# Nonlinear System Identification Using Wiener Basis Functions and Multiple-Variance Perfect Sequences☆

Alberto Carini[1]

*DIA - University of Trieste - 34127 Trieste, Italy*

Simone Orcioni

*DII - Università Politecnica delle Marche - 60131 Ancona, Italy*

Alessandro Terenzi

*DII - Università Politecnica delle Marche - 60131 Ancona, Italy*

Stefania Cecchi

*DII - Università Politecnica delle Marche - 60131 Ancona, Italy*

**Abstract**

The paper addresses nonlinear identification using the Wiener series. Differently from the traditional approach, the truncated Wiener series is expressed as a linear combination of basis functions, which are orthogonal for white Gaussian inputs. The coefficients of the basis functions are efficiently estimated with the cross-correlation method, computing the cross-correlation between the basis functions and the system output. Perfect periodic sequences (PPSs), which are periodic sequences guaranteeing the perfect orthogonality of the basis functions over a period, are also developed. The PPSs allow to avoid the estimation problems experienced with the cross-correlation method using stochastic inputs. The Wiener series formulation in terms of basis functions allows also to develop a novel, more efficient, multiple-variance identification method. Multiple-variance methods exploit input signals with multiple variances for estimating the Volterra kernels. They overcome the problem of locality of the solution, i.e., the fact that the identified model well approximates the nonlinear system only for input signal variances close to that used for the identification. Optimal values of the multiple variances are also studied in the paper. Experimental results, involving the identifications of real devices, show that the

---

[1]Corresponding author: Alberto Carini, Tel.: +39 040 5583453.

proposed approach can accurately model the identified system on a wide range of input variances.

## 1. Introduction

The Wiener series composed of the Wiener G-functionals [1, 2] was introduced to overcome one of the main limitations of the Volterra filters, whose polynomial terms are never orthogonal for any input signal. The G-functionals can be derived from the orthogonalization of the Volterra series for white Gaussian inputs. They allow the efficient identification of nonlinear systems with the cross-correlation method, which in the original formulation of Lee-Schetzen [1] requires computing the cross-correlations between products of delayed input samples and the unknown system output. When applied to stochastic inputs, the approach presents many drawbacks at the point that many researchers consider it just a "legacy" method [3, page 77]: (i) Millions of input samples are often needed for an accurate estimation of the nonlinear kernels. (ii) Due to the limitation of the input signal length and to saturation of the input amplitude, an exact white Gaussian input cannot be generated. (iii) The central moments of a Gaussian input deviate from ideal values as the moment order increases [4]. (iv) The input non-idealities affects particularly the estimation of the diagonal points of the kernels [4]. (v) The problem is exacerbated by the errors caused by a model order under-determination.

Some improvements to the original cross-correlation method of Lee-Schetzen were provided in [4, 5], where the problems of input non-ideality in the identification of diagonal points were addressed. A solution to mitigate the identification errors due to model

---

under-determination was proposed in [6].

A known drawback of Volterra and Wiener theory [7] is the problem of "locality of the solution", i.e., the fact that the estimated model well approximates the unknown nonlinear system only for input variances in a neighborhood of that used for the identification [6]. The non-idealities of input signal make the output mean square error (MSE) a function of the input variance [4]. To accurately estimate the higher order kernels, high input variances are needed to excite high order nonlinearities, but they cause high identification errors in lower order kernels. Low input variances can obtain an more accurate estimation of the lower order kernels but they produce an underestimation in high order ones. The problem can be addressed with multiple-variance approaches [6]: low input variances are used to model lower order kernels, while the input variance is gradually increased for higher order kernels. It has to be pointed out that the original multiple-variance identification method of [6] requires recomputing all lower order kernels for each kernel to be estimated. Multiple-variance approaches are very effective in solving the problem of the locality of the solution and could be fruitfully adapted and applied to many areas of nonlinear signal processing, from audio effect emulation [8, 9, 10, 11, 12], to nonlinear acoustic echo cancellation [13, 14, 15, 16], nonlinear active noise control [17, 18, 19, 20], telecommunications [21, 22, 23, 24, 25], biological system modeling [3, 26], and dynamic nonlinear system modeling in general.

In this paper, differently from the classical approach based on G-functionals, the Wiener series is expressed as a linear combination of basis functions, which are orthogonal for white Gaussian inputs, as was proposed in the early conference papers [27, 28]. In analogy to Volterra filters, the Wiener series doubly-truncated with respect to the order and the memory is here called a Wiener Nonlinear (WN) filter. Because of the orthogonality of Wiener basis functions (WBF) for white Gaussian inputs, the WN filter coefficients can be efficiently estimated with the cross-correlation method, applied here by computing the cross-correlation between the basis functions and the system output.

The WN filters are a member of the class of linear-in-the-parameters nonlinear filters, and in particular of the subclass of functional link polynomial (FLiP) filters [29, 30]. As the other members of FLiP filter class having orthogonal basis functions, e.g., Even Mirror Fourier [31, 32], Legendre [33], and Chebyshev [34] nonlinear filters, we show

3

in this paper that WN filters admit perfect periodic sequences (PPSs). The PPSs are periodic sequences that guarantee a perfect orthogonality of the basis functions of a certain nonlinear filter over a period of the sequence. The PPSs were first used for the identification of linear systems [35, 36, 37, 38] and were later applied to nonlinear systems [32, 33]. When applied as input to an unknown nonlinear system, they allow system identification with the cross-correlation approach. The PPSs for WN filters permit to avoid many of the problems encountered in Wiener series identification with the cross-correlation method for stochastic inputs. They realize a perfect orthogonality of the basis functions on a finite period and they have a finite maximum amplitude. Using PPSs, the kernel diagonal points can be accurately estimated without the need to resort to specific algorithms, as in [4, 5, 39]. Using PPSs, it is also possible to easily estimate the most relevant basis functions, according to some information criterion [32].

By exploiting the formulation of the Wiener series in terms of basis functions, a novel multiple-variance approach is further developed in the paper. Compared with the original approach of [6], the proposed approach allows to avoid recomputing all lower order kernels, with a significant reduction in the computational complexity. Moreover, multiple-variance PPSs can be devised to further improve the nonlinear estimation. A procedure for selecting the most relevant basis functions according to some information criterion [40, 41, 42] is also proposed for the multiple variance approach. Another contribution of the paper is a study about the optimal choice of the multiple input variances, both for stochastic inputs and for PPSs.

Experimental results are provided to illustrate the advantages of PPSs for WN filters and of the multiple-variance approach. It is shown that the proposed multiple-variance approach can accurately model real nonlinear devices on a wide range of input variances.

The proposed system identification approaches have been developed considering in particular a nonlinear system emulation scenario but could be applied in many other scenarios requiring system modeling, e.g., in nonlinear system equalization or linearization, in vibration and noise control, in measurements robust towards nonlinearities, in biological system modelling, and many others.

The rest of the paper is organized as follows. The Wiener basis functions and WN filters are introduced in Section 2. System identification using Wiener basis functions is

addressed in Section 3. PPSs for WN filters are discussed in Section 4. Multiple-variance system identification is the topic of Section 5. The optimal choice of basis functions and input variances is discussed in the same Section. Experimental results are given in Section 6. Eventually, Section 7 provides the conclusions of the paper.

Throughout the paper the following notation is used: $E[\cdot]$ denotes mathematical expectation, $< \cdot >_L$ denotes average over an interval of $L$ samples, $\mathcal{N}(0, \sigma_x^2)$ indicates the zero mean, variance $\sigma_x^2$, normal distribution, $\mathbb{R}^+$ is the set of positive real numbers, $\propto$ means "proportional to".

## 2. The Wiener basis functions

In this section, the Wiener series is expressed as a linear combination of basis functions. The Wiener basis functions $w_s(n)$ are a set of polynomial functions, orthogonal for any white Gaussian input signal $x(n) \in \mathcal{N}(0, \sigma_x^2)$, i.e., $E[w_{s_1}(n)w_{s_2}(n)] = 0$ for any $s_1 \neq s_2$, that can arbitrarily well approximate any discrete time, time-invariant, finite memory, continuous, nonlinear system,

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

for any memory $N$, where $f$ is a continuous functions from $\mathbb{R}^N$ to $\mathbb{R}$.

To develop the set of basis functions, we can first consider the case of $N = 1$. The nonlinear system

$$y(n) = f[x(n)] \tag{2}$$

can be arbitrarily well approximated with the set of monomials

$$\{1, \ x(n), \ x^2(n), \ x^3(n), \ \ldots\}, \tag{3}$$

which are not orthogonal for $x(n) \in \mathcal{N}(0, \sigma_x^2)$. Nevertheless, by applying the Gram-Schmidt orthogonalization to the set in (3), a set of orthogonal polynomials for $x(n) \in \mathcal{N}(0, \sigma_x^2)$ can be obtained as follows

$$\{1, \ x(n), \ x^2(n) - \sigma_x^2, \ x^3(n) - 3\sigma_x^2 x(n), \ \ldots\}. \tag{4}$$

In the orthogonalization procedure input moments naturally appears, determining the coefficients of the polynomials in (4). These polynomials are related to the Hermite polynomials [43] and in particular are probabilists' Hermite polynomials of variance $\sigma_x^2$, according to the definition in [44]. The polynomials can be generated with the following relation

$$H_{j+1}(x) = xH_j(x) - j\sigma_x^2 H_{j-1}(x), \tag{5}$$

where the recursion is initialized with $H_0(x) = 1$ and $H_1(x) = x$, and $H_j(x)$ is the Hermite polynomial of degree $j$. In what follows, for compactness the Hermite polynomials of order 0 and 1, will be indicated as 1 and $x$, respectively, while the other polynomials will be indicated as $H_j(x)$, with $j = 2, 3, ....$

To develop the basis functions for $N > 1$, we follow the same procedure of [33], [45], and [46]. The Hermite polynomials are first written for $x(n), x(n-1), \ldots, x(n-N+1)$,

$$1, \quad x(n), \quad H_2[x(n)], \quad H_3[x(n)], \quad \ldots$$

$$1, \quad x(n-1), \quad H_2[x(n-1)] \quad H_3[x(n-1)], \quad \ldots$$

$$\vdots$$

$$1, \quad x(n-N+1), \quad H_2[x(n-N+1)] \quad H_3[x(n-N+1)], \quad \ldots$$

Then, the polynomials of different variable are multiplied in any possible manner, taking care of avoiding repetitions. The resulting set of basis functions and their linear combinations form an algebra on any compact $[-A, +A]^N \in \mathbb{R}^N$, with $A \in \mathbb{R}^+$, because the set is closed under addition, multiplication, and scalar multiplication. This algebra satisfies all requirements of the Stone-Weierstrass theorem [47]:

> "Let $\mathcal{A}$ be an algebra of real continuous functions on a compact set $K$. If $\mathcal{A}$ separates points on $K$ and if $\mathcal{A}$ vanishes at no point of $S$, then the uniform closure $\mathcal{B}$ of $\mathcal{A}$ consists of all real continuous functions on $S$".

The algebra of Wiener basis functions separates points and vanishes at no point, and can thus arbitrarily well approximate the system in (1).

The Wiener basis functions of order from 0 to 3 and memory $N$ are summarized in Table 1.

Table 1: The Wiener basis functions

| Order 0 |
|---|
| 1 |

| Order 1 | |
|---|---|
| $x(n-t)$ | for $t = 0, \ldots, N-1$ |

| Order 2 | |
|---|---|
| $H_2[x(n-t)]$ | for $t = 0, \ldots, N-1$ |
| $x(n-t)x(n-t-u)$ | for $u = 1, \ldots, N-1$, and $t = 0, \ldots, N-1-u,$ |

| Order 3 | |
|---|---|
| $H_3[x(n-t)]$ | for $t = 0, \ldots, N-1$ |
| $H_2[x(n-t)]x(n-t-u)$ | for $u = 1, \ldots, N-1$, and $t = 0, \ldots, N-1-u,$ |
| $x(n-t)H_2[x(n-t-u)]$ | for $u = 1, \ldots, N-1$, and $t = 0, \ldots, N-1-u,$ |
| $x(n-t)x(n-t-u)x(n-t-v)$ | for $u = 1, \ldots, N-2$, and $v = u+1, \ldots, N-1$, $t = 0, \ldots, N-1-v,$ |

A Wiener nonlinear (WN) filter of order $P$, memory $N$, is a linear combination of the Wiener basis functions up to the order $P$ and memory $N$. To reduce the number of basis functions of the filter, it is a common practice to exploit the diagonal representation of the filter [48] and to limit the diagonal number. The diagonal number is the maximum time difference between the samples involved in each basis function. The limitation of the diagonal number finds justification in the experimental observation that in real-world nonlinear systems the "energy" of the nonlinear kernels tends to concentrate around the main diagonals, as was observed in nonlinear acoustic echo cancellation [13, 15, 49], nonlinear active noise control [50, 51], identification of nonlinear systems [52, 39]. For example, a WN filter of order 3, memory $N$, and diagonal numbers $D_2$ and $D_3$ for the

second and third order basis functions, respectively, has the following diagonal form,

$$
\begin{aligned}
y(n) = k_0 &+ \sum_{t=0}^{N-1} k_{1,t} x(n-t) + \sum_{t=0}^{N-1} k_{2,t,t} H_2[x(n-t)] + \\
&+ \sum_{u=1}^{D_2} \sum_{t=0}^{N-1-u} k_{2,t,t+u} x(n-t) x(n-t-u) + \\
&+ \sum_{t=0}^{N-1} k_{3,t,t,t} H_3[x(n-t)] + \\
&+ \sum_{u=1}^{D_3} \sum_{t=0}^{N-1-u} k_{3,t,t,t+u} H_2[x(n-t)] x(n-t-u) + \\
&+ \sum_{u=1}^{D_3} \sum_{t=0}^{N-1-u} k_{3,t,t+u,t+u} x(n-t) H_2[x(n-t-u)] + \\
&+ \sum_{u=1}^{D_3-1} \sum_{v=u+1}^{D_3} \sum_{t=0}^{N-1-v} k_{3,t,t+u,t+v} x(n-t) x(n-t-u) x(n-t-v).
\end{aligned}
\tag{6}
$$

For $D_2 = D_3 = N - 1$, a full WN filter of order 3 and memory $N$ is obtained. The set of coefficients $k_{l,...}$ of equal order $l$ forms the so-called $l$-th kernel of the WN filter. It should be noted that the expression in (6) is a truncated Wiener series of order 3 and memory $N$, and the only difference with the standard formulation [2] is the different arrangement of terms produced by the basis functions. In what follows, when a short notation is needed, the $s$-th Wiener basis functions in (6) will be indicated with $w_s(n)$, with $s = 0, 1, \ldots$, and the corresponding coefficient will be denoted with $k_s$. If we need to specify also the order $i$, the basis function and the corresponding coefficient will be denoted with $w_{i,s}(n)$ and $k_{i,s}$, respectively.

The WN filter can be transformed into a Volterra filter, by equating the polynomial terms of equal degree. For example, the WN filter in (6) can be converted into the

following Volterra filter of order 3 and memory $N$,

$$
\begin{aligned}
y(n) = h_0 &+ \sum_{t=0}^{N-1} h_{1,t} x(n-t) + \sum_{t=0}^{N-1} h_{2,t,t} x^2(n-t) + \\
&+ \sum_{u=1}^{D_2} \sum_{t=0}^{N-1-u} h_{2,t,t+u} x(n-t)x(n-t-u) + \\
&+ \sum_{t=0}^{N-1} h_{3,t,t,t} x^3(n-t) + \\
&+ \sum_{u=1}^{D_3} \sum_{t=0}^{N-1-u} h_{3,t,t,t+u} x^2(n-t)x(n-t-u) + \\
&+ \sum_{u=1}^{D_3} \sum_{t=0}^{N-1-u} h_{3,t,t+u,t+u} x(n-t)x^2(n-t-u) + \\
+ \sum_{u=1}^{D_3-1} \sum_{v=u+1}^{D_3} \sum_{t=0}^{N-1-v} &h_{3,t,t+u,t+v} x(n-t)x(n-t-u)x(n-t-v).
\end{aligned}
\tag{7}
$$

By direct inspection, we find that the following conversion formulas hold for any $t, u, v$:

$$
h_{3,t,t+u,t+v} = k_{3,t,t+u,t+v}, \tag{8}
$$

$$
h_{2,t,t+u} = k_{2,t,t+u} \tag{9}
$$

$$
h_{1,t} = k_{1,t} - \sigma_x^2 k_{3,t,t,t} - \sigma_x^2 \sum_{u=0}^{\min(D_3, N-1-t)} k_{3,t,t+u,t+u} \tag{10}
$$

$$
- \sigma_x^2 \sum_{u=0}^{\min(D_3,t)} k_{3,t-u,t-u,t}, 
$$

$$
h_0 = k_0 - \sigma_x^2 \sum_{r=0}^{N-1} k_{2,r,r}. \tag{11}
$$

## 3. System identification using Wiener basis functions

By construction the Wiener basis functions are orthogonal for a white Gaussian input signal $x(n) \in \mathcal{N}(0, \sigma_x^2)$, i.e., $E[w_{s_1}(n)w_{s_2}(n)] = 0$ for any $s_1 \neq s_2$. Thus, the coefficients $k_s$ in (6) can be estimated with the classical cross-correlation approach as follows:

$$
k_s = \frac{E[y(n)w_s(n)]}{E[w_s^2(n)]}, \tag{12}
$$

where $y(n)$ is the unknown nonlinear system output. In (12), the expectations are usually estimated with time averages over a finite period.

The most relevant basis functions, i.e., the basis functions that guarantee the most compact representation according to some information criterion in principle can also be estimated. Indeed, exploiting the orthogonality of the basis functions, the mean square error (MSE) reduction provided by $w_s(n)$ is

$$\delta \mathrm{MSE}_s = \frac{E^2[y(n)w_s(n)]}{E[w_s^2(n)]}. \tag{13}$$

The basis functions can be ranked according to the MSE reduction they produce and thus the most important basis functions can be selected by minimizing an information criterion [40, 41, 42]. In Section 6, we will provide some experimental results using the Bayesian information criterion [41], i.e., minimizing

$$\mathrm{BIC} = L \ln[\sigma_\epsilon^2(n_p)] + n_p \ln[L], \tag{14}$$

where $\sigma_\epsilon^2(n_p)$ is the variance of the residual error for the first $n_p$ most relevant basis functions and $L$ is the number of samples used in the identification. The Bayesian information criterion is one of the most popular and selective information criteria [41, 32]. Other information criteria, like the Akaike's information criterion [40], the final prediction error [40], Khundrin's law of iterated logarithm criterion [42], could also be used. It should be noted that the basis function selection according to an information criterion requires the perfect orthogonality of the basis functions and often provides erratic results when estimating the expectations of the white Gaussian signals with time averages over finite lengths.

Once a WN filter has been identified, it can be directly used, e.g., to emulate a nonlinear device, or it can be transformed into another form for different uses. It is also possible to transform it into a Volterra filter, e.g., with (8)–(11) in case of an order 3 filter.

The cross-correlation method obtained by implementing (12) using Gaussian input signals requires a huge number of input samples, in the order of millions or more, for guaranteeing an approximate orthogonality of the basis functions and a reasonable ac-

curacy in the coefficients. The problem can be solved using perfect periodic sequences (PPSs), which are periodic sequences that guarantee a perfect orthogonality of the basis functions over a period.

## 4. Perfect periodic sequences

In this Section, we want to develop a PPS $x_{\mathrm{p}}(n)$ of period $L$ suitable for the identification of the WN filters up to an order $P$ and memory $N$ with Gaussian variance $\sigma_x^2$. For sake of simplicity, we consider a full WN filter, but everything presented holds also when we limit the diagonal number of the nonlinear kernels. The PPS should to be bounded by 1, i.e., $|x_{\mathrm{p}}(n)| < 1$ for all $n$, to be able to faithfully reproduce the sequence using digital to analog converters.

The approach followed in this paper differs from [32, 46, 53], where PPSs for even mirror Fourier, Legendre, and Chebyshev filters were obtained.

The cross-product $E[w_{s_1}(n)w_{s_2}(n)]$ between any two basis functions of a WN filter of order $P$ and memory $N$ is a linear combination of joint moments of the input signal $x(n)$ up to the memory $N$ and order $2P$, i.e., of

$$E[x^{r_0}(n) \cdot x^{r_1}(n-1) \cdot \ldots \cdot x^{r_{N-1}}(n-N+1)]$$

with $r_0, r_1, \ldots, r_{N-1} \in \mathbb{N}$ and $r_0 + \ldots + r_{N-1} \leq 2P$. The same observation applies also to the cross-products $< w_{s_1}(n)w_{s_2}(n) >_L$ with the expectations replaced with time averages. The orthogonality of the basis functions implies that the joint moments are the same of a Gaussian distribution, but also the reverse holds. Consider a periodic sequence $x_{\mathrm{p}}(n)$ that over the period $L$ has the same joint moments up to memory $N$ and order $2P$ of a white Gaussian signal $\mathcal{N}(0, \sigma_x^2)$. The periodic sequence guarantees over the period the orthogonality of all basis functions and is perfect for the identification of the WN filter. Thus, a PPS for WN filters of memory up to $N$ and order up to $P$ can be obtained by imposing the following system of nonlinear equations:

$$< x_{\mathrm{p}}^{r_0}(n) \cdot x_{\mathrm{p}}^{r_1}(n-1) \cdot \ldots \cdot x_{\mathrm{p}}^{r_{N-1}}(n-N+1) >_L = \mu_{r_0} \cdot \mu_{r_1} \cdot \ldots \cdot \mu_{r_{N-1}}, \qquad (15)$$

for all $r_0, r_1, \ldots, r_{N-1} \in \mathbb{N}$ with $r_0 > 0$ (for the periodicity of the sequence) and $r_0 + r_1 +$

$\ldots + r_{N-1} \le 2P$. The quantity $\mu_r$ is the $r$-th moment of the Gaussian process $\mathcal{N}(0, \sigma_x^2)$,

$$\mu_r = E[x^r(n)] = \begin{cases} 0 & \text{for } r \text{ odd,} \\ (r-1)!! \, \sigma_x^r & \text{for } r \text{ even,} \end{cases} \tag{16}$$

with $q!!$ the double factorial [54] and $q!! = q \cdot (q-2) \cdot (q-4) \cdot \ldots \cdot 1 = \frac{q!}{2^{\frac{q-1}{2}} \left( \frac{q-1}{2} \right)!}$.

The number of equations $Q$ of the nonlinear system (15) is equal to the number of different basis functions of a Volterra filter of order $2P-1$ and memory $N$ (since $r_0 > 0$, all equations has a factor $x_{\mathrm{p}}(n)$ ), i.e., $Q = \binom{N+2P-1}{N}$. For sufficiently large $L$, the system is underdetermined and may have infinite solutions in the variables $x_{\mathrm{p}}(n)$. To solve it, we have found particularly effective the Newton-Raphson method, which has been implemented as described in [55, ch. 9.7], starting from a random Gaussian distribution of the variables with variance $\sigma_x^2$, with the Jacobian matrix computed analytically. The variables $x_{\mathrm{p}}(n)$ were reflected in $[-1, +1]$ every time they exceeded the range to obtain a sequence bounded by $-1$ and $+1$, as desired. The Newton-Raphson method converges to a solution only if the signal power $\sigma_x^2$ is sufficiently small. Indeed, the PPS converges to a distribution similar to the Gaussian and convergence is possible only when the probability of finding samples outside the range $[-1, +1]$ is sufficiently small. Anyway, a solution was always found for $L$ ranging between $3Q$ and $4Q$ and for $\sigma_x^2 \le 1/10$. The number of iterations of the Newton-Raphson method depends on the ratio $L/Q$ and on the signal power $\sigma_x^2$. The larger the ratio $L/Q$ or the smaller the signal power $\sigma_x^2$, the faster is the speed of convergence of the Newton-Raphson method. Since the Newton-Raphson is a numerical method, the solution is only approximate but the precision can be arbitrarily improved acting on the stop-condition of the method. Imposing $\sigma_x^2 = 1/12$, and $L = 4Q$, and as stop condition the maximum absolute deviation of the moments to be lower than $10^{-13}$, the Newton-Raphson method in our experiments has converged in around 100 iterations.

A problem of the system in (15) is the large number of equations and variables also for low orders $P$ and memories $N$. Indeed, $Q$ depends exponentially on the order $P$ and geometrically on the memory $N$. The number of equations and thus variables can be reduced by imposing specific structures to the PPS, as done in [32] and [33]. The following conditions can almost halve the number of equations and variables:

12

- Symmetry: when in the PPS for any $N$-tuple of samples $a_1, a_2, \ldots, a_N$, there is also the reversed one $a_N, a_{N-1}, \ldots, a_1$, for every couple of symmetric joint moments (e.g., $< x(n)x^3(n-1) >_L$ and $< x^3(n)x(n-1) >_L$), it suffices to consider only one of them.

- Oddness: when in the PPS for any $N$-tuple of samples $a_1, a_2, \ldots, a_N$, there is also the negated one $-a_1, -a_2, \ldots, -a_N$, all odd joint moments are *a priori* zero.

- Oddness-1: when in the PPS for any $N$-tuple of samples $a_1, a_2, \ldots, a_N$, there is also the one obtained by alternatively negating one every two terms $a_1, -a_2, a_3, \ldots, -a_N$, all odd-1 joint moments are *a priori* zero.

By definition, odd-1 are all those joint moments that change sign by alternatively negating the sign of the samples, as for example $< x(n)x(n-1) >_L$. Odd-2, Odd-4, Odd-$2^k$, joint moments, with $k \in \mathbb{N}^+$, are similarly defined and could also be exploited to reduce the number of equations and variables. Multiple structural conditions can be imposed at the same time and imposing them is often fundamental for being able to find a solution to the system in (15) in acceptable time. Indeed, the Newton-Raphson algorithm has memory and processing time requirements that grow with $Q^3$.

PPSs with different variances can be obtained by properly scaling a PPS of variance $\sigma_x^2$. Indeed, if we scale the PPS by a factor $c$, any order $r$ joint moment in (15), with $r = r_0 + r_1 + \ldots + r_{N-1}$, is scaled by a factor $c^r$ and the sequence is still a PPS suitable for the identification of WN filters but for Gaussian variance $c^2\sigma_x^2$. This property allows the use of PPSs for the multiple-variance system identification approach of Section 5.

The interested reader can download PPSs for WN filters of order 3, signal power $\sigma_x^2 = 1/12$, and memories $N$ ranging from 5 to 20, from the website [56].

A PPS input signal allows to accurately estimate the coefficients of the WN filter in (6) with the cross-correlation approach, implemented now computing time averages over a period $L$ of the sequence, i.e.,

$$k_s = \frac{< y(n)w_s(n) >_L}{< w_s^2(n) >_L}. \tag{17}$$

13

Equation (13) still holds, with the expectations replaced by time averages,

$$\delta\text{MSE}_s = \frac{< y(n)w_s(n) >_L^2}{< w_s^2(n) >_L},\tag{18}$$

thus allowing a simple ranking of the basis functions. The most relevant basis functions can be selected by minimizing an information criterion, e.g., minimizing (14) for the Bayesian information criterion.

## 5. Multiple-variance system identification

The multiple-variance method is used to contrast the problem of locality of the identified system. When applying the cross-correlation method with stochastic input signals, the input non-idealities affect the kernel estimation, and the identification of the kernel diagonal points is particularly affected [4]. Any difference between the input joint moments and the corresponding ideal values causes errors in the identification, even if the unknown system and the model have the same memory and order. Other sources of identification error are the under-estimation of the unknown system memory or order and the effect of the output noise. The effect of these errors is particularly evident in the phenomenon of the locality of the solution. The normalized MSE (NMSE) between the output of the unknown system and the identified model is a function of the input signal power. Often the NMSE provides reasonable results only in a limited range of input signal powers, around the input variance used to estimate the model. If the system is identified with an high input variance, the high order nonlinearities are well excited and a low estimation error is obtained in the high order kernels, but high errors are caused in the estimation of the low order kernels [6]. The use of a low input variances can allow a more accurate estimate of the lower order kernels, but at the detriment of the higher orders kernel estimation. The multiple-variance method improve the accuracy in the estimation of the different kernels by using input signals with multiple-variances: the variance is gradually increased for estimating the higher order kernels. In what follows, we first present the multiple-variance approach and then we discuss how the most relevant basis functions can be selected according to some information criterion.

14

*5.1. The multiple-variance approach*

In the multiple-variance approach, each nonlinear kernel is estimated with a possibly different input signal variance $\sigma^2_{x,(i)}$ of the white Gaussian noise (WGN) or PPS input. Let us indicated with $k^{(i)}_{j,\dots}$ the $j-th$ kernel estimated with input variance $\sigma^2_{x,(i)}$, and with $x^{(i)}$ and $y^{(i)}(n)$ the input and output signal, respectively, for the same input variance $\sigma^2_{x,(i)}$. In the original approach of [6], each kernel $i$ is estimated from the expectation between $y^{(i)}(n)$ and a product of delayed input samples $x^{(i)}(n)$. For example $k^{(i)}_{3,t,t+u,t+v}$ is estimated from $E[y^{(i)}(n)x^{(i)}(n-t)x^{(i)}(n-t-u)x^{(i)}(n-t-v)]$. Nevertheless, the approach of [6] requires recomputing each odd/even lower order kernel for each odd/even kernel to be estimated, since

$$E[y^{(2)}(n)x^{(2)}(n-t)x^{(2)}(n-t)] = 3\sigma^4_{x,(2)}k^{(2)}_{2,t,t} + k^{(2)}_0\sigma^2_{x,(2)}, \qquad (19)$$

and

$$E[y^{(3)}(n)x^{(3)}(n-t)x^{(3)}(n-t)x^{(3)}(n-t)] = 6\sigma^4_{x,(3)}k^{(3)}_{3,t,t,t} + 3k^{(3)}_{1,t}\sigma^2_{x,(3)}. \qquad (20)$$

Similar formulas can be found also for larger order kernels.

On the contrary, if we consider the unknown system a linear combination of Wiener basis functions, each kernel can be separately estimated with (12), without the necessity of recomputing the lower order kernels, with a significant computational saving.

Clearly, the Wiener basis functions and the WN filter coefficients change with the variance of the input signal. Given the multiple-variance kernels $\{k^{(0)}_0, k^{(1)}_{1,.}, k^{(2)}_{2,.,.}, k^{(3)}_{3,.,.,.}, \dots\}$, we want to estimate the corresponding Volterra kernels, which are independent of the input variance. For a nonlinear system of order 3, by equating (6) and (7), it can be noticed that the two largest order kernels of the WN filter are equal to the corresponding kernels of the Volterra filter for any input variance $\sigma^2_x$. Thus, for $u = 0,..,D_3 - 1$, $v = u,..,D_3$, $t = 0,..,N-1-v$:

$$h_{3,t,t+u,t+v} = k^{(3)}_{3,t,t+u,t+v} \qquad (21)$$

and for $u = 0,..,D_2$, $t = 0,..,N-1-u$:

$$h_{2,t,t+u} = k^{(2)}_{2,t,t+u}. \qquad (22)$$

15

Equating the first order terms in (6) and (7) for $\sigma_x = \sigma_{x,(1)}$, since $k^{(1)}_{3,t,t+u,t+v} = k^{(3)}_{3,t,t+u,t+v}$ for all $u$, $v$, and $t$, it results

$$h_{1,t} = k^{(1)}_{1,t} - \sigma^2_{x,(1)} k^{(3)}_{3,t,t,t} - \sigma^2_{x,(1)} \sum_{u=0}^{\min(D_3, N-1-t)} k^{(3)}_{3,t,t+u,t+u} - \sigma^2_{x,(1)} \sum_{u=0}^{\min(D_3, t)} k^{(3)}_{3,t-u,t-u,t}.$$

(23)

Equating the constant terms in (6) and (7) for $\sigma_x = \sigma_{x,(0)}$, since $k^{(0)}_{2,t,t+u} = k^{(2)}_{2,t,t+u}$ for all $u$ and $t$, we have

$$h_0 = k^{(0)}_0 - \sigma^2_{x,(0)} \sum_{r=0}^{N-1} k^{(2)}_{2,r,r}.$$

(24)

This procedure can be applied also for filters of order greater than 3 obtaining similar formulas.

### 5.2. The most relevant basis functions

Selecting the most relevant basis functions according to some information criterion in the multiple variance approach is not a trivial task. Equations (18) and (14) cannot be directly applied since in multiple variance approaches there are multiple inputs and outputs, and each kernel is separately estimated. Moreover, as shown in the experimental results for the single variance case, the number of selected basis functions varies with the input signal power. To address these problems, the basis functions selection was separately performed on each kernel using (18) and (14), starting from the lowest order kernel till the highest order one. When processing the kernel of order $i$, in (14) the initial value of $n_p$ accounts for all lower order basis functions previously selected, and the initial value of $\sigma^2_\epsilon(n_p)$, say $\sigma^2_{\epsilon,(i)}$, is set equal to the output signal power $P_{y,(i)} = \langle [y^{(i)}(n)]^2 \rangle_L$ minus the power of all previously selected basis functions for input signal variance $\sigma^2_{x,(i)}$, i.e., it is

$$\sigma^2_{\epsilon,(1)} = P_{y,(1)} - (k^{(0)}_0)^2,$$

(25)

$$\sigma^2_{\epsilon,(2)} = P_{y,(2)} - \sum_{t=0}^{N-1} (k^{(1)}_{1,t})^2 \sigma^2_{x,(2)} - (k^{(0)}_0)^2,$$

(26)

$$\sigma^2_{\epsilon,(3)} = P_{y,(3)} - \sum_{t=0}^{N-1} (k^{(2)}_{2,t,t})^2 2\sigma^4_{x,(3)} - \sum_{u=1}^{D_2} \sum_{t=0}^{N-1-u} (k^{(2)}_{2,t,t+u})^2 \sigma^4_{x,(3)}$$

$$-\sum_{t=0}^{N-1}(k_{1,t}^{(1)})^2\sigma_{x,(3)}^2 - \left[k_0^{(0)} + \sum_{t=0}^{N-1}k_{2,t,t}^{(2)}(\sigma_{x,(3)}^2 - \sigma_{x,(2)}^2)\right]^2, \tag{27}$$

for $i = 1$, 2, and 3, respectively, with the coefficients $k_{1,t}$ and $k_{2,t,u}$ set to zero if the corresponding basis functions have not been selected. The formulas for $\sigma_{\epsilon,(i)}^2$ in (25)–(27), have been derived by expressing each model of order $i-1$ in terms of the basis functions for $\sigma_{x,(i)}^2$ and computing the powers of these basis functions.

The selection procedure can be easily implemented with the multiple variance approach discussed in this section, and has proved very effective in all our experimental results. It allows to reduce the number of basis functions used to model a nonlinear system, while keeping almost unaltered the performance of the model at the various input powers.

### 5.3. Optimal choice of the multiple variances

An important problem, which has been often overlooked in the literature, is the optimal choice of the multiple variances used to estimate the nonlinear system. Different optimality criteria could be used to find the multiple variances. A reasonable criterion estimates each kernel at the input signal variance that minimizes the error on the kernel coefficients. Thus, in the following we look for the input variance $\sigma_{x,(i)}^2$ which minimizes the Mean Square Deviation (MSD) between the estimated coefficients of the $i$-th kernel and the desired value of these coefficients. In order to account for an unknown system under estimation, which affects most of practical measurements, we assume the unknown nonlinear system to have order $R$ and memory $M$ possibly larger than those of the model, which has order $P$ and memory $N$. We also assume the output of the system to be affected by an additive white Gaussian noise of variance $\sigma_\nu^2$. In these conditions, in the Appendix we estimate the MSD in the identification of the $i$-th kernel,

$$\mathrm{MSD}_i = \sum_s E\left[(k_{i,s} - \overline{k}_{i,s})^2\right] \tag{28}$$

where $k_{i,s}$ is a measured coefficient using time averages,

$$k_{i,s} = \frac{<y(n)w_{i,s}(n)>_L}{<w_{i,s}^2(n)>_L}, \tag{29}$$

and $\overline{k}_{i,s}$ is its true value,

$$\overline{k}_{i,s} = \frac{E[y(n)w_{i,s}(n)]}{E[w_{i,s}^2(n)]}. \tag{30}$$

Both in the case of a Gaussian input and of a PPS input, it is shown in the Appendix that $\text{MSD}_i$ assumes the same form, i.e.,

$$\text{MSD}_i = \frac{u_{i,0}}{\sigma_{x,(i)}^{2i}} + \frac{u_{i,1}}{\sigma_{x,(i)}^{2(i-1)}} + \frac{u_{i,2}}{\sigma_{x,(i)}^{2(i-2)}} + \ldots + u_{i,i} + u_{i,i+1}\sigma_{x,(i)}^2 + \ldots$$

$$+ u_{i,R}\sigma_{x,(i)}^{2(R-i)} + \frac{u_{i,\nu}\sigma_\nu^2}{\sigma_{x,(i)}^{2i}}, \tag{31}$$

where $u_{i,0}$, ..., $u_{i,R}$ are proportionality constants and $u_{i,0} = 0$ for a PPS input.

Using the conversion formulas (21)–(24), it is shown in Appendix that (31) applies (with different proportionality constants) also to the Volterra filter coefficients.

The MSD expression in (31) highlights the different dependence on $\sigma_{x,(i)}^2$ in the various terms. The largest is $\sigma_{x,(i)}^2$, the lowest is the influence of noise and of the error terms of order lower than $i$ (i.e., of the terms depending on $u_{i,0}, u_{i,1}, \ldots, u_{i,i-1}$) . On the other hand, the largest is $\sigma_{x,(i)}^2$, the highest is the influence of the error terms of order larger than $i$ (i.e., of the terms depending on $u_{i,i+1}, \ldots, u_{i,R}$). The choice of $\sigma_{x,(i)}^2$ should compromise these contrasting effects.

Ideally, each kernel of order $i$ should be estimated at the input variance $\sigma_{x,(i)}^2$ that minimizes (31). In reality, the coefficients $u_{i,j}$ are rarely known and the equation (31) cannot be solved. Nevertheless, this equation is very useful for guiding the choice of the multiple variances used to estimate the different kernels.

According to (31), the kernel of order 0 should be estimated at the lowest possible input variance, since

$$\text{MSD}_0 = u_{0,0} + u_{0,1}\sigma_{x,(0)}^2 + u_{0,2}\sigma_{x,(0)}^4 + \ldots + u_{0,R}\sigma_{x,(0)}^{2R} + u_{0,\nu}\sigma_\nu^2. \tag{32}$$

Neglecting $u_{1,0}$, the kernel of order 1 should be estimated at the lowest possible input variance for which the effect of noise is still negligible, since

$$\text{MSD}_1 = u_{1,1} + u_{1,2}\sigma_{x,(1)}^2 + \ldots + u_{1,R}\sigma_{x,(1)}^{2R-2} + u_{1,\nu}\frac{\sigma_\nu^2}{\sigma_{x,(1)}^2}. \tag{33}$$

18

Table 2: Memory and diagonal numbers of the nonlinear kernels of the PPS and of the identified Volterra and Wiener models. $M_j$ is the memory of kernel $j$, and $D_j$ is the diagonal number of kernel $j$.

| Memory | Diagonal Number |
|---|---|
| $M_1 = 64$ | |
| $M_2 = 40$ | $D_2 = 15$ |
| $M_3 = 32$ | $D_3 = 12$ |

If all the kernels of order larger than $P$ are negligible, then the kernel of order $P$ should be estimated at the largest possible input variance, since

$$\text{MSD}_P = \frac{u_{P,0}}{\sigma_{x,(P)}^{2P}} + \frac{u_{P,1}}{\sigma_{x,(P)}^{2(P-1)}} + \ldots + u_{P,P} + \frac{u_{P,\nu}\sigma_\nu^2}{\sigma_{x,(P)}^{2P}}. \tag{34}$$

Eventually, the kernels of order $2, \ldots, P-1$ should be estimated with input variances ranging between $\sigma_{x,(1)}^2$, and $\sigma_{x,(P)}^2$.

## 6. Experimental results

In this Section we provide a set of experimental results that illustrate the proposed identification procedure and the achievable performances. In what follows, we first report a description of the experimental set-up and then a discussion on the obtained results.

### 6.1. The experimental set-up

The identification of a vacuum-tube audio device, i.e., the Synthesis Roma 27AC amplifier is here considered. The amplifier was loaded with a pair of loadspeakers, as suggested in the literature [57]. In particular a pair of Auna 501 passive 2-way speakers with a RMS rated power of 100 W, a nominal impedance of 8 $\Omega$, and a frequency range of 92 Hz – 20 kHz was adopted. The volume was set at the highest value and after half an hour of warm up, the input signal was applied and the output signal was measured across the amplifier output terminals as an electric signal since the loudspeakers have been used only as load of the amplifier. At the maximum used amplitude the second, third and total harmonic distortion in percent on a tone at 1 kHz were respectively, 0.47%, 0.99% and 1.097%.

A sampling frequency of 44.1 kHz was adopted since it is sufficient for the characterization the audible part of the output signal [58]. The amplifier was fed with four WGN sequences and four PPS sequences at different variances and the corresponding output signals were recorded and used to identify the model. The chosen input variances $\{\sigma^2_{x,(1)}, \sigma^2_{x,(2)}, \sigma^2_{x,(3)}, \sigma^2_{x,(4)}\}$ were the same for the two set of sequences and were equal to

$$\{\sigma^2_{x,(4)}/64,\ \sigma^2_{x,(4)}/16,\ \sigma^2_{x,(4)}/2,\ \sigma^2_{x,(4)}\}, \tag{35}$$

respectively, where $\sigma^2_{x,(4)} = 1/12$. The output signal to noise ratio (SNR) at the maximum used input variance was 38 dB and at the minimum used input variance was 28 dB. The PPSs were all scaled version of a PPS of period 2 097 656, suitable for the identification of a WN filter of order 3, with the following characteristics: memory of 1st order kernel 64; memory and diagonal number of 2nd order kernel 40 and 15, respectively; memory and diagonal number of 3rd order kernel 32 and 12, respectively, as also reported in Table 2.

To model the system, four WN filters were identified with the cross-correlation algorithm using PPSs with a different variances belonging to the set in (35). Then, a single WN filter was identified with the proposed multiple-variance method and with the original method of [6], estimating each kernel with one of the variances reported in (35). The WN models were eventually converted into Volterra models using the formulas in (8)–(11) for a single variance input, or (21)–(24) for the multiple-variance case. The same identification was performed with the white Gaussian sequences. In all the experiments, the same results were obtained with the proposed multiple-variance approach and with the approach of [6], showing the equivalence of the approaches apart from the different computational complexity.

The Volterra and Wiener models identified with the set of WGN and PPS inputs, have order 3 and memory and diagonal number of the different kernels again reported in Table 2. The constant kernel of order 0 has also been considered, and thus these models have a total of 2705 basis functions and coefficients. For the PPS inputs, reduced models were then determined by sorting the basis functions according to the MSE they produce and applying the Bayesian information criterion in (14). The reduced models were determined also for the WGN inputs, but the basis function selection process pro-

Table 3: Number of selected basis functions at different orders for the complete models and the reduced models identified with the multiple variance (MV) approach and single variance (SV) approach with variances $\sigma^2_{x,(1)}, \sigma^2_{x,(2)}, \sigma^2_{x,(3)}, \sigma^2_{x,(4)}$.

|  | Complete Model | MV | SV $\sigma^2_{x,(1)}$ | SV $\sigma^2_{x,(2)}$ | SV $\sigma^2_{x,(3)}$ | SV $\sigma^2_{x,(4)}$ |
|---|---|---|---|---|---|---|
| Order 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| Order 1 | 64 | 61 | 57 | 61 | 63 | 63 |
| Order 2 | 520 | 44 | 27 | 56 | 114 | 136 |
| Order 3 | 2120 | 118 | 38 | 53 | 100 | 254 |
| Total | 2705 | 224 | 123 | 170 | 277 | 453 |

vided poor results because of the not perfect orthogonality of the basis functions over an interval of two million samples: a much larger number of basis functions were included in the model and (25)–(27) often produced negative values. Thus, the results of reduced models with WGN inputs are not included in the paper. Table 3 provides the number of selected basis functions of the complete models and of the reduced models obtained with the Bayesian information criterion using the single variance approach and the multiple variance approach of Section 5. Note that in the single variance approach the number of selected basis functions changes with the power of the input signal, because of the different importance of the basis functions at different input variances.

The identified models were then tested using as input signals 13 WGN sequences of 2 seconds with variance ranging between $2.0345 \cdot 10^{-5}$ and $8.33 \cdot 10^{-2}$, and 13 music sequences of 2 seconds (i.e., pop music track "I Know It's Over" by Mario Biondi) again with the same variances. The performance of the different models has been compared using the normalized mean square error (NMSE) defined over the output spectrogram. The NMSE can be defined formally by considering the Short Time Fourier Transform (STFT) of the amplifier output and the model output,

$$Y(\tau, \omega) = \text{STFT}[y(n)] \tag{36}$$

$$\hat{Y}(\tau, \omega) = \text{STFT}[\hat{y}(n)], \tag{37}$$

where $y(n)$ is the amplifier output and $\hat{y}(n)$ is the model output. The STFT is calculated using a Hamming window and dividing the input signal into eight segments with 50%

overlap. The mean square error between two matrices can be defined as

$$\text{MSE}[\hat{Y}, Y] = \frac{1}{RC} \sum_i \sum_j [\hat{Y}(i,j) - Y(i,j)]^2, \tag{38}$$

where $RC$ is the product between the number of rows and columns of the matrices. The NMSE used to assess the performance of the different models is defined as

$$\text{NMSE} = \text{MSE}(|\hat{Y}|, |Y|)/\text{MSE}(|Y|, 0), \tag{39}$$

where $|.|$ denotes the magnitude applied to each element of the matrix.

*6.2. Discussion*

Figure 1 shows the test results of the Volterra model identified with white Gaussian sequences, and tested with white Gaussian noise and music test inputs, respectively. On the contrary, Figure 2 shows the test results of the Volterra model identified with PPSs, again for the same white Gaussian noise and music test inputs. Figure 1 and 2 show the NMSE of both the single-variance models and of the multiple-variance models, as a function of standard deviation of test signals.

In Figure 1.(a) we can see that the error curve related to the multiple-variance models interpolates the inferior values of the curves of single-variance models, as the multiple-variance method was meant to do. The single-variance models have similar performance to the multiple-variance one only in a short interval around the variance used for the identification, while outside this interval the error increases up to 15 dB. In Figure 1.(b) the performance of the multiple-variance model is almost equal to that of the single-variance model $\sigma_{x,(4)}$, which is superior to all other single-variance models. In Figure 1.(b), for high input variances all models provide poor identification results. According to Table 3, the complete model is overdetermined and the errors in the coefficients estimation due to imperfect white Gaussian input manifest especially at high input variances.

In Figure 2, the multiple variance-model identified with PPSs provides very good results for both the white Gaussian and music test inputs. Also with PPS, the single variance models can provide slightly better results than the multiple variance model, around the variances used in the measurements, but for the other variances the multiple
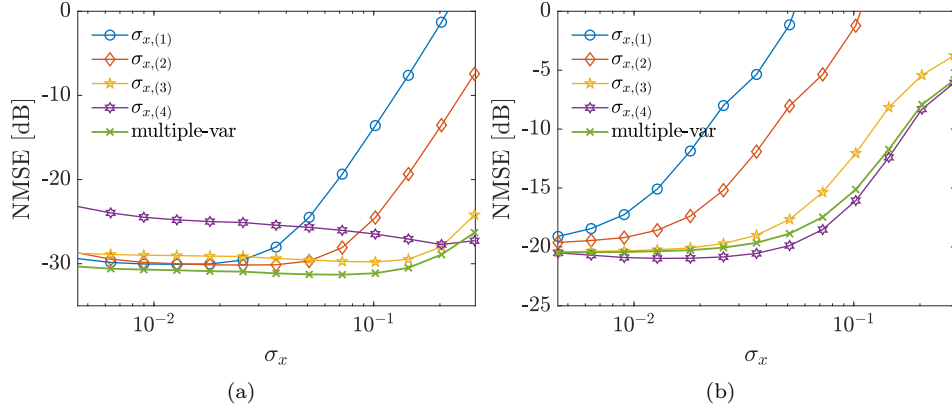
22

Figure 1: NMSE of Volterra models, identified with WGN. (a) WGN test input, (b) music test input.
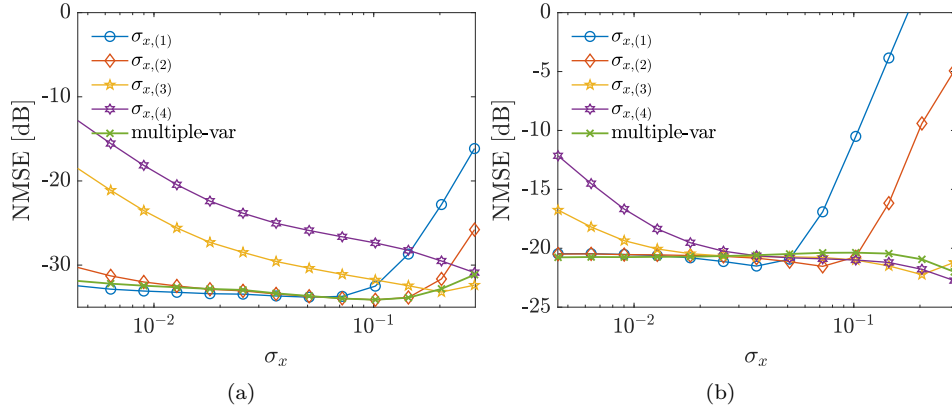


Figure 2: NMSE of Volterra models identified with PPS. (a) WGN test input, (b) music test input.

variance model is always far superior. It should be noticed that the multiple-variance model identified with PPSs has a NMSE always lower than $-22$ dB.

Comparing Figures 1 and 2, the improvement of the NMSE in the multiple variance approach using PPS inputs can be clearly appreciated. Also with PPS, the single variance approaches obtain a better NMSE for the variances used in the measurement, but a worse performance could result at other input variances.

To better visualize the relation between the NMSE, shown in the previous Figures and the real outputs, Figure 3 reports a comparison between amplifier and Volterra model outputs, for the case of identification with WGN and PPS input, when as test input is applied the same music track used for the realization of Figures 1.(b) and 2.(b), with the
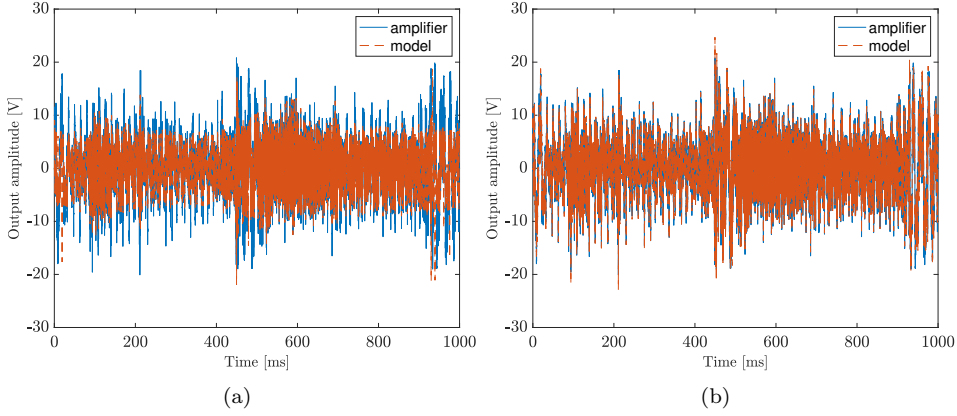
Figure 3: Outputs of amplifier and of (a) multiple-variance WGN Volterra model, (b) multiple-variance PPS Volterra model, for the music test input.
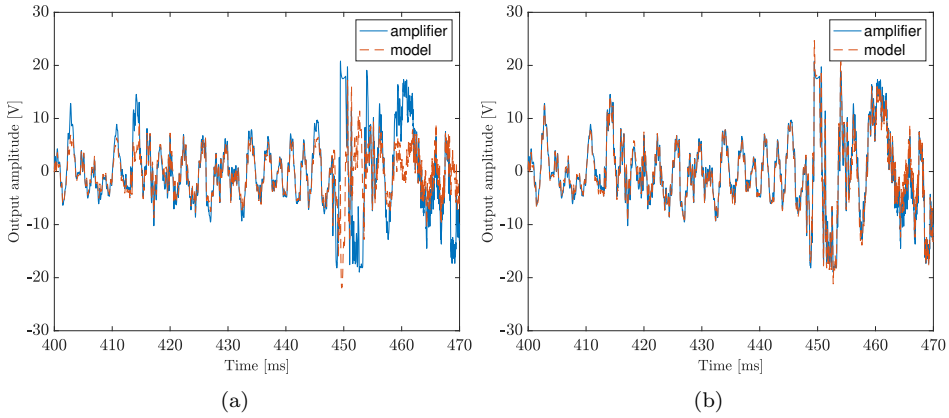


Figure 4: Zoom in time (from 400 ms to 470 ms) of outputs of amplifier and of (a) multiple-variance WGN Volterra model, (b) multiple-variance PPS Volterra model, for the music test input.

maximum variance. Only the first second of music track is reported to make the Figure clearer. In Figure 4, to better visualize the difference between the two identified Volterra models, a zoom in time of Figure 3 is reported. As can be clearly seen from both the Figures 3 and 4, in the case of the music input with highest variance the Volterra model identified with WGN exhibits a higher error with respect to the model identified with PPS, as also shown from the NMSE reported in Figures 1 and 2.

Figure 5 reports a time zoom, equivalent to Figure 4, but in the case of WGN test input. Since the WGN input has higher frequency contents with respect to the music
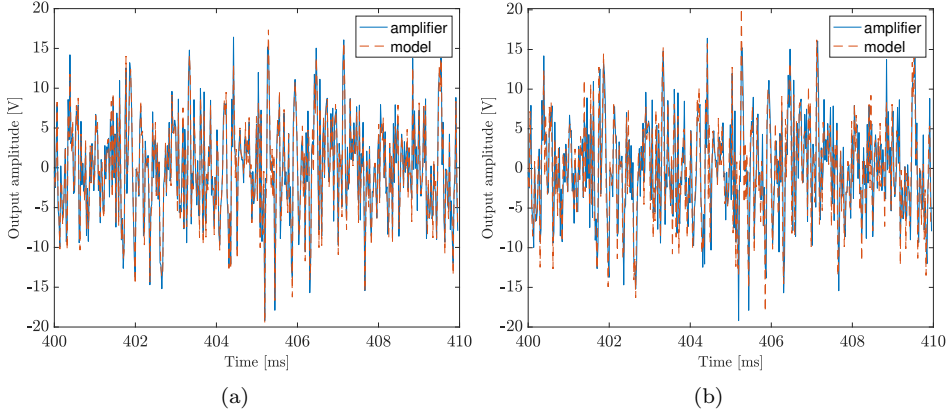
Figure 5: Zoom in time (from 400 ms to 410 ms) of outputs of amplifier and of (a) multiple-variance WGN Volterra model, (b) multiple-variance PPS Volterra model, for the WGN test input.
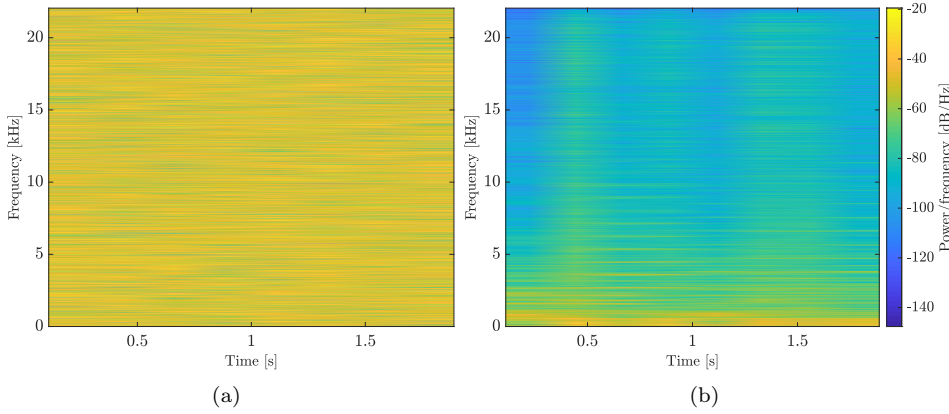


Figure 6: Input spectrogram: (a) WGN, (b) music input.

input, a shorter time segment is shown to make the Figure more readable[3]. The two models show comparable performance in the case of WGN input, as shown also in the NMSE Figures 1.(a) and 2.(a), with a slight difference in favour of the PPS Volterra model.

Finally, Figure 7 shows a comparison between the output spectrograms of the amplifier, of the WGN Volterra model, and of the PPS Volterra model when the input is a WGN or music at the highest variance (whose spectrograms are reported in Figure 6).

---

[3]A Figure with a 1 s duration WGN input, equivalent to Figure 3, is not reported because indistinguishable.
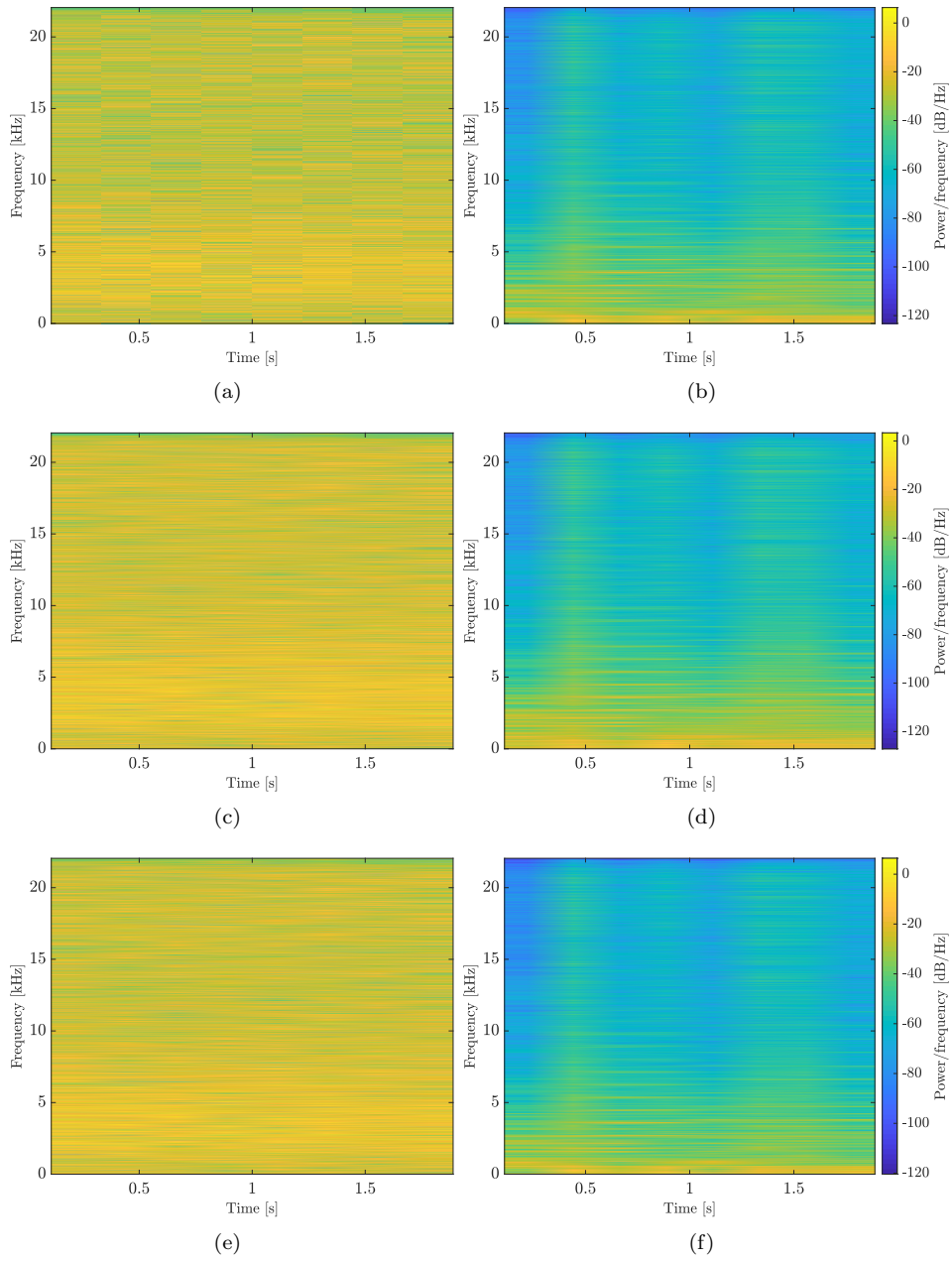
Figure 7: Output spectrogram with WGN (first column) and music input (second column) of (a,b) the amplifier, (c,d) multiple-variance WGN Volterra model, (e,f) multiple-variance PPS Volterra model.

In Figures 7, the spectrograms of Volterra models are closer together but differences can still be seen in favour of PPS Volterra model.
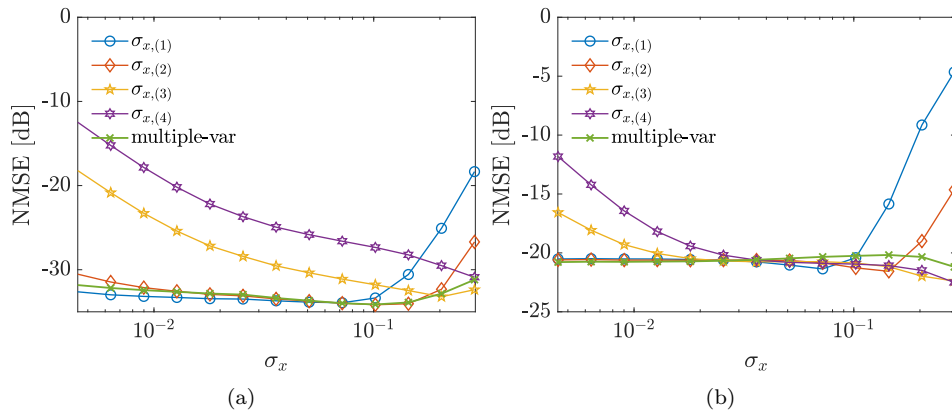
Figure 8: NMSE of Volterra models identified with PPS after basis functions selection according to the Bayesian information criterion. (a) WGN test input, (b) music test input.

The effect of the application of an information criteria to the selection of the most relevant basis function, presented in Sec. 5.2 to reduce the computational complexity of the model, is shown in Figure 8.

This Figure shows the results for the reduced modeled obtained in the single variance and the multiple variance approach using PPSs inputs and applying the Bayesian information criterion. In the single variance approach, the reduction in the number of selected basis functions often improves the estimation performances at the high variances, because it reduces the influence of error terms in the higher order kernels. This behavior is visible especially with $\sigma^2_{x,(1)}$ and $\sigma^2_{x,(2)}$, where the higher order kernels are most affected by the low variance of the estimation. In the multiple variance approach, the reduction in the number of selected basis functions by a factor larger than ten, shown in Table 3, provides almost no alteration in the curves of NMSE. Overall, the multiple variance approach is able to provide the best performances for the various input signal powers, also for the reduced models. Figures equivalent to Figure 4-7 are not shown in this case because the performance is almost identical to the PPS identification case.

## 7. Conclusion

The paper has discussed the identification of nonlinear systems exploiting the Wiener basis functions and WN filters. The Wiener basis functions are a set of polynomial functions that are orthogonal for any white Gaussian input signal of variance $\sigma^2_x$. The

WN filters are a linear combination of the Wiener basis functions. Formulas for converting the WN filters into Volterra filters have been provided. We have shown that WN filters admit PPSs. PPSs guarantee the orthogonality of the basis functions over a period of the sequence and allow the identification of the filter with the cross-correlation method. Exploiting the properties of Wiener basis functions, we have proposed a novel multiple-variance approach, which allows to identify nonlinear systems contrasting the locality of the solution. The proposed multiple-variance approach can be applied also with white Gaussian input signals, but is particularly effective when the input signal is formed by PPSs with different variances. In that case, a procedure for selecting the most relevant basis functions according to an information criterion has been proposed. The optimal choice of the input variances has been theoretically studied and has taken us to simple rules to guide the choice of the variances. The experimental results illustrate the proposed approach and its effectiveness in modeling a real audio device on a wide range of input signal powers.

## Appendix A. MSD in multiple variance identification

When the nonlinear system input $x(n)$ is a Gaussian noise or a PPS of variance $\sigma^2_{x,(i)}$, the output of the unknown system can be expressed in the following form:

$$y(n) = f_0(x)(n) + f_1(x)(n) + \ldots + f_i(x)(n) + \ldots + f_R(x)(n) + \nu(n), \qquad \text{(A.1)}$$

where $\nu(n)$ is the additive Gaussian noise, $f_j(x)(n)$ for $j = 0, ..., R$ are functionals of $x(n)$, specifically are linear combinations of Wiener basis functions of order $j$, and $f_0(x)(n)$ is a constant term. Note that changing the input variance $\sigma^2_{x,(i)}$, the Wiener basis functions and their coefficients change, and also the functionals $f_j(x)(n)$ change. In what follows, for sake of compactness we will indicate $f_j(x)(n)$ as $f_j(n)$. Thus, it is

$$y(n) = f_0(n) + f_1(n) + \ldots + f_i(n) + \ldots + f_R(n) + \nu(n). \qquad \text{(A.2)}$$

We will distinguish between the case of a Gaussian input signal and of a PPS input signal.

Let us assume we want to identify the coefficient $\overline{k}_{i,s}$ of the basis function of order $i$ $w_{i,s}(n)$. The exact value of this coefficient is

$$\overline{k}_{i,s} = \frac{E[y(n)w_{i,s}(n)]}{E[w_{i,s}^2(n)]} = \frac{E[f_i(n)w_{i,s}(n)]}{E[w_{i,s}^2(n)]}, \tag{A.3}$$

while the measured value using time averages is

$$k_{i,s} = \frac{< y(n)w_{i,s}(n) >_L}{< w_{i,s}^2(n) >_L}. \tag{A.4}$$

The measured value is affected by different errors due to noise and to the imperfect orthogonality of the basis functions over a finite time interval.

On a finite time interval of length $L$, $< w_{i,s}^2(n) >_L$ is in general different from the ideal value, but in the following we neglect this error and consider

$$< w_{i,s}^2(n) >_L = E[w_{i,s}^2(n)]. \tag{A.5}$$

In order to estimate the error on $k_{i,s}$, let us decompose $f_i(n)$ into

$$f_i(n) = \overline{k}_{i,s}w_{i,s}(n) + \overline{f}_i(n) + e_i(n), \tag{A.6}$$

and, for any other $l \neq i$, $f_l(n)$ into

$$f_l(n) = \overline{f}_l(n) + e_l(n). \tag{A.7}$$

In (A.6) and (A.7), for all $j$, $\overline{f}_j(n)$ is the component of $f_j(n)$ orthogonal to $w_{i,s}(n)$ over the time interval $L$, i.e., $< \overline{f}_j(n)w_{i,s}(n) >_L = 0$ for all $j$. On the other hand, $e_j(n)$ is the error term caused by the imperfect orthogonality of the basis functions, an thus in general $< e_j(n)w_{i,s}(n) >_L \neq 0$ for all $j$.

Note that in WN filters, the basis functions of order $j$ have power proportional to $\sigma_{x,(i)}^{2j}$, thus, the functionals $f_j(n)$ and the error terms $e_j(n)$ have power proportional to $\sigma_{x,(i)}^{2j}$, since they can be expressed as a linear combination of basis functions of order $j$.

Considering (A.2), (A.6), and (A.7), $k_{i,s}$ is

$$k_{i,s} = \overline{k}_{i,s} + \frac{<e_0(n)w_{i,s}(n)>_L}{<w_{i,s}^2(n)>_L} + \frac{<e_1(n)w_{i,s}(n)>_L}{<w_{i,s}^2(n)>_L} + \ldots$$

$$+ \frac{<e_R(n)w_{i,s}(n)>_L}{<w_{i,s}^2(n)>_L} + \frac{<\nu(n)w_{i,s}(n)>_L}{<w_{i,s}^2(n)>_L}. \tag{A.8}$$

According to the authors' experience, the effect of the error term of order 0 is often negligible, since in most cases $<e_0(n)w_{i,s}(n)>_L \simeq 0$.

We can now estimate the MSD of the measured coefficients of order $i$, which is

$$\mathrm{MSD}_i = \sum_s E\left[(k_{i,s} - \overline{k}_{i,s})^2\right]. \tag{A.9}$$

We assume the error terms $e_j(n)$ to be uncorrelated with each other, with $\nu(n)$, and with $w_{i,s}(n)$. The values of $w_{i,s}(n)$ at different time instants are uncorrelated, because of the orthogonality of the basis functions. In these conditions,

$$\mathrm{MSD}_i = \sum_s \left[ \frac{E[e_0^2(n)w_{i,s}^2(n)]}{E^2[w_{i,s}^2(n)]} + \frac{E[e_1^2(n)w_{i,s}^2(n)]}{E^2[w_{i,s}^2(n)]} + \ldots \right.$$

$$\left. + \frac{E[e_R^2(n)w_{i,s}^2(n)]}{E^2[w_{i,s}^2(n)]} + \frac{E[\nu^2(n)w_{i,s}^2(n)]}{E^2[w_{i,s}^2(n)]} \right]. \tag{A.10}$$

Note that

$$E[w_{i,s}^2(n)] \quad \propto \quad \sigma_{x,(i)}^{2i}, \tag{A.11}$$

$$E[e_j^2(n)w_{i,s}^2(n)] \quad \propto \quad \sigma_{x,(i)}^{2j}\sigma_{x,(i)}^{2i}, \tag{A.12}$$

$$E[\nu^2(n)w_{i,s}^2(n)] \quad \propto \quad \sigma_\nu^2\sigma_{x,(i)}^{2i}. \tag{A.13}$$

Thus,

$$\mathrm{MSD}_i = u_{0,i}\sigma_{x,(i)}^{-2i} + u_{1,i}\sigma_{x,(i)}^{2(1-i)} + u_{2,i}\sigma_{x,(i)}^{2(2-i)} + \cdots + u_{R,i}\sigma_{x,(i)}^{2(R-i)} + u_{i,\nu}\sigma_{x,(i)}^{-2i}\sigma_\nu^2, \tag{A.14}$$

where $u_{i,j}$ for $j = 0, \ldots, R$, and $u_{i,\nu}$ are coefficients of proportionality which depend on the specific kernel $i$, on the unknown system and also on $\sigma_{x,(i)}^2$, for large variations of the

30

input variance.

When the input signal is a PPS, the only stochastic term in (A.2) is $\nu(n)$, all other quantities are deterministic. Thanks to the PPS, we have a perfect orthogonality of the basis functions of the model over a period of the sequence. Nevertheless, in most cases the model underestimates the order and the memory of the unknown system. The basis functions that have been erroneously neglected in the definition of the model and the additive noise $\nu(n)$ are the cause of errors in the estimation of the model coefficients.

Let $\overline{k}_{i,s}$ be the desired value of coefficient of the basis function $w_{i,s}(n)$, given in (A.3) for a Gaussian input of variance $\sigma_{x,(i)}^2$. The corresponding measured coefficients is

$$k_{i,s} = \frac{< y(n)w_{i,s}(n) >_L}{< w_{i,s}^2(n) >_L},\qquad\text{(A.15)}$$

where $L$ is the PPS period or a multiple of the PPS period.

For the construction rules of the PPS, the denominator of (A.15) is a constant equal to the power of the basis function for a Gaussian input of variance $\sigma_{x,(i)}^2$, i.e., $E[w_{i,s}^2(n)]$.

We can decompose $f_i(n)$ into

$$f_i(n) = \overline{k}_{i,s}w_{i,s}(n) + \overline{f}_i(n) + e_i(n),\qquad\text{(A.16)}$$

and, for any other $l \neq i$,

$$f_l(n) = \overline{f}_l(n) + e_l(n).\qquad\text{(A.17)}$$

For all $j$, $\overline{f}_j(n)$ is formed by all basis functions of order $j$ that have been neglected in the model and that are orthogonal with $w_{i,s}(n)$, while $e_j(n)$ is formed by all the basis function of order $j$ neglected in the model that are not orthogonal with $w_{i,s}(n)$. It should be noted that $e_0(n) = 0$, since $f_0(n)$ is a constant and for the construction rules of the PPS $< w_{i,s}(n) >_L = 0$.

Thus, replacing (A.16) and (A.17) in (A.15),

$$k_{i,s} = \overline{k}_{i,s} + \frac{< e_1(n)w_{i,s}(n) >_L}{< w_{i,s}^2(n) >_L} + \frac{< e_2(n)w_{i,s}(n) >_L}{< w_{i,s}^2(n) >_L} + \dots$$

$$+ \frac{<e_R(n)w_{i,s}(n)>_L}{<w_{i,s}^2(n)>_L} + \frac{<\nu(n)w_{i,s}(n)>_L}{<w_{i,s}^2(n)>_L}. \tag{A.18}$$

In this expression, the only stochastic term is $\nu(n)$. Nevertheless, since the PPS has a sample distribution which is similar to a Gaussian distribution, in the following we assume the terms $e_j(n)$ are stochastic, uncorrelated with each other, with $\nu(n)$, and with $w_{i,s}(n)$. Under these assumptions we estimate the MSD of the $i$-th kernel. Estimating (A.9), and considering that also for a PPS input (A.11), (A.12), and (A.13) hold,

$$\text{MSD}_i = u_{1,i}\sigma_{x,(i)}^{2(1-i)} + u_{2,i}\sigma_{x,(i)}^{2(2-i)} + \cdots + u_{R,i}\sigma_{x,(i)}^{2(R-i)} + u_{i,\nu}\sigma_{x,(i)}^{-2i}\sigma_\nu^2, \tag{A.19}$$

where $u_{i,j}$ for $j = 1, \ldots, R$, and $u_{i,\nu}$ are coefficients of proportionality which depends on the specific kernel $i$, on the unknown system and also on $\sigma_{x,(i)}^2$, for large variations of the input variance.

*Appendix A.3. MSD of Volterra coefficients*

In this subsection, we discuss the MSD of the coefficients of the Volterra filter obtained from the WN to Volterra conversion.

We have shown in Section 2 that the two highest order Volterra kernels are equal to the corresponding Wiener kernels. Thus, also the MSDs of these kernels are the same. On the contrary, in the lower order kernels the MSD of the Volterra coefficients of order $i$ depends on the MSD of the WN filter coefficients of the same order and of order $i + 2$, $i + 4$, ..., till the maximum order $R$. As a matter of fact, let us consider the case of a third order Volterra and WN filter, for which the conversion formulas in (21)–(24) hold. According to (21) and (22), the MSD of the kernels of order 3 and 2 of Volterra and WN filter coincides, and is given by (A.14) and (A.19) for a Gaussian and PPS input signal, respectively. Let us consider the MSD of the kernel of order 0,

$$E[(h_0 - \overline{h}_0)^2], \tag{A.20}$$

where $h_0$ is the measured coefficient given by (24), and $\overline{h}_0$ is its true value, with

$$\overline{h}_0 = \overline{k}_0 - \sigma_{x,(0)}^2 \sum_{r=0}^{N-1} \overline{k}_{2,r,r}. \tag{A.21}$$

32

It results that

$$E[(h_0 - \overline{h}_0)^2] = E\left\{[k_0 - \overline{k}_0 - \sigma^2_{x,(0)} \sum_{r=0}^{N-1} (k_{2,r,r} - \overline{k}_{2,r,r})]^2\right\}. \qquad \text{(A.22)}$$

Replacing (A.8) in (A.22), for $i = 0$ and $i = 2$, and using the same assumptions of the previous subsections, it is possible to prove that

$$E[(h_0 - \overline{h}_0)^2] = E[(k_0 - \overline{k}_0)^2] + \sigma^2_{x,(0)} \sum_{r=0}^{N-1} E[(k_{2,r,r} - \overline{k}_{2,r,r})^2] = \text{MSD}_0 + \sigma^2_{x,(0)} \text{MSD}'_2,$$
$$\text{(A.23)}$$

where $\text{MSD}_0$ is given by (A.14) or (A.19), and $\text{MSD}'_2$ is the MSD of the main diagonal of the WN second order kernel, and, similarly to (A.14) and (A.19),

$$\text{MSD}'_2 = \frac{u'_{i,0}}{\sigma^4_{x,(2)}} + \frac{u'_{2,1}}{\sigma^2_{x,(2)}} + \frac{u'_{2,2}}{\sigma^0_{x,(2)}} + u'_{2,3}\sigma^2_{x,(2)} + \frac{u'_{2,\nu}\sigma^2_\nu}{\sigma^4_{x,(2)}}, \qquad \text{(A.24)}$$

with $u'_{2,0}$, ...., $u'_{2,3}$ and $u'_{2,\nu}$ that are coefficients of proportionality.

Posing $\sigma^2_{x,(2)} = r_2\sigma^2_{x,(0)}$ for some constant $r_2$, it is possible to show that (A.22) assume the same form of (A.14) and (A.19); only the coefficients of proportionality $u_{j,0}$ and $u_{0,\nu}$ change.

The same result can be proved also for the MSD of the first order kernel, $\sum_{t=0}^{n-1} E[(h_{1,t} - \overline{h}_{1,t})^2]$.

## References

[1] M. Schetzen, The Volterra and Wiener Theories of Nonlinear Systems, Krieger Publishing Company, Malabar, FL, 2006.

[2] V. J. Mathews, G. L. Sicuranza, Polynomial Signal Processing, Wiley, New York, 2000.

[3] V. Marmarelis, Nonlinear Dynamic Modeling of Physiological Systems, John Wiley & Sons, Hoboken, NJ, USA, 2004.

[4] S. Orcioni, M. Pirani, C. Turchetti, Advances in Lee-Schetzen method for Volterra filter identification, Multidimensional Systems and Signal Processing 16 (3) (2005) 265–284.

[5] M. Pirani, S. Orcioni, C. Turchetti, Diagonal kernel point estimation of n-th order discrete Volterra-Wiener systems, EURASIP Journal on Applied Signal Processing 2004 (12) (2004) 1807–1816.

[6] S. Orcioni, Improving the approximation ability of Volterra series identified with a cross-correlation method, Nonlinear Dynamics 78 (4) (2014) 2861–2869.

[7] W. Silva, Identification of nonlinear aeroelastic systems based on the Volterra theory: Progress and opportunities, Nonlinear Dynamics 39 (2005) 25–62.

[8] J. Pakarinen, D. T. Yeh, A review of digital techniques for modeling vacuum-tube guitar amplifiers, Computer Music Journal 33 (2) (2009) 85–100. doi:10.1162/comj.2009.33.2.85.

[9] L. Tronchin, V. L. Coli, Further investigations in the emulation of nonlinear systems with Volterra series, Journal of the Audio Engineering Society 63 (9) (2015) 671–683. doi:10.17743/jaes.2015.0065.

[10] L. Tronchin, The emulation of nonlinear time-invariant audio systems with memory by means of Volterra series, Journal of the Audio Engineering Society 60 (12) (2012) 984–996.

[11] U. Zölzer, DAFX: Digital Audio Effects, John Wiley & Sons, New York, USA, 2011.

[12] F. Eichas, S. Möller, U. Zölzer, Block-oriented gray box modeling of guitar amplifiers, in: Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17), Edinburgh, UK, 2017, pp. 184–191.

[13] L. Azpicueta-Ruiz, M. Zeller, A. Figueiras-Vidal, J. Arenas-Garcia, W. Kellermann, Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancellation, IEEE Trans. Audio, Speech and Language Processing 19 (11) (2011) 97–110.

[14] T. G. Burton, R. A. Goubran, A generalized proportional subband adaptive second order Volterra filter for acoustic echo cancellation in changing environments, IEEE Trans. Audio, Speech and Language Processing 19 (8) (2011) 2364–2373.

[15] C. Hofmann, C. Huemmer, M. Guenther, W. Kellermann, Significance-aware filtering for nonlinear acoustic echo cancellation, EURASIP Journal on Advances in Signal Processing 2016 (1) (2016) 113.

[16] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, A. Uncini, Full proportionate functional link adaptive filters for nonlinear acoustic echo cancellation, in: Signal Processing Conference (EUSIPCO), 2017 25th European, IEEE, 2017, pp. 1145–1149.

[17] L. Tan, J. Jiang, Adaptive Volterra filters for active noise control of nonlinear processes, IEEE Trans. Signal Processing 49 (8) (2001) 1667–1676.

[18] D. P. Das, G. Panda, Active mitigation of nonlinear noise processes using a novel filtered-s LMS algorithm, IEEE Trans. Speech and Audio Processing 12 (3) (2004) 313–322.

[19] N. V. George, G. Panda, Advances in active noise control: A survey, with emphasis on recent nonlinear techniques, Signal Processing 93 (2) (2013) 363–377.

[20] V. Patel, D. Comminiello, M. Scarpiniti, N. V. George, A. Uncini, Design of hybrid nonlinear spline adaptive filters for active noise control, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE, 2016, pp. 3420–3425.

[21] J. Patra, R. Pal, R. Baliarsingh, G. Panda, Nonlinear channel equalization for QAM signal constellation using artificial neural networks, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 29 (2) (1999) 262–271. doi:10.1109/3477.752798.

[22] J. Kim, K. Konstantinou, Digital predistortion of wideband signals based on power amplifier model with memory, Electronics Letters 37 (2001) 1417–1418.

[23] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, J. Pastalan, A generalized memory polynomial model

for digital predistortion of RF power amplifiers, IEEE Trans. Signal Processing 54 (2006) 3852–3860.

[24] F.-L. Luo (Ed.), Digital Front-End in Wireless Communications and Broadcasting, Cambridge University Press, Cambridge, UK, 2011.

[25] R. Braithwaite, Digital predistortion of an RF power amplifier using a reduced Volterra series model with a memory polynomial estimator, IEEE Transactions on Microwave Theory and Techniques 65 (10) (2017) 3613–3623.

[26] V. Marmarelis, Analysis of physiological systems: The white-noise approach, Springer Science & Business Media, 2012.

[27] A. Carini, S. Cecchi, L. Romoli, S. Orcioni, Perfect periodic sequences for nonlinear Wiener filters, in: 2016 24th European Signal Processing Conference (EUSIPCO), 2016, pp. 1788–1792.

[28] S. Orcioni, S. Cecchi, A. Carini, Multivariance nonlinear system identification using Wiener basis functions and perfect sequences, in: 2017 25th European Signal Processing Conference (EUSIPCO), 2017, pp. 2679–2683.

[29] G. L. Sicuranza, A. Carini, Nonlinear system identification using quasi-perfect periodic sequences, Signal Processing 120 (2016) 174–184.

[30] A. Carini, S. Cecchi, S. Orcioni, Orthogonal LIP nonlinear filters, in: D. Comminello, J. C. Príncipe (Eds.), Adaptive learning methods for nonlinear system modeling, Elsevier, 2018, Ch. 2.

[31] A. Carini, G. L. Sicuranza, Fourier nonlinear filters, Signal Processing 94 (2014) 183–194.

[32] A. Carini, G. L. Sicuranza, Perfect periodic sequences for even mirror Fourier nonlinear filters, Signal Processing 104 (2014) 80–93.

[33] A. Carini, S. Cecchi, L. Romoli, G. L. Sicuranza, Legendre nonlinear filters, Signal Processing 109 (2015) 84–94.

[34] A. Carini, G. L. Sicuranza, A study about Chebyshev nonlinear filters, Signal Processing 122 (2016) 24–32.

[35] C. Antweiler, M. Dörbecker, Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control, Annales des Telecommunications 49 (7-8) (1994) 386–397.

[36] C. Antweiler, M. Antweiler, System identification with perfect sequences based on the NLMS algorithm, International Journal of Electronics and Communications (AEU) 49 (3) (1995) 129–134.

[37] C. Antweiler, Multi-channel system identification with perfect sequences, in: R. Martin, U. Heute, C. Antweiler (Eds.), Advances in Digital Speech Transmission, SPIE, John Wiley & Sons, 2008, pp. 171–198.

[38] C. Antweiler, G. Enzner, Perfect sequence LMS for rapid acquisition of continuous-azimuth head related impulse responses, in: Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09. IEEE Workshop on, 2009, pp. 281–284.

[39] S. Orcioni, A. Terenzi, S. Cecchi, F. Piazza, A. Carini, Identification of Volterra models of tube audio devices using multiple-variance method, Journal of the Audio Engineering Society 66 (10) (2018) 823–838.

[40] H. Akaike, A new look at the statistical model identification, IEEE Trans. Signal Processing 19 (6) (1974) 716–723.

35

[41] G. Schwartz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 416–464.

[42] E. J. Hannan, B. G. Quinn, The determination of the order of an autoregression, J. R. Statist. Soc. B 41 (2) (1979) 190–195.

[43] M. Abramowitz, I. A. Stegun, et al., Handbook of mathematical functions with formulas, graphs, and mathematical tables, Vol. 9, Dover, New York, 1972.

[44] S. Roman, The umbral calculus (1984).

[45] A. Carini, G. L. Sicuranza, Fourier nonlinear filters, Signal Processing 94 (1) (2014) 183–194.

[46] A. Carini, G. L. Sicuranza, A study about Chebyshev nonlinear filters, Signal Processing 122 (2016) 24–32.

[47] W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York, 1976.

[48] G. M. Raz, B. D. V. Veen, Baseband Volterra filters for implementing carrier based nonlinearities, IEEE Trans. Signal Processing 46 (1) (1998) 103–114.

[49] A. Fermo, A. Carini, G. Sicuranza, Low complexity nonlinear adaptive filters for acoustic echo cancellation, European Transactions on Telecommunications 14 (2) (2003) 161–169.

[50] N. V. George, G. Panda, Advances in active noise control: A survey, with emphasis on recent nonlinear techniques, Signal processing 93 (2) (2013) 363–377.

[51] G. L. Sicuranza, A. Carini, A new recursive controller for nonlinear active noise control, in: Proc. ISPA 2013, Trieste, Italy, 2013, pp. 626–631.

[52] A. Carini, G. L. Sicuranza, Recursive even mirror Fourier nonlinear filters and simplified structures, IEEE Trans. Signal Processing 62 (24) (2014) 6534–6544.

[53] A. Carini, S. Cecchi, L. Romoli, G. L. Sicuranza, Legendre nonlinear filters, Signal Processing 109 (2015) 84–94.

[54] D. Zwillinger, Standard mathematical tables and formulae, CRC Press, Boca Raton, 1996.

[55] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical recipes in C : the art of scientific computing, Cambridge University Press, New York, NY, USA, 1995.

[56] A. Carini, Perfect periodic sequences (2016).
URL http://www2.units.it/ipl/res_PSeqs.htm

[57] E. Benjamin, Audio power amplifiers for loudspeaker loads, J. Audio Eng. Soc 42 (9) (1994) 670–683.

[58] R. J. Martin, Volterra system identification and Kramer's sampling theorem, IEEE Transactions on Signal Processing 47 (11) (1999) 3152–3155.