Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 21, Iss. 1, pp. 5-18, 2023
© Brazilian Society on Computational Intelligence

ESTIMATION OF KAUTZ POLES IN WIENER-VOLTERRA MODELS USING LEVENBERG-MARQUARDT ALGORITHM

Higor de Souza Serafin, Elder Oroski, and André Eugenio Lazzaretti
Universidade Tecnológica Federal do Paraná - Curitiba
serafin@alunos.utfpr.edu.br, {oroski,lazzaretti}@utfpr.edu.br

Abstract – This work approaches the problem of estimating the Kautz optimal poles in kernel expansion in Wiener-Volterra models. The analytical solution for the suboptimal case is already established in the literature. However, the solution for the two parameters that compose the poles is still open. In this paper, an optimization strategy using the Levenberg-Marquardt is presented. This algorithm is used to find kernel expansion parameters, with the same base for all dimensions. The construction of bases using digital filter is considered. To validate the implemented algorithm, data collected from the excitation of an electrically coupled drive system was used to analyze the impact of the search space thresholds and the behavior of Levenberg-Marquardt's parameters. It was also analyzed the impact on the model accuracy, as the number of functions in the base is increased. As a result, the models determined have achieved better results than the works found in the literature.
Keywords – Wiener-Volterra, Levenberg-Marquardt, System Identification, Kautz function.

1. INTRODUTION

As a result of the industrial advances in the last decades, the production processes became increasingly complex, and, as a consequence, industrial plants started to require more robust control systems, restricting the performance of the models in use due to their capacity to represent them [1]. In this way, developing techniques that meet industrial needs is necessary through more sophisticated strategies. Among these techniques, system identification can be highlighted.

According to [2], system identification aims to elaborate mathematical models for dynamic systems, based on input data and observed output data. This methodology, based on data observations, has its roots directed to the areas of statistics and time series analysis [3] and is currently extended to the most diversified branches of studies.

There are several techniques to perform the identification of systems, as is shown in the works of [2, 4–6]. These works use the parametric system identification technique, via Orthonormal Basis Function (OBF), to identify dynamic, nonlinear, causal, and time-invariant systems.

The use of the OBF has shown an increasing interest in several studies [1, 7–9] in the areas of identification of dynamic systems. These works are based on the systems' representation using orthonormal basis functions in a given space. Since the property of orthonormality enables: (i) incorporate a priori knowledge of the system; and (ii) a simpler representative models [9].In this context, applying theses functions, it is possible to specify more suitable models for systems to be identified.

The most common function bases to be used are: (i) Laguerre; and (ii) Kautz, as these functions have a great capacity to represent systems with dynamics of first and second-order [10]. In order to represent higher-order systems, the most appropriate is the use of the Generalized Orthonormal Basis Functions (GOBF) [6].

Laguerre's functions are parameterized by a real pole and are more suitable for modeling super-damped systems [1], systems are those with a damping coefficient greater than 1, $\xi > 1$. Its optimal pole choice, in an analytical way, has already been explored in the literature. It has well-defined solutions for continuous and discrete cases, as well as for linear and nonlinear systems.

However, for modeling sub-damping systems, Laguerre's function bases is required a high number of functions, making these functions less interesting for those cases. Thus, Kautz functions appear as an alternative, since resonant poles parameterize them. In other words, a pair of complex conjugate poles, facilitating the representation of oscillatory systems [11].

An important work to highlight those properties is [11], which uses Kautz functions to model linear systems, and, through linear regression methods, a strategy to reduce the number of parameters to be estimated was explored. In [12], a general formulation for the construction of the basis of Laguerre and Kautz functions is developed. As a result, that work presents some precision properties of the estimated model.

With those results, [13] developed a strategy for optimizing the expansion of the Volterra series (using Laguerre bases) applied to discrete-time systems. The main achievement of that work was the definition of an analytical solution for the optimal expansion of the Volterra kernel with the Laguerre bases, both for using individual bases, for each kernel, and a single base for all kernels. However, the results cannot be extrapolated to the Kautz functions, thus leaving the solutions for under-damped systems open.

In [14], the authors complemented their previous work by extending the results to symmetric kernels. This result is significant because it is necessary to estimate fewer parameters due to the symmetry of the kernel formed by Laguerre functions.

Still linked to the same problem (Volterra kernel expansion), [15] used Kautz functions to expand kernel and, thus, nonlinear model discrete systems. In that work, the search for optimal poles is approached using a single base for all kernels. Hence, an analytical solution for the suboptimal expansion of the Volterra kernel (with a set of Kautz functions) was developed. As a complement to their results, the same authors, in [16], developed a generalization of the previously obtained analytical model for the

case of kernel expansion with a basis of Kautz functions for each kernel dimension. Estimating a base for each kernel dimension is essential because it increases the model's representation capacity. In general terms, these expansions for the suboptimal case consist of fixing one of the pole parameters and calculating the best complementary parameter. However, depending on the value chosen for the first parameter, the result for the second parameter can lead to unsatisfactory modeling.

Another work that seeks to optimize kernel expansion with orthonormal functions (in modeling discrete systems) is [17]. The author uses GOBF with internal functions in the model kernels. From that, an analytical gradient of the model is calculated, both for symmetrical and asymmetrical kernels. Hence, the obtained values allow more accurate models without *a priori* knowledge of the system characteristics. An extension of the results, presented in [17], is discussed in [18]. The authors proposed expanding the Volterra kernel with a different GOBF base for each dimension, reducing the modeling error.

A different approach is taken in the work of [19], in which the parameters are estimated by updating a finite element model. Thus, a Sequential Quadratic Programming (SQP) optimization routine is used to estimate the Kautz poles. However, this strategy ends up not finding the optimal Kautz poles, but manages to minimize the model's prediction error. Another work that uses the SQP to estimate the Kautz poles is the [20], the results obtained are similar to the previous one, the difference is in the way in which this study excites the system and analyzes the kernels, since this study does not uses forced exit and parses the kernels separately and sums the contributions. With this, the authors are able to make the analysis simpler to compute.

Another work that presents a new treatment in the search for the Kautz poles is [21], in which a heuristic algorithm is used to determine the Kautz poles. The focus of the work is to compare the modeling of a nonlinear system with a NARX-OBF model and a Volterra-OBF model. In both models, the OBF used are the Kautz functions and the implemented heuristic algorithm is a variation of the Genetic Algorithm (GA) with the combination of two different local search operators, namely the Least Mean Squares (LMS) and Nelder Mead (NM). Through this variation, it was possible to accelerate the convergence of the algorithm search compared to its classical version. As a final result of the implementation, in this work, the NARX-OBF model excelled in relation to the Volterra-OBF in the modeling of the nonlinear system. However, as it is a heuristic optimization approach, it does not guarantee the optimal solution of the Kautz poles.

Returning to the analytical discussions, the work of [22] uses an extension of the optimal solution for the Laguerre pole to develop a method that estimates the two real parameters of the Kautz model and, from these, estimates the poles. In general, a value is fixed for the first real parameter and then optimization conditions are applied to find the second parameter. After that, a geometric average is applied to find the poles of the Kautz function. This approach gives a satisfactory approximation of the optimal Kautz poles especially when the poles are not very far from each other or are close to zero. However, it does not behave very well for when the poles are sparse.

Another new approach can be found in the work of [23], which treats Volterra kernels as a stochastic process, with the Kautz filter parameters (poles) being understood as independent random variables. In this way, the only information about these random variables is assumed to be within a region [a,b] and the principle of maximum entropy is applied to develop a probabilistic model for the poles (uniform distribution). With this, the authors obtained a model with robustness to the uncertainties and presents itself as an alternative to the more expensive deterministic models. However, neither of those works allow extending the estimation results to obtain the optimal Kautz pole.

Therefore, the selection of Kautz's optimal poles is still an open problem in the literature. In this sense, the main original contribution of this work is to present a strategy for estimating the poles using the Levenberg-Marquardt algorithm. This approach is based on the decomposition of Volterra Kernels into sets of orthonormal bases. It is parameterized by a pair of Kautz conjugated poles and associated with the kernel dynamics. The solution is based on minimizing the error resulting from the approximation of the system's response to be modeled.

The outline of this paper is as follows. Section 2 discusses the Winer-Volterra model, providing the basis for how this model is structured. Section 3 describes how orthonormal functions are formulated, detailing three types of orthonormal functions and their characteristics. In turn, Section 4 exposes how the Levenberg-Marquardt method works for optimization in the estimation of function parameters. Then, Section 5 brings the mathematical formulation of the problem addressed in this work. Next, the simulations and results achieved are detailed in Section 6, and the conclusions are presented in Section 7.

2. WIENER-VOLTERRA MODEL

The Volterra's kernels are a generalized representation of the unitary impulse response [1]. In discrete-time domain, a Volterra model represents the output of the *y(t)* system using only the *u(t)* inputs, as follows:

$$\hat{y}(k) = \sum_{\eta=1}^{\infty} \sum_{\tau_1=0}^{\infty} \dots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta) \prod_{m=1}^{\eta} u(k-\tau_m), \tag{1}$$

in which $h_\eta(\tau_1, \dots, \tau_\eta)$ is the $\eta - th$ kernel of Volterra and τ is operator delay. Although this model can describe a wide range of nonlinear systems, it needs to estimate hundreds of parameters [5], which can be considered a drawback of this method.

To circumvent this problem, Wiener developed in his work [24], a model to approximate the Volterra kernel employing orthonormal series with a truncation at a certain point, called the Wiener-Volterra model. This representation consists of linear dynamics followed by nonlinear static mapping. The strategy used in the Wiener-Volterra model basically consists of describing the h_n kernel using bases' functions $\{\phi_{n_m}\}$, as shown in (2):

© Brazilian Society on Computational Intelligence

$$h_\eta(\tau_1,...,\tau_\eta) = \sum_{n_1=1}^{\infty} ... \sum_{n_\eta=1}^{\infty} a_{n_1,...,n_\eta} \prod_{m=1}^{\eta} \phi_{n_m}(\tau_m), \tag{2}$$

in which $\alpha_{(n_\eta)}$ are scalar gains. For this kernel, it is assumed that it is absolutely summable in the range $]0,\infty)$. This ensures that the model is stable.

For the case in which the input signal $u(k)$ is limited, that is, $|u(k)| < 1, \forall k \in \mathbb{N}$, then high-order kernels can be ignored, allowing truncation in the Volterra model in an order N [25]. In addition, the number of functions that make up the bases' functions is limited to M for computational reasons. Therefore, the kernel is approximated using Eq. (3):

$$\hat{h}_\eta(\tau_1,...,\tau_\eta) = \sum_{n_1=1}^{M_{n_1}} ... \sum_{n_\eta=1}^{M_{n_\eta}} a_{n_1,...,n_\eta} \prod_{m=1}^{\eta} \phi_{n_m}(\tau_m), \tag{3}$$

as can be seen, once the orthonormal base function is defined, the kernel's estimation depends only on obtaining the scalar parameters. Due to its linear structure in the parameters, the estimation can be done using the well-known least-squares method. Thus, a system can be approximated by:

$$\hat{y}(k) = \sum_{\eta=1}^{N} \sum_{n_1=0}^{M_{n_1}} ... \sum_{n_\eta=0}^{M_{n_\eta}} a_{n_1,...,n_\eta} \prod_{m=1}^{\eta} \phi_{n_m}(\tau_m) u(k-\tau_m). \tag{4}$$

Note that the system is represented by a sum of kernels, in which the first kernel is a 1-dimensional array, the second kernel is a 2-dimensional array, and the $N-th$ kernel is an $N-dimensional$ array. From what is stated in the literature, for non-linear systems, the use of the first two kernels is enough to represent most systems [1, 15, 17]. Thus, to represent a system with the Wiener-Volterra model, it is necessary to define which orthonormal functions one should use to approximate the kernel, which is the subject of the next section.

3. ORTHONORMAL FUNCTIONS

The sequences of orthonormal functions were initially developed in the 1920s, published in the works of [26] and [27]. However, applications in systems identification emerged, with greater relevance, in the works of [7] and [8] only in the late twentieth century.

According to [28], orthonormal functions can be understood as vectors (in a specific vector space of functions) that have a null inner product (with each other) and unitary norm. This inner product is defined in [29] as being:

$$\langle \phi_p(z), \phi_q(z) \rangle = \frac{1}{2\pi i} \oint_C \phi_p(z) \phi_q^* \left(\frac{1}{z^*} \right) \frac{dz}{z}, \tag{5}$$

in which $\phi_p(z)$ and $\phi_q(z)$ are orthonormal functions, C is the circle of unit radius, i is the imaginary unit, p, $q \in \mathbb{N}$, $z \in \mathbb{C}$, and * is the conjugate operator. In its discrete form, this inner product can be represented by (6):

$$\langle \phi_p(k), \phi_q(k) \rangle = \sum_{k=0}^{\infty} \phi_p(k) \phi_q(k)^*. \tag{6}$$

In this way, the orthonormal functions belong to the Hilbert space, related to the quadratically summable functions, also called space L^2 [1]. An implication of this characteristic for the identification of linear systems, stated by [1, 30], is that: given a quadratically summable $h(k) : \mathbb{N} \to \mathbb{R}$ function in [0,∞), there will be an integer $N > 0$ which, for any error (e) greater than zero, has:

$$\sum_{k=0}^{\infty} \left(h(k) - \sum_{n=1}^{N} \alpha_n \phi_n(k) \right)^2 < e, \tag{7}$$

in which $\phi_1(k)$,...,$\phi_n(k)$ are the n orthonormal functions, and $\alpha_1, ..., \alpha_n$ are scalar and real gains. In this way, it can be shown that, for any k, Eq. (8) converges to the original function for an infinite n:

$$\hat{h}(k) = \lim_{n \to \infty} \sum_{n=0}^{N} \alpha_n \phi_n(k). \tag{8}$$

For linear dynamic systems, the representation through OBF is to develop the impulse response with a basis of orthonormal functions. This is only possible for stable Bounded Input Bounded Output (BIBO) systems because their impulse response is absolutely and, therefore, quadratically summable [1].

In their most generalized form (GOBF), the orthonormal basis functions are defined in the Z domain as:

$$\phi_n(z) = \frac{\sqrt{1-|\xi_n|^2}}{z-\xi_n} \prod_{m=1}^{n-1} \left(\frac{1-\overline{\xi}_m z}{z-\xi_m} \right) \quad n \in \mathbb{N} \tag{9}$$

in which ξ_n, $\overline{\xi}_n \in \mathbb{C}$ are the GOBF poles.

3.1 Laguerre's Functions

When the poles of the Eq. (9) are real, one can have the bases of Laguerre [31]:

$$\phi_n(z) = \frac{\sqrt{1-\xi_n^2}}{z-\xi_n}\left(\frac{1-\xi_n z}{z-\xi_n}\right)^{n-1}, \tag{10}$$

being $n \in \mathbb{N}$, and ξ_n the real pole of the base.

3.2 Kautz's Functions

Another particular case of GOBFs is when the bases are parameterized by a pair of complex conjugate poles, which results in the Kautz [11] functions. These bases are defined by two Kautz functions, called even and odd functions, and can be seen in the Eq. (11) and (12):

$$\phi_{2n}(z) = \frac{\sqrt{(1-c^2)(1-b^2)}}{z^2+b(c-1)z-c}\left(\frac{-cz^2+b(c-1)z+1}{z^2+b(c-1)z-c}\right)^{(n-1)}, \tag{11}$$

$$\phi_{2n-1}(z) = \frac{(z-b)\sqrt{1-c^2}}{z^2+b(c-1)z-c}\left(\frac{-cz^2+b(c-1)z+1}{z^2+b(c-1)z-c}\right)^{(n-1)}. \tag{12}$$

in which $|b| < 1$ and $|c| < 1$, and correspond, respectively, to the terms shown in Eq. 13 and 14:

$$b = \frac{(\xi+\overline{\xi})}{1+\xi\overline{\xi}}, \tag{13}$$

$$c = -\xi\overline{\xi}, \tag{14}$$

being $\overline{\xi}$ and ξ the complex conjugate poles that parameterize the Kautz functions, and $n \in \mathbb{N}$.

The Kautz functions can represent the dynamics of linear systems as linear combinations of the signals $w_n(k)$, as expressed by Eq. (15):

$$\hat{y}(k) = \sum_{n=1}^{N} \alpha_n w_n(k), \tag{15}$$

being $\hat{y}(k) \in \mathbb{R}$ the estimated output. In this context, one can consider $w_n(k) \in \mathbb{R}$ as the input $u(k) \in \mathbb{R}$ filtered by the Kautz functions, and one can define this filtering process by Eq. (16), being that $\alpha_n \in \mathbb{R}$ represents the gains. Hence, $w_n(k)$ can be defined as follows:

$$w_n(k) = \sum_{\tau=0}^{k} u(k-\tau)\phi_n(\tau). \tag{16}$$

In summary, if the system is super-damped and has poles close to the circle of unit radius, therefore, slow poles, it is more appropriate to use the structures of the Laguerre model. On the other hand, in the cases in which the behavior of the system is underdamped, i.e., with oscillations behaviors and with complex poles, [1] indicates that the most appropriate function for the modeling is Kautz's functions. However, for systems with an order greater than two, the GOBFs have a greater capacity for representation because they enable the combination of real and complex poles.

As seen, knowledge of the system's behavior is important when choosing the orthonormal function. It makes it possible to obtain a better cost-benefit ratio in the representation of the systems. However, defining the pole or poles that will parameterize the functions is challenging regardless of the chosen function. Although there is already a solution in the literature for the optimal case in Laguerre's case, for the Kautz functions, it is an open problem. Thus, the next section will discuss the Levenberg-Marquardt algorithm for optimizing the choice of Kautz conjugate poles.

4. LEVENBERG-MARQUARDT

The Levenberg-Marquardt method approaches the nonlinear least-squares minimization problem. Then, the evaluation function (or objective function) could be represented by Eq. (17):

$$f(x) = \frac{1}{2}\sum_{n=1}^{N} r_n^2(x), \tag{17}$$

in which r is called the residual vector.

Taking advantage of the developments of the gradient descent and Gauss-Newton methods, in [32], a strategy to guarantee the decrease of the Gauss-Newton gradient is proposed. Hence, the addition of a scalar regularization term $\lambda \geq 0$ is considered. In this way, the method is described by:

$$\delta = -\left[J_{r(x)}J_{r(x)}^{\intercal} + \lambda I\right]^{-1} J_{r(x)}r(x), \tag{18}$$

in which $J_{r(x)}$ is the Jacobian matrix of residuals and I is the identity matrix.

The λ factor is updated on each new iteration. If the approximation error decreases, λ is reduced, decreasing the influence of the gradient, and Eq. (18) tends to the Gauss-Newton method. However, if λ is too large, the gradient loses representativeness. An improvement to this is presented in [33], in which the identity matrix I is replaced by the diagonal of the gradient $J_{r(x)}J_{r(x)}^{\mathsf{T}}$, like can be observed in the Eq. 19, arriving in the Levenberg-Marquardt method:

$$\delta = -\left[J_{r(x)}J_{r(x)}^{\mathsf{T}} + \lambda \times diag\left[J_{r(x)}J_{r(x)}^{\mathsf{T}}\right]\right]^{-1} J_{r(x)}r(x). \tag{19}$$

In this way, the update of the parameter x_{k+1} in each iteration k, is given by:

$$x_{k+1} = x_k + \delta_k. \tag{20}$$

In other words, in this method, the gradient moves in large steps in directions where the curvature is small and in small steps when the curvature is accentuated.

5. PROBLEM FORMULATION

Consider a dynamic, causal, underdamped, and nonlinear system, in which the input $u(k) \in \mathbb{R}$ and the output $y(k) \in \mathbb{R}$ are known. One can model this system with a Wiener-Volterra model. Due to its underdamped behavior, the Kautz functions will be chosen to model the kernel. For the representation, one can consider that the first two kernels are satisfactory for modeling. In this way, the system can be approximated by Eq. (21):

$$\begin{aligned}\hat{y}(k) = \sum_{n=1}^{M} \alpha_n u(k-\tau_n)\phi_n(\tau_n) + \\ \sum_{n_1=1}^{M}\sum_{n_2=1}^{M} \alpha_{n_1,n_2} u(k-\tau_{n_1})\phi_{n_1}(\tau_{n_1})u(k-\tau_{n_2})\phi_{n_2}(\tau_{n_2}).\end{aligned} \tag{21}$$

Considering, for simplicity, without losing generality, that the system has a unit delay, that $\tau_n = \tau_{n_1} = \tau_{n_2} = \tau$ and

$$u(k-\tau_n)\phi_n(\tau_n) = w_n(k), \tag{22}$$

thereby,

$$\hat{y}(k) = d^{\mathsf{T}}\psi(k) + \psi(k)^{\mathsf{T}} D \psi(k). \tag{23}$$

Wherein

$$d^{\mathsf{T}} = \begin{bmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_M\end{bmatrix}, \tag{24}$$

$$D = \begin{bmatrix} \alpha_{1,1} & 0 & 0 & \dots & 0 \\ \alpha_{2,1} & \alpha_{2,2} & 0 & \dots & 0 \\ \vdots & & & & \\ \alpha_{M,1} & \alpha_{M,2} & \alpha_{M,3} & \dots & \alpha_{M,M} \end{bmatrix}, \tag{25}$$

$$\psi(k) = \begin{bmatrix} w_1(k) \\ \vdots \\ w_M(k) \\ w_{1,1}(k) \\ w_{2,1}(k) \\ w_{2,2}(k) \\ \vdots \\ w_{M,1}(k) \\ \vdots \\ w_{M,M}(k) \end{bmatrix}. \tag{26}$$

Eq. (23) can be rewritten to be linear in the parameters, as shown in Eq. (27):

$$\hat{y}(k) = \Theta^{\mathsf{T}}\psi(k), \tag{27}$$

whereupon:

$$\Theta = \begin{bmatrix}\alpha_1 & \dots & \alpha_M & \alpha_{1,1} & \alpha_{2,1} & \alpha_{2,2} & \dots & \alpha_{M,1} & \dots & \alpha_{M,M}\end{bmatrix}. \tag{28}$$

In this way, the scalar parameters α_n can be estimated using least squares, as shown in Eq. (29):

$$\Theta = [\psi^{\mathsf{T}}(k)\psi(k)]^{-1}\psi^{\mathsf{T}}(k)y(k). \tag{29}$$

The number of parameters to be estimated, considering the symmetric kernel, is

$$S_\Theta = M + \frac{M \times (M-1)}{2}. \tag{30}$$

Once the solution for the scalar parameters is found, the estimation of b and c that determine the Kautz poles remains. To do so, one can define the function $\hat{y}(b,c)$ which represents the Eq. (27) with the base functions formed with the parameters b and c.

Initially, the parameters only have the constraints exposed in Eq. (13) and Eq. (14). However, as shown in [18], to guarantee the stability of the poles, it is necessary that:

$$-1 < c < 0. \tag{31}$$

In addition, to ensure that the poles are complex conjugates, it is necessary that:

$$b^2(1-c)^2 + 4c \leq 0. \tag{32}$$

Thus, the chosen evaluation function is the one that minimizes the mean square error (MSE), given by Eq. (33):

$$\begin{aligned} \min \quad & \frac{1}{N}\sum_{n=1}^{N}\left(\hat{y}_n(b,c) - y_n\right)^2 \\ s.t. \quad & -1 < c < 0; \\ & |b| < 1; \\ & b^2(1-c)^2 + 4c \leq 0. \end{aligned} \tag{33}$$

From this formulation, it is possible to apply the Levenberg-Marquardt algorithm with constraints to estimate which parameters will optimize the modeling of the system. Fig. 1 represent the implemented algorithm, simulations and discussions will be covered in the subsequent section.

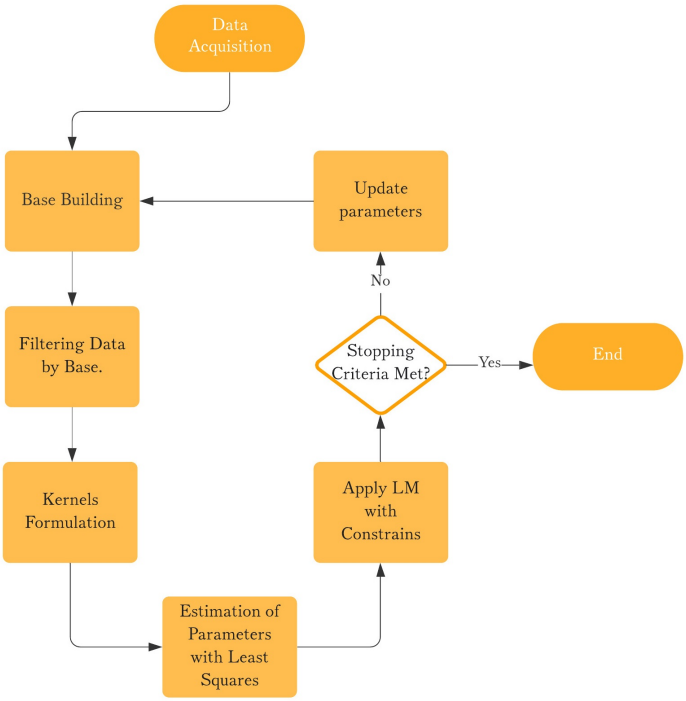

Figure 1: Implemented algorithm flowchart.

6. SIMULATION RESULTS AND DISCUSSION

In order to validate the proposed algorithm, a dataset available in [34] of a coupling system composed of two electric motors interconnected in a pulley employing a belt was used. The pulley is fixed on a spring, which allows the system to be damped. The tension and speed of the belt are controlled through the individual drive of the motors. The objective of the experiment is to collect the speed of the system. Therefore, a pulse sensor was installed on the pulley, and its signal goes through low-pass and anti-aliasing filters. Then, it is collected on a computer. Fig. 2 illustrates the system.

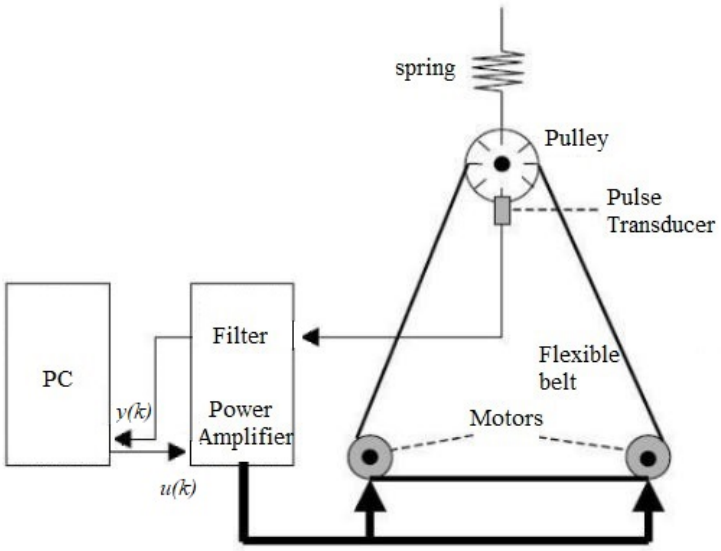

Figure 2: System used in data collection - [34].

To excite the system, a voltage input signal $u(k)$, between -1.5[V] and $+2.5$[V] was applied to the motors, which was multiplied by a random scalar uniform distributed between 0 and 1, e.g., $U(0,1)$. With that, it was possible to explore the system's nonlinearities, both in frequency and amplitude. The speed result was converted into a voltage signal $y(k)$, and the system responses were also collected. The experimentation interval was 10[s], with a 20[ms] sampling period, resulting in 500 samples. The system's input and response signal can be seen in Fig. 3 and Fig. 4.

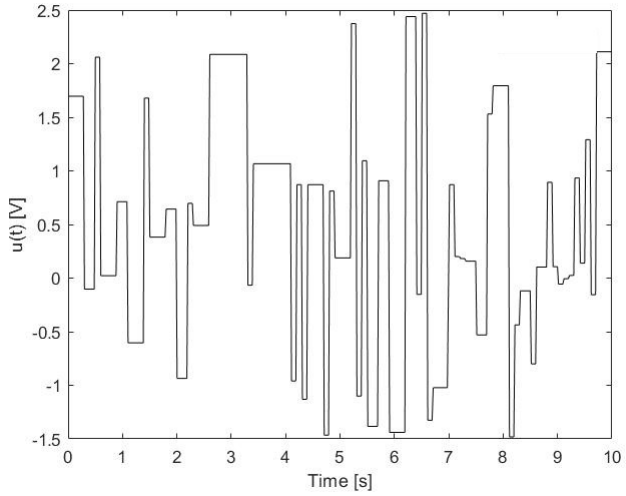

Figure 3: System excitation signal$(u(k))$.

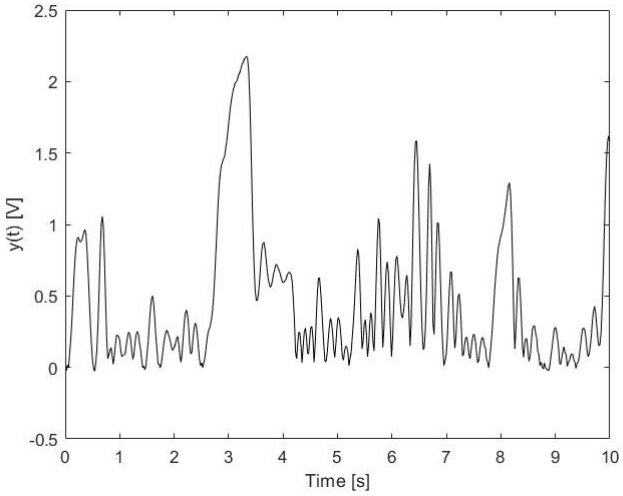

Figure 4: System response signal$(y(k))$.

In order to find out the system's model, the software Matlab was used. The Levenberg-Marquardt algorithm, the development of the Kautz base (as digital filters) were also executed in it, as shown in Fig. 5. In this representation, the $u(k)$ inputs are filtered by each even and odd Kautz function resulting in $w_n(k)$ and, in sequence, the operator $\odot$ symbolizes the static mapping that is done, as can be seen in Figure 5. This mapping is performed through the structure of the Volterra series with the first and second-order kernel. Thus, the estimated output $\hat{y}(k)$ will be modeled using the Wiener-Volterra model.

The initial parameters of the Levenberg-Marquardt algorithm are shown in Table 1. They were adjusted empirically seeking the convergence of the algorithm.

Then, the algorithm was run with 20 different starting points and with two functions on the Kautz basis. The results achieved are shown in Table 3, where b_0 and c_0 are the initial values for the parameters and b_f and c_f are the parameters found by the algorithm. The initial values already meet the three constraints for minimization.

As can be seen, only the initial points that have close values converged to poles with similar values. This is due to the Levenberg-Marquardt algorithm converging to a local minima. Additionally, regarding the MSE values, they are not expressive

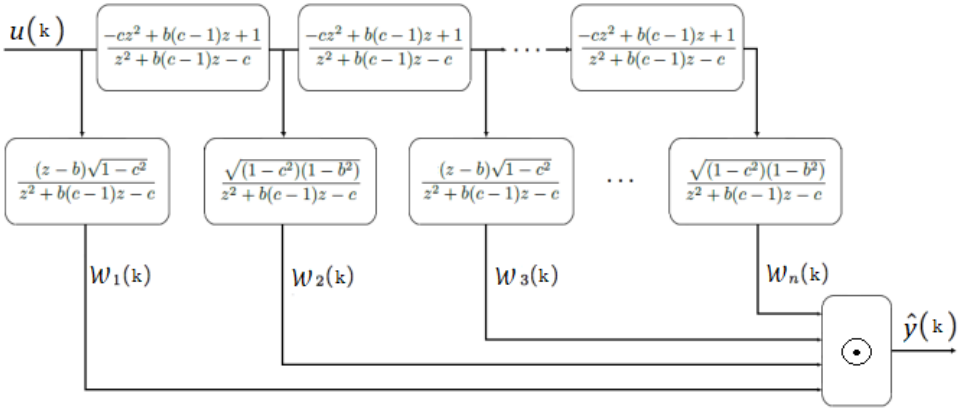

Figure 5: Kautz base implemented as a set of digital filters, and its association composing a Volterra-OBF model.

Table 1: Parameters of the Levenberg-Marquardt algorithm.

Parameters	Value
Iterations	200
Gradient Convergence Tolerance	10^{-4}
Parameter Convergence Tolerance	10^{-3}
MSE Tolerance	10^{-4}
Step Tolerance	10^{-6}
Initial Value of λ	10^{-3}
Growth Factor of λ	10
Decrement Factor of λ	10

since, in the literature, the modeling results become interesting from values in the order of magnitude of 10^{-3}.

Making use of what is expressed in Eq. (7), with the increase in the number of functions that make up the base, the approximation error tends to decrease. Therefore, a sequence of tests was carried out. The number of functions in the base was increased from 2 to 20, with the same poles already found by Levenberg-Marquardt. The results are shown in Fig. 6 and the MSE values for 20 functions in the base are shown in Table 4.

Comparing the values displayed in Tables 3 and 4, it is possible to notice the improvement of all the MSE values of the experiments. It can be observed that in some cases, such as in tests 4, 6, 10, 19, and 20, the modeling could not reach MSE below 10^{-2}, this fact highlights the importance of the initial parameters for the algorithm of Levenberg-Marquardt, that got stuck in an unsatisfactory local minimum, leading to unrepresentative modeling of the system.

For a more in-depth analysis, data from test 7 was used. In Fig. 7 and 8 the results of modeling with 2 and 20 functions, respectively, are illustrated.

As can be seen, the modeling with two functions had difficulty to model the oscillations in the intervals of 1s to 3s, and 5s to 7s. With the use of more functions, the representation capacity was increased, and these problems were solved. However, visually, it is possible to notice that for the representations of the peaks of the consecutive oscillations, as occurs in the range of 4s to 4.5s, and 7.2s to 8s, even with 20 Kautz functions, the model presented a deviation in relation to the collected data. Despite improving the modeling capacity, the increase in the number of functions in the base also implies an increase in the number of parameters to be estimated. That is, there will always be a trade-off between the number of parameters to be estimated and the quality of the modeling. Table 2 shows the relation between the number of parameters as a function of the number of functions in the base.

Another behavior that can be analyzed is the Levenberg-Marquardt algorithm's performance in searching for optimal parameters. In Fig. 9 the behavior of the MSE during the search is seen. Note that the algorithm did not execute the 200 iterations, with the criterion of convergence of the parameter being reached leading to the end of the execution of the algorithm.

When contrasting the results of Fig. 9 with Fig. 10, which represents the behavior of λ during the search, it can be noted that during the first ten iterations, the MSE decreases quickly. Consequently, the λ tends to decrease, leading the algorithm to behave similarly to Gauss-Newton. In later iterations, the variation of the MSE is not very expressive (decreasing from 0.065 to 0.054), which leads to an increase in the value of λ. In this way, the algorithm walked with small steps when the curvature of the MSE was accentuated and with large steps when the MSE had a slight variation.

It is also possible to analyze how the parameters behaved inside the search space. Fig. 11 illustrates this behavior. As can be seen, for the particular case of test 7, the parameters did not have a great variation from the initial value. This is due

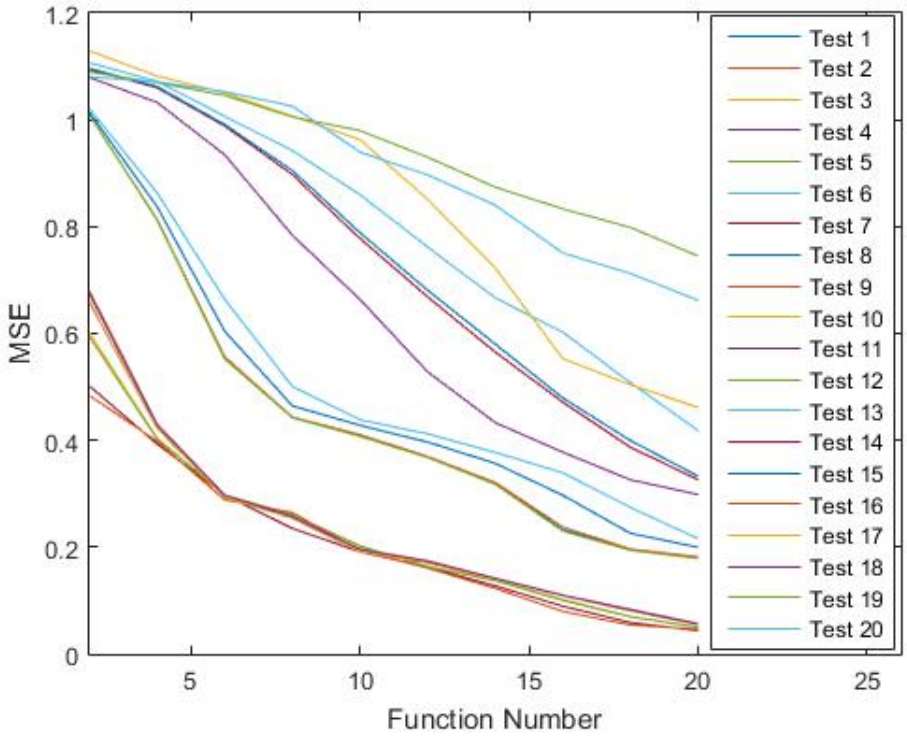

Figure 6: MSE's behavior with increasing the number of orthonormal functions used in the Volterra-OBF model.

Table 2: Number of parameters to be estimated.

Base Function	Parameters
2	3
4	10
6	21
8	36
10	55
12	78
14	105
16	258
18	171
20	210

Table 3: Algorithm results.

Test	b_0	c_0	b_f	c_f	Polo	MSE
1	0.100	-0.500	0.900	-0.390	$0.620 \pm i0.005$	0.095
2	-0.300	-0.100	0.900	-0.400	$0.630 \pm i0.007$	0.094
3	0.400	-0.200	0.920	-0.450	$0.670 \pm i0.022$	0.078
4	0.500	-0.100	-0.490	-0.070	$-0.260 \pm i0.013$	0.250
5	0.200	-0.400	-0.900	-0.400	$0.630 \pm i0.018$	0.094
6	-0.900	-0.600	-0.690	-0.160	$-0.400 \pm i0.003$	0.258
7	0.900	-0.700	-0.960	-0.550	$0.740 \pm i0.032$	0.054
8	0.800	-0.900	0.100	-0.003	$0.050 \pm i0.018$	0.220
9	0.700	-0.700	0.900	-0.390	$0.620 \pm i0.037$	0.099
10	-0.600	-0.800	-0.490	-0.880	$-0.450 \pm i0.822$	0.274
11	0.250	-0.100	0.900	-0.400	$0.630 \pm i0.021$	0.093
12	0.500	-0.300	0.920	-0.450	$0.670 \pm i0.039$	0.077
13	-0.400	-0.200	0.900	-0.400	$0.630 \pm i0.032$	0.095
14	-0.600	-0.300	-0.680	-0.150	$-0.390 \pm i0.012$	0.257
15	-0.700	-0.500	-0.690	-0.160	$-0.400 \pm i0.004$	0.258
16	0.900	-0.800	0.970	-0.630	$0.790 \pm i0.058$	0.050
17	0.600	-0.500	0.900	-0.400	$0.630 \pm i0.008$	0.094
18	0.800	-0.700	0.900	-0.390	$0.620 \pm i0.010$	0.098
19	-0.400	-0.300	-0.004	-0.680	$-0.003 \pm i0.824$	0.255
20	-0.300	-0.700	-0.100	-0.910	$-0.090 \pm i0.950$	0.251

Table 4: MSE results to 20 functions in base.

Test	Polo	MSE
1	$0.620 \pm i0.005$	6.57×10^{-4}
2	$0.630 \pm i0.007$	6.50×10^{-4}
3	$0.670 \pm i0.022$	5.50×10^{-4}
4	$-0.260 \pm i0.013$	1.91×10^{-2}
5	$0.630 \pm i0.018$	6.50×10^{-4}
6	$-0.400 \pm i0.003$	2.42×10^{-2}
7	$0.740 \pm i0.032$	4.37×10^{-4}
8	$0.050 \pm i0.018$	8.60×10^{-3}
9	$0.620 \pm i0.037$	6.88×10^{-4}
10	$-0.450 \pm i0.822$	4.57×10^{-2}
11	$0.630 \pm i0.021$	6.50×10^{-4}
12	$0.670 \pm i0.039$	5.28×10^{-4}
13	$0.630 \pm i0.032$	6.54×10^{-4}
14	$-0.390 \pm i0.012$	2.31×10^{-2}
15	$-0.400 \pm i0.004$	2.38×10^{-2}
16	$0.790 \pm i0.058$	4.99×10^{-4}
17	$0.630 \pm i0.008$	6.55×10^{-4}
18	$0.620 \pm i0.010$	6.73×10^{-4}
19	$-0.003 \pm i0.824$	1.19×10^{-1}
20	$-0.090 \pm i0.950$	9.41×10^{-2}

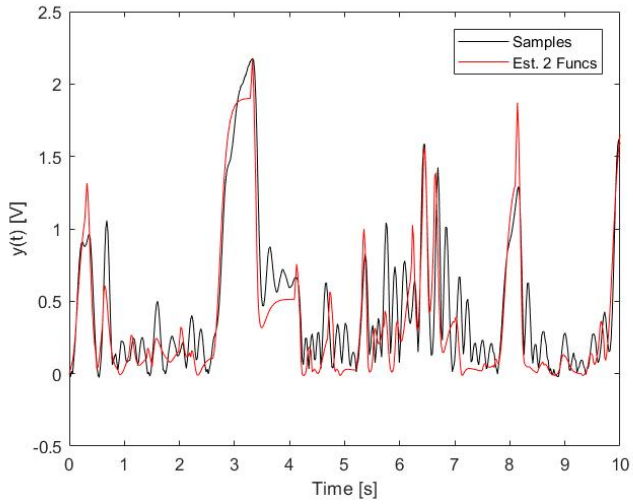

Figure 7: System Identification with 2 functions.

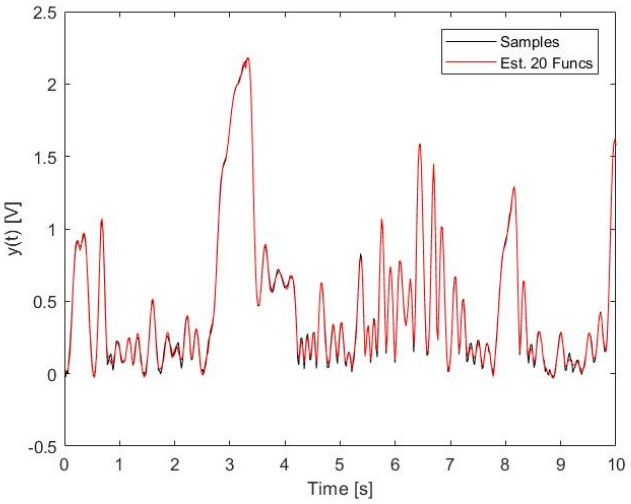

Figure 8: System Identification with 20 functions.

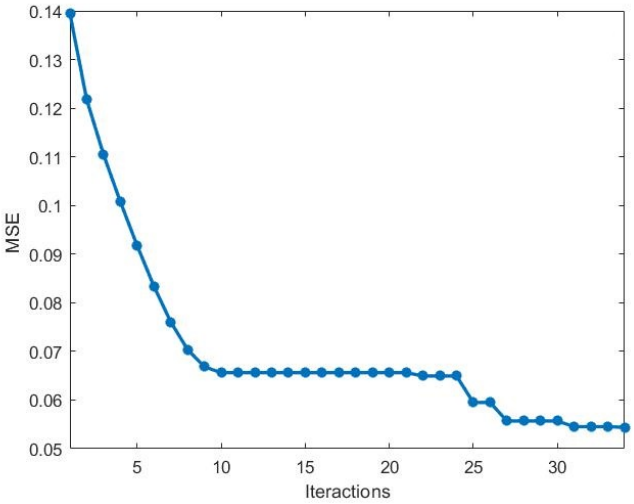

Figure 9: MSE's behavior during Levenberg-Marquardt seeking.

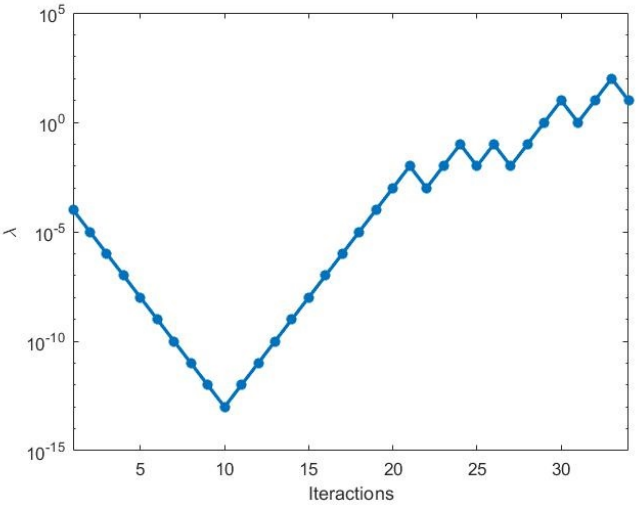

Figure 10: λ's behavior during Levenberg-Marquardt seeking.

to the constraints that were imposed on the algorithm. As shown in Fig. 11, the algorithm optimized the parameters without disrespecting any of the three restrictions. It is also noticeable that the parameters were optimized to the maximum point of the second restriction, which was already inside the search space of the first restriction.

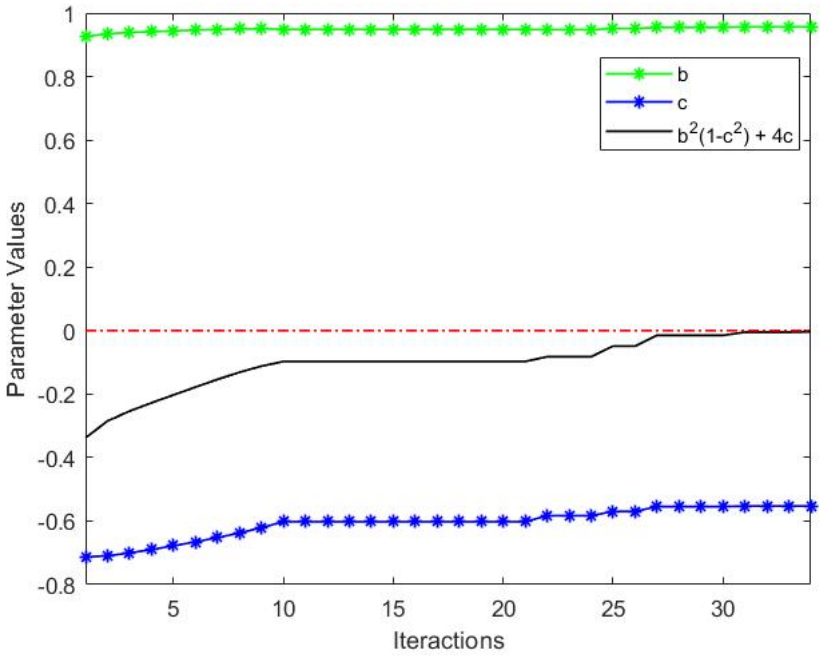

Figure 11: Parameter's behavior during Levenberg-Marquardt seeking.

To compare the results obtained with the presented approach, Table 5 illustrates the results achieved by different modeling methodologies found in the literature and which used the same dataset provided by [34]. In [35], the authors carried out the system identification based on Radial Basis Functions Neural Networks (RBFNNs) models and Free Search Differential Evolution (FSDE) with cascaded evolutionary algorithm paradigm. The best results were achieved with the Lipschitz coefficient for input selection. As in [36], the authors used a nonlinear autoregressive exogenous input (NARX) model for learning a conditional distribution of output based on a finite window of past inputs and outputs, as a tool for this approach was used energy-based (EB) models. The main result of this work was to demonstrate that the approach used is capable of learning both simple and highly complex noise models.

Table 5: Comparison of the MSE among different recent works in the literature.

Approach	MSE
RBFNNs	1.60×10^{-3}
EB-NARX	5.03×10^{-2}
MOODE-NARMAX	1.35×10^{-2}
Present Work	4.37×10^{-4}

In turn, the work of [37] applied a nonlinear autoregressive moving average with exogenous input (NARMAX) structure modified through the use of a multi-objective optimization differential evolution (MOODE) to select the structure. As a main result, the work concluded that the proposed NARMAX presents better output predictions than conventional NARMAX. However, for the three works found, the methodology proposed in this work achieved more promising results, in the order of 10^{-4}.

7. CONCLUSION

The modeling of dynamical systems with the Volterra series is recurrent in the literature, as this series enables the description of a wide range of nonlinear models. However, the use of this series implies the estimation of numerous parameters, mainly for nonlinear systems. A strategy present in the literature is using orthonormal basis functions, expanding the Volterra Kernel, resulting in a significant reduction in terms to be estimated. The orthonormal functions used in this work were Kautz's functions, the main challenge being to find which poles best parameterize the functions. For this, the use of the Levenber-Marquardt algorithm and a structure of digital filters for structuring the function base were proposed.

In order to validate the proposed methodology, data from motors system coupled with a belt were used, and the oscillation speed of a fixed pulley with a spring was measured. With the implemented algorithm, it was noticed that the behavior of the algorithm is strongly linked to the initial values, as can be seen in Tables 3 and 4. Due to this behavior, 20 points were tested to find results in the order of magnitude of 10^{-3}. Another factor that impacts the representation result is the number of functions in

the base, since with the increase the number of orthonormal functions, the representation capacity increases, and the identification result improves, decreasing the MSE. Nonetheless, it is important to emphasize that increasing the number of functions, increases the number of parameters to be estimated. In the experiments performed, by increasing up to 20 functions in the base, several experiments achieved a modeling with an MSE result in the order of 10^{-4}, as in the cases of tests 1, 2, 3 and 7. These results suggest that the strategy proposed in this work is an alternative to solve the Kautz optimal pole estimation problem and managed to model the same problem with more promising results, in terms of MSE, than other system identification methodologies present in the literature.

In future works, another strategy that can be approached is the use of metaheuristic optimization algorithms such as Ant Colony Optmization and Particle Swarm Optmization, as they do not have a starting point dependency, which can improve the way the global search of b and c parameters is performed.

REFERENCES

[1] R. J. Campello, G. H. Oliveira and W. C. Amaral. "Identificação e controle de processos via desenvolvimentos em séries ortonormais. Parte A: Identificação". *Sba: Controle & Automaçao Sociedade Brasileira de Automatica*, vol. 18, no. 3, pp. 301–321, 2007.

[2] L. Ljung. *System Identiflcation: Theory for the User*. Prentice Hall, 1999.

[3] M. Deistler. *Stochastic Theory and Control*. Springer, 2002.

[4] M. V. Corrêa and L. A. Aguirre. "IDENTIFICAÇÃO NÃO-LINEAR CAIXA-CINZA: UMA REVISÃO E NOVOS RESULTADOS". *Controle & Automação*, 2004.

[5] R. Isermann and M. Münchhof. *Identification of Dynamic Systems*. Springer, 2011.

[6] P. S. Heuberger, P. M. van den Hof and B. Wahlberg. *Modelling and identification with rational orthogonal basis functions*. Springer Science & Business Media, 2005.

[7] B. Wahlberg. "System Identification Using Laguerre Models." *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, 1991.

[8] B. Ninness and F. Gustafsson. "A unifying construction of orthonormal bases for system identification". *IEEE Trans. Automat. Control*, 1994.

[9] P. S. Heuberger, T. J. de Hoog, P. M. Van den Hof and B. Wahlberg. "Orthonormal basis functions in time and frequency domain: Hambo transform theory". *SIAM Journal on Control and Optimization*, vol. 42, no. 4, pp. 1347–1373, 2003.

[10] W. Bo and P. Mäkila. "On Approximation of Stable Linear Dynamical Systems using Laguerre and Kautz Functions". In *Automatica*, 1996.

[11] B. Wahlberg. "System Identification Using Kautz Model". In *IEEE TRANSACTIONS AUTOMATIC CONTROL*, 1994.

[12] B. Ninness and F. Gustafsson. "A unifying construction of orthonormal bases for system identification". *IEEE Transactions on automatic control*, vol. 42, no. 4, pp. 515–521, 1997.

[13] R. J. Campello, G. Favier and W. C. Do Amaral. "Optimal expansions of discrete-time Volterra models using Laguerre functions". *Automatica*, vol. 40, no. 5, pp. 815–822, 2004.

[14] R. J. Campello, W. C. Do Amaral and G. Favier. "A note on the optimal expansion of Volterra models using Laguerre functions". *Automatica*, vol. 42, no. 4, pp. 689–693, 2006.

[15] A. Da Rosa, R. J. Campello and W. C. Amaral. "Choice of free parameters in expansions of discrete-time Volterra models using Kautz functions". *Automatica*, vol. 43, no. 6, pp. 1084–1091, 2007.

[16] A. Da Rosa, R. J. Campello and W. C. Amaral. "An optimal expansion of Volterra models using independent Kautz bases for each kernel dimension". *International Journal of Control*, vol. 81, no. 6, pp. 962–975, 2008.

[17] J. B. Machado *et al.*. "Modelagem de sistemas não-lineares por base de funções ortonormais generalizadas com funções internas". 2011.

[18] M. F. Braga *et al.*. "Modelos de Volterra: identificação não paramétrica e robusta utilizando funções ortonormais de Kautz e generalizadas". 2011.

[19] S. B. Shiki, J.-P. Noël, G. Kerschen, V. Lopes Junior and S. da Silva. "Identification of mechanical systems with local nonlinearities through discrete-time Volterra series and Kautz functions". In *11th International Conference on Recent Advances in Structural Dynamics*, 2013.

[20] O. Scussel and S. Da Silva. "Output-only identification of nonlinear systems via volterra series". *Journal of Vibration and Acoustics*, vol. 138, no. 4, 2016.

[21] E. Oroski, R. H. Lopez and A. Bauchspiess. "Identification of a magnetic levitator using NARX-OBF models and genetic algorithm." *Int. J. Model. Identif. Control.*, vol. 28, no. 4, pp. 307–316, 2017.

[22] R. A. L. de Andrade, P. R. Barros and R. B. C. Lima. "Identification of second-order Kautz models by two-step pole location optimisation". In *2019 18th European Control Conference (ECC)*, pp. 836–841. IEEE, 2019.

[23] L. G. G. Villani, S. da Silva and A. Cunha. "An optimizationless stochastic volterra series approach for nonlinear model identification". *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 44, no. 6, pp. 1–8, 2022.

[24] N. Wiener. "Nonlinear problems in random theory". *Nonlinear Problems in Random Theory, by Norbert Wiener, pp. 142. ISBN 0-262-73012-X. Cambridge, Massachusetts, USA: The MIT Press, August 1966.(Paper)*, p. 142, 1966.

[25] P. Eykhoff. "System identification: parameter and state estimation". 1974.

[26] S. Takenaka. "On the orthogonal functions and a new formula of interpolation,". *J. Math.*, 1925.

[27] F. Malmquist. "Sur la determination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de points". *Comptes Rendus du Sixième Congrès des Mathématiciens Scandinaves*, 1925.

[28] G. Strang. *Álgebra Linear e suas Aplicações*. Cengage Learning, 2010.

[29] B. Wahlberg. "ORTHONORMAL BASIS FUNCTION MODELS: A TRANSFORMATION ANALYSIS". In *14th Triennial World Congress*, 1999.

[30] D. Lemma, M. Ramasamy and M. Shuhaimi. "System identification using orthonormal basis filters". *Journal of Applied Sciences(Faisalabad)*, vol. 10, no. 21, pp. 2516–2522, 2010.

[31] Y. Fu and G. Dumont. "An Optimum Time Scale for Discrete Laguerre Network". In *IEEE TRANSACTIONS AUTOMATIC CONTROL*, 1993.

[32] K. Levenberg. "A method for the solution of certain problems in least squares". *Quart. Appl. Math. Vol 2*, 1944.

[33] D. Marquardt. "An algorithm for least-squares estimation of nonlinear parameters". *SIAM J. Appl. Math Vol 1*, 1963.

[34] T. Wigren and M. Schoukens. *Coupled electric drives data set and reference models*. Department of Information Technology, Uppsala Universitet, 2017.

[35] H. V. H. Ayala, L. F. da Cruz, R. Z. Freire and L. dos Santos Coelho. "Cascaded free search differential evolution applied to nonlinear system identification based on correlation functions and neural networks". In *2014 IEEE Symposium on Computational Intelligence in Control and Automation (CICA)*, pp. 1–7. IEEE, 2014.

[36] J. N. Hendriks, F. K. Gustafsson, A. H. Ribeiro, A. G. Wills and T. B. Schön. "Deep energy-based NARX models". *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 505–510, 2021.

[37] M. Z. Zakaria, Z. Mansor, A. M. Nor, M. E. Baharudin and M. S. Saad. "Modeling coupled electric drives systems using a modified NARMAX model". In *AIP Conference Proceedings*, volume 2339, p. 020015. AIP Publishing LLC, 2021.