Available online at www.sciencedirect.com
SCIENCE DIRECT®
Automatica 41 (2005) 491–504
automatica
www.elsevier.com/locate/automatica

Identification of linear systems with nonlinear distortions☆

J. Schoukens[a,*], R. Pintelon[a], T. Dobrowiecki[b], Y. Rolain[a]

[a]*Vrije Universiteit Brussel, Department ELEC, Pleinlaan 2, B1050 Brussels, Belgium*
[b]*Department MIS, Budapest University of Technology & Economics, H-1521 Budapest*

Received 20 October 2003; received in revised form 8 June 2004; accepted 20 October 2004
Available online 15 December 2004

Abstract

This paper studies the impact of nonlinear distortions on linear system identification. It collects a number of previously published methods in a fully integrated approach to measure and model these systems from experimental data. First a theoretical framework is proposed that extends the linear system description to include the impact of nonlinear distortions: the nonlinear system is replaced by a linear model plus a 'nonlinear noise source'. The class of nonlinear systems covered by this approach is described and the properties of the extended linear representation are studied. These results are used to design the experiments; to detect the level of the nonlinear distortions; to measure efficiently the 'best' linear approximation; to reveal the even or odd nature of the nonlinearity; to identify a parametric linear model; and to improve the model selection procedures in the presence of nonlinear distortions.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: System identification; Linear system; Nonlinear system

1. Introduction

Identification of linear systems became a mature scientific discipline over the last decades (Ljung, 1999; Söderström & Stoica, 1989; Pintelon & Schoukens, 2001). It is a very successful method and is applied on a large variety of problems coming from a wide range of different fields. The basic reason for this success is the appealing simplicity of linear models. They give a lot of insight and are often used as the basis for many design techniques. The price for this 'simplicity' is the need for a strong assumption: the system is assumed to behave linearly. However, in practice many systems are not linear. If the linear modelling approach is maintained, the question arises if the whole framework is still valid, and if its results are still reliable. This leads to the questions that are addressed in this paper:

- What is the validity of a linear model that is identified in the presence of nonlinear distortions?
- Can we detect, qualify and quantify the presence of nonlinear distortions?
- What is the best 'engineering practice' to obtain a linear model under these conditions?
- Can the convergence results of the linear identification framework be maintained under these conditions?

Very often the user is not aware of the presence of nonlinear distortions because the classical linear identification framework can be completely fooled: a nonlinear system driven by random excitations can be modelled as a linear system that passes all usual validation tests like a whiteness test of the residuals, or a cross-correlation test between the input and the residuals. This is a dangerous situation, because the user will rely on an invalid model for the rest of his design without being aware of it. In the best case this leads to a loss in performance, but in the worst case it might even lead to unstable control loops. For that reason we strongly advice

☆ This paper was presented at IFAC Symposium on System Identification 2003, Rotterdam. This paper was recommended for publication in revised form by Guest Editors Torsten Söderström, Paul van den Hof, Bo Wahlberg and Siep Weiland. This work is supported by the Flemish government (FWO-ICCoS onderzoeksgemeenschap; Concerted Action ILiNoS; the Bilateral agreement program) and the Belgium government (IUAP-5/22).
* Corresponding author.
E-mail address: johan.schoukens@vub.ac.be (J. Schoukens).

0005-1098/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2004.10.004

to add to each identification experiment a nonlinearity test to reveal the level of the distortions.

If nonlinear distortions are detected, the user can either abandon the linear framework in favour of a nonlinear model, or the user can decide to go on with linear approximations. Both decisions might be valid as long as they are the result of a conscious decision. In practice it is not always an obvious choice to move from linear to nonlinear modelling, because the required additional modelling efforts are large while there is no guarantee for success. With the approach proposed in this paper, the user will get in advance an idea what is the validity of the linear model, and how much might be gained by moving to nonlinear models.

The results of this approximate linear modelling process can be used in all applications where till now a classical linear identification approach is used. The major difference with the actual common practice is that the user gets additional bounds on its model that indicate the validity of the linear assumptions. Since these will also depend upon the actual applied excitation, all results are conditioned on the class of excitation signals that was used during the experiments. This might seem an enormous drawback, but the reader should realize that nowadays this restriction exists also, but most users are unaware of it. They live in peace, ignorant and dangerous.

The paper consists of 3 parts. First the system setup will be defined, followed by a precise definition of the class of nonlinear systems that is considered. Next the impact of nonlinear distortions on the linear identification framework is analysed and an integrated approach is setup how to differentiate between all the different contributions (linear, nonlinear, process noise) to the output. Eventually, the linear identification in the presence of nonlinear distortions is discussed.

2. Setup

Consider the time invariant, single-input, single-output (SISO), continuous or discrete time nonlinear dynamic system g_{NL}:

$$y_0 = g_{\mathrm{NL}}(u_0) \tag{1}$$

and the discrete observations:

$$u_0(t) \text{ and } y(t) = y_0(t) + n_y(t), \quad t = 0, 1, \ldots, N, \tag{2}$$

where $n_y(t)$ is zero mean noise. The input $u_0(t)$ is assumed to be known exactly. The sampling period is normalized to $T_s = 1$.

Assumption 1. *Output noise model.* $n_y(t)$ is filtered white noise $n_y(t) = H_0(q)e_0(t)$ where $e_0(t)$ is a sequence of independent random variables, with zero mean values, variances λ_0, and bounded moments of order $4 + \delta$, for some $\delta > 0$. $H_0(q)$ is a stable and inversely stable, monic filter (Ljung, 1999).

3. A formal framework to describe the nonlinear system

Describing nonlinear systems is a tedious job. Since there does not exist a single model structure that covers all possible nonlinear systems it is necessary to specify what subclass S of nonlinear systems will be considered. The nonlinear systems that can be included in S will depend on: (i) The class of excitation signals that will be allowed (for example single sine excitations or random noise excitations, uniformly or normally distributed noise); (ii) The model class that will be used to describe the input–output relation (Volterra models, neural nets, support vector machines, narmax models, etc.); (iii) The convergence criterion that will be used to match the model and the system output. These three choices are specified below.

3.1. Class of excitation signals E

3.1.1. Introduction

In this paper normally distributed random excitations with a user defined power spectrum $S_u(f)$ will be used. Below we give the precise definitions of the considered excitation signals followed by an illustration of each of these signals in time- and frequency-domain, and some remarks. All excitations are defined by a discrete time sequence $u(t)$. The actual continuous time excitation signal $\widehat{u}(\tau)$ that is applied to the physical system is obtained from this discrete time signal by passing it through a hybrid reconstruction filter: $\widehat{u}(\tau) = \sum_{t=-\infty}^{\infty} u(t)l(\tau - t)$. The most popular reconstruction is to keep the signal constant between two successive samples. This is called the zero-order-hold reconstruction and it corresponds to $l(\tau) = 1$ if $0 \leqslant \tau < 1$, and zero elsewhere. We do not further elaborate on these aspects in this paper.

3.1.2. Definitions

Definition 2. Spectrum generating function. $S_u(f)$ is a uniformly bounded real positive function with a countable number of discontinuities, and $S_u(0) = 0$.

Remark. $S_u(f)$ will be used as the power spectrum for the noise excitations. For periodic excitations it will be used to set the amplitude of the discrete spectral components. The DC value $S_u(0)$ is set equal to zero, the system is described and modelled around its operating point.

Definition 3. Gaussian noise excitation. A random sequence $u(t)$, $t = 0, 1, \ldots, N$ drawn from a zero mean normally distributed process with a user defined power spectrum $S_u(f)$.

Definition 4. Periodic noise. A signal $u(t)$ is a periodic noise excitation if

$$u(t+N)=u(t), \quad \forall t, N \text{ is the period, and } u(t), \\ t=0,\ldots,N-1 \text{ is a Gaussian noise excitation.} \tag{3}$$

Remark. A periodic noise sequence can also be written as a finite Fourier series:

$$u(t)=N^{-1/2}\sum_{k=-N/2}^{N/2}\hat{U}\left(\frac{k}{N}\right)e^{j(2\pi k\,\frac{t}{N}+\varphi_k)}, \tag{4}$$

with $\varphi_{-k}=-\varphi_k$, $\hat{U}(f)\geqslant 0$, N even, and $\hat{U}(f=0)=0\cdot\hat{U}(k/N)$ and φ_k are the realisations of independent (jointly, and over k) random processes satisfying the following condition: $\hat{U}(f)$ has bounded moments of any order ($<\infty$), and $\mathscr{E}\{e^{j\varphi_k}\}=0$, and $\mathscr{E}\{\hat{U}(f)^2\}=S_u(f)$. This could be used as an alternative for Definition 4.

Definition 5. Random (phase) multisine. A signal $u(t)$ is a random phase multisine (also called random multisine) if

$$u(t)=N^{-1/2}\sum_{k=-N/2}^{N/2}\sqrt{S_u\left(\frac{k}{N}\right)}e^{j(2\pi k\,\frac{t}{N}+\varphi_k)}, \tag{5}$$

with $\varphi_{-k}=-\varphi_k$. The phases φ_k are the realisations of an independent (over k) uniformly distributed random process on $[0,2\pi)$.

The set of excitation signals that is considered in this paper is the union of the signals defined before.

Definition 6. Set of excitation signals E: A signal u with power spectrum $S_u(f)$ belongs to E if u is either a Gaussian noise excitation (Definition 3), a periodic noise (Definition 4), or a random phase multisine (Definition 5).

3.1.3. Illustration

In Fig. 1 an example for each of these signals is shown. The Fourier spectrum of an infinite stationary (Gaussian) noise sequence does not exist (only its power spectrum $S_u(f)$ which is the Fourier transform of the auto correlation function $R_u(\tau)=\mathscr{E}\{u(t)u(t-\tau)\}$ is well defined). However, for a finite record $u(k), k=0,1,\ldots,N-1$, we can still calculate the discrete Fourier transform (DFT) $U(k), k=0,1,\ldots,N/2$ and plot $|U|$ on top of the amplitude spectrum $\sqrt{S_u(k/N)}$. This shows also that although the excitation has a continuous spectrum, the resolution of the measurement is still restricted to $1/N$, which is the frequency resolution of the DFT. It is also important to observe that the actual realized amplitude spectrum differs strongly from the underlying amplitude spectrum $\sqrt{S_u(k/N)}$. This results at some frequencies in a significant drop of the signal-to-noise ratio (SNR). A periodically repeated Gaussian noise sequence has a discrete frequency spectrum, but its behaviour is completely similar to that of Gaussian noise. A random phase multisine eliminates the random variations of the amplitude spectrum, the user gets full control over the actual realized amplitude spectrum while the signal keeps still a random behaviour due to its random phase choice.

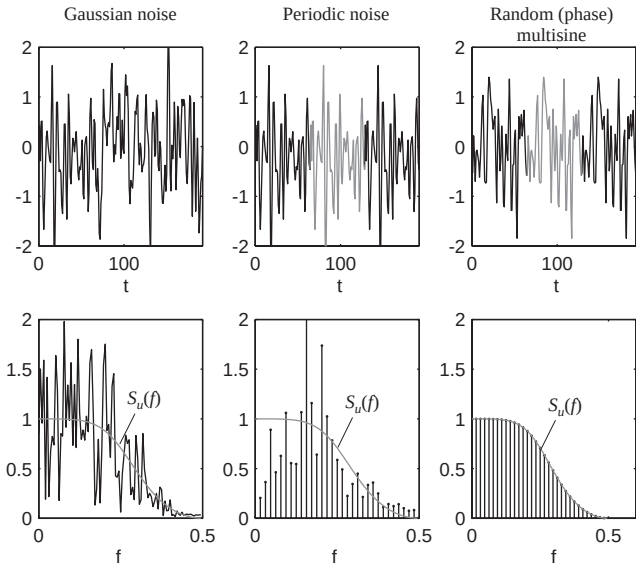

Fig. 1. Illustration of a Gaussian noise, periodic noise and random multisine excitation. Top: time domain; bottom: frequency domain.

3.1.4. Remarks

- The periodic signals (Definitions 4 and 5) are asymptotically ($N\to\infty$) normally distributed in the time domain.
- The phase condition in Definitions 4 and 5 can be relaxed. The phases can be restricted to a discrete set as long as $\mathscr{E}\{e^{j\varphi}\}$ is zero. This allows for example to include orthogonal frequency domain modulation (OFDM) where such random multisines are intensively used (Vandersteen, Verbeeck, Rolain, & Schoukens, 2000).
- The period length N will be sometimes indicated explicitly by using the subscript N, for example E_N.

3.2. Class S of nonlinear systems

In order to define the class of nonlinear systems that will fit into the framework of this paper, we need first to select the model class that will be used as a mathematical vehicle during the proofs. Next also the convergence criterion that will be used to match the model output to the system output should be specified because it has a direct impact on what systems are covered or not. Combination of both selections sets eventually the class of systems that is considered in this paper.

Both choices are discussed below and eventually a short overview is given about the properties of the approximation as a function of the system assumptions (continuity, fading, memory, etc.). It should be emphasized that we are not interested at all in the identification of these models. They are only used to allow for formal proofs of the claimed results.

3.2.1. Nonlinear model

In this paper we focus on Volterra models. For many people, these have a bad reputation because they seem to cover only a very restricted class of systems, and they are difficult to parametrize and to identify. Since we do not intend to identify the Volterra models, the last remarks are not an issue. Moreover, it will turn out that the covered class of systems can be significantly enlarged by selecting a proper convergence criterion as will be shown below.

Definition 7. Class M of nonlinear models: M is the set of Volterra models

$$y_Q(t) = \sum_{n=1}^{Q} y^{n(t)},$$
$$\text{with } y^{n(t)} = \int_0^\infty \cdots \int_0^\infty g_n(\tau_1, \ldots, \tau_n) u(t-\tau_1) \ldots u(t-\tau_n)\, \mathrm{d}\tau_1 \ldots \mathrm{d}\tau_n, \quad (6)$$

for continuous time systems and a similar expression for discrete time systems (replacing the integrals by sums). Because there appears an infinite sum in this definition, it is necessary to consider its convergence. This is discussed in the next section.

3.2.2. Convergence criterion

Depending on the convergence criterion that is selected, a wider or a smaller class of systems can be approximated by the Volterra model. Also the properties of the approximation will change: for example convergence of the model output (and its derivatives) to the system output (and its derivatives), convergence of the model to the system, convergence of the derivatives of the model to the derivatives of the system. In this paper we can only give a simplified discussion. For a more thorough discussion, the reader is referred to the literature mentioned below.

The most general class of systems is retrieved if mean square convergence of the system and model output is selected: this paper considers SISO nonlinear time invariant systems whose output can be approximated arbitrarily well in least squares sense by a Volterra series (6) for the class of excitations E. This leads eventually to the formal definition of the considered class of systems:

Definition 8. Class S of nonlinear systems. S is the set of nonlinear systems for which there exists a Volterra series representation $y_Q(t) = \sum_{n=1}^{Q} y^{n(t)}$ that converges in mean squares sense with probability 1 to $y(t)$ for all excitations $u \in E$:

$$\lim_{Q\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathscr{E}\{|y(t) - y_Q(t)|^2\} = 0, \quad (7)$$

with N the experiment length. The expected value $\mathscr{E}\{\}$ is the ensemble average over the considered class of random inputs u.

For these systems (called Wiener systems, not to be confused with the cascade of a linear dynamic system followed by a static nonlinearity that is also called a Wiener system) a number of statements can be made: (i) the influence of the initial conditions vanishes asymptotically ($N \to \infty$), (ii) the steady state response to a periodic input is a periodic signal with the same period as the input. Phenomena such as bifurcation, chaos, and sub harmonics are excluded, while strongly nonlinear phenomena such as saturation (e.g. amplifiers) and discontinuities (e.g. relays) are allowed. (iii) Only a point wise approximation of the output is obtained (see the Wiener theory in Schetzen, 1980; Doyle, Pearson, & Ogunnaike, 2001). Also the derivatives of the output converge for band limited inputs ($S_u(f) = 0$ for $|f| > f_{max}$). Hence, these models can be used to model for example the output spectrum, but they should not be used to calculate the derivative of model characteristics. (iv) The output of the system at the discontinuities cannot be modelled.

3.2.3. Remarks

A stronger convergence result can be obtained if only fading memory systems are considered. Stating Boyd and Chua (1985): 'Intuitively, an operator has fading memory if two input signals which are close in the recent past, but not necessarily close in the remote past yield present outputs which are close'. Discontinuous nonlinear systems are excluded but hard saturating nonlinear systems can still be modelled. For these systems uniform convergence of the model (output) to the system (output) is shown, while the model derivatives are still not guaranteed to converge. The approximation is valid for bounded inputs, where the bounds can be set by the user. The properties of fading memory systems (or approximately finite memory systems) are extensively discussed in Boyd and Chua (1985), Borys (2000) and the work of Sandberg (1992, 1993, 2002).

Restricting the class of nonlinear systems even more, to those systems having a Volterra series that converges uniformly around a given working point, similar to a convergent Taylor series for a static nonlinear system, gives the strongest convergence results. We call these systems Volterra systems. For these systems uniform convergence of the model (output) and the derivatives is guaranteed. Often these Volterra series exist only in a restricted input domain that cannot be freely chosen by the user. Consider for example the Taylor series of arctan x that exists only for $|x| < 1$.

The previous results are summarized in Table 1.

Table 1
Convergence properties of the different nonlinear model classes

Model class	Properties
Wiener system	- Output convergences in mean square sense, point wise convergence - Discontinuities and saturation allowed - Model valid for the set of Gaussian signals
Fading memory system	- Output converges uniformly - Saturation allowed - Model valid for bounded inputs (bound set by the user)
Volterra system	- Output converges uniformly - Derivatives model converge uniformly - Saturation allowed - Model valid for bounded inputs (bound cannot be set by the user)

It should be emphasized once more that we are not interested in the identification of the Volterra kernels. The Volterra model is only used to allow for formal proofs of the claimed results for a given class of inputs E and systems S.

4. Impact of nonlinear distortions on the linear framework

4.1. Why to use a linear model for a nonlinear system?

Linear models are very popular, even if it is well known that in practice many systems are not perfectly linear, because they offer important advantages: (i) They result in useful models that give the user a lot of intuitive insight in the system behaviour; (ii) Many design techniques are valid for linear models only; (iii) Nonlinear model building is often difficult and time consuming, and often the user cannot afford or is not prepared to make this huge effort; (iiii) No general framework is available for nonlinear systems. Dedicated models are needed, complicating the development/use of general software packages.

For all these reasons, there exists a strong need to use linear models even if it is known that they are erroneous. To allow extension of the linear framework to systems with a dominant linear behaviour, the nonlinear contributions are considered to be a parasitic effect.

4.2. Non parametric linear framework

Although the discussion covers equally well time domain and frequency domain identification methods, we will mostly use a frequency domain representation using the DFT:

$$X(k) = N^{-1/2} \sum_{t=0}^{N-1} x(t) \mathrm{e}^{-\mathrm{j}2\pi kt/N}, \tag{8}$$

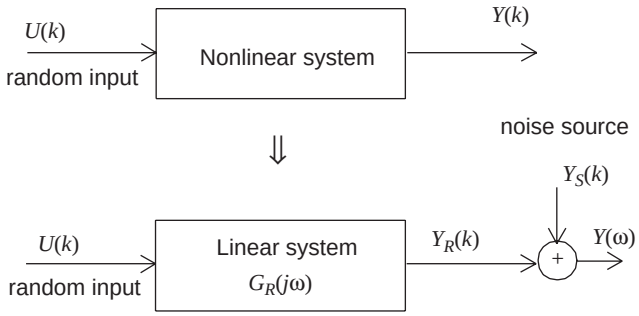

Fig. 2. Representation of a nonlinear system by a linear system for a random input.

where N is the considered record length. For simplicity we do not consider initial conditions and leakage effects because they have no fundamental impact on the interpretations that will be made. Leakage effects in the frequency domain are equivalent to initial- and end-condition effects in the time domain (Pintelon, Schoukens, & Vandersteen, 1997; Schoukens, Pintelon, & Rolain, 1999).

4.2.1. Major result: intuitive presentation

A nonlinear system belonging to S, excited with a random excitation $u \in E$ can be represented by a linear system G_R plus an error term Y_S (see Fig. 2). Consider R experiments $U^{[r]}, Y^{[r]}, r = 1, \ldots, R$, each obtained with a different realization of the random input. Then

$$Y^{[r]}(k) = G_{RN}(\mathrm{j}\omega_k)U^{[r]}(k) + Y_S^{[r]}(k),$$

with

$$\begin{aligned} G_{RN}(\mathrm{j}\omega_k) &= \arg\min_G \sum_{r=1}^{M} |Y^{[r]}(k) - G(\mathrm{j}\omega_k)U^{[r]}(k)|^2 \\ &= \arg\min_G \sum_{r=1}^{M} |Y_S^{[r]}(k)|^2. \end{aligned} \tag{9}$$

This is an exact representation of the nonlinear system, all approximation errors are put into the second term Y_S. The index N in G_{RN} indicates that the estimate is obtained from experiments with a length of N data points. It turns out that for N growing to infinity, $Y_s(k)$ looks like complex Gaussian noise. It changes from one realization of the input to the other. G_{RN} is called the best linear approximation. Both terms are discussed a bit more in detail below.

Discussion of G_{RN}: the best linear approximation

Define,

$$\lim_{N\to\infty} G_{RN}(\mathrm{j}\omega) = G_R(\mathrm{j}\omega), \tag{10}$$

then we have that

$$\mathscr{E}\{G_{RN}(\mathrm{j}\omega)\} = G_R(\mathrm{j}\omega) + O(N^{-1}), \tag{11}$$

where the expected value $\mathscr{E}\{\}$ is taken with respect to the random input phases (averaging over different realizations of the input), and $O(N^{-1})$ indicates a uniform convergence over ω. For that reason the dependency on N will be mostly dropped in the notation G_{RN}.

For Gaussian distributed excitations, $G_R(\mathrm{j}\omega)$ depends solely on the power spectrum S_u of the excitation (Schoukens, Dobrowiecki, & Pintelon, 1998; Pintelon & Schoukens, 2002b; Evans, Rees, & Jones, 1994b), and not on its actual random phase realization. Note that the central limit theorem implies that a random multisine has asymptotically a Gaussian amplitude distribution. Y_R consist of the sum of all those output contributions that are 'coherent' with the input and can be written as

$$Y_R(k) = G_R(\mathrm{j}\omega_k)U(k). \tag{12}$$

To illustrate a typical nonlinear contribution to $Y_R(k)$, the third degree nonlinear kernel $g_n(\tau_1, \tau_2, \tau_3)$ with transfer function $G_3(\mathrm{j}\omega_1, \mathrm{j}\omega_2, \mathrm{j}\omega_3)$ is considered. The contribution of such a kernel to the output appears at $\omega_1 + \omega_2 + \omega_3$ and equals $G_3(\mathrm{j}\omega_1, \mathrm{j}\omega_2, \mathrm{j}\omega_3)U(k_1)U(k_2)U(k_3)$. For $\omega_1 = \omega_l$, $\omega_2 = -\omega_l$, and $\omega_3 = \omega_k$ the contribution to the output at $\omega = \omega_k$ is given by

$$\begin{aligned} &G_3(\mathrm{j}\omega_l, -\mathrm{j}\omega_l, \mathrm{j}\omega_k)U(l)U(-l)U(k) \\ &\quad = \{G_3(\mathrm{j}\omega_l, -\mathrm{j}\omega_l, \mathrm{j}\omega_k)|U(l)|^2\}U(k). \end{aligned} \tag{13}$$

Note that only the phase $\varphi_k = \mathrm{phase}(U(k))$ of the kth input component comes into the output $Y_R(k)$, all the other input phases $(\varphi_l, l \neq k)$ are cancelled in these contributions by making combinations like $U(l)U(-l) = |U(l)|^2$. This results in a nonlinear contribution to the best linear approximation $G_R(\mathrm{j}\omega_k)$ given by

$$\begin{aligned} &\frac{\{G_3(\mathrm{j}\omega_l, -\mathrm{j}\omega_l, \mathrm{j}\omega_k)|U(l)|^2\}U(k)}{U(k)} \\ &\quad = G_3(\mathrm{j}\omega_l, -\mathrm{j}\omega_l, \mathrm{j}\omega_k)|U(l)|^2. \end{aligned} \tag{14}$$

Such combinations can only be found for odd degree nonlinearities $G_{2\alpha+1}$, it is impossible to get such a result for $G_{2\alpha}$ if $U(0) = 0$ (see Definition 4).

The reader should be aware that these results are only valid for inputs belonging to E. Replacing for example a Gaussian excitation by a binary noise source with the same power spectrum can result in a completely different linear approximation.

Discussion of Y_S, the 'nonlinear noise source'

A typical nonlinear contribution to $Y_S(k)$ is for example

$$G_3(\mathrm{j}\omega_{k_1}, \mathrm{j}\omega_{k_2}, \mathrm{j}\omega_{k-k_1-k_2})U(k_1)U(k_1)U(k-k_1-k_2), \quad \text{with } k_1 + k_2 \neq 0. \tag{15}$$

This is a stochastic contribution because the phase of (15) is not within a constant equal to φ_k, it depends explicitly on the input phases at other frequencies. The stochastic contributions $Y_S(k)$ contain all those output contributions for which

$$\begin{aligned} &U(k_1)U(k_1)\ldots U(k-k_1-\cdots-k_{n-1}) \\ &\quad = \alpha \mathrm{e}^{\mathrm{j}\beta}\mathrm{e}^{\mathrm{j}\varphi_k} \\ &\quad \neq |U(k_1)U(k_1)\cdots U(k-k_1-\cdots-k_{n-1})|\mathrm{e}^{\mathrm{j}\varphi_k}. \end{aligned} \tag{16}$$

Hence it is impossible to write this output as $Y_S(k) = G_R(\mathrm{j}\omega_k)U(k)$ without G_R depending on the actual input phases (Pintelon & Schoukens, 2001; Evans & Rees, 2000). For that reason these terms do not fit in the linear system representation (12). Due to the remaining phase dependency β and the random phase choices of the input, these terms also have a random phase. As such these contributions look very similar to noise, and it is hard to distinguish them from process noise. Most linear validation procedures of the classical linear identification framework do not recognize the presence of these nonlinear contributions, they just classify them as 'regular' noise. For that reason we prefer to call these terms 'the nonlinear noise contributions', although formally spoken it are deterministic signals once the input signal is drawn. Their properties will be characterized below in the formal description.

The contributions $Y_S(k)/U(k)$ are called $G_S(\mathrm{j}\omega_k)$.

4.2.2. Major result: formal presentation

In this section the formal statements of the previous results are made. The proofs are outside the scope of this paper, they can be found in the book Pintelon and Schoukens (2001), or in the references given below the theorem. The basic idea is to count the number of possible systematic (13) and stochastic (16) contributions to the output.

Theorem 1. *A nonlinear system represented by its best linear approximation and an error term*:

The nonlinear system: The output of the nonlinear system belonging to S excited with $u \in E$ can be written as

$$Y(k) = G_R(\mathrm{j}\omega_k)U(k) + Y_S(k), \tag{17}$$

with G_R the linear approximation and Y_S the error term:

$$G_R(\mathrm{j}\omega_k) = \arg\min_G \mathscr{E}\{|Y(k) - GU(k)|^2\}, \quad \textit{or}$$

$$G_R(\mathrm{j}\omega_k) = \frac{\mathscr{E}\{Y(k)\overline{U}(k)\}}{\mathscr{E}\{U(k)\overline{U}(k)\}} = S_{yu}(k)/S_u(k). \tag{18}$$

The nonlinear noise source: *The nonlinear noise source Y_S has the following properties for S and $u \in E$*:

(1) *Zero mean*: $\mathscr{E}\{Y_S(k)\} = 0$.
(2) *Uncorrelated with the input* $\mathscr{E}\{Y_S(k)\overline{U}(k)\} = 0$.
(3) *$Y_S(k)$ is asymptotically independent from $U(l)$, $\forall k, l$.*
(4) *$Y_S(k)$ is asymptotically circular complex normally distributed and mixing of arbitrarily order.*

(5) *The even moments do not disappear*:

$$\mathscr{E}\{N|Y_S(k)|^2\} = \sigma^2_{Y_S}(k) = O(N^0) = O(1);$$

$$\mathscr{E}\{N^2(|Y_S(k)|^2 - \sigma^2_{Y_S}(k))(|Y_S(l)|^2 - \sigma^2_{Y_S}(l))\} = O(N^{-1}) \text{ if } k \neq l, \text{ and an } O(1) \text{ if } k = l.$$

(6) *The odd moments converge to zero* ($k \neq l$):

$$\mathscr{E}\{NY_S(l)\overline{Y}_S(k)\} = O(N^{-1}),$$
$$\mathscr{E}\{N^{3/2}Y_S(l)|Y_S(k)|^2\} = O(N^{-1}),$$
$$\mathscr{E}\{N^2(|Y_S(k)|^2 - \sigma^2_{Y_S}(k))(|Y_S(l)|^2 - \sigma^2_{Y_S}(l))\} = O(N^{-1}).$$

The expected value $\mathscr{E}\{\}$ has to be taken over different realizations of the excitation.

Proof. See Pintelon and Schoukens (2001) and Schoukens et al. (1998) for random multisines, and Pintelon and Schoukens (2002a) for the generalization to periodic noise and Gaussian noise.

Remarks.

- In the rest of this paper, we simplify the statement 'Nonlinear system represented by its best linear approximation and an error term' to 'Linear representation of a nonlinear system', not only because it is simpler in daily use, but also because it reflects very well what is almost systematically done in practice: a nonlinear reality is modelled with a linear model. Also the 'nonlinear noise source' will be used instead of 'error term', again because it corresponds to the daily practice where the nonlinear errors are often not recognized as such, but modelled by a filtered white noise source.
- Note that (18) is the classical result for frequency response function (FRF) measurements of linear systems (Bendat & Piersol, 1980). It is also connected to the early results of Bussgang (1952) and the work reported in Brillinger (1981).
- These observations are in agreement with the classical result, showing that the output of a nonlinear system can be split in two parts (Bendat 1990, 1998; Forsell & Ljung, 2000): a first part that is linearly related with the input (in our case leading to G_R), and a second part that is uncorrelated with the input (leading to Y_S).
- The independency claims in the frequency domain $(Y_S(k), U(l))$ are not in conflict with the obvious dependency of $y_S(t)$ on $u(t)$ ($y_S(t)$ is a periodic signal with the same period as $u(t)$). These asymptotic results are valid on a frequency by frequency basis. However, if arbitrary large numbers of such components are combined (as for the inverse Fourier transform to calculate the time domain signals), the independency is not necessarily maintained.

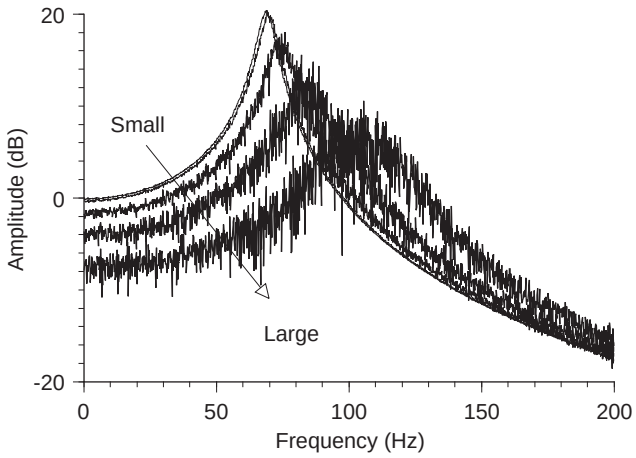

Fig. 3. Evolution of the FRF for growing excitation levels: 34, 54, 127, 253, and 507 mVRMS.

- All these properties can be transferred to G_S by replacing $\sqrt{N}Y_S$ by G_S in the theorem.
- A detailed interpretation of statement 4 is given in Pintelon and Schoukens (2001), p. 437.

4.2.3. Experimental illustration

In Fig. 3 the previously discussed aspects are illustrated by experimental results on a nonlinear circuit (Pintelon & Schoukens, 2001). The FRF of G_R of a nonlinear circuit is measured for different levels of the excitation signal (a random multisine). As the amplitude grows, it is seen that the FRF shifts due to the growing systematic contributions like (13), and at the same time it looks more noisy due to the growing nonlinear noise source contributions like (15). Since the small amplitude measurements are smooth, and the disturbing noise conditions do not change with the excitation level, it is clear that the dominant 'noise' effects are due to the nonlinearity.

4.3. Best engineering practice for FRF measurements: reduction of the nonlinear noise source level.

In this section we show that a good choice of the excitation signal can significantly reduce the disturbances that come from the nonlinear noise source. In the beginning of the paper, we mentioned that all signals in the set E result in the same best linear approximation G_R (Pintelon & Schoukens, 2002a). This is formulated below more formally, next we select within the class E the best signals for FRF measurements. Once it is known that all these signals are equivalent, the question rises if one class of excitations out of all possible choices has a better behaviour than the others? Can a good choice reduce the effect of the nonlinear noise source on the measurement of G_R? To answer this question, we need to introduce even/odd nonlinearities, and odd excitations.

Even/odd nonlinearities: Each static or dynamic nonlinearity can be written as the sum of an even nonlinear system

$$NL(u) = NL_{\text{even}}(u) + NL_{\text{odd}}(u), \tag{19}$$

with $NL_{\text{even}}(-u) = NL_{\text{even}}(u)$ an even nonlinearity, and $NL_{\text{odd}}(-u) = -NL_{\text{odd}}(u)$ an odd nonlinearity.

Even nonlinearities do not contribute to the best linear approximation G_R because it is impossible to make combinations of the form given in (13) that contain always an odd number of input spectral components, while an even nonlinearity combines an even number of spectral components. Hence they only contribute to the nonlinear noise Y_S which acts as an disturbance during the measurement of G_R (Pintelon & Schoukens, 2001).

Odd excitations are periodic signals that excite only the odd frequencies $2k+1$, the amplitudes of the even frequencies $2k$ are put to zero ($U(2k)=0$). Such a signal can be simply generated as an inverse repeated noise excitation $[u\ -u]$, with u a Gaussian noise excitation, or by putting $U(2k)=0$ (Definition 5 of the random multisine). Other possibilities are discussed in Godfrey (1993).

An even nonlinearity transfers all the power of an odd excitation to the even frequencies. Because G_R is only measured at the excited frequencies (the odd frequencies), the measurements will be no longer disturbed by the even nonlinearity, and the uncertainty on G_R drops. These possibilities are extensively studied (Dobrowiecki & Schoukens, 2001b; Pintelon & Schoukens, 2001) where a number of possibilities are proposed to create such odd excitation signals.

4.3.1. Formal result

The index N is put here explicitly in $G_{R,N}$ because the convergence of G_R is discussed.

Theorem 2. *Equivalencies of the excitation signals: Consider Gaussian random noise, periodic noise and the random multisine* (*Definitions* 3–5) *with power spectrum* S_u. *For these three classes of excitation signals and for a nonlinear system belonging to the class S* (*see Definition* 8) *we have that*:

- $G_{R,N}(\mathrm{j}\omega)$, *converges* ($N \to \infty$) *at the rate* $O(N^{-1})$ *to the same limit value* $G_R(\mathrm{j}\omega)$.
- $G_R(\mathrm{j}\omega)$ *depends only on the odd nonlinear contributions, the best linear approximation of an even nonlinear distortion is zero*.
- *The variances* $\mathrm{var}(\sqrt{N}Y_{S,N}(k))$ *of the stochastic nonlinear distortions converge* ($N \to \infty$) *at the rate* $O(N^{-1})$ *to the same limit value* $\sigma_S^2(f)$.
- *The scaled moments specified in Theorem* 1, *point* 6, *converge* ($N \to \infty$) *at the rate* $O(N^{-1})$ *to the same limit value*.
- $G_R(\mathrm{j}\omega)$ *is a continuous function of* ω *with continuous higher order derivatives if the approximating Volterra system and its derivatives are continuous*.

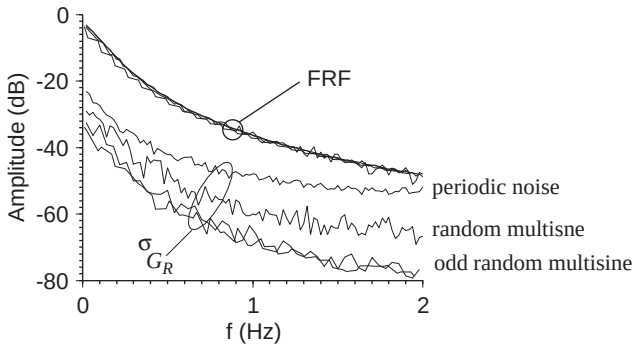

Fig. 4. Hair dryer experiment. Impact of the excitation signal on the uncertainty (σ_{G_R}) due to stochastic nonlinear distortions.

Proof. Pintelon and Schoukens (2002a).

4.3.2. Experimental illustration

The equivalency between the excitation signals with respect to the measurement of the FRF, and the possibility to reduce the uncertainty due to the stochastic nonlinear distortions is illustrated in Fig. 4 on a hair dryer device. The temperature at the outlet is measured as a function of the set point variations at the input. The heating of the hair dryer is controlled by a thyristor with an important even nonlinear contribution in its characteristic. The measured FRF and its uncertainty are shown for different kinds of excitation signals (Németh & Vargha, 1999). From the previous section it turns out that it should be possible to reduce the uncertainty on the FRF measurement without changing its expected value by using odd excitations. In this experiment three excitation signals were used. The first is a periodic noise excitation. It gives very poor results due to the drops in its actual realized power spectrum (see Section 3.1). Better results are obtained by using random multisines. The 'full' multisine excites all frequencies, including the even, while the odd and special odd multisines only excite the odd frequencies. As could be expected from the previous discussion, the expected value of the FRFs does not depend on the specific class of excitations (upper traces in Fig. 4), while the uncertainties are significantly different (lower traces). In this case the uncertainty is reduced with 20 dB or more, which corresponds to a reduction in measurement of a factor 100 or more! This illustrates nicely that a good understanding of the nonlinear behaviour can result in enormous reductions of measurement time or disturbances levels.

4.4. Detection, qualification and quantification of nonlinear distortions

The level of the nonlinear distortions provides valuable information for the user, even in a linear modelling framework, because it will give natural bounds on the validity of the linear models. However, to be useful, not too much time should be lost to get this information, most of the time

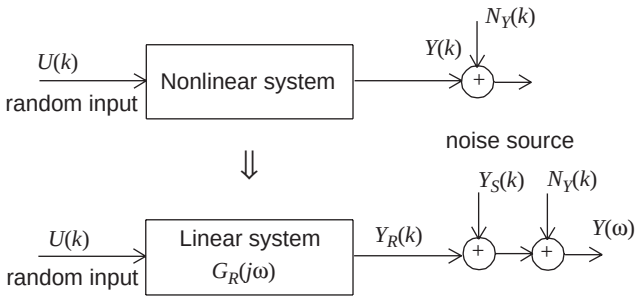

Fig. 5. Representation of a nonlinear system by a linear system for a random input in the presence of process noise.

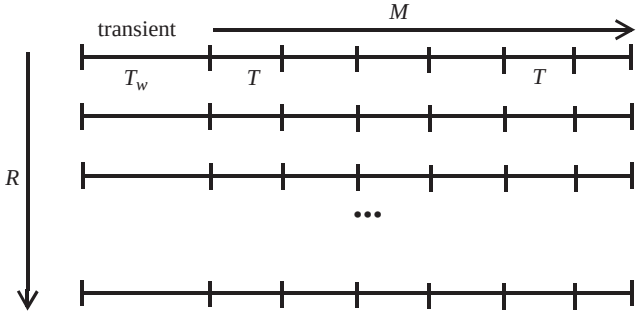

Fig. 6. Applying R realizations of the excitation, and measuring each time M periods after a waiting time T_W.

should be spent on the ultimate goal which is to obtain a (linear) model.

In the full problem, two distortions are faced at the same time: the measurement/process noise N_Y and the nonlinear noise source Y_S (see Fig. 5) should be separated from each other. At the same time the FRF of G_R should be measured.

A large number of methods are developed to detect the presence of nonlinear distortions. An extensive overview is given in (Vanhoenacker, Schoukens, Swevers, & Vaes, 2002). Many of these are very time consuming and require dedicated experiments. Only few give detailed information about the distortion levels. Here we present two simple methods that allow to measure explicitly the nonlinear and the disturbing noise levels using periodic excitation signals (periodic random or random multisines) while most of the experiment time is still used to measure the FRF. The first method allows to detect the level of the nonlinear distortions, the second not only detects the level but also qualifies the nonlinearity as even or odd.

4.4.1. Detection of the level of the stochastic nonlinearities (Schoukens, Swevers, Pintelon, & Van der Auweraer, 2002a)

The basic idea is to apply R realizations of a periodic signal and to measure the response to each input over $M \geqslant 2$ periods once the transients disappeared as shown in Fig. 6. Two variances, $\sigma^2_{G,p}(k)$ and $\sigma^2_{G,r}(k)$, are calculated. The first is the sample variance of the FRF measured over the successive periods for a single realization of the input, the other is the variance measured over the different realizations.

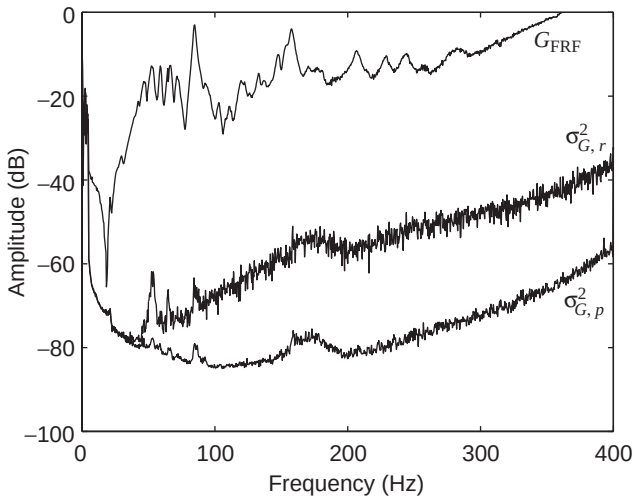

Fig. 7. Detection of nonlinear distortions on a body in white.

Define $G^{[r,m]}$ as the FRF measured in period m of realization r : $G^{[r,m]}(\mathrm{j}\omega_k) = Y^{[r,m]}(k)/U^{[r,m]}(k)$, with $U^{[r,m]}$, $Y^{[r,m]}$ the DFT spectra of the corresponding input/output data blocks. Then

$$\sigma^2_{G,p}(k) = \frac{1}{R(M-1)} \sum_{r=1}^{R} \sum_{m=1}^{M} |G^{[r,m]} - G^{[r]}_m|^2,$$

$$\text{with } G^{[r]}_m = \frac{1}{M} \sum_{m=1}^{M} G^{[r,m]} \tag{20}$$

and

$$\sigma^2_{G,r}(k) = \frac{1}{RM-1} \sum_{r=1}^{R} \sum_{m=1}^{M} |G^{[r,m]} - G_m|^2,$$

$$\text{with } G_m = \frac{1}{RM} \sum_{r=1}^{R} \sum_{m=1}^{M} G^{[r,m]}. \tag{21}$$

Since we consider systems for which a periodic input results in a periodic output with the same period, it is clear that $\sigma^2_{G,p}(k)$ depends only on the variations from one period to the other which are due to the disturbing noise n_y. $\sigma^2_{G,r}(k)$ is calculated over the different input realizations and so it depends on both noise sources. Comparing $\sigma^2_{Y,p}(k)$ and $\sigma^2_{Y,r}(k)$ gives immediately an idea about the disturbing noise and the nonlinear noise levels (see Schoukens et al., 2002a for the details, and D'Haene, Pintelon, Schoukens, & Van Gheem, 2004 for an extended version).

Example. In Fig. 7, the method is applied on a car body in white (only the metal frame of the car, without seats, shock absorbers, motor, wheels, etc.) with the following settings: $R = 8$, $M = 15$. The structure is excited with random multisines up to 400 Hz. The impact of the different noise sources is clearly visible. The reader should be aware that the nonlinearities might not only be due to a nonlinear behaviour of the car, also nonlinearities in the measurement setup will

be accounted for (Peeters, Van der Auweraer, Schoukens, & Pintelon, 2003).

4.4.2. Separation of the even and the odd stochastic nonlinearities

A second possibility to detect nonlinear distortions makes explicitly use of the flexibility of random multisines: only a selected set of harmonics (called measurement lines) is excited and the nonlinearities at the output are detected by measuring the output levels at the non excited frequencies (called detection lines). This idea was already suggested by Evans, Rees, and Jones (1994a) and McCormack, Godfrey, and Flower (1994) and is further elaborated in (Vanhoenacker & Schoukens, 2001; Schoukens, Pintelon, & Dobrowiecki, 2002b) where the choice of the excitation and detection frequencies is studied. Consider for example a random multisine that excites the system at the frequencies $k/N, k=1, 3, 9, 11, 17, 19, \ldots, 4p+1, 4p+3, \ldots, p \in \mathbb{N}$ (such a signal is called special odd). In that case the even nonlinearities are detected at the even lines in the output spectrum $y(2l), l=1, \ldots, N/2$, and the odd nonlinearities at the non excited odd lines $Y(l), l=5, 7, \ldots, 4p+5, 4p+7, \ldots, p \in \mathbb{N}$.

Discussion: In practice some additional problems can occur during this test.

- The nonlinear interaction between generator and plant can also generate unwanted excitation lines at the detection frequencies which should remain zero in the ideal situation. Under these conditions it is no longer clear what part of the output should be assigned to the linear behaviour, and what part is due to the nonlinear distortions. A first order correction can be made to reduce the problem (Vanhoenacker & Schoukens, 2003), but this method is less robust compared to the previous method (Section A) where such interaction is not disturbing at all the results.
- It turns out that the level of the nonlinear noise source measured at the non excited odd frequencies underestimates the level of the nonlinear noise source contributions at the measurement lines if periodic grids of excited and unexcited spectral lines are used (Dobrowiecki & Schoukens, 2001a; Vanhoenacker, Dobrowiecki, & Schoukens, 2001). If the detection lines are randomly chosen, the extrapolation factor is 1.
- Another price to be paid is the loss in frequency resolution caused by the non-excited lines which increases the required measurement time if a given resolution should be respected.

Example. This method is illustrated on the hair dryer using a special odd multisine (Németh & Vargha, 1999). In Fig. 8, the output amplitude spectrum is shown. Notice again that not only the level of the nonlinearities is detected, also the process noise levels are available from the periodic repetitions. In this example it is clear that the even nonlinearities

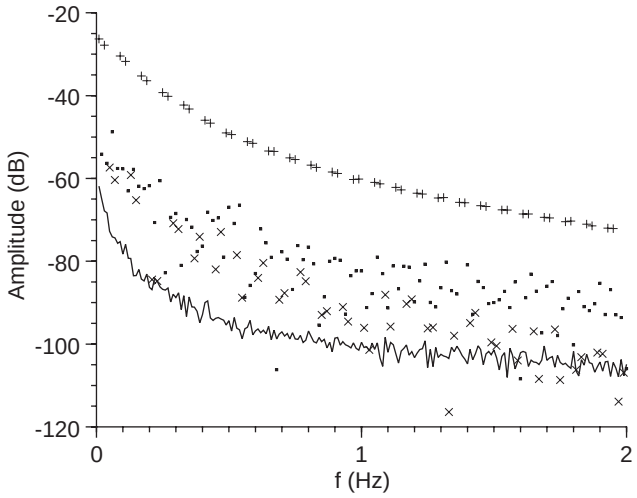

Fig. 8. Detection of nonlinear distortions: + output level at the excitation lines, . even nonlinearities, x odd nonlinearieties, thin line: noise level.

are the dominating distortion. For this reason odd excitations will reduce σ_{G_R} significantly, as was observed in Fig. 4.

Conclusion: This test provides more information than the previous one, but an experienced user is needed to perform it well. Moreover, the test is less robust with respect to the experimental setup and heuristic extrapolation factors are needed during the interpretation of the results.

4.4.3. Estimating the level of the nonlinear contributions to G_R

There is no direct access possible to the nonlinear contributions to G_R. These contributions cannot be separated from the rest of the signal as was done for the nonlinear noise source contributions Y_S. It is tempting to use the levels of Y_S also to bound the nonlinear contributions to G_R. There is yet no formal theoretical framework available to support this idea, although some insights are available (Dobrowiecki & Schoukens, 2001a). Using Y_S as an indication of the nonlinear contributions to G_R is an extrapolation. The bias is usually under estimated and the level depends upon the choice of the excitation and the system. Depending upon the situation extrapolation factors of ± 3 dB up to ± 20 dB are observed in the examples. The latter appear if only a small fraction of the total excitation power is in the pass band of the system.

The reader might be a bit disappointed about these 'loose' results, but she/he should realize that it is better to get at least a rough idea about the possible errors, than being even not aware about their existence.

4.5. Impact on the linear identification practice: parametric linear models

In this section we study the impact of nonlinear distortions on the parametric identification of linear models. Two approaches are considered. The first one is the classical pre-

diction error approach (Söderström & Stoica, 1989; Ljung, 1999) where a parametric plant and noise model are simultaneously estimated. In the second approach a non parametric noise model is identified during the preprocessing of the raw data. Next a parametric plant model is identified using the previous noise model as a weighting function (Schoukens, Pintelon, Vandersteen, & Guillaume, 1997; Pintelon & Schoukens, 2001; Pintelon & Schoukens, 2002b).

4.5.1. Combined identification of plant and noise model

In the 'classical' identification approach, a linear model $G_0(q,\theta)$ is estimated together with a noise model $H_0(q,\theta)$

$$y(t) = G_0(q,\theta)u_0(t) + H_0(q,\theta)e(t) + T_G(\theta). \tag{22}$$

$T_G(\theta)$ models the plant transients. Nowadays, $T_G(\theta)$ is usually estimated together with the plant model (estimate the initial conditions).

The model parameters θ are estimated by minimizing the prediction error, leading to the following definition for the estimates:

Definition 9. The prediction error estimates (estimated plant and noise model) are given by

$$\hat{\theta}_{\mathrm{PE}}(N) = \arg\min_{\theta} \frac{1}{N}\sum_{k=1}^{N} |H^{-1}(q,\theta)(y(t) - G(q,\theta)u(t) - T_G(\theta))|^2. \tag{23}$$

Because we deal here with identification in the presence of model errors, consistency should be replaced by convergence to the model (parameters) that would be obtained on the "exact" data G_R.

Definition 10. Best linear parametric approximation:

$$\hat{\theta}*(N) = \arg\min_{\theta} \frac{2}{N}\sum_{k=1}^{N/2} \frac{\left|G_R\left(\mathrm{j}2\pi\frac{k}{N}\right) - G(\mathrm{e}^{\mathrm{j}2\pi\frac{k}{N}}, \theta(N))\right|^2}{\sigma^2_{Y_S}(k)+\sigma^2_{Y_n}(k)}. \tag{24}$$

Because Theorem 1 guarantees that G_R and $\sigma^2_{Y_S}(k)$ are smooth functions, these can be approximated arbitrary well by a rational form if the order of the models is large enough. Hence it is always possible to balance the model errors and the noise errors using a classical model selection tool like the AIC or MDL criterion (Ljung, 1999; Akaike, 1974; Rissanen, 1978). For such a well selected model order and under the classical identifiability assumptions (Ljung, 1999; persistent excitation, existence of a unique minimum), the following result is obtained:

Theorem 3. *Consider a system belonging to the set S, excited with an excitation $u_N \in E$. If the noise Assumption* 1 *is met, $\hat{\theta}_{\mathrm{PE}}(N)$ converges in probability to $\hat{\theta}*(N)$:*

$$\operatorname*{plim}_{N\to\infty} (\hat{\theta}_{\mathrm{PE}}(N) - \hat{\theta}*(N)) = 0. \tag{25}$$

Proof. See Schoukens et al. (1998), Pintelon and Schoukens (2001).

From Theorem 3 it follows that the 'model errors' $|G_R(\mathrm{j}2\pi k/N) - G(\mathrm{e}^{\mathrm{j}2\pi k/N}, \hat{\theta}(N))|$ can be made arbitrarily small compared to the disturbing noise and the stochastic nonlinear contributions. Similar, the noise model $H(\mathrm{e}^{\mathrm{j}2\pi k/N}, \hat{\theta}(N))$ can follow arbitrarily well $\sigma^2_{Y_S}(k)+\sigma^2_{Y_n}(k)$. As a consequence this estimated plant/noise model will pass all 2nd order moment based validation tests. This is an unwanted and dangerous situation because the user gets no warning at all that a serious problem is hidden in the data. For example the uncertainty bounds that are calculated from this model are not valid. These bounds decrease to zero as an $O(N^{-1/2})$, while it is clear that this is not true for the nonlinear distortions and their induced errors (Pintelon & Schoukens, 2001).

The basic reason for this failure is that the noise model H is shaped to whiten the sum of the disturbing noise and the stochastic nonlinearities, while the variance λ of the driving white noise source is scaled to match the observed levels. This situation changes completely if the noise model is obtained from a prior analysis of the data as is discussed in the next section.

4.5.2. Separated identification of the noise models

If periodic excitations are used, it is known from Section 4.4 that the disturbing noise variances $\sigma^2_Y(k)$ can be obtained separately from the variance of the stochastic nonlinear disturbances even before the identification process starts. Using this non parametric noise model $\sigma^2_{Y_S}(k) + \sigma^2_{Y_n}(k)$ or $\sigma^2_{Y_n}(k)$ as weighting function, the following frequency domain identification scheme can be defined:

$$V_{\mathrm{SML}}(\theta,U,Y) = \frac{2}{N}\sum_{k=1}^{N/2} \frac{|Y(k) - G(\mathrm{j}2\pi k/N,\theta)U(k)|^2}{\hat{\sigma}^2_{Y_S}(k) + \hat{\sigma}^2_{Y_n}(k)}$$

or

$$V_{\mathrm{SML}}(\theta,U,Y) = \frac{2}{N}\sum_{k=1}^{N/2} \frac{|Y(k) - G(\mathrm{j}2\pi k/N,\theta)U(k)|^2}{\hat{\sigma}^2_{Y_n}(k)}. \tag{26}$$

In the first case, the estimate is weighted with respect to all distortions (if their variance is available), in the second case only the process noise is considered.

There is a full equivalence with the time domain identification framework (Schoukens et al., 1999). Remark that in (26) the exact noise variances $\sigma^2_{Y_S}(k)$, $\sigma^2_{Y_n}(k)$ are replaced by the estimated one $\hat{\sigma}^2_{Y_S}(k)$, $\hat{\sigma}^2_{Y_n}(k)$ obtained from measuring M successive periods of the input/output signals. The properties of this estimator, the sample maximum likelihood estimator (sample MLE) are known (Schoukens et al., 1997). The estimator remains consistent (convergence to the noiseless solution in case of model errors) if $M \geqslant 4$. However, a small loss in efficiency appears: the covariance matrix on the

Table 2
Recommendations for the model selection process for V_{SML} with a weighting $\sigma^2_{Y_n}(k)$

	White residuals → best linear approximation	Coloured residuals → still unmodelled dynamics (model errors)
The cost function is too large	Nonlinear distortions present It makes no sense to increase the model order	Increase the model order
The cost function is not significantly different from the expected value	This is the ideal situation Best linear model No model errors detectable	Increase the model order check the noise analysis
The cost function is too small	Check the noise analysis	Increase the model order Check the noise analysis

parameters increases with $((M-2)/(M-3))$ $(M \geqslant 6)$. The parameters are asymptotical normally distributed if $M \geqslant 7$ (Pintelon & Schoukens, 2002a).

Although both estimation procedures (prediction error, sample MLE) converge to the same limit model, the behaviour of the model validation process is completely different. In the latter case, the cost function can be absolutely interpreted since the noise model is fixed. Consider the situation where the weighting in (26) is $\hat{\sigma}^2_{Y_n}(k)$ (the non parametric noise model is extracted from a number of successive periods only, no averaging over different realizations of the excitation is made). In the absence of model errors, the expected value of the cost function is

$$\mathscr{E}\{V_{\mathrm{SML}}(\theta, U, Y)\} = \frac{M-1}{M-2}\left(1 - \frac{n_\theta}{N}\right), \tag{27}$$

with n_θ the number of free parameters. A cost function that is too large compared to (27) indicates errors that are not explained by the observed noise levels. If these errors are white, the best linear approximation is found. Correlated residuals point to unmodelled dynamics, hence it makes sense to increase the model order. This leads to the model selection/validation process given in Table 2. If the full weighting $\hat{\sigma}^2_{Y_S}(k) + \hat{\sigma}^2_{Y_n}(k)$ is used, the classical rules apply again (the nonlinearities are in that case detected from the procedure described in Section 4.4).

There exist a number of tools like the AIC and MDL criteria that are used to choose between different models (Akaike, 1974; Rissanen, 1978). These rules should be adapted to the situation of a fixed/estimated noise model. For a fixed noise model, the criterion should be reformulated to include the effect of model errors. For the standard prediction error identification methods this is done implicitly during the estimation of the noise variance λ starting from the residuals (Söderström & Stoica, 1989; Ljung, 1999). For the sample MLE, a modified criterion is needed. In the end the same criterion is found for both situations (Schoukens, Rolain, & Pintelon, 2002c):

$$V_N(\theta, Z)\left(1 + 2\,\frac{n_\theta}{N}\right), \tag{28}$$

with $V_N(\theta, Z)$ equal to $V_{PE}(\theta, Z)$ or $V_{\mathrm{SML}}(\theta, Z)$.

5. Conclusions

In this overview paper we proposed a framework to extend the linear system identification framework to nonlinear systems with a dominant linear behaviour. The focus is completely on the concepts and not on the mathematical details that can be found in the references. Nonlinear distortions are explicitly included into the framework and three major results are presented.

A first result is that a nonlinear system can be replaced by a linear system plus a nonlinear noise source. The properties of this representation are studied for randomized excitations. It turns out that the linear system is the best linear approximation, while the nonlinear noise source behaves as normally distributed disturbing noise.

In a second step, these insights are used to line out a best engineering practice to measure the FRF of the best linear approximation.

The last major result is a better understanding of the behaviour of the classic identification schemes, for example the prediction error framework and the frequency domain identification schemes. It turns out that the properties of these identification schemes are maintained, but the validation process is strongly affected by the presence of the nonlinear distortions. If a parametric noise model is identified simultaneously with the plant model, the presence of the nonlinear distortions can be completely missed using the classical whiteness and cross-correlation tests. The user gets no indication at all that something is going seriously wrong. If a non parametric noise model is extracted from a prior analysis, making use of periodic excitations, an alternative model selection and validation procedure is formulated that reveals the presence of nonlinear distortions.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Bendat, J. S. (1990). *Nonlinear system analysis & identification from random data*. New York: Wiley.

Bendat, J. S. (1998). *Nonlinear systems. Techniques and applications*. New York: Wiley.

Bendat, J. S., & Piersol, A. G. (1980). *Engineering applications of correlation and spectral analysis*. New York: Wiley.

Borys, A. (2000). *Nonlinear aspects of telecommunications. Discrete Volterra series and nonlinear echo cancellation*. Boca Raton: CRC Press.

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

Brillinger, D. R. (1981). *Time series. Data analysis and theory*. New York: McGraw-Hill.

Bussgang, J. J. (1952). *Crosscorrelation functions of amplitude-distorted Gaussian signals*. MIT Technical report. P.No 216.

D'Haene, T., Pintelon, R., Schoukens, J., & Van Gheem, E. (2004). Variance analysis of frequency response function measurements using periodic excitations. *Proceedings of 21th IEEE instrumentation and measurement technology conference* (pp. 1595–1600). Como, Italy.

Dobrowiecki, T., & Schoukens, J. (2001a). Bounds on Modelling Errors due to Weak Non-Linear Distortions. IMTC/01. *Proceedings of the 18th IEEE instrumentation and measurement technology conference*. Budapest, Hungary, May 21–24.

Dobrowiecki, T., & Schoukens, J. (2001b). Practical choices in the FRF measurement in presence of nonlinear distortions. *IEEE Transactions on Instrumentation and Measurements*, *50*, 2–7.

Doyle, F. J., Pearson, R. K., & Ogunnaike, B. A. (2001). *Identification and control using Volterra models*. Berlin: Springer.

Evans, C., & Rees, D. (2000). Nonlinear distortions and multisine signals—Part I: Measuring the best linear approximation. *IEEE Transactions on Instrumentation and Measurements*, *49*, 602–609.

Evans, C., Rees, D., & Jones, D. L. (1994a). Identifying linear models of systems suffering nonlinear distortions. *Proceedings IEE control'94* (pp. 288–296), Coventry.

Evans, C., Rees, D., & Jones, D. L. (1994b). Nonlinear disturbance errors in system-identification using multisine test signals. *IEEE Transactions on Instrumentation and Measurements*, *43*, 238–244.

Forsell, U., & Ljung, L. (2000). A projection method for closed-loop identification. *IEEE Transactions on Automatic Control*, *45*, 2101–2105.

Godfrey, K. (1993). *Perturbation signals for system identification*. Englewood Cliffs: Prentice-Hall.

Ljung, L. (1999). *System identification. Theory for the user*. Englewood Cliffs: Prentice-Hall.

McCormack, A. S., Godfrey, K. R., & Flower, J. O. (1994). The detection of and compensation for nonlinear effects using periodic input signals. *Proceedings IEE control '94* (pp. 297–302). Coventry, UK.

Németh, J., & Vargha, B. (1999). *Experiments with the hair dryer setup*. Internal report BIL99, Vrije Universiteit Brussel. Contact Johan.Schoukens@vub.ac.be.

Peeters, B., Van der Auweraer, H., Schoukens, J., & Pintelon, R. (2003). Using multisines to assess nonlinear distortions in vibrating mechanical structures. *20th Benelux-meeting on systems and control, 2003*.

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

Pintelon, R., & Schoukens, J. (2002a). Measurement and modelling of linear systems in the presence of non-linear distortions. *Mechanical Systems and Signal Processing*, *16*, 785–801.

Pintelon, R., & Schoukens, J. (2002b). Some peculiarities of identification in the presence of model errors. *Automatica*, *38*, 1683–1693.

Pintelon, R., Schoukens, J., & Vandersteen, G. (1997). Frequency domain system identification using arbitrary signals. *IEEE Transactions on Automatic Control*, *42*, 1717–1720.

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

Sandberg, I. W. (1992). Approximately-finite memory and input-output maps. *IEEE Transactions on Circuits and Systems-I*, *39*, 549–556.

Sandberg, I. W. (1993). Uniform approximation and the circle criterion. *IEEE Transactions on Automatic Control*, *38*, 1450–1458.

Sandberg, I. W. (2002). $\mathbb{R}_+$ Fading memory and extensions of input–output maps. *IEEE Transactions on Circuits and Systems-I*, *49*, 1586–1591.

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

Schoukens, J., Dobrowiecki, T., & Pintelon, R. (1998). Identification of linear systems in the presence of nonlinear distortions. A frequency domain approach. *IEEE Transactions on Automatic Control*, *43*(2), 176–190.

Schoukens, J., Swevers, J., Pintelon, R., & Van der Auweraer, H. (2002a). Excitation design for FRF measurements in the presence of nonlinear distortions. *Proceedings of ISMA2002* (Vol. 2, pp. 951–958).

Schoukens, J., Pintelon, R., & Dobrowiecki, T. (2002b). Linear modeling in the presence of nonlinear distortions. *IEEE Transactions on Instrumentation and Measurements*, *51*, 786–792.

Schoukens, J., Rolain, Y., & Pintelon, R. (2002c). Modified AIC rule for model selection in combination with prior estimated noise models. *Automatica*, *38*, 903–906.

Schoukens, J., Pintelon, R., & Rolain, Y. (1999). Study of conditional ML estimators in time and frequency-domain system identification. *Automatica*, *35*, 91–100.

Schoukens, J., Pintelon, R., Vandersteen, G., & Guillaume, P. (1997). Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, *33*, 1073–1086.

Söderström, T., & Stoica, P. (1989). *System identification*. (pp. 256) Englewood Cliffs: Prentice-Hall.

Vandersteen, G., Verbeeck, J., Rolain, Y., & Schoukens, J. (2000). Accurate bit-error-rate estimation for OFDM based telecommunication systems schemes in the presence of nonlinear distortions. *Proceedings of the 17th IEEE instrumentation and measurements conference IMTC* (pp. 80–85). Baltimore, MD.

Vanhoenacker, K., Dobrowiecki, T., & Schoukens, J. (2001). Design of multisine excitations to characterize the nonlinear distortions during FRF-measurements. *IEEE Transactions on Instrumentation and Measurements*, *50*, 1097–1102.

Vanhoenacker, K., Schoukens, J., Swevers, J. & Vaes, D. (2002). Summary and comparing overview of techniques of the detection of non-linear distortions. *Proceedings of ISMA2002* (Vol. 2, pp. 1241–1256).

Vanhoenacker, K., & Schoukens, J. (2003). Detection of nonlinear distortions with multisine excitations in the case of nonideal behaviour of the input signal. *IEEE Transactions on Instrumentation and Measurement*, *52*(3), 748–753.

Johan Schoukens received the degree of engineer in 1980, the degree of doctor in applied sciences in 1985, all from the Vrije Universiteit Brussel (VUB). He is presently a professor at the VUB. The prime factors of his research are in the field of system identification for linear and non-linear systems, and growing tomatos in his greenhouse.

Rik Pintelon received the degree of engineer in 1982, the degree of doctor in applied science in 1988, all from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. He is presently 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.

Tadeusz Dobrowiecki was born in Poland in 1952. He received the M.Sc. degree in electrical engineering in 1975 from the Technical University of Budapest, and the Ph.D. (candidate of sciences) degree in 1981 from the Hungarian Academy of Sciences, Budapest. Currently, he is an Associate Professor with the Budapest University of Technology&Economics. His research interests concern advanced signal processing algorithms and technical applications of knowledge-based techniques (measurement and system identification in particular).

Yves Rolain is with the Electrical Measurement Department (ELEC), Vrije Universiteit Brussel (VUB), Brussels, Belgium. His main research interests are nonlinear microwave measurement techniques, applied digital signal processing, parameter estimation/system identification, and biological agriculture.