Multidimensional Systems and Signal Processing manuscript No.
(will be inserted by the editor)

A method for informed selection of memory-length and nonlinearity-order parameters in Volterra-Wiener systems from exponential sweep excitations

J. L. Blanco-Murillo · V. Yagüe-Jiménez

Received: date / Accepted: date

Abstract We prove that through the analysis of the static, nonlinear system output to an exponential sweep excitation, it is possible to make an easy and informed estimation of memory-length and nonlinearity order parameters, prior to complete identification. The selection of these parameters has been a matter of discussion for many years since most identification techniques require early estimates of those. This work extends the results of previous contributions concerning a constrained set of memory- and order-fading nonlinear systems with diagonal kernels, onto the more general case of memory- and order-fading nonlinear systems that admit a Volterra series representation. A full mathematical formulation of the system output is provided, based on the input. Estimates on the parameters may be obtained by simple inspection on a single, filtered, output to the exponential sweep excitation. This happens to be a projection of the multidimensional, order kernels, as we prove. Examples are provided to evidence the simplicity and reliability of the proposed method.

Keywords Nonlinear systems identification · Volterra series representation · parameters selection · memory-length · nonlinearity order · exponential sweep

This work is partially supported by the Spanish Ministry of Economy and Comptitivity, Grant No. TEC2012-38402-C04-01.

J. L. Blanco-Murillo
Signal Processing Applications Group, Univ. Politecnica de Madrid
Tel.: +34-915-49-57-00 ext. 4044
Fax: +34-913-36-73-50
E-mail: jlblanco@gaps.ssr.upm.es

V. Yagüe-Jiménez
Institute of Physical and Information Technologies (ITEFI), Spanish National Research Council, E-mail: virginia.yague@csic.es

1 Introduction

Non-linear systems' are ubiquitously present among engineering applications, particularly in those involving electronics (e.g., Crespo-Cadenas et al (2014); Soumitro Banerjee (2001)) and optics (e.g., Agrawal (2007)). For this reason they receive great attention from the signal processing community. Still, the number of open questions related to them is overwhelming.

Authors have proposed multiple solutions for the evaluation and characterization of these systems through-out the years. From the Volterra formulation of the nonlinear convolution integral and Wiener's analysis Schetzen (1980); Rugh (1981); Parente (1970); Casey and Walnut (1994); van Hemmen et al (2000); Orcioni et al (2005), to the use of pseudorandom sequences for non-parametric identification Bose and Mitra (1992); Nowak and Van Veen (1994); improvements have continuously been reported. All of these focus on the ability of Volterra and Wiener series to represent nonlinear systems matching definitions in Boyd and Chua (1985). To these we will refer as Volterra-Wiener systems.

Nowadays, reliable models may be computed that remarkably resemble the system under test. However, regardless of the selected technique Schetzen (1980); Lin (2006); Jing (2012), identification still requires early estimates of at least two parameters: the nonlinearity (N) and the duration of the system memory (Θ). In Schetzen (1980) the author already discussed the connection between these two and the duration of test sequences needed for identification. Considering a discretized system (i.e., $M = \lceil \Theta / F_s \rceil$ with sampling frequency F_s), the total number of samples required for the shortest possible input sequence is given by $\sum_{n=1}^{N} \binom{M+n}{n}$. When addressing nonparametric identification this number varies; but anyhow, in all cases this must be increased. To avoid both time-alliasing, caused by underestimating M, and inter-kernel aliasing introducing artificial intermodulation products, due to an underestimated N. Restimation of N and M has been addressed, but still demands the reiterated complete identification of the system starting from an initial guess, Billings (1980).

In this work we focus on nonlinear systems amenable to the multidimensional, Volterra representation Sharma (2010), covering most static, physical processes Sandberg (1990). A number of different structures match this definition, including those combining memory and continuous nonlinearity blocks —e.g., Hammerstein and Wiener systems or sandwich structures. Readers may refer to Cheng et al (2017) for an exhaustive review on these, including references to applications in electrical and mechanical engineering. Specifically on Hammerstein systems with a weak nonlinearity, in Farina (2000) the author reported that outputs to an exponential sweep excitation are highly informative of the system structure itself, and a simple identification procedure was proposed. Authors in Novák et al (2010) and Rébillat et al (2011) extended this work and suggested that the entire identification procedure could be covered by such analysis on an extended definition of Hammerstein systems.

Here we extend previous contributions on these (generalized) Hammerstein systems, and prove that for any system admitting a multidimensional, Volterra series expansion, one may obtain rough, but reliable, estimates on the nonlinearity order and system memory length, based on its output to an exponential sweep excitation. Through the analysis of these signals, one may identify individual, order-dependent,

frequency-bounded packets of information in the form of chirplets. These chirplets may only be revealed through the instantaneous analysis of input and output, provide reliable insights on the system under test and are being partially parameterized by the input excitation.

To extract system-specific pieces of information we first need to evaluate the output of a Volterra-Wiener system to an exponential sweep. In Sect. 2 we derive the instantaneous form of the system output under such excitation, and include time and frequency analyses. From these we define a series of input- and system-dependent parameters that support the proposed estimation procedures. System output is post-processed in Sect. 3 attending to the input excitation and focusing on instantaneus time and frequency content. We derive closed forms for all signals involved, including the post-processed output. By reformulating the Volterra integral in the latter expression, in Sect. 4 we proof the existence and reveal the nature of the aforementioned, time-shifted chirplets Mann and Haykin (1995), which in essence are similar to functionals in the work by Schetzen (1985). We prove that the number of these relates to the order of the nonlinearity, and their duration to the memory length. In Sect. 5 we comment on the input excitation parameters' selection and the estimation algorithms. One must adequately choose these for the procedure to deliver meaningful results, and be able to conveniently rearrange them when failing. Tests running on the proposed procedure are included in Sect. 6 to assess performance, by addressing different scenarios. Finally, we describe our results in Sect. 7 and discuss them in Sect. 8 including our final remarks on memory- and order-fading systems.

2 System output analysis for exponential sweep excitations

We start with an exponential sweep of duration T, covering the frequency range $f_0 \leq f \leq f_1$, which can be generated using a signal generator or a general purpose DAC. This is fully mathematically described by the following equations:

$$x(t) = \cos(\phi(t)), \quad \phi(t) = \frac{2\pi f_0}{\ln(k)}(k^{t-t_0} - 1) + \phi_0 \tag{1}$$

where t_0 and ϕ_0 are the initial values for time and phase, respectively; and k is the rate of exponential increase, $k = (f_1/f_0)^{1/T}$, instantaneously producing a single mono-component at frequency $f(t) = f_0 \cdot k^{t-t_0}$.

The output that is to be expected from a causal system amenable to the Volterra model to some excitation $x(t)$ attending to its definition may be written as

$$y(t) = h_0 + \sum_{n=1}^{N} \int_0^{+\infty} h_n(\theta_1, \ldots, \theta_n) \prod_{i=1}^{n} x(t - \theta_i)\, \mathrm{d}\bar{\theta} \tag{2}$$

where $h_n(\theta_1, \ldots, \theta_n) = h_n(\bar{\theta})$ is the n-th Volterra kernel defined as n-dimensional tensor, and $\{\theta_i\}$ are all time dimensions considered, which we group into a single vector $\bar{\theta} = (\theta_1, \ldots, \theta_n)$ for easy of notation.

In deriving an specific form for $y(t)$ when $x(t)$ is an exponential sweep, let us first focus on the product of time-shifted versions of the input, to the right of the integral definition. Based on Abramowitz and Stegun (1964), for an exponential sweep excitation this may be rewritten as follows

$$\prod_{i=1}^{n} x(t-\theta_i) = \tfrac{1}{2^{n-1}} \sum_{\bar{\varepsilon}\in\mathscr{S}_n} \cos\left(\phi_{\bar{\varepsilon}}(t,\bar{\theta})\right) \tag{3}$$

Here $\bar{\varepsilon} \in \mathscr{S}_n$ is a vector formed from all possible binary multicombinations of length n, the elements of which are: $\varepsilon_1 = 1$, $\varepsilon_i = \pm 1$, $1 < i \leq n$; and phase function $\phi_{\bar{\varepsilon}}(t,\bar{\theta})$ are given by

$$\phi_{\bar{\varepsilon}}(t,\bar{\theta}) = \frac{2\pi f_0}{\ln(k)} \left(\zeta_{\bar{\varepsilon}} \cdot k^{t+\tau_{\bar{\varepsilon}}} - 1\right) + \phi_{0,\bar{\varepsilon}} \tag{4}$$

including an initial phase term of the form

$$\phi_{0,\bar{\varepsilon}} = \left[\phi_0 - \tfrac{2\pi f_0}{\ln(k)}\right] \sum_{i=2}^{n} \varepsilon_i + \phi_0 \tag{5}$$

One may rewrite phase functions in (4) by means of the exponential sweep *fundamental phase property* derived in Rébillat et al (2011)

$$\phi_{\bar{\varepsilon}}(t,\bar{\theta}) = \zeta_{\bar{\varepsilon}} \cdot \phi(t+\hat{\tau}_{\bar{\varepsilon}}) + \hat{\phi}_{0,\bar{\varepsilon}} \tag{6}$$

where we introduce (i) time-shifts

$$\tau_{\bar{\varepsilon}} = \hat{\tau}_{\bar{\varepsilon}} - t_0 = \log_k \left| \sum_{i=1}^{n} \varepsilon_i k^{-\theta_i} \right| - t_0 \tag{7}$$

which may be written attending to

$$\begin{aligned} \mathscr{N}_{\bar{\varepsilon}}(\bar{\theta}) &= \textstyle\sum_{i=1}^{n} \varepsilon_i k^{-\theta_i} \\ \tau_{\bar{\varepsilon}} &= \log_k \left| \mathscr{N}_{\bar{\varepsilon}}(\bar{\theta}) \right| - t_0 \end{aligned} \tag{8}$$

and (ii) the initial, constant phase

$$\hat{\phi}_{0,\bar{\varepsilon}} = \left[\phi_0 - \frac{2\pi f_0}{\ln(k)}\right] \cdot \left(\sum_{i=1}^{n} \varepsilon_i - \zeta_{\bar{\varepsilon}}\right) \tag{9}$$

with scalar $\zeta_{\bar{\varepsilon}}(\mathscr{N}) = 1$ if $\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta}) \geq 0$ and -1 otherwise.

Hence, applying this into (2), it appears that the system output is itself a linear combination of multiple, phase-transformed, exponential sweeps. The argument of the cosine function has been rewritten here attending to the latter's even symmetry property and (6)

$$y(t) = h_0 + \sum_{n=1}^{N} \int_{0}^{+\infty} \frac{h_n(\bar{\theta})}{2^{n-1}} \sum_{\bar{\varepsilon}\in\mathscr{S}_n} \cos\left(\phi(t+\hat{\tau}_{\bar{\varepsilon}}) + \zeta_{\bar{\varepsilon}} \cdot \hat{\phi}_{0,\bar{\varepsilon}}\right) \mathrm{d}\bar{\theta} \tag{10}$$

This daunting formulation simply addresses the general form of the system output, which otherwise would naturally be produced by recording the system's output to some exponential sweep excitation. However, it is simply too complex to extract information directly from it. To do so one must analyse the instantaneous information embedded in model parameters: $\bar{\varepsilon}$, $\zeta_{\bar{\varepsilon}}$, $\hat{\phi}_{0,\bar{\varepsilon}}$ and $\hat{\tau}_{\bar{\varepsilon}}$; and elaborate it to gain insights into energy distribution on the system output, both along time and frequency domains.

2.1 Interpretation of model parameters: vector $\bar{\varepsilon}$

In (3) we introduced vector $\bar{\varepsilon} \in \mathscr{S}_n$ with elements $\varepsilon_1 = 1$, $\varepsilon_i = \pm 1$ for $1 < i \leq n$, to reformulate the product of time-shifted inputs in the Volterra definition (2), resulting in a summation of (i) phase- ($\zeta_{\bar{\varepsilon}} \cdot \hat{\phi}_{0,\bar{\varepsilon}}$) as well as (ii) time- ($\hat{\tau}_{\bar{\varepsilon}}$) shifted tones. Due to these and sweeps' intrinsic time-frequency dependence, these two automatically result in (iii) $\bar{\varepsilon}$-dependent frequency variations.

$\bar{\varepsilon}$-dependent monocomponents appearing in (10) play a similar role as sinusoidal monocomponents would do for a linear system. However, in this case each order-dependent impulse response —i.e., the Volterra kernels— impacts on different components. Vectors $\bar{\varepsilon}$ characterize each multi-dimensional, cross-product delivered by the system nonlinearity. And by covering set $\mathscr{S}_n$ we ensure that all multicombinations up to nonlinearity order n are addressed. Properties of $\bar{\varepsilon} \in \mathscr{S}_n$ directly impact on the forespoken monocomponents and must be worked out to later discuss their individual phase, time and frequency contributions.

First, the cardinality of $\mathscr{S}_n$ is 2^{n-1}; but set $\mathscr{S}_n$ may be partitioned attending to vectors' sums. Let us define $m = \sum_{i=1}^{n} \varepsilon_i$ with the following properties

1. there are only n possible values for m, which are $\{n, n-2, n-4, \ldots, 2-n\}$;
2. m displays the same parity as n.

From this we define subsets

$$\mathscr{S}_{n\langle m\rangle} = \left\{ \bar{\varepsilon} \in \mathscr{S}_n \middle| \sum_{i=1}^{n} \varepsilon_i = m \right\} \tag{11}$$

Trivially, one may prove that for some given n:

1. $\mathscr{S}_{n\langle n\rangle}$ contains only one vector that matches an all-ones sequence;
2. sets $\{\mathscr{S}_{n\langle m\rangle}\}$ form a complete partition of $\mathscr{S}_n$:

$$\cup_{m=1}^{n} \mathscr{S}_{n\langle m\rangle} = \mathscr{S}_n \quad \text{and} \quad \mathscr{S}_{n\langle m\rangle} \cap \mathscr{S}_{n\langle r\rangle} = \emptyset, \quad \text{with} \quad r \neq n$$

3. the cardinality of the k-th subset, associated to the k-th value of m, is $\binom{n-1}{k-1}$;
4. for some $n' > n$, same parity as n, cardinality of $\mathscr{S}_{n\langle m\rangle}$ is lesser than that of $\mathscr{S}_{n'\langle m\rangle}$.

Second, for any $\bar{\varepsilon} \in \mathscr{S}_{n\langle m\rangle}$, as long as $k > 1$ (this is, $f_1 > f_0$), the following inequalities hold for $\bar{\theta} \in \mathbb{R}^n$

$$\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta}) = \sum_{i=1}^{n} \varepsilon_i k^{-\theta_i} \leq m \tag{12}$$

$$\left(1-\sum_{i=2}^{n} k^{\theta_1-\theta_i}\right) \leq \frac{\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})}{k^{-\theta_1}} \leq \left(1+\sum_{i=2}^{n} k^{\theta_1-\theta_i}\right) \tag{13}$$

Here we assume, without loss of generality, that time dimension $\theta_1 \in [0,\Theta]$ displays the largest memory, and take it as a reference. Hereafter, the memory-length of the system, Θ, is assimilated to the longest duration of a Volterra kernel $h_n(\bar{\theta})$ along a single time dimension, θ_1. Thus, we may restrict our analysis to ranges $0 \leq \theta_i \leq \Theta$ for all $1 \leq i \leq n$, assuming that the Volterra kernel outside this compact ball has no effective response. Further discussion on Θ is included in Sect. 8, accompanying the proposed system memory estimation.

Attending to (13), rather than considering each individual time dimension, we may evaluate differences between these and the reference one. In doing so, we define $\bar{\Delta} = (0,\Delta_2,\Delta_3,\ldots,\Delta_n)$ with $\Delta_i = \theta_1 - \theta_i$, and introduce

$$\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta}) = \sum_{i=2}^{n} \varepsilon_i k^{\Delta_i} \tag{14}$$

Note that $(\theta_1,\bar{\Delta})$ formulation has direct implications on the integration of the Volterra kernels in (10). For instance, integration along vector $\bar{\Delta} = 0_n$ (all-zeros) represents integration along the n-th Volterra kernel's diagonal. All other dimensions for integration in (10) may also be encoded using $(\theta_1,\bar{\Delta})$, satisfying constraint $|\Delta_i| < \Theta$. From these we derive the following inequalities regarding $\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta})$.

$$\left(m_+ k^{-\Theta} - m_- k^{\Theta}\right) \leq \mathscr{R}_{\bar{\varepsilon}}\left(\bar{\Delta}\right) \leq \left(m_+ k^{\Theta} - m_- k^{-\Theta}\right) \tag{15}$$

where $m_\pm$ are respectively the number of ± 1 in vector $\bar{\varepsilon}$, so $m = m_+ - m_-$ and $n = m_+ + m_-$. Hence,

$$\left(m\mathscr{R}_+ + n\mathscr{R}_-\right)/2 \leq \mathscr{R}_{\bar{\varepsilon}}\left(\bar{\Delta}\right) \leq \left(m\mathscr{R}_+ - n\mathscr{R}_-\right)/2 \qquad \mathscr{R}_\pm = k^{-\Theta} \pm k^{+\Theta} \tag{16}$$

When computed on different vectors we observe relevant trends. For $\bar{\varepsilon}_1,\bar{\varepsilon}_2 \in \mathscr{S}_n$, one of the following conditions must hold:

1. if $\bar{\varepsilon}_1 = \bar{\varepsilon}_2$, then:

$$\mathscr{R}_{\bar{\varepsilon}_1} - \mathscr{R}_{\bar{\varepsilon}_2} = 0 \tag{17a}$$

2. if $\bar{\varepsilon}_1 \neq \bar{\varepsilon}_2$ but $\bar{\varepsilon}_1,\bar{\varepsilon}_2 \in \mathscr{S}_{n\langle m\rangle}$, then:

$$|\mathscr{R}_{\bar{\varepsilon}_1} - \mathscr{R}_{\bar{\varepsilon}_2}| \leq n\mathscr{R}_- \tag{17b}$$

3. if $\bar{\varepsilon}_1 \in \mathscr{S}_{n\langle m\rangle}$, $\bar{\varepsilon}_2 \in \mathscr{S}_{n\langle m-p\rangle}$ with $p > 0$, then:

$$[\mathscr{R}_{\bar{\varepsilon}_1} - \mathscr{R}_{\bar{\varepsilon}_2}] \geq p\mathscr{R}_+/2 - n\mathscr{R}_- \tag{17c}$$

Thus, attending to (17b) values of $\mathscr{R}_{\bar{\varepsilon}}$ for vectors with the same m cluster unevenly around some value. And based on (17c), for a given order n and any two consecutive subsets ($p = 2$), there may exist a gap in the values of $\mathscr{R}_{\bar{\varepsilon}}$, which separates the two clusters. Furthermore, when comparing two consecutive orders, n and $n+1$, values for m may differ by $p = 1$; and gaps among clusters may still exist.

These results are better explained by looking at the probability distribution of the values for $\mathscr{R}_{\bar{\varepsilon}}$. Figure 1a displays histograms corresponding to these for nonlinearity orders 2 to 6, some simulated, Volterra-Wiener system and specific values of the input sweep parameters. Vectors $\bar{\varepsilon}$ are deterministic and $\{\Delta_i\}$ assumed to be uniformly and independently distributed along time dimensions $\{\theta_i\}$, producing in Figure 1a the expected clusters with their respective centres of gravity for $\mathscr{R}_{\bar{\varepsilon}}$ clusters locate at

$$\bar{\mathscr{R}}(m) = (m-1)\cdot\bar{r} \tag{18}$$

with $\bar{r} = (\mathscr{R}_+ - 2)/(\Theta\ln(k))^2$. This does not directly depend on order, n, but on memory length Θ and the excitation parameters through k. Differently, their variances (clusters' deviations), given by

$$\mathrm{Var}\left[\mathscr{R}_{\bar{\varepsilon}}\right] = \bar{r}^2\cdot\left[(n-1) + 2\sum_{i=2}^{n-1}\sum_{j=i+1}^{n}\varepsilon_i\varepsilon_j\right] \tag{19}$$

are order- and $\bar{\varepsilon}$-dependent, as one may notice on Figure 1a. This reinforces the same idea as (17b) that gaps in $\mathscr{R}_{\bar{\varepsilon}}$ (low probability regions on the histograms) are order-dependent. Probability distributions in Figure 1a become gaussian-like as order increases, as suggested by the central-limit theorem; while spreading in the form established by (19).

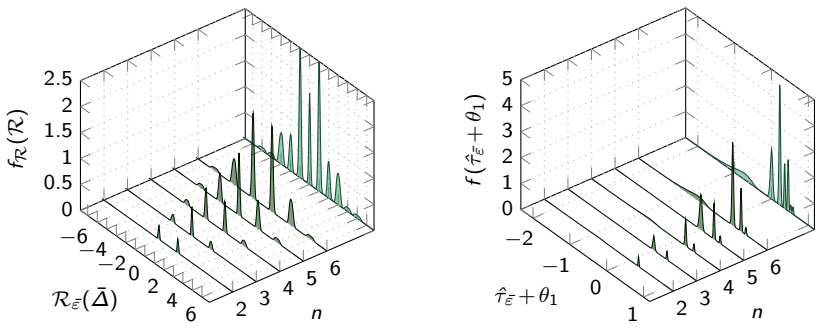
Clusters are slightly asymmetrical towards the larger values of $|\Delta_i|$ and the relative maxima on the histogram are proportional to the cardinality of $\mathscr{S}_{n\langle m\rangle}$. Gaps appear along every histogram depicted in Figure 1a, for the selected values of the system and excitation parameters. However, while comparing consecutive orders through the aggregated distribution on the back plane, it appears that for $n \geq 6$ and $|m| \geq 5$ values of $\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta})$ do overlap —see the very close, outer humps. In this case clusters, are just slightly overlapped. One could possibly adapt the excitation parameters to avoid this. For instance by increasing T, rate k will change, at the expense of larger excitations and realizing that for some larger m this exact situation may reproduce.

Bearing this in mind, we now address the monocomponents in (10). In particular, considering (i) phase shifts, as well as (ii) time (iii) instantaneous frequency content.

2.2 Phase shifts: $\hat{\phi}_{0,\bar{\varepsilon}}$

Considering our previous discussion on $\bar{\varepsilon}$ and the expression for phases $\hat{\phi}_{0,\bar{\varepsilon}}$ in (9), the interpretation of the latter is straightforward: these are order-dependent, initial phase values depending on $\bar{\varepsilon}$. One may reformulate these in terms of m, noticing that only two different values are allowed for a given m; and considering that the sign of $\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})$ is the same as the sign of $1+\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta})$.

$$\zeta_{\bar{\varepsilon}}\cdot\hat{\phi}_{0,\bar{\varepsilon}} = \left[\phi_0 - \frac{2\pi f_0}{\ln(k)}\right]\cdot(\zeta_{\bar{\varepsilon}}m - 1) \tag{20}$$



(a) $\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta})$ (b) $\hat{\tau}_{\bar{\varepsilon}}+\theta_1$

Fig. 1 Histograms for different nonlinearity orders. The aggregated histogram is displayed on the back plane. The excitation parameters used were: $f_0 = 200\,\text{Hz}$, $f_1 = 20\,\text{kHz}$ and $T = 1500\,\text{ms}$; and the system parameters: $\Theta = 50\,\text{ms}$, covering nonlinearity orders n 2 to 6. Histograms were computed assuming i.i.d. dimensions $\{\Delta_i\}$ resulting from uniformly distributed $\{\theta_i\}$. In (a), histograms on $\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta})$, all are comb-like functions including n peaks that actually group values for vectors from the same subset $\mathscr{S}_{n\langle m\rangle}$. In (b), histograms on $\hat{\tau}_{\bar{\varepsilon}}+\theta_1$, we get $\lfloor n/2\rfloor+1$, unevenly-distributed peaks, grouping vectors from subsets $\mathscr{S}_{n\langle m\rangle}$ and $\mathscr{S}_{n\langle -m\rangle}$. Also notice the wide, low-probability hump towards negative values of $\hat{\tau}_{\bar{\varepsilon}}+\theta_1$.

with $\zeta_{\bar{\varepsilon}} = 1$ if $\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta}) \geq 1$ and -1 otherwise. Only when ϕ_0 equals $2\pi f_0/\ln(k)$ order dependence disappears and we have no phase shift.

2.3 Time analysis: $\hat{\tau}_{\bar{\varepsilon}}(\bar{\theta})$ shifts

Identified times shifts along the individual monocomponents in (10) also depend on $\bar{\varepsilon}$. One may reformulate these based on (7) and (14) as

$$\hat{\tau}_{\bar{\varepsilon}}(\bar{\theta}) = \log_k |1+\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta})| - \theta_1 \tag{21}$$

According to (17b)-(17c) and histograms in Figure 1a, we know that the values for $\mathscr{R}_{\bar{\varepsilon}}(\bar{\Delta})$ tend to cluster. This result reappears on the distribution of $\hat{\tau}_{\bar{\varepsilon}}(\bar{\theta})$, which is better understood considering $\hat{\tau}_{\bar{\varepsilon}}(\bar{\theta})+\theta_1$. Histograms on these are depicted in Figure 1b. Peaks are still present as in Figure 1a, but in a different number and distribution. Their locations flow logarithmically, and their number is reduced due to the absolute-value function down to $\lfloor n/2\rfloor+1$. In other words, we get $\lfloor n/2\rfloor+1$ groups, which combine clusters formed from vectors $\bar{\varepsilon}$ in subsets $\mathscr{S}_{n\langle m\rangle}$ and $\mathscr{S}_{n\langle -m\rangle}$.

Gaps between consecutive clusters may still exist, depending of the actual values of the system and excitation parameters. Attending to (15) and the monotonous, logarithmic transformation in (21), for $\bar{\varepsilon}_1 \in \mathscr{S}_{n\langle m\rangle}$, $\bar{\varepsilon}_2 \in \mathscr{S}_{n\langle m-p\rangle}$ with $p > 0$, distance between consecutive clusters is

$$D_{n,m}(p) = \log_k \left| \frac{m\mathscr{R}_+ + n\mathscr{R}_-}{(m-p)\mathscr{R}_+ - n\mathscr{R}_-} \right| \tag{22}$$

Thus, one may arrange the excitations' parameters to produce dead regions (gaps) among the values of $\hat{\tau}_{\bar{\varepsilon}}+\theta_1$, with some specific width (D) and for some order (m).

Parameter k could potentially take any positive value by adequately selecting the excitation parameters: f_0, f_1 and T. In particular, assuming that we are interested in a limited frequency band, by monotonically increasing T, $D_{n,m}(p)$ increases and dead zones become wider.

Additionally, attending to (22), one may proof that

$$\forall p \;\; D_{n.m}(p) \geq D_{n',m}(1)\big|_{n'>n} \tag{23}$$

implying that consecutive ($p = 1$) dead zones' widths monotonically decrease, and their minimum is given by the separations associated to the highest order, ultimately the nonlinearity order it-self. By increasing T one may then produce the desired gaps up to any nonlinearity order, and observe the m-dependent clusters.

Finally, the fact that $\hat{\tau}_{\bar{\varepsilon}}$ incorporates additively the reference time dimension implies that while computing the θ_1 time-integral in (10), some form of linear phase shift is produced. This very particular effect is specific of exponential sweeps and is only revealed while addressing the instantaneous time response of underlying Volterra-Wiener systems.

Time shifts are particularly easy to observe on purely diagonal Volterra kernels, by following the analysis in Rébillat et al (2011). This is, when $\bar{\Delta} = \bar{0}_n$, causing $\mathrm{Var}[\mathscr{R}_{\bar{\varepsilon}}] = 0$. As an example, considering the simplest, n-order, memory-less system of this kind, $y(t) = x^n(t)$, then

$$\begin{aligned} y(t) &= \cos^n\left(\tfrac{2\pi f_0}{\ln(k)}(k^{t-t_0}-1)+\phi_0\right) \\ &= \sum_{m=1}^{n} c_{n,m}\cdot\cos\left(m\cdot\left(\tfrac{2\pi f_0}{\ln(k)}(k^{t-t_0}-1)+\phi_0\right)\right) \end{aligned} \tag{24}$$

where $c_{n,m} = \frac{1}{2^{n-1}}\sum_{m=1}^{n}\binom{n-1}{m-1}$, which coincide with the Chebyshev polinomials' coefficients as derived in Rébillat et al (2011). From this

$$\begin{aligned} m\cdot\phi(t) &= \tfrac{2\pi f_0}{\ln(k)}m(k^{-t_0}-1)+m\phi_0 \\ &= \tfrac{2\pi f_0}{\ln(k)}(k^{t-t_0+\log_k m}-1)+\phi_{0,m}, \qquad t_0 \leq t \leq (t_o+T) \end{aligned} \tag{25}$$

with $\hat{\tau}_m = \log_k m$ and $\phi_{0,m} = m\phi_0 - (m-1)$. Here, through the instantaneous analysis of the signals, one may realize that systematic deviations are to be found on the phase of the individual contributors to the system output. These we note throughout this contribution in the convenient form of time shifts $\hat{\tau}_{\bar{\varepsilon}}$.

2.4 Instantaneous frequency analysis

Focusing on the standard definition of the instantaneous frequency for sinusoids, one may evaluate from (10) the excited frequencies on the individual contributors. These equate the normalized first-time derivative of the phase, as opposed to Fouriers' frequency definition. Starting from $\phi(t)$, we already obtained

$$f(t) = \frac{1}{2\pi}\left|\frac{\partial\phi_{\bar{\varepsilon}}}{\partial t}(t)\right| = f_0\cdot k^{t-t_0} \qquad t_0 \leq t \leq t_0+T \tag{26}$$

and for $\phi_{\bar{\varepsilon}}(t,\bar{\theta})$

$$f_{\bar{\varepsilon}}(t) = \frac{1}{2\pi}\left|\frac{\partial\,\phi_{\bar{\varepsilon}}}{\partial\,t}(t,\bar{\theta})\right| = f_0 \cdot k^{t-t_0+\hat{\tau}_{\bar{\varepsilon}}} \qquad t_0 \le t \le t_0+T \tag{27}$$

We verify that the individual monocomponents are exponential sweeps (i.e., narrow-band signals at any specific time stamp), only affected by time shifts $\hat{\tau}_{\bar{\varepsilon}}(\bar{\theta})$. Additionally, baring in mind (7) one realizes that this is merely equal to

$$f_{\bar{\varepsilon}}(t) = \left|\mathscr{N}_{\bar{\varepsilon}}\right| \cdot f_0 \cdot k^{t-t_0} \qquad t_0 \le t \le t_0+T \tag{28}$$

matching the exact frequency ranges derived in Lang and Billings (1996) for Volterra-Wiener systems' responses.

Results derived in Section 2 evidence the connection between the outcoming signal ($y(t)$) and both the exciting sweep ($x(t)$) and the system under test (SUT). Hereafter, we like to disentangle these expressions to focus just on the system; and only to derive informed estimators for the memory-length and nonlinearity order parameters.

3 Output postprocessing

The expected system output to an exponential sweep given in (10) is ultimately an integral form in N-dimensional time, $\bar{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)$, involving functions $\hat{\tau}_{\bar{\varepsilon}}(\bar{\theta})$ and $\zeta_{\bar{\varepsilon}}\hat{\phi}_{0,\bar{\varepsilon}}$. Through the post-processing of the system output one could possibly extract relevant information on the SUT. We propose the following procedure based on instantaneous, time and frequency analysis, and involving an adapted filter. The latter is derived hereafter.

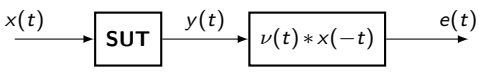


Fig. 2 Block diagram for the evaluation procedure: $x(t)$ is the sweep used to excite the system under test (SUT), and $\nu(t)$ is the impulse response used to decolorize the system output.

Figure 2 describes the structure of our test. The system output, $y(t)$, is here post-processed based on a linear filter that conveys the effect of the time-reversed input, $x(-t)$, and an inverse filter, $\nu(t)$. The latter only depends on the system input and its role is to enhance energy compaction along time and whiten frequency response. For the rest of this section, we analyze the two of these hereafter, to derive the instantaneous form of $e(t)$. From this we will extract the desired system parameters and their estimators. Readers may direct to Section 5 for the final form of the postprocessed output and the proposed estimation methods.

Table 1 Instantaneous frequencies estimated for the individual phase contributors in $\alpha_{\bar{\varepsilon}-}(t,\bar{\theta})$.

Time interval	Phase functions	Inst. freq. ($f_{i,\bar{\varepsilon}}$)	$f_{i,\bar{\varepsilon}}(t=-T)$	$f_{i,\bar{\varepsilon}}(t=-\hat{\tau}_{\bar{\varepsilon}})$	$f_{i,\bar{\varepsilon}}(t=0)$	$f_{i,\bar{\varepsilon}}(t=+T)$
$-T \le t \le 0$	$\phi_{1,\bar{\varepsilon}} = \alpha_{\bar{\varepsilon}-}(t)\cdot k^{T+t}$	$f_{1,\bar{\varepsilon}} = f_1 \cdot k^{t+\hat{\tau}_{\bar{\varepsilon}}}$	$\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})\cdot f_0$	f_1	$\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})\cdot f_1$	—
	$\phi_{2,\bar{\varepsilon}} = \alpha_{\bar{\varepsilon}-}(t)$	$f_{2,\bar{\varepsilon}} = f_0 \cdot k^{-t}$	f_1	$\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})\cdot f_0$	f_0	—
$0 \le t \le +T$	$\phi_{1,\bar{\varepsilon}} = \alpha_{\bar{\varepsilon}-}(t)\cdot k^{T}$	$f_{1,\bar{\varepsilon}} = f_1 \cdot k^{-t}$	—	—	f_1	f_0
	$\phi_{2,\bar{\varepsilon}} = \alpha_{\bar{\varepsilon}-}(t)\cdot k^{t}$	$f_{2,\bar{\varepsilon}} = f_0 \cdot k^{t+\hat{\tau}_{\bar{\varepsilon}}}$	—	—	$\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})\cdot f_0$	$\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})\cdot f_1$

3.1 Post-processed output formulation

Focusing on the effect of the time-reversed sweep, we evaluate

$$\tilde{e}(t) = y(t) * x(-t) = h_0 + \sum_{n=1}^{N} \int_0^{+\infty} \frac{h_n\left(\bar{\theta}\right)}{2^{n-1}} \sum_{\bar{\varepsilon}\in\mathscr{S}_n} u_{\bar{\varepsilon}}(t,\bar{\theta})\,\mathrm{d}\bar{\theta} \tag{29}$$

where $u_{\bar{\varepsilon}}(t,\bar{\theta})$ are windowed functions given by:

$$\begin{aligned} u_{\bar{\varepsilon}}(t,\bar{\theta}) &= w(t)\cdot \overbrace{\cos\left(\phi(t+\hat{\tau}_{\bar{\varepsilon}}) + \zeta_{\bar{\varepsilon}}\cdot\hat{\phi}_{0,\bar{\varepsilon}}\right)}^{\text{output contributor}} * \overbrace{\cos\left(\phi(-t)\right)}^{\text{time-reversed input}} = \\ &= \int_{-\infty}^{+\infty} w(\tau-t)\cdot\cos(\phi(\tau-t))\cdot\prod_{i=1}^{n} w(\tau-\theta_i)\cdot \\ &\quad\cdot\left[c_{\bar{\varepsilon}}\cos\left(\phi\left(\tau+\hat{\tau}_{\bar{\varepsilon}}\right)\right) - s_{\bar{\varepsilon}}\sin\left(\phi\left(\tau+\hat{\tau}_{\bar{\varepsilon}}\right)\right)\right]\mathrm{d}\tau \end{aligned} \tag{30}$$

Here we introduce τ-independent functions $c_{\bar{\varepsilon}} = \cos(\zeta_{\bar{\varepsilon}}\hat{\phi}_{0,\bar{\varepsilon}})$, $s_{\bar{\varepsilon}} = \sin(\zeta_{\bar{\varepsilon}}\hat{\phi}_{0,\bar{\varepsilon}})$, and auxiliary rectangular window, $w(t)$, being the boxcar function with limits matching those of the input excitation $t_0 \le t \le T + t_0$ and satisfying $w^i(t) = w(t)$, $\forall i > 0$.

Attending to all this, the integral in (30) may be formulated as

$$u_{\bar{\varepsilon}}(t,\bar{\theta})\cdot[2\ln(k)] = \begin{cases} c_{\bar{\varepsilon}}\left(T-\tau_{\bar{\varepsilon}}\right)\cdot\ln(k) & \text{if } t=-\tau_{\bar{\varepsilon}} \\ \Big(c_{\bar{\varepsilon}}\left[\mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}k^{T+t}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}\right) + \right. & \\ \quad \left. + \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}+}k^{T+t}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}+}\right)\right] + & \text{if} \\ + s_{\bar{\varepsilon}}\left[\mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}k^{T+t}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}\right) + \right. & -T<t<0 \\ \quad \left. + \mathrm{Si}\left(\alpha_{\bar{\varepsilon}+}k^{T+t}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}+}\right)\right]\Big) & \\ \Big(c_{\bar{\varepsilon}}\left[\mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}k^{T}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}k^{t}\right) + \right. & \\ \quad \left. + \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}+}k^{T}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}+}k^{t}\right)\right] + & \text{if} \\ + s_{\bar{\varepsilon}}\left[\mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}k^{T}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}k^{t}\right) + \right. & 0 \le t < T \\ \quad \left. + \mathrm{Si}\left(\alpha_{\bar{\varepsilon}+}k^{T}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}+}k^{t}\right)\right]\Big) & \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

with $\bar{\varepsilon}$-dependent, phase functions $\alpha_{\bar{\varepsilon}\pm}(t,\bar{\theta})$ and $\beta_{\bar{\varepsilon}}(t,\bar{\theta} = (t+\hat{\tau}_{\bar{\varepsilon}})/2$,

$$\alpha_{\bar{\varepsilon}\pm}(t,\bar{\theta}) = 2\pi f_0 k^{+\beta_{\bar{\varepsilon}}-t}\left(k^{+\beta_{\bar{\varepsilon}}} \pm k^{-\beta_{\bar{\varepsilon}}}\right)/\ln(k).$$

Operating on $u_{\bar{\varepsilon}}(t,\bar{\theta})$ one may rewrite these to separately address $\alpha_{\bar{\varepsilon}\pm}(t,\bar{\theta})$ as in

$$u_{\bar{\varepsilon}}(t,\bar{\theta}) = \tilde{u}_{\bar{\varepsilon}}(t,\bar{\theta}) + \rho_{\bar{\varepsilon}}(t,\bar{\theta}).$$

Terms in $\alpha_{\bar{\varepsilon}+}(t)$ are grouped into $\rho_{\bar{\varepsilon}}(t,\bar{\theta})$, and those in $\alpha_{\bar{\varepsilon}-}(t)$ go to $\tilde{u}_{\bar{\varepsilon}}(t,\bar{\theta})$. The rationale for this lays with the properties of these functions which we include in Appendix A (see also Abramowitz and Stegun (1964), p. 231). Due to the negligible effect of $\rho_{\bar{\varepsilon}}(t,\bar{\theta})$, one may indistinctively refer to $u_{\bar{\varepsilon}}(t,\bar{\theta})$ or $\tilde{u}_{\bar{\varepsilon}}(t,\bar{\theta})$.

Functions $\{\tilde{u}_{\bar{\varepsilon}}(t,\bar{\theta})\}$ are impulsive chirplets, oscillating at a varying frequency along $-T \leq t \leq T$, and may be assimilated to amplitude and phase modulated sinusoids. Differences among them are due to (i) time-shifts $\hat{\tau}_{\bar{\varepsilon}}$ (see Sect. 2.3), which also affect (ii) instantaneous frequency content (see Sect. 2.4); while they do not display spectral zeros within the frequency range of interest. Table 1 includes representative values of their instantaneous frequency content after the τ-integral has been evaluated. Dependence on $\mathscr{N}_{\bar{\varepsilon}}(\bar{\theta})$ is then evident. Moreover, their envelopes display an impulsive shape with *maxima* at $\hat{\tau}_{\bar{\varepsilon}}$ causing them to cluster following rules in (17).

Attending to this last effect, let us now define:

$$u_{\langle m\rangle}(t,\bar{\theta}) = \left\{u_{\bar{\varepsilon}}(t,\bar{\theta}) \mid \bar{\varepsilon} \in \mathscr{S}_{n\langle m\rangle}\right\} \qquad [m\text{-th cluster}] \tag{32}$$

$$u_m(t,\bar{\theta}) = \left\{u_{\bar{\varepsilon}}(t,\bar{\theta}) | \bar{\varepsilon} = \bar{1}_m\right\} = w(t)\cdot\cos\left(m\,\phi(t)\right) * \cos\left(\phi(-t)\right) \tag{33}$$

$$\hat{\tau}_m = \left\{\hat{\tau}_{\bar{\varepsilon}} \mid \bar{\varepsilon} = \bar{1}_m\right\} = -\log_k m \qquad [m\text{-th cluster center}] \tag{34}$$

From the definition of $\bar{\varepsilon} \in \{\mathscr{S}_{n,\langle m\rangle} \forall n | m\}$ we know that $u_m(t,\bar{\theta})$ is unique for order m, has an associated time shift $\hat{\tau}_m$, displays the property $\mathscr{N}_n(\bar{\theta}) \leq n$ and may be computed in advance from the input excitation, regardless of the SUT. Contrarily, $_{<m>}(t,\bar{\theta})$ defines a clustered, SUT-dependent set fully determined by corresponding vectors $\bar{\varepsilon} \in \{\mathscr{S}_{n,\langle m\rangle} | \sum_{i=1}^{n} \varepsilon_i = m\}$. Interestingly, each of these sets contains a single $u_m(t,\bar{\theta})$ with $\bar{\varepsilon} = \bar{1}_m$, playing a relevant role in the coming discussions.

In summary, one may identify a sets of functions $\{\tilde{u}_{\bar{\varepsilon}}(t,\bar{\theta})\}$ depending both on the underlying system and the input used to excite it. These are frequency-bounded, impulsive functions, clustering around $\{\hat{\tau}_m\}$.

3.2 Inverse Filtering

In (29), spectral coloring due to $u_{\bar{\varepsilon}}(t,\bar{\theta})$ may be mitigated by considering some form of inverse filtering. In doing so, we may be able to focus straightly on the time-shifted, Volterra kernels, i.e., the structural elements of the model for the system under test. Without loss of generality we hereafter focus on the cosine part of (31), affected by $c_{\bar{\varepsilon}}$. By appropriately selecting the parameters of the excitation —in this case, the initial value of the excitation's phase (ϕ_0), and/or the duration of the excitation (T)— one may cause that $\hat{\phi}_{0,\bar{\varepsilon}} = 0$ eliminating the sine part. Authors in Novák et al (2010) and Rébillat et al (2011) also discuss these parameters with a similar interest. The same conclusions included hereafter apply to the sine part and the aggregated function. We shall only focus on the former for simplicity.

Let us define time- and frequency-fading signal $\nu(t)$, covering $-T \leq t \leq +T$, $f_0 \leq f \leq f_1$ and satisfying definition

$$\nu(t) * u_{\bar{\varepsilon}}(t,\bar{\theta}) = d_{\bar{\varepsilon}}\left(t + \tau_{\bar{\varepsilon}}(\bar{\theta})\right) \tag{35}$$

so that $\{d_{\bar{\varepsilon}}(t+\tau_{\bar{\varepsilon}})\}$ retain the same clustered structure as $\{u_{\bar{\varepsilon}}(t,\bar{\theta})\}$ and when applied to the test output as in Figure 2 the cosine part in (29) equals

$$e_c(t) = h_0 + \sum_{n=1}^{N} \int_0^{+\infty} \frac{h_n(\bar{\theta})}{2^n} \sum_{m=1}^{n} \underbrace{\sum_{\bar{\varepsilon}\in\mathscr{S}_n\langle m\rangle} c_{\bar{\varepsilon}}\, d_{\bar{\varepsilon}}\left(t+\tau_{\bar{\varepsilon}}(\bar{\theta})\right)}_{\text{order } n \text{ contributors to } m\text{-th cluster}} d\bar{\theta} \tag{36}$$

The result holds for any system amenable to the Volterra representation; and resembles a convolution integral involving the multidimensional Volterra kernels and a set of multidimensional, time-shifted functions. Thus, in looking for $\nu(t)$, we look for a multidimensional linear convolution, with $d_{\bar{\varepsilon}}(t)$ playing a similar role to Dirac's delta $\delta(t)$ to minimize dependence of (36) on the excitation.

Considering the specifications for $\nu(t)$ and the properties of $\{u_{\bar{\varepsilon}}(t,\bar{\theta})\}$, it is simply not possible to obtain a filter such that $d_{\bar{\varepsilon}}\left(t+\tau_{\bar{\varepsilon}}(\bar{\theta})\right)$ equal pure, $\tau_{\bar{\varepsilon}}$-shifted Dirac deltas —see Theorems 2.8 and 3.1 of Casey and Walnut (1994). Nevertheless, under the assumption that the system frequency response falls into $f_0 \le f \le f_1$, one could still formulate a filter that is limited to this frequency range, and that produces an impulsive output, at least for one of these functions.

If $\nu(t)$ is a good approximation to $u_1^{-1}(t,\bar{\theta})$ and goes to zero outside the frequency range of interest, previous conditions hold for the first order ($m=1$), producing

$$\nu(t) * u_1(t,\bar{\theta}) \approx \delta(t). \tag{37}$$

Still, we must check whether this may be computed, and evaluate $\nu(t) * \tilde{u}_{\bar{\varepsilon}}(t,\bar{\theta})$ for any $\bar{\varepsilon} \in \mathscr{S}_{n\langle m\rangle}$ to assess its impact on (36).

1. this is a continuous, monocomponent, completely defined in $-T \le t \le +T$,
2. it displays no spectral zeros or singularities within $f_0 \le f \le f_1$, and has a smooth response within this range.

Consequently, $u_1(t,\bar{\theta})$ is invertible, and $\nu(t)$ could always be computed, for instance in Fourier's domain, down to an arbitrary degree of precision. Furthermore, regarding the result of $\nu(t) * \tilde{u}_{\bar{\varepsilon}}(t,\bar{\theta})$, due to the frequency-bounded responses of the signals involved, (i) output response must go to zero outside frequency range $(\mathscr{N}_{\bar{\varepsilon}} \cdot f_0, \mathscr{N}_{\bar{\varepsilon}} \cdot f_1) \cap (f_0, f_1)$. Additionally, (ii) since $u_{\bar{\varepsilon}}(t,\bar{\theta})$ are smooth, continuous and of the same form, their individual frequency responses differ by their respective support functions, and an $\bar{\varepsilon}$-dependent, t-independent time-shift.

From these we conclude that functions $\{d_{\bar{\varepsilon}}(t,\bar{\theta})\}$ are frequency-bounded, $\tau_{\bar{\varepsilon}}$-shifted, whitened, impulsive functions; and that they also cluster around $\tau_{\bar{\varepsilon}}$. Still, (36) is a nonlinear convolution in $\bar{\theta}$. We must reformulate the integral to derive the desired closed, seemingly linear form.

4 Reformulation of the test output

The daunting form of the n-dimensional nonlinear convolution in (36), implicitly computed during the system output postprocessing, may be overtaken by means of

Table 2 Summary of the transformations applied to the n-dimensional hyper-space

Coordinate system	Transformation	Range
$\bar{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$	———	$\theta_i \in (0, \Theta)$
$(\theta_1, \bar{\Delta}) = (\theta_1, \Delta_2, \ldots, \Delta_n)$	$\Delta_i = \theta_i - \theta_1, \forall i > 1$	$\theta_1, \Delta_i \in (-\Theta, \Theta)$
$(\theta_1, r, \bar{\varphi}) = (\theta_1, r, \varphi_1, \ldots, \varphi_{n-1})$	$\begin{cases} k^{\Delta_1/2} = r\cos(\varphi_1) \\ k^{\Delta_j/2} = r\prod_{m=1}^{j-2}\sin(\varphi_m)\cos(\varphi_{j-1}) \\ k^{\Delta_n/2} = r\sin(\varphi_1)\cdots\sin(\varphi_{n-1}) \end{cases}$	$\begin{cases} \theta_1 \in (0, \Theta) \\ r \in (r_m, r_M)^{\text{a}} \\ \varphi_i \in (0, \pi/2) \end{cases}$
$(\theta_1, s, \bar{\varphi}) = (\theta_1, s, \varphi_1, \ldots, \varphi_{n-1})$	$\begin{cases} s = 2\log_k(r) \\ \sigma_{\bar{\varepsilon}}(\bar{\varphi}) = \log_k \left\vert \begin{array}{l} \cos^2(\varphi_1) + \varepsilon_n \prod_{m=1}^{n-1}\sin^2(\varphi_m) + \\ + \sum_{i=2}^{n} \varepsilon_i \cdot \cos^2(\varphi_{i-1}) \cdot \prod_{m=1}^{i-2}\sin^2(\varphi_m) \end{array} \right. \end{cases}$	$\begin{cases} \theta_1 \in (0, \Theta) \\ s \in (\log_k r_m, \log_k r_M) \\ \varphi_i \in (0, \pi/2) \end{cases}$

[a] Constants r_m and r_M are given by (15) and equate, $\sqrt{|m\mathscr{R}_+ \pm n\mathscr{R}_-|}$, respectively.

a complete and informed transformation of the integration space. In the transformed space, the same finite number of time-frequency contributors should be expected, and time shifts ought to be equispaced. In this section we define this series of transformations only to analyze the properties of (36) in a more convenient space.

Let us consider the following sequence. First, focusing on time dimension θ_1, we already subtracted its influence from the rest of variables, introducing $\bar{\Delta}$. We regard these as deviations of the time dimensions, centered around $\bar{\Delta} = \bar{0}$.

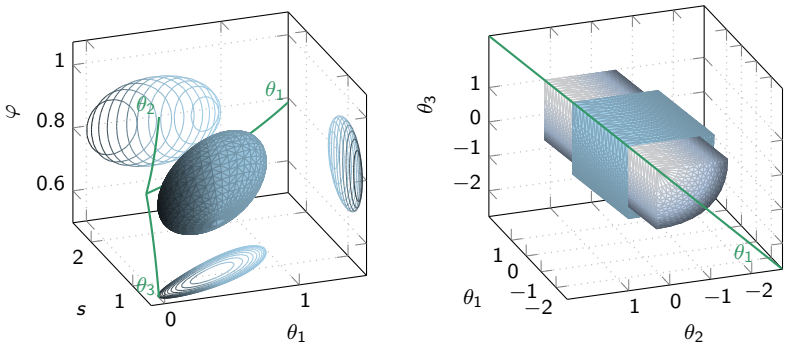
Second, based on the structure of the resulting exponential space, one may rely on the generalized n-dimensional spherical coordinate system to express the former time deviations. Taking one dimension to measure magnitude —in this case, the magnitude of time deviation— in the form of a non-negative radius, r, on which a logarithmic transformation is then applied; and the remaining $n-1$ dimensions as the generalized positioning angles $\bar{\phi} = (\phi_1, \ldots, \phi_{n-1})$, we obtain a convenient form for time delays in (36). The complete set of transformations are summarized in Table 2.

For ease of interpretation we now describe how these would affect two basic volumes for a 3rd-order system. In Figure 3a we present transformations on a 3D sphere originated in the $\bar{\theta}$-space transformed into the (θ_1, s, φ)-space. The result is a deformed sphere in the first quadrant, as expected for a causal system. In Figure 3b we represent the result of the inverse transformations on a set of cubes in the transformed space. The result is a cylindrical structure evidencing coefficients concentration around kernel's diagonal.

In applying all these transformations into (36) we have that $\tau_{\bar{\varepsilon}} = -\theta_1 + s + \sigma_{\bar{\varepsilon}}(\bar{\varphi}) - t_0$, where $\sigma_{\bar{\varepsilon}}(\bar{\varphi})$ groups the combined effects of $\bar{\varphi}$ and $\bar{\varepsilon}$, including (i) variability due to the nonlinearity and (ii) rotational symmetry in kernel coefficients. Specifically, s is an order-dependent, linear time dimension describing kernels' sparsity outside the diagonal.

From these, instead of the n-volume integral in $\bar{\theta}$, one may write (36) as:

$$e_c(t) = h_0 + \sum_{n=1}^{N}\sum_{m=1}^{n}\sum_{\bar{\varepsilon}\in\mathscr{S}_{n\langle m\rangle}} \hat{c}_{\bar{\varepsilon}} \int_0^{\infty}\int_0^{\frac{\pi}{2}}\int_0^{\infty} \frac{h_n(\theta_1, s, \bar{\varphi}) \cdot d_{\bar{\varepsilon}}(t - t_0 - \theta_1 + s + \sigma_{\bar{\varepsilon}})}{\sin(2\varphi_1)\cdots\sin(2\varphi_{n-1})} \mathrm{d}\theta_1 \mathrm{d}\bar{\varphi} \mathrm{d}s \tag{38}$$



(a) Direct transform (b) Inverse transform

Fig. 3 Results of the proposed transformations for $k = 1.5$, applied on: (a) a sphere of radius $\frac{1}{2}$ centered at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ in the $\bar{\theta}$-space, (b) a set of shifted cubes in the transformed space: $\varphi \in [0, \frac{\pi}{2}]$, $s \in [0, \frac{2}{5}]$; extremes, $s \in [0, \frac{3}{5}]$: central, shifted along θ_1.

The L_2-projections of the order-dependent response kernels, $\{h_n(\theta_1, s, \bar{\phi})\}$ for $1 \leq n \leq N$, on integration kernels $\{d_{\bar{\varepsilon}}(t, \theta_1, s, \bar{\phi})\}$, $\bar{\varepsilon} \in \mathscr{S}_n$, are easy to interpret whenever these could be used to expand the transformed system response Larson et al (2008). Particularly in this case, where the latter functions are intrinsically impulsive, time-shifted and frequency-bounded functions. Attending to this very specific nature, one may write

$$e_c(t) = h_0 + \sum_{m=1}^{N} \left[\sum_{n \leq m} \sum_{\bar{\varepsilon} \in \mathscr{S}_{n\langle m \rangle}} \int_0^{+\infty} \hat{c}_{\bar{\varepsilon}} \left(\int_0^{\pi/2} \frac{\tilde{h}_n(t - t_0 + s + \sigma_{\bar{\varepsilon}}, s, \bar{\varphi})}{\sin(2\varphi_1) \cdots \sin(2\varphi_{n-1})} \, \mathrm{d}\bar{\varphi} \right] \mathrm{d}s \right) \tag{39}$$

where $\hat{c}_{\bar{\varepsilon}}$ proportional to $c_{\bar{\varepsilon}}$ and $(\ln k)^{1-n}$. By integrating along $\bar{\varphi}$ and s in (39), we obtain linear combinations of chirplets in m, resulting from all $(n-1)$-multicombinations due to the NLS structure. Each of this is centered around fixed time stamp $\hat{\tau}_{\bar{\varepsilon}}$, related to $(t_0 - s)$ and $\sigma_{\bar{\varepsilon}}$, clusters around some $\hat{\tau}_m$ and are intimately related to the Volterra kernels, providing direct access to relevant information on the underlying system.

5 Informed estimation of order and memory

Attending to (39), the output of this procedure, $e(t)$, may be described as N consecutive, order-dependent packets, located around fixed time stamps $\hat{\tau}_m = \log_k(m)$ depending on the input excitation parameters, as well as the SUT memory (Θ) and order (N). We now intend to compute rough, but reliable estimates on the latter, bearing in mind the values assigned at the input to f_0, f_1, T and ϕ_0; only when the following conditions hold:

1. $e(t)$ resembles a sequence of discrete packets located around $\hat{\tau}_m$,
2. we observe dead-zones (energy gaps) between packets —i.e., kernels have finite responses.

In any other circumstance, the parameters of the excitation must be rearranged, aiming for the required gaps. Typically, by increasing the duration of the excitation (T) and/or varying the frequency range (decreasing $\ln(f_1/f_0)$), values for $\hat{\tau}_m$ readjust, widening the gaps between order-dependent packets. Only when kernel responses may not be finitely approximated, rearrangements will not produce the desired result and the proposed scheme will fail.

Assuming that we appropriately adjusted our excitation, we may proceed with the estimation. Let us formulate $e(t)$ in compact form based on our previous analysis

$$e(t) = \sum_{m=1}^{N} g_m(t+\tau_m) = h_0 + \sum_{m=1}^{N} \left[\sum_{\bar{\varepsilon}\in\mathscr{U}_m} \hat{c}_{\bar{\varepsilon}} \int_0^{+\infty} \int_0^{\pi/2} \frac{\tilde{h}_n(t-t_0+s+\sigma_{\bar{\varepsilon}}, s, \bar{\varphi})}{\sin(2\varphi_1)\cdots\sin(2\varphi_{n-1})} \, \mathrm{d}\bar{\varphi} \, \mathrm{d}s \right] \tag{40}$$

where $\tau_m = \log_k(m) - t_0$, coinciding with the time instants around which functions $u_m(t,\bar{\theta}) = w(t)\cdot\cos(m\,\phi(t)) * \cos(\phi(-t))$ cluster; and $\mathscr{U}_m = \left\{\bar{\varepsilon} \in \cup_{n=1}^{N} \mathscr{S}_{n\langle m\rangle}\right\}$. Attending to this, one may formulate the estimation of N and Θ as a pair of detection and estimation problems, stating that:

1. nonlinearity order equals N iff (i) we detect packet $g_N(t+\tau_N)$ in $e(t)$ and (ii) there is no higher value of m for which a packet is detected;
2. the memory-length Θ is upper bounded by packets' durations.

Despite the actual Volterra kernels being completely unknown to us, the signal formed from chirplets $\{d_m(t)\}$ may be precomputed off-line. Non-overlapping dead zones must exist around the order-dependent packets in intervals:

$$\begin{aligned} -[\tau_{m+}] \le t \le -[\tau_{m-} - \Theta] \qquad & \text{(packet)} \\ -[\hat{\tau}_{m+1-} - \Theta] \le t \le -[\hat{\tau}_{m+}] \qquad & \text{(dead zone)} \end{aligned} \tag{41}$$

where $1 \le m \le N$ indexes the consecutive packets, and the upper and lower bounds are given by:

$$\tau_{m\mp} = \log_k\left(\left|m\mathscr{R}_+ \pm n\mathscr{R}_-\right|/2\right) + t_0 \tag{42}$$

Figure 4 depicts an example on these upper and lower bounds using the same system and excitation parameters as on Figure 1a. Results evidence how higher orders in n project on the lower ones depending on m. Additionally, as n or Θ increase, clusters become wider, strangling the dead zones. Contrarily, adjustments on k also affect time warping (i.e., the base of the dead zone triangles in Figure 4), but may be used to enhance separability by simply increasing sweep duration T. In Schetzen (1980) the author already commented on the need to monotonically increase the duration of excitation sequences to mitigate the impact of estimation error during identification. The key here is to ensure that the process does not require complete identification of the system to produce reliable estimates on N and Θ.

Time boundaries for packets corresponding to orders l and $l+1$ for a nonlinearity of order N, $\hat{\tau}_N(l)$, are marked in Figure 4 along with their associated memory values, $\Theta_N^*(l)$. These were computed as the non-negative roots of $p_{m,N}(r)$ with $r = k^{\Theta_N^*(l)}$ (see dashed line in Figure 4) involving third-order polynomial

$$p_{m,n}(r) = (m+n)\,r^3 + (n-m-1)\,r^2 - (n-m)\,r - (m+n+1) \tag{43}$$

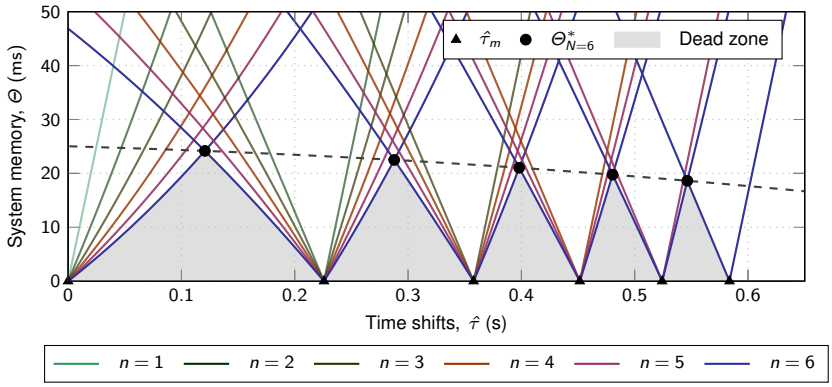


Fig. 4 Boundaries of the order-dependent time-shifts' distributions. Upper and lower bounds were computed according to (41), (42) for the same configuration as Figures 1a and 1b. Dead zones' widths depend on both the system memory and nonlinearity order, explaining the overlap in Figure 1b –see outer humps. Hard boundaries may be relaxed for low probability regions.

From these, the duration of the m-th packet, $g_m(t)$, must be

$$\tilde{T}_m = \Delta\,\hat{\tau}_m + \Theta \qquad 0 \leq \hat{\tau}_m \leq [\hat{\tau}_{m+}] - [\hat{\tau}_{m-}] \tag{44}$$

where $\Delta\,\hat{\tau}_m$ depends on the nonlinear structure of the underlying system and the system memory itself. It follows from our analysis that the shape of $g_m(t)$, and its duration $\Delta\,\hat{\tau}_m$ as well, are affected by all contributors projecting on a particular order. For some order n involving all multicombinations delivering vectors $\bar{\varepsilon} \in \mathscr{S}_n$, each vector belongs to one, and only one subset $\mathscr{S}_{n\langle m\rangle}$, which contributes to the m-th packet. This is centered around time delay $\hat{\tau}_m$ and involves multiple contributors up to a total of

$$2\sum_{n=1}^{N}\binom{n-1}{\frac{n-m}{2}}\left[(n+m+1) \pmod 2\right]. \tag{45}$$

In particular, the higher orders (even or odd) project on either one of the lower two (i.e., $n \in \{1,2\}$). Consequently, $\{\tilde{T}_1, \tilde{T}_2\}$ must provide rough but reliable upper bounds to the system memory

$$\Theta \leq \max_{1\leq m\leq N}\left\{\tilde{T}_m\right\} = \max_{m\in\{1,2\}}\left\{\tilde{T}_m\right\}. \tag{46}$$

Packets' average power is also affected by the number of contributors. This necessarily affects packet detection probability, suggesting that a joint approach should be followed for nonlinearity order and memory length estimation.

5.1 The packet detection approach

Attending to the order-dependent decomposition in $e(t)$ given by (40), order-dependent packets, $g_m(t+\hat{\tau}_m)$, mark the nonlinearity order (presence) and memory (duration) for the SUT, which we intend to characterize.

Hereafter, we formulate order (N) and memory (Θ) estimation in terms of an energy detection problem for arbitrary signals in noise. In doing so we consider time and energy distributions of $e(t)$, formed from order- and excitation-dependent pulses (41). Other approaches may exist that outperform the one described here. Anyhow, our only aim is to take advantage of the structure identified in this contribution, and provide reliable and early estimates on these parameters that are of practical use, all through a supervised procedure.

Let us consider $\tilde{e}(t)$, the result of the test for the system output in noise —i.e., $y(t)+\eta(t)$:

$$\begin{aligned}\tilde{e}(t) &= [y(t)+\eta(t)]*[\nu(t)*x(-t)] = e(t)+\eta(t)*[\nu(t)*x(-t)] = \\ &= \textstyle\sum_{m=1}^{N} g_m(t+\tau_m)+\tilde{\eta}(t)\end{aligned} \tag{47}$$

We assume that the perturbation at the system output, $\eta(t)$, is some additive noise, and $\tilde{\eta}(t)$ is some form of coloured noise, affected by the post-processing.

Let us focus on test output $e(t)$ along the interval between the m-th and $(m+1)$-th packets; within which an energy gap is assumed to exist. Attending to (47), the test output within this interval may be decomposed into the following

$$\Big[g_m(t+\tau_m)+g_{m+1}(t+\tau_{m+1})\Big]+\sum_l g_l(t+\tau_l)+\tilde{\eta}(t) \qquad -\hat{\tau}_{m+1}\le t\le -\hat{\tau}_m \tag{48}$$

where $1\le l\le N$ and $l\neq\{m,m+1\}$. Three major contributors are here present: the two associated to the m-th and $(m+1)$-th packets, respectively; and the energy gap (dead zone, or low energy region) in between those, degraded by the underlying noise and neighboring packets' tails.

We assume that excitation parameters were selected so that energy gaps may be observed. Contrarily, the packet-related events may appear or not depending on the SUT, and no cross-dependence exists on projections with different parities. Hence, within time interval $-\tau_{m+1}\le t\le -\tau_m$ one of three different scenarios takes place: that the energy gap is observed, or not; and that packets, m-th and/or $(m+1)$-th, are present, or not.

We propose a sequential approach based on the analysis of the energy distribution around consecutive packet pairs —and the adjoining dead zone. One may evaluate energy distribution via the following integral, energy operator,

$$\mathcal{E}(s'_{m+1},s_m)=\int_{s'_{m+1}}^{s_m}\tilde{e}^2(t)\,\mathrm{d}t \tag{49}$$

where $-\tau_{m+1}\le s'_{m+1}\le t\le s_m\le -\tau_m\le 0$; $0<\mathcal{E}(s'_{m+1},s_m)<\infty$ —see sect. *Operator Theory and Modulation Spaces* in Larson et al (2008). Whenever integration limits approach values $\{\tau_m\}$ from the left ($s'_{m+1}\to -\tau_{m+1}$) or the right ($s_m\to -\tau_m$), one will incorporate energy content from the order-dependent packets. Contrarily, while moving outside their influences mitigate, and low energy levels should be expected. We hereafter assume that as long as (s'_{m+1},s_m) define intervals of significant width within dead zones defined by (41), energy levels must certainly be low —i.e., fading-memory. Then, by considering interval $\bar{s}\mp\Delta s$ one may evaluate $\mathcal{E}(s'_{m+1},s_m)$ along Δs and compute

$$\Lambda\,(\bar{s},\Delta s)=10\log_{10}\left(\mathcal{E}(\bar{s}-\Delta s,\bar{s}+\Delta s)/\mathcal{E}_\infty\right) \tag{50}$$

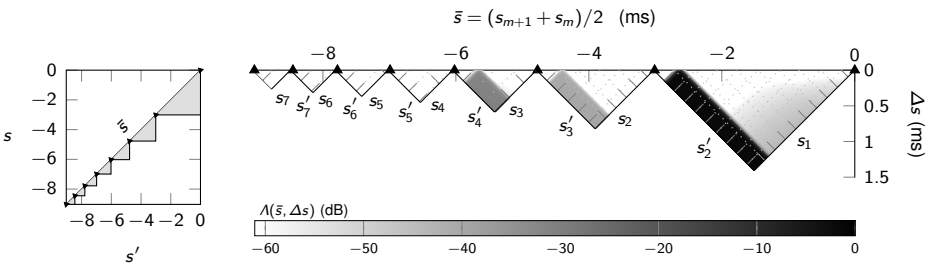


Fig. 5 Output of the energy detection test for nonlinearities orders up to $N^* = 7$. These are represented along axis $\bar{s}$, while considering individually the intervals between order-dependent packets. As expected values obtained for the higher orders or around the center of the triangles are below threshold, signaling contributors separability and detectability. Attending to this, one may state the order $N = 4$, as this is the higher order for which energy is detected, and look at the curves obtained for $s'_{2,1}$ constants to determine packets' durations.

where $\mathcal{E}_\infty$ is obtained from (49) for $s_{(m+1),m} \to \mp\infty$, representing the total amount of energy in $\tilde{e}(t)$, fractions of which we associate to the individual orders. This measure (i) is upper-bounded by 0 dB, representing the total energy of the signal; (ii) displays its maxima around axes (s'_{m+1}, s_m), and (iii) it is lower-bounded by the minimum detectable energy, λ, depending on the representativeness of the order-dependent packets and experimentation conditions. Attending to this one may proceed as follows:

1. extract interval $-\tau_{m+1} \leq t \leq -\tau_m$ out of $\tilde{e}(t)$,
2. evaluate $0 > \Lambda(\bar{s}, \Delta s) \lesseqgtr \lambda^*$, assuming that contributions below the threshold are negligible,
3. associate contributions to either one of the two order-dependent packets.

We shall then discuss the selection of λ^* and the association of the contributions to individual packets.

5.2 Practical use: the supervised procedure

For the practical use of the proposed solution one must turn to fuzzy concepts such as estimates' reliability and relevance towards system modelling.

In real world applications, limitations concerning every parameter of a system are naturally present. Bearing this in mind, our intend is to minimize estimation error while providing the simplest (less demanding) possible model parameters that are descriptive of the system behavior. Those *criteria* require that we (i) maximize the utility of memory representation (i.e., we use the fewer coefficients possible to represent the system), and (ii) ensure that all relevant contributors are present in our model —i.e., include as many nonlinearity orders as required.

Performance is then conditioned by energy and time dispersion of the order-dependent kernels, which will be detected or not depending on the significance of their contributions. Hence, we assume that when signals $x(t)$ and $y(t)$ in Figure 2 were respectively generated and recorded, sensible experimentation conditions were

met. This is, low noise level, enough bits for samples' encoding and full vertical range, resulting in a sensible signal-to-noise ratio, ρ.

Let us assume that full range was satisfactorily set, that it took B bits of the analog-to-digital converter with perfect, uniform bit distribution to encode samples, and that we obtained a ρ signal-to-noise power ratio. Then, for an additive noise affecting $y(t)$ in (47), the average noise energy level is

$$10^{\lambda/10} = \int_{\bar{s}-\Delta s}^{\bar{s}+\Delta s} |\tilde{\eta}(t)|^2/\mathcal{E}_\infty \mathrm{d}t \approx \frac{1/(2T)}{1+\rho} \tag{51}$$

where the final equation requires a stationary, noise process. Here we assume that this condition holds, that the number of bits is 'enough', and quantization noise is below measuring noise. By taking $\lambda^* = \lambda + \beta$ as the detector threshold one may adapt the performance of the system considering noise statistics, in account for the unknown signal statistics.

Figure 5 describes the results of the energy detector on a 4-th order, Volterra-Wiener system, for a 20 dB SNR and a 12-bit DAC. The individual triangular areas contain the dead zones between consecutive packets, located around time-stamps given by the input excitation. Within these one may find dark areas eminently parallel to one of the axes, associated to one of the surrounding packets. These appear for the lower orders up to $N = 4$ (higher values of $\tilde{\tau}_m$) but not for the higher ones. Concluding that at least a 4-th order nonlinearity should be used for modelling the system.

Furthermore, the upper bound to the system memory length may be estimated as the maximum duration associated to packets 1 or 2 —see (46). On Figure 5, we assume that the darker area is the contribution of the 2-nd order packet, the duration of which is given by its projection on s_1.

5.3 The proposed procedure

Despite the complexity of previous derivations, the proposed algorithms for memory length and nonlinearity order estimation are actually simple an easy to use. Hereafter we include the list of steps one must follow, and that we used for experiments in the coming sections.

Let us test some nonlinear system that we assume to be representable by Volterra's or Wiener's series. No prior information is available regarding the length of its memory, Θ, or the order of the nonlinearity, N. We may like to focus on some frequency range $f_0 \leq f \leq f_1$, and look for early estimates on these parameters prior to any kind of identification. The procedure is then:

1. make an initial guess on T;
2. compute the rate of exponential increase $k = (f_1/f_0)^{1/T}$;
3. compute phase function and produce the sweep excitation

$$\phi(t) = \tfrac{2\pi f_0}{\ln(k)}(k^{t-t_0} - 1) + \tfrac{2\pi f_0}{\ln(k)} \tag{52}$$

$$x(t) = \cos(\phi(t)) \tag{53}$$

4. compute $u_1(t) = x(t) * x(-t)$ and produce best possible $\nu(t) \approx u_1^{-1}(t)$;

5. excite the system with $x(t)$ and record its output $\tilde{y}(t)$ in the best possible noise conditions while aiming to satisfy Nyquist criterion $F_s \geq N \cdot f_1$;
6. compute the post-processed output $\tilde{e}(t) = \tilde{y}(t) * [\nu(t) * x(-t)]$;
7. evaluate $\tilde{e}(t)$: we look at the order-dependent time shifts $\hat{\tau}_m = \log_k(m)$ for the expected packets: $g_m(t+\tau_m)$, and the desired energy dead zones. If consecutive packets do not display the necessary gaps, the input excitation, T, must be increased and the process continues from Step 1. If the result is positive:
8. estimate signal to noise ratio on $\tilde{y}(t)$ and for order m compute

$$\mathcal{E}(s'_{m+1}, s_m) = \sum_{r=s'_{m+1}}^{s'_m} \tilde{e}^2[r] \tag{54}$$

with $-\lfloor \tau_{m+1}/F_s \rfloor \leq s'_{m+1} \leq r \leq s_m \leq -\lfloor \tau_m/F_s \rfloor$, as well as the total energy of the postprocessed output $\mathcal{E}_\infty = \sum_r \tilde{e}^2[r]$; and along $s'_{m+1} = \bar{s} - \Delta s$ and $s_m = \bar{s} + \Delta s$

$$\Lambda(\bar{s}, \Delta s) = 10 \cdot \log_{10}(\mathcal{E}(\bar{s} - \Delta s, \bar{s} + \Delta s)/\mathcal{E}_\infty) \tag{55}$$

$$\lambda^* = 10 \cdot \log_{10} \tfrac{1/(2T)}{1+\rho}; \tag{56}$$

9. compute $\tilde{T}_m = |\tau_{m+} - \tau_{m-}|$ where one must evaluate $\lfloor \tau_{m+1}/F_s \rfloor \leq \bar{s} \leq \lfloor \tau_{m-1}/F_s \rfloor$, $\tau_0 = T$ and $\tau_{m-} = \min(\bar{s})$, $\tau_{m+} = \max(\bar{s})$ satisfying constraints

$$\begin{array}{c} \Lambda(\bar{s}, \Delta s) \geq \lambda^* \quad \forall(\bar{s}, \Delta s) \\ \tau_{m-} \leq \tau_m \leq \tau_{m+} \end{array}. \tag{57}$$

From $\{\tilde{T}_1, \tilde{T}_2\}$ one may compute a reliable estimate of the system memory length

$$\Theta \leq \max_{m \in \{1,2\}} \{\tilde{T}_m\}; \tag{58}$$

10. determine highest detectable order, this is $\tilde{T}_m > 0$. Then $N = m$.

6 Experiments

To exemplify on the structure of $e(t)$ and test the performance of the proposed scheme, we have simulated a series of Volterra-Wiener systems with specified parameters, and computed the proposed estimations on nonlinearity order and memory length. We report results on two different scenarios: (1) purely illustrative examples involving short, discrete, nonlinear systems, and (2) examples emulating real, continuous-time, nonlinear systems. Since the key to our analysis is the dispersion of coefficients for an unknown system structure, on each scenario we evaluate three different structures corresponding to: (i) a memoryless system, (ii) a Hammerstein structure where coefficients outside the main diagonal equal zero, and (iii) a Volterra-Wiener system displaying maximum kernel dispersion. Readers may refer to scenario 1 to deepen into our work, and to scenario 2 for a broader view on its applications.

On scenario 1, we evaluate $\mathsf{h}_1[n] = [-.25 \;\; .50 \;\; -.25]$, a three coefficients-long, low-pass filter. Contrarily, on scenario 2 we used a longer filter to emulate a continuous-time system. Here, $\mathsf{h}_2[n]$ is a digital filter obtained from the integration of

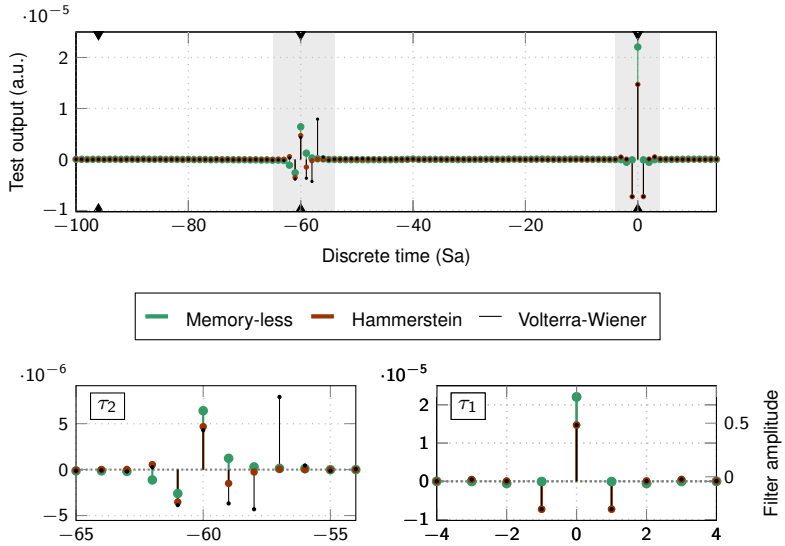


Fig. 6 Outputs of the test described in Figure 2 for several, 2-nd order, nonlinear systems and a particular exponential sweep excitation considering discrete time analysis. The systems implemented are those described in Scenario 1, introduced in Sect. 6, and include: (i) a memory-less, nonlinear system, (ii) a Hammerstein system, (iii) a Volterra-Wiener system. All outputs are depicted in the upper graph, where one may identify the two packets predicted by (39), located around $\{m_i\}_{1,2}$ samples —marked by the triangles. The individual packets corresponding to orders $m = 1,2$ for the Volterra-Wiener systems are depicted below. Differences are noticeable on the second order, and are explained by the different underlying nonlinear structures.

the 127-order (128 coefficients-long), band-pass, Butterworth filter, with a 1 dB ripple, 60 dB stop-band attenuation and cut-off frequencies $f_0 = f_s/800$ and $f_1 = f_s/8$. The systems implemented for scenarios 1 and 2 were obtained attending to Table 3 and Table 4, respectively. Additionally, for scenario 2 on Volterra-Wiener systems coefficients truncation was applied to evidence the role of the memory length system parameter and to test our estimation procedure. Interested readers may refer to Appendix B for an exhaustive description on the systems implemented, for both scenarios and all structures, including the exact coefficients of the filters.

Table 3 Summary of nonlinear systems implemented for Scenario 1

	Memory-less	Hammerstein	Volterra-Wiener
y[n]	$x[n]+x^2[n]$	$z_1[n]+z_2[n] = \mathrm{h}_1[n] * \left(x[n]+x^2[n]\right)$	$z_1[n]+z_1[n]\cdot z_1[n-6]+z_2[n]$

with $z_i[n] = \mathrm{h}_1[n] * x^i[n]$ used to produce Hammerstein's and Wiener's structures.

Figure 6 displays signals $e(t)$ for scenario 1, computed according to the proposed methodology (steps 1 to 6). The discrete-time input sequence was produced to include

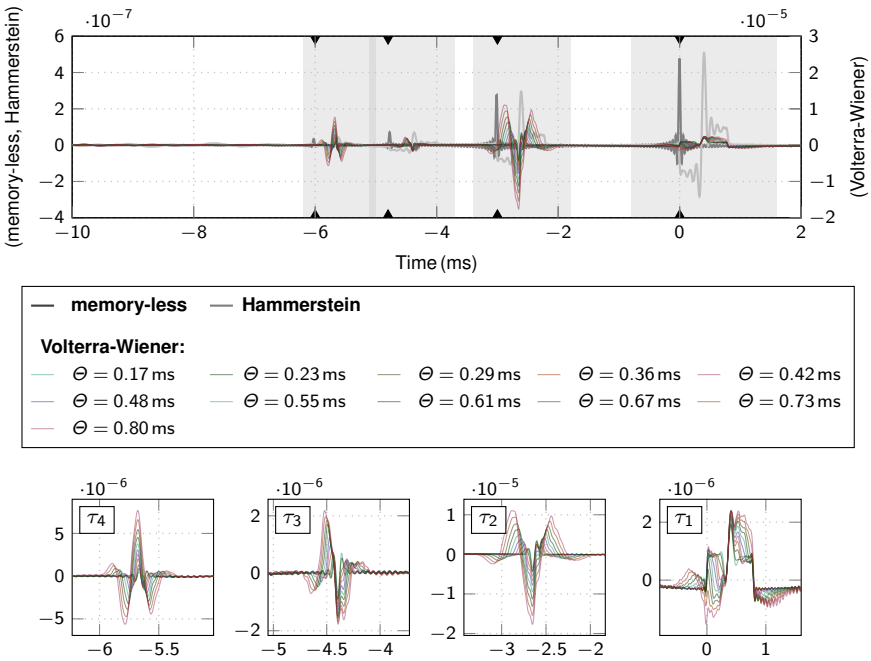


Fig. 7 Outputs of the test described in Figure 2 for various 4-th order, nonlinear systems and a particular exponential sweep excitation. Systems are: (i) a memory-less, nonlinear system, (ii) a Hammerstein system, (iii) a set of Volterra-Wiener systems with varying memory lengths. Excitation parameters were set to: $f_0 = 200\,\text{Hz}$, $f_1 = 20\,\text{kHz}$ and $T = 20\,\text{ms}$. All outputs are depicted in the upper graph, where one may identify the four packets predicted by (39), located around $\{\tau_i\}_{1\ldots4}$ time stamps —marked by the triangles. The individual packets corresponding to orders $m = 1,\ldots,4$ for the Volterra-Wiener systems are depicted below. Time bounds for the lower figures are marked by the shadowed areas in the top one. These are, for order m: $\min\{-\hat{\tau}_{\bar{\varepsilon}} | \bar{\varepsilon} \in \mathscr{U}_m\} \leq t \leq \max\{-\hat{\tau}_{\bar{\varepsilon}} | \bar{\varepsilon} \in \mathscr{U}_m\} + \Theta$.

$L = 3200$ coefficients and excite frequencies $\omega_0 = 1/800$ rad/s, $\omega_1 = 1/8\,\text{rad/s}$. Upsampling and downsampling with factor $\times 8$ were introduced during implementation.

In the same way, Figure 7 displays signals $e(t)$ for all simulated systems corresponding to scenario 2, in noiseless conditions, along frequency range 200 Hz $\leq$ f $\leq$ 20 kHz and using a sample rate of $f_s = 160\,\text{kHz}$ for acquisition. One may observe well-defined energy concentrations followed by low-energy valleys. Packets are located around their expected locations regardless of the system structure, whereas

Table 4 Summary of the kernels' construction for Scenario 2. We evaluate $1 \leq i \leq 4$ (i.e., $N = 4$).

	Memory-less	Hammerstein	Volterra-Wiener
$h_i(\bar{\theta})$	$\delta(\bar{\theta})$	$\begin{cases} \mathsf{h}(\theta), & \theta_{1\ldots i} = \theta \\ 0 & \text{otherwise} \end{cases}$	$\times_{j=1}^{i} \mathsf{h}(\theta_j)$

slight shifts towards the left-side appear on Wiener systems, which are absent from the Hammerstein one.

7 Results

On Figures 6 and 7, one may observe the postulated order-depedent packets are present, with the required energy gaps between them to continue our analysis. Should it happen that signals $e(t)$ do not display such pattern, following the proposed procedure T ought to be increased, and steps 1 to 6 repeated.

We now include further analysis on the test outputs attending to SUT structures.

7.1 Memoryless, nonlinear systems

The memoryless nonlinear structure is essentially a mathematical formulation with limited real-life applications due to its infinite bandwidth. Still, is quite informative for the kind of results one should expect from our analysis, as it brings the shortest possible memory estimate for our method.

On both scenarios, 1 and 2, one may observe impulsive peaks on the test output, located at prespecified time-stamps $\hat{\tau}_m$ for $1 \leq m \leq 2$ and $1 \leq m \leq 4$, respectively. According to (40), when the order-dependent kernels tend to match ideal Delta functions, one may observe that the order-dependent chirplets approximate the functions introduced in Sect. 3. This is,

$$g_m(t) = \nu(t) * u_m(t, \bar{\theta}). \tag{59}$$

The exact functions may be computed based on the formulations derived there.

7.2 Generalized Hammerstein systems

Contrarily to the memoryless network, Generalized Hammerstein systems are naturally bounded, and therefore may be used to model real-life systems. On these structures the n-th kernel only displays non-zero values along its diagonal ($\bar{\varepsilon} = 1_n$, $\theta_1 = \ldots = \theta_n$), and these diagonals may display the same values regardless of n —Hammerstein—, or not —generalized Hammerstein. Thus, integration in (40) restricts to $\bar{\Delta} = \bar{0}$, $s = \log_k(n)$, $\sigma_{\bar{\varepsilon}} = 0$.

Consequently, (i) order-dependent replicas are to be found at time stamps $\hat{\tau}_m = \log_k(m)$, just as for the memoryless structure; while (ii) integrands corresponding to packets $\{g_m(t)\}$ resemble the linear combination of the individual kernels' diagonals. This is consistent with results reported in Farina (2000), Novák et al (2010) and Rébillat et al (2011), and explains why one may compute estimates on N and Θ (or M for the discrete examples in scenario 1), from the post-processed output of the system.

7.3 Volterra-Wiener systems

Volterra's and Wiener's series match the very general form of any memory- and order-fading systems. These are used to approximate real-life systems with great success, even though they may involve heavy computations when addressing full identification. Fortunately, it is possible to extract some information without completing this process, as we propose in this work.

Attending to (40) and transformations in Table 2, s is a linear, causal time dimension along which $\bar{\varepsilon}$-dependent time-shifts cluster; while coefficients $\sigma_{\bar{\varepsilon}}$ represent deviations around the former, introduced by the kernel inner structure. From these, one should then expect qualitatively similar results to those on the analysis of Hammerstein systems, as Figures 6 and 7 evidence. Differences in the duration of the individual packets and their shapes are explained by the different contributions to each of the packets. Combinations imposed by the NL structure and differences in the order-dependent kernels explain why these $\{g_m(t)\}$ greatly vary from one example to another.

To provide further evidence, in Figure 7 we included results on systems simulated from truncated versions of $\mathsf{h}_2[n]$. This is, instead of using the 128 coefficients of the filter, we truncated the first and last ones to produce systems with shorter memory.

Once step 7 is completed, we computed estimates for the nonlinearity-order (steps 8 and 9) and memory-length parameters (step 10) attending to the procedure described in Sect. 5.3. Through the detection process, we observe that all systems evaluated could be approximated by a 4-th order nonlinearity for a 20 dB SNR. Actually, results depicted in Figure 5 are those computed for the Volterra-Wiener system with memory $\Theta = 0.17$ ms. By simple inspection on this figure one may identify the contributions of orders 1 to 4, while for the higher ones no contribution is observed above noise level. Results on memory length estimation are displayed in Figure 8 for the same Volterra-Wiener systems and while introducing additive, white, gaussian noise on system output $y(t)$ at different signal-to-noise ratios (ρ). Deviations in memory length are due to the Volterra-Wiener structure, while deviations observed on the computed estimates are due to embedded noise. As noise conditions improve, estimates become closer to the actual values, yet overestimated due to the unavoidable NL structure effect.

8 Discussion and remarks

The identification of non-linear systems is a complex, highly demanding task; especially when little is known about the SUT. The need for simple, yet robust techniques to roughly but easily estimate certain elementary system parameters has been a constant. Among these, memory length and nonlinearity order are two of the most relevant. Both have direct and critical impact on time and frequency responses, conditioning system output acquisition, and therefore any posterior intend of identification.

The analysis described here extends the results from previous contributions concerning weakly nonlinear, Hammerstein Farina (2000) and generalized Hammerstein Rébillat et al (2011) systems, to the more general, memory- and order-fading,

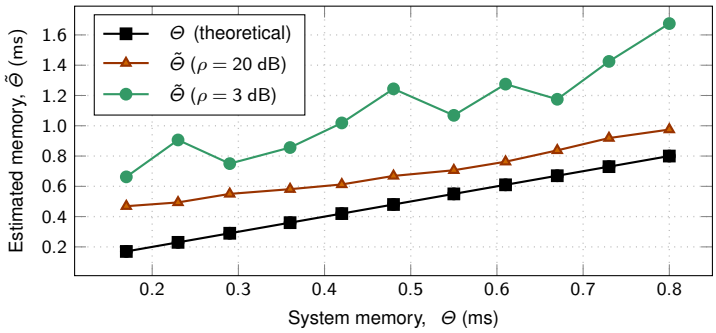


Fig. 8 Estimates on the system memory parameter for the same Volterra-Wiener systems in Figure 7. The actual values are upper bounded by the delivered estimates due to the system structure. Additional deviations due to poor SNR (ρ) also appear.

Volterra-Wiener systems. Despite the complexity of the derivations, the procedure enumerated in Sect. 5.3 only requires the supervised adaptation of the parameters in a phase modulated sinusoid (an exponential sweep) for exciting the system. Output is then post-processed based on the input, and an energy-based detector is used to discover energy packets. The exponential sweep pattern is most frequently incorporated into general purpose, arbitrary pattern generators; but to the authors knowledge, have never been used for the systematic, preliminary evaluation of non-linear systems. Moreover, post-processing of recorded outputs requires a single linear filter, on a signal that may be precomputed from the input excitation.

Even when the initial guess on sweep duration T is insufficient attending to Sect. 5.3, repeating the whole process on a longer signal has a minor computational cost. This is dramatically different for full, kernel-identification techniques, were sequences are fairly long and calculations are computationally expensive. Taking the emulated, continuous-time Voltera-Wiener system, we used a 3200 coefficients-long sequence ($\lceil T \cdot F_s \rceil$) and performed one linear convolution to compute our estimates. Contrarily, according to Schetzen (1980), it would require $\sum_{n=1}^{N} \binom{M+n}{n} > 10^{11}$ coefficients to evaluate nonlinearity order $N = 7$, and only assuming perfect knowledge on the memory-length parameter ($M = 128$).

In addition to its low complexity, the test output is easy to interpret. Particularly compared to the current state of the art for order evaluation (e.g. total harmonic distortion) and memory estimation (e.g. standard correlation tests). Large dependence on the input parameters facilitates on the fly corrections to enhance separability on the order-dependent packets, and ensure good estimated on N and Θ.

Finally, definitions in this contribution describe our interpretation of order and memory estimation. These, along with (41) and Table 1 provide a concrete, interdependent approximation to the abstract notion of memory- and order-fading systems Zang and Iglesias (2003). We say that systems satisfy this definition if kernels:

1. display a systematic energy decrease as order increases, such that for $m = N^*$ the energy of the corresponding packet lays below noise level and system reconstruction considering orders $1 \leq m \leq N^*$ is acceptable in some sense;
2. for $1 \leq m \leq N^*$ the required gaps between orders do exist, implying that kernels' memory length is upper bounded by $\Theta^*_{N^*}(m)$.

Despite the complexity of these definitions and formulations in Sections 2 to 4, the simplicity of the test suggests that it may be used to complement most black-box identification techniques. Thus, providing engineers with a reliable, hands-on tool, to help at the first stages of a systems' identification.

A Evaluation of functions $\{u_{\bar{\varepsilon}}(t,\bar{\theta})\}$

Here we include a series of relevant properties concerning $\{u_{\bar{\varepsilon}}(t,\bar{\theta})\}$ used in this paper.

Property 1 Functions $\{u_{\bar{\varepsilon}}(t)\}$ are time- and energy-bounded. Their envelopes decrease along the warped time axis as a function of k when moving away from $t = \tau_{\bar{\varepsilon}}$. This falls from the following properties:

asymptotic behavior
$$\lim_{t\to\pm\infty} \gamma_{\bar{\varepsilon},2}(t) = 0 \tag{60a}$$
$$\lim_{t\to\pm\infty} \gamma_{\bar{\varepsilon},4}(t) = 0 \tag{60b}$$

centered maximum
$$\gamma_{\bar{\varepsilon},2}(\tau_{\bar{\varepsilon}}) \geq \gamma_{\bar{\varepsilon},2}(t)\ \forall t \tag{61}$$

spectral envelope bounds
$$\begin{gathered} |\Gamma_{\bar{\varepsilon},\{1,3\}}(f)| \ll |\Gamma_{\bar{\varepsilon},\{2,4\}}(f)| \\ f_0 \leq f \leq f_1 \end{gathered} \tag{62}$$

Proof Based on (31), we are interested in all pairs of the *sine* and *cosine* integrals and $\{\alpha_{\bar{\varepsilon}+}, \alpha_{\bar{\varepsilon}-}\}$.

$$\gamma_{\bar{\varepsilon},i}(t) = \int_{-\infty}^{+\infty} W(\tau,t)\cos(\alpha_{\bar{\varepsilon}\bullet}k^{\tau})\,\mathrm{d}\tau = \frac{1}{\ln(k)}\begin{cases} \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}\bullet}k^{T+t}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}\bullet}\right) & -T < t < 0 \\ \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}\bullet}k^{T}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}\bullet}k^{t}\right) & 0 \leq t < T \end{cases}$$
$$\gamma_{\bar{\varepsilon},j}(t) = \int_{-\infty}^{+\infty} W(\tau,t)\sin(\alpha_{\bar{\varepsilon}\bullet}k^{\tau})\,\mathrm{d}\tau = \frac{1}{\ln(k)}\begin{cases} \mathrm{Si}\left(\alpha_{\bar{\varepsilon}\bullet}k^{T+t}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}\bullet}\right) & -T < t < 0 \\ \mathrm{Si}\left(\alpha_{\bar{\varepsilon}\bullet}k^{T}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}\bullet}k^{t}\right) & 0 \leq t < T \end{cases}$$

where $\bullet$ takes the positive subindex for $i = 1$, $j = 3$, and negative for $i = 2$, $j = 4$; and from the boxcar function $w(t)$ restrited to $t_0 \leq t \leq T + T_0$, $W(\tau,t,\bar{\theta}) = w(\tau - t)\cdot\prod_{i=1}^{n} w(\tau - \theta_i)$ to sketch the boundaries of time integrals in (29).

Asymptotic behavior. For an arbitrary duration T while moving away from $t = \tau_{\bar{\varepsilon}}$, combining *cosine* and *sine* integrals by pairs we prove (60a) and (60b):

$$\lim_{t\to-\infty} \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}k^{T+t}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}\right) = 0 \quad \lim_{t\to+\infty} \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}k^{T}\right) - \mathrm{Ci}\left(\alpha_{\bar{\varepsilon}-}k^{t}\right) = 0$$
$$\lim_{t\to-\infty} \mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}k^{T+t}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}\right) = 0 \quad \lim_{t\to+\infty} \mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}k^{T}\right) - \mathrm{Si}\left(\alpha_{\bar{\varepsilon}-}k^{t}\right) = 0$$

Time and energy bounds. We evaluate $\gamma_{\bar{\varepsilon},\{1,2\}}(t)$ while approaching zero argument, where a singularity is expected. On $\gamma_{\bar{\varepsilon},1}(t)$ function $\alpha_{\bar{\varepsilon}+}(t)$ prevents the argument from reaching zero argument; whereas on $\gamma_{\bar{\varepsilon},2}(t)$ we have that $\cos(\alpha_{\bar{\varepsilon}-}(t)\cdot k^{\tau}) \to 1$ when $t \to \tau_{\bar{\varepsilon}}$. Consequently the whole integral goes to a finite value. There is no singularity at $t = \tau_{\bar{\varepsilon}}$.

Since there are no singularities in the phase function, T is bounded and all integral forms are time-bounded according to $W(t,\tau)$, we conclude that functions $\gamma_{\bar{\varepsilon},1}(t)$ are energy-bounded. Finally, one may formulate $u_{\bar{\varepsilon}}(t)$ as a combination of functions $\gamma_{\bar{\varepsilon},m}(t)$, implying that $\{u_{\bar{\varepsilon}}(t)\}$ are also time- and energy-bounded.

Maxima and spectral content. Each of these integral forms may (asymptotically) be approximated by a linear combination of amplitude- and phase-modulated contributors attending to the trigonometric integrals' properties (see Abramowitz and Stegun (1964)) as follows:

$$\mathrm{Ci}(x) = p(x)\sin(x) - q(x)\cos(x) \tag{63}$$

$$\mathrm{Si}(x) = \tfrac{\pi}{2} - p(x)\cos(x) - q(x)\sin(x) \tag{64}$$

These are linear combinations of sines and cosines of an arbitrary argument, x, incorporating amplitude modulations that depend on x as well. We know from Abramowitz and Stegun (1964) that $|p(x)|$ and $|q(x)|$: (i) both tend to a constant value when $t \to \pm\infty$, (ii) their supreme is found at $x = 0$, and (iii) are monotonously decreasing while moving away from $t = 0$. Additionally, they provide the foundations for functions' evaluation. For instance, one may prove by derivation that $\gamma_{\bar{\varepsilon},2}(t)$ displays its absolute maximum at $t = \tau_{\bar{\varepsilon}}$ —i.e., has a centered maximum, (61).

Moreover, based on the definition of $\alpha_{\bar{\varepsilon}\bullet}(t)$, for any linear function of time $r(t)$ it holds:

$$\alpha_{\bar{\varepsilon}+}k^r = \alpha_{\bar{\varepsilon}-}k^r + \tfrac{4\pi f_0}{\ln(k)}k^{r-\tau_{\bar{\varepsilon}}} \tag{65}$$

relating phase functions of $\gamma_{\bar{\varepsilon},\{1,3\}}$ and $\gamma_{\bar{\varepsilon},\{2,4\}}$. From these we evaluate the spectral content of $\{\gamma_{\bar{\varepsilon},l}(t)\}$ based on the complex, analytic signal and the analysis in Cohen (2000). We extract relevant information on the spectral envelopes of $\{\gamma_{\bar{\varepsilon},l}(t)\}$, namely, $|\Gamma_{\bar{\varepsilon},l}(f)|$ from the properties of $|p(x)|$ and $|q(x)|$:

1. as the absolute value of the phase increases, amplitude decreases. Hence:

$$|p(\alpha_{\bar{\varepsilon}+}k^r)| < |p(\alpha_{\bar{\varepsilon}-}k^r)|\,,\ |q(\alpha_{\bar{\varepsilon}+}k^r)| < |q(\alpha_{\bar{\varepsilon}-}k^r)|$$

   so the numerators of the envelopes corresponding to $\Gamma_{\bar{\varepsilon},\{2,4\}}$ are larger than those of $\Gamma_{\bar{\varepsilon},\{1,3\}}$.
2. From (65) one may compute the second-order time derivative of the phase functions:

$$\frac{\mathrm{d}^2\,\alpha_{\bar{\varepsilon}+}k^r}{\mathrm{d}t^2} = \frac{\mathrm{d}^2\,\alpha_{\bar{\varepsilon}-}k^r}{\mathrm{d}t^2} + 4\pi f_0 \ln(k)\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 k^{r-\tau_{\bar{\varepsilon}}}$$

   where the second derivatives in $\alpha_{\bar{\varepsilon}\bullet}(t)$ happen to be non-negative; and the equating addend is always positive. Hence:

$$\left|\mathrm{d}^2\,\alpha_{\bar{\varepsilon}-}k^r/\mathrm{d}t^2\right| < \left|\mathrm{d}^2\,\alpha_{\bar{\varepsilon}+}k^r/\mathrm{d}t^2\right| \tag{66}$$

   In other words, the envelopes' denominators are smaller for $\Gamma_{\bar{\varepsilon},\{2,4\}}$ than on $\Gamma_{\bar{\varepsilon},\{1,3\}}$.
3. Equations (63) and (64) are coherent summations, hence the resulting spectral envelope is the direct summation of the spectral envelopes.

   In summary, attending to numerators and denominators for $\Gamma_{\bar{\varepsilon},\{1,\ldots,4\}}(f)$ we have,

$$|\Gamma_{\bar{\varepsilon},\{1,3\}}(f)| \ll |\Gamma_{\bar{\varepsilon},\{2,4\}}(f)|$$

Property 2 Approximating $u_{\bar{\varepsilon}}(t)$ through $\tilde{u}_{\bar{\varepsilon}}(t)$ is acceptable. Based on function $u_{\bar{\varepsilon}}(t) = \tilde{u}_{\bar{\varepsilon}}(t) + \rho_{\bar{\varepsilon}}(t)$, we say that $\tilde{u}_{\bar{\varepsilon}}(t)$ is a good approximation iff $|P_{\bar{\varepsilon}}(f)| \ll |U_{\bar{\varepsilon}}(f)|$ within $f_0 \le f \le f_1$. The approximation term as derived from $u_{\bar{\varepsilon}}(t)$ is:

$$\tilde{u}_{\bar{\varepsilon}}(t) = \int_{-\infty}^{+\infty} \frac{W(\tau,t)}{2}\left(c_{\bar{\varepsilon}}\cos(\alpha_{\bar{\varepsilon}-}k^\tau) + s_{\bar{\varepsilon}}\sin(\alpha_{\bar{\varepsilon}-}k^\tau)\right)\mathrm{d}\tau$$

and the deviation term:

$$\begin{aligned}\rho_{\bar{\varepsilon}}(t) = \frac{1}{2}\int_{-\infty}^{+\infty} W(\tau,t)\big(&\cos(\alpha_{\bar{\varepsilon}+}k^\tau)\left[c_{\bar{\varepsilon}}\cos 2\Delta\phi + s_{\bar{\varepsilon}}\sin 2\Delta\phi\right] + \\ &+ \sin(\alpha_{\bar{\varepsilon}+}k^\tau)\left[c_{\bar{\varepsilon}}\sin 2\Delta\phi - s_{\bar{\varepsilon}}\cos 2\Delta\phi\right]\big)\mathrm{d}\tau\end{aligned}$$

Proof Both the approximation and the deviation terms can be written as a linear combination of $\{\gamma_{\bar{\varepsilon},l}(t)\}$, defined in Property 1,

$$\begin{aligned}\rho_{\bar{\varepsilon}}(t) &= \frac{1}{2}\left[\cos\left(-(1+m)\Delta\phi\right)\gamma_{\bar{\varepsilon},1} - \sin\left(1+m)\Delta\phi\right)\gamma_{\bar{\varepsilon},3}\right] \\ \tilde{u}_{\bar{\varepsilon}}(t) &= \frac{1}{2}\left[\cos\left((1-m)\Delta\phi\right)\gamma_{\bar{\varepsilon},2} + \sin\left((1-m)\Delta\phi\right)\gamma_{\bar{\varepsilon},4}\right]\end{aligned} \tag{67}$$

Furthermore, $\rho_{\tilde{\varepsilon}}(t)$ may be rewritten, and due to linearity of the Fourier transform:

$$\begin{aligned} P_{\tilde{\varepsilon}}(f) &= c_{\tilde{\varepsilon}}\ [\cos(2\Delta\phi)\,\Gamma_{\tilde{\varepsilon},1}(f) - \sin(2\Delta\phi)\,\Gamma_{\tilde{\varepsilon},3}(f)] + \\ &+ s_{\tilde{\varepsilon}}\ [\sin(2\Delta\phi)\,\Gamma_{\tilde{\varepsilon},1}(f) + \cos(2\Delta\phi)\,\Gamma_{\tilde{\varepsilon},3}(f)] \\ U_{\tilde{\varepsilon}}(f) &= c_{\tilde{\varepsilon}}\,\Gamma_{\tilde{\varepsilon},2}(f) + s_{\tilde{\varepsilon}}\,\Gamma_{\tilde{\varepsilon},4}(f) \end{aligned}$$

Hence, $|\Gamma_{\tilde{\varepsilon},\{1,3\}}(f)| \ll |\Gamma_{\tilde{\varepsilon},\{2,4\}}(f)|$. Thus, provided that Property 1 holds and noticing that the trigonometric constants are bounded, necessarily:

$$|P_{\tilde{\varepsilon}}(f)| \ll |U_{\tilde{\varepsilon}}(f)|.$$

B Description of the implemented nonlinear systems

We report results on three different configurations, including two different scenarios: one (1) purely illustrative, another (2) simulating a continuous-time system; as well as three different nonlinear structures: memoryless, generalized Hammerstein and Volterra-Wiener. Hereafter we include the exact definitions of the discrete systems used, including the coefficients for the discrete linear filters; along with the implementations for the system outputs. We refer by $x[n]$ to the input excitation.

B.1 Scenario 1

We used a discrete filter $\mathsf{h}_1[n]$ to emulate the system memory. The coefficients for these were

$$\begin{aligned} \mathsf{h}_1[n] &= -.25\ \delta[n-1] + .50\ \delta[n] - .25\ \delta[n+1] = \\ &= [-.25 \quad .50 \quad -.25] \end{aligned} \tag{68}$$

B.1.1 Memoryless nonlinear system

The output for the memoryless structure may be computed as:

$$y[n] = x[n] + x^2[n] \tag{69}$$

B.1.2 Generalized Hammerstein system

For the Generalized Hammerstein structure, we introduced $z_i[n] = \mathsf{h}_1[n] * x^i[n]$. The output of the system was then computed as:

$$\begin{aligned} y[n] &= z_1[n] + z_2[n] = \mathsf{h}_1[n] * \big(x[n] + x^2[n]\big) \\ &= -.25\ \big(x[n-1] + x^2[n-1]\big) + .50\ \big(x[n] + x^2[n]\big) - .25\ \big(x[n+1] + x^2[n+1]\big) \end{aligned} \tag{70}$$

B.1.3 Volterra-Wiener system

Continuing on $z_i[n] = \mathsf{h}_1[n] * x^i[n]$, the output of the system was then computed as:

$$\begin{aligned} y[n] &= z_1[n] + z_1[n] \cdot z_1[n-6] + z_2[n] = \\ &= \mathsf{h}_1[n] * \big(x[n] + x^2[n]\big) + (\mathsf{h}_1[n] * x[n]) \cdot (\mathsf{h}_1[n-6] * x[n]) = \\ &= -.25\ \big(x[n-1] + x^2[n-1]\big) + .50\ \big(x[n] + x^2[n]\big) - .25\ \big(x[n+1] + x^2[n+1]\big) + \\ &\quad +.0625\ (x[n-1] \cdot x[n-7] + x[n+1] \cdot x[n-7] + x[n-1] \cdot x[n-5] + x[n+1] \cdot x[n-5]) - \\ &\quad -.125\ (x[n] \cdot x[n-7] + x[n-1] \cdot x[n-6] + x[n+1] \cdot x[n-6] + x[n] \cdot x[n-5]) + \\ &\quad +.25\ x[n] \cdot x[n-6] \end{aligned} \tag{71}$$

Table 5 Coefficients in $\mathsf{h}_2[n]$.

	1	2	3	4	5	6	7	8
0	-0.076	-0.075	-0.076	-0.076	-0.076	-0.075	-0.075	-0.074
8	-0.073	-0.071	-0.071	-0.070	-0.070	-0.071	-0.073	-0.075
16	-0.078	-0.081	-0.084	-0.086	-0.087	-0.087	-0.085	-0.083
24	-0.079	-0.075	-0.071	-0.068	-0.067	-0.067	-0.070	-0.074
32	-0.080	-0.086	-0.092	-0.097	-0.100	-0.100	-0.096	-0.090
40	-0.081	-0.072	-0.062	-0.055	-0.051	-0.052	-0.057	-0.068
48	-0.084	-0.101	-0.118	-0.133	-0.142	-0.143	-0.133	-0.112
56	-0.080	-0.038	0.011	0.064	0.118	0.167	0.208	0.239
64	0.256	0.258	0.247	0.224	0.192	0.155	0.115	0.077
72	0.045	0.020	0.004	-0.002	-0.001	0.008	0.021	0.036
80	0.051	0.064	0.072	0.076	0.076	0.071	0.064	0.055
88	0.046	0.038	0.032	0.029	0.029	0.032	0.036	0.042
96	0.047	0.052	0.055	0.057	0.057	0.055	0.051	0.048
104	0.044	0.040	0.037	0.036	0.036	0.037	0.038	0.040
112	0.043	0.045	0.047	0.048	0.049	0.049	0.048	0.047
120	0.046	0.046	0.045	0.044	0.044	0.044	0.044	0.044

B.2 Scenario 2

We used a long, discrete filter $\mathsf{h}_2[n]$ to emulate the memory effect in the continuous-time systems. The coefficients are listed in Table 5. The position on this table of the coefficient may be simply computed as $N_{\text{col.}} + \text{row}$. For instance, the coefficient in column 4 and row tagged 56 is the $4+56=60$-th coefficient.

B.2.1 Memoryless nonlinear system

The output for the memoryless structure may be computed as

$$y[n] = x[n] + x^2[n] + x^3[n] + x^4[n]. \tag{72}$$

B.2.2 Generalized Hammerstein system

For the Generalized Hammerstein structure, the output of the system was then computed as

$$y[n] = \mathsf{h}_2[n] * \left(x[n] + x^2[n] + x^3[n] + x^4[n]\right). \tag{73}$$

Alternatively, one may prefer to use some implementation of the nonlinear convolution. This would require the computation of Volterra kernels $h_i[n]$, which be zero outside their respective diagonals, and take the following values along the latter

$$\begin{aligned} h_1[n] &= \mathsf{h}_2[n] \\ h_2[n,n] &= \mathsf{h}_2[n] \\ h_3[n,n,n] &= \mathsf{h}_2[n] \\ h_4[n,n,n,n] &= \mathsf{h}_2[n] \end{aligned} \tag{74}$$

B.2.3 Volterra-Wiener system

For this we used an standard implementation of the nonlinear convolution, which requires the computation of the set of order-dependent Volterra kernels, $h_i[n]$. These we computed as follows

$$\begin{aligned} h_1[n] &= \mathsf{h}_2[n] \\ h_2[n_1,n_2] &= \mathsf{h}_2[n_1] \times \mathsf{h}_2[n_2] \\ h_3[n_1,n_2,n_3] &= \mathsf{h}_2[n_1] \times \mathsf{h}_2[n_2] \times \mathsf{h}_2[n_3] \\ h_4[n_1,n_2,n_3,n_4] &= \mathsf{h}_2[n_1] \times \mathsf{h}_2[n_2] \times \mathsf{h}_2[n_3] \times \mathsf{h}_2[n_4] \end{aligned} \tag{75}$$

where $\times$ stands for the cartesian product. Additionally, when evaluating the memory-length effect, we focused on the truncated kernels. This is, considering discrete filter $\mathsf{h}_2[n]$ of length $L = 128$, truncation to a sequence of length $L' < L$ requires that we operate on $\mathsf{h}_2[n'_i]$ instead, where

$$n_i = 1, \ldots, L \qquad \text{while} \qquad n'_i = 1, \ldots, L'. \tag{76}$$

References

Abramowitz M, Stegun IA (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth Dover printing, tenth GPO printing edn. Dover, New York, doi: 10.1119/1.1972842

Agrawal G (2007) *Nonlinear fiber optics*. Quantum electronics–principles and applications, Elsevier / Academic Press

Billings S (1980) Identification of nonlinear systems - A survey. *Control Theory and Applications, IEE Proceedings D* 127(6):272–285

Bose T, Mitra S (1992) System identification and filtering using pseudo random binary inputs. *Journal of the Franklin Institute* 329(4):765 – 774, doi: 10.1016/0016-0032(92)90087-W

Boyd S, Chua LO (1985) Fading memory and the problem of approximating nonlinear operators with Volterra series. *Circuits and Systems, IEEE Transactions on* 32(11):1150–1161, doi: 10.1109/TCS.1985.1085649

Casey SD, Walnut DF (1994) Systems of Convolution Equations, Deconvolution, Shannon Sampling, and the Wavelet and Gabor Transforms. *SIAM Review* 36(4):537–577, doi: 10.1137/1036140

Cheng C, Peng Z, Zhang W, Meng G (2017) Volterra-series-based nonlinear system modeling and its engineering applications: A state-of-the-art review. *Mechanical Systems and Signal Processing* 87:340–364, doi: 10.1016/j.ymssp.2016.10.029

Cohen L (2000) Instantaneous frequency and group delay of a filtered signal. *Journal of the Franklin Institute* 337(4):329 – 346, doi: 10.1016/S0016-0032(00)00039-9

Crespo-Cadenas C, Aguilera-Bonet P, Becerra-González J, Cruces S (2014) On nonlinear amplifier modeling and identification using baseband Volterra-PARAFAC models. *Signal Processing* 96, Part B(0):401 – 405, doi: 10.1016/j.sigpro.2013.09.028

Farina A (2000) Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. In: Audio Engineering Society Convention 108

van Hemmen JL, Kistler WM, Thomas EGF (2000) Calculation of Volterra Kernels for Solutions of Nonlinear Differential Equations. *SIAM Journal on Applied Mathematics* 61(1):1–21, doi: 10.1137/S0036139999336037

Jing X (2012) Truncation order and its effect in a class of nonlinear systems. *Automatica* 48(11):2978 – 2985, doi: 10.1016/j.automatica.2012.08.004

Lang ZQ, Billings S (1996) Output frequency characteristics of nonlinear systems. *International Journal of Control* 64(6):1049–1067, doi: 10.1080/00207179608921674

Larson DR, Massopust P, Nashed Z, Nguyen MC, Papadakis M, Zayed A (2008) *Frames and Operator Theory in Analysis and Signal Processing*. Contemporary Mathematics, American Mathematical Society

Lin JG (2006) Modeling test responses by multivariable polynomials of higher degrees. *SIAM Journal on Scientific Computing* 28(3):832–867, doi: 10.1137/040603954

Mann S, Haykin S (1995) The chirplet transform: physical considerations. *Signal Processing, IEEE Transactions on* 43(11):2745–2761, doi: 10.1109/78.482123

Novák A, Simon L, Kadlec F, Lotton P (2010) Nonlinear System Identification Using Exponential Swept-Sine Signal. *Instrumentation and Measurement, IEEE Transactions on* 59(8):2220–2229, doi: 10.1109/TIM.2009.2031836

Nowak R, Van Veen B (1994) Random and pseudorandom inputs for Volterra filter identification. *Signal Processing, IEEE Transactions on* 42(8):2124–2135, doi: 10.1109/78.301847

Orcioni S, Pirani M, Turchetti C (2005) Advances in Lee–Schetzen Method for Volterra Filter Identification. *Multidimensional Systems and Signal Processing* 16(3):265–284, doi: 10.1007/s11045-004-1677-7

Parente RB (1970) Nonlinear Differential Equations and Analytic System Theory. *SIAM Journal on Applied Mathematics* 18(1):41–66, doi: 10.1137/0118007

Rébillat M, Hennequin R, Corteel É, Katz BF (2011) Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices. *Journal of Sound and Vibration* 330(5):1018 – 1038

Rugh W (1981) *Nonlinear system theory: the Volterra/Wiener approach*. Johns Hopkins series in information sciences and systems, Johns Hopkins University Press

Sandberg IW (1990) g-representations and differential equations. *Multidimensional Systems and Signal Processing* 1(2):217–233, doi: 10.1007/BF01816550

Schetzen M (1980) *The Volterra and Wiener Theories of Nonlinear Systems*. John Wiley & Sons

Schetzen M (1985) Multilinear theory of nonlinear networks. *Journal of the Franklin Institute* 320(5):221 – 247, doi: 10.1016/0016-0032(85)90023-7

Sharma SN (2010) A connection between multi-linear and Volterra systems. *Applied Mathematics and Computation* 216(7):1918–1922, doi: 10.1016/j.amc.2010.01.120

Soumitro Banerjee GCV (2001) *Nonlinear Phenomena in Power Electronics: Bifurcations, Chaos, Control, and Applications*. Wiley-IEEE Press

Zang G, Iglesias PA (2003) Fading memory and stability. *Journal of the Franklin Institute* 340(6-7):489–502, doi: 10.1016/j.jfranklin.2003.11.002