INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2025, VOL. 71, NO. 2, PP. 361–368
Manuscript received January 31, 2025; revised May, 2025. doi: 10.24425/ijet.2025.153581

Dimensionality reduction in kernel-based identification of Wiener system by cyclostationary excitations

Gabriel Maik, and Grzegorz Mzyk

***Abstract*—The topic of nonparametric estimation of nonlinear characteristics in the Wiener system is examined. In this regard, the traditional kernel algorithm faces difficulties stemming from the dimensionality associated with the memory length of the dynamic block. A particular class of input sequences has been proposed, which aids in reducing dimensionality and consequently improves the convergence rate of the estimator to the true characteristics. A theoretical analysis of the suggested method is presented.**

***Keywords*—Nonlinear systems; Identification; Kernel estimation; Wiener structure; Curse of dimensionality; Cyclostationary signals; Nonparametric regression**

I. Introduction

DESCRIBING reality using mathematical models is a common practice and is of fundamental importance in forecasting, making optimal decisions, automatic control, simulation, fault detection, noise reduction, etc. Even when trying to describe seemingly simple processes, we do it ineptly, because reality reveals various nuances, such as dynamic memory, strong nonlinear dependencies, nonstationarity (variability of process behavior over time), or uncertainty connected with a random nature of the observed processes. Taking into account all these factors leads to very complicated descriptions, which in turn increases the computation time. The concept of the so-called predictive control, i.e. one in which a forecast of the behavior of a given phenomenon is determined on an ongoing basis, and then the optimal control/decision is obtained based on it. Of course, for such a controller to be able to work online [1], an appropriate speed of computation is required to be ready with a decision on time. In automation systems, times are often measured in milliseconds. It is therefore necessary to simplify the actual process and approximate it using a model from a specified a priori class. Since the 1980s, models with block-oriented structures, consisting of not only linear dynamic (L) but also nonlinear static (N) blocks, have gained common acceptance. Among them, the Wiener-type structure (L-N) is considered one of the most promising due to its good approximation properties of many real phenomena. This is confirmed by numerous publications in automatic control, telecommunications, and econometrics (see e.g. [2] and references cited therein). This work investigates an approach to estimate nonparametrically the nonlinearity in the Wiener system based solely on the external signals' measurements in the presence of random disturbances. Importantly, the proposed method fits into the so-called nonparametric identification [3]–[5], where a wide class of nonlinear characteristics is allowed, not necessarily described by a finite number of parameters. Compared to the Hammerstein (N-L) system, the Wiener system has had a much larger number of spectacular practical applications. They cover not only automatic control problems, but also issues related to telecommunications, biocybernetics, econometrics, chemistry, mechanics, geology, etc. [6], [7]. Many years of experience of engineers indicate that the L-N structure better reflects most real phenomena. Unfortunately, the problem of identifying the Wiener system is extremely difficult from a theoretical point of view. In particular, when the input excitation is treated as a random process, it is transferred through a linear dynamic filter, and therefore the interactive internal signal (not available for direct measurement) is correlated. This entails malicious problems in constructing estimators of the impulse response and nonlinear characteristics and analyzing their asymptotic consistency. Modern approaches emphasize the most general description of individual model blocks. A linear dynamic subsystem is usually described in terms of its impulse response, while a nonlinear static subsystem is treated nonparametrically. This leads to the formulation of a problem that is essentially an optimization problem in a multidimensional space, and the number of dimensions is equal to the number of considered impulse response elements. Therefore, in light of the preceding discussion, increasing the complexity of the model causes an exponential increase in its variance. This is due to the fact that in a multidimensional space the likelihood of getting measurements close to the operating point declines rapidly. This phenomenon, called the 'curse of dimensionality', means that the identification algorithms with nonparametric nonlinearity developed so far have a very limited range of applicability. Applying them to systems with long memory is rather out of the question. Reducing the dimensionality is one of the crucial challenges in the field of optimization, which has been intensively analyzed in the past decades.

Solutions range from ones intended for linear modeling [8] to nonlinear manifold learning [9]–[12], where for the latter, such methods as Isomap and locally linear embedding are often considered. Dimensionality reduction is also significant

G. Maik and G. Mzyk are with Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Wrocław, Poland (e-mail: {gabriel.maik, grzegorz.mzyk}@pwr.edu.pl).

in dynamic systems identification [13], including nonlinear systems [14], [15] similar to this paper.

The main intention of this work is therefore to formulate classes of specific input signals, which are postulated to guarantee dimensionality reduction and have practical justification.

The article addresses the prevalent issue of input dimensionality in the modeling of data. In the contexts of system identification [16] and pattern recognition [17], the objective is to minimize either the model's order or the number of features taken into account. Specifically, when employing a nonparametric approach to model nonlinear phenomena, the likelihood of encountering input sequences that are proximate to the reference point of interest diminishes noticeably as input dimensionality rises. For instance, in the kernel method, this leads to a limited selection of cases, which in turn causes a significant increase in the estimates' variances. On the contrary, in methodologies that utilize the nearest neighbors technique, an increase in the number of features tends to amplify the bias of the estimators.

Significant efforts have been dedicated to address this issue. Notable methodologies in this area include the Akaike information criterion [18] and L1 lasso regularization [19]. The application of small parametric models facilitates variance reduction even with a limited dataset, however, this approach is inherently linked to a systematic non-zero approximation error (bias) in the asymptotic sense. To eliminate this asymptotical approximation error, one can enhance the model's complexity as long as the number of observations increases. This strategy is referred to as nonparametric [4], [5]. Techniques such as kernel algorithms and orthogonal series expansion algorithms exemplify such an approach. Nevertheless, the challenge is the curse of dimensionality, which often results in a slow rate of increase in model order (or kernel narrowing) necessary to ensure asymptotic consistency. In the context of Wiener systems, numerous strategies have been proposed in the literature to tackle this obstacle, primarily focusing on the generation of specialized input sequences, such as piecewise constants periodic inputs [20], multisinusoids [21], or cyclostationary processes [6]. Generally, it is assumed that the input is contained within a manifold characterized by a reduced internal dimension [12], [22], [23].

The primary contribution of this work lies in the usage of a supplemental ARMA (autoregressive moving average) filter to construct special sequences of the system's excitation. The identification of the Wiener system is performed under conditions, where the autoregressive component functions in a free state, meaning that the generated input is solely derived from the filter's initial state. In this context, we demonstrate the feasibility of constructing a kernel estimator for nonlinear characteristics that operates within a lower-dimensional space, which is determined by the autoregression's order of the employed filter. While a similar approach is discussed in references [24] and [25], this study extends the input signal's class to include any order ARMA time series.

The structure of the paper is outlined thusly. In Section II, the problem is articulated comprehensively, including the foundational assumptions regarding the input signal and the system being identified. Subsequently, Section III elaborates on the proposed algorithm and presents a theorem explaining to which extent the dimension of the problem is reduced. Section IV shows possible ways to aggregate the rescaled estimators. The properties of the proposed identification algorithm are illustrated with the help of the numerical experiment in Section V. A concise discourse along with final observations is presented in Section VI. Finally, the formulated theorem is proven formally within the Appendix in Section VII.

II. Statement of the Problem

This work addresses the task of identifying nonparametrically the static nonlinearity of the dynamic-nonlinear system with the Wiener structure. The entire procedure is performed based on solely N-samples long sequences of the input, $\{u_k\}$, and the output, $\{y_k\}$, while N represents the data set's size.

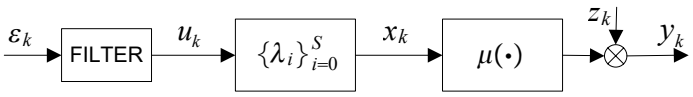

Fig. 1. Wiener system with input signal obtained through filtering

.

[System] The dynamic component in the Wiener system is represented by a finite impulse response filter, characterized by the coefficients $\lambda = [\lambda_0, \lambda_1, ..., \lambda_S]^T$. The length of the memory, denoted as S, is assumed to be beforehand known. In turn, the static nonlinearity is defined by a function $\mu(\cdot)$, that is $\mathcal{L}$-Lipschitz continuous. In summary, the unavailable for measurement internal signal $\{x_k\}$, and the noiseless output $\{v_k\}$ can be described in the following manner

$$x_k = \sum_{j=0}^{S} \lambda_j u_{k-j}, \qquad v_k = \mu(x_k). \tag{1}$$

[Input signal] The sequence $\{u_k\}$ is generated based on unknown process $\{\varepsilon_k\}$ filtered by stable[1] ARMA(r, s) object as shown in Fig. 1. An order of the autoregression should not exceed the memory length of the entire system, such that $r \leq S$. Otherwise, the system could still be identified, but no longer with reduced dimensionality.

The signal $\{u_k\}$ contains $\mathcal{P}$ segments, where each of them consist of T_p samples, while $0 \leq p \leq \mathcal{P}-1$. Each interval is initiated by $T_p^{(\text{on})}$ samples of stochastic part of the $\{\varepsilon_k\}$ process, where the remaining $T_p^{(\text{off})}$-samples long subsequence within the period is characterized by zeroed ε_k. Stochastic part of $\{\varepsilon_k\}$ originating from the entire input sequence constitutes a finite idependent identically distributed (i.i.d.) series of random variables, formally defined as $\forall_k \mathrm{P}(|\varepsilon_k| \leq \varepsilon_{\max} < \infty) = 1$. Additionally, $\forall_p (r \leq T_p^{(\text{on})} \leq T_{\max}^{(\text{on})} < \infty \wedge S+s+1 \leq T_p^{(\text{off})} \leq T_{\max}^{(\text{off})} < \infty)$, where $T_p^{(\text{on})}$ and $T_p^{(\text{off})}$ generally are i.i.d. In special case, $T_p^{(\text{on})}$ as well as $T_p^{(\text{off})}$ could remain constant, and, as a result, the entire $\{u_k\}$ process would be a strict-sense cyclostationarity. For all $0 \leq p \leq \mathcal{P}$, intervals' lengths, T_p, are known beforehand, including values of $T_p^{(\text{on})}$ and $T_p^{(\text{off})}$.

[1]Proposed algorithm could also be used when the ARMA object is unstable, but its state is set to initial every time a new period begins.

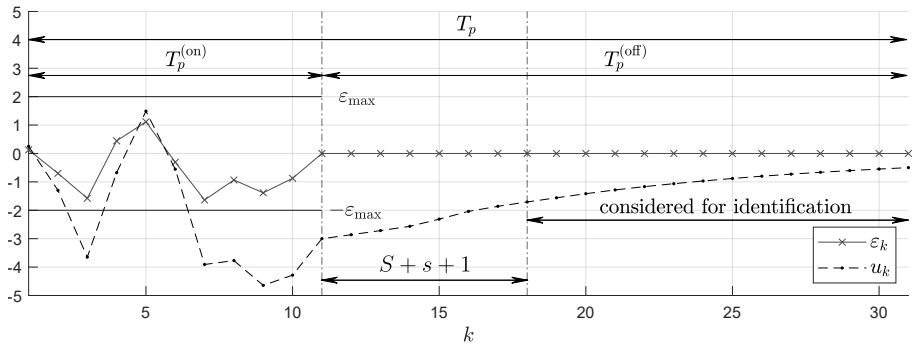

Fig. 2. Illustration of assumed properties of the input signal along with adopted notation: T_p, $T_p^{(\text{on})}$, $T_p^{(\text{off})}$, $\varepsilon_{\max}$.

The size of the data set, denoted as N, equals the total length of the input/output sequence, such that, $N = \sum_{p=0}^{\mathcal{P}-1} T_p = \sum_{p=0}^{\mathcal{P}-1}(T_p^{(\text{on})} + T_p^{(\text{off})})$.

ARMA object condition should belong initially to the set of all states achievable at the start of the interval in the asymptotical sense, i.e., when $p \to \infty$.

Assumption II can be interpreted with the help of the Fig. 2, in which a single interval is visualized. Additionally, Fig. 3 provides insight into how trajectory can be positioned in the state space. The ARMA object is understood as the surroundings before the considered system, where the latter shall be identified when the input-forming object works in a free state. Visually, vectors of considered sequences in state space can be compared to the ones from the previous works as in Fig. 4. Adequate directions in state space are not limited to impulses [26], piece-wise constants [27], or randomly initiated cyclo-exponential (RICE) signals [24].

[Disturbances] The process $\{z_k\}$ is an error in output observations, $y_k = v_k + z_k$, and consists of random variables, which are i.i.d. and do not depend on the input sequence $\{u_k\}$. The noise has a zero expected value, $\mathrm{E}\{z_k\} = 0$, and its variance is finite, $\mathrm{var}\{z_k\} < \infty$.

III. The Algorithm

The main concept originates from the observation that merely r newest samples of $\{u_k\}$ process can determine unambiguously not only the state of the ARMA object at the input, but also the state of the entire system under certain condition[2]. Therefore, the state vector, denoted as $\varphi_k = [u_k, u_{k-1}, ..., u_{k-r+1}]^T$, is changing through the entire process according to the rule shown below

$$\varphi_k = \underbrace{\begin{bmatrix} a^T \\ I_{r-1} \quad \mathbf{0}_{r-1} \end{bmatrix}}_{=A} \varphi_{k-1} + \underbrace{\begin{bmatrix} b^T \\ \mathbf{0}_{r-1,s} \end{bmatrix}}_{=B} \varepsilon_{k,s}, \tag{2}$$

where $a = [a_1, a_2, ..., a_r]^T$ denotes the vector of the autoregression parameters, I_{r-1} is the identity matrix of $r-1$ order, $\mathbf{0}_{r-1}$ is the $(r-1)$-dimensional column vector of zeros, $\mathbf{0}_{r-1,s}$ is the $(r-1)\times s$ matrix of zeros, and $b = [b_0, b_1, ..., b_s]^T$ is the moving average component parameters' vector associated with the excitation vector $\varepsilon_{k,s} = [\varepsilon_k, \varepsilon_{k-1}, ..., \varepsilon_{k-s}]^T$. To enhance the clarity of the presentation, the initial state is described in the following equation

$$\varphi_0 = \alpha u. \tag{3}$$

In this context, α represents a unit r-dimensional column vector of direction within a state space, characterized by the property that $\|\alpha\|_2 = 1$. Simultaneously, the variable u functions as a scaling factor. In the case of the sequence $\{\varepsilon_k\}$ consisting entirely of zeros, the following holds true

$$\varphi_k = A^k \varphi_0, \tag{4}$$

$$u_k = [1, \mathbf{0}_{r-1}^T]\varphi_k = [1, \mathbf{0}_{r-1}^T] A^k \alpha u, \tag{5}$$

as well as

$$\begin{aligned} \mathrm{E}\{y_S | \varphi_0 = \alpha u\} &= \mu\left(\sum_{i=0}^{S} \lambda_i u_{S-i} \middle| \varphi_0 = \alpha u \right) \\ &= \mu\Bigg(\underbrace{\sum_{i=0}^{S} \lambda_i [1, 0_{r-1}^T] A^{S-i} \alpha}_{=\text{const.}} u \Bigg), \end{aligned} \tag{6}$$

which can be generalized as follows

$$\mathrm{E}\{y_S | \varphi_{S-\Delta k}\} = \mu\Big(d_{\Delta k}^{(\alpha)} u \Big), \tag{7}$$

where $\Delta k = 0, 1, ..., S$, and constant is defined below

$$d_{\Delta k}^{(\alpha)} = \sum_{i=0}^{S} \lambda_i [1, 0_{r-1}^T] A^{\Delta k - i} \alpha. \tag{8}$$

In other words, any r neighboring samples can be chosen from the sequence of $S+r$ samples (where $0 \le k \le S+1$, and additional $r-1$ samples are drawn from initial state φ_0). This is because, based on any subsequence of r consecutive samples, the entire $S+r$ long sequence can be reconstructed. Matrix A is always invertible, because its determinant is equal to $a_r(-1)^{r+1}$, what can be shown by using a standard method of Laplace expansion. When $a_r = 0$, the problem is simplified to the one with a lower order of autoregression.

Finally, the estimator of static nonlinear block characteristics is proposed

$$\widehat{\mu}_N^{(\alpha)}(x) = \sum_{k \in K_N} y_k w_k, \tag{9}$$

[2]Condition of free state operation equivalent to zeroing the $\{\varepsilon_k\}$ for $S+s+1$ samples in a row, formally $\forall_{i:(0\le i\le S+s)}\ \varepsilon_{k-i} = 0$.

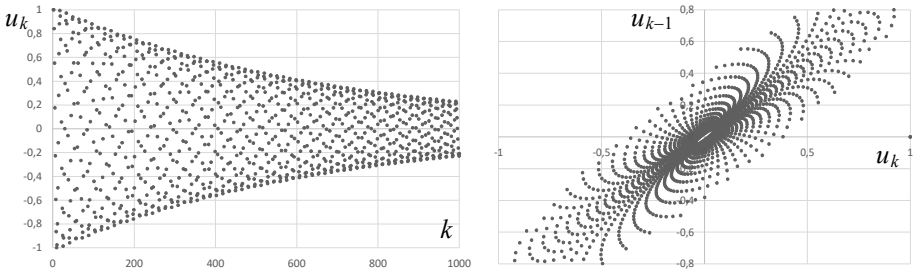

Fig. 3. Exemplary sequence entering the system $\{u_k\}$, when autoregression depends on two last samples, i.e. $r = 2$, and trajectory in state space $\{[u_k, u_{k-1}]\}$.

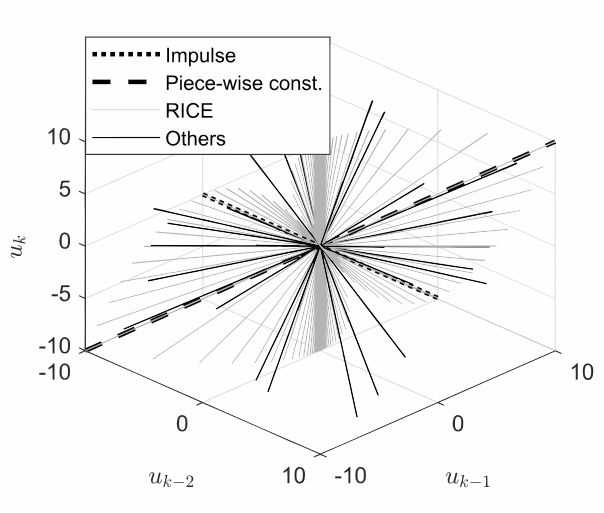

Fig. 4. Directions of scaling in multidimensional space (3-dimensional example) including impulses [26], piecewise constants [27], randomly initiated cyclo-exponential (RICE) signals [24], and others.

with the kernel weights

$$w_k = \frac{\mathcal{K}\left(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\right)}{\sum_{k\in K_N}\mathcal{K}\left(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\right)}$$

where set $K_N = \{k : \varepsilon_{k,S+s} = \mathbf{0}_{S+s+1} \wedge N \geq k \geq 1\}$, and h is bandwidth parameter depending on the data set's size, i.e., it is a function of N. Kernel function $\mathcal{K}(\cdot)$ may be choosen arbitrarily as long as $\mathcal{K}(x) = 0$ for $|x| > 1$, and $0 < \mathcal{K}_{\min} \leq \mathcal{K}(x) \leq \mathcal{K}_{\max} < \infty$ when $|x| \leq 1$. Estimate $\widehat{\mu}_N^{(\alpha)}(x)$ converges uniformly to the true nonlinearity up to the scale factor, i.e. to $\mu^{(\alpha)}(x) = \mu(d_0^{(\alpha)}x)$. When the denominator is equal to zero, the value of the estimator is assumed to be also zero.

To define the considered domain of identifiability, first, the supplementary notation is introduced. Function $\widetilde{p}(k)$ returns a number of interval containing k'th sample, i.e.,

$$\widetilde{p}(k) = \arg\min_{p_1}\left(\sum_{p=0}^{p_1} T_p \geq k\right). \tag{10}$$

Based on that, function $\widetilde{k}_0(k)$ provide a number of sample within period $\widetilde{p}(k)$ as follows

$$\widetilde{k}_0(k) = k - \sum_{p=0}^{\widetilde{p}(k)-1} T_p, \tag{11}$$

with possible values from 1 to $T_{\widetilde{p}(N)}$. When the upper bound of the sum is lower than zero, i.e., $\widetilde{p}(k) - 1 < 0$, then the sum is empty and equal to zero. In the next step, a subset of K_N consisting of samples' numbers belonging to only one considered interval is defined as

$$K_{N,p} = \{k : \widetilde{p}(k) = p \wedge k \in K_N\}. \tag{12}$$

With the help of the introduced notation, state φ_k can be expressed as a sum of the component depending on the $\{\varepsilon_k\}$ process samples from the current interval, $\varphi_k^{(\text{new})}$, and the impact of previous intervals, $\varphi_k^{(\text{old})}$, i.e.,

$$\varphi_k = \underbrace{\overbrace{A^k\varphi_0}^{\text{Initial Condition}} + \overbrace{\sum_{j=\widetilde{k}_0(k)}^{k-1} A^j B\varepsilon_{k-j,s}}^{\text{Impact of old } \varepsilon_k}}_{\stackrel{\text{def}}{=}\varphi_k^{(\text{old})}} + \underbrace{\overbrace{\sum_{j=0}^{\widetilde{k}_0(k)-1} A^j B\varepsilon_{k-j,s}}^{\text{Impact of new } \varepsilon_k}}_{\stackrel{\text{def}}{=}\varphi_k^{(\text{new})}}. \tag{13}$$

Consequently, the contribution of previous intervals to the current state can be rewritten as follows

$$\varphi_k^{(\text{old})} = A^{\widetilde{k}_0(k)-1}\varphi_{\text{init},\widetilde{p}(k)}^{(\text{old})}, \tag{14}$$

where $\varphi_{\text{init},p}^{(\text{old})}$ is the value of $\varphi_k^{(\text{old})}$ when the p'th interval begins, with the formal definition specified below

$$\varphi_{\text{init},p}^{(\text{old})} = \varphi_k^{(\text{old})} : \widetilde{k}_0(k) = 1 \ \wedge \ \widetilde{p}(k) = p. \tag{15}$$

All possible values of $\varphi_{\text{init},p}^{(\text{old})}$, in asymptotical sense for $p \to \infty$ and $N \to \infty$, form the set φ, which can be defined as follows

$$\varphi = \{\varphi^{(\text{old})} : \forall_{h>0} \lim_{N\to\infty} \mathrm{P}\left(\|\varphi_{\text{init},\widetilde{p}(N)}^{(\text{old})} - \varphi^{(\text{old})}\|_\infty \leq h\right) > 0\}. \tag{16}$$

Finally, the domain of identifiability can be described with the help of the subsequent equation

$$\mathbb{D}^{(\alpha)} = \Big\{x : \forall_{h>0, k\in K_{N,\widetilde{p}(N)}, \varphi_{\text{init}}^{(\text{old})}\in\varphi} \lim_{N\to\infty} \mathrm{P}\Big(\big\|\varphi_k^{(\text{new})} + A^{\widetilde{k}_0(k)-1}\varphi_{\text{init}}^{(\text{old})} - \alpha x\big\|_\infty \leq h\Big) > 0\Big\}. \tag{17}$$

Nonlinearity could also be identified at certain points beyond the defined domain of identifiability, but under more restrictive requirements, which imply a slower rate of convergence. Nevertheless, most commonly, $\varphi_k^{(\text{old})}$ is approximately equal to $\mathbf{0}_r$ due to being affected only by older non-zero samples of ε_k with indices that are at least $r + 2(S + s + 1)$

smaller than the current index. The above results in the dissimilarity among the previously mentioned domains being minimal. As a result of the previously discussed factors, those kinds of scenarios are not considered further in the paper. The theorem is hereby introduced in a formal manner, with the necessary conditions for achieving the convergence. If $Nh^r \to \infty$, and $h \to 0$, then $\widehat{\mu}_N^{(\alpha)}(x)$ converges to $\mu^{(\alpha)}(x)$ in the mean squared sense,

$$\mathrm{MSE}\left\{\widehat{\mu}_N^{(\alpha)}(x)\right\} = \mathrm{E}\left\{\left(\widehat{\mu}_N^{(\alpha)}(x) - \mu^{(\alpha)}(x)\right)^2\right\} \to 0, \quad (18)$$

in each point $x \in \mathbb{D}^{(\alpha)}$, for any given α, as $N \to \infty$.

Proof. See Appendix A. □

[Wiener-Hammerstein system identification] Based on the same concept, the set of scaled values of the interconnection signal $\{x_k\}$ can be obtained and the entire Hammerstein-like part could be identified up to the scale factor (instead of just a nonlinear block), which implies potential application in Wiener-Hammerstein system identification.

IV. Aggregation

Note that the proposed estimator (9) uses only a narrow subset of available data, i.e., which are in proximity of the linear space spanned by the direction vector α. Therefore, the gain resulting from dimensionality reduction comes at the cost of impoverishing the set of data. From the perspective of estimation efficiency, it is therefore crucial to find a way to aggregate many estimators with different values of the α vector. Intuitively, using more estimators promises to reduce the variance of the final result.

In this section, we show an idea for aggregating many identified nonlinearities. Nevertheless, for simplicity of the presentation, the reasoning is limited to only two estimators, $\widehat{\mu}_N^{(\alpha_1)}(x)$ and $\widehat{\mu}_N^{(\alpha_2)}(x)$, with unknown scaling factors $d_0^{(\alpha_1)}$ and $d_0^{(\alpha_2)}$, respectively, i.e., with the limits

$$\widehat{\mu}_N^{(\alpha_1)}(x) \to \mu(d_0^{(\alpha_1)}x) \quad \text{and} \quad \widehat{\mu}_N^{(\alpha_2)}(x) \to \mu(d_0^{(\alpha_2)}x).$$

Without losing the generality of the approach, we can assume that $d_0^{(\alpha_1)} = 1$. When the bias and errors' dependency on the estimators are negligible, the following weighted average is optimal in terms of variance reduction

$$\widehat{\mu}_N^{(\mathrm{aggr})}(x) = \frac{n_1 \widehat{\mu}_N^{(\alpha_1)}(x) + n_2 \widehat{\mu}_N^{(\alpha_2)}\left(\frac{x}{d_0^{(\alpha_2)}}\right)}{n_1 + n_2}, \quad (19)$$

where

$$n_1 = \sum_{k \in K_N} \mathcal{K}\left(\frac{1}{h}\|\varphi_k - \alpha_1 x\|_\infty\right),$$
$$n_2 = \sum_{k \in K_N} \mathcal{K}\left(\frac{1}{h}\left\|\varphi_k - \alpha_2 \frac{x}{d_0^{(\alpha_2)}}\right\|_\infty\right).$$

Implementing the idea presented in (19) requires knowledge of the $d_0^{(\alpha_2)}$ scale. Below, we propose two intuitive methods for estimating the scale factor. At this stage, we present them without formal justification and treat their analysis as open problems.

Fitting curves by numerical minimization of the following mean square error measure

$$\widehat{d}_0^{(\alpha_2)} = \arg\min_c \frac{1}{\|\mathcal{D}\|} \int_{\mathcal{D}} \left(\widehat{\mu}_N^{(\alpha_1)}(x) - \widehat{\mu}_N^{(\alpha_2)}(cx)\right)^2 dx. \quad (20)$$

The set $\mathcal{D}$ should be determined so that both $\widehat{\mu}_N^{(\alpha_1)}(x)$ and $\widehat{\mu}_N^{(\alpha_2)}(cx)$ are well defined within the considered range of the scaling variable c. When nonlinearity exhibits some fractal-like patterns, there may be more than one solution.

Estimation of directional derivatives of functions at some point $\varphi = \varphi_{\mathrm{ref}}$ and comparison of them with each other, where $\mu^{(\alpha)}(x) = \mu(d_0^{(\alpha)}x) = \mu(V^T\alpha x) = \mu(V^T\varphi)$, and V^T is a constant part of (8) without α. This approach, for $k \in K_N$, leads to

$$\nabla_\alpha \mu(V^T\varphi) = \alpha^T \nabla \mu(V^T\varphi) = \underbrace{\alpha^T V}_{=d_0^{(\alpha)}} \mu'(V^T\varphi), \quad (21)$$

and

$$\frac{\nabla_{\alpha_2}\left[\mu(V^T\varphi)\right]_{\varphi=\varphi_{\mathrm{ref}}}}{\nabla_{\alpha_1}\left[\mu(V^T\varphi)\right]_{\varphi=\varphi_{\mathrm{ref}}}} = \frac{d_0^{(\alpha_2)}}{d_0^{(\alpha_1)}} = d_0^{(\alpha_2)} \ \text{ if } \ \mu'(V^T\varphi_{\mathrm{ref}}) \neq 0. \quad (22)$$

The estimation of the derivative of the nonlinearity in a system with the Wiener structure can be realized, for example, using the local linearization as in the method described in [6].

V. Numerical Experiment

In this section, the convergence of the algorithm is illustrated with the help of the numerical experiment. The method considered in the paper is suited for the special class of excitations, and is intended for the identification of nonlinear blocks in the Wiener systems. Identification of the entire Wiener system based on a known nonlinear block is not a trivial problem, which is not considered in this paper. Due to the aforementioned facts, finding a reasonable method for comparison is difficult. Other approaches often focus, in the first place, on the identification of the dynamic block in the Wiener system, the knowledge of which makes the identification of the entire system straightforward. In contrast, in this paper, the advantage of the opposite approach, i.e., dimensionality reduction, is considered.

The settings of the experiment are shown in the Tab. I. In Fig. 5, two identified nonlinear characteristics for different directions in a state space, which differ in scale factor, are presented. The same scale factors were used to show a reference, i.e., scaled versions of true nonlinearity. In Fig. 6, the relationship between the RMSE (root mean squared error) and the number of data is shown, and consequently, the convergence of the algorithm is illustrated. The RMSE was computed based on identified nonlinearities and their references, using the grid of 201 uniformly distributed points, within the interval $x \in [-1, 1]$, where distances between neighboring points were equal to 0.01.

TABLE I
EXPERIMENT SETTINGS

T_p	$T_p^{(\mathrm{on})}$	$T_p^{(\mathrm{off})}$	S	λ_i	$\mu(x)$	a (AR)	b (MA)	z_k	$\mathcal{P}$	N	h
25 (const.)	10 (const.)	15 (const.)	6	$\frac{S+1-i}{S+1}$	$0.5x^2$	$[0.6, -0.39]^T$	$[1, 3]^T$	$\sim \mathcal{U}[-1, 1]$	40000	1000000	$N^{-1/4}$

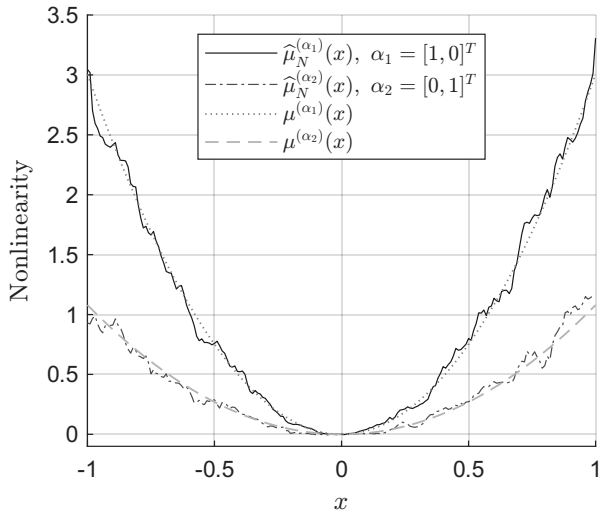

Fig. 5. Nonlinear characteristics identified for different directions in state space α, and number of data equalt to $N = 10^6$.

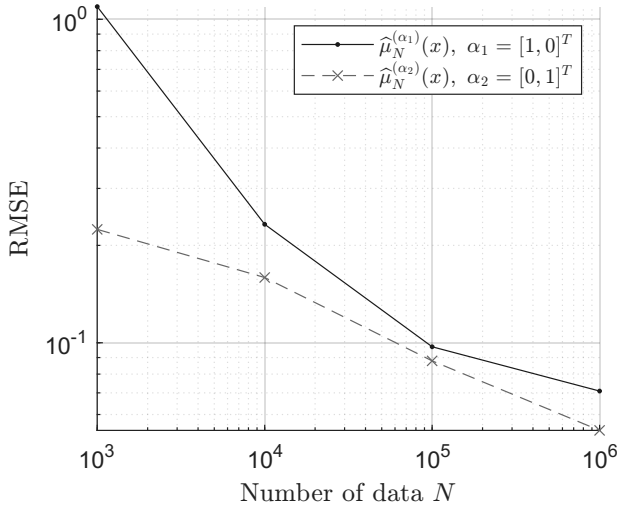

Fig. 6. RMSE vs. number of data.

VI. SUMMARY

The proposed algorithm appears to hold great potential for practical applications. Usually, the input signal is the output of another dynamic system. When such a preceding object is an ARMA filter then, disabling thereof enables the identification to be performed with a higher convergence speed.

A relevant generalization of the proposed method involves extending its applicability to the identification of systems with Wiener-Hammerstein structure, which offers enhanced approximation capabilities for dynamic-nonlinear phenomena modeling.

Another generalization worth consideration is the aggregation of the estimators obtained under the conditions of many unique direction vectors in the state space. Specified problem is challenging as each nonlinearity estimator has an unknown, and in general, different scale. Furthermore, identified characteristics also model true nonlinearity in different parts of its domain, which are unknown too, except not affected by scaling point zero. The presence of estimation errors makes fitting curves to each other even more difficult. Some nonlinearities with fractal-like patterns can lead to ambiguity and multiple solutions when attempting to estimate scale factors.

VII. APPENDICES

APPENDIX

Proof. Let $K_N^{(\alpha,h,x)} = \{k : \mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big) > 0 \wedge k \in K_N\}$ be a set of data indices selected by the kernel for given x, h, and α. Considering the set $\mathbb{D}^{(\alpha)}$, guaranteeing identifiability, the number of selected samples is random, and asymptotically, i.e., for $N \to \infty$ and $h \to 0$, its expected value behaves like

$$
\begin{aligned}
&\mathrm{E}\Big\{\#K_N^{(\alpha,h,x)}\Big\} = \mathcal{P}\mathrm{E}\Big\{\#\Big(K_N^{(\alpha,h,x)} \cap K_{N,p}\Big)\Big\} \\
&= \mathcal{P}\mathrm{E}\Bigg\{\sum_{\tau=1}^{T_{\max}^{(\mathrm{off})}-S-s} \#\Bigg\{k : k \in \Big(K_N^{(\alpha,h,x)} \cap K_{N,p}\Big) \wedge \\
&k = \underbrace{\sum_{q=0}^{p-1} T_q + T_p^{(\mathrm{on})} + S + s + \tau}_{\stackrel{\mathrm{def}}{=}\kappa(p,\tau)}\Bigg\}\Bigg\} = \underbrace{\mathcal{P}}_{=c_1 N} \sum_{\tau=1}^{T_{\max}^{(\mathrm{off})}-S-s} \\
&\mathrm{P}\Big(T_p^{(\mathrm{off})} - S - s \geq \tau \ \wedge\ \mathcal{K}\big(\tfrac{1}{h}\|\varphi_{\kappa(p,\tau)} - \alpha x\|_\infty\big) > 0\Big) \\
&= c_1 N \sum_{\tau=1}^{T_{\max}^{(\mathrm{off})}-S-s} \underbrace{\mathrm{P}\Big(T_p^{(\mathrm{off})} - S - s \geq \tau\Big)}_{=c_{2,1,\tau}} \cdot \\
&\underbrace{\mathrm{P}\Big(\mathcal{K}\big(\tfrac{1}{h}\|\varphi_{\kappa(p,\tau)} - \alpha x\|_\infty\big) > 0 \Big| T_p^{(\mathrm{off})} - S - s \geq \tau\Big)}_{=c_{2,2,\tau}h^r} \\
&= c_1 N \sum_{\tau=1}^{T_{\max}^{(\mathrm{off})}-S-s} c_{2,\tau} h^r = c_1 c_2 N h^r = c_3 N h^r, \qquad (23)
\end{aligned}
$$

where c_1, c_2, and c_3 denote some unknown constants. Obviously, $\mathrm{E}\{\#K_N^{(\alpha,h,x)}\} \to \infty$, provided that $Nh^r \to \infty$. Assuming that the set of selected observations is not empty,

i.e.,$\#\widetilde{K}_N^{(\alpha,h,x)} \geq 1$, where $\widetilde{K}_N^{(\alpha,h,x)}$ detotes realization of $K_N^{(\alpha,h,x)}$, we get asymptotically

$$\begin{aligned}
&\widehat{\mu}_N^{(\alpha)}(x) - \mu^{(\alpha)}(x) = \\
&\frac{\sum_{k\in K_N} (\overbrace{\mu(x_k)+z_k}^{=y_k} - \mu^{(\alpha)}(x))\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)}{\sum_{k\in K_N}\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)} \\
&= \underbrace{\frac{\sum_{k\in K_N} z_k\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)}{\sum_{k\in K_N}\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)}}_{\stackrel{\text{def}}{=}\delta_N^{(1)}} + \\
&\underbrace{\frac{\sum_{k\in K_N} \big(\mu(x_k) - \mu^{(\alpha)}(x)\big)\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)}{\sum_{k\in K_N}\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)}}_{\stackrel{\text{def}}{=}\delta_N^{(2)}}
\end{aligned} \tag{24}$$

and since the disturbance is independent of the inputs, we obtain

$$\begin{aligned}
&\text{MSE}\Big\{\widehat{\mu}_N^{(\alpha)}(x)\Big\} = \text{E}\Big\{\delta_N^{(1)} + \delta_N^{(2)}\Big\}^2 = \text{E}\Big\{{\delta_N^{(1)}}^2\Big\} + \\
&2\underbrace{\text{E}\Big\{\delta_N^{(1)}\Big\}}_{=0}\text{E}\Big\{\delta_N^{(2)}\Big\} + \text{E}\Big\{{\delta_N^{(2)}}^2\Big\}.
\end{aligned} \tag{25}$$

Asymptotically, the first term in eqn. (25) can be evaluated in the following way

$$\begin{aligned}
\text{E}\Big\{{\delta_N^{(1)}}^2\Big\} &\leq \text{E}\left\{\frac{\sum_{k\in K_N^{(\alpha,h,x)}} z_k^2\mathcal{K}_{\max}^2}{\big(\sum_{k\in K_N^{(\alpha,h,x)}}\mathcal{K}_{\min}\big)^2}\right\} \\
&= \text{var}\{z_k\}\frac{\mathcal{K}_{\max}^2}{\mathcal{K}_{\min}^2}\text{E}\left\{\frac{1}{\#K_N^{(\alpha,h,x)}}\right\}.
\end{aligned} \tag{26}$$

Then, let us introduce the following decomposition of the number of active measurements with respect to p

$$\#K_N^{(\alpha,h,x)} = \sum_{p=0}^{\mathcal{P}-1}\mathbb{K}_p, \tag{27}$$

where $\mathbb{K}_p \stackrel{\text{def}}{=} \#\big\{K_N^{(\alpha,h,x)} \cap K_{N,p}\big\}$.
With the above notation we can write that $\frac{1}{\mathcal{P}h^r}\#K_N^{(\alpha,h,x)} = \frac{1}{\mathcal{P}}\sum_{p=0}^{\mathcal{P}-1}\frac{1}{h^r}\mathbb{K}_p$, and based on the formula (23), for $h\to 0$ we obtain asymptotically

$$\text{E}\big\{\tfrac{1}{h^r}\mathbb{K}_p\big\} = \tfrac{1}{h^r}\underbrace{\text{E}\big\{\mathbb{K}_p\big\}}_{=c_4h^r} = c_4, \tag{28}$$

and

$$\begin{aligned}
&\text{var}\big\{\tfrac{1}{h^r}\mathbb{K}_p\big\} = \tfrac{1}{h^{2r}}\text{var}\big\{\mathbb{K}_p\big|\mathbb{K}_p = 0\big\}\text{P}\big(\mathbb{K}_p = 0\big) + \\
&\tfrac{1}{h^{2r}}\text{var}\big\{\mathbb{K}_p\big|\mathbb{K}_p \geq 1\big\}\text{P}\big(\mathbb{K}_p \geq 1\big) \\
&< \tfrac{1}{h^{2r}}\big(0 - \text{E}\big\{\mathbb{K}_p\big\}\big)^2\big(1 - c_5h^r\big) + \\
&\tfrac{1}{h^{2r}}\big(\max\big\{\mathbb{K}_p\big\} - \min\big\{\mathbb{K}_p\big\}\big)^2c_5h^r \\
&\leq \tfrac{1}{h^{2r}}c_4^2h^{2r}\big(1 - c_5h^r\big) + \tfrac{1}{h^{2r}}\big(T_{\max}^{(\text{off})} - S - s\big)^2c_5h^r \\
&\leq \frac{c_6}{h^r}
\end{aligned} \tag{29}$$

Moreover, since $\{\mathbb{K}_p\}_{p=0}^{\mathcal{P}}$ constitutes an i.i.d. sequence, and

$$\sum_{p=0}^{\mathcal{P}-1}\frac{\text{var}\big\{\mathbb{K}_p\big\}}{\mathcal{P}^2} < \frac{\mathcal{P}c_6}{\mathcal{P}^2h^r} = \frac{c_6}{\mathcal{P}h^r}, \tag{30}$$

the strong law of large numbers holds, and consequently

$$\frac{1}{\mathcal{P}h^r}\#K_N^{(\alpha,h,x)} = \frac{1}{\mathcal{P}}\sum_{p=0}^{\mathcal{P}-1}\frac{1}{h^r}\mathbb{K}_p \to c_4 \tag{31}$$

with probability one as $\mathcal{P}h^r \to \infty$, which is fulfilled for $Nh^r \to \infty$ as $\mathcal{P} = c_1N$ for N large. Since $\frac{1}{Nh^r}\#K_N^{(\alpha,h,x)}$ becomes constant with probability one, by the Slutsky's theorem we have that

$$c_7\text{E}\left\{\frac{1}{\#K_N^{(\alpha,h,x)}}\right\} = c_7\text{E}\left\{\frac{\frac{1}{Nh^r}}{\frac{1}{Nh^r}\#K_N^{(\alpha,h,x)}}\right\} = \frac{c_8}{Nh^r}, \tag{32}$$

which leads to

$$\lim_{N\to\infty}\text{E}\Big\{{\delta_N^{(1)}}^2\Big\} \leq \lim_{N\to\infty}\frac{c_8}{Nh^r} = 0 \iff Nh^r \to \infty. \tag{33}$$

To convergence of second term, $\text{E}\Big\{{\delta_N^{(2)}}^2\Big\}$, will be proven using the Lipschitz condition

$$\big|\mu(d_k^{(\alpha,h,x)}x_k) - \mu(d_0^{(\alpha)}x)\big| \leq \mathcal{L}\big|d_k^{(\alpha,h,x)}x_k - d_0^{(\alpha)}x\big|, \tag{34}$$

where $d_k^{(\alpha,h,x)}$ denotes the scale with respect to k'th sample. For $k \in K_N^{(\alpha,h,x)} \wedge \#\widetilde{K}_N^{(\alpha,h,x)} \geq 1$, the following properties take place

$$\begin{aligned}
&\big|d_k^{(\alpha,h,x)}x_k - d_0^{(\alpha)}x\big| \leq \big|\overbrace{(d_k^{(\alpha,h,x)} - d_0^{(\alpha)})}^{\stackrel{\text{def}}{=}\Delta d_k^{(\alpha,h,x)}}x_k\big| + \\
&\big|d_0^{(\alpha)}(x_k - x)\big| \leq \big|\Delta d_k^{(\alpha,h,x)}x_k\big| + c_9h,
\end{aligned} \tag{35}$$

and $|\mu(x_k) - \mu^{(\alpha)}(x)| \leq \mathcal{L}\big(\big|\Delta d_k^{(\alpha,h,x)}x_k\big| + c_9h\big) \stackrel{\text{def}}{=} \beta_k^{(\alpha,h,x)}$, and further

$$\begin{aligned}
&\text{E}\Big\{{\delta_N^{(2)}}^2\Big\} \\
&\leq \text{E}\left\{\left|\frac{\sum_{k\in K_N}\beta_k^{(\alpha,h,x)}\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)}{\sum_{k\in K_N}\mathcal{K}\big(\frac{1}{h}\|\varphi_k - \alpha x\|_\infty\big)}\right|^2\right\} \\
&\leq \max_{k\in K_N}{\beta_k^{(\alpha,h,x)}}^2.
\end{aligned} \tag{36}$$

Regarding (8) we get

$$\Delta d_k^{(\alpha,h,x)} = d_k^{(\alpha,h,x)} - d_0^{(\alpha)} = V^T\alpha_{\text{eff}} + V^T\alpha = V^T(\alpha_{\text{eff}} - \alpha), \tag{37}$$

where V^T is a $1\times r$ vector of constants. Similarly to formula (3), vector α_{eff} determines selected direction in r-dimensional state space. For $x \neq 0$ and $h \to 0$, it holds that $\alpha_{\text{eff}} \to \alpha$, and consequently $\Delta d_k^{(\alpha,h,x)} \to 0$. Moreover, for $x = 0$ and $h \to 0$, $x_k \to 0$ for each selected observetion. Based on the above conclusions, $\beta_k^{(\alpha,h,x)} \to 0$ as $h \to 0$, and

$$\lim_{N\to\infty}\text{E}\Big\{{\delta_N^{(2)}}^2\Big\} \leq \lim_{N\to\infty}\max_{k\in K_N}{\beta_k^{(\alpha,h,x)}}^2 = 0, \tag{38}$$

when $h \to 0$, and $\#\widetilde{K}_N^{(\alpha,h,x)} \geq 1$. Obviously, since $Nh^r \to \infty$ we get finally

$$\lim_{N\to\infty} \mathrm{MSE}\left\{\widehat{\mu}_N^{(\alpha)}(x)\right\} = \lim_{N\to\infty} \mathrm{E}\left\{{\delta_N^{(1)}}^2\right\} + \lim_{N\to\infty} \mathrm{E}\left\{{\delta_N^{(2)}}^2\right\} = 0, \tag{39}$$

as $h \to 0$ what ends the proof. □

REFERENCES

[1] M. Ławryńczuk, *Computationally Efficient Model Predictive Control Algorithms*, ser. Studies in Systems, Decision and Control. Springer Cham, 2014, vol. 3. [Online]. Available: https://doi.org/10.1007/978-3-319-04229-9

[2] F. Giri and E.-W. Bai, *Block-oriented nonlinear system identification*. Springer, 2010, vol. 1. [Online]. Available: https://doi.org/10.1007/978-1-84996-513-2

[3] W. Härdle, *Applied Nonparametric Regression*, ser. Econometric Society Monographs. Cambridge University Press, 1990. [Online]. Available: https://doi.org/10.1017/CCOL0521382483

[4] L. Györfi, M. Kohler, A. Krzyżak, H. Walk *et al.*, *A distribution-free theory of nonparametric regression*. Springer, 2002, vol. 1. [Online]. Available: https://doi.org/10.1007/b97848

[5] W. Greblicki and M. Pawlak, *Nonparametric system identification*. Cambridge University Press, Cambridge, 2008, vol. 10. [Online]. Available: https://doi.org/10.1017/CBO9780511536687

[6] G. Mzyk and G. Maik, "Nonparametric identification of Wiener system with a subclass of wide-sense cyclostationary excitations," *International Journal of Adaptive Control and Signal Processing*, vol. 38, no. 1, pp. 323–341, 2024. [Online]. Available: https://doi.org/10.1002/acs.3702

[7] G. Mzyk, "Identification of time-varying non-linear systems with adaptive bootstrap-based tracking," *Mechanical Systems and Signal Processing*, vol. 223, p. 111896, 2025. [Online]. Available: https://doi.org/10.1016/j.ymssp.2024.111896

[8] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 2859–2900, 2015. [Online]. Available: https://dl.acm.org/doi/10.5555/2789272.2912091

[9] J. P. Cunningham and B. M. Yu, "Dimensionality reduction for large-scale neural recordings," *Nature neuroscience*, vol. 17, no. 11, pp. 1500–1509, 2014. [Online]. Available: https://doi.org/10.1038/nn.3776

[10] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2663–2693, 2022. [Online]. Available: https://doi.org/10.1007/s40747-021-00637-x

[11] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010. [Online]. Available: https://doi.org/10.1016/j.arcontrol.2009.12.001

[12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000. [Online]. Available: https://doi.org/10.1126/science.290.5500.2323

[13] G. Rega and H. Troger, "Dimension reduction of dynamical systems: methods, models, applications," *Nonlinear Dynamics*, vol. 41, pp. 1–15, 2005. [Online]. Available: https://doi.org/10.1007/s11071-005-2790-3

[14] M. Cenedese, J. Axås, H. Yang, M. Eriten, and G. Haller, "Data-driven nonlinear model reduction to spectral submanifolds in mechanical systems," *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2229, p. 20210194, 2022. [Online]. Available: https://doi.org/10.1098/rsta.2021.0194

[15] J. Li, Y. Wang, X. Jin, Z. Huang, and I. Elishakoff, "Data-driven method for dimension reduction of nonlinear randomly vibrating systems," *Nonlinear Dynamics*, vol. 105, no. 2, pp. 1297–1311, 2021. [Online]. Available: https://doi.org/10.1007/s11071-021-06601-1

[16] A. Juditsky, H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg, and Q. Zhang, "Nonlinear black-box models in system identification: Mathematical foundations," *Automatica*, vol. 31, no. 12, pp. 1725–1750, 1995. [Online]. Available: https://doi.org/10.1016/0005-1098(95)00119-1

[17] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005. [Online]. Available: https://doi.org/10.1109/TPAMI.2005.159

[18] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974. [Online]. Available: https://doi.org/10.1109/TAC.1974.1100705

[19] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[20] G. Mzyk, Z. Hasiewicz, and P. Mielcarek, "Kernel identification of non-linear systems with general structure," *Algorithms*, vol. 13, no. 12, p. 328, 2020. [Online]. Available: https://doi.org/10.3390/a13120328

[21] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach*. John Wiley & Sons, 2012. [Online]. Available: https://doi.org/10.1002/9781118287422

[22] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1802.03426

[23] P. Wachel, P. Śliwiński, and Z. Hasiewicz, "Nonparametric identification of MISO Hammerstein system from structured data," *Journal of Systems Science and Systems Engineering*, vol. 24, no. 1, pp. 68–80, 2015. [Online]. Available: https://doi.org/10.1007/s11518-014-5256-7

[24] G. Maik, G. Mzyk, and P. Wachel, "Exponential excitations for effective identification of Wiener system," *International Journal of Control*, vol. 97, no. 9, pp. 1956–1966, 2024. [Online]. Available: https://doi.org/10.1080/00207179.2023.2246070

[25] G. Maik and G. Mzyk, "Randomly initiated cyclostationary excitations for dimensionality reduction in Wiener system identification," in *System Dependability - Theory and Applications*, W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, and J. Kacprzyk, Eds. Cham: Springer Nature Switzerland, 2024, pp. 143–151. [Online]. Available: https://doi.org/10.1007/978-3-031-61857-4_14

[26] G. Mzyk, "A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 10, pp. 897–901, 2007. [Online]. Available: https://doi.org/10.1109/TCSII.2007.901634

[27] G. Mzyk and P. Wachel, "Kernel-based identification of Wiener–Hammerstein system," *Automatica*, vol. 83, pp. 275–281, 2017. [Online]. Available: https://doi.org/10.1016/j.automatica.2017.06.038