Automatica 71 (2016) 179–186

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Convergence analysis for recursive Hammerstein identification☆

Per Mattsson [1], Torbjörn Wigren

Department of Information Technology, Uppsala University, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 4 March 2015
Received in revised form
25 November 2015
Accepted 30 March 2016
Available online 31 May 2016

Keywords:
Averaging
Cement mill
Convergence analysis
Hammerstein system
Recursive identification

ABSTRACT

This paper derives a recursive prediction error identification method based on the Hammerstein model structure. The convergence properties of the algorithm are analysed by application of Ljung's associated differential equation method. It is proved that the algorithm can only converge to stable stationary points of the associated ordinary differential equation. General conditions for local convergence to the true parameter vector are given, and the cases with piecewise linear and polynomial nonlinearities are treated in detail. The derived identification method is illustrated in a numerical study that treats identification of a subsystem of a cement mill.

1. Introduction

In this paper recursive identification of the single-input–single-output (SISO) Hammerstein model is considered. Identification of Hammerstein models is a well studied subject. A common approach is to transform the Hammerstein model into a linear model, using overparametrization, see e.g. Chang and Luus (1971). If the noise is correlated, instrumental-variable methods can be used to construct a consistent estimator (Stoica & Söderström, 1982). In Boutayeb, Aubry, and Darouach (1996) overparameterization is combined with pseudo-inverse techniques to construct a recursive estimator for the original parameters. Another approach is the (non-recursive) iterative method (Narendra & Gallman, 1966), for which the convergence properties are studied in Bai and Li (2004). Common for all these methods is that the static nonlinearity is assumed to be a linear combination of known basis functions. In Chen (2004) and Zhao and Chen (2006), a nonparametric model is used for the nonlinearity and a consistent estimator is constructed. However, the method only works if the input is white, which is restrictive in practice.

The algorithm of the present paper was previously presented in Mattsson and Wigren (2014), and is based on direct optimization of a SISO Hammerstein model, as in Ding, Liu, and Liu (2011) and Ding, Xinggao, and Chu (2013). By not using overparameterization accuracy can be improved. The method uses a static nonlinearity that is expressed as a sum of basis functions. The algorithm avoids the computational complexity associated with e.g. particle filtering.

The main contribution of the paper is an analysis of the algorithm with Ljung's associated differential equation method of Ljung (1977) and Ljung and Söderström (1983). That method ties the convergence of the recursive identification algorithm to the stability of the associated ordinary differential equation (ODE). More specifically, local convergence of the algorithm follows if the ODE is locally stable, and global convergence is implied if the associated ODE is globally Lyapunov stable. Conditions under which the algorithm is guaranteed to converge are established. An analysis of the case where the system is in the model set proves local convergence to the true parameter vector under relatively mild assumptions. To illustrate the performance in non-ideal situations where the system to be identified is not in the model set, identification of a cement mill subsystem is simulated. The results indicate that the algorithm operates well also in such practical situations.

The paper is organized with development of the model and algorithm in Sections 2 and 3. The convergence analysis appears in Sections 4 and 5 with proofs in the appendices. The cement mill based numerical example is treated in Section 6, followed by conclusions in Section 7.

2. The Hammerstein model

The model considered here is described by

$$\hat{y}(t,\theta) = \frac{B(q^{-1})}{A(q^{-1})} f(\theta_n, u(t)) = \bar{\theta}_l^{\mathrm{T}} \varphi(t,\theta) \tag{1}$$

☆ This research has been supported by the European Research Council (ERC) via the Advanced Grant 247035. The material in this paper was partially presented at the 2014 American Control Conference, June 4–6, 2014, Portland, OR, USA. This paper was recommended for publication in revised form by Associate Editor Andrea Garulli under the direction of Editor Torsten Söderström.

E-mail addresses: per.mattsson@it.uu.se (P. Mattsson), torbjorn.wigren@it.uu.se (T. Wigren).

[1] Tel.: +46 739 540317.

http://dx.doi.org/10.1016/j.automatica.2016.04.014
0005-1098/

where $u(t)$ is the input signal, $\hat{y}(t,\theta)$ is the output of the model and q^{-1} denotes the backward shift operator. The polynomials $A(q^{-1})$ and $B(q^{-1})$ are given by

$$B(q^{-1}) = b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b} \tag{2}$$

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n_a}. \tag{3}$$

The parameter vector $\bar{\theta}$ is partitioned as

$$\bar{\theta} = \begin{bmatrix} \bar{\theta}_n^{\mathrm{T}} & \bar{\theta}_l^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \tag{4}$$

$$\bar{\theta}_l = \begin{bmatrix} a_1 & \dots & a_{n_a} & b_1 & \cdots & b_{n_b} \end{bmatrix}^{\mathrm{T}}, \tag{5}$$

where $\bar{\theta}_n$ are the parameters of the nonlinearity, and

$$\varphi(t,\theta) = [-\hat{y}(t-1,\theta) \quad \cdots \quad -\hat{y}(t-n_a,\theta) \\ f(\theta_n, u(t-1)) \quad \cdots \quad f(\theta_n, u(t-n_b))]^{\mathrm{T}}.$$

As in Ding et al. (2011) and Mattsson and Wigren (2014), the static nonlinearity is considered to be a linear combination of basis functions, so that

$$f(\theta_n, u) = \sum_{i=1}^{n_k} k_i f_i(u) = \bar{\theta}_n^{\mathrm{T}} F(u), \tag{6}$$

$$\bar{\theta}_n = \begin{bmatrix} k_1 & \cdots & k_{n_k} \end{bmatrix}^{\mathrm{T}}, \tag{7}$$

$$F(u) = \begin{bmatrix} f_1(u) & \cdots & f_{n_k}(u) \end{bmatrix}^{\mathrm{T}}. \tag{8}$$

A bias can be included by setting $f_1(u) = 1$.

Because of the cascade structure in (1), it is not possible to separate the gain of the linear block from the gain of the static nonlinearity. Therefore, it is assumed that one element in the vector $\bar{\theta}$ is fixed, and known beforehand. The vector θ of unknown parameters is then given by

$$\theta = I_o \bar{\theta} \tag{9}$$

where I_o is the matrix that removes the row corresponding to the fixed parameter.

3. The recursive identification algorithm

In this section the recursive prediction error algorithm (RPEM) presented in Mattsson and Wigren (2014) is described. It is derived by minimization of the criterion

$$V(\theta) = \frac{1}{2} \mathrm{E}\, \varepsilon^2(t,\theta), \tag{10}$$

where E is the expectation operator, $\varepsilon(t,\theta) = y(t) - \hat{y}(t,\theta)$ is the prediction error, and $y(t)$ is the measured output signal of the plant. In order to perform the derivation, the negative gradient, $\psi(t,\theta)$, of $\varepsilon(t,\theta)$ with respect to the parameter vector is needed. Let

$$\bar{\psi}_l(t,\theta) = \left[\frac{d}{d\bar{\theta}_l}\hat{y}(t,\theta)\right]^{\mathrm{T}} = \frac{1}{A(q^{-1})}\varphi(t,\theta) \tag{11}$$

$$\bar{\psi}_n(t,\theta) = \left[\frac{d}{d\bar{\theta}_n}\hat{y}(t,\theta)\right]^{\mathrm{T}} = \frac{B(q^{-1})}{A(q^{-1})}F(u(t)) \tag{12}$$

$$\bar{\psi}(t,\bar{\theta}) = \begin{bmatrix} \bar{\psi}_n^{\mathrm{T}}(t,\theta) & \bar{\psi}_l^{\mathrm{T}}(t,\theta) \end{bmatrix}^{\mathrm{T}}. \tag{13}$$

The negative gradient, $\psi(t,\theta)$, is then given by

$$\psi(t,\theta) = I_o \bar{\psi}(t,\theta), \tag{14}$$

cf. (9). In order to ensure that the estimated linear dynamic block is stable in each step of the RPEM, a projection algorithm is needed. First define

$$D_s = \left\{ \theta \,\middle|\, \frac{1}{A(q^{-1})} \text{ is asymptotically stable} \right\}. \tag{15}$$

It must also be ensured that the estimates stay in a compact set (Ljung & Söderström, 1983). Therefore, let $D_{\mathcal{M}}$ be a compact subset of D_s. Then the following projection algorithm can be used to ensure that $\hat{\theta}(t)$ stays in $D_{\mathcal{M}}$,

$$\left[\hat{\theta}(t)\right]_{D_{\mathcal{M}}} = \begin{cases} \hat{\theta}(t) & \text{if } \hat{\theta}(t) \in D_{\mathcal{M}} \\ \hat{\theta}(t-1) & \text{otherwise} \end{cases}. \tag{16}$$

The recursive identification algorithm of Gauss–Newton type (Ljung & Söderström, 1983; Mattsson & Wigren, 2014), is then

$$\begin{aligned} \varepsilon(t) &= y(t) - \hat{y}(t) \\ R(t) &= R(t-1) + \gamma(t)(\psi(t)\psi^{\mathrm{T}}(t) - R(t-1)) \\ \hat{\theta}(t) &= \left[\hat{\theta}(t-1) + \gamma(t) R^{-1}(t)\psi(t)\varepsilon(t)\right]_{D_{\mathcal{M}}} \\ \varphi(t+1) &= [-\hat{y}(t) \quad \cdots \quad \hat{y}(t-n_a+1) \\ &\qquad f(\hat{\theta}_n, u(t)) \quad \cdots \quad f(\hat{\theta}_n(t), u(t-n_b+1))]^{\mathrm{T}} \\ \hat{y}(t+1) &= \hat{\bar{\theta}}_l^{\mathrm{T}}(t)\varphi(t+1) \\ \psi(t+1) &= I_o \begin{bmatrix} \dfrac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})} F(u(t+1)) \\ \dfrac{1}{\hat{A}(q^{-1})}\varphi(t+1) \end{bmatrix}. \end{aligned} \tag{17}$$

Here $\hat{\theta}(t)$ is the parameter estimate.

4. Convergence analysis

In Ljung (1977) it was shown how the algorithm in Section 3 can be related to the following ODE:

$$\frac{d}{d\tau}\theta_D(\tau) = R_D^{-1}(\tau) f_A(\theta_D(\tau)) \tag{18}$$

$$\frac{d}{d\tau}R_D(\tau) = G_A(\theta_D(\tau)) - R_D(\tau) \tag{19}$$

where

$$f_A(\theta) \triangleq \lim_{t\to\infty} \mathrm{E}\, \psi(t,\theta)\varepsilon(t,\theta) \tag{20}$$

$$G_A(\theta) \triangleq \lim_{t\to\infty} \mathrm{E}\, \psi(t,\theta)\psi^{\mathrm{T}}(t,\theta), \tag{21}$$

for a fixed θ. In order to use the associated ODE approach, it is required to represent the RPEM on state-space form, this is performed in Section 4.1. In Section 4.2 conditions on data generation that ensure the applicability of the results in Ljung (1977) are presented, while the corresponding theorems are presented in Section 4.3. This provides the foundation for the detailed analysis of the Hammerstein identification algorithm that follows in Sections 4.4–4.5. In these sections conditions that ensure that the true parameter vector is a possible convergence point are given. This requires that detailed conditions on the parameterization and signals are introduced, as in e.g. Wigren (1994). In particular this provides new insight on why it is important to excite the system well in amplitude as well as in frequency, cf. Wigren (2003). Finally, the convergence analysis is summarized in Section 4.6.

4.1. State-space representation of the RPEM

In this section the algorithm in Section 3 is reformulated in a way that corresponds to the general structure in Ljung (1977).

Lemma 1. *In the RPEM (17), $\hat{y}(t)$ and $\psi(t)$ can be generated by a linear time-variant system on the form*

$$\xi(t+1) = A(\hat{\theta}(t))\xi(t) + B(\hat{\theta}(t))z(t) \tag{22}$$

$$\begin{bmatrix} \hat{y}(t) & \psi^{\mathrm{T}}(t) & y(t) \end{bmatrix}^{\mathrm{T}} = C(\hat{\theta}(t-1))\xi(t). \tag{23}$$