Automatica 46 (2010) 647–656

Contents lists available at ScienceDirect

Automatica

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

Identification of nonlinear systems using Polynomial Nonlinear State Space models☆

Johan Paduart [a], Lieve Lauwers [a,*], Jan Swevers [b], Kris Smolders [b], Johan Schoukens [a], Rik Pintelon [a]

[a] Vrije Universiteit Brussel, dep. ELEC, Pleinlaan 2, 1050 Brussels, Belgium

[b] Katholieke Universiteit Leuven, dep. PMA, Celestijnenlaan 300 B, 3001 Heverlee, Belgium

ARTICLE INFO

Article history:
Received 4 August 2008
Received in revised form
30 April 2009
Accepted 4 December 2009
Available online 26 February 2010

Keywords:
System identification
Nonlinear systems
Multivariable systems
Best linear approximation

ABSTRACT

In this paper, we propose a method to model nonlinear systems using polynomial nonlinear state space equations. Obtaining good initial estimates is a major problem in nonlinear modelling. It is solved here by identifying first the best linear approximation of the system under test. The proposed identification procedure is successfully applied to measurements of two physical systems.

1. Introduction

Many real-life systems and phenomena are nonlinear. Their behaviour can often be approximated by linear models which are easy to understand and to interpret. Unfortunately, linear approximations are only valid for a given input range. Hence, during the last decades there has been a tendency towards nonlinear modelling in various application fields. In order to build models for nonlinear devices, system identification methods are employed (Ljung, 1999; Pintelon & Schoukens, 2001; Soderstrom & Stoica, 1989). An excellent starting point for nonlinear modelling is Sjoberg et al. (1995). Other reference works on nonlinear systems and nonlinear modelling include Schetzen (1980), Rugh (1981), Chen and Billings (1989), Boyd and Chua (1985), Khalil (1996), Suykens, De Moor, and Vandewalle (1995), Verdult (2002) and Schrempf (2004). A major drawback is the lack of a general nonlinear framework. However, there exists a class of nonlinear systems which has intensively been studied in the past, and which covers a broad spectrum of 'nice' nonlinear behaviour, namely the class of Wiener systems. This class of systems stems from the Volterra–Wiener theory (Rugh, 1981; Schetzen, 1980) and will be employed here as a framework to develop the initialization procedure of the Polynomial NonLinear State Space (PNLSS) model.

The goal of this paper is to model multivariable nonlinear systems. To achieve this, the system is considered as a black box. The only available information about the system is given by its measured inputs and outputs. This approach usually means that no physical parameters or quantities are estimated. Hence, no physical interpretation whatsoever can be given to the model or its parameters. Furthermore, black box modelling implies the application of a model structure that is as flexible as possible, since no information about the device's internal structure is utilized. Often, this flexibility results in a high number of parameters. In order to cope with Multiple Input, Multiple Output (MIMO) systems, a suitable model structure needs to be chosen. For such systems, it is important that the common dynamics, present in the different outputs of the Device Under Test (DUT), are exploited in such a way that the number of model parameters is small. A model structure that fulfils this requirement in a natural way is the state space representation.

Only discrete-time models will be considered here. A motivation for this choice is that, when looking to control applications, discrete-time descriptions are more suitable since control actions are usually taken at discrete time instances. Furthermore, the estimation of nonlinear continuous-time models is not a trivial task,

☆ This work was funded by the Methusalem grant of the Flemish Government (METH-1), and the Belgian program on interuniversity poles of attraction initiated by the Belgian State, Prime Minister's Office, Science Policy programming (IUAP-VI/4). The support of the following projects is also gratefully acknowledged: FWO projects G.0528.04 and G.0422.08. The material in this paper was partially presented at SYSID conference 2006, Australia. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +32 26292953; fax: +32 26292850.
E-mail addresses: jpaduart@gmail.com (J. Paduart), lieve.lauwers@vub.ac.be (L. Lauwers).

0005-1098/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2010.01.001

and can be computationally involved, because it may imply the calculation of time-derivatives or integrals of sophisticated nonlinear functions of the measured signals (Unbehauen & Rao, 1998). Finally, it should be noted that a continuous-time approach is not strictly necessary, since we are not interested in the estimation of physical system parameters. In this paper, an identification procedure is presented for nonlinear systems using a nonlinear state space approach. First, a linear state space model is fit on the measured data, starting from the Best Linear Approximation of the DUT. Next, this linear model is extended to a Polynomial Nonlinear State Space model which is optimized to capture also the system's nonlinear behaviour.

The structure of the paper is the following: First, the Best Linear Approximation of a nonlinear system is defined. Then, the state space framework and the Polynomial Nonlinear State Space model is presented. Next, the identification procedure is explained in detail. Finally, the proposed identification method is applied to measurement data from two physical systems: a Single Input, Single Output (SISO) system and a Multiple Input, Single Output (MISO) system.

2. Best linear approximation

2.1. Definition

Consider a discrete-time SISO nonlinear system S with input $u(t)$ and output $y(t)$.

Definition 1. The Best Linear Approximation (BLA) of a nonlinear system S is defined as the model G belonging to the set of linear models $\mathcal{G}$, such that

$$G_{BLA} = \underset{G\in\mathcal{G}}{\operatorname{argmin}} E\left\{|y(t) - G(u(t))|^2\right\}. \tag{1}$$

In general, the Best Linear Approximation G_{BLA} of a nonlinear system depends on the amplitude distribution, the power spectrum, and the higher order moments of the stochastic input $u(t)$ (Enqvist, 2005; Enqvist & Ljung, 2005; Pintelon & Schoukens, 2001). It corresponds to the classic idea of linearizing a nonlinear system around its operating point. Notice that G_{BLA} can be unstable or noncausal.

2.2. Class of excitations

Since G_{BLA} depends on the properties of the input signal, it is important to define the class of excitation signals that are considered in this paper.

Assumption 1 (*Class of Excitations* S_E). The input signal $u(t)$ is stationary and belongs to the extended class of Gaussian excitation signals with a user-defined power spectrum.

The extended class of Gaussian(-like) signals includes Gaussian noise and random phase multisines, defined below.

Definition 2. A Gaussian noise signal is a random sequence drawn from a Gaussian distribution with a user-defined power spectral density.

Definition 3. A random phase multisine is a periodic signal, defined as a sum of harmonically related sine waves:

$$u(t) = \frac{1}{\sqrt{N}} \sum_{k=-N}^{N} U_k e^{j(2\pi f_{max}\frac{k}{N}t+\phi_k)} \tag{2}$$

with $\phi_{-k} = -\phi_k$, $U_k = U_{-k} = U(kf_{max}/N)$, and f_{max} the maximum frequency of the excitation signal. The amplitudes $U(f)$ are chosen in a custom way, according to the user-defined power spectrum that should be realized. The phases ϕ_k are the realizations of an independent (over k) distributed random process such that $E\left\{e^{j\phi_k}\right\} = 0$. A typical choice is to take ϕ_k uniformly distributed over $[0, 2\pi)$.

The factor $1/\sqrt{N}$ serves as normalization such that, asymptotically ($N \to \infty$), the power of the multisine remains finite, and its Root Mean Square (RMS) value stays constant as N increases. Note that the sum over k goes from $-N$ to N in order to obtain a real signal.

The random phase multisine is asymptotically normally distributed, i.e., for the number of excited frequency components going to infinity ($N \to \infty$), but in practice 20 excited lines works already very well for smoothly varying amplitude distributions U_k (Pintelon & Schoukens, 2001; Schoukens, Dobrowiecki, & Pintelon, 1998).

2.3. Properties

The nonlinear system S needs to fulfil some conditions in order to define its Best Linear Approximation.

Assumption 2 (*Class of Systems*). For the system S and for $u(t) \in S_E$, there exists a uniformly bounded Volterra series of which the output converges in mean square sense to the output $y(t)$ of S.

The class of considered systems is also called Wiener or PISPOT (Periodic Input, Same Period OutpuT) systems. It includes discontinuities like quantizers or relays, and excludes chaotic behaviour or systems with bifurcations.

Theorem 1. *If the system S satisfies Assumption 2, it can be modelled as the sum of a linear system $G_{BLA}(j\omega)$ and a noise source y_S. This noise source represents that part of the output y that cannot be captured by the linear model $G_{BLA}(j\omega)$. Hence, for frequency ω_k the output is written as*

$$Y(j\omega_k) = G_{BLA}(j\omega_k)U(j\omega_k) + Y_S(j\omega_k). \tag{3}$$

$Y_S(j\omega_k)$ depends on the particular input realization and exhibits a stochastic behaviour from realization to realization, with $E\{Y_S(j\omega_k)\} = 0$. The Best Linear Approximation is calculated as the solution of (2),

$$G_{BLA}(j\omega) = \frac{S_{yu}(j\omega)}{S_{uu}(j\omega)} \tag{4}$$

where $S_{uu}(j\omega)$ is the auto-power spectrum of the input, and $S_{yu}(j\omega)$ the cross-power spectrum between the output and the input.

Relation (4) is obtained by calculating the Fourier transform of the Wiener–Hopf equation, which is the solution of Eq. (1) (see for instance Eykhoff (1974) for this classic result).

For random phase multisines (see Definition 3), Eq. (4) reduces to the average of the measured Frequency Response Functions (FRFs) over several realizations of the input signal (see Section 5.1).

In Dobrowiecki and Schoukens (2007), the Best Linear Approximation framework was also developed for MIMO systems. For the sake of brevity, the discussion is restricted here to the SISO BLA.

3. State space framework

The most natural way to represent systems with multiple inputs and outputs is to use the state space framework.

Definition 4. An n_a-th order discrete-time state space model is generally expressed as

$$\begin{cases} x(t+1) = f(x(t), u(t)) \\ y(t) = g(x(t), u(t)) \end{cases} \tag{5}$$

with $u(t) \in \mathbb{R}^{n_u}$ the vector containing the n_u input values at time instance t, and $y(t) \in \mathbb{R}^{n_y}$ the vector of the n_y outputs. The state vector $x(t) \in \mathbb{R}^{n_a}$ represents the memory of the system, and includes the common dynamics present in the different outputs.

The first equation of (5) is referred to as the state equation. It describes the evolution of the state as a function of the input and the previous state. The second equation of (5) is called the output equation. It relates the system output with the state and the input.

The state space representation is not unique. By means of a similarity transform, the model equations in (5) can be converted into a new model that exhibits exactly the same input/output behaviour. The similarity transform $x_T(t) = T^{-1}x(t)$ with an arbitrary nonsingular square matrix T yields

$$\begin{cases} x_T(t+1) = T^{-1}f(Tx_T(t), u(t)) = f_T(x_T(t), u(t)) \\ y(t) = g(Tx_T(t), u(t)) = g_T(x_T(t), u(t)). \end{cases} \tag{6}$$

4. Polynomial Nonlinear State Space model

4.1. Definition

Consider the general state space model in (5) and apply a functional expansion of the functions f and g. In principle, various kinds of basis functions can be used for this purpose. In this paper, a set of polynomial basis functions is chosen. The main advantage of polynomials is that they are straightforward to compute, and easy to apply in a multivariable framework.

Definition 5. The Polynomial NonLinear State Space (PNLSS) model is defined as

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) + E\zeta(t) \\ y(t) = Cx(t) + Du(t) + F\eta(t). \end{cases} \tag{7}$$

The coefficients of the linear terms in $x(t)$ and $u(t)$ are given by the matrices $A \in \mathbb{R}^{n_a \times n_a}$ and $B \in \mathbb{R}^{n_a \times n_u}$ in the state equation, and $C \in \mathbb{R}^{n_y \times n_a}$ and $D \in \mathbb{R}^{n_y \times n_u}$ in the output equation. The vectors $\zeta(t) \in \mathbb{R}^{n_\zeta}$ and $\eta(t) \in \mathbb{R}^{n_\eta}$ contain nonlinear monomials in $x(t)$ and $u(t)$ of degree two up to a chosen degree p. The coefficients associated with these nonlinear terms are given by the matrices $E \in \mathbb{R}^{n_a \times n_\zeta}$ and $F \in \mathbb{R}^{n_y \times n_\eta}$. Note that the monomials of degree one are included in the linear part of the PNLSS model structure.

The separation in a linear and a nonlinear part is of no importance for the behaviour of the model. However, this distinction will turn out to be very practical, since the first step of the identification procedure consists in estimating a linear model.

When a full polynomial expansion of (5) is carried out, all monomials up to degree p must be taken into account. First, $\xi(t)$ is defined as the concatenation of the state vector and the input vector:

$$\xi(t) = [\ x_1(t) \quad \dots \quad x_{n_a}(t) \quad u_1(t) \quad \dots \quad u_{n_u}(t)\ ]^T. \tag{8}$$

As a consequence, the dimension of the vector $\xi(t)$ is given by $n = n_a + n_u$. Then, using the notation explained in Appendix A, $\zeta(t)$ and $\eta(t)$ in (7) are defined as

$$\zeta(t) = \eta(t) = \xi(t)_{\{p\}}. \tag{9}$$

This corresponds to considering all the distinct nonlinear combinations of degree p, which is the default choice for the PNLSS model structure. The total number of parameters required by the model in (7) is given by

$$\left[\binom{n_a + n_u + p}{p} - 1\right](n_a + n_y). \tag{10}$$

4.2. Approximation properties

It is known that bilinear state space models are universal approximators for continuous-time systems. Any continuous causal functional can be approximated arbitrarily well by a bilinear state space model within a bounded time interval (Fliess & Normand-Cyrot, 1982). Unfortunately, this universal approximation property does not hold in the discrete-time setting: in discrete-time, it is not possible to approximate any continuous functional arbitrarily well by a discrete-time bilinear state space model. This is due to the fact that the set of discrete-time bilinear systems is not closed with respect to the product operation. In other words, the product of two discrete-time bilinear state space systems is not necessarily bilinear again (Fliess & Normand-Cyrot, 1982). A more general class of state space models than the bilinear one is necessary in order to formulate a universal approximation property for discrete-time systems, namely state affine models. These models were introduced in Sontag (1979). A state affine model of degree N is defined as

$$\begin{cases} x(t+1) = \sum_{i=0}^{N-1} A_i u^i(t) x(t) + \sum_{i=1}^{N} B_i u^i(t) \\ y(t) = \sum_{i=0}^{N-1} C_i u^i(t) x(t) + \sum_{i=1}^{N} D_i u^i(t). \end{cases} \tag{11}$$

These models pop up in a natural way when describing sampled continuous-time bilinear state space systems (Rugh, 1981). On a finite time interval and with bounded inputs, they can arbitrarily well approximate any continuous, discrete-time system (Fliess & Normand-Cyrot, 1982). The advantage of this type of model is that the states $x(t)$ appear linearly in the state and output equations. As a consequence, subspace identification techniques can be used to estimate the model parameters. As pointed out in Schrempf (2004), the drawback of the state affine approach is that the required state dimension for a good approximation can become quite high, especially when the nonlinearities are concentrated in the states. Note that from model equations (7) and (11), it can be seen that state affine models form a subset of the PNLSS model class. Hence, the approximation capabilities proven for the state affine approach can be adopted, having the additional advantage that the PNLSS approach does not suffer from the high-dimensionality problem discussed in Schrempf (2004).

5. Identification procedure

The identification procedure for the PNLSS model in (7) consists of three major steps. First, G_{BLA} of the DUT is determined nonparametrically in mean square sense. Then, a parametric linear model is estimated from the Best Linear Approximation using frequency domain subspace identification methods (McKelvey, Akcay & Ljung, 1996; Pintelon, 2002). This is followed by a nonlinear optimization of the linear model. The last step consists in estimating the full nonlinear model by using again a nonlinear search routine.

5.1. Best linear approximation

From the measured input/output data, the system's G_{BLA} is extracted by performing classical FRF measurements. The Best Linear Approximation $\hat{G}_{BLA}$ and its sample covariance $\hat{C}_G$ can be determined in a straightforward way using periodic excitations (D'haene, Pintelon, Schoukens & Van Gheem, 2005). In that case, these quantities are calculated as follows:

$$\begin{aligned}\hat{G}_{BLA}(j\omega_k) &= \frac{1}{M}\sum_{m=1}^{M}\hat{G}^{[m]}(j\omega_k)\\ \hat{C}_G(j\omega_k) &= \frac{1}{M(M-1)}\sum_{m=1}^{M}|\hat{G}^{[m]}(j\omega_k)-\hat{G}_{BLA}(j\omega_k)|^2\end{aligned} \tag{12}$$

where, for every experiment m, the FRF estimate $\hat{G}^{[m]}(j\omega_k) \in \mathbb{C}$ is defined as

$$\hat{G}^{[m]}(j\omega_k) = \frac{\hat{Y}^{[m]}(k)}{\hat{U}^{[m]}(k)} \tag{13}$$

with

$$\begin{aligned}\hat{U}^{[m]}(k) &= \frac{1}{P}\sum_{p=1}^{P}U^{[m,p]}(k)\\ \hat{Y}^{[m]}(k) &= \frac{1}{P}\sum_{p=1}^{P}Y^{[m,p]}(k).\end{aligned} \tag{14}$$

$U^{[m,p]}(k)$ and $Y^{[m,p]}(k)$ are the DFT of the pth period of experiment m in a periodic excitation set-up. Note that Eq. (13) is equivalent to (4) for periodic excitations. In order to graphically represent the covariance $\hat{C}_G(j\omega_k)$ in the experimental section, only the square root of the diagonal elements of $\hat{C}_G(j\omega_k)$ will be plotted. This quantity will be referred to as the standard deviation. Explicit expressions to calculate the quantities $\hat{G}_{BLA}$ and $\hat{C}_G$ in the case of nonperiodic excitations can be found in Paduart (2008).

The reduction of the input/output data to a compact FRF and covariance form offers a number of advantages. First of all, the Signal-to-Noise Ratio (SNR) is enhanced. Secondly, it allows the user to select, in a straightforward way, a frequency band of interest. Finally, when periodic data are available, the measurement noise and the effect of the nonlinear behaviour can be separated (Schoukens et al., 1998).

5.2. Frequency domain subspace identification

Next, the nonparametric estimate $\hat{G}_{BLA}(j\omega_k)$ is transformed into a parametric model. The goal is to estimate a linear, discrete-time state space model, taking into account the sample covariance matrix $\hat{C}_G(j\omega_k)$. The state space matrices ($ABCD$) can be retrieved by using the frequency domain subspace identification algorithm described in McKelvey et al. (1996) and by relying on the results presented in Pintelon (2002). In the latter, the stochastic properties of this algorithm are analysed for the case in which the sample covariance matrix is employed instead of the true covariance matrix.

5.3. Nonlinear optimization of the linear model

The Weighted Least Squares (WLS) cost function V_{WLS} is defined as (McKelvey et al., 1996):

$$V_{WLS} = \sum_{k=1}^{F}\varepsilon^H(k)C_G^{-1}(k)\varepsilon(k) \tag{15}$$

where

$$\varepsilon(k,\theta) = vec(G_{SS}(A,B,C,D,j\omega_k) - \hat{G}_{BLA}(j\omega_k)) \tag{16}$$

$$G_{SS}(A,B,C,D,j\omega_k) = C(z_k I_{n_a} - A)^{-1}B + D \tag{17}$$

with $z_k = e^{j2\pi(k/N)}$, I_{n_a} the identity matrix of size n_a, and θ the vector containing all entries of A, B, C, and D. This cost function V_{WLS} is a measure of the fit between the model G_{SS} and the measured $\hat{G}_{BLA}$. According to this measure, the frequency domain subspace algorithm generates reasonable model estimates. However, in practical applications V_{WLS} strongly depends on the dimension parameter r chosen in the identification procedure. A first action that can be taken to improve the model estimates, is to apply the subspace algorithm for different values of r, for instance $r = n_a + 1, \ldots, 6n_a$, and to select the model that corresponds to the lowest V_{WLS}.

Secondly, V_{WLS} can be minimized with respect to all the linear parameters ($ABCD$). This nonlinear problem can be solved using the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). It requires the computation of the Jacobian of the model error $\varepsilon(k,\theta)$ with respect to the model parameters. From (16) and (17), the following expressions are calculated:

$$\begin{cases}\dfrac{\partial\varepsilon(k,\theta)}{\partial A_{ij}} = vec\left(C(z_k I_{n_a} - A)^{-1} I_{ij}^{n_a\times n_a}(z_k I_{n_a} - A)^{-1}B\right)\\ \dfrac{\partial\varepsilon(k,\theta)}{\partial B_{ij}} = vec\left(C(z_k I_{n_a} - A)^{-1} I_{ij}^{n_a\times n_u}\right)\\ \dfrac{\partial\varepsilon(k,\theta)}{\partial C_{ij}} = vec\left(I_{ij}^{n_y\times n_a}(z_k I_{n_a} - A)^{-1}B\right)\\ \dfrac{\partial\varepsilon(k,\theta)}{\partial D_{ij}} = vec\left(I_{ij}^{n_y\times n_u}\right)\end{cases} \tag{18}$$

where $I_{ij}^{m\times n}$ is defined in (33) of Appendix B. Note that special measures need to be taken during the optimization since the state space representation is overparametrized (see Section 5.4.4). The subspace method in McKelvey et al. (1996) is used to generate a number of different linear models which are then used as starting values for the nonlinear optimization procedure. In this way, there is a higher probability to end up in a global minimum of V_{WLS}, or at least in a good local minimum. Finally, the model that corresponds to the lowest cost function V_{WLS} is selected. The estimation procedure is carried out for different model orders n_a. Next, a model selection criterion such as AIC (Akaike, 1974) or MDL (Rissanen, 1978) can be used to choose the best linear model.

5.4. Estimation of the full nonlinear model

The last step in the identification process is to estimate the full nonlinear model

$$\begin{cases}x(t+1) = Ax(t) + Bu(t) + E\zeta(t)\\ y(t) = Cx(t) + Du(t) + F\eta(t) + e(t)\end{cases} \tag{19}$$

with the initial state given by $x(1) = x_0$, and where $e(t)$ is the output noise. In order to keep the parameter estimates unbiased, an additional assumption is required concerning the input signal:

Assumption 3 (*Noise Model*). The input $u(t)$ is assumed to be noiseless, i.e., it is observed without any errors and independent of the output noise.

Consider the weighted least squares cost function

$$V_{WLS}(\theta) = \sum_{k=1}^{F}\varepsilon^H(k,\theta)W(k)\varepsilon(k,\theta) \tag{20}$$

with $W(k) \in \mathbb{C}^{n_y \times n_y}$ a user-chosen, frequency domain weighting matrix. Typically, this matrix is chosen equal to the inverse covariance matrix of the output $\hat{C}_Y^{-1}(k)$ which can straightforwardly be obtained using periodic excitation signals. By choosing $W(k)$ properly, it is also possible to put more weight in a frequency band of interest. When no covariance information is available and no specific weighting is required by the user, a constant weighting is employed ($W(k) = 1$, for $k = 1, \ldots, F$). Furthermore, the model error $\varepsilon(k, \theta) \in \mathbb{C}^{n_y}$ is defined as

$$\varepsilon(k, \theta) = Y(k, \theta) - Y(k) \tag{21}$$

where $Y(k, \theta)$ and $Y(k)$ are the DFT of the modelled and the measured output, respectively.

5.4.1. Calculation of the Jacobian

$V_{WLS}(\theta)$ is minimized with respect to the model parameters $\theta = [vec(A); vec(B); vec(C); vec(D); vec(E); vec(F)]$ via the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). This requires the computation of the Jacobian $J(k, \theta)$ of the modelled output with respect to the model parameters:

$$J(k, \theta) = \frac{\partial \varepsilon(k, \theta)}{\partial \theta} = \frac{\partial Y(k, \theta)}{\partial \theta}. \tag{22}$$

Since it is impractical to calculate $Y(k, \theta)$ and, hence, $J(k, \theta)$ directly in the frequency domain, the calculations are performed in the time domain, followed by a DFT. It is known that the calculation of the Jacobian for a nonlinear state space model boils down to computing the output of another nonlinear model. The dynamics of this new nonlinear model are closely related to the dynamics of the original model (e.g., see Narendra & Parthasarathy, 1990; Suykens, Vandewalle, & De Moor, 1996). For the model equations in (19), explicit expressions for the Jacobian are derived in Appendix B.

5.4.2. Initial conditions

When computing the state sequence $x(t)$, the initial state x_0 of the model in (19) should be taken into account. For this, three possible approaches are distinguished. The simplest, but rather inefficient way, is to calculate the Jacobian for the full data set, and then to discard the first N_{trans} transient samples of both the Jacobian and the model error. However, in this way a part of the data is not used for the model estimation. The second method is only applicable when using periodic excitations. In this case, it suffices to calculate the Jacobian for several periods, and to select a period for which the transients become negligible. The last method, which is suitable for both periodic and nonperiodic excitations, is to estimate the initial conditions x_0 as if they were ordinary model parameters. Note that this corresponds to the estimation of an extra column in the state space matrix B and to an extra value 1 in the input vector.

5.4.3. Starting values

The last obstacle before starting the nonlinear optimization is to choose good starting values for θ. The estimates obtained from the parametric linear state space model are used as initial values for the A, B, C, and D matrices. The other state space matrices E and F are initially set to zero. The idea of using the parametric BLA as the initial nonlinear model offers two important advantages. First of all, it guarantees that the estimated nonlinear model performs at least as good as the best linear model. Secondly, for the model structure in (19), this principle results in a rough estimate of the model order n_a.

5.4.4. Similarity transform

Another issue that needs to be addressed is the rank deficiency of the Jacobian, which is present due to the non-uniqueness of the state space representation. As mentioned before, the similarity transform $x_T(t) = T^{-1}x(t)$ leaves the input/output behaviour unaffected. The n_a^2 elements of the transformation matrix T can be chosen freely, under the condition that T is nonsingular. Consequently, the parameter space has at least n_a^2 unnecessary dimensions. This poses a problem for the gradient-based identification of the model parameters $\theta \in \mathbb{R}^{n_\theta}$: the Jacobian will not be of full rank and, hence, an infinite number of equivalent solutions will exist. To deal with the rank deficiency of the Jacobian when calculating the parameter update, a pseudo-inverse is used. This is achieved by a truncated Singular Value Decomposition (SVD) (Golub & Van Loan, 1996; Pintelon, Schoukens, Vandersteen, & Rolain, 1999; Wills & Ninness, 2008); it allows a full parametrization of the model.

5.4.5. Overfitting and validation

The nonlinear search should be pursued until the cost function in (20) stops decreasing. However, as it is often the case for model structures with many parameters, overfitting can occur during the nonlinear optimization. In order to avoid this effect, the so-called stopped search (Sjoberg et al., 1995) is used. The model quality of every estimated model is evaluated on a test set, and then the model that achieves the best result is selected. This method is a form of implicit regularization, because unnecessary parameters are not explicitly removed.

5.5. Summary

An outline of the identification procedure is given here:

1. From the measured input/output data, compute the Best Linear Approximation $\hat{G}_{BLA}$ and its covariance $\hat{C}_G$.
2. Identify a linear state space model:
   (a) Estimate initial values using the frequency domain subspace method (McKelvey et al., 1996) by scanning over the dimension parameter r, and different model orders n_a.
   (b) Optimize all obtained linear models with a nonlinear search method, and then select the best model according to a model selection criterion.
3. Identify the PNLSS model:
   (a) Choose a degree for the polynomial equations. The default choice is a full expansion with degree $p = 3$ in both the state and the output equations (see also Section 6.3).
   (b) Use the best linear model from step (2), and use a nonlinear optimization to estimate the PNLSS model parameters.

6. Experimental results

The proposed identification procedure is now applied to measurements from two physical systems. The first system is an example of a SISO system; the second one of a MISO system. More applications of the PNLSS approach can be found in Paduart (2008).

6.1. Silverbox

6.1.1. Description

The Silverbox is an electronic circuit that emulates the behaviour of a mass-spring-damper system. The theoretical differential equation that describes this system is given by

$$m\ddot{y}(t) + d\dot{y}(t) + k_1 y(t) + k_3 y^3(t) = u(t). \tag{23}$$

In reality, additional higher order terms are present. The input u is the force applied to the mass m; the output y represents the mass

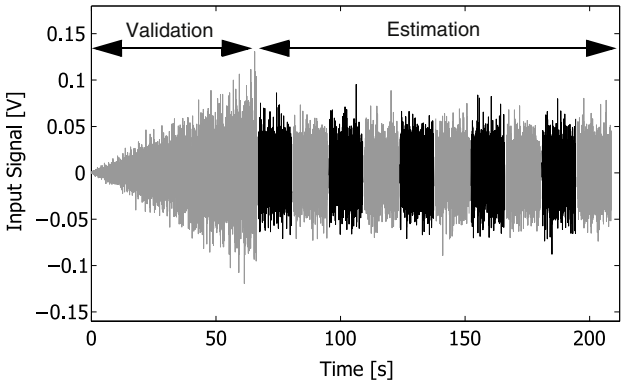

Fig. 1. Excitation signal containing an estimation and a validation set.

displacement. The spring acts nonlinearly and is characterized by the parameters k_1 and k_3. The parameter d determines the damping which is present in the system.

The applied excitation signal consists of two parts (see Fig. 1). The first part of the signal is filtered Gaussian white noise with a linearly increasing RMS value as a function of time. This sequence consists of 40 700 samples and has a bandwidth of 200 Hz. The average RMS value of the signal is 22.3 mV. This data set will be used to validate the models. The second part of the excitation signal contains 10 realizations of an odd random phase multisine (i.e., only the odd frequency lines are excited) with 8192 samples and 500 transient points per realization. The bandwidth of this excitation signal is also 200 Hz and its RMS value is 22.3 mV. This sequence will be used to estimate the models. In all experiments, the input and output signals are measured at a sampling frequency of 610.35 Hz.

6.1.2. Linear model

To obtain a nonparametric estimate of the BLA, the FRF is determined for every phase realization of the estimation data set. $\hat{G}_{BLA}(j\omega_k)$ is then calculated by averaging those FRFs. Next, a parametric second order linear state space model is estimated according to the method described in Section 5.2. The amplitude and the phase of the BLA are plotted in Fig. 2 (solid black line). The solid grey line represents the linear model. The standard deviation of the BLA is also given (black dashed line), together with the model error (dashed grey line), i.e., the difference between the measured BLA and the linear model.

From Fig. 2, it is clear that the linear model is of good quality: the model error coincides with the standard deviation up to a frequency of 120 Hz. A statistically significant, but small model error is present in the frequency band between 120 Hz and 200 Hz. This error is due to the fact that a band-limited set-up was employed during the measurements (Pintelon & Schoukens, 2001). Hence, a discrete-time, second order model does not suffice to perfectly model the continuous-time system in (23). However, in Section 6.1.3, it will be shown that the second order nonlinear model achieves satisfying results. As such, it is not necessary to use a higher order model. The second data set is now used to validate the linear model, using a free-run simulation. The RMS value of the simulation error (RMSE) is 13.7 mV. This number should be compared to the RMS output level that measures 53.4 mV.

6.1.3. Nonlinear model

From the second order linear model, initial values are extracted in order to estimate a number of nonlinear models. First, a second order PNLSS model is estimated with the following settings:

$$\xi(t) = \begin{bmatrix} x_1(t) & x_2(t) & u(t) \end{bmatrix}^T \tag{24}$$

and

$$\zeta(t) = \xi(t)_{\{3\}} \qquad \eta(t) = 0. \tag{25}$$

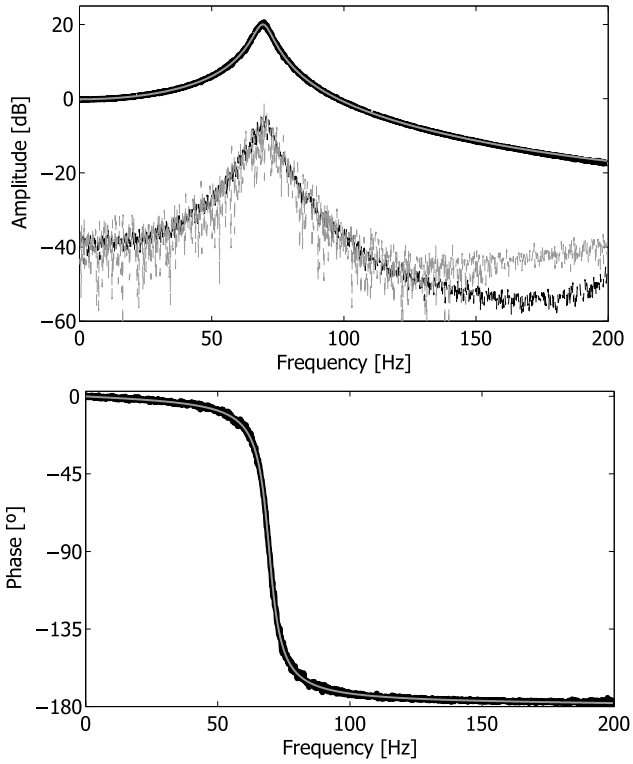

Fig. 2. BLA of the Silverbox (solid black line); standard deviation of the BLA (black dashed line); second order linear model (solid grey line); model error (dashed grey line).

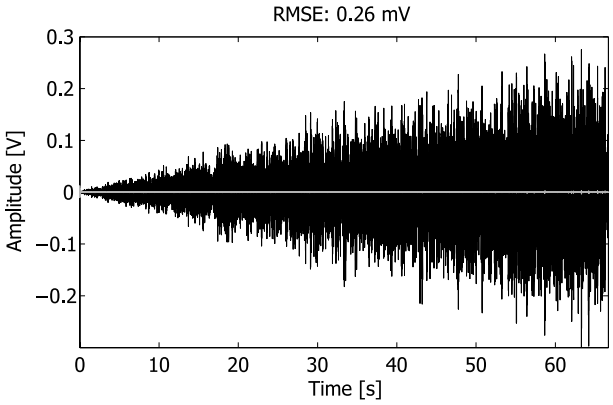

Fig. 3. Validation result for the identified nonlinear model: measured output (black) and model free-run simulation error (grey).

Hence, the nonlinear degree in the state equation is 3, and all cross products of the states and the input are included. In the output equation, only the linear terms are present. This results in a nonlinear model that contains 37 parameters. Via trial and error, it turned out that the settings in (24) and (25) gave the best outcome. The validation results for this nonlinear model are shown in Fig. 3. The measured output signal is denoted by the black line; the simulation error of the nonlinear model is plotted in grey. The RMS value of the model error has dropped significantly from 13.7 mV for the linear model to 0.26 mV for the nonlinear model. Hence, the second order PNLSS model performs more than a factor 50 better than the linear one.

The spectra of the measured validation output signal (black), the linear simulation error (light grey), and the nonlinear simulation error (dark grey) are shown in Fig. 4. The errors of the linear model are particularly present around the resonance frequency (approximately 60 Hz) for large signal amplitudes. The errors of the

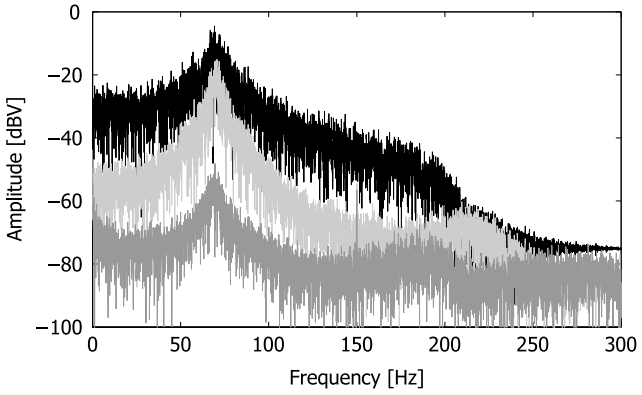

Fig. 4. DFT spectra of the measured validation output signal (black), linear simulation error (light grey), and nonlinear simulation error (dark grey).

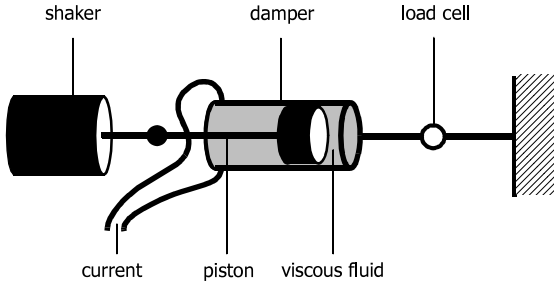

Fig. 5. Measurement set-up of the magneto-rheological damper.

nonlinear model are concentrated around the resonance frequency and close to DC. Higher model orders and degrees of nonlinearity were also tried out, but none of them gave better results than this second order nonlinear model.

6.2. Magneto-rheological damper

6.2.1. Description

This application concerns the modelling of a magneto-rheological (MR) damper which is illustrated in Fig. 5. The damper is a LORD RD-1097-01 friction damper described in Chrzan and Carlson (2001). Two quantities serve as input to this system: the reference signal applied to the PID controller to regulate the piston position via the shaker, and the current which determines the magnetic field over the viscous fluid. The system output is the force over the damper which is measured by a load cell. Three realizations of a full grid, random phase multisine were applied to the DUT. The first two realizations are used for the estimation of the models, and the third realization for the validation. The multisines were excited in a frequency band between 0.12 Hz and 10 Hz, and 6 periods per realization were measured with 65 536 samples per period. In all the measurements, a sampling frequency f_s of 2000 Hz was used. After removal of the DC levels, the signals applied to the first (piston reference) and second input (damper current) of the DUT have a RMS value of 39 mV and 194 mV, respectively.

6.2.2. Linear model

Unfortunately, only two multisine realizations are available for the model estimation. For a dual input system, this is sufficient to calculate the BLA, but not enough to determine an estimate of the covariance. To solve this problem, the data are treated as nonperiodic, and split into subrecords. First, the device's Best Linear Approximation is estimated via the auto- and cross-power spectra (see Eq. (4)), using a diff window (Schoukens, Rolain, & Pintelon, 2006). Then, a number of linear models with different

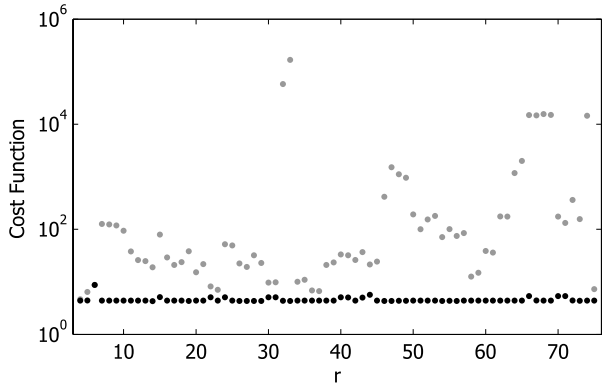

Fig. 6. Identification of a third order linear model for the MR Damper: subspace estimates (grey dots); estimates after applying a nonlinear optimization (black dots).

model orders (second to fifth order) are estimated from the BLA and its covariance matrix, using the subspace method followed by the nonlinear optimization. Fig. 6 shows the cost function for the third order estimation as a function of the subspace parameter r. The grey dots indicate the cost function for the subspace estimates for r ranging from 4 to 75, and exemplify the craggy behaviour of V_{WLS} as a function of r. The black dots denote the cost function (15) after optimizing these estimates using the nonlinear search routine. They show the importance of scanning different r values: for some choices of r, the optimization ends up in a local minimum (e.g. $r = 15, 22, 24, 30, 31, \ldots$). This demonstrates that it is useful to perform a sweep over r, since this diminishes the chance of ending up in some local minimum (e.g. by using a fixed choice of r).

Fig. 7 shows the MISO BLA (solid black line) and its standard deviation (dashed black line), together with the third order linear model (solid grey line) and the amplitude of the complex model error (dashed grey line). G_{11} is the transfer function from the piston reference (input 1) to the measured force (output 1); G_{12} is the transfer function from the damper current (input 2) to the measured force (output 1). G_{11} behaves like expected for a damper: ideally, the force over the damper should be proportional to the velocity of the piston, i.e., $j\omega$ times the displacement. This is, indeed, roughly what is observed for G_{11}. Furthermore, from the top plots it can be seen that the relative uncertainty on G_{12} is high compared with the one on G_{11}. Hence, the estimated linear model is mainly determined by G_{11}. Next, a validation test is carried out with the third order linear model. The RMS value of the model error (34 mV) is quite high compared with the RMS value of the measured output (71 mV). This error can be reduced using a nonlinear model.

6.2.3. Nonlinear model

As starting values for the nonlinear modelling, the second to fifth order linear models obtained in the previous step are used. For the PNLSS models, it is observed that the use of a nonlinear relation for both the state and the output equation always yields better modelling results. Hence, only nonlinear state and output equations are considered in what follows. Furthermore, the elements in the nonlinear vectors $\zeta(t)$ and $\eta(t)$ were restricted to the states only, i.e., without considering the inputs. For a nonlinear degree of 3, this corresponds to $\zeta(t) = \eta(t) = x(t)_{\{3\}}$ and results in a RMSE of 6.6 mV on the validation set. This is a reduction of the model error with a factor 5 compared with the linear model. This result should also be compared with the

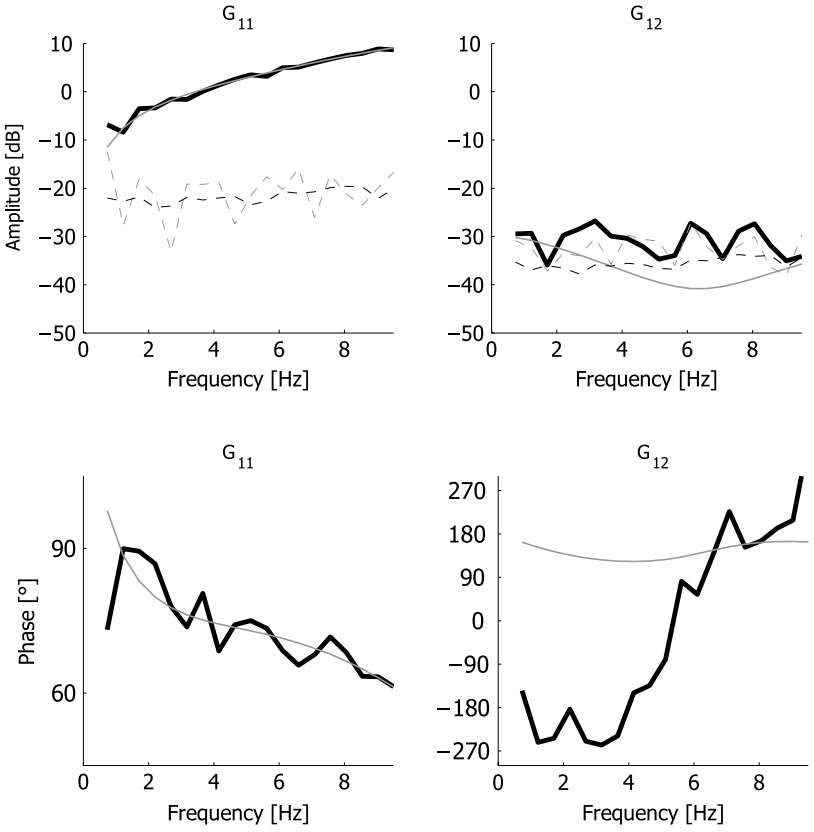

Fig. 7. The MISO BLA (G_{11} and G_{12}) of the MR damper (solid black line); third order linear model (solid grey line); standard deviation of the BLA (dashed black line); model error (dashed grey line).

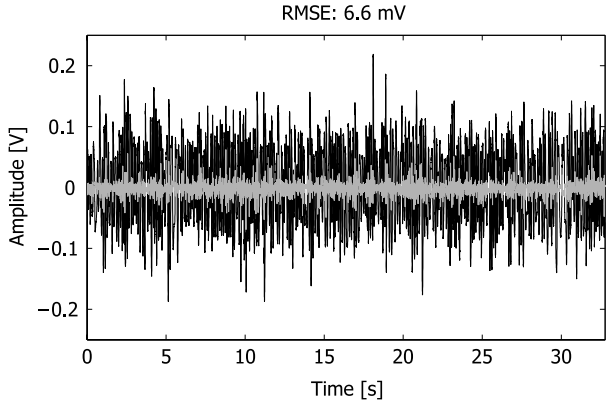

Fig. 8. Validation result for the best nonlinear model: measured output (black) and model free-run simulation error (grey).

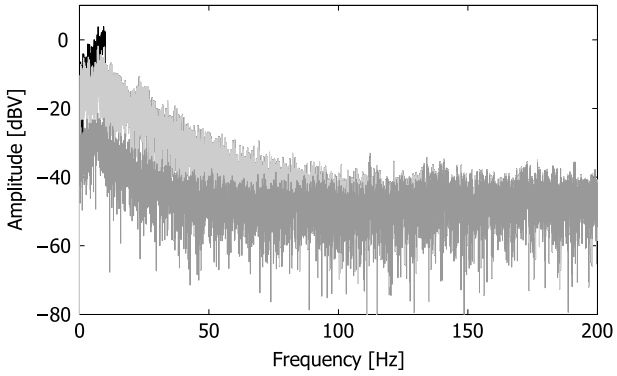

Fig. 9. DFT spectra of the measured validation output signal (black), linear simulation error (light grey), and nonlinear simulation error (dark grey).

noise level (1.8 mV), which can easily be determined since several periods of the measured data are available. In theory, it should be possible to reduce the model error with an additional factor 3 to 1.8 mV. However, it was not possible to achieve this with the PNLSS approach. Maybe the nonlinear optimization got stuck in a local minimum, or the model order/degree should be increased further in order to obtain better results. The validation error for the best nonlinear model is plotted in the time domain in Fig. 8 (grey), together with the measured output (black). Fig. 9 shows the spectra of the measured validation output signal (black), the linear simulation error (light grey), and the nonlinear simulation error (dark grey). This plot illustrates that the nonlinear model squeezes down the model error over a broad frequency range.

6.3. Discussion

One of the main advantages of the PNLSS approach is that no difficult identification settings have to be chosen by the user during the identification, such as the number of input and output time lags, the number of neurons,... An estimate of the model order is determined easily from the BLA when estimating a linear model. The standard PNLSS model usually delivers satisfying results with a moderate nonlinear degree (typically three). However, it is possible to further improve the model by tweaking the nonlinear state and output equations. Typically, a linear relation can be used in one of these equations. Furthermore, the monomials in the polynomial expansion can be restricted to some state or input variables. Naturally, there are limitations to the proposed approach. First of all, the PNLSS model is only suitable to handle low order systems (as a rule of thumb: $n_a < 10$). For higher order

systems, the combinatorial explosion of the number of parameters becomes too restrictive in order to get good modelling results. This problem can be overcome, for instance by tweaking the model equations as explained above.

The second disadvantage resides in the nonlinear search during the estimation of the model parameters: the risk of getting trapped in a local minimum is always imminent. However, it must be said that this weak spot is common to many identification methods, even in the case of linear modelling. Finally, it is difficult to guarantee the stability of the estimated models. But again, the risk of instability is inherent to any recursive model. Sometimes it is possible to add constraints in order to keep the model stable, but this may have a negative influence on the modelling performance. These three limitations constitute interesting topics for future research.

7. Conclusion

It was illustrated that when a general, black box model of a nonlinear device is required, the PNLSS model is a perfect tool to achieve this goal. It is a logical extension of the well known linear identification framework. The ability to cope with multivariable systems is included in a natural way. Furthermore, the PNLSS identification procedure is straightforward, and consists of only three simple steps: (1) compute the Best Linear Approximation, (2) estimate a linear model, and (3) solve a standard nonlinear optimization problem. The two identified DUTs in this paper demonstrate that the identification procedure works well in practice. In both test cases, a significant model error reduction was achieved compared with the linear models.

Appendix A. Notational issue

In order to denote monomials in an uncomplicated way, the n-dimensional multi-index α is defined which contains the powers of a multivariable monomial:

$$\alpha = \begin{bmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_n\end{bmatrix} \tag{26}$$

with $\alpha_i \in \mathbb{N}$. A monomial composed of the components from the vector $\xi \in \mathbb{R}^n$ is then simply written as

$$\xi^\alpha = \prod_{i=1}^{n} \xi_i^{\alpha_i} \tag{27}$$

where ξ_i is the ith component of ξ. The total degree of the monomial is given by

$$|\alpha| = \sum_{i=1}^{n} \alpha_i \tag{28}$$

and the factorial function of the multi-index α is defined as

$$\alpha! = \alpha_1!\alpha_2! \dots \alpha_n!. \tag{29}$$

Furthermore, $\xi_{(p)}$ is defined as the column vector of all the distinct monomials of degree p (i.e., with multi-index $|\alpha| = p$) composed from the elements of vector ξ. The number of elements in vector $\xi_{(p)}$ is given by the following binomial coefficient:

$$\binom{n+p-1}{p}. \tag{30}$$

Finally, the vector $\xi_{\{p\}}$ is defined as the column vector containing all the monomials of degree two up to degree p. The notations introduced above can now be used to express the multinomial expansion theorem (Weisstein, 2009).

Theorem 2. *The multinomial expansion theorem gives an expression for the power of a sum, as a function of the powers of the terms:*

$$\left(\sum_{i=1}^{n} \xi_i\right)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} \xi^\alpha. \tag{31}$$

Appendix B. Expressions for the PNLSS Jacobian

In this appendix, explicit expressions for the derivatives of the model (19) with respect to the parameters θ are computed. First, the matrices $\zeta'(t) \in \mathbb{R}^{n_\zeta \times n_a}$ and $\eta'(t) \in \mathbb{R}^{n_\eta \times n_a}$ are defined as

$$\begin{aligned}\zeta'(t) &= \frac{\partial \zeta(t)}{\partial x(t)} = \begin{bmatrix}\frac{\partial \zeta(t)}{\partial x_1(t)} & \dots & \frac{\partial \zeta(t)}{\partial x_{n_a}(t)}\end{bmatrix} \\ \eta'(t) &= \frac{\partial \eta(t)}{\partial x(t)} = \begin{bmatrix}\frac{\partial \eta(t)}{\partial x_1(t)} & \dots & \frac{\partial \eta(t)}{\partial x_{n_a}(t)}\end{bmatrix}\end{aligned} \tag{32}$$

$I_{ij}^{m\times n} \in \mathbb{R}^{m\times n}$ denotes a zero matrix with a single element equal to one at entry (i, j):

$$I_{ij}^{m\times n} = \begin{bmatrix}0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0\end{bmatrix} \begin{matrix}\\ \\ i \\ \\ \\ \end{matrix} \\ \qquad\qquad\quad j \tag{33}$$

Next, the Jacobian is computed with respect to the elements A_{ij} of the state space matrix A. The derivative of the output equation with respect to A_{ij} is given by

$$\begin{aligned}\frac{\partial y(t)}{\partial A_{ij}} &= \frac{\partial \left(Cx(t) + Du(t) + F\eta(t)\right)}{\partial A_{ij}} \\ &= C\frac{\partial x(t)}{\partial A_{ij}} + F\eta'(t)\frac{\partial x(t)}{\partial A_{ij}}.\end{aligned} \tag{34}$$

In order to determine the right hand side of (34), the derivatives of the state equation are needed. They are given by

$$\frac{\partial x(t+1)}{\partial A_{ij}} = \frac{\partial \left(Ax(t) + Bu(t) + E\zeta(t)\right)}{\partial A_{ij}}. \tag{35}$$

Furthermore, $x_{A_{ij}}(t) \in \mathbb{R}^{n_a}$ is defined as

$$x_{A_{ij}}(t) = \frac{\partial x(t)}{\partial A_{ij}}. \tag{36}$$

Then, Eq. (35) can be rewritten as

$$x_{A_{ij}}(t+1) = I_{ij}^{n_a\times n_a} x(t) + \left(A + E\zeta'(t)\right) x_{A_{ij}}(t). \tag{37}$$

Combining Eqs. (34) and (37) results in

$$\begin{cases} x_{A_{ij}}(t+1) = I_{ij}^{n_a\times n_a} x(t) + \left(A + E\zeta'(t)\right) x_{A_{ij}}(t) \\ J_{A_{ij}}(t) = \left(C + F\eta'(t)\right) x_{A_{ij}}(t). \end{cases} \tag{38}$$

where $J_{A_{ij}}(t) \in \mathbb{R}^{n_y}$ is defined as

$$J_{A_{ij}}(t) = \frac{\partial y(t)}{\partial A_{ij}}. \tag{39}$$

The Jacobian of the other model parameters is computed in a similar way and is summarized below.

$$\begin{cases} x_{B_{ij}}(t+1) = I_{ij}^{n_a\times n_u} u(t) + \left(A + E\zeta'(t)\right) x_{B_{ij}}(t) \\ J_{B_{ij}}(t) = \left(C + F\eta'(t)\right) x_{B_{ij}}(t) \\ x_{E_{ij}}(t+1) = I_{ij}^{n_a\times n_\zeta} \zeta(t) + \left(A + E\zeta'(t)\right) x_{E_{ij}}(t) \\ J_{E_{ij}}(t) = \left(C + F\eta'(t)\right) x_{E_{ij}}(t) \end{cases} \tag{40}$$

and

$$\begin{aligned} J_{C_{ij}}(t) &= I_{ij}^{n_y\times n_a} x(t) \\ J_{D_{ij}}(t) &= I_{ij}^{n_y\times n_u} u(t) \\ J_{F_{ij}}(t) &= I_{ij}^{n_y\times n_\eta} \eta(t). \end{aligned} \tag{41}$$

References

Akaike, H. (1974). A new look at statistical model identification. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, *19*(6), 716–723.

Boyd, S., & Chua, L. (1985). Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, *32*(11), 1150–1161.

Chen, S., & Billings, S. A. (1989). Representations of non-linear systems: The NARMAX model. *International Journal of Control*, *49*(3), 1013–1032.

Chrzan, M. J., & Carlson, J. D. (2001). MR fluid sponge devices and their use in vibration control of washing machines. 4331. pp. 370–378.

D'haene, T., Pintelon, R., Schoukens, J., & Van Gheem, E. (2005). Variance analysis of frequency response function measurements using periodic excitations. *IEEE Transactions on Instrumentation and Measurement*, *54*(4), 1452–1456.

Dobrowiecki, T., & Schoukens, J. (2007). Measuring a linear approximation to weakly nonlinear MIMO systems. *Automatica*, *43*(10), 1737–1751.

Enqvist, M. (2005). Linear models of nonlinear systems. *Ph.D. thesis*. Linkoping University, Linkoping., Sweden..

Enqvist, M., & Ljung, L. (2005). Linear approximations of nonlinear FIR systems for separable input processes. *Automatica*, *41*(3), 459–473.

Eykhoff, P. (1974). *System identification. Parameter and state estimation*. New York: Wiley.

Fliess, M., & Normand-Cyrot, D. (1982). On the approximation of nonlinear systems by some simple state-space models. In *Proceedings of the IFAC identification and parameter estimation conference*. USA. pp. 511–514.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations* (3rd ed.). Baltimore: John Hopkins University Press.

Khalil, H. K. (1996). *Nonlinear systems* (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, *2*, 164–168.

Ljung, L. (1999). *System identification: Theory for the user* (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, *11*, 431–441.

McKelvey, T., Akcay, H., & Ljung, L. (1996). Subspace-based multivariable system identification from frequency response data. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, *41*(7), 960–979.

Narendra, K. S., & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, *1*(1), 4–27.

Paduart, J. (2008). Identification of nonlinear systems using polynomial nonlinear state space models. *Ph.D. thesis*. Vrije Universiteit Brussel.

Pintelon, R. (2002). Frequency-domain subspace system identification using non-parametric noise models. *Automatica*, *38*(8), 1295–1311.

Pintelon, R., & Schoukens, J. (2001). *System identification: A frequency domain approach*. NJ, IEEE Press: Piscataway.

Pintelon, R., Schoukens, J., Vandersteen, G., & Rolain, Y. (1999). Identification of invariants of (over)parameterized models: Finite sample results. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, *44*(5), 1073–1077.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*(5), 465–471.

Rugh, W. J. (1981). *Nonlinear system theory. The Volterra/Wiener approach*. The John Hopkins University Press.

Schetzen, M. (1980). *The Volterra and Wiener theories of nonlinear systems*. New York: Wiley.

Schoukens, J., Dobrowiecki, T., & Pintelon, R. (1998). Parametric and nonparametric identification of linear systems in the presence of nonlinear distortions - A frequency domain approach. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, *43*(2), 176–190.

Schoukens, J., Rolain, Y., & Pintelon, R. (2006). Analysis of windowing/leakage effects in frequency response function measurements. *Automatica*, *42*(1), 27–38.

Schrempf, A. (2004). Identification of extended state-affine systems. Johannes Kepler Universitat. *Ph.D. thesis*. Austria.

Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., et al. (1995). Nonlinear black-box modeling in system identification: A unified overview. *Automatica*, *31*(12), 1691–1724.

Soderstrom, T., & Stoica, P. (1989). In E. Cliffs (Ed.)., *System identification*. Prentice Hall.

Sontag, E. D. (1979). Realization theory of discrete-time nonlinear systems: Part I: The bounded case. *Institute of Electrical and Electronics Engineers. Transactions on on Circuits and Systems*, *26*(5), 342–356.

Suykens, J. A. K., De Moor, B., & Vandewalle, J. (1995). Nonlinear system identification using neural state-space models, applicable to robust-control design. *International Journal of Control*, *62*(1), 129–152.

Suykens, J. A. K., Vandewalle, J., & De Moor, B. (1996). *Artificial neural networks for modelling and control of non-linear systems and their application to control*. Kluwer Academic Publishers.

Unbehauen, H., & Rao, G. P. (1998). A review of identification in continuous-time systems. *Annual Reviews in Control*, *22*, 145–171.

Verdult, V. (2002). Nonlinear system identification: A state-space approach. University of Twente. *Ph.D. thesis*. The Netherlands.

Weisstein, E. W. (2009). Multinomial series. From MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/MultinomialSeries.html.

Wills, A., & Ninness, B. (2008). On gradient-based search for multivariable system estimates. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, *53*(1), 298–306.

Johan Paduart was born in Belgium in 1980. He obtained the degree of Electrical Engineering in 2003 and the Ph.D. degree in applied sciences in 2008, both from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. His Ph.D. topic was based on the nonlinear system identification at the department of Fundamental Electricity and Instrumentation (ELEC) at the VUB.

Lieve Lauwers was born in Belgium in 1982. She received the degree of Electrical Engineering in 2005 from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. She joined the department of Fundamental Electricity and Instrumentation (ELEC) at the VUB, where she is currently pursuing the Ph.D. degree. Her research interests are in the field of nonlinear system identification.

Jan Swevers received the M.Sc. degree in Electrical Engineering and the Ph.D. degree in Mechanical Engineering from the Katholieke Universiteit Leuven (KULeuven), Belgium, in 1986 and 1992, respectively. Currently, he is a professor at the department of Mechanical Engineering, division Production Engineering, Machine Design and Automation (PMA), of the KULeuven. His research interests include modelling, identification, control and optimization of mechatronics systems.

Kris Smolders was born in Belgium in 1979. He obtained the degree of Mechanical Engineering in 2002 and the Ph.D. degree in Mechanical Engineering in 2007, both from the Katholieke Universiteit Leuven (KULeuven), Belgium. His research topic was based on the iterative learning control and optimization of nonlinear systems at the KULeuven, department of Mechanical Engineering, division Production Engineering, Machine Design and Automation (PMA).

Johan Schoukens (M'90-SM'92-F'97) was born in Belgium in 1957. He received the engineering degree in 1980 and the Ph.D. degree in applied sciences in 1985, both from the Vrije Universiteit Brussel, Brussels, Belgium. He is presently a professor at the Vrije Universiteit Brussel. His prime interest lies in the field of system identification for linear and nonlinear systems, and growing tomatoes in his green house.

Rik Pintelon (M'90-SM'96-F'98) was born in Gent, Belgium, on December 4, 1959. He received the electrical engineering degree in 1982, the Ph.D. degree in applied sciences in 1988, both from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. From 1982 to 2000, he was a researcher for the Fund for Scientific Research - Flanders at the VUB. Currently, he is a professor at the VUB in the Electrical Measurement Department (ELEC). His main research interests are in the field of parameter estimation, system identification, and signal processing.