Automatica 50 (2014) 277–281

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Quasiconvexity analysis of the Hammerstein model☆

Mohammad Rasouli, David Westwick [1], William Rosehart

Department of Electrical and Computer Engineering, Schulich School of Engineering, Calgary, Canada

ARTICLE INFO

Article history:
Received 10 March 2012
Received in revised form
16 May 2013
Accepted 5 October 2013
Available online 9 December 2013

Keywords:
Hammerstein
Identification
Convex
Quasiconvex

ABSTRACT

In this paper, the Hammerstein identification problem with correlated inputs is studied in a prediction error framework using separable least squares methods. Thus, the identification is recast as an optimization over the parameters used to describe the nonlinearity. A sufficient condition is derived that guarantees that the identification problem is quasiconvex with respect to the parameters that describe the nonlinearity. Simulations using both IID and correlated inputs are used to illustrate the result.

1. Introduction

Many of the identification problems in which the system is nonlinear in the parameters are non-convex. Hence, iterative methods can get stuck in one of the local minima of the criterion function (Ljung, 2008). An often successful heuristic approach, which does not guarantee the global solution, is random initialization of the parameter vector. This difficulty can be avoided if the identification can be formulated as a convex or quasiconvex optimization.

The Hammerstein model, a memoryless nonlinearity followed by a dynamic linear element, has been used to model a variety of nonlinear systems including heat exchangers (Eskinat, Johnson, & Luyben, 1991), stretch reflexes (Dempsey & Westwick, 2004), satellite and microwave communication links (Prakriya & Hatzinakos, 1997), and solid oxide fuel cells (Jurado, 2006). Its identification has been extensively studied. The earliest contributions were an iterative method, proposed by Narendra and Gallman (1966), and a non-iterative but overparameterized approach (Chang & Luus, 1971). Methods based on the properties of separable processes, which allow the linear element to be recovered using essentially linear system identification methods, have also been extensively studied (Billings & Fakhouri, 1982; Enqvist, 2010). Finally, prediction methods based on separable nonlinear least squares (SNLS) optimization (Golub & Pereyra, 1973; Sjoberg & Viberg, 1997) have been proposed for certain classes of Hammerstein models (Bai, 2002; Dempsey & Westwick, 2004; Westwick & Kearney, 2001).

The convergence of the iterative algorithm for the Hammerstein model was extensively studied by Bai and Li (2004), who showed that the appropriate use of a normalization step can guarantee convergence. Furthermore, they discussed the convexity of the identification problem and proved that all of the estimation problems encountered in the iterative algorithm are convex, although they did not address the convexity of the overall identification problem and the proof was limited to Independent and Identically Distributed (IID) inputs.

Cai and Bai (2011) showed that by using an SNLS approach, the loss-function in the Hammerstein identification problem could be expressed as a function of the inner product between the system and model parameter vectors. In this paper, we will extend the results from Cai and Bai (2011) in two ways. First, the FIR linear element will be replaced with an expansion onto a basis of IIR filters, such as the rational orthogonal basis functions considered in Heuberger, van den Hof, and Wahlberg (2005). Second, the restriction to IID inputs will be relaxed to include correlated inputs. We will show that a result similar to that in Cai and Bai (2011) can be obtained; however under correlated inputs, the resulting inner product involves an orthogonal projection onto a subspace defined by the properties of the input and of the basis elements used to represent the linear and nonlinear elements.

Preliminary results have been previously published in the conference paper (Rasouli, Westwick, & Rosehart, 2011). The present contribution extends those results, removing the requirement for

☆ The material in this paper was partially presented at the 18th IFAC World Congress, August 28–September 2, 2011, Milano, Italy. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

E-mail addresses: mrasouli@ucalgary.ca (M. Rasouli), dwestwic@ucalgary.ca (D. Westwick), rosehart@ucalgary.ca (W. Rosehart).

[1] Tel.: +1 403 220 2725; fax: +1 403 282 6855.

0005-1098/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.automatica.2013.11.004

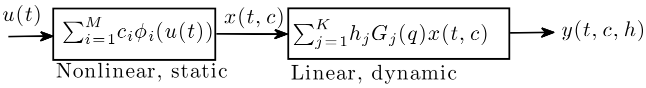

Fig. 1. Block diagram of the Hammerstein model consisting of static nonlinearity followed by a linear dynamic system.

an orthogonal expansion basis for the nonlinearity, and establishing the conditions necessary to invoke the law of large numbers. New simulations using an IIR linear element have been included.

The outline of this paper is as follows: Section 2 will introduce the Hammerstein model used throughout the paper and list the assumptions made regarding the model's inputs and elements. The problem formulation and identification of the Hammerstein model will be described in Section 3. Section 4 will provide sufficient conditions under which the problem is quasiconvex for correlated inputs. Simulation results for clarifying the presented materials are shown in Section 5. Section 6 will conclude the paper.

1.1. Notation

This paper deals with discrete time models. Thus, the time index, t, is an integer, and q is the forward shift operator. Thus, $q^{-1}x(t) = x(t-1)$. Bold lower and uppercase roman letters will denote vectors and matrices, respectively. The vector $\boldsymbol{x}$ will be understood to contain elements of the signal, $x(t)$. Similarly, $G(q)\boldsymbol{x}$ denotes a vector that contains the signal $G(q)x(t)$, and $\phi(\mathbf{u})$ denotes a vector obtained by applying the function $\phi(\cdot)$ to the individual elements of the vector $\mathbf{u}$.

2. Hammerstein model

A Hammerstein model comprises a memoryless nonlinearity followed by a linear filter as shown in Fig. 1. In this study, the nonlinearity will be represented by a basis expansion of known maximum degree, M. Thus,

$$x(t, \boldsymbol{c}) = \sum_{i=1}^{M} c_i \phi_i(u(t)), \tag{1}$$

where $u(t)$ is the input signal, $x(t)$ is the intermediate signal, the c_k are the expansion coefficients contained in the vector $\boldsymbol{c}$, and the $\phi_k(\cdot)$ are a set of basis functions. Such expansions, including polynomials and various linear and cubic splines, have been widely used in nonlinear block structured models, such as the Hammerstein cascade (Dempsey & Westwick, 2004; Van Pelt & Bernstein, 2001).

Similarly, the linear element will be modeled using an expansion

$$y(t, \boldsymbol{c}, \boldsymbol{h}) = \sum_{j=1}^{K} h_j G_j(q) x(t, \boldsymbol{c}) \tag{2}$$

where the $G_j(q)$ are a bank of linear time-invariant filters. Common choices for these filters include delays, $G_j(q) = q^{-j}$, or discrete Laguerre or Kautz filters (Heuberger et al., 2005).

2.1. Model class

The model may be represented by the parameters of the nonlinearity and linear filter, $[\boldsymbol{c}^T \boldsymbol{h}^T]^T$. Thus, the output of the Hammerstein model can be written as

$$y(t, \boldsymbol{c}, \boldsymbol{h}) = \sum_{i=1}^{M} \sum_{j=1}^{K} c_i h_j G_j(q) \phi_i(u(t)). \tag{3}$$

To ensure that all the signals are well behaved, we assume the following:

MC1 The filters $G_j(q)$ are exponentially stable.
MC2 The basis functions $\phi_i(\cdot)$ are continuous.
MC3 The parameter vectors have finite norms, $\|\boldsymbol{c}\| < K_c < \infty$ and $\|\boldsymbol{h}\| < K_h < \infty$ for some finite constants K_c and K_h.

Note, however, that this representation contains a redundant degree of freedom, in that the models $[\boldsymbol{c}^T \boldsymbol{h}^T]^T$ and $[K\boldsymbol{c}^T \frac{1}{K}\boldsymbol{h}^T]^T$, where K is any non-zero real number, will produce identical outputs, given the same input. Thus, it is customary to normalize either $\boldsymbol{c}$ or $\boldsymbol{h}$.

2.2. Assumptions

A1 The input, $u(t)$, is the output of a stable linear time-invariant filter that is being driven by an IID sequence.
A2 The input probability density is such that the output of each nonlinear basis function, $\phi_i(u(t))$ for $i = 1 \dots M$, has finite fourth-order moments.
A3 The true system is within the model class described above. Let $\boldsymbol{c_0}$ and $\boldsymbol{h_0}$ be the parameter vectors containing the nonlinearity coefficients and filter weights, respectively. These are both assumed to be non-zero, $\|\boldsymbol{c_0}\| > 0$ and $\|\boldsymbol{h_0}\| > 0$.
A4 The output is corrupted by an additive, Gaussian IID noise process of finite variance, σ_n^2, that is independent of the input, $u(t)$.

3. Problem formulation and identification

The system identification problem is to estimate $\boldsymbol{c}$ and $\boldsymbol{h}$, to within the unknown internal gain described above, given measurements of the input, and the possibly noise corrupted output, $y(t)$.

Let $\boldsymbol{y_0} = \boldsymbol{y}(\boldsymbol{c_0}, \boldsymbol{h_0})$ be the noise-free output of the true system. Since the measurement noise is assumed to be Gaussian and IID (A4), and there is no undermodelling (A3), minimizing the sum of squared prediction errors produces the maximum likelihood estimate of the model parameters. The prediction errors are given by

$$\begin{aligned} \boldsymbol{\epsilon}(\boldsymbol{c}, \boldsymbol{h}) &= \boldsymbol{y_0} + \boldsymbol{n} - \boldsymbol{y}(\boldsymbol{c}, \boldsymbol{h}) \\ &= \boldsymbol{\epsilon_m} + \boldsymbol{n} \end{aligned} \tag{4}$$

where $\boldsymbol{\epsilon_m}$ is the modeling error. This leads to the loss function

$$\begin{aligned} V_N(\boldsymbol{c}, \boldsymbol{h}) &= \frac{1}{N}[\boldsymbol{\epsilon}^T(\boldsymbol{c}, \boldsymbol{h})\boldsymbol{\epsilon}(\boldsymbol{c}, \boldsymbol{h})] \\ &= \frac{1}{N}[\boldsymbol{\epsilon}_{\boldsymbol{m}}^T(\boldsymbol{c}, \boldsymbol{h})\boldsymbol{\epsilon_m}(\boldsymbol{c}, \boldsymbol{h}) + 2\boldsymbol{\epsilon}_{\boldsymbol{m}}^T(\boldsymbol{c}, \boldsymbol{h})\boldsymbol{n} + \boldsymbol{n}^T\boldsymbol{n}]. \end{aligned} \tag{5}$$

Lemma 3.1. *For models in the model class defined by (3) under conditions* MC1–MC3*, and under Assumptions* A1, A2 *and* A4*, the loss function converges uniformly to*

$$\lim_{N\to\infty} V_N(\boldsymbol{c}, \boldsymbol{h}) = E[\epsilon_m^2(\boldsymbol{c}, \boldsymbol{h}, t)] + \sigma_n^2$$

as $N \to \infty$.

Proof. Given the model class, Assumptions A1 and A2 guarantee the conditions necessary to apply Lemma 3.1 from Ljung (1978), so that

$$\lim_{N\to\infty} V_N(\boldsymbol{c}, \boldsymbol{h}) = E[V_N(\boldsymbol{c}, \boldsymbol{h})]$$

with probability 1. By A4 the noise is independent of the input, all cross-terms vanish in expectation, and hence in the limiting average. □