Automatica 94 (2018) 349–360

Contents lists available at ScienceDirect

Automatica

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

An improved method for Wiener–Hammerstein system identification based on the Fractional Approach☆

Giuseppe Giordano *, Sébastien Gros, Jonas Sjöberg

Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 26 May 2017
Received in revised form 13 December 2017
Accepted 11 April 2018
Available online 29 May 2018

Keywords:
System identification
Wiener–Hammerstein
Conditioning
Identifiability
Lifting
Newton methods

ABSTRACT

This paper develops and analyses a novel method for identifying Wiener–Hammerstein models, i.e. models consisting of two linear dynamic parts with a static non-linearity in between. Starting from the best linear model, which is a consistent estimate of the system dynamics for Gaussian excitation, the identification problem includes the partitioning of the poles and zeros between the two linear parts. This partitioning can be formulated as a discrete optimization problem. The *fractional approach* considers a relaxation of this problem into a continuous one, by parameterizing the partition of each pole and zero in a fractional way, and carrying out the computations in the frequency domain. In this paper it is shown that the fractional approach becomes ill-conditioned for some configurations of the poles and zeros of the linear dynamic parts, causing identifiability issues. A modification of the original fractional approach is then introduced, based on series expansion of the fractional transfer functions. This modification shares most of the properties of the fractional approach. However, it is shown that it provides an implicit regularization of the identification problem. It addresses the ill-conditioning problem while preserving meaningful statistical properties of the estimation. Furthermore, a lifted formulation of the estimation problem is proposed, which improves the algorithmic performance in the framework of Newton-based methods.

1. Introduction

Linear models are often used in system identification due to their simplicity, and because they allow for simple and systematic analysis of the properties of the resulting model. Moreover, the theory underlying the identification of linear systems is fairly mature, and well covered in the literature. E.g., Ljung (1999) and Söderström and Stoica (1988) focus on time-domain methods and Pintelon and Schoukens (2001) focus on frequency-domain methods. However, real systems are non-linear to some extent, and justify the need for more advanced models. General non-linear model structures, such as Volterra series or neural networks combined with dynamical models can approximate almost any non-linear system. However, they often yield overly flexible model structures, which are prone to over-fitting, often hard to analyse, and difficult to deploy in practice. A more structured approach is to limit the flexibility of the model by considering block-oriented modelling based on combining blocks that are either linear dynamic systems or static nonlinear functions, (Giri and Bai, 2010; Lauwers, Schoukens, Pintelon, Moer, and Gomme, 2007). Constraining the dynamic part of the model to be linear still allows to deploy some aspects of the linear system theory to analyse the model properties. The simplest block-oriented models are the Hammerstein model, consisting of a static non-linearity followed by a linear dynamic block, and the Wiener model, where the two blocks are in the reversed order. These models can be generalized into Hammerstein–Wiener (HW) models, which have two static nonlinear blocks with a linear dynamic block in between, and Wiener–Hammerstein (WH) models with two linear blocks with a static nonlinear block in the middle, see Fig. 1.

A Wiener–Hammerstein model is parameterized with parameters describing the two linear parts, e.g. their poles and zeros, and parameters describing the non-linearity. To estimate the parameters it is straightforward to apply the prediction error method (PEM). It provides a consistent estimate when only measurement noise affects the system, see Ljung (1999) and Söderström and Stoica (1988). However, the cost function underlying the PEM can have multiple minima and an iterative search of the minimum is required. This difficulty is e.g. pointed out in Wills and Ninness (2012) where PEM is used to a Wiener–Hammerstein benchmark problem starting from random chosen poles and zeros. The existence of local minima entails that the initial guess provided

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Martin Enqvist under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: giuseppe.giordano@chalmers.se (G. Giordano), grosse@chalmers.se (S. Gros), jonas.sjoberg@chalmers.se (J. Sjöberg).

https://doi.org/10.1016/j.automatica.2018.04.046
0005-1098/

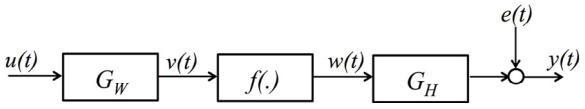

Fig. 1. The Wiener–Hammerstein system consists of the interconnection of a first LTI part G_W, a static nonlinearity f, and a second LTI part G_H. It is assumed that the output of the system is corrupted by measurement noise $e(t)$. The intermediate signals $v(t)$ and $w(t)$ are not measurable.

to the optimization algorithm is crucial for the success of the identification.

Many approaches to obtain an initial guess for the parameters rely on the fact that the Best Linear Approximation (BLA) is a consistent estimate of the concatenation of the two linear dynamic blocks, when the input belongs to the Riemann equivalence class of asymptotically normally distributed signals (Enqvist & Ljung, 2005; Pintelon & Schoukens, 2001). An extensive study on block-oriented approaches based on the BLA can be found in Schoukens and Tiels (2016). Therefore, the problem of retrieving a good initial guess for the parameters of the two linear parts is reduced to a partition problem, where it is required to correctly divide the dynamics contained in the BLA between the two linear parts. In Sjöberg and Schoukens (2012), an exhaustive search approach is suggested where all possible partitionings are considered. This approach appears to be computationally tractable for the partitioning problem up to models of order 10. Indeed, since the static non-linearity is linearly parameterized, estimating the non-linearity for a given partition of the poles and zeros requires solving a small-scale Linear Least Squares (LLS) problem. Nevertheless, the computational burden of the exhaustive search approach increases dramatically for large systems. Several alternatives have been proposed in the literature. In Sjöberg, Lauwers, and Schoukens (2012), the linear parts are parameterized via basis functions. The poles of the BLA are used to define the basis functions of the first linear part and the zeros are used to build the *inverse* of the second linear part. The model structure is over-parameterized in the first stage, and the fitting of the parameters is done by filtering the input through the first linear part and the output through the inverse of the second linear part. An SVD is then used to perform a model reduction. Unfortunately, this method does not allow any right half-plane zeros, and the resulting estimation is biased as the measurement noise is filtered by the poles.

An alternative method, labelled *fractional approach*, is presented in Vanbeylen (2014). The method describes both linear systems as the BLA, and integer exponents in the set $\{0, 1\}$ are introduced for every pole/zero for describing the partition between the two linear parts. A relaxation of the set $\{0, 1\}$ to the continuous interval $[0, 1]$ is then performed. The resulting real-valued exponents can then be identified via a continuous optimization problem. When all the identified real-valued exponents are close to their integer values $\{0, 1\}$, a partition of the pole or the zero can be decided. Iterative methods can be used to solve the continuous problem efficiently. Compared to the methods in Sjöberg et al. (2012) and Sjöberg and Schoukens (2012), with the fractional approach, both exhaustive search and over-parameterization are avoided, and right-half plane zeros are allowed. Furthermore, the original discrete partitioning problem becomes a continuous one, resulting in a reduced complexity.

In this paper the estimation of the exponents in the fractional approach and its identifiability issues are analysed. In general, the WH model can have identifiability deficiencies in two cases. The first one occurs when the static non-linearity is close to being linear; the second, when the BLA has poles and/or zeros near the origin. The intuitive explanation is obvious for these cases. If the static block is linear, the placement of the poles and zeros have no influence on the system output. The same it is true for the second case, since poles/zeros at the origin corresponds to pure time shifts. We show that identifiability issues can also arise from the continuous relaxation of the discrete problem of the fractional approach. More specifically, the following pole/zero configurations can lead to ill-conditioning: poles (or zeros) are being close to one another, and belong to the same linear part; a pole and a zero are close to one another, and belong to different linear parts.

In the fractional approach, the continuous relaxation of the set of the integer exponents yields fractional dynamics, which can be treated via series expansions (Giordano & Sjöberg, 2015). We propose a novel expansion method, supported by a sound theoretical framework, which is exact for integer values of the exponents. This novel expansion naturally introduces a form of regularization in the estimation problem, which alleviates its potential ill-conditioning. Moreover, since no artificial regularization is introduced, the identification problem retains a meaningful description of the (local) statistical properties of the estimation. Finally, a novel formulation of the identification problem based on lifting techniques (Albersmeyer & Diehl, 2010) is proposed, yielding advantageous properties in the resulting continuous optimization algorithm, which allow for a faster and more reliable convergence to the solution when using Newton-type methods.

The paper is organized as follows: Section 2 introduces the model parametrization for the Wiener–Hammerstein system and the parameter estimation problem. In Section 3, the fractional approach is introduced as an initialization algorithm for the estimation problem and its issues are analysed. An expanded formulation of the fractional approach is introduced in Section 4. In Section 5 the properties of the expanded formulation are presented and convergence aspects of the identification algorithm are addressed. Section 6 illustrates the conditioning problem and its solutions via simulation and experimental examples, and Section 7 concludes the paper. Appendices A, B, and C contain the proofs of the theorems related to the conditioning and convergence results.

2. Modelling framework and initialization problem

We recall here the general framework for WH system identification. First, the Wiener–Hammerstein system and the assumptions on the data used for estimation are presented. Then the model parametrization and the estimation problem are introduced. The need of a good initial estimate is discussed and, finally, the BLA is formally defined.

2.1. Wiener–Hammerstein system and data

A Wiener–Hammerstein (WH) system is defined, in discrete-time, as

$$y_0(t) = G_H^0(q)w(t), \tag{1a}$$

$$w(t) = f(v(t)), \tag{1b}$$

$$v(t) = G_W^0(q)u(t), \tag{1c}$$

where q denotes the forward time-shift operator ($qx(t) = x(t+1)$). We assume that the measurement data are generated by a WH system as defined in (1). The input is a realization of a normally distributed random process and it is known, while the output is corrupted by additive white Gaussian noise $e(t)$ with zero mean and variance σ_e^2,

$$y(t) = y_0(t) + e(t). \tag{2}$$

The input is persistently exciting, see Ljung (1999), the LTI systems and the static nonlinearity. A set of N data is assumed to be available for estimation, $\{u(t), y(t)\}_{t=1}^N$. The intermediate signals $v(t)$ and $w(t)$ are unknown and, hence, not available for identification.