Improved Initialization for Nonlinear State-Space Modeling

Anna Marconato, *Member, IEEE*, Jonas Sjöberg, Johan A. K. Suykens, *Senior Member, IEEE*, and Johan Schoukens, *Fellow, IEEE*

***Abstract*— This paper discusses a novel initialization algorithm for the estimation of nonlinear state-space models. Good initial values for the model parameters are obtained by identifying separately the linear dynamics and the nonlinear terms in the model. In particular, the nonlinear dynamic problem is transformed into an approximate static formulation, and simple regression methods are applied to obtain the solution in a fast and efficient way. The proposed method is validated by means of two measurement examples: the Wiener–Hammerstein benchmark problem and the identification of a crystal detector.**

***Index Terms*— Multilayer perceptrons, nonlinear dynamical systems, nonlinear modeling, state-space models, system identification.**

I. Introduction

IN THE instrumentation and measurement community, there is an ever-increasing demand for good models. Measuring and modeling are in fact closely linked together because modeling techniques rely on raw data to build an analytical description of the system from which those data are generated. One can think of a very well-known example in electrical engineering, Ohm's law; this simple model explains the relation between two measured quantities, the voltage difference at two points of a conductor and the current flowing between the two points.

Having a look at more recent applications, accurate models are needed to tackle many problems in electrical engineering: to describe the behavior of sensors [1], to characterize $\Delta\Sigma$ modulators and study their stability [2], to better understand the functioning of wireless transmitters [3], for fault diagnosis in analog circuits [4], just to mention a few examples.

Moreover, good models are very important in all branches of engineering, since they can be used to test a device in the design phase, to develop control units for industrial processes, and in general to replace expensive and time-consuming experiments in the analysis and simulation of a given system.

Manuscript received June 3, 2013; revised August 19, 2013; accepted August 21, 2013. Date of publication October 9, 2013; date of current version March 6, 2014. This work was supported in part by the Fund for Scientific Research, in part by the Flemish Government, in part by the Belgian Government through the Inter University Poles of Attraction Program, and in part by the ERC Advanced Grant SNLSID. The work of J. Suykens was supported in part by Katholieke Universiteit Leuven, in part by the Flemish government, in part by FWO, in part by the Belgian Federal Science Policy Office, and in part by the European Research Council. The Associate Editor coordinating the review process was Dr. Dario Petri.

A. Marconato and J. Schoukens are with Department ELEC, Vrije Universiteit Brussel, Brussels 1050, Belgium (e-mail: anna.marconato@vub.ac.be).

J. Sjöberg is with the Department of Signals and Systems, Chalmers University of Technology, Göteborg SE-412 96, Sweden.

J. Suykens is with the Department ESAT-SCD-SISTA, Katholieke Universiteit Leuven, Leuven 3001, Belgium.

Digital Object Identifier 10.1109/TIM.2013.2283553

Most real life systems (like the examples given above) are, to a certain extent, nonlinear. In particular, to get a thorough description of a given system, one needs to be able to model the dynamics, as well as the (static) nonlinear behavior.

To address the challenge of identifying nonlinear dynamic systems, one can consider nonlinear state-space (NLSS) models (here expressed in the discrete time domain)

$$x(t+1) = f(x(t), u(t)) \tag{1}$$

$$y(t) = g(x(t), u(t)) \tag{2}$$

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 show some advantages: 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 [5], [6], the identification of NLSS models is a far more complex task. In particular, a difficulty that is encountered when estimating NLSS models relates to the nature of the optimization problem. Nonlinear functions f and g are in general 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 [5], [6].

Therefore, as for all iterative nonlinear optimization schemes, good initial values for the model parameters are needed, so that the time to convergence is reduced, and the problem of getting stuck in bad local minima is avoided. A possibility, often used in practice, is to initialize the nonlinear model with a simple linear description. However, a linear initialization does not always allow one to converge to a good local minimum, and, therefore, more advanced techniques are needed.

The goal of this paper is to propose a nonlinear initialization algorithm for the estimation of NLSS models.

As will be explained in more detail in the next sections, good initial values are obtained by transforming the nonlinear dynamic identification problem in (1) and (2) into a general nonlinear regression problem

$$z(t) = h(\xi(t)) \tag{3}$$

0018-9456 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

where $\xi(t)$ and $z(t)$ are the input and output signals, respectively, and $h(\cdot)$ is the nonlinear function to be estimated, which is in general expressed as a basis function expansion [7]. This setting encompasses many different choices for the nonlinear basis functions: polynomials, splines, radial basis functions (RBFs), sigmoids, and so on.

Note that, since by applying regression methods the initialization procedure is speeded up considerably, different model structures for f and g can be tested more efficiently.

The main contributions of this paper are:

1) obtaining good starting values for the optimization of NLSS models;
2) identifying separately the linear dynamics and the nonlinear terms in the model;
3) cutting the recursion in (1) to obtain an approximate (static) version of the estimation problem, to which efficient regression methods can be applied.

This paper is organized as follows: the considered problem is presented in more detail in Section II, together with an overview on the state-of-the-art. The different steps of the proposed identification algorithm are described in Section III. The proposed method is validated by means of two challenging measurement problems: the Wiener–Hammerstein benchmark problem, and the identification of a crystal detector. The first example will be considered to illustrate the different aspects of the proposed approach in Section IV. The second real data problem is discussed in Section V, together with a comparison of the obtained results with the ones given by other state-of-the-art methods. Concluding remarks are provided in Section VI.

II. Problem Description

The nonlinear dynamics in (1) and (2) are assumed to be modeled as

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

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

where $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$, $D \in \mathbb{R}^{n_y \times n_u}$ and $f_{\mathrm{NL}}(\cdot)$ and $g_{\mathrm{NL}}(\cdot)$ 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, C, and D is obtained; see Section III for more details), only the deviation from the linear model needs to be modeled, which has been proposed by several authors; see [8] and [9]. Note that the proposed approach targets systems for which the dynamics can be captured by the BLA and systems that are assumed to have only one equilibrium point.

The separation between the identification of the linear dynamics on one hand, and the estimation of the nonlinear terms on the other hand allows one to employ *ad hoc* techniques specifically conceived to address the two different tasks, and to combine the advantages of both worlds. In particular, available linear system identification techniques (in the time or in the frequency domain) are used to determine the BLA, and nonlinear regression algorithms are employed to model the static nonlinearities [10].

An estimate of the (unknown) nonlinear state is first computed, to cut the recursion loop in (1). Due to this step, the NLSS estimation problem is transformed into a nonlinear regression problem of the form (3).

Note that typically the nonlinear function $h(\cdot)$ is represented in terms of basis functions, for which one can choose among a large variety of possibilities: polynomials, splines, piecewise linear functions, RBFs, sigmoids, and in general any kind of static nonlinear function that one can think of. The formulation of the problem that is proposed in this paper to obtain an initial estimate of the nonlinear terms f_{NL} and g_{NL} in (4) and (5) is general enough to encompass many model structures from statistical learning. Support vector machines (SVMs), least squares (LS) SVMs, and neural network (NN) paradigms, such as RBF networks, multilayer perceptrons (MLPs), extreme learning machines, and so on, are all examples included in this general model class.

In the two measurement examples discussed in this paper, MLPs are considered to estimate the nonlinear terms in the state-space model to provide a simple illustration of the proposed algorithm, but the same approach could be followed employing a different nonlinear model structure.

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 characterized 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 $\hat{y}(t, \theta)$. For the algorithm presented here, beside the parameters characterizing the linear part of the model, the parameters appearing in the basis function expansion of the nonlinear terms mentioned above need to be estimated. Therefore, here θ contains both sets of parameters.

Following the LS approach, optimal values of θ are found that minimize a LS cost function V, typically the mean square error of the modeled outputs with respect to the true output values

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

where

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

Since in most situations, 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 paper is to obtain good initial values of the parameter vector θ, by combining system identification techniques to model the dynamics of the system and regression methods to estimate the nonlinearities. In this way, when fitting the parameters of model (1) and (2), one hopes to end up in a good (local) minimum to increase the quality of the final solution.

A. Related Work

An example of NLSS models that have been successfully applied to identify several real life systems is given by polynomial NLSS (PNLSS) models [9]. In that approach, the nonlinear model is initialized simply by a linear model (BLA).

A different approach is followed in [11], where an NN architecture characterized by a state-space structure is proposed and compared with the system input–output structure. Sines and cosines are used as activation functions in the network.

State-space models parametrized as feedforward NNs with one hidden layer and tanh(·) as activation function are presented in [12] and [13]. For these neural state-space models, prediction error learning algorithms are discussed. In particular, as an alternative to the extended Kalman filter, the Kalman gain is directly parametrized to obtain simpler expressions. In this way, the gradient of the cost function can be computed in a straightforward manner by following Narendra's sensitivity model approach. To deal with the local minima problems, two different heuristics are proposed: the initialization of nonlinear neural state-space models as linear state-space models, and learning complex models starting from less complex ones, by increasing the number of hidden neurons in the model.

To make the recurrent nature of the problem vanish, in [14] a method based on kernel canonical correlation analysis (KCCA) is employed to build the state sequence. The obtained state is then used together with the available input–output data to estimate the state-space model. Thanks to a LS SVMs approach, a form of regularization is embedded within KCCA.

The use of KCCA is suggested also in [15] for the identification of nonlinear MIMO Hammerstein–Wiener systems. A subspace identification technique is used to estimate the nonlinearities separately from system dynamics.

Other approaches for the identification of NLSS models include the method proposed in [16], where knowledge on the state sequence is assumed to be available and is used to build the data records.

The expectation–maximization (EM) algorithm for maximum likelihood parameter estimation is instead used in [17], where extended Kalman smoothing is employed in the expectation step to get an estimate of the state.

More recently, a combination of multilayer perceptron NNs, the EM algorithm and particle smoothing is suggested in [18] for joint parameter and state estimation.

Other examples of techniques that employ the EM algorithm in a maximum likelihood framework, and that use particle smoothing are presented in [19] and [20]. The same authors propose in [21] a different approach that makes use of Fisher's identity and particle smoothing for gradient computation.

Gaussian processes are used in [22] for inference and learning, again using EM.

B. Advantages and Novelty of the Proposed Method

The identification method presented in this paper for the estimation of NLSS models, and more specifically the proposed initialization algorithm, shows a number of differences and novel aspects compared with the other approaches mentioned above:

1) the separation between the linear dynamics and the static nonlinearities in the formulation of the problem allows one to identify the different parts independently, in a more effective way;
2) the initialization scheme is based on a combination of ideas coming from different worlds–system identification and nonlinear regression/statistical learning–exploiting the advantages of both;
3) the BLA is used to incorporate dynamics in methods that are essentially designed to model static nonlinearities;
4) the proposed algorithm is a general scheme that can be used in combination with many different choices of nonlinear model structures.

III. Proposed Algorithm

In this section, the proposed method for the estimation of NLSS models is presented. First of all, the initialization algorithm is explained, then the next steps of the method are discussed, followed by the description of the considered model structure. Finally, some comments on time saving issues are provided.

A. Initialization Scheme

The proposed scheme for the initialization of NLSS models consists of three main steps:

1) obtain a linear model to capture the dynamics of the system;
2) estimate the nonlinear state;
3) model the nonlinearities.

In this section, all the different steps are described in detail.

1) Obtain a Linear Model: First of all, the nonlinear input–output behavior is approximated with a linear model, by estimating the BLA [6]. Among the possible choices of linear models that one can use, the BLA is defined to be optimal in least square sense. More in detail, in the set of linear models $\mathscr{G}$, the BLA is defined as the model G such that

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

where $u(t)$ and $y(t)$ are the input and output of the nonlinear system and $\mathscr{E}$ is the expected value with respect to the input [6], [23]. In this way, matrices $\hat{A}$, $\hat{B}$, $\hat{C}$, and $\hat{D}$ can be determined, obtaining the following linear model:

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

$$y(t) = \hat{C}x(t) + \hat{D}u(t). \tag{7}$$

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

2) Estimate $\hat{x}_{\mathrm{LS}}$: A crucial aspect in the proposed approach is based on the fact that if the state $x(t)$ 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. As the nonlinear state is 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 $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{D}$ obtained in the previous step, the

nonlinear state is approximated as a tradeoff between the linear model and the data fit, by solving the following LS problem:

$$\begin{aligned}\hat{x}_{\mathrm{LS}}(t) &= \arg\min_{\{x(t)\}} \sum_t (y(t) - \hat{C}x(t) - \hat{D}u(t))^2 \\ &\quad + \lambda \sum_t (x(t+1) - \hat{A}x(t) - \hat{B}u(t))^2 \\ &= \arg\min_{\{x(t)\}} E_y + \lambda E_x. \end{aligned} \tag{8}$$

The first term E_y of the cost function represents the data fit, while the second term E_x represents the linear model fit; λ is the tradeoff parameter that needs to be tuned to change the emphasis given on the two criteria. By tuning λ, a deviation from the linear state (resulting from the BLA estimates $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{D}$) is allowed, considering the nonlinear terms in (4) and (5). In practice, the optimal value for the tradeoff parameter λ is chosen from a grid of values as the one minimizing the root-mean-square error (RMSE) (on the validation data) of the obtained initialized nonlinear models. The tuning of the tradeoff parameter λ will be illustrated in more detail on the Wiener–Hammerstein example in Section IV.

Problem (8) could be replaced by a Kalman filter that would change the approximation slightly [24].

3) Estimate Nonlinear Functions f and g: Once the estimate $\{\hat{x}_{\mathrm{LS}}(t)\}_{t=1}^{N}$ of the nonlinear state is available, one obtains the following approximate static problem:

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

$$\begin{aligned} y(t) &= g(\hat{x}_{\mathrm{LS}}(t), u(t)) + e_{\mathrm{LS}}(t) = \hat{C}\hat{x}_{\mathrm{LS}}(t) + \hat{D}u(t) \\ &\quad + g_{\mathrm{NL}}(\hat{x}_{\mathrm{LS}}(t), u(t)) + e_{\mathrm{LS}}(t) \end{aligned} \tag{10}$$

where $r_{\mathrm{LS}}(t)$ and $e_{\mathrm{LS}}(t)$ are error terms resulting from the fact that here the approximated nonlinear state is introduced in the problem. Equations (9) and (10) represent two static regression problems that can be solved independently employing simple regression methods. Note that at this stage, the recursion in the state equation is not present anymore, since the state sequence is now assumed to be "known". Therefore, both functions $f_{\mathrm{NL}}(\hat{x}_{\mathrm{LS}}(t), u(t))$ and $g_{\mathrm{NL}}(\hat{x}_{\mathrm{LS}}(t), u(t))$ can be estimated as basis function expansions.

B. Simulation of the Initialized Model and Optimization

The two estimated nonlinearities $\hat{f}_{\mathrm{NL}}$ and $\hat{g}_{\mathrm{NL}}$ can then be included in a general NLSS structure; at this point, the dynamics are again considered, and one can simulate the obtained initialized nonlinear 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) and (10) to (4) and (5), which is the original problem one wants to solve.

Finally, the obtained initial estimate of the nonlinear model can be further fitted to data, using an iterative nonlinear optimization routine, e.g., the Levenberg–Marquardt algorithm.

C. Model Structure

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

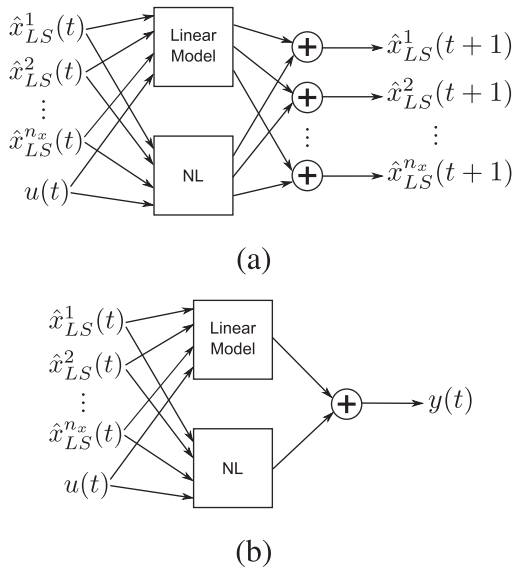


Fig. 1. Model structure. (a) State equation. (b) Output equation, in the case of one input, one output and n_x states. The nonlinear block is added in parallel to the linear model.

D. Time Saving

This paper presents a nonlinear initialization algorithm for the estimation of NLSS models. Another possibility, which is often used in system identification, to initialize a nonlinear model before the optimization step is to use a simple linear initialization. Note that by "linear initialization" we mean that a nonlinear model is estimated by starting from a linear model in parallel with nonlinear terms that are initialized to zero; all parameters (of both the linear and the nonlinear part) are then optimized [8].

As will be illustrated in the next section with the Wiener–Hammerstein example, a linear initialization does not always guarantee convergence to a good local minimum of the cost function. In most cases, when starting from a linear model, the optimization routine gets stuck in bad local minima.

Moreover, by starting from better initial estimates, the time to convergence needed when fitting the model parameters will be reduced (or, dually, given the same amount of computational time the final fitted model is likely to be characterized by lower error values).

Beside this advantage, there are also other aspects of the initialization algorithm presented in this paper that allow one to reduce the computational time.

First, time can be saved since one can choose to perform model (structure) selection already after the initialization phase. This is obviously not the case if a linear initialization

is applied, since all initialized linear models result in the same error value. This means that with the proposed scheme, if a number of the possible nonlinear models can be discarded after initialization, one would not need to optimize all initialized models to be able to determine their performance. This aspect will be discussed further in Section IV.

Secondly, here the initial estimates of the nonlinear functions f and g are computed separately, i.e., if many nonlinear model structures need to be tried out this approach allows one to carry out all estimations in much less time.

IV. Wiener–Hammerstein Benchmark Problem

We illustrate the different aspects of the proposed initialization algorithm for NLSS models on the Wiener–Hammerstein benchmark example. One hidden layer MLPs with $\tanh(\cdot)$ as neuron activation function are chosen to model the nonlinear terms. The considered model structure is then the one shown in Fig. 1, where the two nonlinear blocks are two MLPs.

During the training phase, the parameters of the MLP are typically first randomly initialized so that the active regions of the neurons span the input space, and then optimized, e.g., by means of the backpropagation algorithm [25]. Therefore, in the following many different initial MLP configurations are considered, to show also the effect of this variability source on the obtained model performance.

A. Description of the Data

The data are generated from a nonlinear electronic system characterized by a Wiener–Hammerstein structure. The two linear blocks are a third-order Chebyshev low-pass filter with 0.5 dB ripple and cutoff frequency at 4.4 kHz, and a third-order inverse Chebyshev low-pass filter with a -40 dB stopband from 5 kHz, respectively. The static nonlinearity is built using two resistors and a diode.

The goal is to identify a model to describe this nonlinear dynamic system, based on a set of real input/output measurements. A difficulty that pops up in the identification of this system is given by the fact that it is not straightforward to model the nonlinear part of the system, because this is not directly accessible from either the input or the output.

The data were collected by exciting the system with a filtered Gaussian input signal (cutoff frequency at 10 kHz). The sampling frequency is equal to 51.2 kHz. A very large dataset is available[1] (188 000 input/output samples) to estimate and validate the model, but in this paper shorter records are considered to reduce the computational load. To illustrate the different aspects of the proposed algorithm, an estimation set of 2 500 samples is considered to build the NLSS model, and a validation set of 10 000 samples is used to evaluate the performance of the obtained model on fresh data. A very large validation set is chosen to eliminate the stochastic variations in this step, so that the model quality is clearly visible. In real life applications, one would instead choose to use a larger portion of the available data to estimate the model.

[1]Download at http://tc.ifac-control.org/1/1/Data%20Repository/sysid-2009-wiener-hammerstein-benchmark.

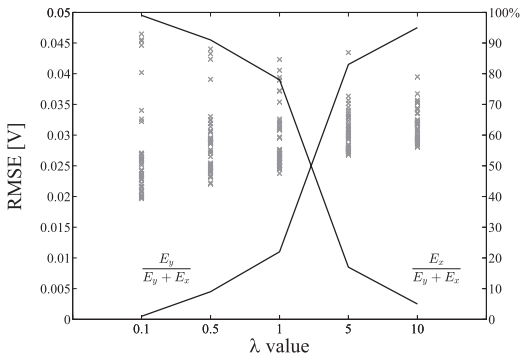


Fig. 2. Wiener–Hammerstein example. Initialized NLSS models (gray crosses): RMSE values on the validation set, for different values of λ (50 initial MLP configurations, number of neurons for f_{NL} and g_{NL} equal to 3). The relative weight (in percentage) of the two error criteria E_y and E_x is also shown (right side axis) for the different λ values.

B. Initialization

1) Estimate of the Linear Model: The BLA is computed to obtain a first (linear) description of the system. A sixth-order linear state-space model is obtained characterized by an RMSE value of 60 mV on the validation set.

2) Estimate of the Nonlinear State: Equation (8) is used to estimate the nonlinear state sequence. The main issue here concerns the tuning of the tradeoff parameter λ. First of all, one needs to determine a suitable range of values for λ. This is done by balancing the weight of the two criteria E_y and E_x in the tradeoff, i.e., by choosing several values of λ for which either E_y or E_x is dominant, or both are equally important. In this example, λ is chosen among these values: $0.1, 0.5, 1, 5, 10$. The effect of changing λ on the importance of the two criteria is shown in Fig. 2. The choice is made for example by selecting the value for which the RMSE of the resulting initialized NLSS model on the validation set is minimized.

The obtained results on the validation set are plotted in Fig. 2, for 50 different random initial configurations of the MLPs.

The lowest RMSE values are obtained for $\lambda = 0.1$, so if one decides to tune λ after the initialization phase, without optimizing all initialized models, this will be the chosen value. As will be shown later in the part dedicated to the optimization, the final performance of the optimized models does not seem to depend much on the specific value chosen for λ (see Fig. 3).

However, for each λ, the variability due to the different MLP configurations has a significant impact on the performance of the resulting NLSS models; so in general, one might need to initialize a number of different models, and check their performance once all parameters have been optimized. Note that this issue would not be present if a different choice of the basis functions in the estimation of the nonlinear terms was considered.

3) Estimate of the Nonlinear Terms: To complete the initialization step, the two nonlinear terms f_{NL} and g_{NL} need to be estimated, by solving the static regression problems in (9) and (10). As already mentioned, MLPs are used here to model the two nonlinearities, so for both f_{NL} and g_{NL} the

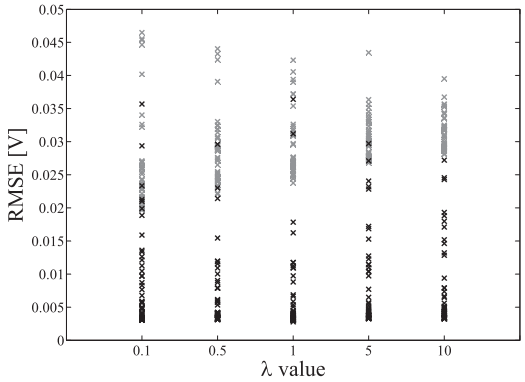


Fig. 3. Wiener–Hammerstein example. Optimized NLSS models (black crosses): RMSE values on the validation set, for different values of λ (50 initial MLP configurations, number of neurons for f_{NL} and g_{NL} equal to 3). The results for the initialized models are repeated here for a comparison (gray crosses).

TABLE I
WIENER–HAMMERSTEIN EXAMPLE. INITIALIZED NLSS MODELS: LOWEST RMSE VALUES OBTAINED ON THE VALIDATION SET FOR DIFFERENT VALUES OF n_f AND n_g ($\lambda = 0.1$)

$n_f = n_g$	1	2	3	4
RMSE [V]	0.0410	0.0256	0.0197	0.0196

TABLE II
WIENER–HAMMERSTEIN EXAMPLE. OPTIMIZED NLSS MODELS: RMSE VALUES ON THE VALIDATION SET AFTER OPTIMIZING THE FOUR INITIALIZED MODELS OF TABLE I FOR DIFFERENT VALUES OF n_f AND n_g

$n_f = n_g$	1	2	3	4
RMSE [V]	0.0129	0.0045	0.0034	0.0032

number of neurons in the hidden layer (n_f and n_g respectively) needs to be specified. Again, this is a model selection problem that can be solved at initialization.

Table I shows the RMSE (on the validation set) obtained using different values for n_f and n_g (for the sake of simplicity, here $n_f = n_g$). For each number of neurons, 20 different MLP configurations are considered, and the lowest RMSE value is shown in the table, to give an indication of how low the error can be pushed with models of increasing complexity.

At this stage, the model obtained with three neurons can be considered the best one, since the small gain in performance using four neurons does not justify the corresponding increase in the number of parameters. Table II shows that the same reasoning holds also for the optimized models resulting from these initializations.

C. Optimization

Although, as already pointed out, the choice of λ, n_f, and n_g can be done after the initialization step, all the initialized models considered above have been optimized, to illustrate the main properties of the proposed algorithm. The obtained results are summarized in Fig. 3 and Table II.

The main conclusion is that, for this example, the choice of λ is not too critical, since it does not affect much the final performance of the optimized models. It is rather the variability due to the different MLP configurations that can yield models with a higher RMSE (Fig. 3).

However, all initialized models improve the result obtained by the BLA, and with the proposed initialization scheme it is possible to obtain (after optimization) NLSS models characterized by low RMSE and low complexity in terms of number of parameters. As for the computational aspects, in this example (run using MATLAB, on a PC with a i7-2600 3.40 GHz processor) the nonlinear initialization takes approximately 8 s, to be compared with a processing time of 40 min needed for the optimization routine. Therefore, the initialization step, even when repeated several times, can be considered inexpensive.

Note that similar conclusions can be drawn also when testing the proposed approach on examples where stronger nonlinearities (such as the absolute value) are present in the system. In those cases, the only difficulty seems to be the fact that it is not always possible to perform model selection already at initialization, and one then needs to optimize several nonlinear models before selecting the optimal one.

D. Comparison With Linear Initialization

To demonstrate the advantage of using the proposed algorithm to initialize NLSS models, two situations are compared:

1) Linear/random initialization: the MLP parameters associated with the neuron amplitudes are set equal to 0, while all other MLP parameters (associated with the neuron positions) are randomly generated;
2) Linear/MLP initialization: the proposed initialization is performed (with $\lambda = 0.1$) to determine the MLP parameters associated with the neuron positions, but the MLP parameters associated with the neuron amplitudes are set equal to 0.

For both initializations, 50 different MLP configurations are considered, with $n_f = n_g = 3$.

All models resulting from these two initializations are linear models (i.e., they have the same performance as the BLA), but in one case (linear/MLP) one exploits the power of the proposed algorithm to fix the positions of the neurons. The mean value and the standard deviation (in V) of the RMSE results (on the validation set) obtained after optimizing all models are listed below.

1) Linear/random: mean 0.0443, std 0.0008;
2) Linear/MLP: mean 0.0068, std 0.0054.

This shows that if no information is given about the positions of the MLP neurons, the optimization routine gets always stuck in a bad local minimum. Only when the proposed algorithm is exploited to determine (at least) the neurons positions, it is possible to converge to lower RMSE values.

A final remark on the initialization: the algorithm discussed in this paper seems to be powerful (in comparison with a linear initialization) in those cases in which the basis functions describing the nonlinearities are characterized by parameters expressing the "position," i.e., like in the MLP example. For "position-free" models, e.g., polynomial models, a linear initialization might be effective enough for the optimization to converge to a good local minimum.

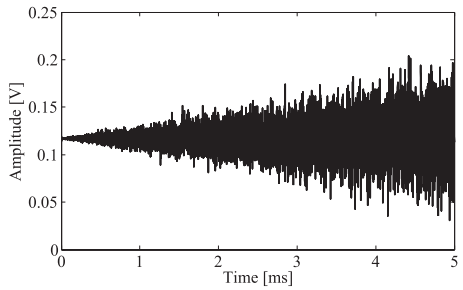


Fig. 4. Crystal detector example. Averaged input signal (for the estimation dataset).

V. Identification of a Crystal Detector

In this section, the task is to model the input/output relationship of a Agilent-HP420C crystal detector, a device that is often used in microwave applications to measure the envelope of a signal [26].

A. Description of the Data

The excitation signal consists of five measured periods of a Gaussian noise sequence (each period is made of 50 000 samples, at a sampling frequency of 10 MHz) characterized by a slowly increasing amplitude. A first dataset was generated by considering a bandwidth of 800 kHz and is used to estimate the model, while a second dataset, used for validation purposes, was generated with a similar excitation signal with bandwidth equal to 400 kHz. The five different periods of the excitation signal are used in the first step of the initialization when estimating the BLA to reduce noise and transient effects, while later on, for the estimation of the nonlinear model, the averaged input signal (over five periods) is considered.

The input signal (averaged over five periods) used for estimation is shown in Fig. 4.

The noise on the input and output measurements is estimated to have a standard deviation of 0.23 mV, to be compared with an RMS value of 15.2 mV for the input, and 13.6 mV for the output (after removing the DC offset).

B. Obtained Results

For the initialization of the NLSS model; first of all, the BLA is estimated nonparametrically to capture the dynamics of the system [6], then a first-order parametric linear model is obtained and transformed into state-space form. The BLA results in an RMSE equal to 1.5 and 1.1 mV on the estimation and validation set, respectively.

The nonlinear state x_{LS} is then approximated solving the problem in (8). Here the optimal value for the tradeoff parameter λ is chosen to be 0.01 (corresponding to $E_y/(E_y+E_x)$ less than 1%), although also in this example, the choice of λ does not seem to have a significant impact on the final result.

To estimate the nonlinear functions f and g, the model structure in Fig. 1 is considered, with the difference that now only one state is present, since a first-order model is sufficient to describe the linear dynamics. One-hidden-layer MLPs with $\tanh(\cdot)$ as activation function are used to estimate the nonlinear terms, and the choice of two neurons for f_{NL} and one neuron for g_{NL} results in an initialized nonlinear model that significantly improves the performance of the BLA (RMSE equal to 0.65 mV on the validation data).

Starting from the initial estimates, all model parameters are then optimized using a Levenberg–Marquardt algorithm. This finally results in a fitted nonlinear model characterized by an RMSE equal to 0.27 mV on the validation set.

All obtained results in terms of RMSE on estimation and validation data are summarized in Table III.

TABLE III
Crystal Detector Example. RMSE Values (in mV) Obtained by the Considered Models

Model	Estimation RMSE	Validation RMSE
Best Linear Approximation	1.50	1.10
Initialized nonlinear model	0.72	0.65
Optimized nonlinear model	0.26	0.27

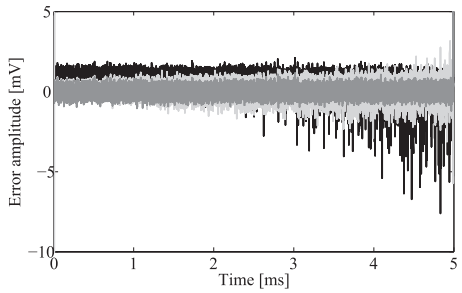


Fig. 5. Crystal detector example. Error (in mV) on the validation set for the BLA (black), the initialized nonlinear model (light gray) and the fitted nonlinear model after optimization (dark gray).

Fig. 5 shows the error obtained on the validation set by the different models. The optimized NLSS model can capture well the nonlinear behavior of the device, as indicated by the fact that the large spikes that characterize the linear model error are not present anymore in the nonlinear model error.

The results obtained by applying the proposed initialization scheme to the identification of the crystal detector can be considered very good. When compared with the error of the BLA, the initialized NLSS model gives an RMSE value on the validation data which is almost 50% lower. Moreover, starting from this initial parameter estimate, the performance of the nonlinear model is further improved in the optimization step, and the final nonlinear model error is very close to the noise level (0.23 mV).

In the final part of this section, the obtained results are compared with the performance of other nonlinear identification methods that have been applied on the same problem.

C. Comparison With Other Methods

A physical description of the crystal detector is developed in [27], where the internal structure is modeled as a nonlinear feedback, with a Wiener–Hammerstein branch in the feedback loop, for which the NL is a ninth-order polynomial. This model (characterized by 14 parameters) results in an RMSE value of 0.30 mV. Therefore, the NLSS model obtained with the proposed algorithm performs better, and its complexity is only slightly higher (17 parameters).

The PNLSS approach [9] has also been applied on the same problem, resulting in an RMSE of 0.26 mV (slightly

better than the proposed algorithm), at the cost of a higher complexity (53 parameters).

Finally, a different method based on the identification of a nonlinear LFR block structure gives an RMSE of 0.286 mV (with 11 parameters) [28].

It can be concluded that the proposed NLSS identification algorithm yields very good results in terms of both RMSE and complexity.

VI. Conclusion

In this paper, a novel initialization scheme for the identification of NLSS models has been presented. The approach was successfully applied on two measurement examples. The proposed initialization procedure has several advantages, as 1) it is a general scheme that can be used in combination with different choices of nonlinear model structures; 2) the separation between system dynamics and nonlinear terms makes it possible to identify them independently; 3) many nonlinear model structures can be tested rapidly on the obtained regression problem; and 4) two different fields—system identification and nonlinear regression/statistical learning—are brought together, combining the advantages of both.

A direction for future research could be to gradually impose more structure on the obtained NLSS models, e.g., by rotation of the state, to reveal the underlying structure of the considered system. Another possibility could be to start up the nonlinear initialization with a larger number of basis functions, and then select a smaller set during optimization; for instance, by adding a regularization term in the cost function.

References

[1] B. Ando, S. Baglio, V. Marletta, and A. R. Bulsara, "A nonlinear electric field sensor that exploits coupled oscillator dynamics: The charge collection mechanism," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 1326–1333, May 2013.

[2] J. Lota, M. Al-Janabi, and I. Kale, "Nonlinear model-based approach for accurate stability prediction of one-bit higher-order $\Delta-\Sigma$ modulators," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 4, pp. 686–692, Apr. 2013.

[3] M. Rawat and F. M. Ghannouchi, "Distributed spatiotemporal neural network for nonlinear dynamic transmitter modeling and adaptive digital predistortion," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 3, pp. 595–608, Mar. 2012.

[4] Y. Deng, Y. B. Shi, and W. Zhang, "An approach to locate parametric faults in nonlinear analog circuits," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 2, pp. 358–367, Feb. 2012.

[5] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.

[6] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, 2nd ed. New York, NY, USA: Wiley, 2012.

[7] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, *et al.*, "Nonlinear black-box modeling in system identification: A unified overview," *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.

[8] J. Sjöberg, "On estimation of nonlinear black-box models: How to obtain a good initialization," in *Proc. IEEE Workshop Neural Netw. Signal Process.*, Amelia Island, FL, USA, Sep. 1997, pp. 72–81.

[9] 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.

[10] A. Marconato, J. Sjöberg, J. A. K. Suykens, and J. Schoukens, "Separate initialization of dynamics and nonlinearities in nonlinear state-space models," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, Graz, Austria, May 2012, pp. 2104–2108.

[11] J. Codina, J. C. Aguado, and J. M. Fuertes, "Neural network identification of non linear systems using state space techniques," in *Proc. IMACS Int. Symp. Signal Process., Robot. Neural Netw.*, Lille, France, 1994, pp. 1–5.

[12] J. A. K. Suykens, B. De Moor, and J. Vandewalle, "Nonlinear system identification using neural state space models, applicable to robust control design," *Int. J. Control*, vol. 62, no. 1, pp. 129–152, 1995.

[13] J. A. K. Suykens, J. Vandewalle, and B. De Moor, *Artificial Neural Networks for Modeling and Control of Non-Linear Systems*. New York, NY, USA: Springer-Verlag, 1996.

[14] V. Verdult, J. A. K. Suykens, J. Boets, I. Goethals, and B. De Moor, "Least squares support vector machines for kernel CCA in nonlinear state-space identification," in *Proc. 16th Int. Symp. MTNS*, Leuven, Belgium, 2004.

[15] J. W. van Wingerden and M. Verhaegen, "Closed-loop subspace identification of Hammerstein-Wiener models," in *Proc. 48th IEEE CDC/CCC*, Shanghai, China, Dec. 2009, pp. 3637–3642.

[16] M. M. Tobenkin, I. R. Manchester, J. Wang, A. Megretski, and R. Tedrake, "Convex optimization in identification of stable non-linear state space models," in *Proc. 49th IEEE CDC*, Atlanta, GA, USA, Dec. 2010, pp. 7232–7237.

[17] Z. Ghahramani and S. T. Roweis, "Learning nonlinear dynamical systems using an EM algorithm," in *Advances in Neural Information Processing Systems*, vol. 11. Cambridge, MA, USA: MIT Press, 1999.

[18] A. A. Gorji and M. B. Menhaj, "Identification of nonlinear state space models using an MLP network trained by the EM algorithm," in *Proc. IJCNN*, Hong Kong, Jun. 2008, pp. 53–60.

[19] F. Lindsten and T. B. Schön, "Identification of mixed linear/nonlinear state-space models," in *Proc. 49th IEEE CDC*, Atlanta, GA, USA, Dec. 2010, pp. 6377–6382.

[20] T. B. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, 2011.

[21] B. Ninness, A. Wills, and T. B. Schön, "Estimation of general nonlinear state-space systems," in *Proc. 49th IEEE CDC*, Atlanta, GA, USA, Dec. 2010, pp. 6371–6376.

[22] R. Turner, M. P. Deisenroth, and C. E. Rasmussen, "State-space inference and learning with Gaussian processes," in *Proc. 13th Int. Conf. Artif. Intell. Stat.*, Sardinia, Italy, 2010, pp. 868–875.

[23] M. Enqvist, "Linear models of nonlinear systems," Ph.D. dissertation, Dept. Electr. Eng., Linköping Univ., Linköping, Sweden, 2005.

[24] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.

[25] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer-Verlag, 2009.

[26] L. Gommé, J. Schoukens, Y. Rolain, and W. Van Moer, "Validation of a crystal detector model for the calibration of the large signal network analyzer," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, Warsaw, Poland, May 2007, pp. 1–6.

[27] J. Schoukens, L. Gommé, W. Van Moer, and Y. Rolain, "Identification of a block-structured nonlinear feedback system, applied to a microwave crystal detector," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1734–1740, Aug. 2008.

[28] A. Van Mulders, L. Vanbeylen, and J. Schoukens, "Identification of a block-structured model with localised nonlinearity," in *Proc. 16th IFAC Symp. Syst. Identificat.*, Brussels, Belgium, 2012, pp. 1125–1130.



Anna Marconato (M'07) received the B.Sc. degree in mathematics and the M.Sc. degree in telecommunication engineering from the University of Trento, Trento, Italy, in 2002 and 2005, respectively, and the joint Ph.D. degree from the University of Trento and the Vrije Universiteit Brussel, Brussel, Belgium, in 2009.

She has been a Post-Doctoral Researcher with the Electrical Engineering Department, Vrije Universiteit Brussel, since 2009. Her current research interests include nonlinear system identification and machine learning.



Jonas Sjöberg received the master's degree in applied physics from Uppsala University, Uppsala, Sweden, in 1989 and the Ph.D. degree in engineering from Linköping University, Linköping, Sweden, in 1995.

He was an Assistant Professor with the Chalmers University of Technology, Göteborg, Sweden, after a Post-Doctoral Researcher with ETH Zurich, Zurich, Switzerland, and after research visits with TU Wien, Vienna, Austria, and Technion, Haifa, Israel, he became a Professor of mechatronics with the Chalmers University of Technology in 2001. From 2005 to 2010, he was the Programme Director for the five-year engineering programme in automation and mechatronics. His interest focus on model-based methods, simulations, system identification, and optimization for design and product development of mechatronic systems. His current research interests include mechatronics and mechatronic related fields, such as signal processing, and control.

Dr. Sjöberg received the Volvo Cars Technology Award in 2011.



Johan A. K. Suykens (SM'04) was born in Willebroek, Belgium, in 1966.

He was a Post-Doctoral Researcher with Fund for Scientific Research FWO Flanders and is currently a Professor with KU Leuven, Leuven, Belgium. He is the author of the books *Artificial Neural Networks for Modeling and Control of Nonlinear Systems* (Kluwer Academic Publishers) and *Least Squares Support Vector Machines* (World Scientific), co-author of the book *Cellular Neural Networks, Multiscroll Chaos and Synchronization* (World Scientific), and editor of the books *Nonlinear Modeling: Advanced Black-Box Techniques* (Kluwer Academic Publishers), and *Advances in Learning Theory: Methods, Models, and Applications* (IOS Press). In 1998, he organized an International Workshop on Nonlinear Modeling with Time-Series Prediction Competition.

Dr. Suykens has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1997 to 1999 and from 2004 to 2007 and the IEEE TRANSACTIONS ON NEURAL NETWORKS from 1998 to 2009. He received the IEEE Signal Processing Society 1999 Best Paper (Senior) Award and several Best Paper Awards at international conferences. He is a recipient of the International Neural Networks Society INNS Young Investigator Award in 2000 for significant contributions in the field of neural networks. He has served as the Director and Organizer of the NATO Advanced Study Institute on Learning Theory and Practice in Leuven in 2002, as a Program Co-Chair for the International Joint Conference on Neural Networks in 2004, and the International Symposium on Nonlinear Theory and its Applications in 2005, as an Organizer of the International Symposium on Synchronization in Complex Networks in 2007, a Co-Organizer of the NIPS 2010 workshop on Tensors, Kernels, and Machine Learning, and Chair of ROKS in 2013. He has been awarded an ERC Advanced Grant 2011.



Johan Schoukens (F'97) received the master's degree in electrical engineering and the Ph.D. degree in engineering from Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1985, and the Doctor Honoris Causa degree from the Budapest University of Technology and Economics, Budapest, Hungary, in 2011.

He was a Researcher of the Belgian National Fund for Scientific Research with the Electrical Engineering Department, VUB, from 1981 to 2000, where he is currently a Full-Time Professor of electrical engineering. His current research interests include system identification, signal processing, and measurement techniques.

Dr. Schoukens was a recipient of the 2002 Andrew R. Chi Best Paper Award from the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, the 2002 Society Distinguished Service Award from the IEEE Instrumentation and Measurement Society, and the 2007 Belgian Francqui Chair at the Université Libre de Bruxelles, Belgium. Since 2010, he has been a member of the Royal Flemish Academy of Belgium for Sciences and the Arts.