Nonlinear System Identification: An Effective Framework Based on the Karhunen–Loève Transform

Claudio Turchetti, *Member, IEEE*, Giorgio Biagetti, *Member, IEEE*, Francesco Gianfelici, and Paolo Crippa, *Member, IEEE*

***Abstract*—This paper proposes, on the basis of a rigorous mathematical formulation, a general framework that is able to define a large class of nonlinear system identifiers. This framework exploits all those relationships that intrinsically characterize a limited set of realizations, obtained by an ensemble of output signals and their parameterized inputs, by means of the separation property of the Karhunen–Loève transform. The generality and the flexibility of the approximating mappings (ranging from traditional approximation techniques to multiresolution decompositions and neural networks) allow the design of a large number of distinct identifiers each displaying a number of properties such as linearity with respect to the parameters, noise rejection, low computational complexity of the approximation procedure. Exhaustive experimentation on specific case studies reports high identification performance for four distinct identifiers based on polynomials, splines, wavelets and radial basis functions. Several comparisons show how these identifiers almost always have higher performance than that obtained by current best practices, as well as very good accuracy, optimal noise rejection, and fast algorithmic elaboration. As an example of a real application, the identification of a voice communication channel, comprising a digital enhanced cordless telecommunications (DECT) cordless phone for wireless communications and a telephone line, is reported and discussed.**

***Index Terms*—Hilbert space, Karhunen–Loève transform (KLT), nonlinear approximation, nonlinear mapping, nonlinear system identification, radial basis functions (RBF), statistical signal processing, wavelet approximation.**

NOMENCLATURE

$\mathbb{R}$	The set of real numbers.
a	Deterministic scalar $a \in \mathbb{R}$.
a	Deterministic vector $\mathrm{a} = [a_1 \cdots a_N]^{\mathrm{T}} \in \mathbb{R}^N$.
A	Deterministic matrix $\mathrm{A} = [\mathrm{a}_1 \ldots \mathrm{a}_M] \in \mathbb{R}^{N \times M}$.
$\mathcal{F}\{\cdot\}$	Functional.
$\boldsymbol{y}$	Scalar random variable or scalar stochastic process.
$\mathbf{y}$	Vector random variable or vector stochastic process $\mathbf{y} = [\boldsymbol{y}_1 \cdots \boldsymbol{y}_N]^{\mathrm{T}}$.
$\mathcal{G}[\cdot]$	Nonlinear operator.
$E\{\cdot\}$	Statistical expectation.
$\mathrm{R}_{\mathbf{yy}}$	Autocorrelation matrix of $\mathbf{y}$.
$\otimes$	Kronecker product.
P^{+}	Moore–Penrose pseudo-inverse of the matrix P.
$\lfloor\cdot\rfloor$	Floor function: greatest integer function.

Manuscript received February 04, 2008; revised September 26, 2008. First published November 07, 2008; current version published January 30, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vitor Heloiz Nascimento.

The authors are with the Dipartimento di Ingegneria Biomedica, Elettronica e Telecomunicazioni, Università Politecnica delle Marche, I-60131 Ancona, Italy (e-mail: c.turchetti@univpm.it; g.biagetti@deit.univpm.it; f.gianfelici@deit.univpm.it; p.crippa@univpm.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.2008964

I. INTRODUCTION

NONLINEAR system identification (NSI) is one of the key issues in the modeling of signals generated by artificial systems and natural phenomena. The reason for the great interest in this field [1] is closely related to the intrinsic nonlinear nature of real phenomena, making the linear hypothesis simply an approximation of real behavior. As applicative examples, we can recall (to mention just a few), the following: i) neuroscience [2] where the pioneering works of Marmarelis *et al.* [3], [4] have had a decisive impact on the study of neuronal activation generated by perceptive *stimuli*; ii) the automatic control sector, where the milestone developed by Lee and Schetzen [5] opened new research directions and effective applications [6]; iii) communications that inspired Lee [7] and where the identification techniques have been used to model channels with high distortions [8], nonlinear amplifiers in the transmission stage [9], pass-band nonlinear channels [10], channel equalizations [11], [12], echo-cancellations in GSM receivers [13], and pre-distortion modeling of radio frequency (RF) amplifiers [14]; iv) image processing, thanks to the works of Ramponi *et al.* in the fields of edge extraction [15], image enhancement [16], and prediction of TV images [17], and of Thurnhofer *et al.* in the field of edge enhancement [18]; and v) signal processing where there are well-known applications in echo-cancellation and active noise control both in mono-channel [19] and multichannel settings [20], [21], and in the traditional problem of the electrodynamic loudspeaker [22].

On the basis of the above considerations, let us classify the current state of the art according to the milestones that have been developed since 1965, when Lee and Schetzen [5] published their decisive technique. In agreement with the above statement, it is possible to state that a first trend was generated by the above-mentioned Lee–Schetzen method that identifies the Volterra kernels of nonlinear systems stimulated by random input with assigned statistics [5], [23]. During recent decades, a large number of different approaches and techniques, also known as *polynomial signal processing*, have been developed, among which the recent works by Carini *et al.* [24], [25] are worth noting. A second trend was stimulated by, on the one hand, the works of Billings *et al.* [26]–[29], that proposed an extension of the linear formulation of systems commonly adopted

in control theory to the nonlinear case, giving rise to the development of the widespread nonlinear auto-regressive moving-average with exogenous input (NARMAX) model, and, on the other hand, by the works of Narendra *et al.* [30], where recurrent neural networks are seen as nonlinear dynamical networks, in which nonlinearity is implemented by means of static multilayer networks in the feedback loop. This approach has determined a dichotomic relationship between identification and the learning process and has led to important contributions [31], [32] on both *real-time* and *adaptive* identification techniques. In this context, it is necessary to remember the intrinsic relationship between identification and regularization which has been extensively studied by Poggio *et al.* [33]. The above-mentioned trends have determined the development of suitable algorithms that are closely related to the adaptive filter theory [34], and are usually based on iterative refinement for finding the optimal values of model parameters such as the least mean square (LMS), recursive least square (RLS), normalized least mean square (NLMS), affine projection (AP), and filtered-X affine projection algorithms [24]. These algorithms have high computational efficiency, but identification has satisfactory results mainly for mildly nonlinear systems with a low number of parameters, and for the low-order kernels of the Volterra series (VSs) [35].

Recent decades of advanced research studies in signal processing and automatic control have not yet solved the main limitations affecting the above-mentioned approaches. These limitations can be divided into the following five categories:

1) the well-known algorithmic complexity which limits the application to the first orders of VSs, with high dependence and great sensitivity with respect to the fading memory, and the signal length (this aspect prevents, on the one hand, their application to all those systems and natural phenomena where it is not possible to select *a priori* only the first VS orders, and on the other hand, their engineering in *embedded systems* with low capabilities of memory and elaboration power);
2) the need for an *a priori* knowledge of the system features and/or the mathematical properties that commonly affect the mathematical modeling, and thus the identification technique that has to be used, such as the well-known distinction between time-variant and time-invariant systems;
3) the need for an *a priori* knowledge of the model order, in order to prevent the well-known effects of under- or over-parameterizations, which has also determined the development of a large number of statistical methodologies that are able to select the optimal model order, based on Akaike's information criterion [36], Rissanen's minimum description length criterion [37], and Hannan–Quinn's criterion [38];
4) the absence of an effective criterion that allows us to establish which systems and/or physical phenomena can be identified with a specific technique based only on the processing of realization *ensembles*;
5) the dearth of general frameworks that are able to define large classes of identification algorithms.

These limitations have affected the development of new identification techniques, and they are known as the current *open problems* of NSI.

The great interest in the intrinsic properties of nonlinear systems and the absence of a rigorous closed-form formulation of these problems represent the starting points for the formalization and development of specific approaches and suitable frameworks, following the decisive contributions that these kinds of methodologies have generated in other contexts [39]. Moreover, these properties are one of the main principles behind the development of new probabilistic characterization on limited *ensembles* of realizations, as clearly stated in [40].

This paper presents, on the basis of a rigorous mathematical formulation, an effective framework for NSI that is able to define a large class of nonlinear system identifiers. Methodologically, we define a general theory based on a stochastic setting where the nonlinear systems analyzed generate nondeterministic signals, i.e., stochastic processes (SPs), from given initial conditions and random parameters of input signals. Based on this premise, the abovementioned geometrical relationships are then extracted in these Hilbert spaces by means of the Karhunen–Loève transform (KLT), thanks to its separability property. The KLT permits a large class of approximating *mappings* to be defined on specific input domains, and thus allows matching system identification algorithms to be developed. This framework[1] allows the best identification to be achieved, with a fixed rank on a chosen ensemble of realizations, for all the systems which show geometrical relationships in Hilbert spaces and with no constraints in terms of model kind and/or order. The generality and the flexibility of approximating *mappings*, ranging from traditional approximation theory to multiresolution decompositions and neural networks, allow the design of a large number of distinct identifiers covering a broad spectrum of properties, such as linearity with respect to the parameters, noise rejection, and computational complexity of the approximation procedure. Moreover, this class of identifiers does not require an *a priori* knowledge of system features or distinctions between time-variant and time-invariant systems, and, unlike other techniques, model selection criteria are not needed.

Exhaustive experimentation based on four distinct identifiers, obtained using polynomial basis functions $(\mathrm{FRM}_{\mathrm{PLY}})$, splines $(\mathrm{FRM}_{\mathrm{SPL}})$, wavelets $(\mathrm{FRM}_{\mathrm{WLT}})$, and radial basis functions $(\mathrm{FRM}_{\mathrm{RBF}})$, was performed on specific case studies showing high identification performance, also in case of limited ensembles of realizations. Several comparisons with the LMS, RLS, and NLMS algorithms demonstrate the good identification performance of the proposed approach. It is also worth noting that comparisons with NARMAX modeling show that better results can be achieved in most of the cases studied, both with or without additional output noise, with an excellent *tradeoff* between superimposed noise and accuracy obtained. Moreover, computational requirements are lower than that of LMS, NLMS, and RLS algorithms, and of the NARMAX identifier.

[1]A patent application of the proposed framework has been deposited [41] by Gianfelici and Turchetti according to the Italian patent law.

As an example of a real application, the identification of a voice communication channel comprising a digital enhanced cordless telecommunications (DECT) cordless phone for wireless communications and a telephone line, is reported and discussed.

This paper is organized as follows. Section II proposes the KLT-based identification framework for nonlinear systems. In Section III some identification algorithms are presented. Section IV reports the experimental results while Section V gives some exhaustive comparisons with the *state of the art*. Section VI discusses the application to a physical system, and Section VII concludes the work.

II. KLT-Based Identification Framework

In this paper we will consider discrete-time nonlinear dynamical systems that can be described by the following relation:

$$y(n) = \mathcal{F}\{e(\eta)|_{[0,n]}, n, \mathrm{c}\}, \quad n \in [0, L-1] \tag{1}$$

where $e(\eta)$ represents the input, with the notation $\cdot|_{[0,n]}$ meaning that η ranges in the interval $[0, n]$, $y(n)$ is the output, c is the vector of the initial conditions, and $\mathcal{F}$ is a functional that describes the input/output relationship in the time interval $[0, L-1]$.

In order to gain simplicity in the notation, we assume u and y are both one-dimensional, the extension to the general case being straightforward. The identification of (1) requires the system must be stimulated by a set $\mathcal{U}$ of input signals $e \in \mathcal{U}$ so that the corresponding y's, which vary within a set $\mathcal{V}$, can be measured at the output. The more the space $\mathcal{U}$ used for exciting the system is wide the more the identification so obtained is close to a complete description of the system. Unfortunately an exhaustive stimulation of the system is in general impracticable. Nevertheless, in most of application problems a complete description of the system is not required, since it suffices to restrict the input space to the subset $\mathcal{E} \subset \mathcal{U}$ of signals actually occurring in the problem under observation.

The identification of systems can thus be more effective if the input signal e is regarded as a stochastic process $\boldsymbol{e}$, and the vector c of the initial conditions is also reckoned as a vector random variable $\mathbf{c}$. As a consequence also the output y is an SP $\boldsymbol{y}$.

Throughout this paper, boldface letters are used to denote random variables (RVs) and SPs, non-boldface letters are used for deterministic quantities and for SP realizations, lowercase upright letters are used to denote vectors, and capital upright letters are used to denote matrices.

In addition let us consider $\mathcal{E}$ is spanned by the signals $\boldsymbol{e}(n, \mathbf{a})$, where $\mathbf{a} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_R]^{\mathrm{T}}$ is a vector of R RVs that parameterize the process $\boldsymbol{e}(n)$, and the function $\boldsymbol{e}(\cdot, \cdot)$ is assumed to be known.

Correspondingly, $\boldsymbol{y}(n)$ varies within a subset $\mathcal{Y} \subset \mathcal{V}$ so that (1) establishes a transformation between SPs given by

$$\boldsymbol{y}(n) = \mathcal{F}\{\boldsymbol{e}(\eta, \mathbf{a})|_{[0,n]}, n, \mathbf{c}\}, \quad e \in \mathcal{E}, y \in \mathcal{Y}\ ,\ n \in [0, L-1] \tag{2}$$

where all the parameters related to both the input process and the initial conditions can be collected in a single vector $\mathbf{x}$ defined as $\mathbf{x} = [\mathbf{a}^{\mathrm{T}}\ \mathbf{c}^{\mathrm{T}}]^{\mathrm{T}}$. As the expression for $\boldsymbol{e}(n, \mathbf{a})$ is substituted in (2), the system output becomes a function of the time n and vector $\mathbf{x}$ alone, and it is no more explicitly depending on the variable $\boldsymbol{e}$, so we can simple write $\boldsymbol{y}(n) = \boldsymbol{y}(n, \mathbf{x})$ instead of (2). This assumption is equivalent, in some cases, to restricting the input space only to a certain class of signals. A trivial example is the class of multicomponent sinusoidal signals represented by

$$\boldsymbol{e}(n, \mathbf{a}) = \sum_{i=1}^{R} A_i \sin(\boldsymbol{\alpha}_i n). \tag{3}$$

In this case assuming the ith sinusoidal component has a fixed amplitude A_i and a random frequency $\boldsymbol{\alpha}_i$, then it results

$$\mathbf{a} = [\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2 \cdots \boldsymbol{\alpha}_R]^{\mathrm{T}}. \tag{4}$$

In some cases, even if this assumption is not true, the partial description of the system so obtained suffices for the application purposes. An approach of this kind is used, for instance, in the speech production modeling, in which the excitation is represented by a multicomponent sinusoidal signal of arbitrary amplitudes, frequencies, and phases [42].

Nevertheless, it is worth noting that, in general, given a class of signals $\boldsymbol{e} \in \mathcal{E}$, if $\mathcal{E}$ is a Hilbert space, they always can be represented by the discrete Karhunen–Loève transform (DKLT) [43], also called *canonical representation*, which is defined by

$$\boldsymbol{e}(n, \mathbf{a}) = \sum_{k=1}^{R} \boldsymbol{a}_k \psi_k(n), \quad n \in [0, L-1] \tag{5}$$

where $\psi_k(n), k = 1, \ldots, R$ are orthonormal functions and $R \leq L$. As a consequence the subset $\mathcal{E}$ is spanned by the signals $\boldsymbol{e}(n, \mathbf{a})$ given by (5) with $\mathbf{a} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_R]^{\mathrm{T}}$ and this proves that the assumption $\boldsymbol{e}(n, \mathbf{a})$ is a general representation of the input space.

As previously mentioned, for the sake of notational simplicity, in the rest of this paper we will refer to the input parameters $\mathbf{a}$ and to the initial conditions $\mathbf{c}$ together, as to vector $\mathbf{x}$. For example, if we consider a simple dynamical system composed of a pendulum with a sinusoidal forcing term of random frequency $\boldsymbol{\alpha}_1$, as in (3) with $R = 1$, the initial conditions $\mathbf{c}$ will be $[\boldsymbol{\vartheta}(0)\ \dot{\boldsymbol{\vartheta}}(0)]^{\mathrm{T}}$, where ϑ is the displacement angle, and the input parameter vector $\mathbf{a}$ will contain just $\boldsymbol{\alpha}_1$. Hence, $\mathbf{x} = [\boldsymbol{\alpha}_1\ \boldsymbol{\vartheta}(0)\ \dot{\boldsymbol{\vartheta}}(0)]^{\mathrm{T}}$ will contain all the non-time varying parameters on which the system output $\boldsymbol{y}$ depends.

With the above considerations in mind, in order to tackle the identification problem, as the output $\boldsymbol{y}(n, \mathbf{x})$ is a stochastic process depending on the random parameter vector $\mathbf{x}$, and belonging to a Hilbert space, it can be represented by the DKLT as follows:

$$\boldsymbol{y}(n, \mathbf{x}) \approx \sum_{j=1}^{D} k_j(\mathbf{x}) \phi_j(n), \quad n \in [0, L-1] \tag{6}$$

where $\phi_j(n),\ j = 1, \cdots, D$ are orthonormal functions, and $D \leq L$. It is worth noting that the KLT is the most efficient representation of the SP if the expansion is truncated to use fewer than L orthonormal basis functions, and is exact if $D = L$.

The transformation (6) and its inverse can be written in matrix form as

$$\begin{cases} \mathbf{y} = \Phi \mathrm{k}(\mathbf{x}) \\ \mathrm{k}(\mathbf{x}) = \Phi^{\mathrm{T}} \mathbf{y} \end{cases} \tag{7}$$

where, denoting with $[.]_i$ the ith component of the vector in brackets, $\mathbf{y}$ is defined as $\mathbf{y} = [\boldsymbol{y}(0,\mathbf{x})\, \boldsymbol{y}(1,\mathbf{x}) \cdots \boldsymbol{y}(L-1,\mathbf{x})]^{\mathrm{T}}$, with its realizations $\mathrm{y} \in \mathbb{R}^{L\times 1}$, and $\mathrm{k}(\mathbf{x}) \in \mathbb{R}^{D\times 1}$ is defined as $[\mathrm{k}(\mathbf{x})]_j = k_j(\mathbf{x})$, with $j = 1, \ldots, D$.

$\Phi \in \mathbb{R}^{L\times D}$ is the reduced orthogonal matrix whose zjth element is $[\Phi]_{zj} = \phi_j(z)$, $z = 1, \ldots, L$, $j = 1, \ldots, D$ so that its columns are the eigenvectors as obtained from the eigenvalue equation

$$\mathrm{R}_{\mathbf{yy}} \Phi = \Phi \Gamma \tag{8}$$

where $\mathrm{R}_{\mathbf{yy}} \in \mathbb{R}^{L\times L}$ is the autocorrelation matrix of $\mathbf{y}$ that is estimated as

$$\mathrm{R}_{\mathbf{yy}} = E\{\mathbf{y}\mathbf{y}^{\mathrm{T}}\} \tag{9}$$

and $\Gamma \in \mathbb{R}^{D\times D}$ is the diagonal matrix with (non-null) eigenvalues on the diagonal. The mean-square error of the approximation (5) is equal to the sum of the residual eigenvalues, i.e., $\sum_{z=D+1}^{L} [\Gamma]_{zz}$, and thus the optimal choice of D depends on the desired degree of accuracy and on the eigenvalue spectrum.

The main benefit of this representation is related to the separation property of KLT. On the basis of this property, the output of the system can be expressed, as a linear combination of products of a function of $\mathbf{x}$ alone and a function of n alone, as it is clear from (6). Since the functions $\phi_j(n)$ are determined by means of $\mathrm{R}_{\mathbf{yy}}$, which can be estimated by the realizations of $\mathbf{y}$, the system identification reduces to modeling the functions $k_j(\mathbf{x})$, $j = 1, \ldots, D$. As $\mathbf{y}$ is a function of $\mathbf{x}$, the terms $k_j(\mathbf{x})$ describe on the space spanned by the columns of Φ the curves $\mathcal{C}^j_{\mathbf{y}}(\mathbf{x})$, which all together characterize the SP $\mathbf{y}$.

Under wide conditions, these curves show a smooth behavior so that they can be reconstructed from an *ensemble* of points extracted by the described approach to perform the identification. Since $\mathrm{k}(\mathbf{x})$ is a no-memory input–output mapping, it can be approximated by a given vector function

$$\mathrm{k}(\mathbf{x}) \approx \mathcal{G}[\mathbf{x}, \mathrm{W}] \tag{10}$$

where $\mathcal{G}[\cdot]$ is a nonlinear operator and $\mathrm{W} \in \mathbb{R}^{D\times M}$ is a matrix, of M parameters per eigenfunction, to be estimated.

On the basis of these results, a general *framework* can be defined that allows the use of any of the many known function approximation techniques to obtain a model for the operator $\mathcal{G}[\cdot]$, representing the starting point for the formalization and development of a large class of distinct identifiers. Indeed, it can be noted that (7) defines an isomorphism, that is, a one-to-one linear transformation from the space of $\mathrm{k}(\mathbf{x})$ onto the space of $\mathbf{y}$ that preserves the inner product. In fact, it results

$$\mathbf{y}^{\mathrm{T}}\mathbf{y} = [\Phi \mathrm{k}(\mathbf{x})]^{\mathrm{T}} \Phi \mathrm{k}(\mathbf{x}) = [\mathrm{k}(\mathbf{x})]^{\mathrm{T}} \Phi^{\mathrm{T}} \Phi \mathrm{k}(\mathbf{x}) = [\mathrm{k}(\mathbf{x})]^{\mathrm{T}} \mathrm{k}(\mathbf{x}) \tag{11}$$

being Φ a unitary matrix. This property is important to ensure that reducing the error in approximating $\mathrm{k}(\mathbf{x})$ consequently reduces the error in approximating $\mathbf{y}$.

This aspect allows a great flexibility in the definition of the approximation techniques that can be used to finalize the identification framework, as these techniques need only focus on the minimization of the error $\epsilon = ||\mathrm{k}(\mathbf{x}) - \mathcal{G}[\mathbf{x}, \mathrm{W}]||$. Their application naturally leads to the definition of various different algorithms within this framework. These algorithms can be classified according to several properties, such as linearity with respect to the parameters, the number of parameters that have to be estimated, i.e., the dimension of W, noise rejection, and the computational complexity of the estimation procedure.

Promising algorithms with optimal identification capabilities and covering a broad spectrum of the aforementioned properties can be designed using: i) traditional polynomial-based approximation techniques, ii) multiresolution decompositions, and iii) neural networks (NNs). Among these, neural networks are also particularly suitable to face this problem due to their ability to approximate complex nonlinear mappings directly from the input samples [44]. Earlier works have demonstrated that multilayer perceptrons (MLPs) [45]–[47] and radial basis function (RBF) networks [48] possess this property with reference to some classes of functions. These results show that NNs of these kinds are capable of approximating, arbitrarily well, any function belonging to these classes, with the degree of accuracy depending on the learning algorithm and the number of neurons available.

Without detriment to generality, four examples of identifiers originating from the proposed framework will be given in the following section.

III. Identification Algorithms

With the above considerations in mind, it can be stated that once the structure of the functional $\mathcal{G}[\mathbf{x}, \mathrm{W}]$ has been defined, the identification of the nonlinear system is equivalent to the estimation of the matrix W from an *ensemble* of the system's input-output pairs.

In order to derive several different identification algorithms, it is necessary to relate the stochastic setting (that allowed the development of the general theory) to the available *ensemble* of realizations. Methodologically, let us then refer to these N realizations of $\mathbf{x}$ as $\mathrm{x}^{(i)} \in \mathbb{R}^{M_{\mathrm{x}}\times 1}$, M_{x} being the dimension of vector $\mathbf{x}$, and to the corresponding realizations of $\mathbf{y}$ as $\mathrm{y}^{(i)} \in \mathbb{R}^{L\times 1}$, with $i = 1, \ldots, N$. Both can be put in matrix form as $\mathrm{X} = [\mathrm{x}^{(1)}\ \mathrm{x}^{(2)} \cdots\ \mathrm{x}^{(N)}]$ and $\mathrm{Y} = [\mathrm{y}^{(1)}\ \mathrm{y}^{(2)} \cdots\ \mathrm{y}^{(N)}]$, where $\mathrm{X} \in \mathbb{R}^{M_{\mathrm{x}}\times N}$ and $\mathrm{Y} \in \mathbb{R}^{L\times N}$. A currently used estimation $\hat{\mathrm{R}}_{\mathbf{yy}}$ of the autocorrelation matrix is given by

$$\mathrm{R}_{\mathbf{yy}} \approx \hat{\mathrm{R}}_{\mathbf{yy}} = \frac{1}{N} \mathrm{Y}\mathrm{Y}^{\mathrm{T}} \tag{12}$$

where $\mathrm{R}_{\mathbf{yy}} \in \mathbb{R}^{L\times L}$, and its spectral representation is

$$\hat{\mathrm{R}}_{\mathbf{yy}} \mathrm{U} = \mathrm{U}\Lambda \tag{13}$$

Algorithm initialization:
$\mathrm{Y} \leftarrow$ output realizations ▷ (one realization per column)
$\mathrm{X} \leftarrow$ input parameters ▷ (one realization per column)
KLT analysis:
$\mathrm{m} \leftarrow$ mean of the columns of Y
$\mathrm{Y} \leftarrow \mathrm{Y} - \mathrm{m}$ ▷ (for each column)
$\mathrm{U\,S\,V}^{\mathrm{T}} \leftarrow \mathrm{svd(Y)}$
trim matrix U to preserve only the first D columns
$\mathrm{K} \leftarrow \mathrm{U}^{\mathrm{T}}\,\mathrm{Y}$
Trajectory interpolation:
$\mathrm{W} \leftarrow \mathrm{FRM}_{\mathrm{alg}}\mathrm{\_fit(K, X)}$
System output reconstruction:
$\mathrm{k} \leftarrow \mathrm{FRM}_{\mathrm{alg}}\mathrm{\_eval(W, x)}$
$\mathrm{y} \leftarrow \mathrm{m} + \mathrm{U\,k}$

Fig. 1. Pseudocode of the KLT-based identification framework.

where $\mathrm{U} = [\mathrm{u}_1\ \mathrm{u}_2\ \cdots\ \mathrm{u}_D] \in \mathbb{R}^{L\times D}$ is the matrix of eigenvectors and $\Lambda \in \mathbb{R}^{D\times D}$ the matrix of eigenvalues, where now, because of the approximation (12), $D \leq \min(L, N)$.

By projecting all the N realizations onto the basis U, we obtain the KLT representation

$$\begin{cases} \mathrm{y}^{(i)} = \mathrm{U}\mathrm{k}^{(i)} \\ \mathrm{k}^{(i)} = \mathrm{U}^{\mathrm{T}}\mathrm{y}^{(i)} \end{cases} \quad i = 1, \ldots, N \tag{14}$$

and, in matrix form

$$\mathrm{K} = \mathrm{U}^{\mathrm{T}}\,\mathrm{Y} \tag{15}$$

where $\mathrm{K} = [\mathrm{k}^{(1)}\ \mathrm{k}^{(2)} \cdots\ \mathrm{k}^{(N)}] \in \mathbb{R}^{D\times N}$.

Once these projections have been obtained, the problem of approximating $\mathrm{k}(\mathbf{x})$ by a given function $\mathcal{G}[\mathbf{x}, \mathrm{W}]$ corresponds to find the parameters W that make the following approximation:

$$\mathrm{k}^{(i)} \approx \mathcal{G}[\mathrm{x}^{(i)}, \mathrm{W}], \quad i = 1, \ldots, N \tag{16}$$

hold.

The estimation algorithm is particularly simple if $\mathcal{G}[\mathbf{x}, \mathrm{W}]$ is a linear function of W, namely

$$\mathcal{G}[\mathbf{x}, \mathrm{W}] = \mathrm{W}\,\mathrm{g}(\mathbf{x}) \tag{17}$$

with $\mathrm{g}(\mathbf{x})$ being an M-dimensional vector of suitable functions. By sampling these functions in correspondence of the N realizations of the parameterized input values and initial conditions vector, $\mathrm{x}^{(i)}$, $i = 1, \ldots, N$, we obtain the matrix $\mathrm{G} = [\mathrm{g}(\mathrm{x}^{(1)})\ \mathrm{g}(\mathrm{x}^{(2)})\ \cdots\ \mathrm{g}(\mathrm{x}^{(N)})] \in \mathbb{R}^{M\times N}$ and the estimation problem reduces to estimating W so that

$$\mathrm{K} \approx \mathrm{W\,G}. \tag{18}$$

Different choices of the nonlinear functions $\mathrm{g}(\mathbf{x})$ will therefore lead to the design of distinct identifiers. Their mathematical treatment and the exact form which the resulting matrix G assumes in their context will be presented in the following, while the main flow of the identification procedure is shown in pseudocode form in Fig. 1, which invokes one of the interpolation techniques that will be described later and that are here generically denoted by the symbols $\mathrm{FRM}_{\mathrm{alg}}$_fit and $\mathrm{FRM}_{\mathrm{alg}}$_eval. Here, SVD was used to implement KLT, as to the best of the authors' knowledge, it is one of the most efficient techniques to implement it.

A. *Polynomial Approximation-Based Identifier*—$\mathrm{FRM}_{\mathrm{PLY}}$

The first identification algorithm, called $\mathrm{FRM}_{\mathrm{PLY}}$, is defined by means of an approximating *mapping* based on a traditional approximation technique—the polynomial interpolation. Due to the well-known approximating properties of polynomials, they might represent a suitable choice as basis functions to model the nonlinearity in the identification problem, if the number of oscillations of the function $\mathrm{k}(\mathbf{x})$, and therefore the polynomial degree, is not too large. Therefore, a polynomial basis of maximum degree p for the function $\mathrm{g}(\mathbf{x})$ in (17) has been adopted, that is

$$\mathrm{g}(\mathbf{x}) = [1\ \mathbf{x}^{\mathrm{T}}\ (\mathbf{x}^2)^{\mathrm{T}}\ \cdots\ (\mathbf{x}^p)^{\mathrm{T}}]^{\mathrm{T}}. \tag{19}$$

Here, powers of vectors are assumed to be defined as the iterated application of the Kronecker product to the vector itself

$$\mathrm{x}^p = \underbrace{\mathrm{x} \otimes \mathrm{x} \otimes \cdots \otimes \mathrm{x}}_{p \text{ times}} \tag{20}$$

where the Kronecker product $\otimes$ of two generic vectors $\mathrm{a} \in \mathbb{R}^{L_{\mathrm{a}}\times 1}$ and $\mathrm{b} \in \mathbb{R}^{L_{\mathrm{b}}\times 1}$ is a vector $(\mathrm{a} \otimes \mathrm{b}) \in \mathbb{R}^{(L_{\mathrm{a}}\,L_{\mathrm{b}})\times 1}$ with $[\mathrm{a} \otimes \mathrm{b}]_{j+(i-1)L_{\mathrm{b}}} = [\mathrm{a}]_i\,[\mathrm{b}]_j$, $i = 1, \ldots, L_{\mathrm{a}}$, $j = 1, \ldots, L_{\mathrm{b}}$.

The matrix $\mathrm{W} = [\mathrm{w}_1\ \mathrm{w}_2\ \cdots\ \mathrm{w}_D]^{\mathrm{T}}$ can easily be determined by solving the overdetermined linear system $\mathrm{K} \approx \mathrm{W\,G}$ with the least mean square approach, that in this case corresponds to the back-propagation algorithm. To that end, minimizing the mean square error ϵ_j for the jth component

$$\epsilon_j = \sum_{i=1}^{N} \left[\varepsilon_j^{(i)}\right]^2 \quad \text{with} \quad \varepsilon_j^{(i)} = \mathrm{u}_j^{\mathrm{T}}\,\mathrm{y}^{(i)} - \mathrm{w}_j^{\mathrm{T}}\,\mathrm{g}(\mathrm{x}^{(i)}), \quad j = 1, \ldots, D, i = 1, \ldots, N \tag{21}$$

where $\mathrm{w}_j \in \mathbb{R}^{M\times 1}$ and $\mathrm{u}_j \in \mathbb{R}^{L\times 1}$ yields the following equation:

$$\frac{\partial \epsilon_j}{\partial \mathrm{w}_j} = 0 \tag{22}$$

which reduces to

$$\mathrm{Y}^{\mathrm{T}}\mathrm{u}_j = \mathrm{G}^{\mathrm{T}}\mathrm{w}_j, \quad j = 1, \ldots, D. \tag{23}$$

Finally the weights w_j are estimated by solving the D linear matrix (23), as detailed in the pseudocode shown in Fig. 2, where, for the sake of simplicity, the code is only reported for the case of a scalar variable x, as will be for the other two linear-in-the-parameters interpolators that follows.

B. *Spline Approximation-Based Identifier*—$\mathrm{FRM}_{\mathrm{SPL}}$

The second identification algorithm, called $\mathrm{FRM}_{\mathrm{SPL}}$, is defined by means of an approximating *mapping* based on another traditional approximation technique—the splines. This linear-in-the-parameter identifier is an extension of $\mathrm{FRM}_{\mathrm{PLY}}$, able to capture more complicated dynamics without incurring in the numerical instabilities of high-order polynomials. In fact, B-splines [49] proved to be very well suited to fit the smooth curves shown in the Hilbert space by the nonlinear systems we considered. For the sake of notational simplicity, we limit here

function $\mathrm{FRM_{PLY}}$_fit(K, X) ▷ Interpolation
  for all $\ell \in \{0, \ldots, p\}$ **do**
    for all $i \in \{1, \ldots, N\}$ **do**
      $[\mathrm{G}]_{\ell i} \leftarrow (\mathrm{x}^{(i)})^{\ell}$ ▷ Vandermonde matrix
    end for
  end for
  $\mathrm{U\,S\,V^T} \leftarrow \mathrm{svd(G)}$
  trim matrices U, S, and V, removing unnecessary rows and columns corresponding to null singular values
  $\mathrm{W} \leftarrow \mathrm{K\,V\,S^{-1}\,U^T}$
end function

function $\mathrm{FRM_{PLY}}$_eval(W, x) ▷ Reconstruction
  $\mathrm{k} \leftarrow \mathrm{W}\,[1\ \mathrm{x}^{\mathrm{T}}\ (\mathrm{x}^2)^{\mathrm{T}}\ \cdots\ (\mathrm{x}^p)^{\mathrm{T}}]^{\mathrm{T}}$
end function

Fig. 2. Pseudocode of the polynomial interpolation technique $\mathrm{FRM_{PLY}}$.

function $\mathrm{FRM_{SPL}}$_fit(K, X) ▷ Interpolation
  select knot sequence $\{\mathrm{x}^{(\ell)}, \ell = 1, \ldots, M\}$
  for all $\ell \in \{1, \ldots, M\}$ **do**
    for all $i \in \{1, \ldots, N\}$ **do**
      $[\mathrm{G}]_{\ell i} \leftarrow B_{\ell,p}(\mathrm{x}^{(i)})$ ▷ As defined in (25)
    end for
  end for
  $\mathrm{U\,S\,V^T} \leftarrow \mathrm{svd(G)}$
  trim matrices U, S, and V ▷ As in $\mathrm{FRM_{PLY}}$_fit
  $\mathrm{W} \leftarrow \mathrm{K\,V\,S^{-1}\,U^T}$
end function

function $\mathrm{FRM_{SPL}}$_eval(W, x) ▷ Reconstruction
  for all $j \in \{1, \ldots, D\}$ **do**
    $[\mathrm{k}]_j \leftarrow \sum_{\ell=1}^{M} [\mathrm{W}]_{j\ell}\, B_{\ell,p}(\mathrm{x})$
  end for
end function

Fig. 3. Pseudocode of the spline-based interpolation technique $\mathrm{FRM_{SPL}}$.

the definition of $\mathrm{FRM_{SPL}}$ to the case of a one-dimensional RV $\mathbf{x}$.

Starting from the framework introduced by (17), let us assume that, without detriment to generality, the realizations $\mathrm{x}^{(i)}$ are ordered, so that $\mathrm{x}^{(i)} < \mathrm{x}^{(i+1)}$. The spline-based mapping can then be obtained by letting

$$\mathrm{g}(\mathbf{x}) = [B_{1,p}(\mathbf{x})\ B_{2,p}(\mathbf{x})\ \cdots\ B_{N-p,p}(\mathbf{x})]^{\mathrm{T}} \tag{24}$$

where $B_{i,p}(\mathrm{x})$ is the pth-order B-spline (e.g., $p = 4$ for a cubic B-spline) defined over the knot sequence $\mathrm{x}^{(i)}, \ldots, \mathrm{x}^{(i+p)}$ and that is zero for $\mathrm{x} < \mathrm{x}^{(i)}$ and $\mathrm{x} > \mathrm{x}^{(i+p)}$, defined by the recurrent formula:

$$B_{i,p}(\mathrm{x}) = \frac{\mathrm{x} - \mathrm{x}^{(i)}}{\mathrm{x}^{(i+p-1)} - \mathrm{x}^{(i)}}\, B_{i,p-1}(\mathrm{x}) + \frac{\mathrm{x}^{(i+p)} - \mathrm{x}}{\mathrm{x}^{(i+p)} - \mathrm{x}^{(i+1)}}\, B_{i+1,p-1}(\mathrm{x}) \tag{25}$$

for $p > 1$, and

$$B_{i,1} = \begin{cases} 1, & \mathrm{x}^{(i)} \le \mathrm{x} < \mathrm{x}^{(i+1)} \\ 0, & \text{elsewhere.} \end{cases} \tag{26}$$

Thus the weight matrix W, being the matrix G completely defined by the sampling of the B-splines on the available realizations of $\mathbf{x}$, can then be determined exactly as for the polynomial case by solving the linear system $\mathrm{K} \approx \mathrm{W\,G}$, or, that is the same, by applying (21)–(23), as detailed in the pseudocode shown in Fig. 3.

C. *Wavelet Approximation-Based Identifier*—$\mathrm{FRM_{WLT}}$

The third identification algorithm, called $\mathrm{FRM_{WLT}}$, is defined by means of an approximating *mapping* based on multiresolution decompositions. This linear-in-the-parameter identifier is based on wavelets that bring the advantage of tunable regularization and denoising parameters, and also allow regularization and denoising to be performed simultaneously [50] in the same coefficient space. This allows the identifier to be defined as follows, where again for the sake of simplicity the mathematical formulation has been restricted to the case of a one-dimensional RV $\mathbf{x}$.

Let $\mathrm{H} = [\mathrm{h}_1\ \mathrm{h}_2\ \cdots\ \mathrm{h}_M] \in \mathbb{R}^{M\times M}$ be an orthonormal discrete wavelet transformation matrix for vectors of dimension M, ordered so that the lth row contains a scaled and shifted version of the mother wavelet at scale s_l. M is a power of two and must be large enough for the approximation $\mathrm{x}^{(i)} \approx x_L + (l^{(i)}/M)(x_U - x_L)$ to be reasonably accurate for $i = 1, \ldots, N$, with $l^{(i)} \in \{1, \ldots, M\}$ and x_L, x_U appropriate bounds such that $x_L \le \mathrm{x} \le x_U$. Let then the matrix $\mathrm{G} \in \mathbb{R}^{M\times N}$ defined in (18) be the "sampled" version of H, sampled in correspondence to the values of $\mathrm{x}^{(i)}$,

$$\mathrm{G} = [\mathrm{h}_{l^{(1)}}\ \mathrm{h}_{l^{(2)}}\ \cdots\ \mathrm{h}_{l^{(N)}}] \tag{27}$$

so that the approximation problem (18) can be restated as the problem of determining the coefficients w_j in the following inverse wavelet transform:

$$\mathrm{k}_j = \mathrm{G}^{\mathrm{T}}\,\mathrm{w}_j, \quad j = 1, \ldots, D \tag{28}$$

where $\mathrm{k}_j \in \mathbb{R}^{N\times 1}$ is the jth column of matrix K^{T}. The problem (28) is clearly underdetermined (M must usually be much greater than N for the sampling to be sufficiently accurate), so that a regularization criterion can be used to select the best solution.

As shown in [50], regularization can easily be performed in the wavelet domain by considering functions belonging to Sobolev spaces [51], [52]. These are spaces in which the definition of the norm also involves the derivatives of the functions up to a certain order, so that functions associated to smaller norms can be deemed to be "smoother." The norm of a function in a Sobolev space also corresponds to the Euclidean norm of its wavelet coefficients after suitable scaling, hence the usefulness of a wavelet representation for regularization.

Let $\mathrm{A} \in \mathbb{R}^{M\times M}$ be a scaling matrix

$$\mathrm{A} = \operatorname*{diag}_{l=1,\ldots,M} \{2^{\delta s_l}\} \tag{29}$$

where δ is a regularization parameter, with higher values corresponding to smoother functions, and the $\mathrm{diag}_{l=1,\ldots,M}\{\cdot\}$ operator is used to denote an $M \times M$ matrix with the specified elements on the main diagonal. The Sobolev norm then equals the Euclidean norm of $\mathrm{A}^{-1}\,\mathrm{w}_j$, and regularization corresponds to minimizing this quantity, that is, the interpolation problem reduces to a simple weighted least-squares problem. To solve it, it is enough to solve the system (28) by means of

function $\mathrm{FRM_{WLT}\_fit(K, X)}$ ▷ Interpolation
  for all $i \in \{1, \ldots, N\}$ **do** ▷ Compute indices
    $l^{(i)} \leftarrow 1 + \lfloor (M-1) \cdot (\mathrm{x}^{(i)} - x_L)/(x_U - x_L) + 1/2 \rfloor$
  end for
  detect and remove duplicates from $l^{(i)}$, $i = 1, \ldots, N$
  $\mathrm{G} \leftarrow [\mathrm{h}_{l^{(1)}} \cdots \mathrm{h}_{l^{(N)}}]$
  $\mathrm{P} \leftarrow \mathrm{G}^{\mathrm{T}}\mathrm{A}$
  $\mathrm{U\,S\,V}^{\mathrm{T}} \leftarrow \mathrm{svd(P)}$ ▷ for Moore-Penrose p.inv.
  trim matrices U, S, and V ▷ As in $\mathrm{FRM_{PLY}\_fit}$
  $\mathrm{P}^{+} \leftarrow \mathrm{V\,S}^{-1}\,\mathrm{U}^{\mathrm{T}}$
  $\mathrm{W}^{\mathrm{T}} \leftarrow \mathrm{A\,P}^{+}\,\mathrm{K}^{\mathrm{T}}$
  optionally perform denoising by thresholding W
end function

function $\mathrm{FRM_{WLT}\_eval(W, x)}$ ▷ Reconstruction
  $l^{(\star)} \leftarrow 1 + \lfloor (M-1) \cdot (\mathrm{x} - x_L)/(x_U - x_L) + 1/2 \rfloor$
  $\mathrm{h}^{(\star)} \leftarrow l^{(\star)}$-th column of H^{T}
  $\mathrm{k} \leftarrow \mathrm{W\,h}^{(\star)}$
end function

One-time initialization of wavelet transform matrices:
for all $l \in \{1, \ldots, M\}$ **do** ▷ Compute wavelet matrix
  $\mathrm{b}_l \leftarrow l$-th row of an $M \times M$ identity matrix
  $\mathrm{h}_l \leftarrow \mathrm{wavelet(b}_l\mathrm{)}$ ▷ Inverse discrete wavelet transform
  $s_l \leftarrow 1 + \min(s_U, \lfloor \log_2(M/l) \rfloor)$
end for
$\mathrm{A} \leftarrow \mathrm{diag}_{l=1,\ldots,M}\left\{2^{\delta s_l}\right\}$

Fig. 4. Pseudocode of the wavelet-based interpolation technique $\mathrm{FRM_{WLT}}$.

the Moore-Penrose pseudo-inverse $\mathrm{P}^{+} = \mathrm{P}^{\mathrm{T}}(\mathrm{P\,P}^{\mathrm{T}})^{-1}$ of the matrix $\mathrm{P} = \mathrm{G}^{\mathrm{T}}\mathrm{A}$, since the Moore–Penrose pseudoinverse is the transformation that gives the minimum-norm solution to an underdetermined system, thus obtaining

$$\mathrm{w}_j = \mathrm{A\,P}^{+}\,\mathrm{k}_j, \quad j = 1, \ldots, D \tag{30}$$

where $\mathrm{P} \in \mathbb{R}^{N \times M}$ and $\mathrm{P}^{+} \in \mathbb{R}^{M \times N}$. If the data are noisy, this formulation of $\mathrm{FRM_{WLT}}$ also allows denoising to be easily performed on the coefficients w_j by simple thresholding techniques.

A summary of the steps involved by this algorithm is shown in pseudocode form in Fig. 4.

D. RBF Approximation-Based Identifier—$\mathrm{FRM_{RBF}}$

The fourth identification algorithm, called $\mathrm{FRM_{RBF}}$, is defined by means of an approximating *mapping* based on neural networks. This nonlinear-in-the-parameter identifier is based on radial basis function networks [53], so that the jth component of the functional $\mathcal{G}[\mathbf{x}, \mathrm{W}]$ defined in (10) can be put in the following form:

$$[\mathcal{G}[\mathbf{x}, \mathrm{W}]]_j = \sum_{l=1}^{M_{\mathrm{n}}} [\omega_j]_l \exp\left(-[\chi_j]_l \sum_{m=1}^{M_{\mathrm{x}}} ([\mathbf{x}]_m - [\Xi_j]_{l,m})^2 \right) \tag{31}$$

for $j = 1, \ldots, D$, and where M_{n} is the number of neurons in the RBF network and M_{x} is the dimension of vector $\mathbf{x}$. The parameters $\omega_j \in \mathbb{R}^{M_{\mathrm{n}} \times 1}$, $\chi_j \in \mathbb{R}^{M_{\mathrm{n}} \times 1}$, and $\Xi_j = [\xi_j^1\ \xi_j^2\ \cdots\ \xi_j^{M_{\mathrm{x}}}] \in \mathbb{R}^{M_{\mathrm{n}} \times M_{\mathrm{x}}}$, with $\xi_j^m \in \mathbb{R}^{M_{\mathrm{n}} \times 1}$, $m = 1, \ldots, M_{\mathrm{x}}$, are vectors or matrices of weights. These weights, since the output of a feed-forward RBF network can be viewed as the superposition of many bell-shaped basis functions, actually define the shape and position of the basis functions used to build the network: ω_j contains the heights, ξ_j^m the coordinate of the centers along the mth dimension of the space of the $\mathbf{x}$'s realizations and χ_j controls the width of these functions, their radius at half height being exactly $\sqrt{\ln(2)}/\sqrt{[\chi_j]_l}$ for the lth RBF.

function $\mathrm{FRM_{RBF}\_fit(K, X)}$ ▷ Interpolation
  for all $j \in \{1, \ldots, D\}$ **do**
    $\mathrm{k} \leftarrow j$-th row of K
    $\mathrm{w}_j \leftarrow \mathrm{rbf\_train(X, k)}$
  end for
end function

function $\mathrm{FRM_{RBF}\_eval(W, x)}$ ▷ Reconstruction
  for all $j \in \{1, \ldots, D\}$ **do**
    $[\mathrm{k}]_j \leftarrow \mathrm{rbf\_sim(w}_j\mathrm{,\ x)}$
  end for
end function

Fig. 5. Pseudocode of the RBF-based interpolation technique $\mathrm{FRM_{RBF}}$.

These parameters amount to a total of $M = M_{\mathrm{n}}\,(M_{\mathrm{x}} + 2)$ weights to be estimated, and that can be included in our framework by defining $\mathrm{W} = [\mathrm{w}_1\ \mathrm{w}_2\ \cdots\ \mathrm{w}_D]^{\mathrm{T}} \in \mathbb{R}^{D \times M}$, with $\mathrm{w}_j = [\omega_j^{\mathrm{T}}\ \chi_j^{\mathrm{T}}\ (\xi_j^1)^{\mathrm{T}}\ (\xi_j^2)^{\mathrm{T}}\ \cdots\ (\xi_j^{M_{\mathrm{x}}})^{\mathrm{T}}]^{\mathrm{T}}$. Of course, since RBFs are able to approximate any function arbitrarily well, as in any neural network training endeavour, overfitting must be avoided. To regularize the network output, the number of neurons must then be chosen appropriately, and constraints on their parameters must carefully be applied, possibly with the usage of cross-validation techniques.

Fig. 5 only sketches a pseudocode for this identification algorithm, since any neural network package would certainly possess function approximation primitives, such as "rbf_train" and "rbf_sim", used to train and to simulate a neural network, respectively, and readily adaptable to the context at hand.

Being a nonlinear mapping, learning the proper weight matrix W is usually computationally much more expensive than solving the corresponding linear problems that ensue from the previous approaches, but the neural network-based approximations allow for greater flexibility in the choice of the number of free parameters and scale more gracefully when M_{x} increases, thus posing themselves as an interesting option in many circumstances.

IV. Experimental Results

Nonlinear systems are very common in nature and form a large class of time-variant and time-invariant systems. Many natural phenomena that generate signals, images, and biometric information belong to this class.

Before digging into too much detail, an overview of the kind of systems being considered may be helpful. To this end, let us consider a few quite commonly encountered forced differential equations, such as Duffing's [54] and Narendra's [30], as will be detailed in the following Sections IV-A to IV-C. From them, a number of output realizations were computed. For simplicity, only a single input parameter $\boldsymbol{a}$ was varied, and the initial conditions held constant, so that $\mathbf{x} \equiv \boldsymbol{a}$, and the projections of their outputs on several eigenvectors, as functions of the parameterized input, are reported in Fig. 6, which shows points along the curves $\mathcal{C}_{\mathbf{y}}^j(\mathbf{x})$, for $j = 1, \ldots, 9$. Black dots represent projections

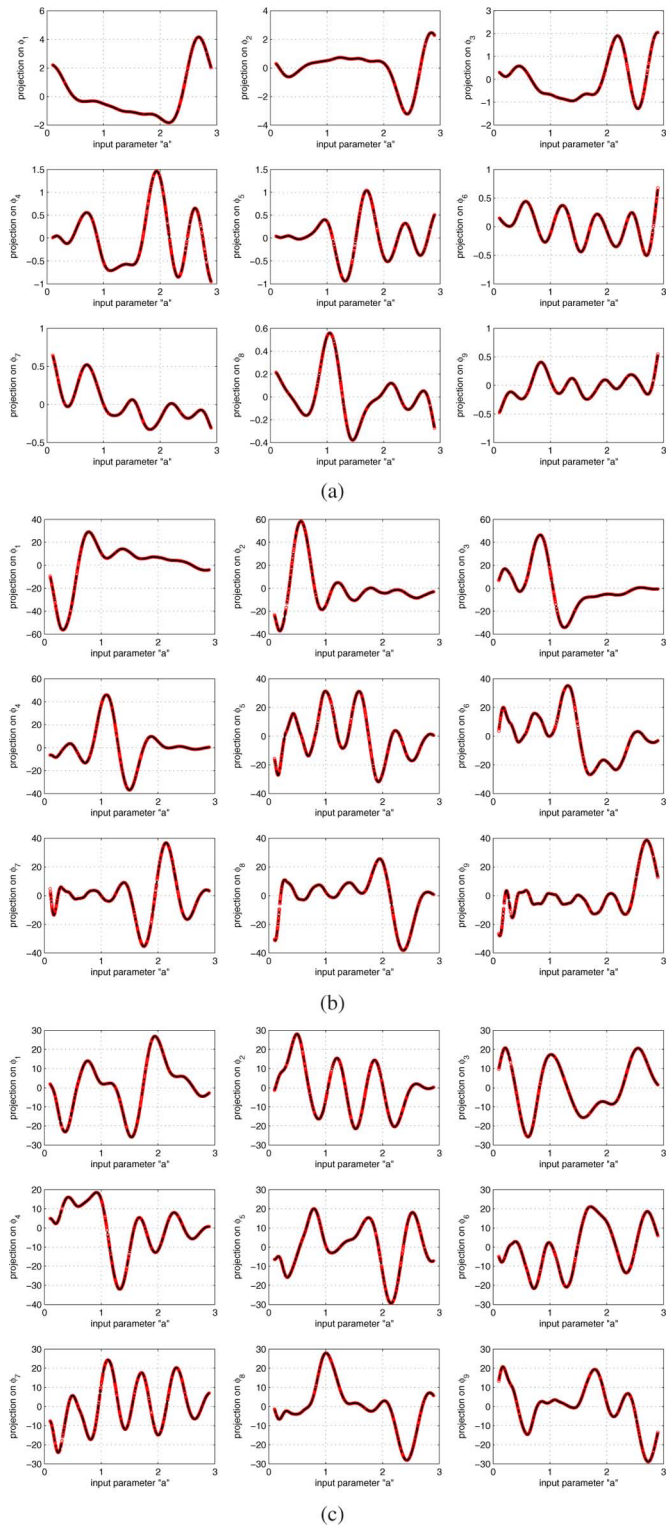

Fig. 6. Projections of the output signals of three forced differential equations onto their first nine eigenvectors $\phi_1(n), \ldots, \phi_9(n)$, plotted versus the parameter $\boldsymbol{a}$ characterizing the input domain. Black dots are the projections from the training set, red circles from the testing set. The results for the (a) synthetic system 1 (Section IV-A), (b) synthetic system 2 (Section IV-B), and (c) synthetic system 3 (Section IV-C), are shown.

from the training set, consisting of 500 realizations, while red circles come from the 2000 realizations of the testing set.

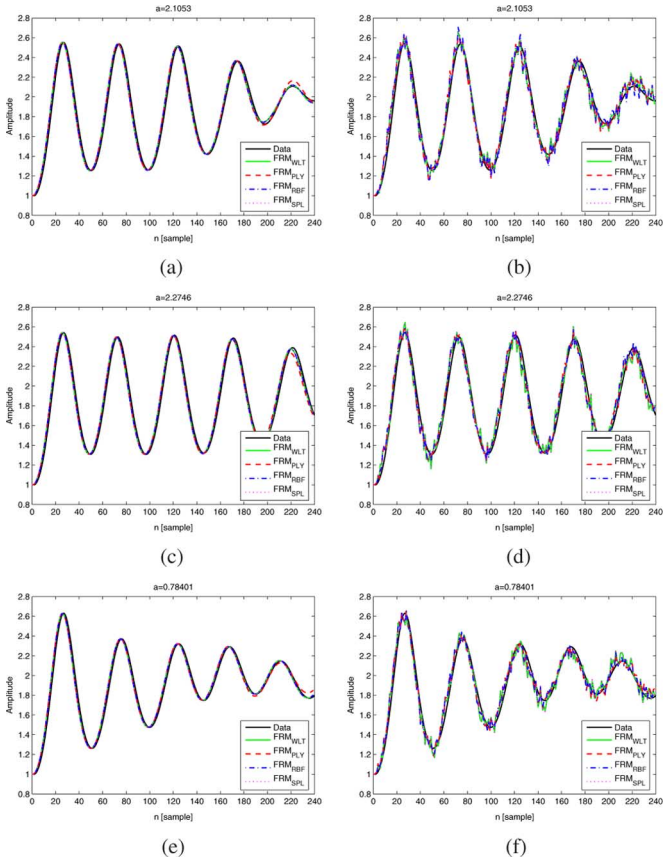

Fig. 7. Identification of the synthetic system 1 ($N = 50$, $L = 240$). (a), (c), and (e) show the identification results for the noiseless case, and (b), (d), and (f) show the results for the corresponding noisy cases.

In the remaining of this section some experimental results on nonlinear system identification, using our approach, are presented. Several examples with *noiseless* and *noisy* output signals are considered and presented in the first three parts into which this section is methodologically divided.

A. Synthetic System 1

The first example of a system to be identified is described by the difference equation:

$$\begin{cases} \boldsymbol{e}(n) = \gamma \cos(\boldsymbol{a}\,\Delta t\, n) \\ \boldsymbol{y}(n+1) = \dfrac{\Delta t^2[\boldsymbol{e}(n) - \beta \boldsymbol{y}^3(n)] - \boldsymbol{y}(n-1) + \delta\, \boldsymbol{y}(n)}{(1+\kappa \Delta t)} \end{cases} \tag{32}$$

with $\delta = \kappa \Delta t - \alpha \Delta t^2 + 2$, where Δt, κ, α, β, γ are constant parameters, and $\boldsymbol{a}$ is a random variable. (32) is the discrete-time version of the well-known Duffing equation with $\Delta t = 0.05$, $\kappa = 0.3$, $\alpha = -4$, $\beta = 1$, $\gamma = 0.5$, $\boldsymbol{a}$ being an RV uniformly distributed in the interval $[0.1, 2.9]$ with mean $E\{\boldsymbol{a}\} = 1.5$, and the initial conditions being $y(0) = y(1) = 1$.

The ability of the proposed framework to represent nonlinear input-output mappings is proved by the results in Fig. 7, which compares the output of the modeled system, identified with the four algorithms mentioned previously, to the output of the original system, both for the noiseless case and with additive white Gaussian noise at a 25 dB SNR. The experiment was performed with a training of $N = 50$ realizations of length

$L = 240$, using three randomly chosen realizations of the RV $\boldsymbol{a}$, namely $a = 2.1053$ [Fig. 7(a) and (b)], $a = 2.2746$ [Fig. 7(c) and (d)], and $a = 0.78401$ [Fig. 7(e) and (f)]. The approximation was limited to the first $D = 10$ eigenvectors, which proved to be enough to get a relative approximation error better than -100 dB. The results can be deemed to be excellent with all the employed function approximation techniques, and the noise rejection capability also proved to be very good. For the $\mathrm{FRM_{WLT}}$ algorithm we used a 6-level decomposition with Daubechies' DB8 wavelets, employing the interpolation technique for nonuniformly sampled data [50] with an adaptive soft-thresholding denoising based on Stein's unbiased risk estimate. The $\mathrm{FRM_{PLY}}$ used polynomial interpolators of order $p = 12$. For the $\mathrm{FRM_{RBF}}$ we used an RBF network with 15 or 40 neurons (for the noisy and noiseless case, respectively), and the $\mathrm{FRM_{SPL}}$ made use of a cubic spline. The results appeared to be excellent for all the four algorithms of the proposed framework.

Finally, it can be noted that the identification error slowly increases towards the end of the fixed interval $[0, L-1]$. This phenomenon does not usually occur in time-domain adaptive filter-based identification techniques, and it is due to the increasing variance of the system output with time, being the initial conditions fixed to the assigned value.

B. Synthetic System 2

The second example is more complex. The system to be identified is governed by difference equation proposed by Narendra *et al.* in [30]:

$$\boldsymbol{y}(n+1) = 0.3\,\boldsymbol{y}(n) + 0.6\,\boldsymbol{y}(n-1) + f[\boldsymbol{e}(n)] \tag{33}$$

where

$$f(\boldsymbol{e}) = 0.6\,\sin(\pi \boldsymbol{e}) + 0.3\,\sin(3\pi \boldsymbol{e}) + 0.1\,\sin(5\pi \boldsymbol{e}) \tag{34}$$

$$\boldsymbol{e}(n) = \sin[(1+\boldsymbol{a})\,\omega_0\, n] \tag{35}$$

with $\omega_0 = 2\pi/250$, and where $\boldsymbol{a}$ is an RV uniformly distributed in the interval $[0.1, 2.9]$ with $E\{\boldsymbol{a}\} = 1.5$, and the initial conditions are $y(0) = y(1) = 1$.

Fig. 8 shows the identification capabilities of the proposed framework for this system, with $N = 50$, $L = 240$, and $D = N$, comparing several trajectories of (33) with those of the approximating model. It reports three cases, selected as in the previous example, showing that the dynamics of the approximating signals are very close to those of the system to be identified, even when the signals are corrupted by noise. The $\mathrm{FRM_{WLT}}$, $\mathrm{FRM_{PLY}}$, $\mathrm{FRM_{RBF}}$, and $\mathrm{FRM_{SPL}}$ algorithms were implemented exactly as per the previous example. Also in this case no perceivable difference in the performance of the four techniques could be observed, as all of them provided an excellent modeling capability.

C. Synthetic System 3

The third example of system to be identified is described by the difference equation also proposed by Narendra *et al.* in [30]:

$$\boldsymbol{y}(n+1) = \frac{\boldsymbol{y}(n)\,\boldsymbol{y}(n-1)[\boldsymbol{y}(n)-2.5]}{1+\boldsymbol{y}^2(n)+\boldsymbol{y}^2(n-1)} + \boldsymbol{e}(n) \tag{36}$$

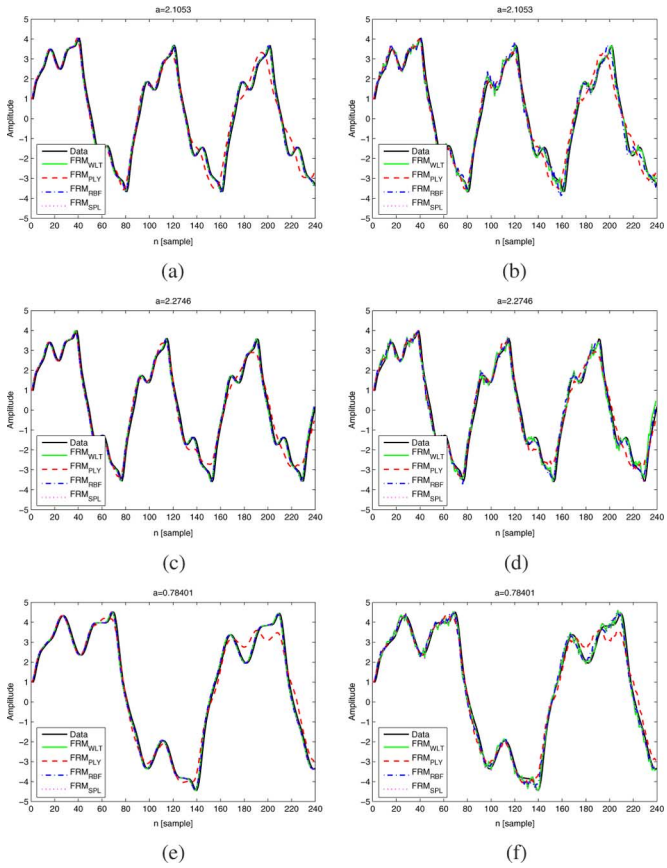

Fig. 8. Identification of the synthetic system 2 ($N = 50$, $L = 240$). (a), (c), and (e) show the identification results for the noiseless case, and (b), (d), and (f) show the results for the corresponding noisy cases.

with

$$\boldsymbol{e}(n) = \sin[(1+\boldsymbol{a})\,\omega_0\, n] \tag{37}$$

where $\omega_0 = 2\pi/250$, and $\boldsymbol{a}$ is an RV uniformly distributed in the interval $[0.1, 2.9]$ with $E\{\boldsymbol{a}\} = 1.5$, and the initial conditions are $y(0) = y(1) = 1$.

Fig. 9 shows the identification results, also in this case with $N = 50$, $L = 240$, $D = N$, and the same algorithms and parameters as those used in Section IV-A.

The approximated dynamics follow the signals generated by (36) quite accurately. The initial transitions, and the micro and macro dynamics of signals are generally captured by this method. No clear advantage of any technique over the others used in the proposed framework can be noted for this example.

V. Comparison with the State of the Art

In order to clearly show the effectiveness of our framework, several comparisons with current best practices were considered. In this analysis, adaptive filtering-based approaches [34], [35], such as the LMS, RLS, and NLMS algorithms, and NARMAX models employing polynomial nonlinearities were taken into account. Methodologically, the performance evaluation was divided into two categories: identification accuracy and processing times.

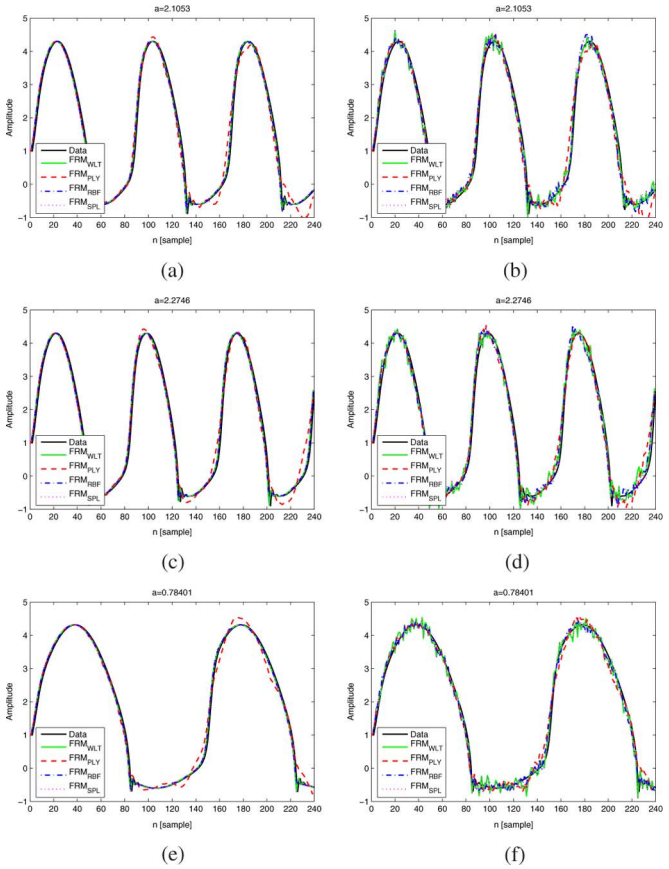

Fig. 9. Identification of the synthetic system 3 ($N = 50$, $L = 240$). (a), (c), and (e) show the identification results for the noiseless case, and (b), (d), and (f) show the results for the corresponding noisy cases.

The identification accuracy of the proposed approach was assessed by comparing its results with those of a number of other techniques.

Table I shows the median of the relative mean square errors, where the relative mean square error of a signal is defined as the ratio between the mean square error (MSE) and the signal power. The results are obtained by different identification techniques, for signal lengths L ranging from 120 to 480 samples, and training set sizes N ranging from 10 to 50 realizations.

The adaptive filters were used to train a model of the following type:

$$\begin{aligned} y(n) = {} & w_0 \\ & + w_y\, f_{p_y}(y(n-1), y(n-2), \ldots, y(n-M_y)) \\ & + w_e\, f_{p_e}(e(n-1), e(n-2), \ldots, e(n-M_e)) \end{aligned} \quad (38)$$

where f_{p_y} and f_{p_e} are polynomials of degree p_y and p_e, respectively, and w_0, w_y, w_e are the filter parameters to be adapted. The predictor orders M_y and M_e used in the LMS, RLS, and NLMS algorithms were set to the real model orders (second order for the output signal, i.e., $M_y = 2$, and first order for the input signal, i.e., $M_e = 1$), and the polynomial coefficients inside of the functions f_{p_y} and f_{p_e} were fixed at those values resulting from the training of the NARMAX model, reported next. Ideally, one might want the adaptive filter to also adapt the coefficients inside of the polynomials, but in our experiments this almost invariably led to instable filters, hence our choice of sticking them to a known good solution. The polynomial order used were also selected to obtain the best possible results, while ensuring the stability of the adapted filter, and were $p_y = 3$ and $p_e = 1$ for the synthetic system 1, $p_y = 1$ and $p_e = 12$ for the synthetic system 2, and $p_y = 4$ and $p_e = 1$ for the synthetic system 3. The step sizes used for adaptation were $\mu = 0.005$ for LMS and $\mu = 0.1$ for NLMS, while the forgetting factor for RLS was $\lambda = 0.9999$ and the parameter δ was set to 0.0001. The filters were adapted by presenting to them $\lfloor 1000/N \rfloor$ epochs of the training set, except for the RLS case, which, having a convergence rate much faster than that of the other two techniques, needed only $\lfloor 100/N \rfloor$ adaptation epochs to reach convergence.

Similar results are reported in Table II, where the median relative MSE at different output signal-to-noise ratios is compared for a fixed value of $L = 240$ and different values of N.

Additionally, the proposed framework was tested against the NARMAX algorithm. According to Billings' original formulation [26, p. 1015], a model of a nonlinear system can be written as

$$\begin{aligned} y(n) = f(&y(n-1), y(n-2), \ldots, y(n-M_y), \\ & e(n-1), e(n-2), \ldots, e(n-M_e)) \end{aligned} \quad (39)$$

where M_y and M_e are the auto-regressive and moving-average model orders, respectively, and f is a nonlinear function to be estimated.

It is worth noting that the proposed framework is based on substantially different principles than NARMAX models, so that a number of assumptions had to be made to compare the results. In fact, our framework is based on a spectral representation and not on a recursive modeling in the time domain; it only uses a parameterized value of the input and not the actual input samples; NARMAX in its original formulation can not be applied to nonstationary systems while this limitation is intrinsically absent in our framework; as model order selection criteria are not needed at all in our approach whilst they play a fundamental role for the identification performance of NARMAX.

Different techniques have been employed to estimate the function f after the development of NARMAX. In this work, we found that good results can be obtained with polynomials of appropriate order. Their coefficient was trained using data arranged in Toeplitz matrices according to the structure mandated by (39).

The NARMAX identifier was applied to all the proposed examples with the model order fixed to the real order of the examples, i.e., $M_y = 2$ and $M_e = 1$, and a polynomial order ranging from 3 to 12 according to the specific example, with and without additional white Gaussian noise with SNRs between 0 dB and 50 dB. The results of the comparison are reported in Tables I and II for the noiseless and noisy cases, respectively.

In all these tests we used our framework complemented with a simple spline-based function approximator ($\mathrm{FRM_{SPL}}$), and despite its simplicity it was able to almost always outperform the other techniques in the moderately noisy conditions. In noiseless conditions, the synthetic system 1, which exhibits a simple polynomial nonlinearity, was unsurprisingly best modeled by the

TABLE I
PERFORMANCE COMPARISON OF VARIOUS IDENTIFIERS APPLIED TO NOISELESS NSI FOR THE THREE SYNTHETIC NONLINEAR SYSTEMS CONSIDERED AS EXAMPLES AND AS A FUNCTION OF L AND N, VERSUS ONE OF THE ALGORITHMS OF THE PROPOSED FRAMEWORK ($\mathrm{FRM_{SPL}}$). VALUES ARE RELATIVE MSES EXPRESSED IN DECIBEL

		$L=120$					$L=240$					$L=360$					$L=480$				
Ex.	N	LMS	NLMS	RLS	NARMAX	FRM	LMS	NLMS	RLS	NARMAX	FRM	LMS	NLMS	RLS	NARMAX	FRM	LMS	NLMS	RLS	NARMAX	FRM
System 1	10	-127	-125	-115	**-205**	-53.9	-124	-122	-122	**-204**	-36.0	-124	-122	-124	**-195**	-30.1	-124	-123	-126	**-196**	-27.5
	20	-127	-124	-116	**-190**	-78.2	-124	-124	-121	**-201**	-49.4	-124	-123	-124	**-192**	-36.7	-124	-124	-126	**-195**	-32.7
	30	-126	-126	-114	**-190**	-92.3	-125	-125	-119	**-197**	-50.6	-126	-127	-123	**-194**	-37.6	-127	-129	-126	**-212**	-33.5
	40	-126	-124	-113	**-207**	-98.5	-124	-122	-118	**-196**	-50.9	-123	-122	-122	**-194**	-38.2	-124	-122	-125	**-202**	-33.8
	50	-127	-127	-115	**-192**	-101	-127	-127	-121	**-192**	-51.1	-128	-128	-124	**-200**	-38.5	-128	-129	-126	**-206**	-33.9
System 2	10	-15.9	-12.7	-13.2	-13.0	**-19.8**	-9.8	-10.5	-12.8	**-17.5**	-14.7	-8.5	-8.2	-12.7	**-18.0**	-13.1	-9.3	-8.8	-12.7	**-17.7**	-12.6
	20	-15.4	1.8	-12.8	-13.5	**-29.5**	-9.7	-10.2	-12.5	-17.9	**-19.7**	-10.0	-9.6	-12.6	**-17.3**	-16.5	-13.2	-10.7	-12.4	-14.3	**-14.9**
	30	-9.2	-11.6	-13.2	-18.1	**-38.0**	-8.6	-11.5	-12.9	-17.0	**-24.5**	-6.3	-6.5	-12.9	-13.1	**-19.5**	-10.2	-9.6	-12.8	-12.9	**-17.1**
	40	-12.9	-9.8	-12.9	-18.5	**-45.6**	-14.6	-9.2	-12.6	-14.3	**-29.4**	-12.9	-9.3	-12.7	-13.0	**-22.9**	-14.1	-7.2	-12.6	-12.6	**-19.6**
	50	-9.5	-11.8	-12.9	-17.5	**-51.4**	-10.8	-10.9	-12.7	-13.1	**-33.2**	-9.0	-10.4	-12.8	-12.7	**-25.9**	-8.0	-7.3	-12.7	-12.5	**-21.8**
System 3	10	-13.3	-14.1	-13.4	-12.1	**-23.6**	-12.3	-13.6	-12.1	-12.9	**-17.4**	-11.8	-14.1	-11.6	-12.7	**-15.2**	-11.6	-14.1	-11.3	-13.9	**-14.3**
	20	-12.9	-9.1	-13.2	-14.0	**-29.9**	-12.4	-12.9	-12.2	-14.3	**-23.0**	-12.0	-13.4	-11.6	-14.4	**-19.3**	-10.7	-7.2	-11.2	-14.1	**-17.3**
	30	-13.3	-12.4	-13.3	-15.5	**-34.9**	-12.5	-13.9	-12.1	-14.9	**-27.0**	-11.8	-10.6	-11.6	-14.7	**-23.0**	-11.0	-6.1	-11.3	-14.4	**-20.3**
	40	-13.5	-13.3	-13.3	-17.3	**-39.6**	-12.1	-13.5	-12.1	-14.9	**-29.7**	-11.3	-5.8	-11.6	-14.5	**-25.7**	-11.6	-13.6	-11.3	-14.2	**-23.0**
	50	-13.4	-15.1	-13.4	-15.8	**-43.4**	-11.8	-6.4	-12.1	-14.8	**-31.7**	-12.0	-12.8	-11.6	-14.6	**-27.5**	-11.5	-11.8	-11.3	-14.3	**-24.8**

TABLE II
PERFORMANCE COMPARISON OF VARIOUS IDENTIFIERS APPLIED TO NOISY NSI FOR THE THREE SYNTHETIC NONLINEAR SYSTEMS CONSIDERED AS EXAMPLES AND AS A FUNCTION OF N WITH $L=240$, VERSUS ONE OF THE ALGORITHMS OF THE PROPOSED FRAMEWORK ($\mathrm{FRM_{SPL}}$). VALUES ARE RELATIVE MSES EXPRESSED IN DECIBELS

		SNR $=12.5$ dB					SNR $=25.0$ dB					SNR $=37.5$ dB					SNR $=50.0$ dB				
Ex.	N	LMS	NLMS	RLS	NARMAX	FRM	LMS	NLMS	RLS	NARMAX	FRM	LMS	NLMS	RLS	NARMAX	FRM	LMS	NLMS	RLS	NARMAX	FRM
System 1	10	-14.7	-13.8	-14.7	**-14.8**	-11.5	-14.6	-14.6	-14.6	-14.7	**-21.3**	-13.6	-13.3	-13.9	-14.4	**-29.4**	-22.3	-18.1	-23.1	-17.0	**-33.7**
	20	-14.7	-14.5	-14.7	**-14.9**	-13.0	-14.6	-14.6	-14.7	-15.2	**-24.2**	-13.6	-13.6	-13.8	-14.9	**-37.0**	-23.4	-21.8	-22.2	-18.1	**-46.1**
	30	-14.5	-13.2	-14.7	**-14.9**	-14.1	-14.3	-12.9	-14.6	-15.0	**-26.2**	-13.6	-12.6	-13.9	-14.8	**-39.2**	-20.7	-16.6	-21.9	-17.6	**-47.8**
	40	-14.7	-13.2	-14.7	**-14.9**	-14.7	-14.6	-14.5	-14.7	-15.2	**-27.0**	-13.6	-13.6	-13.8	-14.9	**-40.2**	-21.9	-18.9	-22.0	-17.9	**-48.1**
	50	-14.2	-11.6	-14.7	-15.0	**-15.3**	-14.3	-11.9	-14.6	-15.2	**-27.9**	-14.0	-12.3	-13.9	-14.9	**-40.9**	-22.3	-22.2	-22.2	-17.7	**-48.4**
System 2	10	**-14.9**	-10.0	-13.6	-4.1	-10.5	**-18.8**	-18.2	-18.5	-15.1	-14.1	-12.8	-15.0	-15.4	**-17.3**	-14.6	-10.0	-10.9	-13.0	**-17.5**	-14.7
	20	-10.5	-8.3	**-13.6**	-3.6	-10.9	-18.5	-7.1	**-18.6**	-15.3	-17.0	-14.0	-10.0	-15.9	-17.9	**-19.0**	-10.1	-10.9	-12.8	-17.9	**-19.6**
	30	-12.0	-10.6	**-13.4**	-3.6	-11.0	-18.1	-10.9	-18.6	-14.4	**-19.0**	-11.8	-11.9	-15.8	-16.9	**-23.0**	-8.8	-11.9	-13.1	-17.0	**-24.2**
	40	-7.5	-6.8	**-13.4**	-3.5	-10.8	-11.1	-13.5	-18.6	-12.5	**-20.0**	-13.8	-12.2	-16.0	-14.3	**-26.1**	-14.6	-9.5	-12.9	-14.3	**-28.6**
	50	-12.7	-12.0	**-13.4**	-3.4	-11.0	-18.8	-14.1	-18.6	-11.6	**-20.7**	-14.7	-12.7	-16.0	-13.1	**-28.4**	-11.1	-11.3	-13.0	-13.2	**-31.8**
System 3	10	**-12.7**	-8.7	-12.5	-11.2	-11.1	-12.5	-14.8	-12.1	-12.5	**-16.1**	-12.3	-13.8	-12.1	-12.8	**-17.1**	-12.3	-13.7	-12.1	-12.9	**-17.3**
	20	-12.4	**-13.3**	-12.4	-12.9	-11.5	-12.6	-13.8	-12.3	-13.8	**-18.9**	-12.4	-13.1	-12.2	-14.2	**-22.0**	-12.4	-13.0	-12.2	-14.3	**-22.8**
	30	-12.5	**-14.0**	-12.5	-13.0	-11.5	-12.7	-15.4	-12.2	-14.6	**-20.4**	-12.6	-14.3	-12.2	-14.9	**-25.4**	-12.6	-14.0	-12.1	-14.9	**-26.7**
	40	-12.1	**-13.9**	-12.5	-12.8	-11.4	-11.9	-13.1	-12.2	-14.6	**-21.1**	-12.0	-13.3	-12.1	-14.9	**-27.1**	-12.1	-13.4	-12.1	-14.9	**-29.1**
	50	-12.4	**-14.7**	-12.5	-12.8	-11.5	-11.9	-7.2	-12.2	-14.5	**-21.4**	-11.8	-6.5	-12.2	-14.8	**-28.5**	-11.8	-6.4	-12.1	-14.8	**-31.0**

NARMAX identifier, which employs a polynomial model that is clearly able to exactly match these test signals. But moving to more complex nonlinearities, as those displayed by synthetic systems 2 and 3, the advantage of the proposed techniques becomes more evident.

The performance of the proposed approach with interpolation techniques such as $\mathrm{FRM_{PLY}}$, $\mathrm{FRM_{WLT}}$ and $\mathrm{FRM_{RBF}}$, other than spline approximation ($\mathrm{FRM_{SPL}}$), are reported in Table III. Extensive simulations were performed, and the performance, evaluated in terms of the relative MSE expressed in decibel, statistically analyzed over a fixed testing set of 2000 realizations and averaged over different choices of training sets. The identification was deemed to be successful when the root MSE was less than a half of the signal root mean square amplitude, that is, the threshold performance used to define the *yield* of the reported techniques was fixed at -6 dB. Q_1, Q_2, and Q_3 are the first three quartiles of the distribution of the relative MSE expressed in decibel, that is, the values at which the cumulative distribution function equals 0.25, 0.50, and 0.75, respectively. IQR, the interquartile range, is Q_3–Q_1, and gives a measure of the variability of the performance around the median (Q_2). Results above the yield threshold have been excluded from the statistics. In order to improve readability, the best results have been highlighted in bold, while the worst are in italics.

From the data, it is apparent that the simplest technique, $\mathrm{FRM_{PLY}}$, also provides the worst results. It only gets the best Q_2 values for very small training sets, but this is accompanied by low yield values. On the contrary, $\mathrm{FRM_{WLT}}$ never wins in terms of median values, but invariably shows the highest yields, as a consequence of the regularization it alone performs, and provides reasonably good Q_2s. $\mathrm{FRM_{SPL}}$ and $\mathrm{FRM_{RBF}}$ have similar performance, and often the best, in terms of Q_2, with $\mathrm{FRM_{SPL}}$ usually providing higher yields.

Comparisons of the processing times, with respect to adaptive filter-based approaches and NARMAX identifiers, are shown in Figs. 10 and 11. The simulation results were obtained with several signal lengths and training set sizes. The results show that the identification times of our approach are significantly lower

TABLE III
PERFORMANCE OF THE PROPOSED FRAMEWORK WITH DIFFERENT APPROXIMATION TECHNIQUES AS A FUNCTION OF N, WITH $L = 240$. VALUES FOR YIELD ARE NORMALIZED TO 1, ALL THE OTHERS ARE RELATIVE MSES EXPRESSED IN DECIBELS. RBF STANDS FOR 40-NEURON $\mathrm{FRM}_{\mathrm{RBF}}$, SPL FOR CUBIC SPLINE $\mathrm{FRM}_{\mathrm{SPL}}$, PLY FOR TWELFTH-ORDER $\mathrm{FRM}_{\mathrm{PLY}}$, AND WLT FOR $\mathrm{FRM}_{\mathrm{WLT}}$ USING DAUBECHIES' DB8 WAVELETS

		PLY					SPL					WLT					RBF				
Ex.	N	Yield	Q_2	Q_1	Q_3	IQR	Yield	Q_2	Q_1	Q_3	IQR	Yield	Q_2	Q_1	Q_3	IQR	Yield	Q_2	Q_1	Q_3	IQR
System 1	10	0.78	**-36.4**	-50.8	-23.4	27.4	0.95	-36.0	-48.8	-26.8	22.0	1.00	*-33.7*	-45.6	-26.0	19.6	0.99	-35.9	-49.3	-25.0	24.3
	20	0.92	*-44.7*	-51.0	-36.1	15.0	0.99	-49.4	-57.1	-39.5	17.7	1.00	-46.5	-53.9	-35.9	18.0	1.00	**-50.3**	-57.8	-40.4	17.4
	30	0.97	*-42.6*	-46.4	-38.0	8.4	1.00	-50.6	-55.9	-45.4	10.5	1.00	-48.8	-54.0	-43.4	10.7	1.00	**-51.0**	-56.1	-45.8	10.3
	40	0.98	*-41.2*	-44.8	-37.8	7.0	1.00	-50.9	-55.6	-46.7	8.9	1.00	-49.4	-54.1	-45.1	9.0	1.00	**-51.2**	-55.7	-46.9	8.8
	50	0.99	*-40.7*	-43.8	-37.6	6.2	1.00	-51.1	-55.2	-46.9	8.2	1.00	-49.9	-54.0	-45.6	8.3	1.00	**-51.2**	-55.3	-47.1	8.2
System 2	10	0.35	**-15.7**	-24.0	-10.7	13.3	0.63	-14.7	-22.3	-9.9	12.4	0.67	*-14.2*	-21.5	-9.6	11.9	0.50	-15.4	-23.3	-10.5	12.8
	20	0.70	*-16.1*	-20.1	-12.2	8.0	0.86	-19.7	-28.9	-13.7	15.2	0.90	-18.3	-26.9	-12.3	14.6	0.46	**-20.4**	-30.5	-13.1	17.4
	30	0.88	*-15.0*	-17.9	-12.2	5.7	0.94	**-24.5**	-34.6	-17.0	17.6	0.97	-22.5	-31.0	-15.3	15.7	0.45	-24.1	-35.4	-15.0	20.4
	40	0.92	*-15.0*	-17.3	-12.6	4.7	0.96	**-29.4**	-40.0	-20.4	19.6	0.99	-26.2	-33.8	-18.0	15.8	0.61	-28.6	-39.7	-18.3	21.4
	50	0.95	*-14.8*	-17.0	-12.7	4.3	0.98	**-33.2**	-44.9	-23.9	21.0	0.99	-29.0	-35.6	-21.0	14.6	0.86	-29.9	-37.6	-21.4	16.2
System 3	10	0.44	**-17.9**	-25.3	-12.1	13.2	0.80	-17.4	-24.1	-12.1	11.9	0.83	*-16.8*	-23.6	-11.6	12.0	0.61	-17.7	-24.9	-12.3	12.6
	20	0.80	*-20.9*	-24.9	-16.1	8.8	0.93	**-23.0**	-29.9	-17.3	12.6	0.97	-21.9	-28.9	-16.1	12.8	0.47	-21.7	-30.3	-14.1	16.1
	30	0.93	*-20.7*	-23.5	-17.8	5.6	0.96	**-27.0**	-33.9	-21.4	12.5	0.99	-25.8	-32.3	-19.9	12.4	0.40	-22.8	-32.6	-14.4	18.2
	40	0.95	*-20.8*	-23.0	-18.5	4.5	0.97	**-29.7**	-37.0	-24.1	12.9	1.00	-28.3	-34.7	-22.5	12.2	0.58	-26.4	-35.2	-17.4	17.7
	50	0.97	*-20.7*	-22.8	-18.6	4.3	0.98	**-31.7**	-39.9	-26.3	13.6	1.00	-30.1	-36.4	-24.8	11.6	0.86	-29.4	-36.1	-22.2	13.9

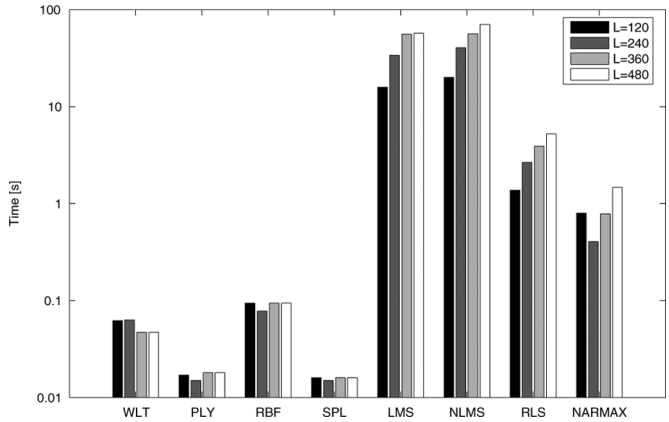

Fig. 10. Elaboration times of the proposed identification framework, along with the LMS, RLS, NLMS, and NARMAX algorithms as functions of the signal length L, for a fixed $N = 10$.

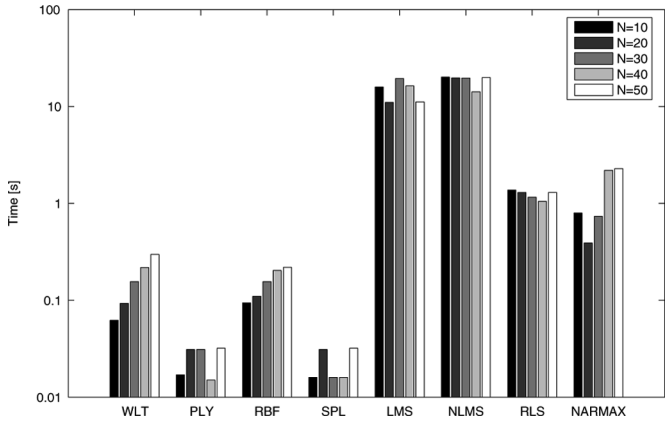

Fig. 11. Elaboration times of the proposed identification framework, along with the LMS, RLS, NLMS, and NARMAX algorithms as functions of the training set size N, for a fixed $L = 240$.

than that of the other algorithms. Indeed, our approach has a computation time that is almost proportional to N (or D, the two being closely related), since it employs block algorithms that operate on the signals as a whole and not on a per-sample basis. The algorithms used for the adaptive filters need to perform a number of iterations proportional to the number of samples in the training set, so have computation times proportional to LN. For the adaptive filters, the number of epochs used to train them also plays a fundamental role in determining their computation time. The amount of training actually performed was chosen so as to ensure the convergence of the adapted filters in most of our tests. Ensuring convergence in all the tests would have implied unreasonably long computational times for LMS, NLMS, and RLS. Also the NARMAX identifier, despite being based on somewhat different premises, showed computation times roughly proportional to LN, with LN being the amount of sample points available for its training.

VI. APPLICATION TO A PHYSICAL SYSTEM

In this section, an application of the proposed technique for the identification of a real nonlinear system is presented. As an example we consider a voice communication channel over a telephone system, since microphone distortions and overall system nonlinearities are known to be a major cause of performance degradation in tasks such as speech or speaker recognition [55], [56]. As reported in these studies, the availability of an accurate nonlinear model of the system would help the development of appropriate compensation techniques.

The system studied is a voice communication channel comprising a DECT cordless phone, a telephone line, and the interface circuitry needed to operate the system, as depicted in Fig. 12.

The test signals used are made up of a sequence of fixed frequency sine waves, each 1 s long and with a random amplitude, for an overall duration of 1 h. In order to probe the system under different conditions within the telephone passband three different frequencies of 500 Hz, 1 kHz, and 2 kHz were used.

The excitation signals were generated by a computer system, used to control the experiment, and reproduced by a high-quality loudspeaker with the gain adjusted so that the mean output amplitude matched the loudness of the typical human voice. A cordless handset was placed in front of the loudspeaker, in a

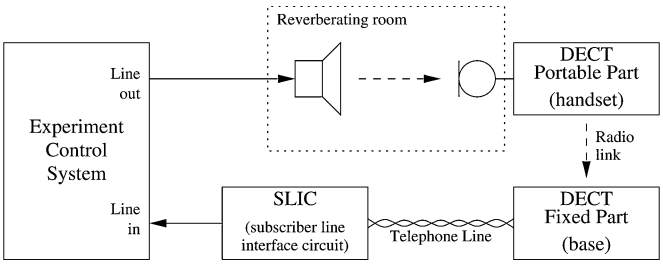

Fig. 12. Sketch of setup used for the telephone system identification experiment.

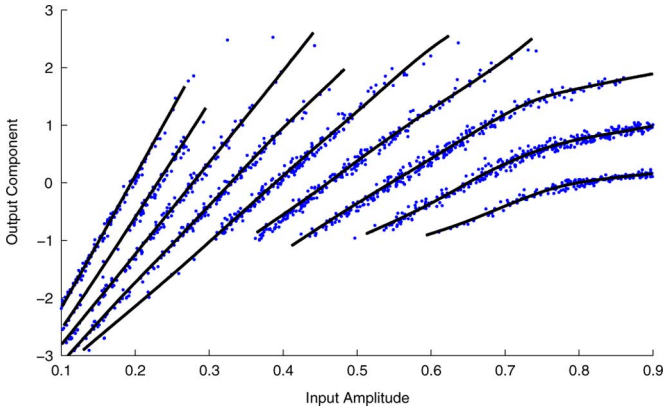

Fig. 13. Projections of the 1 kHz telephone test set over the first eigenvector. The solid lines are the fitted model.

reverberating room with a level of background noise typical of working environments, so as to match as closely as possible the typical operating conditions of the mobile handset.

The output of the DECT phone was then conveyed back to the computer by means of a subscriber-line interface circuit (SLIC), used to generate the power and signaling levels on the telephone line and to separate the audio to and from the telephone line, and was then digitized with a sampling rate of 48 kHz with a 16-bit per sample resolution.

A 10 ms window for each recorded wave in the sequence was extracted after about 950 ms from the beginning of the wave itself, so that transient effects, room reverberation, and automatic gain control (AGC) in the phone could be considered stable. The population considered was thus composed of 3600 signals of 480 samples each, and was divided into a training set and a test set of equal sizes.

The results of the experiment for a frequency of 1 kHz are reported in Figs. 13–15. The variables used for the parameterization are the wave amplitude, uniformly distributed in the normalized range $[0.1, 0.9]$, and the AGC gain level. In fact, as can be seen from Fig. 13, which shows the projection of the test set over the first eigenvector, the operation of the telephone AGC is immediately apparent, with nine easily distinguishable discrete gain levels. Although the nonlinearity shown by each curve in Fig. 13 is not very pronounced, the effectiveness of the identification method in this case is mostly related to its capability in modeling the amplitude quantization effect, as clearly appears by projecting the output realizations over the eigenvectors.

Since the eigenvalue spectrum rapidly converges to the noise plateau, as shown in Fig. 14, only the projections onto the first

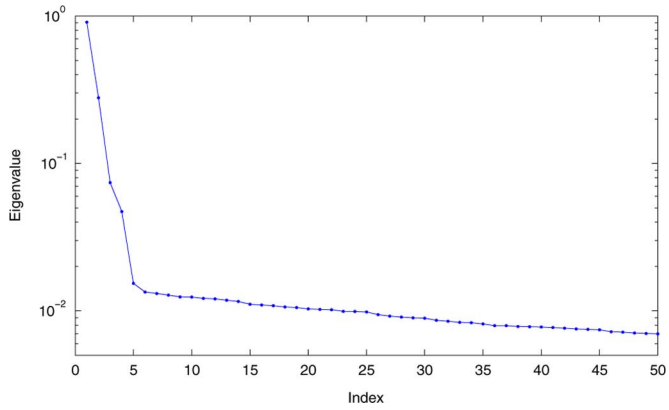

Fig. 14. Eigenvalue spectrum for the 1 kHz telephone experiment.

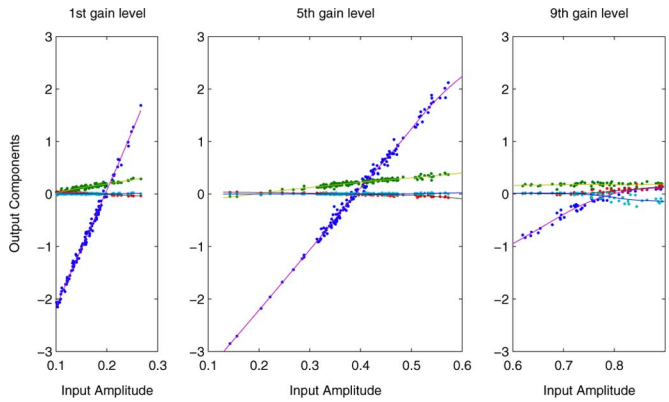

Fig. 15. Projections of the 1 kHz telephone test set over the first four eigenvectors. The solid lines are the polynomial model fitted on the corresponding training set. From left to right results for three different values of the AGC gain are shown.

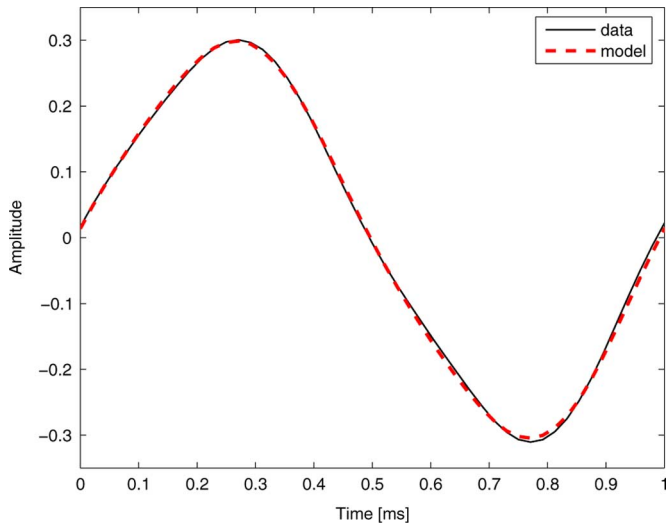

Fig. 16. Portion of a measured signal and its corresponding synthesized version.

four eigenvectors were considered for the modeling, i.e., D was set to 4.

A separate model for each gain level was thus extracted, leading to a family of curves like those shown in Fig. 15 where the projection of the test set onto the first four eigenvectors, as

well as the polynomial model extracted from the training set, are plotted for a few gain levels of the telephone AGC.

The overall modeling result is shown in Fig. 16, which displays a portion of a recorded signal together with the signal rebuilt using the extracted model. The relative MSE averaged over the whole test set was only -31.7 dB.

VII. Conclusion

This paper presents an innovative identification framework that allows designing a number of distinct identifiers for nonlinear systems, which generate geometrical relationships in the Hilbert space. The identifiers resulting from this framework are based on approximating mappings of the complete set of the *trajectories* with a limited *ensemble* of realizations. Exhaustive experimentation shows the effectiveness of the proposed technique with both noise and noiseless output signals. Several comparisons with the state of the art show how this framework almost always has a higher identification accuracy than the current best practices and excellent elaboration times. Finally, it is worth noting that, unlike other techniques, an *a priori* knowledge (or the use of estimation criteria) for the system order and/or its mathematical properties is not needed thus making the proposed technique one of the current best practices in this field.

References

[1] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Process.*, vol. 81, pp. 533–580, Mar. 2001.

[2] J. D. Victor, "Analyzing receptive fields, classification images and functional images: Challenges with opportunities for synergy," *Nature Neurosci.*, vol. 8, pp. 1651–1656, Dec. 2005.

[3] P. Z. Marmarelis and K. Naka, "White-noise analysis of a neuron chain: An application of the Wiener theory," *Science*, vol. 175, pp. 1276–1278, Mar. 1972.

[4] P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems*. New York: Plenum, 1978.

[5] Y. W. Lee and M. Schetzen, "Measurement of the Wiener kernels of a non-linear system by crosscorrelation," *Int. J. Control*, vol. 2, no. 3, pp. 237–254, Feb. 1965.

[6] M. Inagaki and E. Mochizuki, "Bilinear system identification by Volterra kernels estimation," *IEEE Trans. Autom. Control*, vol. 29, pp. 746–749, Aug. 1984.

[7] Y. Lee, *Statistical Theory of Communication*. New York: Wiley, 1960.

[8] J. Stapleton and S. Bass, "Adaptive noise cancellation for a class of nonlinear, dynamic reference channels," *IEEE Trans. Signal Process.*, vol. SP-32, pp. 143–150, Feb. 1985.

[9] S. Pupolin and L. Greenstein, "Performance analysis of digital radio links with nonlinear transmit amplifier," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, pp. 534–546, Apr. 1987.

[10] E. Biglieri, S. Barberis, and M. Catena, "Analysis and compensation of nonlinearities in digital transmission systems," *IEEE J. Sel. Areas Commun.*, vol. 6, pp. 42–51, Jan. 1988.

[11] D. Falconer, "Adaptive equalization of channel nonlinearities in QAM data transmission systems," *Bell Syst. Tech. J.*, vol. 57, pp. 2589–2611, Sep. 1978.

[12] O. Agazzi, D. Messerschmitt, and D. Hodges, "Nonlinear echo cancellation of data signals," *IEEE Trans. Commun.*, vol. COM-30, pp. 2421–2433, Nov. 1982.

[13] A. Fermo, A. Carini, and G. Sicuranza, "Simplified Volterra filters for acoustic echo cancellation in GSM receivers," in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*, Sep. 2000, pp. 1443–1463.

[14] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

[15] G. Ramponi, "Edge extraction by a class of second-order nonlinear filters," *Electron. Lett.*, vol. 22, pp. 482–484, Apr. 1986.

[16] G. Ramponi and G. Sicuranza, "Quadratic digital filters for image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 937–939, Jun. 1988.

[17] G. Sicuranza and G. Ramponi, "Adaptive nonlinear prediction of TV image sequences," *Electron. Lett.*, vol. 25, pp. 526–527, Apr. 1989.

[18] S. Thurnhofer and S. K. Mitra, "A general framework for quadratic Volterra filters for edge enhancement," *IEEE Trans. Signal Process.*, vol. 5, pp. 950–963, Jun. 1996.

[19] A. Carini and G. L. Sicuranza, "Filtered-X affine projection algorithm for active noise control using second-order Volterra filters," *EURASIP J. Appl. Signal Process.*, vol. 12, pp. 1841–1848, Sep. 2004.

[20] M. Bouchard, "Multichannel affine and fast affine projection algorithms for active noise control and acoustic equalization systems," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 1, pp. 54–60, Jan. 2003.

[21] G. L. Sicuranza and A. Carini, "Filtered-X affine projection algorithm for multichannel active noise control using second-order Volterra filter," *IEEE Signal Process. Lett.*, vol. 11, pp. 853–857, Nov. 2004.

[22] A. J. M. Kaizer, "Modeling of the nonlinear response of an electrodynamic loudspeaker by a Volterra series expansion," *J. Audio Eng. Soc.*, vol. 35, pp. 421–433, Jun. 1987.

[23] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proc. IEEE*, vol. 69, no. 12, pp. 1557–1573, Dec. 1981.

[24] A. Carini and G. L. Sicuranza, "Transient and steady-state analysis of filtered-X affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 665–678, Feb. 2006.

[25] A. Carini and G. L. Sicuranza, "Optimal regularization parameter of the multichannel filtered-X affine projection algorithm," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4482–4895, Oct. 2007.

[26] S. Chen and S. Billings, "Representation of non-linear systems: The NARMAX model," *Int. J. Control*, vol. 49, no. 3, pp. 1013–1032, 1989.

[27] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[28] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems. Part I: Deterministic non-linear systems," *Int. J. Control*, vol. 41, no. 2, pp. 303–328, 1985.

[29] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems. Part II: Stochastic non-linear systems," *Int. J. Control*, vol. 41, no. 2, pp. 328–344, 1985.

[30] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Jan. 1990.

[31] S. Lu and T. Basar, "Robust nonlinear system identification using neural-network models," *IEEE Trans. Neural Netw.*, vol. 9, no. 3, pp. 407–429, Mar. 1998.

[32] L. Chen and K. S. Narendra, "Identification and control of a nonlinear discrete-time system based on its linearization: A unified framework," *IEEE Trans. Neural Netw.*, vol. 15, no. 3, pp. 663–673, May 2004.

[33] T. Poggio, R. Rifkin, S. Mukherjee, and P. Niyogi, "General conditions for predictivity in learning theory," *Nature*, vol. 428, pp. 419–422, Dec. 2004.

[34] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[35] T. Ogunfunmi, *Adaptive Nonlinear System Identification—The Volterra and Wiener Model Approaches*. New York: Springer, 2007.

[36] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[37] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.

[38] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *J. Roy. Statist. Soc. Series B*, vol. 41, no. 2, pp. 190–195, 1979.

[39] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 10, pp. 2319–2323, Dec. 2000.

[40] C. Tomasi, "Past performance and future results," *Nature*, vol. 428, p. 378, Dec. 2004.
[41] C. Turchetti and F. Gianfelici, "A novel identification technique for nonlinear systems with high computational performance," Ufficio Brevetti—Ministero Attività Produttive Italian Patent (in Internationalization Phase) Progr. 6193, Cod. Reg. RIC1, N. Vend. 2006/7988, Nov. 2006.
[42] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 4, pp. 744–754, Aug. 1986.
[43] E. R. Dougherty, *Random Processes for Image and Signal Processing*, ser. SPIE-IEEE Series on Imaging Science and Engineering. Bellingham, WA: SPIE, 1998.
[44] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, Sep. 1990.
[45] G. Cybenko, "Approximation by superpositions of sigmoidal function," *Math. Control, Syst., Signal*, vol. 2, pp. 303–314, Feb. 1989.
[46] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Netw.*, vol. 2, no. 3, pp. 183–192, 1989.
[47] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 3, pp. 395–303, 1989.
[48] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Comput.*, vol. 2, no. 3, pp. 246–257, 1991.
[49] C. de Boor, *A Practical Guide to Splines*, ser. Applied Mathematical Sciences, rev. ed no. 27 ed. New York: Springer, 2001.
[50] H. Choi and R. Baraniuk, "Interpolation and denoising of nonuniformly sampled data using wavelet-domain processing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Mar. 1999, vol. 3, pp. 1645–1648.
[51] R. A. Adams, *Sobolev Spaces*, 1st ed. New York: Academic, 1975.
[52] J. Dubinskij, *Sobolev Spaces of Infinite Order and Differential Equations*, 2nd ed. Dordrecht, Germany: D. Reidel, 1986.
[53] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
[54] A. Nayfeh and D. Mook, *Nonlinear Oscillations*, 1st ed. New York: Wiley, 1990.
[55] T. F. Quatieri, D. A. Reynolds, and G. C. O'Leary, "Estimation of handset nonlinearity with application to speaker recognition," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 567–584, Sep. 2000.
[56] D. A. Reynolds, M. A. Zissman, T. F. Quatieri, G. C. O'Leary, and B. A. Carlson, "The effects of telephone transmission degradations on speaker recognition performance," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 1995, vol. 1, pp. 329–332.

Claudio Turchetti (M'86) received the Electronics Engineering degree from the University of Ancona, Italy, in 1979.

He joined the Università Politecnica delle Marche, Ancona, Italy, in 1980, where he was Director of the Department of Electronics, Artificial Intelligence and Telecommunications (DEIT) and is currently Full Professor of Electronics at the Department of Biomedical, Electronics, Telecommunication Engineering. He has published more than 140 journal and conference papers and two books. His research interests are in computational intelligence, algorithms for signal processing, pattern recognition, system identification, algorithms for simulation of VLSI integrated circuits, statistical analysis of submicron integrated circuits, analog and digital CMOS ICs for artificial neural networks and fuzzy circuits, and RF integrated circuits.

Prof. Turchetti has served as a program committee member of several conferences and as a reviewer of several scientific journals. He has been project leader of several international and national research projects.

Giorgio Biagetti (S'03–M'05) received the Laurea degree (*summa cum laude*) in electronics engineering from the Università degli Studi di Ancona, Italy, in 2000 and the Ph.D. degree in electronics and telecommunications engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2004.

In 2004, he joined, as a Research Assistant, the Department of Electronics, Artificial Intelligence, and Telecommunications (DEIT), now the Department of Biomedical Engineering, Electronics, and Telecommunications (DIBET), at the Università Politecnica delle Marche, Ancona, Italy. His research interests include statistical and high-level simulation of analog and mixed-mode integrated circuits, statistical modeling, coding, synthesis, and automatic recognition of speech, together with wireless systems and networks for data gathering and processing.

Dr. Biagetti has served as a reviewer for a number of international journals and conferences in the fields of electronics and signal processing and has coauthored dozens of scientific papers in international journals, edited books, and conference proceedings.

Francesco Gianfelici was born in 1979. He received the Laurea degree in electronics engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2003 and the Ph.D. degree in electronics, informatics, and telecommunications engineering from the Department of Electronics, Artificial Intelligence, and Telecommunications (DEIT), the Università Politecnica delle Marche, in 2006.

He is currently a Research Assistant at the Department of Biomedical Engineering, Electronics, and Telecommunications (DIBET), Università Politecnica delle Marche, Ancona, Italy. His current research interests include signal processing theory and methods, statistical signal processing, signal processing for wireless communications, machine learning for signal processing, neural networks for signal processing, audio and speech processing, and information theory.

Dr. Gianfelici was a member of many International Technical Committees and several international boards. He served as a reviewer for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE SIGNAL PROCESSING LETTERS, the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING, *Signal Processing* (Elsevier), the IEEE Conference on Audio, Speech, and Signal Processing (ICASSP), the IEEE Workshop on Statistical Signal Processing (SSP), the IEEE International Symposium on Circuits and Systems (ISCAS), the IEEE International Conference of Signal Processing and Communications, the IEEE World Congress on Computational Intelligence, and the European Signal Processing Conference (EUSIPCO).

Paolo Crippa (M'02) received the Laurea degree in electronics engineering (*summa cum laude*) from the Università degli Studi di Ancona, Italy, in 1994 and the Ph.D. degree in electronics engineering from the Polytechnic of Bari, Italy, in 1999.

From 1994 to 1999, he was a Research Fellow at the Department of Electronics and Automation, Università degli Studi di Ancona, where in 1999 he was appointed Research Assistant as a Member of the Technical Staff. In 2006, he joined the Department of Electronics, Artificial Intelligence, and Telecommunications (DEIT), now the Department of Biomedical Engineering, Electronics, and Telecommunications (DIBET), at the Università Politecnica delle Marche, Ancona, Italy, as an Assistant Professor. His current research interests include statistical integrated circuit design, statistical device modeling and simulation, mixed-signal and RF circuit design, neural networks, stochastic processes, and areas of signal processing involving coding, synthesis, and automatic recognition of speech.

Dr. Crippa serves as a reviewer for a number of international journals and conferences in the fields of electronics, signal processing, and computational intelligence. He has published more than 50 papers in international journals, edited books, and conference proceedings. He is an Associate Editor of the *International Journal of Computational Intelligence Studies*. He is a Member of the Federazione Italiana di Elettrotecnica, Elettronica, Automazione, Informatica e Telecomunicazioni (AEIT).