Automatica 46 (2010) 1675–1681

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Two nonlinear optimization methods for black box identification compared[☆]

Anne Van Mulders[a,*], Johan Schoukens[a], Marnix Volckaert[b], Moritz Diehl[c]

[a] Vrije Universiteit Brussel, Department ELEC, Pleinlaan 2, B1050 Brussels, Belgium
[b] K.U.Leuven, Department MECH, Celestijnenlaan 300b, B3001 Leuven, Belgium
[c] K.U.Leuven, Department ESAT, Kasteelpark Arenberg 10, B3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 16 June 2009
Received in revised form
16 April 2010
Accepted 27 May 2010
Available online 24 July 2010

Keywords:
Identification algorithms
Parameter estimation
Nonlinear systems
Nonlinear models
State-space models
Constraints

ABSTRACT

In this paper, two nonlinear optimization methods for the identification of nonlinear systems are compared. Both methods estimate the parameters of e.g. a polynomial nonlinear state-space model by means of a nonlinear least-squares optimization of the same cost function. While the first method does not estimate the states explicitly, the second method estimates both states and parameters adding an extra constraint equation. Both methods are introduced and their similarities and differences are discussed utilizing simulation data. The unconstrained method appears to be faster and more memory efficient, but the constrained method has a significant advantage as well: it is robust for unstable systems of which bounded input–output data can be measured (e.g. a system captured in a stabilizing feedback loop). Both methods have successfully been applied on real-life measurement data.

1. Introduction

Most real-life systems are to some extent nonlinear. For such systems, nonlinear modelling can improve the identification results, i.e. reduce the residual error. There exist several types of nonlinear models, among them black box models such as Volterra systems, block structured models, neural networks and fuzzy models. For more information, we refer to Giannakis and Serpedin (2001).

In this paper, the model structure is a conventional linear state-space model extended with polynomial nonlinear terms. It is called a Polynomial Nonlinear State-space Model (PNLSS) (Paduart, 2008). The PNLSS model serves as an example, but the ideas could be applied to other nonlinear state-space models as well.

The goal is to estimate the parameters of the nonlinear model given the exact input and the noisy output measurements. The states are assumed to be unknown and in order to solve the problem we will minimize the least square error between the measured and modelled output.

The contributions of this work are the following:

- Construction of a constrained optimization method that allows one to estimate the PNLSS model parameters from input–output data. The approach is related to multiple shooting methods for parameter estimation (Bock, 1987), but is now applied to black box system identification with a procedure that provides initial estimates of the parameters;
- Evaluation of its properties, such as the estimation of an unstable system based on bounded input–output data, by means of simulation examples;
- Comparison of the performances (least square errors) of the two optimization methods.

The structure of the paper is the following: in Section 2, we present the model structure. In Section 3, we explain briefly the generation of initial estimates for the parameters to be estimated. In Section 4, the two nonlinear optimization methods are discussed and Section 5 shows the simulation results. The last section recapitulates the main conclusions of this paper. Although the paper does not contain any experimental results, both methods have already been used in practice, such as the benchmark session of SYSID2009 (Paduart, Lauwers, Pintelon, & Schoukens, 2009; Schoukens, Suykens, & Ljung, 2009; Van Mulders, Volckaert, Diehl, & Schoukens, 2009).

☆ The material in this paper was partially presented at 15th IFAC Symposium on System Identification, SYSID 2009, July 6–8, 2009, Saint-Malo, France. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +32 2 629 36 65; fax: +32 2 629 28 50.
E-mail addresses: Anne.Vanmulders@vub.ac.be (A. Van Mulders), Johan.Schoukens@vub.ac.be (J. Schoukens), Marnix.Volckaert@mech.kuleuven.be (M. Volckaert), Moritz.Diehl@esat.kuleuven.be (M. Diehl).

0005-1098/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2010.06.021

2. Model structure

2.1. Class of discrete time systems considered

As is known very well, state-space models are particularly well suited for multiple-input multiple-output (MIMO) systems. Let n_u and n_y represent resp. the number of input and output signals. In general, a discrete time nonlinear state-space model can be formulated as:

$$\begin{aligned} x(t+1) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)). \end{aligned} \tag{1}$$

Herein, $t = [0 \cdots N-1]$ is the discrete time instant, $x \in \mathbb{R}^{n \times N}$ are the states and $u \in \mathbb{R}^{n_u \times N}$ and $y \in \mathbb{R}^{n_y \times N}$ are the input and output. n is the model order and N is the total number of time instants. The upper equation is called the state equation and describes the evolution of the states. The lower equation is called the output equation and describes the output as function of the states and inputs.

In our case, we assume that the exact description of the nonlinear system is of the form:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + E\zeta(x(t), u(t)) \\ y(t) &= Cx(t) + Du(t) + F\eta(x(t), u(t)). \end{aligned} \tag{2}$$

The vectors $\zeta \in \mathbb{R}^{n_\zeta}$ and $\eta \in \mathbb{R}^{n_\eta}$ contain monomials in $x(t)$ and $u(t)$; the matrices $E \in \mathbb{R}^{n \times n_\zeta}$ and $F \in \mathbb{R}^{n_y \times n_\eta}$ contain the coefficients associated with those monomials. n_η and n_ζ are the number of monomials in resp. η and ζ.

The above mentioned model is called a polynomial nonlinear state-space model (PNLSS). It consists of a classical linear state-space model with nonlinear terms $E\zeta$ and $F\eta$. The coefficients of the linear terms in $x(t)$ (the states) and $u(t)$ (the inputs) are given by the coefficient matrices $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n_u}$ in the state equation, and $C \in \mathbb{R}^{n_y \times n}$ and $D \in \mathbb{R}^{n_y \times n_u}$ in the output equation.

The monomials can be any chosen set of combinations of $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} u_1^{\beta_1} u_2^{\beta_2} \ldots u_{n_u}^{\beta_{n_u}}$ with $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_{n_u} \in \mathbb{N}$ and $\sum_j \alpha_j + \sum_i \beta_i \leq d$. Herein, $d \in \mathbb{N}$ is called nonlinear degree and has to be chosen by the user. The n state equations and n_y output equations of a linear state-space system are extended by adding a polynomial to every equation. The major advantage of the PNLSS model is its capability of describing a very large class of nonlinear systems, such as bilinear models, affine models, nonlinear models with only nonlinearities in the states, nonlinear models with only nonlinearities in the input and certain block-structured nonlinear models (Wiener, Hammerstein, Wiener–Hammerstein and nonlinear feedback) (Paduart, 2008). In this reference, the model has been successfully used on several application examples. Consequently, it can be stated that the PNLSS model (2) is a generic "all-purpose" black-box model (although its approximation capabilities are quite large, e.g. nonsmooth state evolutions can not be adequately represented). One drawback is that, in practice, when a full parameterisation is used, with a high nonlinear degree, the number of parameters grows very large. Current research focuses on reducing the amount of parameters by means of similarity transforms.

2.2. Parameterization

Despite the ease with which the state space model structure can handle MIMO systems, we will restrict ourselves without loss of generality to single-input single-output (SISO) systems ($n_u = n_y = 1$) in order to focus on the main topic of this paper.

Define $\theta \in \mathbb{R}^{n_\theta}$ as a vector containing all the model parameters:

$$\theta^T = \begin{bmatrix} vec(A)^T & B^T & C & D & vec(E)^T & F \end{bmatrix} \tag{3}$$

with vec an operator that stacks the columns of a matrix onto each other. Since all model parameters are included, the model is overparameterised. This is a consequence of similarity transforms on the states that do not influence the input–output behaviour. Both linear and nonlinear transforms can exist. This problem is taken care of in methods A and B, respectively by means of a pseudo-inverse and some kind of Levenberg–Marquardt term. Other approaches exist, such as the use of canonical forms or Data Driven Local Coordinates (DDLC). The latter approach avoids the numerical ill-conditioning of the estimation problem in the case of a canonical parameterisation (McKelvey, Helmersson, & Ribarits, 2004). The DDLC approach has in fact been proven to be equivalent to the pseudo-inverse (as is used in method A) (Wills & Ninness, 2008). The advantage of the pseudo-inverse is that it can easily be implemented, while the DDLC method is up to now only feasible for linear, bilinear or LPV state space models. The disadvantage is that all parameters need to be identified, although this is less important for nonlinear models with many parameters, since the relative gain is then small. It has been shown (Pintelon, Schoukens, McKelvey, & Rolain, 1996) that the choice of parameterisation does not affect the stochastic properties (i.e. the minimum variance bounds).

2.3. Stochastic framework

The input is assumed to be known exactly (without noise). If the model is capable of describing the system, the output measurements y_m are related to the system output $y(t, \theta_0)$:

$$y_m(t) = y(t, \theta_0) + v(t) \tag{4}$$

with θ_0 the true parameter values and $v(t)$ the output measurement noise, which is here for simplicity assumed to be white Gaussian, zero mean and with finite variance.

Under these assumptions, the least-squares estimator corresponds to the maximum-likelihood estimator, which is asymptotically consistent, efficient and normally distributed (Kendall & Stuart, 1979). The noise condition can be relaxed to filtered white noise with existing second and fourth order moments.

3. Initial estimates

In both cases, the initial estimates for the linear parameters (A, B, C, D) are found by a two-step procedure (Paduart, 2008).

We choose to use a frequency domain approach. This has two advantages: bounded initial estimates can also be obtained for unstable systems and nonparametric weighting is easier than in the time domain. This robustifies the method significantly.

First, the Best Linear Approximation (BLA) of the system is estimated (Pintelon & Schoukens, 2001)

$$\hat{G}_{BLA}(k) = \frac{\hat{S}_{YU}(k)}{\hat{S}_{UU}(k)} \tag{5}$$

with $\hat{G}_{BLA}$ the estimated frequency response function, k the frequency line, $\hat{S}_{YU}$ the estimated cross-power spectrum between output and input and with $\hat{S}_{UU}$ the estimated auto-power spectrum of the input. The BLA minimizes the output error in least squares sense. Also the variance $\hat{\sigma}^2_{\hat{G}_{BLA}}(k)$ can be estimated to enhance the second step by using a weighted least-squares method.

This first step offers a number of advantages: the signal to noise ratio (SNR) is enhanced, the user can select – in a straightforward way – a frequency band of interest and, when periodic data are available, the measurement noise and the effect of the nonlinear behaviour can be separated. If the total variance lies close to the noise variance (i.e. the nonlinear variance is small), a linear model is sufficient, otherwise, a nonlinear model is needed.

The second step is to convert this nonparametric model into a linear parametric state-space model using the Frequency Domain Subspace identification method (McKelvey, Akçay, & Ljung, 1996; Pintelon, 2002).