CHAPTER

ORTHOGONAL LIP NONLINEAR FILTERS

2

Alberto Carini*, Stefania Cecchi†, Simone Orcioni†

*University of Urbino Carlo Bo, DiSPeA, Urbino, Italy †Università Politecnica delle Marche, DII, Ancona, Italy

CONTENTS

Adaptive Learning Methods for Nonlinear System Modeling. DOI: 10.1016/B978-0-12-812976-0.00003-8

CHAPTER POINTS

- The chapter provides an overview of orthogonal linear-in-the-parameters (LIP) nonlinear filters.
- It presents different classes of orthogonal LIP nonlinear filters, under the unified framework of functional link polynomial (FLiP) filters.
- It discusses how perfect periodic sequences (PPSs) for FLiP filters can be developed and how they can be used for system identification using the cross-correlation method.
- Multiple-variance identification techniques, which allow one to contrast the problem of locality of the solution, are also discussed.

2.1 INTRODUCTION

The history of linear-in-the-parameters (LIP) nonlinear filters starts at the end of the 19th century, when the Italian mathematician Vito Volterra developed his *functional power series*, later called Volterra series, as an extension to functionals of the Taylor series expansion [1,2]. In 1910 Fréchet showed that a sum of *homogeneous functionals of degree n* can approximate, arbitrarily well, any continuous real-valued functional, defined on compact sets as, for example, the set of functions defined over intervals [3]. Fréchet obtained his results as a generalization of the Weierstrass theorem, which states that the set of polynomial functions is complete. Fréchet showed in a similar manner that the set of Volterra functionals is complete.

The restriction to a compact domain is an important limitation when the input signals are defined over an infinite (or semiinfinite) time interval (*i.e.*, periodic stimuli). Years later, this gave rise to many works on the so-called fading memory systems [4,5]. However, this is not a limitation in practical cases when most systems are implemented in the form of discrete-time, finite-memory filters.

A key factor for spreading the use of polynomial series was the work of Wiener on Brownian motion linear transformation, on higher degree functionals of Brownian motion and on their orthogonalization [6], published during his years at the MIT Research Laboratory of Electronics. Along with the work of his colleagues [7], Wiener continued to develop his theories by combining Volterra series and his previous works, finally defining the so called Wiener series.

In real applications, the implementation of Wiener's method was difficult due to the use of Brownian motion. A further step in the usability of Volterra and Wiener series was due to the work of Lee and Schetzen [8], who developed a method that uses white Gaussian noise (WGN) as inputs and cross-correlation for parameter estimation of Wiener series. This was possible because, differently from what happens in the Volterra series, the Wiener functionals of any order are orthogonal to each other when the input is a WGN. Exploiting the relationship between Wiener and Volterra coefficients, it is also possible to obtain an equivalent Volterra representation of Wiener series. A rigorous proof on the range of validity of this method for continuous-time time-invariant nonlinear systems can be found in [9]. The same paper showed unavoidable difficulties in the continuous-time identification of terms with at least two equal indices, called diagonal elements, of kernels of order greater than or equal to two. Fortunately, these theoretical difficulties disappear in most real applications concerning discrete-time systems [10], but a great inaccuracy in the identification of diagonal elements still remains [11]. Effective solutions that overcome this problem have been proposed in the literature, for series up to the third order in [12] and for a generic order in [13], where a comparison between the two methods is also provided.

While in analytical power series (infinite sum of elements) the identification with cross-correlation is independent of the input variance, this is no more true with truncated power series, where the approximation error depends on the variance used in the identification: a Volterra series is optimal only for inputs with variances in a neighborhood of that used for identification [14]. An improved cross-correlation method for nonlinear system identification based on multiple variances has been proposed for Wiener–Volterra series and Gaussian noise in [14]: low input variances are used to identify lower-order Wiener kernels, while the input variance is gradually increased for higher-order kernels.

The truncated Volterra filter, which derives from a double truncation with respect to the order and memory of the Volterra series, has been one of the first LIP nonlinear filters to be used in theory and applications. It is still today one of the most popular LIP filters. Its success derives from (i) the simplicity of the LIP input–output relationship and (ii) the property of being a universal approximator according to the Stone–Weierstrass theorem. On the other hand, the main limitations of Volterra filters are (i) the very large number of coefficients involved, which increases exponentially with the filter order and geometrically with the memory length, and (ii) the difficulties in coefficient estimation. For example, the gradient descent adaptation algorithms provide for any input signal slow convergence speed, mainly due to the poor conditioning of the input autocorrelation matrix. The algorithms used to overcome these convergence problems have often a high computational complexity.

It is possible to facilitate the nonlinear system identification with the use of Wiener nonlinear (WN) filters, which derive from the double truncation of the Wiener series. The WN filter is a first example of nonlinear filters with orthogonal basis functions, in particular, orthogonal for a white Gaussian input signal. The orthogonality of the basis functions allows the efficient identification of the filter coefficients with the cross-correlation method, as in [8]. Moreover, the orthogonality guarantees fast convergence of gradient descent algorithms. Deterministic periodic signals, called perfect periodic sequence (PPS), that guarantee the orthogonality of the basis functions on a finite period can also be developed. Expressing the WN filter as a linear combination of basis functions and using a PPS input signal, problems in the estimation of the kernel diagonal points can be avoided [15].

To reduce the model complexity, different block structured models have been proposed in the literature. They consist of a nonlinear memoryless block preceded or followed by a linear system and can often be considered subclasses of the Volterra or Wiener model. To this class belong the Hammerstein filters [16–20], the memory polynomial filters [21,22] and functional link artificial neural networks (FLANNs) based on polynomial expansions [23–30]. All these filters involve nonlinearities of the input samples at the same time instant, without the so-called cross-terms, *i.e.*, products of samples at different instants.

Block structured models based on nonpolynomial functions have also been proposed, such as FLANNs based on orthogonal trigonometric expansions of the input samples [23] and the radial basis function networks [31].

Among the block structured models involving also cross-terms, the generalized memory polynomial filters [22], the Wiener systems, the Hammerstein–Wiener model and generalized FLANN filters [32] can be mentioned.

Many of the block structured models previously considered are LIP filters. Very few of them, *e.g.*, the FLANN filters based on orthogonal expansions, have orthogonal basis functions. In most cases, the reduction in the computational complexity obtained from the use of a block structure is paid with the loss of the property of universal approximation. Thus, the block structure models cannot arbitrarily well approximate any nonlinear system.

Recently, different families of LIP nonlinear filters have been proposed considering a set of basis functions that can arbitrarily well approximate the N-dimensional nonlinear function defining a nonlinear system of memory N. These LIP filters are universal approximators for discrete-time, causal, time-invariant, finite-memory, continuous nonlinear systems and have orthogonal basis functions for appropriate distributions of the input signal. To this class of LIP nonlinear filters belong the Fourier nonlinear (FN) filters [33], the even mirror FN (EMFN) filters [33,34], the Legendre nonlinear (LN) filters [35,36] and the Chebyshev nonlinear (CN) filters [37].

It is also possible to develop perfect periodic sequences (PPSs) that guarantee the orthogonality of the basis functions of EMFN, LN and CN filters on a finite period [38–40,36,37]. Using PPSs as input signals, the filter coefficients can be estimated with the cross-correlation method. It is worth noting that also adaptive identification algorithms can benefit from the use of PPS inputs since the orthogonality of the basis functions guarantees a fast convergence of gradient descent algorithms [32].

Volterra, WN, EMFN, LN and CN filters belong to the general class of functional link polynomial (FLiP) filters [41], which include also many of the block structured models previously mentioned.

In what follows, this chapter aims to:

- review the theory of the principal LIP nonlinear filters that are universal approximators according to the Stone–Weierstrass theorem;
- present Volterra, WN, EMFN, LN and CN filters under the unified framework of FLiP filters;
- provide an overview of orthogonal LIP nonlinear filters and in particular of orthogonal FLiP filters;
- discuss some recent identification techniques for orthogonal LIP nonlinear filters.

The rest of the chapter is organized as follows. Section 2.2 provides a review of LIP nonlinear filters, of the Stone–Weierstrass theorem and of the Volterra and Wiener theory, discusses FLiP filters and orthogonal FLiP filters and provides simplified structures that can be used to reduce the computational complexity of the filters. Classical identification methods, like minimum mean square estimation and the cross-correlation method, are also introduced in this section. Section 2.3 presents an overview of new identification techniques that can be applied to orthogonal LIP nonlinear filters. Specifically, identification based on PPSs and the multiple-variance method is discussed. Section 2.4 provides experimental results that illustrate the ability of the discussed nonlinear filters and identification techniques to model the nonlinearities encountered in real nonlinear systems. Concluding remarks are provided in Section 2.5.

The following notation is used throughout the chapter. Intervals are represented with square brackets, $\mathbb{N}$ is the set of natural numbers, $\mathbb{R}$ is the set of real numbers, $\mathbb{R}^+$ is the set of positive real numbers, $\mathbb{R}_1$ is the unit interval $[-1, +1]$, $< x(n) >_L$ indicates time average over L successive samples of $x(n)$, $\binom{n}{m}$ denotes the number of combinations of n things taken m at a time, δ_{ij} is equal to 1 when $i = j$ and is equal to 0 otherwise, $E(\cdot)$ indicates mathematical expectation, $\mathcal{N}(m_x, \sigma_x^2)$ is the normal distribution with mean m_x and variance σ_x^2 and $(Hx)(n)$ is an operator acting on the input sequence. Sets are represented with curly brackets and intervals with square brackets, while the following convention for brackets $\{[(\ldots)]\}$ is used elsewhere.

2.2 LIP NONLINEAR FILTERS

In this section, we first discuss the approximation of nonlinear filters using LIP nonlinear filters and we review the Stone–Weierstrass theorem reporting the classical Volterra and Wiener theories. Then we introduce the class of FLiPs, which includes many families of nonlinear filters currently used in common practice. Some families of orthogonal FLiP filters are also discussed. Simplified structures for an efficient implementation of FLiP filters are eventually presented.

2.2.1 NONLINEAR FILTERS AND STONE–WEIERSTRASS THEOREM

We are interested in developing families of LIP nonlinear filters capable of arbitrarily well approximating any discrete-time, time-invariant, finite-memory, causal, continuous nonlinear system. The input–output relationship of the nonlinear system can be expressed in the following form:

$$y(n) = f[x(n), x(n-1), \dots, x(n-N+1)], \tag{2.1}$$

where $x(n)$ is the input signal, which is assumed to belong to the unit interval, *i.e.*, $x(n) \in \mathbb{R}_1$, $y(n) \in \mathbb{R}$ is the output signal and N is the system memory length. The function f in Eq. (2.1) can be interpreted as a multidimensional function over the space $\mathbb{R}_1^N$. Each delayed input sample corresponds to a different dimension. Taking into account this notation, it is possible to expand the function f in Eq. (2.1) with a series of basis functions f_i,

$$f[x(n), x(n-1), \dots, x(n-N+1)] = \sum_{i=1}^{+\infty} c_i f_i[x(n), x(n-1), \dots, x(n-N+1)], \tag{2.2}$$

where $c_i \in \mathbb{R}$ and each f_i is a continuous function from $\mathbb{R}_1^N$ to $\mathbb{R}$. Every choice of the set of basis functions f_i leads to a different family of LIP nonlinear filters, which can be used to identify or to approximate the nonlinear systems in Eq. (2.1). For compactness, in what follows $f_i[x(n), x(n-1), \dots, x(n-N+1)]$ will be indicated as $f_i(n)$.

We are particularly interested in sets of basis functions that satisfy the conditions of the Stone–Weierstrass theorem [42].

Theorem 1 (Stone–Weierstrass Theorem)**.** *Let $\mathcal{A}$ be an algebra of real continuous functions on a compact set K. If (i) $\mathcal{A}$ separates points on K and if (ii) $\mathcal{A}$ vanishes at no point of K, then the uniform closure $\mathcal{B}$ of $\mathcal{A}$ consists of all real continuous functions on K.*

According to this theorem, every algebra of real continuous functions on the compact $\mathbb{R}_1^N$ that separates points and vanishes at no point is able to arbitrarily well approximate the function f in Eq. (2.1).

Definition 1. A family $\mathcal{A}$ of real functions is said to be an *algebra* if $\mathcal{A}$ is closed under addition, multiplication and scalar multiplication, *i.e.*, if (i) $f + g \in \mathcal{A}$, (ii) $f \cdot g \in \mathcal{A}$ and (iii) $cf \in \mathcal{A}$, for all $f \in \mathcal{A}$, $g \in \mathcal{A}$ and for any real constant c.

For example, the set of basis functions formed by all possible products of the samples $x(n)$, $x(n-1)$, ..., $x(n-N+1)$ form an algebra that satisfies all conditions of the Stone–Weierstrass theorem. The linear combination of these basis functions defines the class of truncated Volterra filters.

2.2.2 THE VOLTERRA AND WIENER THEORY

Before introducing the class of FLiP filters, which encompasses many of the LIP filters currently used in practice, we first review the classical theory of discrete-time Volterra and Wiener filters.

The Volterra series can be defined as a summation of nonlinear Volterra operators of order p, as

$$y(n) = (Vx)(n) = \sum_{p=0}^{P} (H_p x)(n), \tag{2.3}$$

where the Volterra operator of order p is defined as

$$(H_p x)(n) = \sum_{\tau_1=0}^{M} \sum_{\tau_2=0}^{M} \cdots \sum_{\tau_p=0}^{M} h_{p,\tau_1,\tau_2\ldots,\tau_p} x(n-\tau_1)x(n-\tau_2)\cdots x(n-\tau_p), \tag{2.4}$$

with $P = M = +\infty$ in the most general formulation and $(H_0 x)(n) = h_0$ being a constant term. In Eq. (2.4), the summations start from zero since only causal systems are considered. The set of coefficients $h_{p,\tau_1,\tau_2,\ldots,\tau_p}$ is the pth-order Volterra kernel.

Truncated Volterra filters derive from the double truncation of the Volterra series in Eq. (2.3). Specifically, they are obtained by truncating the order of the series, setting $P = K < +\infty$, and by limiting the memory of the Volterra operators in (2.4) to N samples, *i.e.*, considering $M = N - 1 < +\infty$.

Without loss of generality, the kernel $h_{p,\tau_1,\ldots,\tau_p}$ can always be considered symmetrical. In fact, for the commutativity of the multiplication it is always possible to symmetrize the kernel without changing $(H_p x)$. It is also possible to reduce the summation in Eq. (2.4) to an upper triangular form, considering

$$(H_p x)(n) = \sum_{\tau_1=0}^{M} \sum_{\tau_2=\tau_1}^{M} \cdots \sum_{\tau_p=\tau_{p-1}}^{M} h_{p,\tau_1,\tau_2\ldots,\tau_p} x(n-\tau_1)x(n-\tau_2)\cdots x(n-\tau_p), \tag{2.5}$$

for a proper choice of the coefficients $h_{p,\tau_1,\tau_2\ldots,\tau_p}$, obtaining the so called *triangular representation* of Volterra filters.

Once established with the Stone–Weierstrass theorem that any discrete-time, causal, time-invariant, continuous system can be arbitrarily well approximated with a truncated Volterra series, the problem of estimating its parameters comes up. To this end we will present the Wiener's orthogonal classic representation and its estimation by means of the cross-correlation method by Lee and Schetzen [8]. The WN filters, which derive from the double truncation of the Wiener series, represent one of the first examples of nonlinear filters with orthogonal basis functions proposed in the literature.

2.2.2.1 Orthogonalization

In order to allow identification by a cross-correlation method, the Volterra series must be rearranged in terms of orthogonal operators:

$$y(n) = \sum_{p=0}^{P} (H_p x)(n) \equiv \sum_{p=0}^{P} (G_p x)(n). \tag{2.6}$$

These operators, commonly known as Wiener G–functionals, are orthogonal to each other and, also, to Volterra operators of lower orders if the input $x(n)$ is a stationary white-noise signal with zero mean and variance σ_x^2, *i.e.*,

$$\begin{aligned} E[(H_p x)(n)(G_q x)(n)] &= 0, \quad p < q, \\ E[(G_p x)(n)(G_q x)(n)] &= 0, \quad p \neq q, \end{aligned} \tag{2.7}$$

whenever $(H_p x)(n)$ is an arbitrary homogeneous Volterra operator and $x(n)$ is a stationary white noise with zero mean and variance σ_x^2, *i.e.*, $x(n) \in \mathcal{N}(0, \sigma_x^2)$. By means of the Gram–Schmidt orthogonalization method, the following expressions for the first Wiener G–functionals can be obtained:

$$\begin{aligned} (G_0 x)(n) &= k_0, \\ (G_1 x)(n) &= \sum_{\tau_1=0}^{M} k_{1,\tau_1} x(n-\tau_1), \\ (G_2 x)(n) &= \sum_{\tau_1=0}^{M} \sum_{\tau_2=0}^{M} k_{2,\tau_1,\tau_2} x(n-\tau_1) x(n-\tau_2) - \sigma_x^2 \sum_{\tau_1=0}^{M} k_{2,\tau_1,\tau_1}, \\ (G_3 x)(n) &= \sum_{\tau_1=0}^{M} \sum_{\tau_2=0}^{M} \sum_{\tau_3=0}^{M} k_{3,\tau_1,\tau_2,\tau_3} x(n-\tau_1) x(n-\tau_2) x(n-\tau_3) + \\ &\quad - 3\sigma_x^2 \sum_{\tau_1=0}^{M} \sum_{\tau_2=0}^{M} k_{3,\tau_1,\tau_2,\tau_2} x(n-\tau_1). \end{aligned} \tag{2.8}$$

As can be seen, the Wiener G–functionals are nonhomogeneous operators. The order P and the memory M are infinite in the case of the Wiener series, while they are truncated to finite values, $P = K < +\infty$ and $M = N - 1 < +\infty$, in the case of a WN filter. For simplicity, the Wiener kernels k_p have been assumed symmetric, but the triangular representation is also possible.

From the expression of the Wiener G–functionals, it is not immediate to determine the basis functions of WN filters. In Section 2.2.3, we will derive an alternative representation of the WN filters under the framework of FLiP filters. Moreover, it will be shown that the basis functions are orthogonal for $x(n) \in \mathcal{N}(0, \sigma_x^2)$.

The Volterra kernels of a truncated series can be derived from the Wiener kernels by means of simple equivalence formulas. In the following, the Wiener–Volterra formulas valid in the case of a third-order Volterra system are reported:

$$\begin{aligned} h_3 &= k_3, \\ h_2 &= k_2, \\ h_1 &= k_1 - 3\sigma_x^2 \sum_{\tau_2=0}^{M} k_{3,\tau_1,\tau_2,\tau_2}, \\ h_0 &= k_0 - \sigma_x^2 \sum_{\tau_1=0}^{M} k_{2,\tau_1,\tau_1}. \end{aligned} \tag{2.9}$$

The reader is referred to [14] for higher-order formulas.

2.2.2.2 Cross-Correlation

The Wiener kernels can be estimated with the cross-correlation method. Recalling that every Volterra functional is orthogonal to all Wiener functionals of greater order and considering the following Volterra functional:

$$(H_p^* x)(n) = \prod_{j=1}^{p} x(n-\tau_j), \tag{2.10}$$

it is possible to write

$$E\left[y(n)(H_p^* x)(n)\right] = E\left[\sum_{q=0}^{P} (G_q x)(n)(H_p^* x)(n)\right]. \tag{2.11}$$

If $x(n) \in \mathcal{N}(0, \sigma_x^2)$, $\tau_1 \neq \tau_2 \neq \ldots \neq \tau_p$, *i.e.*, excluding the so called diagonal terms,

$$E\left[y(n)\prod_{j=1}^{p} x(n-\tau_j)\right] = E\left[(G_p x)(n)\prod_{j=1}^{p} x(n-\tau_j)\right] = p!\sigma_x^{2p} k_{p,\tau_1,\ldots,\tau_p}$$

and thus

$$k_{p,\tau_1,\ldots,\tau_p} = \frac{E\left\{y(n)x(n-\tau_1)\cdots x(n-\tau_p)\right\}}{p!\sigma_x^{2p}}.$$

To consider also the diagonal elements, Lee and Schetzen proposed to subtract from the output $y(n)$ the Wiener operators of lower orders that, in correlation with the operator of Eq. (2.10), give raise to impulses. However, if in Eq. (2.11) we write explicitly the expectation of the product between the G functionals and products of delay input samples, we obtain

$$\begin{aligned} E[y(n)\,x(n-\tau_1)x(n-\tau_2)] &= 2!\sigma_x^4 k_{2,\tau_1,\tau_2} - \sigma_x^2 k_0 \delta_{\tau_1\tau_2}, \\ E[y(n)\,x(n-\tau_1)x(n-\tau_2)x(n-\tau_3)] &= 3!\sigma_x^6 k_{3,\tau_1,\tau_2,\tau_3} \\ &\quad -\sigma_x^4\left(k_{1,\tau_1}\delta_{\tau_2\tau_3} + k_{1,\tau_2}\delta_{\tau_1\tau_3} + k_{1,\tau_3}\delta_{\tau_1\tau_2}\right), \end{aligned} \tag{2.12}$$

where $\delta_{\tau_i\tau_j} \triangleq \delta(\tau_i - \tau_j)$ is the unitary impulse delayed by $\tau_i - \tau_j$.

By properly considering these impulses, in [13] the authors proposed a formula for an efficient implementation of the Lee–Schetzen method for the estimation of Wiener series of any order. The formula simplifies to the classic Lee–Schetzen one if nondiagonal terms of the Wiener kernel are identified. The reader is referred to [13] for the implementation details.

In the cross-correlation method described above, the nonlinear kernels are estimated computing the cross-correlation between the unknown system output and products of input samples and solving a system of linear equations as in Eq. (2.12). In the next section, we introduce a large class of nonlinear systems, the FLiP filters, which includes many subclasses of filters with orthogonal basis functions. In the orthogonal FLiP filters, the nonlinear kernels can be directly estimated with the aforementioned cross-correlation method.

2.2.3 FLIP FILTERS AND ORTHOGONAL FLIP FILTERS

FLiP filters are a subclass of LIP filters whose basis functions are polynomials of nonlinear expansions of delayed input samples that follow the constructive rule of the triangular representation of Volterra filters. All families of FLiP nonlinear filters are universal approximators. As mentioned in the introduction, Volterra, WN, EMFN, LN and CN filters belong to the general class of FLiP filters. EMFN filters, like FN, are based on trigonometric function expansions of the input signal samples and do not include a linear term among the basis functions. EMFN filters should be preferred to FN filters, because they often provide a much more compact representation of nonlinear systems. Nevertheless, also FN filters have been found useful for modeling nonlinear systems [43]. LN, CN and WN filters are based on Legendre, Chebyshev and Hermite polynomial expansions, respectively, and have a linear term formed by the first-order basis functions. The basis functions of FN, EMFN, LN, CN and WN filters form algebras that satisfy all the requirements of the Stone–Weierstrass approximation theorem.

Let us consider an ordered set of one-dimensional basis functions that satisfies all requirements of the Stone–Weierstrass theorem on $\mathbb{R}_1$,

$$\{g_0[x(n)], g_1[x(n)], g_2[x(n)], g_3[x(n)], \ldots\}, \tag{2.13}$$

where $g_0[x(n)]$ is the basis function of order 0, usually equal to 1, $g_{2k}[x(n)]$ is an even basis function of order $2k$, $g_{2k+1}[x(n)]$ is an odd basis function of order $2k+1$, for any k in $\mathbb{N}$.

This set of basis functions can arbitrarily well approximate the system in Eq. (2.1) for $N = 1$. To develop a set of basis functions for the more general case of $N > 1$, we first write the basis functions $g_0, g_1, g_2, \ldots$, for $x(n), x(n-1), x(n-2), \ldots$,

$$\begin{gathered}
g_0[x(n)], g_0[x(n-1)], g_0[x(n-2)], \ldots \\
g_1[x(n)], g_1[x(n-1)], g_1[x(n-2)], \ldots \\
g_2[x(n)], g_2[x(n-1)], g_2[x(n-2)], \ldots \\
\vdots
\end{gathered}$$

Then the functions g_j of different variables are multiplied with each other in any possible manner, avoiding repetitions, as in the triangular representation of Volterra filters. The resulting basis functions f_i are formed by cross-products of the functions g_j. The basis functions and their linear combinations satisfy all requirements of the Stone–Weierstrass theorem on $\mathbb{R}_1^N$ and thus FLiP filters are universal approximators. Indeed, by construction the set is an algebra since it is closed under addition, multiplication and scalar multiplication, *i.e.*, it satisfies properties (i), (ii) and (iii) of Definition 1. It is closed under multiplication because any product of the functions $g_j[x(n)]$ can be expanded in a linear combination of the functions $g_j[x(n)]$ and, consequently, any product of the functions f_i can be expanded in a linear combination of the functions f_i. The set of basis functions vanishes at no point, thanks to $g_0[x(n)] = 1$, satisfying property (ii) of Theorem 1. As required in property (i) of Theorem 1, the set separates points, because two different points must have at least one different coordinate $x(n-k)$ and the functions $g_j[x(n-k)]$ separates it. Thus, the algebra of FLiP basis functions can arbitrarily well approximate the nonlinear system in Eq. (2.1).

The order of each N-dimensional basis function $f_i(n)$ is defined as the sum of the orders of the constituent one-dimensional basis functions. Since any one-dimensional basis function $g_{2k}[x(n)]$ is

Table 2.1 Basis functions $f_i(n)$ of FLiP filters

Order 0: $$g_0[x(n)] = 1.$$
Order 1: $$g_1[x(n)], \ldots, g_1[x(n-N+1)].$$
Order 2: $$\begin{gathered} g_2[x(n)], \ldots, g_2[x(n-N+1)], \\ g_1[x(n)]g_1[x(n-1)], \ldots, g_1[x(n-N+2)]g_1[x(n-N+1)], \\ g_1[x(n)]g_1[x(n-2)], \ldots, g_1[x(n-N+3)]g_1[x(n-N+1)], \\ \vdots \\ g_1[x(n)]g_1[x(n-N+1)]. \end{gathered}$$
Order 3: $$\begin{gathered} g_3[x(n)], \ldots, g_3[x(n-N+1)], \\ g_2[x(n)]g_1[x(n-1)], \ldots, g_2[x(x-N+2)]g_1[x(n-N+1)], \\ \vdots \\ g_2[x(n)]g_1[x(n-N+1)], \\ g_1[x(n)]g_2[x(n-1)], \ldots, g_1[x(n-N+2)]g_2[x(n-N+1)], \\ \vdots \\ g_1[x(n)]g_2[x(n-N+1)], \\ g_1[x(n)]g_1[x(n-1)]g_1[x(n-2)], \ldots \\ g_1[x(n-N+3)]g_1[x(n-N+2)]g_1[x(n-N+1)], \\ \vdots \\ g_1[x(n)]g_1[x(n-N+2)]g_1[x(n-N+1)]. \end{gathered}$$

even and any basis function $g_{2k+1}[x(n)]$ is odd, for any $k \in \mathbb{N}$, by construction all N-dimensional basis functions $f_i(n)$ of even order are even and all N-dimensional basis functions $f_i(n)$ of odd order are odd.

A FLiP filter of order K, memory N, is originated by the linear combination of all basis functions f_i having orders from 0 to K, which depends on $x(n), x(n-1), \ldots, x(n-N+1)$.

For example, Table 2.1 summarizes the basis functions f_i of a FLiP filter of order $K = 3$ and memory N. In particular, in Table 2.1 $f_0(n) = g_0[x(n)]$, $f_1(n) = g_1[x(n)]$, $f_2(n) = g_1[x(n-1)]$, ..., $f_N(n) = g_1[x(n-N+1)]$, $f_{N+1}(n) = g_2[x(n)]$, and so on for the other basis functions of order 2 and 3, till $f_{N_{\mathrm{T}}-1}(n) = g_1[x(n)]g_1[x(n-N+2)]g_1[x(n-N+1)]$, where N_{T} is the total number of basis functions.

Any choice of the set of one-dimensional basis functions $g_j[x(n)]$ defines a different family of FLiP filters. Many families of LIP nonlinear filters studied in the literature or used in applications belong to the class of FLiP filters. For example, the truncated Volterra filter is a FLiP filter obtained using as $g_j[x(n)]$ the monomials

$$\left\{1, x(n), x^2(n), x^3(n), \ldots\right\}. \tag{2.14}$$

In what follows, we discuss four families of nonlinear filters with orthogonal basis functions. Specifically, we consider:

1. the EMFN filters [33],
2. the LN filters [36],
3. the CN filters [37] and
4. the WN filters [15].

For each of these filters, the expression of the one-dimensional basis functions g_j is provided and their orthogonality properties are discussed. At the end of the section, it is also explained how to develop a family of polynomial FLiP filters whose basis functions are orthogonal for any desired distribution of the input samples.

2.2.3.1 Even-Mirror Fourier Nonlinear Filters

EMFN filters are defined using as $g_j[x(n)]$ the following trigonometric functions:

$$\left\{1, \sin\left[\frac{1}{2}\pi x(n)\right], \cos[\pi x(n)], \sin\left[\frac{3}{2}\pi x(n)\right], \ldots, \cos[j\pi x(n)], \sin\left[\frac{2j+1}{2}\pi x(n)\right], \ldots\right\}, \tag{2.15}$$

where 1 is the basis function of order 0, $\sin[\frac{2j+1}{2}\pi x(n)]$ is a basis function of order $2j+1$, and $\cos[j\pi x(n)]$ is a basis function of order $2j$, with $j \in \mathbb{N}$. Note that all even-basis functions are even and all odd-basis functions are odd. The basis functions $g_j[x(n)]$ and, consequently, the corresponding FLiP basis functions f_i are even mirror-symmetric, which explains the name of the filters.

The basis functions g_j and their linear combinations form an algebra that satisfies all the requirements of the Stone–Weierstrass theorem on the compact $\mathbb{R}_1$. Indeed, the set is complete under addition, scalar multiplication and multiplication, since

$$\sin\left(\frac{2i+1}{2}\pi x\right)\cos(j\pi x) = \frac{1}{2}\sin\left[\frac{2(i-j)+1}{2}\pi x\right] + \frac{1}{2}\sin\left[\frac{2(i+j)+1}{2}\pi x\right], \tag{2.16}$$

$$\cos(i\pi x)\cos(j\pi x) = \frac{1}{2}\cos\left[(i-j)\pi x\right] + \frac{1}{2}\cos\left[(i+j)\pi x\right], \tag{2.17}$$

$$\sin\left(\frac{2i+1}{2}\pi x\right)\sin\left(\frac{2j+1}{2}\pi x\right) = \frac{1}{2}\cos\left[(i-j)\pi x\right] + \frac{1}{2}\cos\left[(i+j+1)\pi x\right], \tag{2.18}$$

it separates points (*e.g.*, with $\sin\left[\frac{1}{2}\pi x(n)\right]$) and it vanishes at no point (*e.g.*, for $g_0 = 1$).

The basis functions are orthogonal for a white uniform distribution in $\mathbb{R}_1$, *i.e.*,

$$\int_{-1}^{+1} \cdots \int_{-1}^{+1} f_i[x(n), \ldots, x(n-N+1)] f_l[x(n), \ldots, x(n-N+1)] \cdot \\ \cdot p[x(n), \ldots, x(n-N+1)] \mathrm{d}x(n) \cdots \mathrm{d}x(n-N+1) = 0, \tag{2.19}$$

where $i \neq l$ and $p[x(n), \ldots, x(n-N+1)]$ is the probability density function of the N-tuple $[x(n), \ldots, x(n-N+1)]$, equal to the constant $\dfrac{1}{2^N}$ for a white uniform distribution of the input signal

Table 2.2 Legendre polynomials

$\text{leg}_0(x) = 1$
$\text{leg}_1(x) = x$
$\text{leg}_2(x) = \frac{1}{2}(3x^2 - 1)$
$\text{leg}_3(x) = \frac{1}{2}x(5x^2 - 3)$
$\text{leg}_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$
$\text{leg}_5(x) = \frac{1}{8}x(63x^4 - 70x^2 + 15)$

$x(n)$ in $\mathbb{R}_1$. Indeed, by constructions, the basis functions f_i and f_l are products of one-dimensional basis functions g_j. For any $j \in \mathbb{N}$, it is

$$\int_{-1}^{+1} \sin\left(\frac{2j+1}{2}\pi x\right) \mathrm{d}x = 0 \tag{2.20}$$

and for any $j \in \mathbb{N}^+$ it is

$$\int_{-1}^{+1} \cos(j\pi x)\, \mathrm{d}x = 0. \tag{2.21}$$

Since the product of two different one-dimensional basis functions g_j, according to Eqs. (2.16), (2.17) and (2.18), is a linear combination of basis functions g_j different from g_0, it immediately follows Eq. (2.19).

2.2.3.2 Legendre Nonlinear Filters

LN filters are defined by using as $g_j[x(n)]$ the Legendre polynomials, so

$$\text{leg}_0[x(n)], \text{leg}_1[x(n)], \text{leg}_2[x(n)], \text{leg}_3[x(n)], \ldots. \tag{2.22}$$

The Legendre polynomials are obtained from the recursive relation

$$\text{leg}_{j+1}[x(n)] = \frac{2j+1}{j+1} x(n)\text{leg}_j[x(n)] - \frac{j}{j+1}\text{leg}_{j-1}[x(n)], \tag{2.23}$$

where $\text{leg}_0[x(n)] = 1$, $\text{leg}_1[x(n)] = x(n)$ and $j \in \mathbb{N}$ is the order of the basis function.

Table 2.2 summarizes the first Legendre polynomials.

The product of two Legendre polynomials of order i and j, respectively, can be expressed as a linear combination of Legendre polynomials up to the order $i + j$ [44]:

$$\text{leg}_i(x)\text{leg}_j(x) = \sum_{m=0}^{\min(i,j)} \frac{A(j-m)A(i-m)A(m)}{A(i+j-m)} \frac{2i+2j-4m+1}{2i+2j-2m+1} \text{leg}_{i+j-2m}(x), \tag{2.24}$$

where

$$A(t) = \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2t-1)}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot t} = \frac{(2t)!}{2^t (t!)^2} = \frac{1}{2^t}\binom{2t}{t}. \tag{2.25}$$

Exploiting this property it can be proved that the set of Legendre polynomials and their linear combinations satisfy all the requirements of the Stone–Weierstrass theorem on the compact $\mathbb{R}_1$.

The Legendre polynomials are orthogonal in $\mathbb{R}_1$, since

$$\int_{-1}^{1} \mathrm{leg}_i(x)\mathrm{leg}_j(x)\mathrm{d}x = \frac{2}{2i+1}\delta_{ij} \tag{2.26}$$

and, thus, the LN basis functions satisfy Eq. (2.19) and are orthogonal for a white uniform distribution of the input signal $x(n)$ in $\mathbb{R}_1$.

2.2.3.3 Chebyshev Nonlinear Filters

In CN filters, the one-dimensional basis functions $g_j[x(n)]$ are Chebyshev polynomials of the first kind [45,46], *i.e.*, $g_j[x(n)] = T_j[x(n)]$, and T_j are the Chebyshev polynomials of order j. Chebyshev polynomials of the first kind are orthogonal polynomials generated by the following recursive relation:

$$T_{j+1}(x) = 2xT_j(x) - T_{j-1}(x), \tag{2.27}$$

with $T_0(x) = 1$, $T_1(x) = x$. The polynomials are orthogonal in $\mathbb{R}_1$ with respect to the weighting function $\dfrac{1}{\pi\sqrt{1-x^2}}$, since

$$\int_{-1}^{+1} T_j(x)T_k(x)\frac{1}{\pi\sqrt{1-x^2}}\mathrm{d}x = \begin{cases} 0, & n \neq m, \\ 1, & n = m = 0, \\ 1/2, & n = m \neq 0. \end{cases} \tag{2.28}$$

For any $x \in \mathbb{R}_1$ also $T_j(x) \in \mathbb{R}_1$ and it has multiple maxima and minima equal to $+1$ and -1, respectively. Thus, the $T_j(x)$ are equiripple functions in $\mathbb{R}_1$.

The Chebyshev polynomials and their linear combinations form an algebra on the compact $\mathbb{R}_1$ that satisfies all the requirements of the Stone–Weierstrass theorem, as can be proved since the product of two polynomials of order i and j is

$$2T_i(x)T_j(x) = T_{i+j}(x) + T_{|i-j|}(x). \tag{2.29}$$

It is interesting to note that the approximation of any continuous function $f(x) = \sum_{n=1}^{+\infty} c_n T_n(x)$ with a linear combination $h(x)$ of Chebyshev polynomials up to a degree N is very close to a min–max approximation [46]. Indeed, the approximation error can be expressed as

$$\epsilon(x) = f(x) - h(x) = \sum_{N+1}^{+\infty} c_n T_n(x). \tag{2.30}$$

If the function is continuous and differentiable, the sequence of coefficients c_n converges rapidly to 0, such that $\epsilon(x) \simeq c_{N+1}T_{N+1}(x)$, which is an equiripple function for the properties of Chebyshev polynomials.

Table 2.3 summarizes the first Chebyshev polynomials of the first kind.

By exploiting the orthogonality property of the Chebyshev polynomials in Eq. (2.28), it can be verified that the CN basis functions $f_i[x(n), x(n-1), \ldots, x(n-N+1)]$ are orthogonal in $\mathbb{R}_1^N$ with

Table 2.3 Chebyshev polynomials of the first kind

$T_0(x) = 1$
$T_1(x) = x$
$T_2(x) = 2x^2 - 1$
$T_3(x) = 4x^3 - 3x$
$T_4(x) = 8x^4 - 8x^2 + 1$
$T_5(x) = 16x^5 - 20x^3 + 5x$

weighting function $\frac{1}{\pi\sqrt{1-x^2(n)}} \cdots \frac{1}{\pi\sqrt{1-x^2(n-N+1)}}$. Taking two different basis functions f_i and f_j, according to Eq. (2.28), we have

$$\int_{-1}^{+1} \cdots \int_{-1}^{+1} f_i[x(n), \ldots, x(n-N+1)] \cdot f_j[x(n), \ldots, x(n-N+1)] \cdot \frac{\mathrm{d}x(n) \cdots \mathrm{d}x(n-N+1)}{\pi\sqrt{1-x^2(n)} \cdots \pi\sqrt{1-x^2(n-N+1)}} = 0. \tag{2.31}$$

Thus, the basis functions are orthogonal for a white distribution of the input signal in $\mathbb{R}_1$ having the probability density function

$$p_x(x) = \frac{1}{\pi\sqrt{1-x^2}}. \tag{2.32}$$

Indeed, in Eq. (2.31) the factor

$$\frac{1}{\pi\sqrt{1-x^2(n)} \cdots \pi\sqrt{1-x^2(n-N+1)}}$$

can be interpreted as the joint probability density function of the N-tuple $[x(n), \ldots, x(n-N+1)]$.

An input signal $x(n)$ with probability density function (2.32) can be obtained by transforming a signal $u(n)$, white and uniformly distributed in $\mathbb{R}_1$, with the following mapping:

$$x(n) = \sin[\frac{\pi}{2}u(n)]. \tag{2.33}$$

This can be derived considering that the probability density of $x(n)$, $p_x(x)$, is related to the probability density of $u(n)$, $p_u(u)$, as follows:

$$p_x(x) = \left|\frac{\mathrm{d}u}{\mathrm{d}x}\right| p_u(u). \tag{2.34}$$

By inserting in Eq. (2.34) the inverse of Eq. (2.33),

$$u(n) = \frac{2}{\pi}\arcsin[x(n)], \tag{2.35}$$

and considering that $p_u(u) = \frac{1}{2}$, it is immediate to obtain Eq. (2.32).

Table 2.4 Hermite polynomials of variance σ_x^2

$W_0(x) = 1$
$W_1(x) = x$
$W_2(x) = x^2 - \sigma_x^2$
$W_3(x) = x^3 - 3\sigma_x^2 x$
$W_4(x) = x^4 - 6\sigma_x^2 x^2 + 3\sigma_x^4$
$W_5(x) = x^5 - 10\sigma_x^2 x^3 + 15\sigma_x^4 x$

2.2.3.4 Wiener Nonlinear Filters

The WN nonlinear filters defined in Section 2.2.2 are polynomial filters with orthogonal basis functions for any white Gaussian input signal $x(n) \in \mathcal{N}(0, \sigma_x^2)$. The corresponding one-dimensional basis functions g_j can be determined by applying the Gram–Schmidt orthogonalization to the set of polynomials

$$\{1, \quad x, \quad x^2, \quad x^3, \quad \dots\}, \tag{2.36}$$

taking into account that, for $x \in \mathcal{N}(0, \sigma_x^2)$, the kth moment of x is

$$\mu_k = E[x^k] = \begin{cases} 0 & \text{for } k \text{ odd}, \\ \sigma_x^k (k-1)!! & \text{for } k \text{ even}, \end{cases} \tag{2.37}$$

with $q!! = q \cdot (q-2) \cdot (q-4) \cdot \dots \cdot 1$. The resulting polynomials, here labeled as $W_j(x)$ for order j, are reported in Table 2.4 and are related to the Hermite polynomials [47]. In particular, they are "Hermite polynomials of variance σ_x^2" according to the definition reported in [48]. They can be generated with the following recursive relation:

$$W_{j+1}(x) = x W_j(x) - j\sigma_x^2 W_{j-1}(x), \tag{2.38}$$

with $W_0(x) = 1$ and $W_1(x) = x$. The set of polynomials $W_j(x)$ and their linear combinations satisfy the conditions of the Stone–Weierstrass theorem on any compact interval $[-A, +A]$, with $A \in \mathbb{R}^+$. Thus, also the resulting N-dimensional WN basis functions f_i satisfy the condition of the Stone–Weierstrass theorem on the compact $[-A, +A]^N$.

By constructions, the polynomials $W_j(x)$ are orthogonal for $x \in \mathcal{N}(0, \sigma_x^2)$, *i.e.*, for any i, j with $i \neq j$ we have

$$\int_{-\infty}^{+\infty} W_i(x) W_j(x) \frac{e^{\frac{x^2}{2\sigma_x^2}}}{\sigma_x \sqrt{2\pi}} \mathrm{d}x = 0. \tag{2.39}$$

From Eq. (2.39), the orthogonality of the WN basis function f_i for any white Gaussian input signal $x(n) \in \mathcal{N}(0, \sigma_x^2)$ also follows.

2.2.3.5 Orthogonal FLiP Filters for a Generic Probability Distribution $\mathcal{P}$

The procedure followed to derive the polynomials $W_j(x)$ can easily be applied to derive a family of orthogonal polynomials for any probability distribution $\mathcal{P}$. Once the moments $\mu_k = E[x^k]$ are computed, the Gram–Schmidt orthogonalization procedure can easily be applied to the polynomial set (2.36), deriving a family of polynomials $g_j(x)$ orthogonal for $x \in \mathcal{P}$. From this, it is possible to define a family of FLiP filters orthogonal for any white input signal $x(n) \in \mathcal{P}$.

2.2.3.6 Least Square Error Estimation

The input–output relationship of any FLiP filter of order K, with memory of N samples, can be written as

$$y(n) = \sum_{i=0}^{N_{\mathrm{T}}-1} c_i f_i(n), \tag{2.40}$$

where $f_i(n), i = 0, \ldots, N_{\mathrm{T}} - 1$, are N-dimensional basis functions taken from Table 2.1. Using vector notation, Eq. (2.40) becomes

$$y(n) = \mathbf{c}^{\mathrm{T}} \mathbf{f}_n, \tag{2.41}$$

where

$$\mathbf{c} = [c_0, c_1, \ldots, c_{N_{\mathrm{T}}-1}]^{\mathrm{T}}$$

and

$$\mathbf{f}_n = [f_0(n), f_1(n), \ldots, f_{N_{\mathrm{T}}-1}(n)]^{\mathrm{T}}.$$

Then the minimum mean square error (MSE) estimation gives as optimal solution of (2.41)

$$\mathbf{c}_o = \mathbf{R}_{ff}^{-1} \mathbf{p}_{yf}, \tag{2.42}$$

since the filter output is linear with respect to the coefficients [49]. In Eq. (2.42), $\mathbf{R}_{ff}$ is the $N_{\mathrm{T}} \times N_{\mathrm{T}}$ auto-correlation matrix of the basis functions $f_i(n)$ and $\mathbf{p}_{yf}$ is the $N_{\mathrm{T}} \times 1$ cross-correlation vector between the output $y(n)$ of the unknown system and the basis functions f_i, evaluated using the given input sequence $x(n)$. When the elements of $\mathbf{R}_{ff}$ and $\mathbf{p}_{yf}$ are computed by time averages over L samples of the input $x(n)$, $\mathbf{c}_o$ is called the optimal least square solution of (2.41). When the basis functions are orthogonal, the autocorrelation matrix is diagonal and Eq. (2.42) reduces to the formulas of the cross-correlation method.

2.2.3.7 Cross-Correlation Method

The orthogonality of the basis functions for a stochastic signal makes it possible to estimate the coefficients of the FLiP filter with the cross-correlation method. The method is here applied computing the cross-correlation between the basis functions and the system output. Exploiting orthogonality of the basis functions $f_i(n)$, the ith entry of Eq. (2.42) reduces to

$$c_i = \frac{E[y(n) f_i(n)]}{E[f_i^2(n)]}. \tag{2.43}$$

The expectations can be estimated using time averages, providing a feasible identification method.

The main problem of the cross-correlation method for stochastic inputs is the huge number of samples (in the order of millions) often necessary to obtain an accurate estimate of the coefficients c_i.

Moreover, it can be difficult to generate white input signals with the desired distribution. For example, for white Gaussian inputs the maximum amplitude and the maximum bandwidth have to be limited, altering the orthogonality of the basis functions of the Wiener nonlinear filter. The problem will be solved in Section 2.3, where some deterministic signals, the PPSs, capable of guaranteeing the orthogonality of the basis functions on a finite period, will be developed.

2.2.4 SIMPLIFIED FLIP FILTERS

All FLiP filters discussed in the previous subsection have been obtained following the constructive rule of the triangular representation of Volterra filters and have the same number of basis functions and coefficients. For a FLiP filter of order K, memory N, this number is equal to

$$N_{\mathrm{T}} = N_T(K, N) = \binom{N+K}{N}, \tag{2.44}$$

as in the triangular representation of truncated Volterra filters.

According to Eq. (2.44), the number of coefficients of a FLiP filter increases geometrically with the memory of the filter N and exponentially with the order K of the filter. Even for small values of K (*e.g.*, 2 or 3) and moderate values of N, the number of coefficients can be prohibitively large. This problem is considered the main drawback of all FLiP filters. To alleviate this difficulty, it is possible to resort to the same strategy of simplified Volterra filters [50]. The FLiP filter can be first expressed according to its diagonal representation, which was first introduced for finite-memory symmetric Volterra filters in [51]. In the diagonal representation rather than using Cartesian coordinates, the filter is expressed considering coordinates aligned with the diagonals of the hypercubes representing the nonlinear kernels. For comparison, the second-order kernel of a FLiP filter of memory N in the Cartesian representation is

$$\sum_{i=0}^{N-1} c_{i,i}\, g_2[x(n-i)] + \sum_{i=0}^{N-1} \sum_{j=i+1}^{N-1} c_{i,j}\, g_1[x(n-i)] g_1[x(n-j)], \tag{2.45}$$

while in the diagonal representation it is

$$\sum_{i=0}^{N-1} c_{i,i}\, g_2[x(n-i)] + \sum_{i=1}^{N-1} \sum_{j=0}^{N-1-i} c_{j,i+j}\, g_1[x(n-j)] g_1[x(n-i-j)]. \tag{2.46}$$

The diagonal number of a basis function is defined as the maximum time difference between the samples involved in its expression. For example, $g_1[x(n)]g_1[x(n-2)]$ has diagonal number 2; $g_2[x(n-r)]g_1[x(n-r-3)]$ has diagonal number 3 for any $r \in \mathbb{Z}$. Simplified FLiP filters are obtained by considering only the basis functions that have a *diagonal number* lower than or equal to a certain value D. The diagonal number is always comprised in the range $0 \le D \le N-1$. When $D = 0$, only the basis functions of the principal diagonals of hypercubes representing the nonlinear kernels are selected. When $D = N-1$, all basis functions of nonlinear kernels are selected. The second-order

Table 2.5 Number of coefficients N_D of a FLiP filter of order K, memory N and diagonal number D and number of coefficients N_T of a full FLiP filter with the same order and memory

K	N	N_D					N_T
		D					
		0	2	4	6	8	
2	10	21	38	51	60	65	66
	50	101	198	291	380	465	1326
	100	201	398	591	780	965	5151
3	10	31	90	161	228	275	286
	50	151	490	1001	1668	2475	23426
	100	301	990	2051	3468	5225	176851

kernel of a FLiP filter of memory N and diagonal number D reduces to

$$\sum_{i=0}^{N-1} c_{i,i} g_2[x(n-i)] + \sum_{i=1}^{D} \sum_{j=0}^{N-1-i} c_{j,i+j} g_1[x(n-j)] g_1[x(n-i-j)]. \tag{2.47}$$

The motivation for limiting the diagonal number of the basis functions used to define the FLiP filter derives from the observation that the "energy" of the kernels, used to represent real-world nonlinear systems, is often concentrated around the principal diagonals. This situation has been observed in the fields of nonlinear acoustic echo cancellation [50,52,53], nonlinear active noise control [54] and identification of nonlinear systems [55] such as saturated amplifiers or guitar pedals [38,40].

A FLiP filter of memory N, order K, diagonal number D has a number of coefficients N_D equal to

$$N_D = \binom{D+K+1}{D+1} + \binom{D+K}{D+1} \cdot (N-1-D). \tag{2.48}$$

In the particular case of $D = N-1$, $N_D = N_T$.

Table 2.5 provides a comparison of the numbers of coefficients N_D and N_T of FLiP filters of order 2 and 3, for different memory lengths and diagonal numbers. Reducing the number of diagonals, it is possible to obtain a significant reduction in the computational complexity of the FLiP filter.

2.3 RECENT IDENTIFICATION METHODS FOR ORTHOGONAL FILTERS

Most algorithms suitable for the identification of linear filters can also be used for the identification of LIP nonlinear filters. For example, the adaptive algorithms suitable for the identification of linear FIR filters can easily be adapted to LIP filters. The Least Mean Square (LMS), the Recursive Least Square (RLS), the Affine Projection (AP), the Filtered-X Least Mean Square (FX-LMS) and the Filtered-X Affine Projection (FX-AP) algorithms have derivations and expressions formally identical to the linear case. The reader is referred to [32] for a review of these algorithms in the context of LIP nonlinear filters.

In what follows, we review some recent identification techniques suitable for orthogonal LIP nonlinear filters. In particular, we consider the identification of FLiP filters using PPSs, *i.e.*, deterministic periodic sequences that guarantee the orthogonality of the FLiP basis functions on a finite period. Then we consider the multiple-variance identification method that can be used for the identification of Volterra filters avoiding the problem of locality of the solution, *i.e.*, the fact that the estimated model well approximates the unknown system only at the same input variance of the measurement.

2.3.1 PERFECT PERIODIC SEQUENCES

A recent solution for the identification of linear and nonlinear models exploits deterministic input signals able to guarantee the orthogonality of the basis functions on a finite interval. As a matter of fact, PPSs have been proposed for the identification of linear filters, as shown in [56–58]. A periodic sequence, applied as input to a linear or nonlinear system, is called perfect if all cross-correlations between any two different basis functions of the modeling filter, estimated over a period, are zero. In the case of linear filters, using a PPS input of period L the basis functions $x(n-i)$, with i ranging from 0 to $L-1$, form an orthogonal set. It has been shown in the literature that the PPSs optimize the convergence speed of the NLMS algorithm [59,60]. Moreover, in the absence of output noise, a PPS can identify a linear filter in just one period [61,62]. Exploiting periodic input signals, it is also possible to identify the unknown system using an adaptive algorithm that requires just a multiplication, an addition and a subtraction per time sample [63].

2.3.1.1 Development of Perfect Periodic Sequences

In what follows, we discuss how we can develop PPSs suitable for the identification of an orthogonal FLiP filter. Different strategies have been devised for this purpose. Let us indicate with $f_i(n)$ the ith FLiP basis function with i ranging from 1 to the cardinality of the set it belongs to, with $S_f(K,N)$ being the set of basis functions of order less than or equal to K and memory N and $S_{f,n}(K,N)$ being the subset of $S_f(K,N)$ formed by all basis functions having $x(n)$ as one of their arguments and with L as the PPS period. $S_f(K,N)$ has cardinality $N_{\mathrm{T}}(K,N)$, defined in (2.44), and $S_{f,n}(K,N)$ has cardinality $N_{\mathrm{T}}(K-1,N)$. The L samples of the fundamental period of the PPS are indicated as

$$x_0, x_1, \ldots, x_{L-1}.$$

PPSs for orthogonal FLiP filters can be developed by directly imposing the orthogonality of any couple of basis functions over a period, *i.e.*,

$$< f_i(n) f_j(n) >_L = 0, \tag{2.49}$$

for all $f_i(n) \in S_{f,n}(K,N)$, $f_j(n) \in S_f(K,N)$ with $f_i(n) \neq f_j(n)$. It should be noted that $f_i(n) \in S_{f,n}(K,N)$ is assumed because if Eq. (2.49) is satisfied, then also

$$< f_i(n-l) f_j(n-l) >_L = 0,$$

for all $l \in \mathbb{Z}$.

Together with the conditions in (2.49), it is also convenient to impose the condition that

$$< f_i^2(n) >_L = P_i, \tag{2.50}$$

for all $f_i(n) \in S_{f,n}(K, N)$, with P_i being the power of the basis functions $f_i(n)$ for the white stochastic input signal that guarantees the orthogonality of the basis functions (*i.e.*, for a white uniform input signal in case of EMFN and LN filters, for a white input signal with the probability density function in (2.32) for CN filters and for a white Gaussian input signal with variance σ_x^2 for WN filters). With (2.50), the power of each basis function is *a priori* known. Moreover, the condition in (2.50) allows also to easily form an orthonormal set of basis functions by dividing each basis function $f_i(n)$ by $\sqrt{P_i}$.

The system of nonlinear equations defined in (2.49) and (2.50) can be proved to be equivalent to the following simpler system [39,36]:

$$< f_i(n) >_L = 0, \tag{2.51}$$

for all $f_i(n) \in S_{f,n}(2K, N)$.

For sufficiently large L, Eq. (2.51) provides an under-determined system of nonlinear equations in the variables $x_0, x_1, \ldots, x_{L-1}$, which is expected to have infinite solutions. The system in (2.51) has $Q = N_T(2K - 1, N)$ equations and, for L ranging between $1.5Q$ and $6Q$, depending on the particular FLiP filter, a solution for this system has always been found. The nonlinear equation system in (2.51) has been solved using the Newton–Raphson method, implemented as described in [64, Ch. 9.7], with the Jacobian matrix computed analytically. In the case of EMFN, LN and CN filters, where the input samples belong to $\mathbb{R}_1$, the variables $x_0, x_1, \ldots, x_{L-1}$ have been reflected in $\mathbb{R}_1$ any time they exceeded the range at some iteration.

Another strategy for deriving PPSs for orthogonal polynomial FLiP filters is also possible. We can impose the PPS to have the same joint moments of $\mathcal{P}$, which guarantees the orthogonality of the basis functions, for all variables involved in the system in (2.51), *i.e.*, we can impose

$$< x_n^{i_0} x_{n-1}^{i_1} \cdots\ldots\cdots x_{n-N+1}^{i_{N-1}} >_L = E[x^{i_0}(n) x^{i_1}(n-1) \cdots\ldots\cdots x^{i_{N-1}}(n-N+1)], \tag{2.52}$$

where x_n is the PPS sample at time n, $x(n) \in \mathcal{P}$, $i_0 \in [1, 2K]$, $i_1, i_2, \ldots, i_{N-1} \in [0, 2K]$ and $i_0 + i_1 + \ldots i_{N-1} \le 2K$. For example, this strategy has been applied to develop PPSs for WN filters. The system in (2.52) has the same number of equations as that in (2.51) and can again be solved with the Newton–Raphson method.

The number of equations Q of the nonlinear systems in (2.51) and (2.52) increases exponentially with the order K and geometrically with the memory N. Even for low orders and memory lengths, Q can be unacceptably large. Periodic sequences with specific structures can be adopted to reduce the number of equations and variables in the nonlinear system. For example, the following conditions can almost halve the number of equations and variables [39]:

1. Symmetry: if in the PPS for any N-tuple of samples $a_1, a_2, \ldots, a_N$ there is also the reversed one $a_N, a_{N-1}, \ldots, a_1$, then, for any couple of symmetric basis functions, only one needs to be considered.
2. Oddness: if 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 basis functions have *a priori* zero average and can be discarded from the system.
3. Oddness-1: if in the PPS for any N-tuple of samples $a_1, a_2, \ldots, a_N$ there is also that obtained by alternatively negating one every two terms, $a_1, -a_2, a_3, -a_4, \ldots, -a_N$, all Odd-1 functions have *a priori* zero average and can be discarded from the system.

By definition, Odd-1 are all those basis functions that change their sign by alternatively negating one every two samples, *e.g.*, $f_1[x(n)]f_1[x(n-1)]$. Similarly Odd-2, Odd-4, ... functions can be defined and exploited to reduce the complexity of the nonlinear system in (2.51). Moreover, two or more conditions can be considered together.

The Newton–Raphson algorithm has memory and processing time requirements that grow with Q^3. Thus, the reduction in the number of equations obtained with symmetry and oddness conditions is often fundamental to be able to solve the system in (2.51), but it is paid for with a longer period of the resulting PPS. Another strategy to reduce the computational complexity of the system in (2.51) for large orders and memory lengths resorts to the use of simplified models, as those introduced in Section 2.2.4.

2.3.1.2 System Identification using Perfect Periodic Sequences

Let us assume we want to identify a time-invariant, finite-memory, causal, continuous, nonlinear system. When the input–output relationship of the nonlinear system can be expressed as a linear combination of FLiP basis functions up to order K and memory N as in Eq. (2.40), the coefficients c_i can be estimated using as input signal a PPS suitable for the specific FLiP filter with at least order K and memory N.

The coefficients of the FLiP filter can be estimated with the cross-correlation method, *i.e.*, computing the cross-correlation between the basis functions and the system output over a period of the PPS,

$$\widehat{c_i} = \frac{< f_i(n)y(n) >_L}{< f_i^2(n) >_L}. \tag{2.53}$$

Since the autocorrelations $< f_i^2(n) >_L = P_i$, with P_i *a priori* known, the computational cost of (2.53) is simply a multiplication and an addition per basis function and per input sample.

In what follows, we discuss the errors caused by an under-determination of the unknown system order or memory.

2.3.1.3 Under-Determined Order

Assume that a PPS for FLiP filters of order K, memory N, is used to identify a system formed by a linear combination of FLiP basis functions with memory N but maximum order greater than K, *i.e.*,

$$y(n) = \sum_{i=0}^{N_{\mathrm{T}}} c_i f_i(n) + O_{K+1}(n), \tag{2.54}$$

where $f_i(n)$ has maximum order K and $O_{K+1}(n)$ is a linear combination of basis functions of order greater than K. Assume also $O_{K+1}(n)$ is formed by basis functions of order lower than $2K$ and that the power of the nonlinear kernels reduces with the kernel order. In these conditions, the estimated coefficient $\widehat{c_i}$ is affected by an error generated by $O_{K+1}(n)$. The error influences mainly the coefficients of the higher-order basis functions and, in general, does not influence or influences only mildly the coefficients of the lower-order basis functions. This can be justified by first considering $O_{K+1}(n)$ a linear combination of basis functions of order $K+1$. The estimated coefficients $\widehat{c_i}$ of all basis functions of order less than K are not affected by the error, because for the PPS construction rules the cross-correlation of these basis functions with those of $O_{K+1}(n)$ is zero. In fact, the product of a basis function of an order less than K with a basis function of order $K+1$ is a linear combination of basis functions of order lower than or equal to $2K$, which satisfy (2.51) for a PPS input. Only the coefficients of the basis functions of order K are affected, because their cross-correlation with $O_{K+1}(n)$ is in general different from zero. If $O_{K+1}(n)$ is a linear combination of basis functions of order $K+2$, the estimated coefficients $\widehat{c_i}$ of all basis functions of order less than $K-1$ are not affected by the error, while the coefficients of the basis functions of order $K-1$ and K are affected by an error. Eventually, consider $O_{K+1}(n)$ a linear combination of basis functions of order ranging from $K+1$ to $K+A$, with $A<K$, whose kernels have power decreasing with the order. Then the estimated coefficients $\widehat{c_i}$ of all basis functions of order lower than $K-A+1$ are not affected by the error, while the coefficients of the basis functions of order greater than $K-A$ are affected. The error in the coefficients $\widehat{c_i}$ of the basis functions of order $K-A+1$ depends only on the kernel of order $K+A$ of $O_{K+1}(n)$, which has the lowest power. The error in the coefficients $\widehat{c_i}$ of the basis functions of order $K-A+2$ depends only on the kernels of order $K+A-1$ and $K+A$ of $O_{K+1}(n)$. The error in the coefficients $\widehat{c_i}$ of the basis functions of order K depends on all the kernels of $O_{K+1}(n)$ and can be expected to be the most relevant.

2.3.1.4 Under-Determined Memory

Assume next that a PPS for FLiP filters of order K, memory N, is used to identify a system that is a linear combination of FLiP basis functions with order K but memory greater than N, *i.e.*,

$$y(n)=\sum_{i=0}^{N_{\mathrm{T}}} c_i f_i(n)+M_{N+1}(n), \tag{2.55}$$

where $f_i(n)$ has order K, memory N, and $M_{N+1}(n)$ is a linear combination of basis functions of memory greater than N but lower than $2N$. Also in this case, the estimated coefficients $\widehat{c_i}$ are affected by an aliasing error generated by $M_{N+1}(n)$. The error affects mainly the coefficients of the basis functions associated with the most recent samples, $x(n), x(n-1), \ldots$, and, in general, affects only mildly or does not affect the coefficients of the basis functions associated with the less recent samples, $x(n-N+1), x(n-N+2), \ldots$. To illustrate this property, let us first consider the identification of the system of memory $\widetilde{N}=N+1$,

$$y(n)=\sum_{i=0}^{\widetilde{N}-1} c_i g_1[x(n-i)], \tag{2.56}$$

where g_1 is the one-dimensional basis function of order 1 defining the FLiP filter. The system is identified using a PPS for a FLiP filter of order 1, memory N. In this case, the estimate of c_0 in (2.53) is affected by an error, since we have not imposed $< g_1[x(n)]g_1[x(n-N)] >_L$ to be zero in the construction of the PPS. In contrast, $c_1, \ldots, c_{N-1}$ are not affected by any error since $< g_1[x(n-l)]g_1[x(n-N)] >_L = 0$ for all $1 \le l \le N-1$ according to the construction rules of the PPS. If the system in (2.56) has memory $\widetilde{N}=N+2$, the estimate of c_0 and c_1 in (2.53) is affected by an error, since we have not imposed $< g_1[x(n)]g_1[x(n-N)] >_L = < g_1[x(n-1)]g_1[x(n-N-1)] >_L$ and $< g_1[x(n)]g_1[x(n-N-1)] >_L$ to be zero, and in (2.53)

$$
\begin{aligned}
< y(n) g_1[x(n)] >_L = c_0 < g_1^2[x(n)] >_L + c_N < g_1[x(n)] g_1[x(n-N)] >_L \\
+ c_{N+1} < g_1[x(n)] g_1[x(n-N-1)] >_L ,
\end{aligned} \tag{2.57}
$$

$$
\begin{aligned}
< y(n) g_1[x(n-1)] >_L = c_1 < g_1^2[x(n-1)] >_L \\
+ c_{N+1} < g_1[x(n-1)] g_1[x(n-N-1)] >_L .
\end{aligned} \tag{2.58}
$$

In contrast, $c_2, \dots, c_{N-1}$ are not affected by any error. When the system in (2.56) has memory $\widetilde{N}$, with $N < \widetilde{N} < 2N$, the coefficients $c_0, c_1, \dots, c_{\widetilde{N}-N-1}$ are affected by an error, while the coefficients $c_{\widetilde{N}-N}, \dots, c_{N-1}$ are not affected. Moreover, the error in the estimate of c_0 depends on $c_N, c_{N+1}, \dots, c_{\widetilde{N}}$; the error in the estimate of c_1 depends on $c_{N+1}, \dots, c_{\widetilde{N}}$; and the error in the estimate of $c_{\widetilde{N}-N-1}$ depends only on $c_{\widetilde{N}}$. Also considering that in real systems the coefficients c_i tends to fade for increasing i, we can expect the error to affect mainly the coefficients of the basis functions associated with the most recent samples, $x(n), x(n-1), \dots$ and only mildly those of the basis functions associated with the less recent samples. The same considerations can be applied to more general forms of nonlinear systems than (2.56).

2.3.1.5 Most Relevant Basis Function and Information Criteria

When a FLiP filter is identified with a PPS and the cross-correlation method, its basis functions can easily be ranked according to the MSE reduction they produce. In fact, for the orthogonality property over a period induced by the PPS, the MSE reduction produced by the ith basis function is

$$
\delta \mathrm{MSE}_i = \frac{< f_i(n) y(n) >_L^2}{< f_i^2(n) >_L} . \tag{2.59}
$$

Eqs. (2.53) and (2.59) can be combined with the minimization of an information criterion to determine the number of most relevant basis functions and thus obtain a compact representation for the nonlinear system. The most popular information criteria available in the literature are:

- the Akaike information criterion (AIC) [65]:

$$
\mathrm{AIC}(\alpha) = L \log_e[\sigma_\epsilon^2(n_p)] + \alpha n_p , \tag{2.60}
$$

  with $\alpha > 0$,
- the Final Prediction Error (FPE) [65]:

$$
\mathrm{FPE} = L \log_e[\sigma_\epsilon^2(n_p)] + L \log_e \frac{L + n_p}{L - n_p} , \tag{2.61}
$$

- the Khundrin law of iterated logarithm criterion (LILC) [66]:

$$
\mathrm{LILC} = L \log_e[\sigma_\epsilon^2(n_p)] + 2 n_p \log_e[\log_e(L)] , \tag{2.62}
$$

- the Bayesian information criterion (BIC) or Schwarz criterion [67]:

$$
\mathrm{BIC} = L \log_e[\sigma_\epsilon^2(n_p)] + n_p \log_e[L] , \tag{2.63}
$$

where $\sigma_\epsilon^2(n_p)$ is the variance of the residual estimation error for a model formed by n_p basis functions, *i.e.*, in our case

$$\sigma_\epsilon^2(n_p) = < y^2(n) >_L - \sum_{i=0}^{n_p-1} \delta\text{MSE}_i, \tag{2.64}$$

and L is the number of data used for the model estimation.

2.3.2 MULTIPLE-VARIANCE METHODS

Let us consider again the Volterra and Wiener theory. It has been shown in [11] how the input non-idealities affect kernel estimation. In particular, an explicit expression of residuals that occur when there is a difference between the moments of an ideal input and those produced by a pseudorandom finite length input is given. The residual causes an error in system identification even if the target system itself is a Volterra series having the same order as the model. Another important source of error are the truncation errors that arise if the order of the Volterra filter is lower than that of the system to be identified. These factors cause limitations in Volterra filter approximation capabilities, which are particularly evident in the phenomenon of locality of the solution. In most cases, the model well represents the nonlinear system only in a limited range of input signal powers, close to the variance σ_x^2 used for the model estimate.

The aim of the multiple-variance methods is to lower the MSE of the Volterra filter when signals with variances different from that used in the identification are input of the system. To do this, multiple-variance inputs are used for Wiener kernel identification.

In the traditional orthogonal algorithm, using high σ_x^2 inputs has the advantage of stimulating high-order nonlinearities, so as to achieve more accurate high-order kernel identification. As a drawback, the use of high σ_x^2 values causes higher identification errors in lower-order kernels, as shown in [14].

The rationale should be to use a low σ_x^2 value for the lower-order kernels and to gradually increase it for higher-order kernels. For the peculiar expression of Wiener functionals, see Eq. (2.8) and the resulting Wiener-to-Volterra formulas in Eq. (2.9); an appropriate normalization is needed in Wiener kernel expressions and in Wiener-to-Volterra conversion formulas, for taking into account the use of multiple variances.

New formulas for Wiener kernel identification were proposed in [14], and they are explicitly reported in the following up to the order 3:

$$k_0^{(0)} = E[y^{(0)}(n)], \tag{2.65}$$

$$k_{1,\,\tau_1}^{(1)} = \frac{1}{\sigma_{x,\,(1)}^2} E\left[y^{(1)}(n)\,x^{(1)}(n-\tau_1)\right], \tag{2.66}$$

$$k_{2,\,\tau_1,\tau_2}^{(2)} = \frac{1}{2!\sigma_{x,\,(2)}^4} \left\{E\left[y^{(2)}(n)\,x^{(2)}(n-\tau_1)x^{(2)}(n-\tau_2)\right] - \sigma_{x,\,(2)}^2 k_0^{(2)} \delta_{\tau_1\tau_2}\right\}, \tag{2.67}$$

$$\begin{aligned} k_{3,\,\tau_1,\tau_2,\tau_3}^{(3)} = \frac{1}{3!\sigma_{x,\,(3)}^6} &\left\{E\left[y^{(3)}(n)\,x^{(3)}(n-\tau_1)x^{(3)}(n-\tau_2)x^{(3)}(n-\tau_3)\right]\right. \\ &\left. - \sigma_{x,\,(3)}^4 \left[k_{1,\,\tau_1}^{(3)}\delta_{\tau_2\tau_3} + k_{1,\,\tau_2}^{(3)}\delta_{\tau_1\tau_3} + k_{1,\,\tau_3}^{(3)}\delta_{\tau_1\tau_2}\right]\right\}, \end{aligned} \tag{2.68}$$

where $k_p^{(j)}$ and $y^{(j)}(n)$ are obtained with the input $x^{(j)}(n)$ of variance equal to $\sigma_{x,\,(j)}^2$.

As can be clearly seen in the explicit formulas, the drawback with respect to the classic formula is that, for the identification of a pth-order kernel, all lower kernels must be identified again with the higher variance. However, an outstanding improvement in the output MSE is obtained if the Volterra kernels are obtained from the previous estimated Wiener kernels, using the following Wiener to Volterra formulas:

$$h_3 = k_3^{(3)}, \tag{2.69}$$

$$h_2 = k_2^{(2)}, \tag{2.70}$$

$$h_1 = k_1^{(1)} - 3\sigma_{x,\,(1)}^2 \sum_{\tau_2=1}^{M} k_{3,\,\tau_1,\tau_2,\tau_2}^{(3)}, \tag{2.71}$$

$$h_0 = k_0^{(0)} - \sigma_{x,\,(0)}^2 \sum_{\tau_1=1}^{M} k_{2,\,\tau_1,\tau_1}^{(2)}. \tag{2.72}$$

These formulas are valid for a third-order Volterra filter. By comparing them with Eq. (2.9), we note that in each Volterra kernel the diagonal Wiener coefficients of higher order are multiplied by the variance corresponding to the order of the Volterra kernel instead of the corresponding Wiener kernel.

The multiple-variance method is straightforward when used with the Wiener basis functions of Section 2.2.3.4, since each Wiener kernel can be determined computing only the cross-correlation between the system output and the basis functions of the same order. Using the Wiener basis functions, recomputing the kernels of lower orders as in (2.67) and (2.68) can be avoided [68].

The multiple-variance method can also be used with PPSs as inputs. In this case the residual due to input nonidealities are greatly reduced, if not completely canceled, and the estimation of the kernel diagonals is not a problem anymore. However, the approximation of the estimated model is still affected by truncation errors in the memory or the order of the model, so multiple-variance methods can still be useful. PPSs with multiple variances suitable for WN filters can easily be developed. Indeed, if a PPS for a WN filter of order P, memory N and input variance σ_x^2 is amplified by a factor A, by construction the resulting sequence is a PPS for a WN filter of the same order and memory and variance $A^2\sigma_x^2$. Each Wiener kernel $k_i^{(i)}$ of order i can be separately estimated using a PPS of variance $\sigma_{x,(i)}^2$ using (2.53). Eventually, Eqs. (2.69)–(2.72) can be used to estimate the corresponding Volterra kernels.

2.4 EXPERIMENTAL RESULTS

In this section, we provide a couple of experimental results illustrating the identification and emulation of real nonlinear devices using orthogonal nonlinear filters.

2.4.1 IDENTIFICATION OF NONLINEAR DEVICES USING PERFECT PERIODIC SEQUENCES

In the first experiment, we consider the identification of a real device using different FLiP filters. The experiment aims to illustrate the advantages offered by PPSs for the identification of nonlinear devices and the very compact models that can be obtained by selecting the most significant basis functions

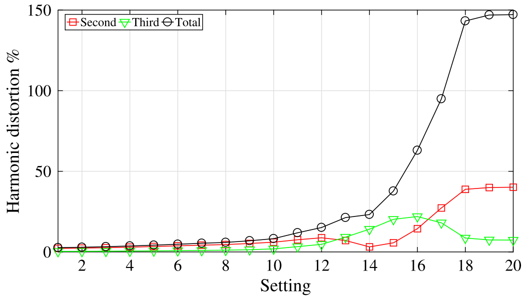

FIGURE 2.1

Second, third and total harmonic distortion.

according to some information criterion. In particular, the identification of an audiophile vacuum tube preamplifier, *i.e.*, Behringer Tube Ultragain Mic 100, is reported. The experiment takes advantage of the preamplifier gain setting that is used for introducing different levels of nonlinear distortion. The results of twenty different distortion conditions are reported. Fig. 2.1 shows the second, third and total harmonic distortions on a 200-Hz signal at the maximum used amplitude at the different settings. Clearly, the last settings that have total harmonic distortion close to or greater than 100% provide extreme distortion conditions. The preamplifier has been identified with EMFN, LN and CN filters using three PPSs, suitable for the three filters with order 3, memory 20 and with a period of 655 408 samples. The coefficients of the filters were estimated with the cross-correlation method and the most relevant basis functions were selected according to the Bayesian information criterion, minimizing Eq. (2.63). The preamplifier has also been identified with a Volterra filter on the same data used for the CN filter identification. The cross-correlation method cannot be applied to the Volterra filter estimation since its basis functions are not orthogonal. Thus, the Volterra filter has been identified with the method of [69], which belongs to the most computationally efficient identification methods for LIP nonlinear systems available in the literature. In all conditions, the signal-to-noise ratio was greater than 65 dB.

Fig. 2.2 shows the number of selected terms and the percentage of unexplained power (*i.e.*, the ratio in percent between the residual MSE and the power of the output signal). The percentage of unexplained power is very low, close to 1%, till settings 16. Then it tends to increase, indicating that a third-order model is inadequate to represent the nonlinear system at those high nonlinear distortions. CN, LN and Volterra filters are all polynomial filters and each filter can be converted into one of the other representations. Thus, for the same input signal the filters should provide very similar results, with just a possible little change in the number of selected basis functions. This is confirmed in Fig. 2.2 for Volterra and CN filters, which are estimated for the same input signal. Also LN filters, which are estimated for a different PPS input, provide very similar results in Volterra and CN filters for small and medium nonlinear distortions. On the contrary, because of the different input signal, they produce larger errors for higher distortions. When the Volterra, LN and CN filters are estimated on the same signal, they provide very similar results. EMFN filters for low and medium distortions originate slightly worse results than the other filters, because they lack a linear term, but for higher distortions they are able to provide better results than the other filters, also when the estimate is performed on the same

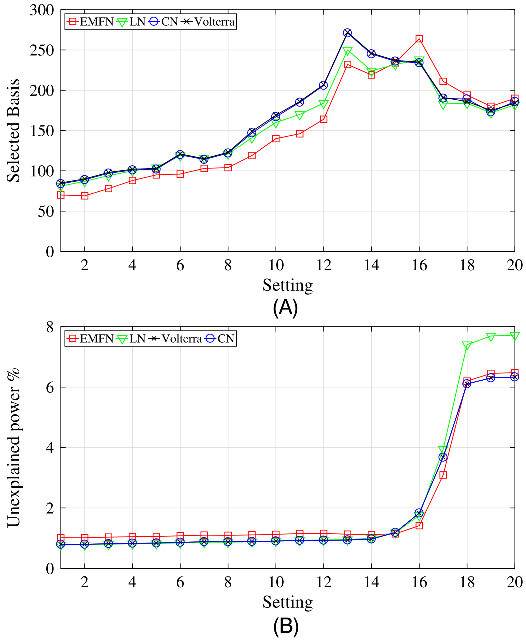

FIGURE 2.2

Number of selected bases (A) and unexplained power (B) for EMFN, LN, CN and Volterra filters.

signal. Note that the number of selected basis functions is very low compared with the total number of basis functions N_T of a FLiP filter of order 3, memory 20, which is $N_T = 1771$.

There is here a significant difference between the effort necessary to estimate the CN, LN and EMFN filters and that for estimating the Volterra filters. Obtaining CN, LN and EMFN filters using PPSs and the cross-correlation method required only a few hours of computer time. On the contrary, computing the results for Volterra filters with the method in [69] required days of simulations on the same computer. As a matter of fact, if T indicates the number of samples used for the identification, B the number of candidate basis functions and S the number of selected basis functions, the computational cost of the method in [69] is of the order of TBS^2 operations, while the cross-correlation method requires an order of TB operations.

2.4.2 MULTIPLE-VARIANCE SYSTEM IDENTIFICATION AND EMULATION

In the second experiment, we illustrate the advantage offered by the multiple-variance methods in reducing the problem of the solution locality and in improving the emulation of nonlinear devices. For this purpose, several measurements have been carried out to test the exploitation of multiple-variance PPSs in the cross-correlation method and making comparisons with multiple-variance WGN [14]. Different test sessions have been accomplished, considering a real-world nonlinear device, modeled using a third-order Volterra series with memory 10 and 25, respectively. The adopted PPS sequence has order $P = 3$, memory $N = 25$, period $L = 1\,393\,024$ and variance $\sigma^2_{\mathrm{pps}} = 1/12$, and the WGN sequence has length L. A sampling frequency $f_{\mathrm{s}} = 44.1$ kHz has been adopted. For the multiple-variance cross-correlation method, since in the considered examples the second-order kernel is dominant at high input variances, $\sigma^2_{x,0} = \sigma^2_{x,1} = \frac{\sigma^2_{\mathrm{pps}}}{16}$ and $\sigma^2_{x,2} = \sigma^2_{x,3} = \sigma^2_{\mathrm{pps}}$ have been assumed. With reference to the traditional cross-correlation method, the same two values for variance have been considered. Therefore, the identification of the model has been performed considering the following settings: (1) WGN with single variance $\sigma^2_{x,0}$, (2) WGN with single variance $\sigma^2_{x,3}$, (3) WGN with multiple variances, (4) PPS with single variance $\sigma^2_{x,0}$, (5) PPS with single variance $\sigma^2_{x,3}$, (6) PPS with multiple variances.

Then a performance evaluation has been performed applying WGN and music of length 44 100 samples to the system under investigation and to the model, assuming several input variances. Results are reported in terms of the normalized MSE (NMSE) in the frequency domain between the output of the system under investigation $y(n)$ and the output of the identified Volterra series $\hat{y}(n)$, according to the following formula:

$$\mathrm{NMSE} = 10 \log_{10} \frac{\sum_{n=1}^{N} \left[|Y(f_n)| - \left| \hat{Y}(f_n) \right| \right]^2}{\sum_{n=1}^{N} |Y(f_n)|^2}. \tag{2.73}$$

The real-world device used is the Presonus TubePRE microphone/instrument tube preamplifier. It provides a drive potentiometer controlling the amount of tube saturation, *i.e.*, the amount of applied distortion. The device has been set to provide a second harmonic distortion of 4.2% and third harmonic distortion of 0.5% on a 1-kHz tone signal. Once the model has been obtained, inputs with variance σ^2_x in the interval $\left[(1/4096, \ldots, 1/4, 1/2, 1)\, \sigma^2_{\mathrm{pps}} \right]$ have been adopted and the output of the real device has been compared with the output of the model. The results reported in Fig. 2.3 show that the exploitation of multiple-variance PPSs provides two advantages at the same time, *i.e.*, the input region in which the error has acceptable values both for noise and music is widened, thus overcoming the locality problem, and accuracy is improved with respect to stochastic input. In particular, the improvements provided by multiple-variance PPSs can be noted especially for music input as reported in Fig. 2.3 B.

2.5 CONCLUDING REMARKS

The chapter has discussed different families of orthogonal LIP nonlinear filters. FLiP filters have provided the common framework for introducing the orthogonal LIP filters. FLiP filters are able to arbitrarily well approximate any causal, time-invariant, finite-memory, continuous, nonlinear system. In orthogonal FLiP filters the basis functions are orthogonal for some stochastic distribution of the input signal. For example, EMFN and LN filters have orthogonal basis functions for a white uniform input signal in $\mathbb{R}_1$, CN filters for a white signal with the distribution in (2.32) in $\mathbb{R}_1$, WN filters for a white Gaussian input signal. The orthogonality of the basis functions guarantees a fast convergence of gradient descent adaptive identification algorithms and allows the identification of nonlinear filters with the cross-correlation method.

The cross-correlation method applied to stochastic inputs requires a huge amount of input samples for an accurate estimation of the coefficients. To solve this problem, PPSs have been proposed since

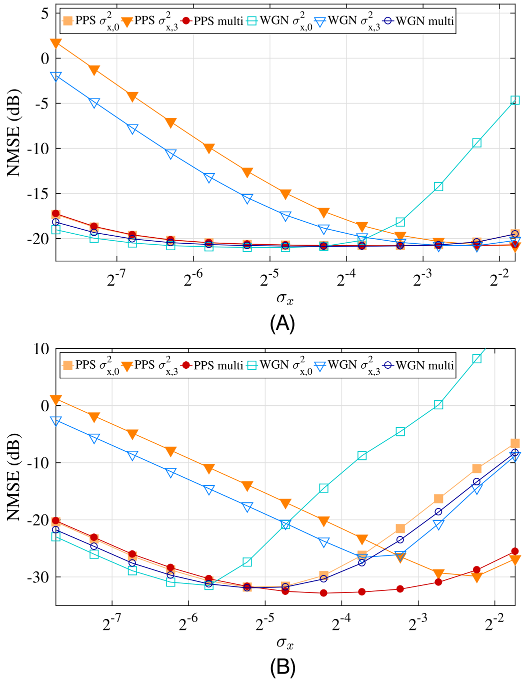

FIGURE 2.3

NMSE obtained for (A) noise and (B) music signals.

they are deterministic sequences that guarantee orthogonality of basis functions on a finite period. The PPSs allow also to avoid problems in the estimation of kernel diagonal points, which afflict the classical cross-correlation methods.

The nonlinear filter estimation is often affected by the problem of locality of the solution: the estimated filter well models the unknown system only at the same input power used for the identification. Multiple-variance methods can be used in conjunction with PPSs to improve the accuracy of the estimation and contrast the solution locality.

Experimental results have illustrated nonlinear system identification with FLiP filters using PPSs and the multiple-variance technique, highlighting the benefits provided by these methodologies.

REFERENCES

[1] V. Volterra, Sopra le funzioni che dipendono da altre funzioni, in: Rend. R. Accademia dei Lincei, vol. III of IV, 1887, pp. 97–105.

[2] V. Volterra, Leçons sur les Fonctions de Lignes, Gauthier-Villars, Paris, 1913.

[3] M. Frechét, Sur les fonctionnelles continues, in: Annales scientifiques de l'École Normale Supérieure, vol. 27 of 3, 1910, pp. 193–216.

[4] S. Boyd, L.O. Chua, Fading memory and the problem of approximating nonlinear operators with Volterra series, IEEE Transactions on Circuits and Systems CAS-32 (11) (1985) 1150–1161.

[5] I.W. Sandberg, $\mathbb{R}_+$ fading memory and extensions of input–output maps, IEEE Transactions on Circuits and Systems I 49 (11) (2002) 1586–1592.

[6] N. Wiener, Nonlinear Problems in Random Theory, John Wiley, New York, 1958.

[7] M.B. Brilliant, Theory of the Analysis of Nonlinear Systems, RLE Technical Report 345, MIT, Cambridge, MA, 1958.

[8] Y.W. Lee, M. Schetzen, Measurement of the Wiener kernels of a nonlinear system by crosscorrelation, International Journal of Control 2 (3) (1965) 237–254.

[9] G. Palm, T. Poggio, The Volterra representation and the Wiener expansion: validity and pitfalls, SIAM Journal of Applied Mathematics 33 (2) (1977) 195–216.

[10] G. Palm, T. Poggio, Stochastic identification methods for nonlinear systems: an extension of the Wiener theory, SIAM Journal of Applied Mathematics 34 (3) (1978) 524–534.

[11] 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.

[12] Y. Goussard, W.C. Krenz, L. Stark, An improvement of the Lee and Schetzen cross-correlation method, IEEE Transactions on Automatic Control AC-30 (9) (1985) 895–898.

[13] 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.

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

[15] A. Carini, S. Cecchi, L. Romoli, S. Orcioni, Perfect periodic sequences for nonlinear Wiener filters, in: Proc. of EUSIPCO 2016, European Signal Processing Conference, Budapest, Hungary, 2016.

[16] K. Narendra, P. Gallman, An iterative method for the identification of nonlinear systems using a Hammerstein model, IEEE Transactions on Automatic Control 11 (3) (1966) 546–550.

[17] E.L.O. Batista, R. Seara, A new perspective on the convergence and stability of NLMS Hammerstein filters, in: Proc. of ISPA 2013, 8th International Symposium on Image and Signal Processing and Analysis, Trieste, Italy, 2013, pp. 336–341.

[18] M. Gasparini, L. Romoli, S. Cecchi, F. Piazza, Identification of Hammerstein model using cubic splines and FIR filtering, in: Proc. of ISPA 2013, 8th International Symposium on Image and Signal Processing and Analysis, Trieste, Italy, 2013, pp. 347–352.

[19] J.M. Gil-Cacho, T. van Waterschoot, M. Moonen, S.H. Jensen, Linear-in-the-Parameters Nonlinear Adaptive Filters for Loudspeaker Modeling in Acoustic Echo Cancellation, Internal Report 13-110, ESAT-SISTA, K.U. Leuven, Leuven, Belgium.

[20] L. Romoli, M. Gasparini, S. Cecchi, A. Primavera, F. Piazza, Adaptive identification of nonlinear models using orthogonal nonlinear functions, in: Proc. AES 48th Int. Conf., Munich, Germany, Sep. 21–23, 2012, pp. 1597–1601.

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

[22] 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 Transactions on Signal Processing 54 (2006) 3852–3860.

[23] Y.H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[24] J.C. Patra, R.N. Pal, A functional link artificial neural network for adaptive channel equalization, Signal Processing 43 (1995) 181–195.

[25] J. Patra, A. Kot, Y.Q. Chen, Chebyshev functional link artificial neural networks for nonlinear dynamic system identification, in: Proc. of IEEE International Conference on Systems, Man, and Cybernetics, 2000, vol. 4, 2000, pp. 2655–2660.

[26] J. Patra, W. Chin, P. Meher, G. Chakraborty, Legendre-FLANN-based nonlinear channel equalization in wireless communication system, in: Proc. of SMC 2008, IEEE International Conference on Systems, Man and Cybernetics, 2008, pp. 1826–1831.

[27] H. Zhao, J. Zhang, Functional link neural network cascaded with Chebyshev orthogonal polynomial for nonlinear channel equalization, Signal Processing 88 (2008) 1946–1957.

[28] H. Zhao, J. Zhang, Adaptively combined FIR and functional link artificial neural network equalizer for nonlinear communication channel, IEEE Transactions on Neural Networks 20 (4) (2009) 665–674.

[29] K. Das, J. Satapathy, Legendre neural network for nonlinear active noise cancellation with nonlinear secondary path, in: Proc. of IMPACT 2011, International Conference on Multimedia, Signal Processing and Communication Technologies, 2011, pp. 40–43.

[30] N. George, G. Panda, A reduced complexity adaptive Legendre neural network for nonlinear active noise control, in: Proc. of IWSSIP 2012, International Conference on Systems, Signals and Image Processing, 2012, pp. 560–563.
[31] M.D. Buhmann, Radial Basis Functions: Theory and Implementations, University Press, Cambridge, 2003.
[32] G.L. Sicuranza, A. Carini, On a class of nonlinear filters, in: M.G.I. Tabus, K. Egiazarian (Eds.), Festschrift in Honor of Jaakko Astola on the Occasion of His 60th Birthday, vol. TICSP Report #47, 2009, pp. 115–144.
[33] A. Carini, G.L. Sicuranza, Fourier nonlinear filters, Signal Processing 94 (2014) 183–194.
[34] A. Carini, G.L. Sicuranza, Even mirror Fourier nonlinear filters, in: Proc. of ICASSP 2013, International Conference on Acoustic, Speech, Signal Processing, Vancouver, Canada, 2013, pp. 5608–5612.
[35] A. Carini, S. Cecchi, M. Gasparini, G.L. Sicuranza, Introducing Legendre nonlinear filters, in: Proc. of ICASSP 2014, International Conference on Acoustic, Speech, Signal Processing, Florence, Italy, 2014, pp. 7989–7993.
[36] A. Carini, S. Cecchi, L. Romoli, G.L. Sicuranza, Legendre nonlinear filters, Signal Processing 109 (2015) 84–94.
[37] A. Carini, G.L. Sicuranza, A study about Chebyshev nonlinear filters, Signal Processing 122 (2016) 24–32.
[38] A. Carini, G.L. Sicuranza, Perfect periodic sequences for identification of even mirror Fourier nonlinear filters, in: Proc. of ICASSP 2014, International Conference on Acoustic, Speech, Signal Processing, Florence, Italy, 2014, pp. 8009–8013.
[39] A. Carini, G.L. Sicuranza, Perfect periodic sequences for even mirror Fourier nonlinear filters, Signal Processing 104 (2014) 80–93.
[40] A. Carini, S. Cecchi, L. Romoli, G.L. Sicuranza, Perfect periodic sequences for Legendre nonlinear filters, in: Proc. of EUSIPCO 2014, European Signal Processing Conference, Lisbon, Portugal, 2014.
[41] G.L. Sicuranza, A. Carini, Nonlinear system identification using quasi-perfect periodic sequences, Signal Processing 120 (2016) 174–184.
[42] W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York, 1976.
[43] N.D. Vanli, S.S. Kozat, A comprehensive approach to universal piecewise nonlinear regression based on trees, IEEE Transactions on Signal Processing 62 (20) (2014) 5471–5485.
[44] J.C. Adams, On the expression of the product of any two Legendre's coefficients by means of a series of Legendre's coefficients, Proceedings of the Royal Society of London 27 (1878) 63–71.
[45] T.J. Rivlin, Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory, John Wiley & Sons, New York, USA, 1990.
[46] A. Gil, J. Segura, N. Temme, Numerical Methods for Special Functions, Society for Industrial and Applied Mathematics, 2007.
[47] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables, Dover Publications, New York, 1972.
[48] S. Roman, The Umbral Calculus, Academic Press, San Diego, 1984.
[49] V.J. Mathews, G.L. Sicuranza, Polynomial Signal Processing, Wiley, New York, 2000.
[50] A. Fermo, A. Carini, G.L. Sicuranza, Simplified Volterra filters for acoustic echo cancellation in GSM receivers, in: Proc. of EUSIPCO 2000, European Signal Processing Conference, Tampere, Finland, Sept. 5–8, 2000.
[51] G.M. Raz, B.D.V. Veen, Baseband Volterra filters for implementing carrier based nonlinearities, IEEE Transactions on Signal Processing 46 (1) (1998) 103–114.
[52] L.A. Azpicueta-Ruiz, M. Zeller, A.R. Figueiras-Vidal, J. Arenas-Garcia, W. Kellermann, Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancellation, IEEE Transactions on Audio, Speech and Language Processing 19 (11) (2011) 97–110.
[53] A. Fermo, A. Carini, G.L. Sicuranza, Low-complexity nonlinear adaptive filters for acoustic echo cancellation in GSM handset receivers, European Transactions on Telecommunications 14 (2) (2003) 161–169.
[54] G.L. Sicuranza, A. Carini, A new recursive controller for nonlinear active noise control, in: Proc. ISPA 2013, 8th International Symposium on Image and Signal Processing and Analysis, Trieste, Italy, 2013, pp. 626–631.
[55] A. Carini, G.L. Sicuranza, Recursive even mirror Fourier nonlinear filters and simplified structures, IEEE Transactions on Signal Processing 62 (24) (2014) 6534–6544.
[56] V. Ipatov, Ternary sequences with ideal periodic autocorrelation properties, Radio Engineering and Electronic Physics 24 (1979) 75–79.
[57] A. Milewski, Periodic sequences with optimal properties for channel estimation and fast start-up equalization, IBM Journal of Research and Development 27 (5) (1983) 426–431.
[58] R.H. Kwong, E.W. Johnston, A variable step size LMS algorithm, IEEE Transactions on Signal Processing 40 (1992) 1633–1642.

[59] 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.
[60] 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.
[61] 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.
[62] C. Antweiler, A. Telle, P. Vary, NLMS-type system identification of MISO systems with shifted perfect sequences, in: Proc. IWAENC 2008, International Workshop on Acoustic Echo and Noise Control, Seattle, Washington, USA, 2008.
[63] A. Carini, Efficient NLMS and RLS algorithms for perfect and imperfect periodic sequences, IEEE Transactions on Signal Processing 58 (4) (2010) 2048–2059.
[64] 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.
[65] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Signal Processing 19 (6) (1974) 716–723.
[66] E.J. Hannan, B.G. Quinn, The determination of the order of an autoregression, Journal of the Royal Statistical Society. Series B 41 (2) (1979) 190–195.
[67] G. Schwartz, Estimating the dimension of a model, The Annals of Statistics 6 (1978) 416–464.
[68] S. Orcioni, L. Romoli, S. Cecchi, A. Carini, Multivariance nonlinear system identification using Wiener basis functions and perfect sequences, in: Proc. of EUSIPCO 2017, European Signal Processing Conference, 2017.
[69] K. Li, J.-X. Peng, G.W. Irwin, A fast nonlinear model identification method, IEEE Transactions on Automatic Control 50 (8) (2005) 1211–1216.