Automatica 85 (2017) 234–247

Contents lists available at ScienceDirect

Automatica

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

A nonparametric kernel-based approach to Hammerstein system identification ☆

Riccardo Sven Risuleo, Giulio Bottegal, Håkan Hjalmarsson

ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 18 September 2016
Received in revised form 22 June 2017
Accepted 30 June 2017

Keywords:
System identification
Hammerstein systems
Nonlinear systems
Kernel-based methods
Gaussian processes

ABSTRACT

Hammerstein systems are the series composition of a static nonlinear function and a linear dynamic system. In this work, we propose a nonparametric method for the identification of Hammerstein systems. We adopt a kernel-based approach to model the two components of the system. In particular, we model the nonlinear function and the impulse response of the linear block as Gaussian processes with suitable kernels. The kernels can be chosen to encode prior information about the nonlinear function and the system. Following the empirical Bayes approach, we estimate the posterior mean of the impulse response using estimates of the nonlinear function, of the hyperparameters, and of the noise variance. These estimates are found by maximizing the marginal likelihood of the data. This maximization problem is solved using an iterative scheme based on the expectation-conditional maximization, which is a variation of the standard expectation–maximization method for solving maximum-likelihood problems. We show the effectiveness of the proposed identification scheme in some simulation experiments.

1. Introduction

Nonlinear system identification is challenging because it is difficult to choose a general model structure to represent data from a nonlinear system (Sjöberg, Zhang, Ljung, Benveniste, Delyon, Glorennec, Hjalmarsson, & Juditsky, 1995). One way to reduce the number of candidate models is to use block-oriented models; in this way, we can use model structures that are applicable to a wide array of problems and for which effective estimation techniques exist (Giri & Bai, 2010). Among the block-oriented models is the Hammerstein structure (HS). It is a nonlinear cascaded model where a linear time-invariant (LTI) dynamical model follows a static nonlinear mapping (see Ljung, 1999).

The HS is capable of modeling a wide range of processes and has therefore been object of numerous studies (Bai, Cai, Dudley-Javorosk, & Shields, 2009; Hunter & Korenberg, 1986; Westwick & Kearney, 2001). Over the years, several identification methods have been proposed. In the following, we characterize the main approaches. In *overparameterization methods*, the problem of identifying the unknown parameters is embedded in the larger problem of identifying a vector containing all the cross products of the parameters (Bai, 1998). This turns the problem into a linear identification problem, to which linear techniques can be applied (for instance, least squares in Bai, 1998, instrumental variables in Han and De Callafon, 2011). Once the overparameterized vector has been identified, the solution of the original problem is recovered by means of some reduction step (for instance, minimum norm in Bai, 1998 and Han and De Callafon, 2011, consistent estimation in Boutayeb, Aubry, and Darouach, 1996). These methods hinge around a rank-one constraint, which is difficult to enforce. Attempts have been made using regularization techniques (see, for instance, Falck, Suykens, Schoukens, and De Moor, 2010 and Risuleo, Bottegal, and Hjalmarsson, 2015b). *Subspace methods* have been extended to HSs. The MOESP subspace method has been extended to HSs by assuming a polynomial model for the static nonlinearity (Verhaegen & Westwick, 1996). N4SID has been adapted to HSs by using support-vector machines (Goethals, Pelckmans, Suykens, & De Moor, 2005). In *Separable Least Squares* methods, the unknowns are divided into two sets and the variables in one set are expressed as functions of the variables in the other set. This reduces the dimensionality of the problem (Golub & Pereyra, 2003; Han & De Callafon, 2012; Westwick & Kearney, 2001). Similarly, in *iterative methods*, the variables are split into two sets, and the problem is solved by alternating the optimization between the two sets of variables (see, for instance, Bai & Li, 2004; Liu & Bai, 2007). In *stochastic methods* the linearity is identified irrespective of the nonlinear transformation using correlation analysis (Billings

☆ This work was supported by the Swedish Research Council via the projects NewLEADS (contract number: 2016-06079) and System identification: Unleashing the algorithms (contract number: 2015-05285). The material in this paper was partially presented at the 17th IFAC Symposium on System Identification (SYSID), October 19-21, 2015, Beijing, China. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Torsten Söderström.

E-mail addresses: risuleo@kth.se (R.S. Risuleo), bottegal@kth.se (G. Bottegal), hjalmars@kth.se (H. Hjalmarsson).

http://dx.doi.org/10.1016/j.automatica.2017.07.055

& Fakhouri, 1978; Greblicki, 2000; Rangan, Wolodkin, & Poolla, 1995). These methods typically rely on certain properties of the input signal, for instance on its whiteness. Similarly, in *blind methods*, techniques from blind system identification are adapted to the identification of the linear component of the cascade (Bai & Fu, 2002; Vanbeylen, Pintelon, & Schoukens, 2009). The *frequency-domain methods* rely on the frequency content of the input and output signals. The output spectrum is a known function of the input frequency and magnitude, with some unknown parameters. By applying various sinusoidal inputs, these parameters can be identified (see, for instance, Baumgartner & Rugh, 1975). This requires that both the order of the polynomial input transformation and the order of the system are known. When these are unavailable, the harmonics of the output signal can be used to derive information about the nonlinear transformation in either a parametric or a nonparametric setting (see, for instance, Pintelon & Schoukens, 2012; Schoukens, Dobrowiecki, & Pintelon, 1998; Schoukens, Pintelon, Dobrowiecki, & Rolain, 2005). In *parametric–nonparametric methods*, mixed descriptions are used for the components of the cascade (in a time-domain setting). For instance, the nonlinear transformation can be described with kernels or orthogonal series and the linear component with polynomial models (see Greblicki, 1989; Greblicki & Pawlak, 1986; Mzyk, 2007). An effective method based on maximum likelihood has been proposed in Wills, Schön, Ljung, and Ninness (2013).

Besides a few exceptions (such as Pillonetto and Chiuso, 2009; Pillonetto, Quang, and Chiuso, 2011b; Risuleo, Bottegal, and Hjalmarsson, 2015a), in most of the aforementioned works the authors consider parametric models for the LTI block of the HS, with known model order. Also, in many cases, the authors consider polynomial models for the nonlinear transformation, with known order. If the model structure and orders are unknown, which might be the case in many applications, we have to use complexity criteria such as AIC, BIC, or cross validation (Ljung, 1999); however, the choice of the best model is in general a very difficult problem.

In this paper, we extend the preliminary work presented in Risuleo et al. (2015a). In that work, we considered HS with FIR models for the linear system and basis-function models for the static nonlinearity. The approach in Risuleo et al. (2015a) is, thus, limited to models where the noise is white and where the static nonlinearity is well represented by a model that is a linear combination of known basis functions. In this work, we relax the first modeling restriction by allowing for models with colored noise for the linear system (such as ARX, ARMAX, and Box–Jenkins models). To allow this relaxation, we use a nonparametric modeling approach where we model the impulse responses of the system in predictor form using Gaussian processes. We also relax the second modeling restriction by modeling the static nonlinearity with a Gaussian process.

To estimate the model, we follow an empirical Bayes (Maritz & Lwin, 1989) approach and we find an approximation of the posterior mean of the predictor impulse responses. This approximation depends on the parameters of the Gaussian-process model and on the noise variance (the *hyperparameters*), as well as on the input nonlinearity. To find these unknowns, we use a variation of the marginal-likelihood criterion. We maximize the likelihood of the data after integrating out the predictor impulse responses; the resulting problem is a joint maximum-a-posteriori/maximum likelihood problem (JMAP-ML, see Yeredor, 2000). The proposed procedure is reminiscent of the marginal-likelihood approach for hyperparameter tuning in LTI system identification (Pillonetto & Chiuso, 2015); however, the presence of the static nonlinearity complicates the solution of the JMAP-ML problem, because of the large number of decision variables. To overcome this, we design a new iterative solution scheme based on an extension of the expectation–maximization (EM) method (McLachlan & Krishnan, 2007), known as expectation-conditional maximization (ECM, see Meng & Rubin, 1993). In this way, we obtain a series of update rules for the estimates of the static nonlinearity and the hyperparameters. Except for one of the kernel hyperparameters, whose update is a scalar optimization over a finite domain, all the updates are available in closed-form.

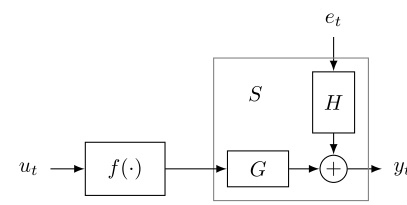

Fig. 1. Block scheme of the HS studied in this paper.

It is worth stressing that Bayesian kernel-based methods using the stable spline kernel are not new in nonlinear system identification (see for instance Pillonetto & Chiuso, 2009; Pillonetto et al., 2011b; Risuleo et al., 2015a). The method in Pillonetto et al. (2011b) is a fully nonparametric Bayesian method that can be used to identify any nonlinear model, without postulating specific structure. It works by identifying the model in nonlinear predictor form as a functional relationship between past inputs and the current output; however, it is not tuned specifically for the HS. Also the method in Pillonetto and Chiuso (2009) is fully nonparametric, this time specific for the Wiener–Hammerstein model structure—that is, a static nonlinearity sandwiched between two linear time-invariant models. In contrast with the present work it does not involve a marginalization step and thus cannot rely on the robustness properties of the empirical Bayes paradigm (see Wahba, 1990, Chap. 4 and Pillonetto and Chiuso, 2015).

The paper is organized as follows. In Section 2, we formulate the problem of system identification for the HS. In Section 3, we review the empirical Bayes approximation for the identification of the linear part of the system. In Section 4, we show the JMAP-ML criterion for the identification of the static nonlinearity and of the hyperparameters. In Section 5, we propose an ECM based algorithm to find the HS estimate. In Section 6, we compare, using simulations, our algorithm to standard system-identification tools for nonlinear system modeling. In Section 7, we give some concluding remarks. Appendix A contains the proofs of the main results.

2. Problem formulation

The HS we consider in this paper is the block-oriented nonlinear system structure given in Fig. 1. It is the cascade composition of a static nonlinear function $f(\cdot)$ and a linear system S. In our formulation the system S is a LTI discrete-time dynamic system (the *linear system*). We assume that the linear system is causal and exponentially stable. The linear system is fed by the input $\{w_t\}$ (unavailable to the experimenter), which is obtained by transforming the known input signal $\{u_t\}$ through the static nonlinear map $f(\cdot)$ (the *static nonlinearity*).

The dynamics of the HS considered in this paper are described by

$$\begin{aligned} y_t &= \sum_{k=0}^{\infty} g_k w_{t-k} + \sum_{k=0}^{\infty} h_k e_{t-k}\,, \quad h_0 = 1, \\ w_t &= f(u_t)\,, \end{aligned} \tag{1}$$

so that the linear system S is a general Box–Jenkins model, where $\{g_k\}_{k=0}^{\infty}$ and $\{h_k\}_{k=0}^{\infty}$ are the impulse responses of the causal and exponentially stable filters G and H (with H minimum phase), and $\{e_t\}$ zero-mean white Gaussian noise with unknown variance.