



CHALMERS

Chalmers Publication Library

Separate initialization of dynamics and nonlinearities in nonlinear state-space models

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

2012 IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2012, Graz, 13-16 May 2012 (ISSN: 1091-5281)

Citation for the published paper:

Marconato, A. ; Sjöberg, J. ; Suykens, J. (2012) "Separate initialization of dynamics and nonlinearities in nonlinear state-space models". 2012 IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2012, Graz, 13-16 May 2012 pp. 2104-2108.

<http://dx.doi.org/10.1109/I2MTC.2012.6229167>

Downloaded from: <http://publications.lib.chalmers.se/publication/163530>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Separate Initialization of Dynamics and Nonlinearities in Nonlinear State-Space Models

Anna Marconato*, Jonas Sjöberg[†], Johan Suykens[‡], Johan Schoukens*

*Dept. ELEC, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

Email: anna.marconato@vub.ac.be

[†]Dept. of Signals and Systems, Chalmers University of Technology, SE412 96 Gothenburg, Sweden

[‡]Dept. ESAT-SCD-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven (Heverlee), Belgium

Abstract—This work focuses on the identification of nonlinear dynamic systems. In particular the problem of obtaining good starting values for the identification of nonlinear state-space models is addressed. A fast and efficient initialization algorithm is proposed, combining the use of methods from the statistical learning community to model the nonlinearities and classic system identification tools to capture system dynamics. The performance of the method is evaluated on simulation examples.

I. INTRODUCTION

Most real-life systems are characterised by a nonlinear behavior, and very often one should also be able to model dynamics to get a thorough description of the system. Hence, modeling of nonlinear dynamical systems represents an interesting challenge in system identification.

In this work the problem is addressed by considering nonlinear state-space models (in the discrete time domain) of the form:

$$x(t+1) = f(x(t), u(t)) \quad (1)$$

$$y(t) = g(x(t)) \quad (2)$$

where $u(t) \in \mathbb{R}^{n_u}$ and $y(t) \in \mathbb{R}^{n_y}$ are the given input and output signal vectors at time instant t , $x(t) \in \mathbb{R}^{n_x}$ is the unknown state vector of the system, and $f(\cdot)$ and $g(\cdot)$ are the nonlinear functions to be estimated.

Among the different model classes that can be used, state-space models exhibit some nice properties which make them often the preferred choice: they are general model structures that allow one to naturally describe system dynamics, and they are particularly suited for Multiple Input Multiple Output (MIMO) systems. Although state-space models have been extensively studied and employed in the context of linear system identification [1], [2], the identification of nonlinear state-space models is a far more complex task.

Nonlinear functions f and g are characterized by a number of parameters that need to be optimized, by minimizing a given criterion. If the problem is nonlinear in the parameters, an iterative search for the cost function minimum is performed, e.g. using a Levenberg-Marquardt technique [1], [2]. Starting from initial parameters, in each iteration several simulations of (1-2) and their derivatives need to be done, resulting in high time costs. The main idea of this work is to transform

the dynamic identification problem into a nonlinear regression problem that can be solved much more efficiently.

The statistical learning community provides us with methods that can be used to identify nonlinear systems, see e.g. [3] and [4]. In that setting, the nonlinear regression problem to be solved is the following:

$$z(t) = h(\varphi(t)) \quad (3)$$

where $\varphi(t)$ and $z(t)$ are the input and output signals respectively, and $h(\cdot)$ is the nonlinear function to be estimated. Notice that $\varphi(t)$ is available from the beginning of the estimation, while the state $x(t)$ in (1-2) is unknown.

The approach proposed in this paper is based on a combination of ideas from statistical learning used to solve nonlinear regression problems on one hand, and methods to handle dynamics from the system identification community on the other hand [2]. More in particular an approximate version of the problem is obtained, by getting an approximation of the nonlinear state $x(t)$ to cut the recursion loop in equation (1), in order to be able to estimate the nonlinear functions f and g individually as static mappings, to bring one back to the statistical learning problem setting in Eq. (3).

By applying this procedure, the aim is to generate good starting values for the model parameters that need to be optimized, in order to reduce the time to convergence and possibly to avoid to get stuck in bad local minima. Moreover, since by applying regression methods the initialization procedure is speeded up, one can test more efficiently many choices for the nonlinear functions f and g .

Examples of related methods are present in the literature (e.g. [5]–[13]), including a combination of kernel canonical correlation analysis (KCCA) and Least Squares Support Vector Machines [5], neural state-space models [6], subspace identification techniques [7], and algorithms based on Expectation-Maximization [10]–[12]. Although these approaches seem to be successful in specific applications, in this paper we address the identification of nonlinear state-space models combining the use of NNs with more classic system identification techniques to capture system dynamics, by keeping the estimation of the nonlinearities separated from the estimation of the linear dynamics.

This paper is organized as follows. The considered problem is presented in more details in Section II, while the different

steps of the proposed approach are described in Section III. The proposed method is illustrated by means of a simulation example (Section IV), and concluding remarks are provided in Section V.

II. PROBLEM DESCRIPTION

The nonlinear dynamics in Eqs. (1-2) are assumed to be modeled as:

$$f(x(t), u(t)) = Ax(t) + Bu(t) + f_{NL}(x(t), u(t)) \quad (4)$$

$$g(x(t)) = Cx(t) + g_{NL}(x(t)) \quad (5)$$

where $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$ and $f_{NL}(\cdot)$ and $g_{NL}(\cdot)$ have n_x and n_y outputs respectively.

In this way, once the Best Linear Approximation (BLA) is estimated (that is, once an initial estimate for the matrices A , B and C is obtained, see next section for more details), only the deviation from the linear model needs to be modeled, which has been proposed by several authors, see e.g. [14] and [15]. Notice that the proposed approach targets systems that are weakly nonlinear (the linear part is considered to be dominant if compared with the nonlinear part) and systems that are assumed to have only one equilibrium point.

By obtaining approximate nonlinear states (based on the BLA and on the available data) the estimation problem is transformed into a nonlinear regression problem of the form (3), and basically any method from the statistical learning community can be used to obtain preliminary estimates of f_{NL} and g_{NL} . To illustrate the idea, multilayer perceptrons (MLP) with one hidden layer are used, i.e. the nonlinear terms f_{NL} and g_{NL} are described as weighted sums of sigmoid functions of the form $s(a) = \frac{1}{1+e^{-a}}$ [3], with parameters characterizing their center position, width and amplitude. This is a general way of describing the nonlinearities in the model, which allows one to extend the method to include other possibilities for the basis function [16].

In the classic (nonlinear) system identification framework, on the basis of a set of N input/output measurement data $\{u(t), y(t)\}_{t=1}^N$, one can build a model characterised by a vector of parameters θ to describe the behavior of the underlying system. The obtained model can then be used to predict the output values $\hat{y}(t, \theta)$. In our case, beside the parameters characterizing the linear part of the model, the parameters of the sigmoid functions mentioned above need to be estimated.

Following the Least Squares approach, optimal values of θ are found that minimize a least squares cost function V , typically the mean square error of the modeled outputs (on the basis of past input/output measurements) with respect to the true output values:

$$\theta_{opt} = \arg \min_{\theta} V(\theta)$$

where

$$V(\theta) = \frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t, \theta))^2$$

Since the resulting problem is nonlinear in the parameters, a numerical optimization is needed.

A typical difficulty that is encountered when minimizing the cost function $V(\theta)$ is the presence of a number of local minima in which the search algorithm may get trapped. Therefore, choosing the starting values for parameters θ represents a crucial issue, since the initialization step has a big impact both on the quality of the final solution and on the time required for convergence. The goal of this work is to obtain better starting values, by employing an initialization procedure that combines system identification techniques to model the dynamics of the system and NNs as a regression method to estimate the nonlinearities. In this way, when fitting the parameters of model (1-2), one hopes to end up in a good (local) minimum to increase the quality of the final solution.

III. PROPOSED ALGORITHM

The proposed scheme for the initialization of nonlinear state-space models consists of three main steps:

- obtain a linear model to capture the dynamics of the system;
- estimate the nonlinear states;
- model the nonlinearities.

In this section all the different steps are described in details.

A. Obtain a linear model

First of all, the nonlinear input-output behavior is approximated with a linear model, by estimating the BLA [2]. Among the possible choices of linear models that one can use, the BLA is defined to be optimal in least square sense. More in details, in the set of linear models \mathcal{G} , the BLA is defined as the model G such that:

$$G_{BLA} = \arg \min_{G \in \mathcal{G}} \{ |y(t) - G(u(t))|^2 \}$$

where $u(t)$ and $y(t)$ are the input and output of the nonlinear system [2], [17]. In this way matrices \hat{A} , \hat{B} and \hat{C} can be determined, obtaining the following linear model:

$$x(t+1) = \hat{A}x(t) + \hat{B}u(t) \quad (6)$$

$$y(t) = \hat{C}x(t) \quad (7)$$

The linear model can then be used to get an approximation of the nonlinear states, as discussed in the next paragraph.

B. Estimate \hat{x}_{LS}

A main point in the proposed approach is based on the fact that if the states $x(t)$ would be exactly known, the problem of obtaining a nonlinear model could be solved much more easily by estimating f and g individually and as static mappings. Since the nonlinear states are in practice not available, one would like to obtain an approximation of $x(t)$, to be able to obtain initial estimates of f and g .

In particular, using the available data $\{u(t), y(t)\}_{t=1}^N$ and the BLA estimates \hat{A} , \hat{B} , \hat{C} obtained in the previous step, the nonlinear states are approximated as a trade-off between the linear model and data fit, by solving the following Least Squares problem:

$$\hat{x}_{LS}(t) = \arg \min_{\{x(t)\}} \sum_t (y(t) - \hat{C}x(t))^2 + \lambda \sum_t (x(t+1) - \hat{A}x(t) - \hat{B}u(t))^2 \quad (8)$$

The first term of the cost function in Eq. (8) represents the data fit, while the second term represents the linear model fit; λ is the trade-off parameter that needs to be tuned to change the emphasis given on the two criteria. By tuning λ a deviation from the linear states (resulting from the BLA estimates \hat{A} , \hat{B} , \hat{C}) is allowed, to take into account the nonlinear terms in Eqs. (4-5). Problem (8) could be replaced by a Kalman filter which would change the approximation slightly.

C. Estimate nonlinear functions f and g

Once the nonlinear states $\{\hat{x}_{LS}(t)\}_{t=1}^N$ are obtained, regression methods are employed to solve the following approximate static problem:

$$\hat{x}_{LS}(t+1) = f(\hat{x}_{LS}(t), u(t)) + r_{LS}(t) = \hat{A}\hat{x}_{LS}(t) + \hat{B}u(t) + f_{NL}(\hat{x}_{LS}(t), u(t)) + r_{LS}(t) \quad (9)$$

$$y(t) = g(\hat{x}_{LS}(t)) + e_{LS}(t) = \hat{C}\hat{x}_{LS}(t) + g_{NL}(\hat{x}_{LS}(t)) + e_{LS}(t) \quad (10)$$

where $r_{LS}(t)$ and $e_{LS}(t)$ are error terms resulting from the fact that here the approximated nonlinear states are introduced in the problem. By solving this nonlinear regression problem without recursion, one can estimate both functions $f(\hat{x}_{LS}(t), u(t))$ and $g(\hat{x}_{LS}(t))$ using the methods of the statistical learning community. As already mentioned, in this work a specific type of nonlinearity is used, namely one-hidden-layer sigmoidal MLPs, since they are proved to be “universal approximators”. This means that such networks can approximate any continuous function arbitrary well [6], and their use seems therefore appealing for our purposes, although local minima issues may pop up when using NNs.

Model structure

The model structure represented in Fig. 1 will be used. In Fig. 1(a) the representation of state equation (1) is given in the case of having one input, one output and two states. In the block scheme the three inputs are the states $\hat{x}_{LS}^1(t)$, $\hat{x}_{LS}^2(t)$ and the input $u(t)$, while the two outputs are the states $\hat{x}_{LS}^1(t+1)$, $\hat{x}_{LS}^2(t+1)$. The sigmoidal MLP block is added in parallel to the linear model, so that the update for both states consists of a linear plus a nonlinear part, see Eq. (4). The same structure is used to describe the output equation (2) in Fig. 1(b).

The MLP blocks, namely f_{NL} and g_{NL} in Eqs. (4-5), consist of the sum of a number of sigmoids, each one characterized by three types of parameters: center position, width and amplitude. These parameters are estimated together with the ones defining the linear model, using the input-output data $\{u(t), y(t)\}_{t=1}^N$ and the nonlinear states $\{\hat{x}_{LS}(t)\}_{t=1}^N$.

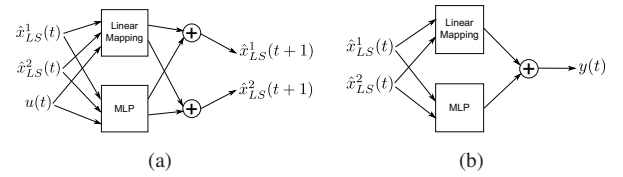


Fig. 1. Model structure used to describe (a) the state equation, (b) the output equation, in the case of one input, one output and two states. The MLP block is added in parallel to the linear model.

Simulation of the initialized nonlinear model

The two estimated blocks \hat{f}_{NL} and \hat{g}_{NL} can then be included in a general nonlinear state-space structure; at this point the dynamics are again taken into account, and one can simulate the obtained initialized model to assess its performance. In other words, the recursion in the state equation is included again, switching back from the approximate initialization obtained using (9-10) to (4-5), which is the problem one wants to solve.

Finally, the obtained initial estimate of the nonlinear model can be further fitted to data.

Time saving

One of the advantages of employing the initialization technique presented in this paper rather than a linear initialization scheme is related to the reduction of time needed for the estimation of the nonlinear model.

In particular, the proposed method allows one to reduce the computational time at different stages of the algorithm, thanks to the fact that f and g are estimated separately. For instance, a considerable amount of time can be saved during model selection since many different choices for the nonlinearities can be tested independently.

IV. SIMULATION EXAMPLES

In this section the proposed initialization technique is applied on a simulation example, which is useful to evaluate the pros and cons of the method. In particular, the case in which a nonlinear term is present only in the state equation is considered, i.e. f is nonlinear and g is linear.

A. Simulation settings

The considered nonlinearity is a sum of sigmoid functions, i.e. the same model structure that has already been described in Section III-C is used.

One input, one output and two states are considered. Data were generated by considering a normally distributed input signal characterized by zero mean and unit variance. Two data records of 2000 samples each were generated, one used for estimation and one for validation purposes. In this simple example, no noise was added to the output. Matrices A, B, C were chosen as follows:

$$A = \begin{pmatrix} 0.4 & -0.2 \\ -0.1 & 0.5 \end{pmatrix} \quad B = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad C = (2 \quad 1)$$

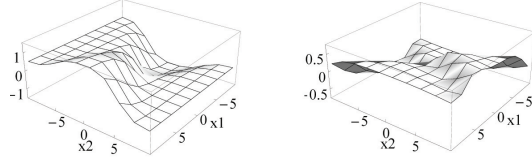


Fig. 2. Sum of sigmoids used for the two states. In both cases the sum of three different sigmoid functions is considered.

An example of a weakly nonlinear system is considered, in particular for each state a sum of three different sigmoids was added to the linear part in the state equation.

The resulting nonlinear terms for the two states are shown in Fig. 2.

B. Simulation results

In the following the results obtained for the considered simulation problem will be discussed. In particular, the gain in performance (e.g. in terms of RMSE) of the proposed initialization technique will be evaluated when compared with a linear initialization. RMSE results both after the initialization phase, and after the fitting of all parameters will be shown.

Different ‘realizations’ of the sigmoidal MLP are taken into account, meaning that the initialization scheme was repeated several times, each time with a different initial position of the sigmoids. The initial center position parameters of the sigmoids are chosen randomly in the domain of the data, so by looking at different initializations a statistical overview on the results is obtained.

In this simple example our input/output data were generated by considering a nonlinear f and a linear g . However, since this information is not available *a priori*, both f and g will be estimated as nonlinear functions with the following structure: a linear block plus a sum of sigmoids.

The results on the validation set obtained when comparing the proposed initialization technique with the linear initialization will now be discussed. Fig. 3 depicts the improvement in terms of RMSE after the initialization phase. Since no information on the nonlinear terms was available, the use of ten sigmoids for f and five sigmoids for g was empirically chosen. 50 different initializations of the MLPs (50 different center positions of the sigmoid functions) are taken into account, which explains the variability of the results. Instead, for the linear initialization, only one RMSE value is obtained.

One can see that, by employing the proposed method, the RMSE values of the initialized models are approximately 50% lower than in the case of an initialized linear model.

Next, the performance of the models obtained when fitting all parameters by means of an iterative Levenberg-Marquardt procedure is compared in Fig. 4. RMSE values of the final fitted models obtained starting from the proposed initialization and starting from the linear initialization are shown.

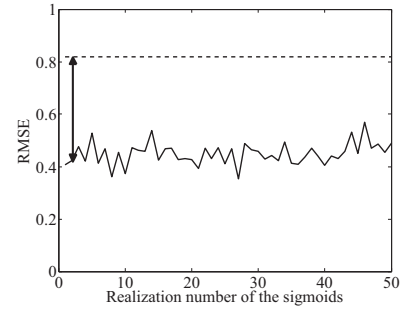


Fig. 3. RMSE values (on validation data) of the initialized models, with the linear initialization (dashed line) and with the proposed approach (solid line). 50 different initializations of the sigmoids are considered.

The proposed initialization scheme gives significantly better results for the 20 best models characterized by low RMSE values (the best model shows an improvement of three orders of magnitude if compared with the linear initialization).

Notice that the RMSE values shown here for the validation set are almost identical to the ones obtained on the estimation data, indicating not only that the proposed approach yields a significant improvement in terms of error reduction, but also that the obtained nonlinear models can generalize very well on previously unseen data.

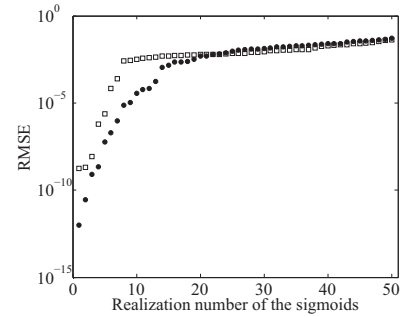


Fig. 4. RMSE values (on validation data) of the final fitted models, obtained by starting with the proposed initialization technique (black circles) and with the linear initialization (white squares), for 50 different initializations of the sigmoids. The RMSE values (in log scale) are sorted in increasing order.

V. CONCLUSION

In this work a novel initialization scheme for the identification of nonlinear state-space models has been presented. The approach was successfully applied on simulation examples. The proposed initialization procedure has several advantages, as (i) the separation between system dynamics and nonlinear terms makes it possible to identify them independently; (ii) many nonlinear model structures can be tested rapidly on the obtained regression problem; (iii) two different fields - system identification and statistical learning - are brought together, combining the advantages of both.

ACKNOWLEDGMENT

This work is sponsored by the Fund for Scientific Research (FWO/Vlaanderen), the Flemish Government (Methusalem

Fund, METH1) and the Belgian Federal Government (IAP VI/4). J. Suykens acknowledges support from K.U.Leuven, the Flemish Government, FWO and the Belgian federal science policy office (CoE EF/05/006, GOA MANET, IUAP DYSCO).

REFERENCES

- [1] L. Ljung, *System Identification: Theory for the User (2nd ed.)*. Prentice Hall, New Jersey, 1999.
- [2] R. Pintelon and J. Schoukens, *System Identification. A Frequency Domain Approach*. IEEE Press, 2001.
- [3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2009.
- [4] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.
- [5] V. Verdult, J. Suykens, J. Boets, I. Goethals, and B. De Moor, "Least squares support vector machines for kernel CCA in nonlinear state-space identification," in *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2004)*, Leuven, Belgium, 2004.
- [6] J. Suykens, J. Vandewalle, and B. De Moor, *Artificial Neural Networks for Modeling and Control of Non-Linear Systems*. Springer, 1996.
- [7] J. van Wingerden and M. Verhaegen, "Closed-loop subspace identification of Hammerstein-Wiener models," in *Proceedings of the 48th IEEE Conference on Decision and Control (CDC 2009)*, Shanghai, China, 2009.
- [8] K. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.
- [9] J. Codina, J. Aguado, and J. Fuertes, "Neural network identification of non linear systems using state space techniques," in *IMACS International Symposium on Signal Processing, Robotics and Neural Networks (SPRANN'94)*, Lille, France, 1994.
- [10] Z. Ghahramani and S. Roweis, "Learning nonlinear dynamical systems using an EM algorithm," *Advances in Neural Information Processing Systems*, vol. 11, 1999.
- [11] A. Gorji and M. Menhaj, "Identification of nonlinear state space models using an MLP network trained by the EM algorithm," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2008), part of the IEEE World Congress on Computational Intelligence (WCCI 2008)*, Hong Kong, China, 2008.
- [12] T. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, 2011.
- [13] B. Ninness, A. Wills, and T. Schön, "Estimation of general nonlinear state-space systems," in *Proceedings of the 49th IEEE Conference on Decision and Control (CDC 2010)*, Atlanta, GA, USA, 2010.
- [14] J. Sjöberg, "On estimation of nonlinear black-box models: How to obtain a good initialization," in *IEEE Workshop on Neural Networks for Signal Processing*, Amelia Island Plantation, FL, USA, 1997.
- [15] J. Paduart, L. Lauwers, J. Swevers, K. Smolders, J. Schoukens, and R. Pintelon, "Identification of nonlinear systems using polynomial nonlinear state space models," *Automatica*, vol. 46, no. 4, pp. 647–656, 2010.
- [16] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorenec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.
- [17] M. Enqvist, *Linear Models of Nonlinear Systems*. PhD Thesis, Linköping University, Sweden, 2005.