Automatica 70 (2016) 21–31

Contents lists available at ScienceDirect

Automatica

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

A new kernel-based approach to hybrid system identification☆

Gianluigi Pillonetto [1]

Department of Information Engineering, University of Padova, Padova, Italy

ARTICLE INFO

Article history:
Received 5 March 2015
Received in revised form
1 September 2015
Accepted 22 February 2016
Available online 13 April 2016

Keywords:
Piecewise affine systems
Model order selection
Kernel-based regularization
Gaussian processes
Marginal likelihood
Markov chain Monte Carlo
Nonlinear system identification
Bias–variance trade off

ABSTRACT

All the approaches for hybrid system identification appeared in the literature assume that model complexity is known. Popular models are e.g. piecewise ARX with a priori fixed orders. Furthermore, the developed numerical procedures have been tested only on simple systems, e.g. composed of ARX subsystems of order 1 or at most 2. This represents a major drawback for real applications. This paper proposes a new regularized technique for identification of piecewise affine systems, namely the hybrid stable spline algorithm (HSS). HSS exploits the recently introduced stable spline kernel to model the submodels impulse responses as zero-mean Gaussian processes, including information on submodels predictor stability. The algorithm consists of a two-step procedure. First, exploiting the Bayesian interpretation of regularization, the problem of classifying and distributing the data to the subsystems is cast as marginal likelihood optimization. We show how an approximated optimization can be efficiently performed by a Markov chain Monte Carlo scheme. Then, the stable spline algorithm is used to reconstruct each subsystem. Numerical experiments on real and simulated data are included to test the new procedure. They show that HSS not only solves all the most popular benchmark problems proposed in the literature without having exact information on ARX subsystems order, but can also identify more complex (high-order) piecewise affine systems. MATLAB code implementing the approach, called Hybrid Stable Spline Toolbox, is also made available.

1. Introduction

Hybrid (switched) systems have been subject of much research in the last years. Their importance stems from the capability of describing in an unified setting several processes evolving through continuous/discrete dynamics and logic rules (Bemporad, Ferrari-Trecate, & Morari, 2000), permitting e.g. to represent linear complementarity systems (Heemels, De Schutter, & Bemporad, 2001) as well as interactions between affine systems and finite automata (Sontag, 1996). Hybrid systems can be used also to approximate (with arbitrary accuracy) nonlinear dynamics by linearization around different working points, see Breiman (1993) and Lin and Unbehauen (1992) for universal approximation properties. Examples and applications of these models can be found in many different fields, including e.g. model predictive/nonlinear systems control, state estimation, computer vision, air traffic management (Bemporad & Morari, 1999; Liberzon, 2003; Paoletti, Juloski, Ferrari-Trecate, & Vidal, 2007).

This paper in particular deals with the identification of a discrete-time hybrid system composed by s affine submodels, each defined by a (column) vector θ_k. A discrete state variable x_t evolves over time t and selects the kth submodel if $x_t = k$. For $t = 1, \ldots, N$, the measurements model is

$$y_t = \rho_t^\top \theta_k + e_t \quad \text{for } x_t = k \tag{1.1}$$

where y_t is the system output corrupted by a zero-mean white Gaussian noise e_t of variance σ^2, i.e. $e_t \sim \mathcal{N}(0, \sigma^2)$, while ρ_t is an observable regression (column) vector. In particular, as a concrete and important example, hereby we assume

$$\rho_t = [1\, y_{t-1} \ \ldots \ y_{t-m}\, u_{t-1} \ \ldots \ u_{t-m}]^\top \tag{1.2}$$

where u_t is the system input measured at instant t and m is the system order/memory. With this definition, the first component of θ_k defines the offset, while the other $2m$ contain two "impulse responses". Examples built using (1.2) are in Table 1 which reports

☆ This research has been partially supported by the MIUR FIRB project RBFR12M3AC-Learning meets time: a new computational approach to learning in dynamic systems and by the Progetto di Ateneo CPDA147754/14-New statistical learning approach for multi-agents adaptive estimation and coverage control. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Er-Wei Bai under the direction of Editor Torsten Söderström. The author would like to thank Prof. Henrik Ohlsson for providing the real data coming from a pick-and-place machine.

E-mail address: giapi@dei.unipd.it.

[1] Tel.: +390498277607.

http://dx.doi.org/10.1016/j.automatica.2016.03.011
0005-1098/

six popular hybrid systems taken from the literature. Starting from the measurements $\{u_t, y_t\}_{t=1}^N$, our problem is to reconstruct the s vectors θ_k.

Switched (also called segmented) models arise e.g. when the state variable follows a deterministic (possibly periodic) trajectory, e.g. see System 1 in Table 1, or when the x_t are modeled as random variables independent of the regressors ρ_t, as in System 2 of Table 1 ($x \sim \mathscr{U}_A$ indicates a random variable x uniform on the set A). The opposite situation is found when the regressor space $\mathscr{X}$ is partitioned into s subsets $\mathscr{X}_k$ and the switching rule becomes $x_t = k \iff \rho_t \in \mathscr{X}_k$. This leads to the popular piecewise affine (PWA) models which, under (1.2), specialize to the important subclass of the piecewise auto-regressive with exogenous input (PWARX) models. Examples are Systems 3–6 contained in Table 1. Combining the condition $x_t = \rho_t$ and (1.1), (1.2), one can see that a PWARX model determines the active model only on the basis of the last m input–output samples. A variant is obtained neglecting the autoregressive part, i.e. $\rho_t = [1\ u_{t-1}\ \dots\ u_{t-m}]^\top$, thus leading to PWFIR models.

The difficulty of hybrid system identification is the need of jointly classifying the data (assigning each regressor to the submodel more likely to be active) and estimating the system parameters. Furthermore, the input–output hybrid map can be discontinuous along the boundaries of the submodels regions. This encumbers the use of standard kernel-based approaches, e.g. support vector regression and regularization/neural networks (Evgeniou, Pontil, & Poggio, 2000; Fausett, 1994; Schölkopf & Smola, 2001) which postulate function smoothness.

The approach in Ferrari-Trecate, Muselli, Liberati, and Morari (2003) faces these difficulties combining clustering, linear identification and pattern recognition technique. In particular, the algorithm is based on the assumption that regressors close each other likely belong to the same ARX submodel. In Roll, Bemporad, and Ljung (2004), mixed-integer linear and quadratic programming is proposed to identify two subclasses of PWARX models. The approach in Bemporad, Garulli, Paoletti, and Vicino (2005) is instead inspired by set-membership identification techniques (Milanese & Vicino, 1991). The identification error is assumed to be bounded by a known quantity, and then the search for a minimum number of feasible subsystems is performed. This problem is however NP-hard and a suboptimal algorithm is proposed based on thermal relaxations. A Bayesian framework is introduced in Juloski, Weiland, and Heemels (2005). Here, the θ_k are random vectors and classification corresponds to extracting data with highest a posteriori probability. This step is performed by designing an approximated Bayes estimator implemented by particle filters (Andrieu, Doucet, & Holenstein, 2010).

Hybrid system identification is faced in an algebraic fashion in Vidal, Chiuso, and Soatto (2002): exploiting polynomial factorization and hyperplane clustering an exact solution is obtained but only in the noiseless case. While a recursive estimation scheme is described in Vidal (2008), more recent approaches rely on convex relaxation and sparse optimization. In particular, in Ohlsson and Ljung (2013) the problem's combinatorial nature is tackled by first introducing an overparametrized model. Then, the submodels parameters are estimated by least squares regularized via a sum-of-norms penalty. A regularization parameter is introduced to balance adherence to experimental data and number of submodels. In Bako (2011), identification is instead performed by solving a sequence of (non regularized) problems defined by weighted (and reweighted) ℓ_1 losses. An analysis of the algorithm is also obtained under noiseless assumptions.

It is worth noticing that all of the aforementioned approaches to hybrid system identification assume known the order m of the ARX submodels. In addition, all the proposed algorithms have been tested only on quite simple hybrid systems (e.g. in Table 1 one has $m = 2$, at most). This appears an important drawback for real applications where systems can be more complex and m is typically unknown. This is a central issue in system identification: it is crucial to find a suitable model structure with the right model complexity yielding a good bias–variance tradeoff (Ljung, 1999; Söderström & Stoica, 1989). In light of this, the aim of this paper is to design a new regularized technique which determines from data also submodels complexity. This will be achieved by extending the *stable spline estimator* proposed in Pillonetto, Chiuso, and De Nicolao (2011) and Pillonetto and De Nicolao (2010) (and further discussed in Chen, Ohlsson, & Ljung, 2012 and Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014). We interpret hybrid system identification as a functional estimation problem, facing its ill-posedness/ill-conditioning in a Bayesian framework (Rasmussen & Williams, 2006). In particular, submodels impulse responses are modeled as zero-mean Gaussian processes with autocovariances equal to the *stable spline kernel*. In this way, information on the exponential stability of the predictor of each isolated subsystem is included in the estimation process.

The stable spline estimator for linear system identification depends on two (unknown) hyperparameters: the scale factor λ and the stability parameter α which regulates how fast the impulse response decays to zero. In comparison with classical parametric approaches, one important feature of this estimator is that the difficult model order selection can be replaced by hyperparameters estimation. In particular, in Chen et al. (2012) and Pillonetto and De Nicolao (2010) λ and α are estimated optimizing the marginal likelihood (ML), i.e. the marginal density of measurements obtained after integrating out the dependence on the impulse response (MacKay, 1992). This operation is also known as Empirical Bayes (Maritz & Lwin, 1989). Several merits of ML are documented in the literature, e.g. the fact that it automatically includes the Occam's razor (MacKay, 1992). Recent studies have also clarified why ML may work well also in presence of deviations from the stochastic model, i.e. when undermodeling affects the kernel-based impulse response description (Pillonetto & Chiuso, 2014, 2015).

However, differently from the linear scenario, in hybrid system identification other N unknown variables have to be considered: the state variables x_t which indicate which submodel is active at every instant t. The main idea explored in this paper is to consider these classification variables as further hyperparameters which can be estimated via ML optimization. Due to its combinatorial nature, this problem would seem unfeasible. We will instead show how an approximated optimization can be efficiently performed through a Markov chain Monte Carlo (MCMC) approach (Gilks, Richardson, & Spiegelhalter, 1996). Our scheme is completely automatic: it generates a Markov chain exploring the ML without the need of specifying any proposal density or tuning parameter. Experimental results show that running few and short Markov chains can already lead to very accurate classifications. Then, once the x_t are determined via ML optimization, s stable spline estimators are used to reconstruct the submodels.

The paper is so organized. In Section 2, we introduce the stable spline model for hybrid systems adopted to classify and distribute data to the submodels. Section 3 then describes how the classification problem is solved by HSS via ML optimization. In particular, an MCMC scheme to efficiently explore the support of ML is designed. In Section 4 the algorithm's description is completed showing how the submodels are reconstructed by HSS once the estimates of x_t are available. Section 5 introduces some indexes related to classification and impulse responses reconstruction. We also present two oracle-based procedures, and related indices, which permit to define useful performance references to assess the effectiveness of a hybrid system identification procedure. Section 6 reports some numerical