can be simply verified that there is a switched linear model with $s = N$ submodels that can reproduce the data. In order to remove the identifiability issue, we will assume in this paper that the orders n_a and n_b are finite, equal for

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Antonio Vicino under the direction of Editor Torsten Söderström.

* Corresponding address: Univ Lille Nord de France, F-59000 Lille, France. Tel.: +33 327 712 127; fax: +33 327 712 917.
E-mail address: laurent.bako@mines-douai.fr.

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.01.036

all submodels and known a priori. With this setting for the structural indices n_a and n_b, the SARX of interest here will be viewed as the one that, among all switched linear models consistent with the data, has a number of submodels that is as small as possible. The interested reader is referred to the paper Petreczky, Bako, and van Schuppen (2010) for a more complete treatment of the identifiability problem in the framework of switched linear state space models.

1.2. Prior work

During the last ten years a number of interesting results have been achieved in the field of hybrid system identification. Examples of such works include, in the case of switched linear models, the algebraic–geometric method (Ma & Vidal, 2005; Vidal, 2008; Vidal, Soatto, Ma, & Sastry, 2003), the product-of-errors based method (Lauer, Vidal, & Bloch, 2009). Other methods such as the mixed integer programming approach (Roll, Bemporad, & Ljung, 2004), the bounded-error approach (Bemporad, Garulli, Paoletti, & Vicino, 2005), the Bayesian learning based procedure (Juloski, Weiland, & Heemels, 2005), the clustering-based strategies (Boukharouba, Bako, & Lecoeuche, 2009; Ferrari-Trecate, Muselli, Liberati, & Morari, 2003; Ferrari-Trecate, Schinkel, & Pnueli, 2003; Nakada, Takaba, & Katayama, 2005) apply to piecewise affine systems, i.e., particular switched linear/affine systems where the switching surfaces are the faces of a set of non-overlapping polyhedra. An excellent survey can be found in Paoletti, Juloski, Ferrari-Trecate, and Vidal (2007) where most of the methods developed prior to 2007 have been summarized. Despite the clear merits of all these pioneering contributions, one can fairly observe that the subject of hybrid system identification is still open on many challenging issues such as computational complexity reduction, optimality and convergence analysis of the proposed methods. Recently, a promising idea has emerged as to what extent some results from sparse optimization based signal recovery can be applied to hybrid system identification. This idea may indeed be an expedient for tackling simultaneously both the clustering and estimation problems that are inherent to hybrid system computation. The work of Ozay, Sznaier, Lagoa, and Camps (2008) exploits successfully such an approach. The identification of the parameter vectors is formulated as the problem of recovering a sparse vector-valued sequence, the instances of which sequence are subsequently agglomerated to reach a minimum number of submodels. The work of Elhamifar and Vidal (2009) also suggests sparse representation as a possible alternative for solving the problem of subspace clustering. More precisely, the authors of Elhamifar and Vidal (2009) consider the problem of estimating bases for a set of linear/affine subspaces from data lying in the union of these subspaces. A limitation of their work however is that the mixed subspaces need to be linearly independent, an assumption which is violated when dealing with the union of more than one hyperplanes.

1.3. Contributions of this paper

The contribution of the paper consists in the development of a new identification method for switched linear systems. Data vectors generated by such systems lie in the union of a finite set of linear hyperplanes. Therefore we pose the identification of a specific submodel as the problem of extracting the hyperplane that contains the largest number of data. The corresponding submodel is hence the one that, among all submodels, achieves, over the whole dataset, the sparsest vector of fitting errors. With this formulation, one submodel can be estimated directly without any prior clustering, by means of sparse optimization, i.e., the minimization of the number of nonzero components in an error vector. Since sparse optimization is in general non-convex, it is classical to consider instead a convex ℓ_1 relaxation of this problem. We then present sufficient conditions under which the ℓ_1 relaxation is guaranteed to recover exactly the solution of the initial sparse optimization problem. In the case when these conditions are not satisfied, we show that all the PVs can still be identified by slightly adapting an iterative reweighted ℓ_1 optimization technique proposed in Candès, Wakin, and Boyd (2008). In contrast to most of the existing methods for hybrid system identification, our method lends itself to a relatively easy analysis. For example, conditions for optimality even though somewhat conservative, can be derived. A number of results from the field of compressed sensing (Candès & Randall, 2006; Candès, Rudelson, Tao, & Vershynin, 2005; Donoho, 2006) can be insightful for this purpose.

1.4. Outline of this paper

The remainder of the paper is organized as follows. We start by presenting in Section 2 the main mathematical terminology used in the paper. We then describe in Section 3 the proposed algorithm for the identification of switched linear systems with arbitrary switchings. Section 4 contains some numerical results that confirm the potential of our method. Section 5 concludes the paper.

2. Mathematical preliminaries

In this preliminary section we introduce some mathematical concepts and notations that will be extensively used throughout the paper. We first introduce a notion of *k-genericity index*.

Definition 1. For a given data matrix $X = \begin{bmatrix} x(1) & \dots & x(N) \end{bmatrix} \in \mathbb{R}^{n\times N}$ with $n \leq N$, and for any integer k verifying $0 < k \leq \text{rank}(X)$, we define the k-genericity index $\nu_k(X)$ of X to be the minimum integer m such that any $n \times m$ submatrix of X has rank k:

$$\nu_k(X) = \min\big\{m : \forall(t_1, \dots, t_m) \text{ with } t_i \neq t_j \text{ for } i \neq j, \quad \text{rank}\big[x(t_1) \cdots x(t_m)\big] = k\big\}. \tag{3}$$

If $k > \text{rank}(X)$, we set by convention $\nu_k(X) = +\infty$ and if $k = 0$ we set $\nu_0(X) = 0$ for all X.

For an overview on the function $\nu_k(\cdot)$, we quickly mention the following two obvious properties.

1. If $\nu_k(X) = k$ then $\nu_p(X) = p$ for all $p \leq k$.
2. For any $k \leq \text{rank}(X)$, it holds that $k \leq \nu_k(X) \leq N$.

Observe additionally that when the data $\{x(t)\}_{t=1}^N$ are in *general position*, i.e., when any subset $\{x(t_1), \dots, x(t_n)\}$ of n data vectors are linearly independent, we have $\nu_n(X) = n$. Hence, the number $\nu_n(X)$ characterizes a property of richness (or linear independence) of the columns of X. For $\nu_n(X)$ to be finite, we need to assume that $\text{rank}(X) = n$. It will be so in all the paper. Other possible indices for measuring linear independence between the columns of a matrix are the so-called *spark* and *mutual coherence* whose formal definitions are recalled from Bruckstein, Donoho, and Elad (2009) and Donoho and Elad (2003).

Definition 2. The spark of a given matrix X denoted $\text{spark}(X)$, is the smallest number σ such that there exists a set of σ columns of X that are linearly dependent. In fact we have

$$\text{spark}(X) = \min_{\substack{z\in\ker(X)\\ z\neq 0}} \|z\|_0 \tag{4}$$

where $\ker(X)$ refers to the kernel subspace of matrix X and $\|z\|_0$ stands for the number of nonzero entries in z. If $\ker(X) = \{0\}$, then $\text{spark}(X)$ will be conventionally set to $+\infty$.