A Polynomial Multiple Variance Method for Volterra Filter Identification

Alberto Carini, *Senior Member, IEEE*, Riccardo Forti, and Simone Orcioni, *Senior Member, IEEE*

***Abstract*— The article presents a methodology for accurately estimating the Volterra kernels of a discrete-time nonlinear system, even when the system order exceeds that of the Volterra series model. The approach involves conducting multiple measurements with the same excitation signal, scaled by different gain factors, and deriving the Volterra kernels via interpolation of the measured data. The methodology is thoroughly discussed, and the mean square deviation (MSD) of the estimated coefficients is calculated to determine the optimal gain factors that minimize the MSD. It is demonstrated that the optimal gains are constrained to a specific set of values, which are provided in the article. Experimental results, using both synthetic and real systems, showcase the effectiveness of the proposed methodology.**

***Index Terms*— Multiple variance method, orthogonal periodic sequences, Volterra filter.**

I. Introduction

VOLTERRA filters are among the most popular models used for identifying nonlinear systems [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]. Their success arises from the simplicity of the input–output relationship, which results from truncating the Volterra series, as they are a sum of homogeneous polynomial operators on the input signal samples. Moreover, Volterra filters exhibit the characteristic of being universal approximators. Based on the Stone–Weierstrass theorem, it is established that any discrete-time, time-invariant, continuous nonlinear system with finite memory can be approximated to any desired degree of accuracy by employing a Volterra filter.

Despite their widespread use in modeling real-world systems, the identification of Volterra filters often involves difficult choices and significant approximations. To identify a Volterra filter, one must select the memory length of the filter and polynomial order, based on the characteristics of the real-world system being modeled. In particular, the polynomial order strongly depends on the input signal power. For very low input powers, nonlinear terms are often negligible, allowing the system to be modeled as linear. However, as input power increases, the nonlinear terms generally become significant, and the model order should increase with the maximum amplitude of the input signal. In fact, as the input signal amplitude grows, higher harmonics of the output signal are excited. Practically, due to the curse of dimensionality, the number of coefficients in the Volterra filter increases geometrically with the filter memory and exponentially with the filter order. For this reason, Volterra models of order greater than 3 are rarely encountered.

The input signal power also influences the accuracy of estimating the Volterra filter coefficients, known as Volterra kernels. As shown in [19], to accurately estimate lower order kernels, the input power must be low enough to mitigate the effects of unmodeled higher order terms, yet high enough to counteract the impact of noise. Conversely, to estimate higher order kernels (e.g., the third-order kernel), a higher input amplitude is needed to properly excite the kernel, but it must still be sufficiently low to minimize the influence of unaccounted higher order kernels (e.g., those of order greater than 3). This highlights the inherent compromise in modeling a nonlinear system, where the resulting model often suffers from a "locality" issue: the model accurately represents the behavior of the nonlinear system only for input powers close to those used during identification. To address this issue, a multiple-variance method was proposed in [19]. In particular, Orcioni investigated the identification of a Wiener model utilizing the Lee–Schetzen approach. To enhance the generalization of the model, various kernels were identified by employing input signals with different variances, aiming to minimize the mean-square error (mse) between the system's output and that of the model.

In this article, a multiple-variance approach is also proposed, but with objectives and methodology that differ entirely from those in [19]. Instead of minimizing the mse between the system and model outputs, the goal is to accurately measure specific kernels of the Volterra model, even when the real-world system exhibits nonlinear terms beyond those considered in the model. The core idea is to take multiple measurements of the model at different input power levels—or to use different input gains when the same input is employed for identification—and then estimate the kernels by performing polynomial interpolation on the coefficients corresponding to these different input gains or power levels. This approach, originally proposed in [20] and [21], has been used for system identification, primarily in the frequency domain [22], [23], [24], [25], [26], though it has not been extensively applied. The motivation stems from the fact that the multiple gains required by the method "must be carefully selected" [27] to achieve optimal results. In this article, this issue is specifi-

Received 23 September 2024; revised 16 December 2024; accepted 26 December 2024. Date of publication 3 March 2025; date of current version 13 March 2025. This work was supported in part by European Union-Next Generation EU through Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, under Grant CUP E63C22002070006; and in part by the Telecommunications of the Future through the RESTART under Grant PE00000001. The Associate Editor coordinating the review process was Dr. Jing Yuan. *(Corresponding author: Alberto Carini.)*

Alberto Carini and Riccardo Forti are with the Department of Engineering and Architecture, University of Trieste, 34127 Trieste, Italy (e-mail: acarini@units.it; riccardo.forti@phd.units.it).

Simone Orcioni is with the Department of Information Engineering, Università Politecnica delle Marche, 60131 Ancona, Italy (e-mail: s.orcioni@staff.univpm.it).

Digital Object Identifier 10.1109/TIM.2025.3546397

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

cally addressed. After outlining the technique, the manuscript examines the effect of noise and the mean square deviation (MSD) of the measured coefficients. An optimal set of gains is developed by minimizing the MSD. Interestingly, this article shows that the optimal gains consist of a limited set of values, which depend on the polynomial order and the number of measurements performed. Since good measurement accuracy and effective rejection of higher order kernel disturbances can also be achieved by averaging multiple measurements using a single low input gain, the manuscript discusses when the proposed technique is advantageous compared to this computational approach.

Any of the linear-in-the-output identification methods for Volterra filters available in the literature can be used for implementing the following technique. The method used in this article is the identification based on orthogonal periodic sequences (OPSs) [28]. OPSs offer one of the most efficient identification methods for Volterra models through the cross correlation method, which involves computing the kernels by cross-correlating the output of a system, excited by a deterministic periodic input signal, with precalculated orthogonal periodic sequences. The cross correlation can be efficiently computed using the fast Fourier transform (FFT), resulting in the OPS identification having a computational cost on the order of $P\log_2(P)$, where P represents the period of the OPS.

The approach discussed in this manuscript and its analysis was previously introduced by the authors in [29] and [30], but only for the identification of the first-order kernel, i.e., for measuring the impulse response of a nonlinear system at small signal amplitudes. In this manuscript, the approach is generalized to higher order kernels, a nontrivial extension. There is, inevitably, some overlap with the theory presented in [29] and [30], as certain equations take a similar form, but in this work, they actually depend on higher order polynomial terms. Despite the similarities, there are also significant differences: the optimal gains for measuring higher-order kernels differ significantly from those for the first-order kernel and vary notably for even-order kernels. The optimal gains for simultaneously measuring multiple kernels are also determined. Furthermore, the conditions under which the proposed technique outperforms averaging multiple measurements with a single low gain are distinct. The experimental tests considered here differ significantly from those in [29] and [30].

The original contributions of this article are as follows.

1) A technique is proposed for accurately measuring the kernels of a Volterra model, even when the order of the model is lower than that of the real-world system being measured.
2) The technique applies a multiple-variance approach, based on taking multiple measurements with different input gains for the Volterra kernels.
3) The MSD of the measured kernels is analyzed and linked to the selection of input gains.
4) The optimal input gains for minimizing the MSD when measuring different kernels are derived.
5) The article discusses the conditions under which the proposed technique is more advantageous than averaging multiple measurements with a single low input gain.

The remainder of the article is organized as follows. Section II reviews Volterra filters and their key properties used in this manuscript. Section III discusses the considered polynomial multiple-variance method and examines the effect of noise on the measurements, providing an expression for the kernel MSD. Section IV derives the optimal gains for the different kernels, demonstrating that these gains take on a limited set of values. Section V offers remarks on the proposed method, highlighting when it is advantageous compared to averaging multiple measurements taken with a single low input gain. Section VI presents experimental results, including both simulations using a synthetic system with known kernels and measurements performed on a real-world device. Finally, Section VII offers concluding remarks.

The following notation is used throughout the article: calligraphic letters represent operators, $E[\cdot]$ denotes expectation, $\lceil\cdot\rceil$ represents the ceiling function (the smallest integer greater than or equal to the argument) and $\lfloor\cdot\rfloor$ the floor function (the largest integer lower than or equal to the argument), bold lowercase letters denote arrays, and bold uppercase letters denote matrices.

II. Volterra Filters

Volterra filters, which are polynomial, are derived by truncating the order and memory of the Volterra series [1]. According to the Stone–Weierstrass theorem, these filters have the ability to approximate, with arbitrary precision, any causal, discrete-time, time-invariant, continuous nonlinear system with finite memory, where the input–output relationship is described by a nonlinear function f of the most recent input samples

$$y(n) = f[x(n), x(n-1), \ldots, x(n-N+1)]. \tag{1}$$

Here, $x(n)$ represents the nth sample of the input signal x, which lies within a compact subset of $\mathbb{R}$, while $y(n)$ denotes the nth sample of the output signal y. N corresponds to the memory length of the filter.

The input–output relationship of a Volterra filter of order K and memory N is

$$y(n) = \sum_{k=0}^{K} \mathcal{H}_k(x)(n). \tag{2}$$

Here, $\mathcal{H}_k(x)$ represents a homogeneous polynomial operator of order k and memory N applied to the input signal x, while $\mathcal{H}_k(x)(n)$ denotes its nth sample. In particular, the 0th-order operator $\mathcal{H}_0(x)(n)$ is a constant $h_0 \in \mathbb{R}$ for all n. $\mathcal{H}_1(x)$ is the linear operator

$$\mathcal{H}_1(x)(n) = \sum_{i=0}^{N-1} h_{1,i}\, x(n-i) \tag{3}$$

where $h_{1,i}$ are the coefficients of the first-order Volterra kernel, i.e., the linear kernel. $\mathcal{H}_2(x)$ is a quadratic operator, and $\mathcal{H}_k(x)$ is a kth-order operator, expressed in the triangular form as follows [1]:

$$\mathcal{H}_k(x)(n) = \sum_{i_1=0}^{N-1} \sum_{i_2=i_1}^{N-1} \cdots \sum_{i_k=i_{k-1}}^{N-1} h_{k,i_1,\ldots,i_k}\, x(n-i_1)\cdots x(n-i_k) \tag{4}$$

where $h_{k,i_1,\ldots,i_k}$ denote the coefficients of the kth-order Volterra kernel.

For simplicity, the coefficients of kth-order Volterra kernel will be denoted as $h_{k,i}$, with $0 < i < N_k$, with N_k the number of coefficients of the kth kernel.

This assumes that all coefficients have been organized in a sequential list.

The approach detailed in Section III is based on the homogeneity property of the operator $\mathcal{H}_k(x)$. Specifically, if the input is scaled by a gain A:

$$\mathcal{H}_k(Ax) = A^k \mathcal{H}_k(x). \tag{5}$$

Furthermore, we leverage the observation that odd-order homogeneous operators are odd, while even-order ones are even.

III. Polynomial Multiple Variance Method

In this section, we initially present the considered multiple-variance method and subsequently investigate the impact of noise on measurements.

A. Multiple Variance Method

Consider dealing with a nonlinear system that can be represented as a Volterra filter with order K, memory N, plus noise

$$y = \mathcal{H}(x) + \nu = \sum_{k=0}^{K} \mathcal{H}_k(x) + \nu. \tag{6}$$

Here, $\mathcal{H}$ represents the Volterra operator as defined in (2), ν denotes an additive noise signal in the output, and the time notation (n) is omitted for conciseness. Our goal is to determine the coefficients of kernels up to a certain order R, where $R \leq K$.

The proposed technique is applicable to any Volterra identification approach functioning as a linear operator from the space of output sequences to $\mathbb{R}$. An illustrative example is the OPS technique employed in the experimental results, where each Volterra coefficient can be derived from the cross correlation of the output sequence and a suitable orthogonal sequence. Another instance is the classical least-square approach, in which the solution of the equation providing the coefficients is a linear function of the output sequence. In the following discussion, we assume the existence of a linear operator $\mathcal{L}_{r,i}$ for each coefficient of order r and index i from the space of sequences to $\mathbb{R}$, such that:

$$\mathcal{L}_{r,i}[\mathcal{H}_r(x)] = h_{r,i}. \tag{7}$$

In other words, applying $\mathcal{L}_{r,i}$ to the rth-order term $\mathcal{H}_r(x)$ yields the coefficient $h_{r,i}$.

Any Volterra identification approach that is a linear function of the output sequence generally furnishes a linear operator $\widehat{\mathcal{L}}_{r,i}$. For a Volterra filter with order R, memory N, and in noise absence, this operator satisfies

$$\widehat{\mathcal{L}}_{r,i}[\mathcal{H}(x)] = \widehat{\mathcal{L}}_{r,i}[y] = h_{r,i}. \tag{8}$$

The operator in (8) also satisfies (7) because $\mathcal{H}_r(x)$ is a Volterra filter of order no greater than R. However, it is important to note that the condition specified in (7) is less restrictive than that in (8).

For compactness, in what follows $\mathcal{L}_{r,i}[\mathcal{H}(x)]$ is abbreviated as $\mathcal{L}_{r,i}\mathcal{H}(x)$, taking advantage of the concatenation of the operators.

Let us apply $\mathcal{L}_{r,i}$ to (6), considering that it is a linear operator

$$\mathcal{L}_{r,i}(y) = \mathcal{L}_{r,i}\mathcal{H}(x) + \mathcal{L}_{r,i}(\nu) = \sum_{k=0}^{K} \mathcal{L}_{r,i}\mathcal{H}_k(x) + \mathcal{L}_{r,i}(\nu). \tag{9}$$

In these circumstances, (9) does not equate to $h_{r,i}$, not only due to the presence of noise ν but more significantly, because of the presence of nonlinear terms $\mathcal{H}_k(x)$ with $k \neq r$.

In what follows we present a multiple-variance identification method capable of estimating $h_{r,i}$, for all $r \in [0, R]$ and for all i. This approach avoids the impact of kernels with orders exceeding R and contrasts the influence of noise. In the considered approach, the input signal x undergoes multiple applications, each time being scaled by different factors A_m. The resultant output signals y_m are used for the estimation of $h_{r,i}$. When considering the homogeneity property in (5), if the input x is multiplied by A_m, the resulting output y_m can be expressed as

$$y_m = \mathcal{H}(A_m x) + \nu_m = \sum_{k=0}^{K} A_m^k \mathcal{H}_k(x) + \nu_m \tag{10}$$

where ν_m is the additive output noise of y_m. Applying the linear operator $\mathcal{L}_{r,i}$ to (10), we derive

$$\mathcal{L}_{r,i}(y_m) = \sum_{k=0}^{K} A_m^k \mathcal{L}_{r,i}\mathcal{H}_k(x) + \mathcal{L}_{r,i}(\nu_m) \tag{11}$$

which is a polynomial in A_m with coefficients $\mathcal{L}_{r,i}\mathcal{H}_k(x)$, excluding the noise term. By employing a sufficiently large set of gains A_m, it becomes feasible to estimate the terms $\mathcal{L}_{r,i}\mathcal{H}_k(x)$ for $k = 0, \ldots, K$ through polynomial fitting. Consequently, this estimation process allows us to measure $h_{r,i}$ based on the estimate of $\mathcal{L}_{r,i}\mathcal{H}_r(x)$.

In the subsequent discussion, we assume that the input signal x is applied M times with gains A_m, and we define the $M \times 1$ vectors

$$\boldsymbol{d}_{\mathcal{L}_{r,i}} = \left[\mathcal{L}_{r,i}(y_1), \mathcal{L}_{r,i}(y_2), \ldots, \mathcal{L}_{r,i}(y_M)\right]^{\mathrm{T}} \tag{12}$$

$$\boldsymbol{v}_{\mathcal{L}_{r,i}} = \left[\mathcal{L}_{r,i}(\nu_1), \mathcal{L}_{r,i}(\nu_2), \ldots, \mathcal{L}_{r,i}(\nu_M)\right]^{\mathrm{T}} \tag{13}$$

the $K \times 1$ vector

$$\boldsymbol{h}_{\mathcal{L}_{r,i}} = \left[\mathcal{L}_{r,i}\mathcal{H}_1(x), \mathcal{L}_{r,i}\mathcal{H}_2(x), \ldots, \mathcal{L}_{r,i}\mathcal{H}_K(x)\right]^{\mathrm{T}} \tag{14}$$

and the $(K+1) \times M$ Vandermonde-like matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ A_1 & A_2 & \ldots & A_M \\ A_1^2 & A_2^2 & \ldots & A_M^2 \\ \vdots & & & \vdots \\ A_1^K & A_2^K & \ldots & A_M^K \end{bmatrix}. \tag{15}$$

Expressing (11) for $m = 1, \ldots, M$ in matrix form yields

$$\boldsymbol{d}_{\mathcal{L}_{r,i}} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{h}_{\mathcal{L}_{r,i}} + \boldsymbol{v}_{\mathcal{L}_{r,i}}. \tag{16}$$

If $M \geq K+1$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is invertible

$$\boldsymbol{h}_{\mathcal{L}_{r,i}} = \left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{d}_{\mathcal{L}_{r,i}} - \left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{v}_{\mathcal{L}_{r,i}}. \tag{17}$$

Disregarding the impact of noise, $\boldsymbol{h}$ can be estimated using

$$\widehat{\boldsymbol{h}} = \left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{d}_{\mathcal{L}_{r,i}}. \tag{18}$$

Consequently, in accordance with (7), $h_{r,i}$ can be determined as

$$\widehat{h}_{r,i} = \boldsymbol{e}_r^{\mathrm{T}}\boldsymbol{h} = \boldsymbol{e}_r^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{d}_{\mathcal{L}_{r,i}} \tag{19}$$

with $\boldsymbol{e}_r$ the rth column of the $(K+1)\times(K+1)$ identity matrix, with $r \in [0, K]$.

This result shows that, for any Volterra identification approach functioning as a linear operator, removing the distortion given by higher order kernels only requires some repeated measures with different gains.

B. Effect of Noise on the Measurement

Upon comparing (17) with (18), it can be noted that in the absence of noise, i.e., for $\boldsymbol{v} = 0$, the measurement in (19) accurately estimates $h_{r,i}$. However, in the presence of noise ($\boldsymbol{v} \neq 0$), the measurement is subject to an error

$$\epsilon_{r,i} = \widehat{h}_{r,i} - h_{r,i} = \boldsymbol{e}_r^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{v}_{\mathcal{L}_{r,i}}. \tag{20}$$

The MSD of the estimated coefficient is

$$\begin{aligned}\mathrm{MSD}_{r,i} &= E\left[\left(\widehat{h}_{r,i} - h_{r,i}\right)^2\right] \\ &= \boldsymbol{e}_r^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}E\left[\boldsymbol{v}_{\mathcal{L}_{r,i}}\boldsymbol{v}_{\mathcal{L}_{r,i}}^{\mathrm{T}}\right]\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{e}_r.\end{aligned} \tag{21}$$

Under the assumption that the noise terms $\mathcal{L}_{r,i}(v_l)$, for $l = 1, \ldots, M$ are uncorrelated, and that they follow a Gaussian distribution with zero mean and variance $\sigma^2_{\mathcal{L}_{r,i}v}$, the MSD simplifies to

$$\mathrm{MSD}_{r,i} = \boldsymbol{e}_r^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{e}_r\sigma^2_{\mathcal{L}_{r,i}v}. \tag{22}$$

Assume we estimate all the coefficients of the Volterra model for all $r \in [0, R]$ and all i. The resulting MSD of the estimated coefficients is

$$\begin{aligned}\mathrm{MSD}_{\mathrm{T}} &= E\left[\sum_{r=0}^{R}\sum_{i=0}^{N_r-1}\left(\widehat{h}_{r,i} - h_{r,i}\right)^2\right] \\ &= \sum_{r=0}^{R}\sum_{i=0}^{N_r-1}\mathrm{MSD}_{r,i} \\ &= \sum_{r=0}^{R}\sum_{i=0}^{N_r-1}\boldsymbol{e}_r^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{e}_r\sigma^2_{\mathcal{L}_{r,i}v} \\ &= \mathrm{diag}\left[\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\right]^{\mathrm{T}} \\ &\quad\cdot\left[\sigma^2_{\mathcal{L}_{0,0}v}, \sum_{i=0}^{N-1}\sigma^2_{\mathcal{L}_{1,i}v}, \ldots, \sum_{i=0}^{N_R-1}\sigma^2_{\mathcal{L}_{R,i}v}\right]^{\mathrm{T}}\end{aligned} \tag{23}$$

where $\mathrm{diag}(\boldsymbol{X})$ denotes the operator that extracts the diagonal of the matrix $\boldsymbol{X}$ and transforms it into a column vector.

IV. Optimal Gains

As can be seen from (15)–(23), the choice of gains A_m has a direct impact on the resulting $\mathrm{MSD}_{r,i}$ and $\mathrm{MSD}_{\mathrm{T}}$. Therefore, there exists an optimal set of gains that minimizes these values. Achieving the best performance involves choosing an optimal set of coefficients for each $h_{r,i}$, i.e., selecting gains A_m that minimize $\mathrm{MSD}_{r,i}$. However, this choice is impractical as it requires a large number of measurements. A more reasonable approach is to use the same set of gains A_m for estimating all coefficients and finding the set of optimal gains A_m that minimize $\mathrm{MSD}_{\mathrm{T}}$. For applications requiring greater accuracy, it is also possible to use different sets of gains A_m for each kernel, finding the optimal gains that minimize $\sum_{i=0}^{N_r-1}\mathrm{MSD}_{r,i}$. This approach may be acceptable since the maximum order of the Volterra model R is often limited to 3.

Different approaches can be used to choose the gains A_m, such as symmetric or asymmetric distributions around zero. A symmetric distribution provides several benefits: it reduces the condition number of the Vandermonde matrix associated with the nodes used for polynomial fitting [31], [32]. Furthermore, it simplifies the estimation process by removing either all even-order or odd-order nonlinear terms, as shown in Appendix A. Thus, in the following analysis, we assume that the input signal x is applied M times, with M being even, and the relationship $A_{m+M/2} = -A_m$ holds for $m = 1, \ldots, M/2$.

It is assumed that the additive noise v in (6) follows a Gaussian distribution with zero mean and variance σ_v^2. For a nonlinear model of order K, with M positive gains and the maximum gain constrained to 1, the optimal gains A_m are determined by minimizing the nonlinear function:

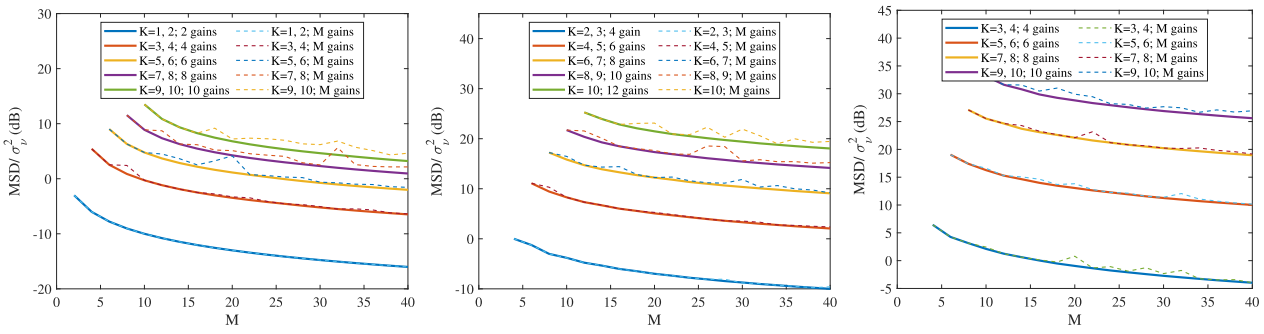
$$\begin{aligned}J(A_1, \ldots, A_M) &= \frac{\mathrm{MSD}_{\mathrm{T}}}{\sigma_v^2} \\ &= \mathrm{diag}\left[\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\right]^{\mathrm{T}} \\ &\quad\cdot\left[\frac{\sigma^2_{\mathcal{L}_{0,0}v}}{\sigma_v^2}, \sum_{i=0}^{N_1-1}\frac{\sigma^2_{\mathcal{L}_{1,i}v}}{\sigma_v^2}, \ldots, \sum_{i=0}^{N_R-1}\frac{\sigma^2_{\mathcal{L}_{R,i}v}}{\sigma_v^2}\right]^{\mathrm{T}}\end{aligned} \tag{24}$$

subject to $0 < A_m \leq 1$ for $m = 1, \ldots, (M/(2-1))$ and $A_{M/2} = 1$. For a particular identification method, which is linear in the output sequence y, the values of $(\sigma^2_{\mathcal{L}r,iv})/(\sigma_v^2)$ for $r \in [0, R]$ are constants. If we assume these constants to be equal, the problem would simplify to minimizing the function

$$\hat{J}(A_1, \ldots, A_M) = \mathrm{diag}\left[\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\right]^{\mathrm{T}}\cdot[1, N_1, \ldots, N_R]^{\mathrm{T}}. \tag{25}$$

Actually, the quantities $(\sigma^2_{\mathcal{L}r,iv})/(\sigma_v^2)$ often exhibit variations among themselves, but they tend to have similar magnitudes or magnitudes that increase with the order, such as in the case of OPSs. Given that the number of coefficients N_r expands with the order r, it is reasonable to anticipate that the solution will closely approximate the solution of the reduced problem

$$J_R(A_1, \ldots, A_M) = \sum_{i=0}^{N_R-1}\frac{\mathrm{MSD}_{R,i}}{\sigma_v^2}$$



(a) (b) (c)

Fig. 1. Minimum value of the cost function in (27) in case of (a) $R = 1$, (b) $R = 2$, and (c) $R = 3$ for different orders K versus the number of gains M. The continuous lines are the results for the gains constrained to $\lceil K/2 \rceil$ gains, while the dashed lines are the results for M distinct gains.

$$= \boldsymbol{e}_R^{\mathrm{T}} \left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1} \boldsymbol{e}_R \sum_{i=0}^{N_R-1} \frac{\sigma_{\mathcal{L}_{R,i}v}}{\sigma_v^2}. \tag{26}$$

Furthermore, as $(\sigma_{\mathcal{L}_{R,i}v})/(\sigma_v^2)$ remains approximately constant for all i, the optimal coefficients are those that minimize the reduced cost function

$$\hat{J}_R(A_1, \ldots, A_M) = \boldsymbol{e}_R^{\mathrm{T}} \left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1} \boldsymbol{e}_R. \tag{27}$$

The cost function in (27) can be further simplified by leveraging the symmetry of the gains, as detailed in Appendix B. In this scenario, the same cost function is obtained for two consecutive orders, and therefore, the same set of gains is optimal for both odd and even consecutive orders.

Any method designed for minimizing multivariable nonlinear functions can be employed to optimize $\hat{J}_R$. For instance, MATLAB[1] provides optimization routines like `fminsearch` and `fmincon` that are particularly suited for this purpose. From preliminary trials concerning the minimization of (27), it became evident that the optimal gains tend to group around a few distinct values. Specifically, when K is odd, the positive optimal gains tend to cluster around exactly $\lceil K/2 \rceil$ values. Conversely, when K is even, the gains cluster around $K/2 + 1$ values, with A_1 close to zero. Moreover, it was observed that by constraining the $M/2$ gains A_m to take at most $\lceil K/2 \rceil$ distinct values for odd K, or $K/2 + 1$ values, with $A_1 = 0$ for even K, the minimum value of J obtained was either equal to or lower than the value found when all M gains were considered independently. Fig. 1 illustrates the minimum of the cost function (27) for $R = 1, 2$, and 3, across various orders K and gains M. Both cases, involving a reduced set of gains (either $\lceil K/2 \rceil$ or $K/2+1$) and M distinct symmetric gains, are presented. It is important to mention that the curves corresponding to the case of distinct gains M were generated by running the optimization algorithm 10 000 times for each combination of R, K, and M, with random initialization of gains in the interval $[0, 1]$, and selecting the smallest outcome. For smaller K and M, the continuous and dashed curves coincide. However, for larger values of K and M, the optimization process using M different values consistently leads to suboptimal solutions. The results depicted in Fig. 1 suggest that a reduced number of gains suffices for the optimization of (27). When $M/2$ exceeds the number of distinct positive gains, some of the gains are repeated, meaning that multiple measurements must be carried out using the same gain value.

The optimal positive gains for $R = 1$, 2, and 3, with $K = 3$, 5, 7, and 9, and M values ranging from 2 to 100, are shown in Fig. 2. In the figure, asterisks indicate gains that are used more than once, while circles represent gains that are used only once. Repeated gains tend to favor lower values, which is expected since measurements at lower gains are generally more sensitive to noise. It is important to highlight that both the continuous curves in Fig. 1 and the optimal gains in Fig. 2 were obtained by optimizing (27) for all potential combinations of repeated gains, and selecting the best solution.

For large M values, the continuous curves in Fig. 1 exhibit a consistent 3 dB reduction for each doubling of M. This implies that the MSD decreases by 3 dB with every twofold increase in M. This behavior mirrors the effect commonly seen in many identification techniques, where increasing the data length or support results in similar reductions in MSD.

To further illustrate this trend, Fig. 3 presents the data with a logarithmic scale applied to the X-axis, where M ranges from 1 to 100. The dashed lines in the figure represent the 3 dB slope support lines, which align with the continuous curves for sufficiently large M. This observation motivates a criterion for selecting the optimal M value [30].

> For each combination of R and K, choose the M value that closely follows the 3 dB slope, apart from a small dB deviation.

This criterion is based on the idea that increasing M beyond this threshold offers diminishing returns in MSD improvement and the same benefit can be achieved by increasing the length or duration of the sequence used in the identification procedure.

Table I presents the optimal value of M for each R and K based on this criterion, along with the corresponding optimal positive gains A_m and their multiplicity, assuming a 1 dB deviation from the 3 dB slope line. For each R and K, the optimal value of M presented in the Table has been determined by minimizing the cost function (27) for different successive values of M and selecting the value that results in a 1 dB deviation from the 3 dB slope line. Similarly, Table II provides this information for a 0.5 dB deviation. It is important to note that, due to the assumption of a symmetric gain distribution,

[1]Registered trademark.

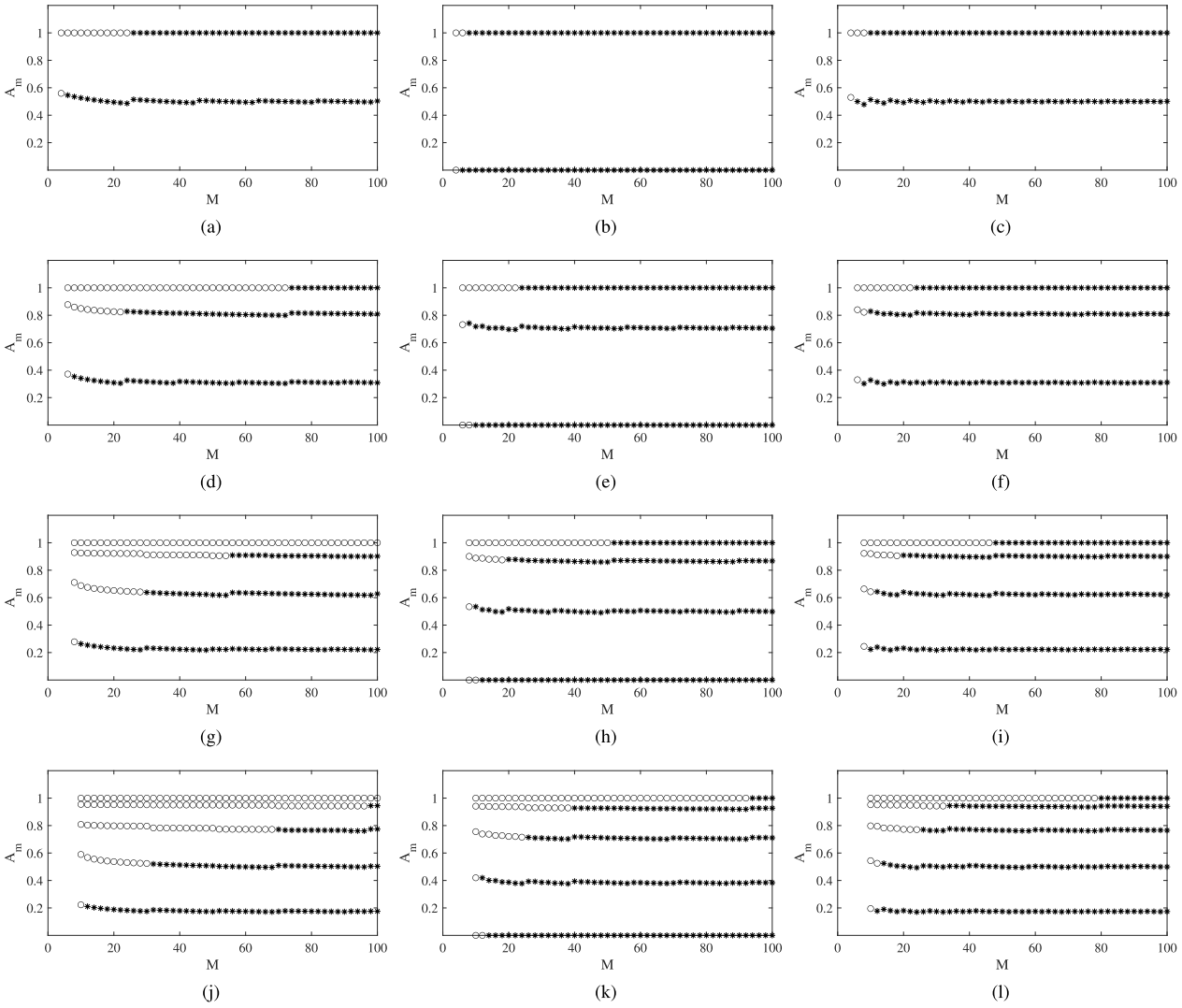


Fig. 2. Optimal value of the positive gains for different R and K versus the number of gains M. (a) $R = 1, K = 3$. (b) $R = 2, K = 3$. (c) $R = 3, K = 3$. (d) $R = 1, K = 5$. (e) $R = 2, K = 5$. (f) $R = 3, K = 5$. (g) $R = 1, K = 7$. (h) $R = 2, K = 7$. (i) $R = 3, K = 7$. (j) $R = 1, K = 9$. (k) $R = 2, K = 9$. (l) $R = 3, K = 9$.

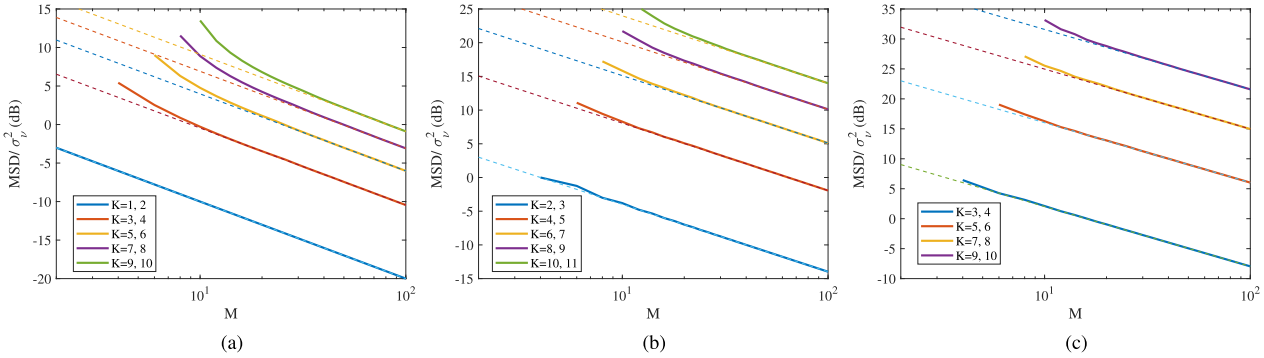


Fig. 3. Minimum value of the cost function in (27) in case of (a) $R = 1$, (b) $R = 2$, and (c) $R = 3$ for different orders K versus the number of gains M in logarithm scale.

when $A_m = 0.0$, the listed multiplicity represents only half of the actual value. The parameter $\widehat{A}$, which appears in Tables I and II, will be defined in the following section.

V. Discussion

The proposed polynomial multiple-variance (PMV) method is designed to measure a Volterra kernel of order $r \leq R$ or a set of Volterra kernels up to a specified maximum order R for a device under test (DUT). This measurement is based on a linear-in-the-output operator that can perfectly estimate the kernel in an ideal scenario—where there is no noise and no nonlinearities of order higher than R.

However, in practical situations, the measurement is often influenced by higher-order nonlinearities up to a certain order $K > R$, as well as by noise—both of which are common. To enhance the signal-to-noise ratio (SNR) and improve

TABLE I

OPTIMAL VALUES OF M AND THE CORRESPONDING POSITIVE GAINS A_m, ALONG WITH THEIR MULTIPLICITIES, FOR A 1 dB DEVIATION FROM THE 3 dB SLOPE LINE FOR $R = 1, 2$, AND 3

R	K	M	A_m	Multiplicy	$\widehat{A}$
1	1, 2	2	1.0000	1	1.0
1	3, 4	6	0.5459	2	0.3049
			1.0000	1	
1	5, 6	10	0.3411	3	0.1827
			0.8491	1	
			1.0000	1	
1	7, 8	14	0.2470	4	0.1295
			0.6670	1	
			0.9235	1	
			1.0000	1	
1	9, 10	18	0.1922	5	0.0998
			0.5419	1	
			0.7989	1	
			0.9523	1	
			1.0000	1	
2	2, 3	4	0.0	1	0.7071
			1.0	1	
2	4, 5	6	0.0	1	0.3374
			0.7319	1	
			1.0	1	
2	6, 7	10	0.0	1	0.2180
			0.5342	2	
			0.8881	1	
			1.0	1	
2	8, 9	12	0.0	1	0.1608
			0.4190	2	
			0.7385	1	
			0.9395	1	
			1.0	1	
2	10, 11	16	0.0	2	0.1342
			0.3288	2	
			0.6234	1	
			0.8325	1	
			0.9605	1	
			1.0	1	
3	3, 4	4	0.5299	1	0.6198
			1.0000	1	
3	5, 6	6	0.3302	1	0.3572
			0.8403	1	
			1.0000	1	
3	7, 8	10	0.2237	2	0.2557
			0.6431	1	
			0.9208	1	
			1.0000	1	
3	9, 10	12	0.1778	2	0.1964
			0.5250	1	
			0.7949	1	
			0.9514	1	
			1.0000	1	

TABLE II

OPTIMAL VALUES OF M AND THE CORRESPONDING POSITIVE GAINS A_m, ALONG WITH THEIR MULTIPLICITIES, FOR A 0.5 dB DEVIATION FROM THE 3 dB SLOPE LINE FOR $R = 1, 2$, AND 3

R	K	M	A_m	Multiplicy	$\widehat{A}$
1	1, 2	2	1.0000	1	1.0
1	3, 4	8	0.5351	3	0.3197
			1.0000	1	
1	5, 6	12	0.3320	4	0.1889
			0.8421	1	
			1.0000	1	
1	7, 8	18	0.2366	6	0.1355
			0.6560	1	
			0.9226	1	
			1.0000	1	
1	9, 10	24	0.1823	8	0.1050
			0.5307	1	
			0.7971	1	
			0.9518	1	
			1.0000	1	
2	2, 3	4	0.0	1	0.7071
			1.0	1	
2	4, 5	8	0.0	1	0.3374
			0.7415	2	
			1.0	1	
2	6, 7	12	0.0	2	0.2180
			0.5127	2	
			0.8872	1	
			1.0	1	
2	8, 9	16	0.0	2	0.1608
			0.3986	3	
			0.7277	1	
			0.9382	1	
			1.0	1	
2	10, 11	20	0.0	3	0.1342
			0.3186	3	
			0.6132	1	
			0.8305	1	
			0.9600	1	
			1.0	1	
3	3, 4	4	0.5299	1	0.6198
			1.0000	1	
3	5, 6	8	0.3022	2	0.3624
			0.8220	1	
			1.0000	1	
3	7, 8	12	0.2396	2	0.2566
			0.6432	2	
			0.9120	1	
			1.0000	1	
3	9, 10	16	0.1804	3	0.2001
			0.5131	2	
			0.7804	1	
			0.9497	1	
			1.0000	1	

measurement accuracy, the input signal amplitude is frequently amplified. This, in turn, introduces nonlinear distortions from orders higher than those included in the model.

The proposed solution is to perform the same measurement multiple times using different input gain levels. For a given kernel order R and a DUT nonlinearity of maximum order K, we have identified the optimal input gains that minimize the MSD. The high-precision Volterra kernels are obtained through interpolation of the measured data. This interpolation effectively translates into a weighted average of the measured data and has a negligible computational cost. The primary cost of the proposed method, in terms of both computational complexity and measurement time, lies in performing the same measurement multiple times.

The proposed approach addresses a classic problem in single-gain nonlinear identification: which gain should be selected for the input signal? This choice is often made through a trial-and-error process by observing the mse performance on a test signal. However, a low mse frequently does not correspond to a low MSD for the kernels, as the influence of neglected higher order kernels affects the results. In contrast, the proposed PMV method provides a straightforward way to determine the optimal gains required to achieve a good performance.

Does the proposed strategy always provide the lowest measurement error, i.e., the lowest MSD, for repeated measurements with the same amount of recorded data? In general, performance depends on the DUT and the measurement conditions. For example, if the noise is very high, such that it obscures the effect of any nonlinear kernel of order greater than R, then applying the proposed strategy is not beneficial. In such cases, it is better to perform repeated measurements

and average the results. Note that under these conditions, the assumption that the kernel of order greater than R affects the measurement does not hold.

Another strategy often followed in common practice involves measuring the DUT with an input gain A sufficiently low such that the effect of any kernel of order greater than R can be neglected. In the absence of noise, this strategy provides the desired kernels with a single measurement, making the application of the PMV method unnecessary. In the presence of noise, M measurements can be performed with the same input gain A.

As shown in Appendix C, with this measurement strategy, the MSD of $h_{r,i}$ is given by

$$\mathrm{MSD}_{r,i} = \frac{\sigma^2_{\mathcal{L}_{r,i}v}}{MA^{2r}}. \tag{28}$$

Comparing (28) with (22), we can see that PMV will provide a better $\mathrm{MSD}_{r,i}$ whenever the input gain A necessary to neglect the effect of any kernel of order greater than R is lower than

$$\widehat{A} = \left(M\boldsymbol{e}_r^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{e}_r\right)^{-\frac{1}{2r}}. \tag{29}$$

Tables I and II provide the value of $\widehat{A}$ in the last column for each of the optimal cases considered. For large values of M, $\widehat{A}$ is between 0.2 and 0.1, indicating reasonable values.

In the measurement process, using a sufficiently large number of input samples—i.e., considering an OPS of adequate length—allows us to handle different SNR values. However, it is important to ensure that the resolution and sensitivity of the sensors, particularly the ADC, are appropriately selected, with sufficient resolution and good linearity. These requirements are identical to those of the original identification method combined with the multiple-variance approach.

VI. Experimental Results

In this section, we first present simulation results that identify a synthetic nonlinear system with known nonlinear kernels under different noise conditions. Then, we identify a real system for which the true model is unknown.

A. Identification of a Synthetic Nonlinear System

In the first experiment, we consider the identification of a synthetic Wiener nonlinear system with known nonlinear kernels. The Wiener system is composed of the cascade of a linear FIR filter and a saturation nonlinearity. The linear FIR filter is a lowpass FIR filter of memory length 25 designed with the window method, using a Hamming window and 6 dB cut-off frequency 0.75π obtained with the MATLAB command `fir1(24,0.75)`. The saturation nonlinearity has input–output relationship

$$y = \frac{4.5}{1+2e^{-2x}} - 1.5 \tag{30}$$

whose Taylor expansion allows us to compute the Volterra kernels

$$y = 2\ x + 2/3x^2 - 4/9x^3 - 10/27x^4 + 28/405x^5 + \cdots \tag{31}$$

On a sinusoidal signal with the maximum amplitude considered in the simulations and a normalized frequency of $1/8\pi$, the nonlinear system introduces second-, third-, fourth-, fifth-, and sixth-order harmonic distortions of 11.8%, 5.0%, 1.4%, 0.24%, and 0.13%, respectively.

The nonlinear system was identified using an OPS for a Volterra system of memory 25, order 3 having period of 131 072 samples using the multiple-variance method with the gains of Table I, for $R = 3$, $K = 7$, $M = 10$, which are used for computing all kernels. We consider a high value of K, even if the fifth and sixth harmonic distortion appear to be very small because some of the measurements consider very high SNRs. For high SNRs, even a small high-order distortion can have an impact on the low-order kernel identification.

We compare the results with measurements performed using the OPS sequence with a single gain, but considering multiple repetitions of the period (i.e., ten repetitions) to compare the performance for the same amount of data. In this case, we use the same gains used for the polynomial multiple-variance method, but also a set of ten gains equally spaced in the range [0.1, 1.0].

In the measurement, the output of the nonlinear system is corrupted with an additive white Gaussian noise with different SNRs. The SNR values indicated in the figures refer to the signal having maximum gain 1.0. An output noise with the same power has been added to all the signals with lower gain.

Fig. 4 shows the MSD for kernels 1, 2, and 3, respectively, at different SNRs using the PMV method and the OPS identification on the same amount of input data for the gains 1.0, 0.92, 0.64, and 0.22 used in the PMV identification. All measurements have the same computational complexity and acquisition time. While the single-gain method relies solely on repeating the measurement to reduce the effect of noise on the kernels, the PMV approach additionally uses this repetition to mitigate distortions caused by higher-order kernels. For low SNRs, the noise obscures the effect of higher order kernels, and the PMV method does not offer any advantage compared to the OPS identification. Conversely, when the SNR is sufficiently high, the effect of higher order nonlinearities becomes evident, and the PMV method provides better results than the original OPS method for any of the considered gains. The OPS measurement of kernel 1, especially at low gains, is very robust, and the PMV method provides an advantage only at very high SNRs. The measurements of kernels 2 and 3 are here more affected by higher order nonlinearities, and with SNRs of 50 dB for the second-order kernel and 70 dB for the third-order kernel, the PMV method shows an improvement in terms of MSD.

We know that for sufficiently low gains, the higher order nonlinearities fade. Thus, for low gains and sufficiently high SNR, the PMV method and the OPS method should give similar results. Therefore, we repeated the experiment with the OPS identification on the same amount of data, considering a set of ten gains equally spaced in the range [0.1, 1.0]. Fig. 5 shows the MSD for kernels 1, 2, and 3, respectively, at different SNRs using the PMV method (the same dashed curve as in Fig. 4) and the OPS identification. For kernel 1, the PMV method and the OPS method with a gain of 0.1 show

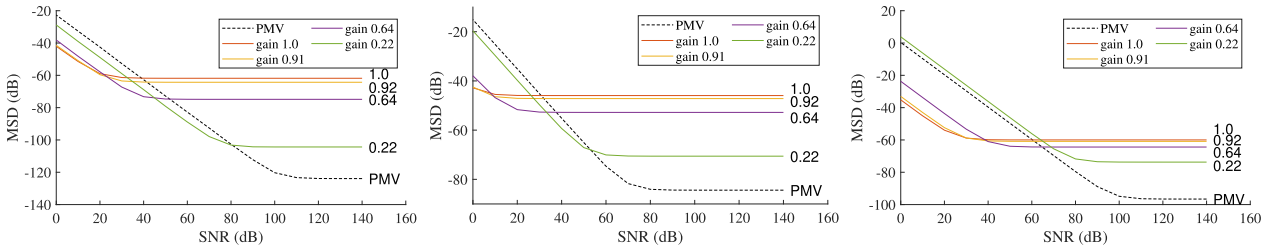


Fig. 4. MSD for kernel 1 (a), kernel 2 (b), and kernel 3 (c), for different SNR with the PMV method and with the OPS identification for the PMV optimal gains 1.0, 0.92, 0.64, and 0.22.

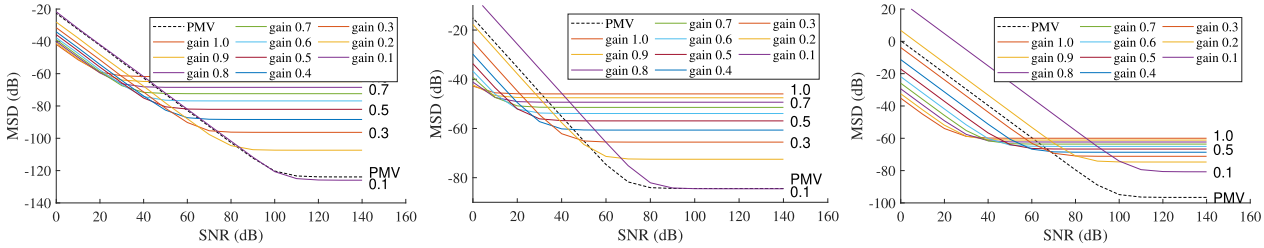


Fig. 5. MSD for kernel 1 (a), kernel 2 (b), and kernel 3 (c), for different SNR with the PMV method and with the OPS identification for the equally spaced gains 1.0, 0.9, 0.8, . . . , 0.1.

very similar results, indicating that under the experimental conditions for this kernel, the OPS measurement is already optimal. The situation changes with kernels 2 and 3, where there is a range of SNRs in which the PMV method consistently provides better results than the OPS measurement. For kernel 2 at SNRs higher than 50 dB and for kernel 3 at SNRs higher than 70 dB, it is always advantageous to choose the PMV method as the identification approach since it provides MSDs lower than or equal to the OPS method, without the need to experiment with different gains to choose the optimal setup.

We also repeated the experiment using the PMV approach to measure kernels 2 and 1, with the gains listed in Table I for $R = 2$, $K = 7$, $M = 10$ and $R = 1$, $K = 7$, $M = 14$, respectively. The MSD curves changed only slightly, indicating good robustness of the optimal gains.

We also repeated all identifications using the least-squares approach instead of the OPS technique. The least-squares approach is one of the most widely used methods for nonlinear identification and has a higher computational complexity than the OPS technique. Nevertheless, the results we obtained are very similar to those shown in Figs. 4 and 5 and have been omitted solely for the sake of space.

B. Identification of a Real System

In the second experiment, we considered the identification of a real nonlinear device, a Behringer MIC500 vacuum tube preamplifier [33], at a sampling frequency of 16 kHz. For the identification, we used the same OPS as in the previous experiment, which is suitable for a Volterra system with a memory of 25 and an order of 3, with a period of 131 072 samples, multiplied by different gains.

The input signals, originally sampled at 16 kHz, were upsampled to 48 kHz in MATLAB to facilitate their playback and subsequent recording using the Focusrite Scarlett 2i2 audio interface [34]. After the recording process, the signals

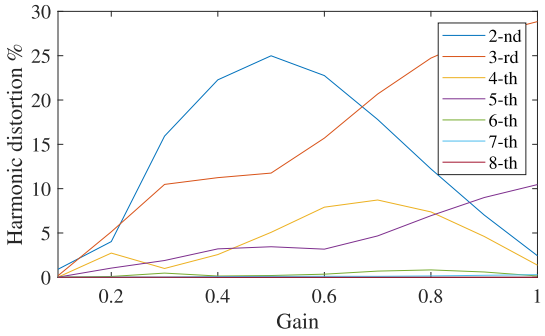


Fig. 6. Harmonic distortion on the MIC100 for gains from 0.1 till 1.0.

were downsampled to their initial rate of 16 kHz. In this experiment, the identified signal processing chain was composed of the upsampler, the Scarlett digital-to-analog converter (DAC), the MIC500 preamplifier, the Scarlett analog-to-digital converter (ADC), and finally, the downsampler. The MIC500 preamplifier was set to preamp mode "valve" with the settings "20 dB PAD," "48 V," "low cut," "phase reverse, " and "direct monitor" disabled. The GAIN setting of the MIC500 was set to a very high value to introduce strong distortions in the output signal. Fig. 6 shows the harmonic distortion on a tone at 1 kHz from the second to the eighth order for input gains ranging from 0.1 to 1.0, with 1.0 corresponding to the highest gain used in the experiments. With the noise introduced by the preamplifier, and in smaller parts by the DAC and ADC, the output SNR for the signal with maximum gain was 70 dB.

Since the harmonic distortion is negligible for orders greater than 5, the system was identified using the multiple-variance method, considering the gains in Table I, with $R = 3$, $K = 5$, and $M = 6$, which were used to compute all kernels. The system was also identified with six repetitions of the same OPS, using 10 gains equally spaced in the range [0.1, 1.0].

The main difficulty in this experiment is the absence of ground truth for comparison. The curse of dimensionality

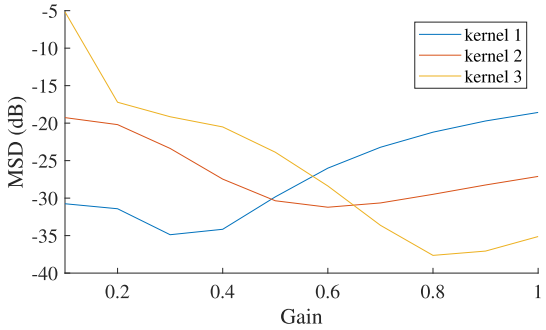


Fig. 7. MSD between the first-, second-, and third-order kernel identified with the polynomial multiple-variance method and with the ten gains from 0.1 till 1.0.

prevents us from identifying the model with a Volterra filter of order 5 and memory 25, as would be necessary in this case. The model would have 142 506 coefficients, making its identification impossible within a reasonable time frame. Nevertheless, it is known from the literature [19] that to accurately identify the highest order kernel, which in this experiment is the third-order kernel, a large gain should be used to sufficiently excite this nonlinearity while avoiding the excitation of higher order nonlinearities. Conversely, to accurately identify the first-order kernel, a low gain should be used that does not excite the nonlinearities but is still sufficiently large to counteract the effect of noise. For the second-order nonlinearity, the highest accuracy is expected from a gain between those used for the third and first orders. Thus, using the same amount of data, we compare the MSD values for the first, second, and third-order kernels identified using the PMV method and a single-gain OPS approach. We consider ten gains ranging from 0.1 to 1.0. Fig. 7 displays the MSD curves for the three kernels. For kernel 1, the curve shows a minimum at a low gain value of 0.3. For kernel 3, the minimum occurs at a high gain value of 0.8, while for kernel 2, the minimum is achieved at an intermediate gain of 0.5. These results are entirely consistent with the previous observations about the choice of gain for accurately identifying the kernels. The results indicate that the polynomial multiple-variance method provides a reasonably accurate estimate of the Volterra kernels.

VII. Conclusion

This work proposes a methodology for the accurate estimation of Volterra kernels in discrete-time nonlinear systems, particularly in cases where the system order exceeds that of the Volterra model. The proposed approach leverages multiple measurements obtained by varying the amplitude of a single excitation signal through different gain levels. Polynomial interpolation was then employed to estimate the system's Volterra kernels from these measurements. An analysis of the noise impact on the technique was performed, establishing a relationship between the MSD of the kernel coefficients and the input gains. Additionally, the optimal input gains that minimize the MSD were determined, revealing that these optimal values form a discrete set, as summarized in Tables I and II. A comparative discussion was provided on when the proposed methodology offers advantages over measurements using a single, low gain value. Finally, the efficacy of the approach was demonstrated through experimental validation involving both a simulated nonlinear system and a real-world nonlinear device.

Appendix A
Exploiting Symmetry in Gains

Consider the input signal x applied M times, with M even, and let $A_{m+M/2} = -A_m$ for $m = 1, \ldots, M/2$.

Given that the operator $\mathcal{H}_k(x)$ is odd for odd k and even for even k, when the input x is multiplied by $A_{m+M/2} = -A_m$, the corresponding system output $y_{m+M/2}$ becomes:

$$y_{m+M/2} = \mathcal{H}(-A_m x) + v_m = \sum_{k=0}^{K} (-1)^k A_m^k \mathcal{H}_k(x) + v_m. \quad (32)$$

The estimation of an odd-order coefficient can be simplified by forming the terms

$$d_{\mathcal{L}_{r,i},m,\mathrm{o}} = \frac{\mathcal{L}_{r,i}(y_m) - \mathcal{L}_{r,i}\left(y_{m+M/2}\right)}{2} = \sum_{p=1}^{\lceil K/2 \rceil} A_m^{2p-1} \mathcal{L}_{r,i} \mathcal{H}_{2p-1}(x) + \frac{\mathcal{L}_{r,i}(v_m) - \mathcal{L}_{r,i}\left(v_{m+M/2}\right)}{2} \quad (33)$$

for $m = 1, \ldots, M/2$ and by exploiting the cancellation of even-order terms. Note that when K is even, at least $K+1$ measurements with different gains A_m are needed for the polynomial fitting in equation (11), as there are $K+1$ coefficients to determine. In contrast, the polynomial fitting in equation (33) only requires K measurements, since there are $\lceil K/2 \rceil$ coefficients to estimate, and two measurements are needed for each term $d_{\mathcal{L}_{r,i},m,\mathrm{o}}$.

Define the $M/2 \times 1$ vectors

$$\boldsymbol{d}_{\mathcal{L}_{r,i},\mathrm{o}} = \left[d_{\mathcal{L}_{r,i},1,\mathrm{o}}, d_{\mathcal{L}_{r,i},2,\mathrm{o}}, \ldots, d_{\mathcal{L}_{r,i},M/2,\mathrm{o}}\right]^{\mathrm{T}} \quad (34)$$

$$\boldsymbol{v}_{\mathcal{L}_{r,i},\mathrm{o}} = \left[\frac{\mathcal{L}_{r,i}(v_1) - \mathcal{L}_{r,i}(v_{M/2+1})}{2}, \frac{\mathcal{L}_{r,i}(v_2) - \mathcal{L}_{r,i}\left(v_{M/2+2}\right)}{2}, \ldots, \frac{\mathcal{L}_{r,i}\left(v_{M/2}\right) - \mathcal{L}_{r,i}(v_M)}{2}\right]^{\mathrm{T}} \quad (35)$$

the $\lceil K/2 \rceil \times 1$ vector

$$\boldsymbol{h}_{\mathcal{L}_{r,i},\mathrm{o}} = \left[\mathcal{L}_{r,i}\mathcal{H}_1(x), \mathcal{L}_{r,i}\mathcal{H}_3(x), \ldots, \mathcal{L}_{r,i}\mathcal{H}_Q(x)\right]^{\mathrm{T}} \quad (36)$$

with $Q = 2\lceil K/2 \rceil - 1$, and the $\lceil K/2 \rceil \times M/2$ Vandermonde-like matrix

$$\boldsymbol{A}_{\mathrm{o}} = \begin{bmatrix} A_1 & A_2 & \ldots & A_{M/2} \\ A_1^3 & A_2^3 & \ldots & A_{M/2}^3 \\ A_1^5 & A_2^5 & \ldots & A_{M/2}^5 \\ \vdots & & & \vdots \\ A_1^Q & A_2^Q & \ldots & A_{M/2}^Q \end{bmatrix}. \quad (37)$$

Writing (33) for $m = 1, \ldots, M/2$ in matrix form results in

$$\boldsymbol{d}_{\mathcal{L}_{r,i},\mathrm{o}} = \boldsymbol{A}_{\mathrm{o}}^{\mathrm{T}} \boldsymbol{h}_{\mathcal{L}_{r,i},\mathrm{o}} + \boldsymbol{v}_{\mathcal{L}_{r,i},\mathrm{o}}. \quad (38)$$

Provided $M/2 \geq \lceil K/2 \rceil$ and $\boldsymbol{A}_{\mathrm{o}}\boldsymbol{A}_{\mathrm{o}}^{\mathrm{T}}$ is invertible, for r odd $h_{r,i}$ can be evaluated as

$$\widehat{h}_{r,i} = \boldsymbol{e}_q^{\mathrm{T}} \boldsymbol{h}_{\mathcal{L}_{r,i},\mathrm{o}} = \boldsymbol{e}_q^{\mathrm{T}} \left(\boldsymbol{A}_{\mathrm{o}}\boldsymbol{A}_{\mathrm{o}}^{\mathrm{T}}\right)^{-1} \boldsymbol{A}_{\mathrm{o}} \boldsymbol{d}_{\mathcal{L}_{r,i},\mathrm{o}} \quad (39)$$

with $q = \lfloor r/2 \rfloor$ and $\boldsymbol{e}_q$ is the qth column of the $\lceil K/2 \rceil \times \lceil K/2 \rceil$ identity matrix.

Similarly, the estimation of an even-order coefficient can be simplified by forming the terms

$$d_{\mathcal{L}_{r,i},m,\mathrm{e}} = \frac{\mathcal{L}_{r,i}(y_m) + \mathcal{L}_{r,i}\left(y_{m+M/2}\right)}{2} = \sum_{p=0}^{\lfloor K/2 \rfloor} A_m^{2p} \mathcal{L}_{r,i} \mathcal{H}_{2p}(x) + \frac{\mathcal{L}_{r,i}(v_m) + \mathcal{L}_{r,i}\left(v_{m+M/2}\right)}{2} \tag{40}$$

for $m = 1, \ldots, M/2$ and by exploiting the cancellation of odd-order terms.

Define the $M/2 \times 1$ vectors

$$\boldsymbol{d}_{\mathcal{L}_{r,i},\mathrm{e}} = \left[d_{\mathcal{L}_{r,i},1,\mathrm{e}}, d_{\mathcal{L}_{r,i},2,\mathrm{e}}, \ldots, d_{\mathcal{L}_{r,i},M/2,\mathrm{e}}\right]^{\mathrm{T}} \tag{41}$$

$$\boldsymbol{v}_{\mathcal{L}_{r,i},\mathrm{e}} = \left[\frac{\mathcal{L}_{r,i}(v_1) + \mathcal{L}_{r,i}(v_{M/2+1})}{2} \frac{\mathcal{L}_{r,i}(v_2) + \mathcal{L}_{r,i}(v_{M/2+2})}{2}, \ldots, \frac{\mathcal{L}_{r,i}(v_{M/2}) + \mathcal{L}_{r,i}(v_M)}{2}\right]^{\mathrm{T}} \tag{42}$$

the $\lfloor K/2 \rfloor + 1 \times 1$ vector

$$\boldsymbol{h}_{\mathcal{L}_{r,i},\mathrm{e}} = \left[\mathcal{L}_{r,i}\mathcal{H}_0(x), \mathcal{L}_{r,i}\mathcal{H}_2(x), \ldots, \mathcal{L}_{r,i}\mathcal{H}_P(x)\right]^{\mathrm{T}} \tag{43}$$

with $P = 2\lfloor K/2 \rfloor$, and the $\lfloor K/2 \rfloor + 1 \times M/2$ Vandermonde-like matrix

$$\boldsymbol{A}_{\mathrm{e}} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ A_1^2 & A_2^2 & \ldots & A_{M/2}^2 \\ A_1^4 & A_2^4 & \ldots & A_{M/2}^4 \\ \vdots & & & \vdots \\ A_1^P & A_2^P & \ldots & A_{M/2}^P \end{bmatrix}. \tag{44}$$

Writing (40) for $m = 1, \ldots, M/2$ in matrix form results in

$$\boldsymbol{d}_{\mathcal{L}_{r,i},\mathrm{e}} = \boldsymbol{A}_{\mathrm{e}}^{\mathrm{T}} \boldsymbol{h}_{\mathcal{L}_{r,i},\mathrm{e}} + \boldsymbol{v}_{\mathcal{L}_{r,i},\mathrm{e}}. \tag{45}$$

Provided $M/2 \geq \lfloor K/2 \rfloor + 1$ and $\boldsymbol{A}_{\mathrm{e}}\boldsymbol{A}_{\mathrm{e}}^{\mathrm{T}}$ is invertible, for r even $h_{r,i}$ can be evaluated as

$$\widehat{h}_{r,i} = \boldsymbol{e}_q^{\mathrm{T}} \boldsymbol{h}_{\mathrm{e}} = \boldsymbol{e}_q^{\mathrm{T}} (\boldsymbol{A}_{\mathrm{e}}\boldsymbol{A}_{\mathrm{e}}^{\mathrm{T}})^{-1} \boldsymbol{A}_{\mathrm{e}} \boldsymbol{d}_{\mathrm{e}} \tag{46}$$

with $q = r/2$ and $\boldsymbol{e}_q$ is the qth column of the $\lfloor K/2 \rfloor + 1 \times \lfloor K/2 \rfloor + 1$ identity matrix.

Appendix B
Cost Functions for Symmetric Gains

For R odd, as can be proved using (39) in the MSD evaluation, the cost function in (27) can be replaced by

$$\hat{J}_{R,o}(A_1, \ldots, A_M) = \boldsymbol{e}_q^{\mathrm{T}} \left(\boldsymbol{A}_o \boldsymbol{A}_o^{\mathrm{T}}\right)^{-1} \boldsymbol{e}_q \tag{47}$$

where $q = \lfloor R/2 \rfloor$, $\boldsymbol{e}_q$ is the qth column of the $\lceil K/2 \rceil \times \lceil K/2 \rceil$ identity matrix, and $\boldsymbol{A}_o$ is defined in (37).

For R even, as can be proved using (46), the cost function in (27) can be replaced by

$$\hat{J}_{R,e}(A_1, \ldots, A_M) = \boldsymbol{e}_q^{\mathrm{T}} \left(\boldsymbol{A}_e \boldsymbol{A}_e^{\mathrm{T}}\right)^{-1} \boldsymbol{e}_q \tag{48}$$

where $q = r/2$ and $\boldsymbol{e}_q$ is the qth column of the $\lfloor K/2 \rfloor + 1 \times \lfloor K/2 \rfloor + 1$ identity matrix, and $\boldsymbol{A}_e$ is defined in (44).

Appendix C
Multiple Measurements With the Same Input Gain

Let us apply the input x multiplied by A to the nonlinear system. According to (10), the resulting output y_m is

$$y_m = \sum_{k=0}^{K} A^k \mathcal{H}_k(x) + v_m \tag{49}$$

and if A is sufficiently low such that $\mathcal{H}_k(x) \simeq 0$ for $k > R$, then

$$y_m = \sum_{k=0}^{R} A^k \mathcal{H}_k(x) + v_m. \tag{50}$$

In these conditions,

$$\mathcal{L}_{r,i}(y_m) = A^r h_{r,i} + \mathcal{L}_{r,i}(v_m), \tag{51}$$

and by repeating the measurement M times and averaging, we can estimate $h_{r,i}$ with

$$\widehat{h}_{r,i} = \frac{\sum_{m=1}^{M} \mathcal{L}_{r,i}(y_m)}{M A^r} \tag{52}$$

with estimation error

$$-\frac{\sum_{m=1}^{M} \mathcal{L}_{r,i}(v_m)}{M A^r}.$$

Under the assumption that the noise terms $\mathcal{L}_{r,i}(v_m)$ are uncorrelated and Gaussian distributed with zero mean and variance $\sigma^2_{\mathcal{L}_{r,i}v}$, the MSD results in the expression of (28).

References

[1] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. Hoboken, NJ, USA: Wiley, 2000.

[2] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, J. Arenas-Garcia, and W. Kellermann, "Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 1, pp. 97–110, Jan. 2011.

[3] M. Faifer, R. Ottoboni, M. Prioli, and S. Toscani, "Simplified modeling and identification of nonlinear systems under quasi-sinusoidal conditions," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 6, pp. 1508–1515, Jun. 2016.

[4] C. M. Cheng, Z. K. Peng, W. M. Zhang, and G. Meng, "Volterra-series-based nonlinear system modeling and its engineering applications: A state-of-the-art review," *Mech. Syst. Signal Process.*, vol. 87, pp. 340–364, Mar. 2017.

[5] M. Alizadeh, S. Amin, and D. Ronnow, "Measurement and analysis of frequency-domain Volterra kernels of nonlinear dynamic 3 × 3 MIMO systems," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1893–1905, Jul. 2017.

[6] D. Comminiello and J. C. Príncipe, *Adaptive Learning Methods for Nonlinear System Modeling*. London, U.K.: Butterworth, 2018.

[7] M. Faifer et al., "Overcoming frequency response measurements of voltage transformers: An approach based on quasi-sinusoidal Volterra models," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2800–2807, Aug. 2019.

[8] G. P. Gibiino, A. M. Angelotti, A. Santarelli, F. Filicori, and P. A. Traverso, "Multitone multiharmonic scattering parameters for the characterization of nonlinear networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[9] A. Chatterjee and H. P. Chintha, "Identification and parameter estimation of cubic nonlinear damping using harmonic probing and Volterra series," *Int. J. Non-Linear Mech.*, vol. 125, Oct. 2020, Art. no. 103518.

[10] S. Chang et al., "Model predictive control for seizure suppression based on nonlinear auto-regressive moving-average Volterra model," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2173–2183, Oct. 2020.

[11] A. Dalla Libera, R. Carli, and G. Pillonetto, "Kernel-based methods for Volterra series identification," *Automatica*, vol. 129, Jul. 2021, Art. no. 109686.
[12] L. Lu et al., "A survey on active noise control in the past decade—Part II: Nonlinear systems," *Signal Process.*, vol. 181, Apr. 2021, Art. no. 107929.
[13] Z. Tian and F. Li, "Network traffic prediction method based on autoregressive integrated moving average and adaptive Volterra filter," *Int. J. Commun. Syst.*, vol. 34, no. 12, p. 4891, Aug. 2021.
[14] Y. Wang, S. Tang, and X. Gu, "Parameter estimation for nonlinear Volterra systems by using the multi-innovation identification theory and tensor decomposition," *J. Franklin Inst.*, vol. 359, no. 2, pp. 1782–1802, Jan. 2022.
[15] G. Gowtham et al., "A family of adaptive Volterra filters based on maximum correntropy criterion for improved active control of impulsive noise," *Circuits, Syst., Signal Process.*, vol. 41, no. 2, pp. 1019–1037, Feb. 2022.
[16] H. Skyvulstad, Ø. W. Petersen, T. Argentini, A. Zasso, and O. Øiseth, "Regularised Volterra series models for modelling of nonlinear self-excited forces on bridge decks," *Nonlinear Dyn.*, vol. 111, no. 14, pp. 12699–12731, Jul. 2023.
[17] Y. Pu, X. Li, and F. Zhang, "Hybrid control of piezoelectric flexible manipulator based on Volterra filtered-xLMS algorithm," *J. Vibrat. Control*, vol. 29, nos. 1–2, pp. 185–199, Jan. 2023.
[18] Y. Liu and H. Liang, "Review on the application of the nonlinear output frequency response functions to mechanical fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
[19] S. Orcioni, "Improving the approximation ability of Volterra series identified with a cross-correlation method," *Nonlinear Dyn.*, vol. 78, no. 4, pp. 2861–2869, Dec. 2014.
[20] A. Halme, J. Orava, and H. Blomberg, "Polynomial operators in non-linear systems theory," *Int. J. Syst. Sci.*, vol. 2, no. 1, pp. 25–47, Jul. 1971.
[21] R. J. Simpson and H. M. Power, "Correlation techniques for the identification of non-linear systems," *Meas. Control*, vol. 5, no. 8, pp. 316–321, Aug. 1972.
[22] S. Boyd, Y. Tang, and L. Chua, "Measuring Volterra kernels," *IEEE Trans. Circuits Syst.*, vol. CS-30, no. 8, pp. 571–577, Aug. 1983.
[23] L. N. O. Chua and Y. Liao, "Measuring Volterra kernels (II)," *Int. J. Circuit Theory Appl.*, vol. 17, no. 2, pp. 151–190, Apr. 1989.
[24] G.-M. Lee, "Estimation of non-linear system parameters using higher-order frequency response functions," *Mech. Syst. Signal Process.*, vol. 11, no. 2, pp. 219–228, Mar. 1997.
[25] D. Bard and G. Sandberg, "Modeling of nonlinearities in electrodynamic loudspeakers," in *Audio Engineering Society Convention 123*. New York, NY, USA: Audio Engineering Society, 2007.
[26] B. Zhang and S. A. Billings, "Volterra series truncation and kernel estimation of nonlinear systems in the frequency domain," *Mech. Syst. Signal Process.*, vol. 84, pp. 39–57, Feb. 2017.
[27] D. Bouvier, T. Hélie, and D. Roze, "Phase-based order separation for Volterra series identification," *Int. J. Control*, vol. 94, no. 8, pp. 2104–2114, Aug. 2021.
[28] A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi, "Orthogonal periodic sequences for the identification of functional link polynomial filters," *IEEE Trans. Signal Process.*, vol. 68, pp. 5308–5321, 2020.
[29] R. Forti, A. Carini, and S. Orcioni, "Polynomial multiple variance impulse response measurement," in *Proc. Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Sep. 2022, pp. 1–5.
[30] A. Carini, R. Forti, and S. Orcioni, "A polynomial multiple variance method for impulse response measurement," *Signal Process.*, vol. 207, Jun. 2023, Art. no. 108960.
[31] W. Gautschi, "Optimally conditioned Vandermonde matrices," *Numerische Math.*, vol. 24, no. 1, pp. 1–12, Feb. 1975.
[32] W. Gautschi, "Optimally scaled and optimally conditioned Vandermonde and vandermonde-like matrices," *BIT Numer. Math.*, vol. 51, no. 1, pp. 103–125, Mar. 2011.
[33] Behringer. *ECM8000 | Behringer Measurement Microphone Product Page*. Accessed: Dec. 2, 2024. [Online]. Available: https://www.behringer.com/product.html?modelCode=0838-AAP
[34] Focusrite. *Scarlett 2i2 User Guide*. Accessed: Dec. 2, 2024. [Online]. Available: https://fael-downloads-prod.focusrite.com/customer/prod/downloads/Scarlett%202i2%203rd%20Gen%20User%20Guide%20V2.pdf

Alberto Carini (Senior Member, IEEE) received the Laurea degree (summa cum laude) in electronic engineering and the Dottorato di Ricerca (Ph.D.) degree in information engineering from the University of Trieste, Trieste, Italy, in 1994 and 1998, respectively.

In 1996 and 1997, during his Ph.D., he was a Visiting Scholar at the University of Utah, Salt Lake City. From 1997 to 2003, he was a DSP Engineer with Telit Mobile Terminals SpA, Trieste. In 2003, he was with Neonseven srl, Trieste, as an Audio and DSP Expert. From 2001 to 2004, he collaborated with the University of Trieste as a Contract Professor of digital signal processing. From 2004 to 2018, he was an Associate Professor with the University of Urbino, Urbino, Italy, and since 2018, he has been an Associate Professor with the University of Trieste. His research interests include system identification, nonlinear filtering, nonlinear equalization, adaptive filtering, acoustic echo cancellation, active noise control, room response equalization, and Li-Ion battery impedance spectroscopy.

Dr. Carini professional activities are as follows: a Member-at-Large, SPS Conference Board from 2009 to 2010; a member, SPS Signal Processing Theory and Methods Technical Committee from 2006 to 2011 and from 2015 to 2020; and the Program Co-Chair signal processing area, 2009 International Symposium on Image and Signal Processing and Analysis (ISPA) and 2013 International Symposium on ISPA; an Editorial Board Member (handling editor) of Elsevier Signal Processing from 2005 to 2021. Since 2022, he is serving as Area Editor for the journal *Digital Signal Processing* and since 2023 is serving as Associate Editor for the *EURASIP Journal on Advances in Signal Processing*. He received the Zoldan Award for the best Laurea degree.

Riccardo Forti received the Laurea degree in electronic engineering from the University of Trieste, Trieste, Italy, in 2021.

From 2021 to 2023, he worked as a firmware engineer for the software and hardware company SMH Technologies, and now he is a Ph.D. student in the field of electronic engineering for quantum computing, mainly working for the Artificial Quantum Systems (ArQuS) laboratory in Trieste, to improve the performances of quantum experiments.

Simone Orcioni (Senior Member, IEEE) received the Laurea and Ph.D. degrees in electronics engineering from Università Politecnica delle Marche, Ancona, Italy, in 1992 and 1995, respectively.

In 2000, he became an Assistant Professor, teaching courses in analog and digital electronics, and publishing a text book. In 2017, he was Adjunct Professor at Ubiquitous Computing Laboratory (UC-Lab), HTWG Konstanz, where he is currently a Guest Researcher. Since 2021, he is Associate Professor at Department of Information Engineering, Università Politecnica delle Marche, where he currently serves as President of the Unified Board of the Electronic Engineering Degrees. He has published more than 50 journal articles and over 100 papers in international conference proceedings and international book chapters.

Dr. Orcioni was Guest Editor for *EURASIP Journal on Embedded Systems Frontiers in Energy Research*, and *MDPI Sensors*. Additionally, he has been a reviewer for 18 international journals, a program committee member for seven international conferences, a program chair for three international conferences, an editor of four international books, and an inventor in two patents. He is featured in the World's Top 2% Scientists by Stanford University, Stanford, CA, USA, in 2023 and 2024 "Annual Influence Ranking" and "Lifetime Scientific Influence Ranking." His research has focused on statistical device modeling and simulation, analog circuit design, cyber-physical system simulation and linear and nonlinear system identification. His current research interests include non-linear digital signal processing and power electronics for renewable energy applications.

Open Access funding provided by 'CARE-CRUI 2024 Overflow' within the CRUI CARE Agreement