Article:
Li, Y., Zhang, J., Cui, W. et al. (2 more authors) (2019) A multiple beta wavelet-based locally regularized ultraorthogonal forward regression algorithm for time-varying system identification with applications to EEG. IEEE Transactions on Instrumentation and Measurement. ISSN 0018-9456

https://doi.org/10.1109/TIM.2019.2907036

A Multiple Beta Wavelet-based Locally Regularized Ultra-Orthogonal Forward Regression Algorithm for Time-Varying System Identification with Applications to EEG

Yang Li, Jing-Bo Zhang, Wei-Gang Cui, Heng Yuan, and Hua-Liang Wei

Abstract—Time-varying nonlinear systems widely exist in various fields of engineering and science. Effective identification and modeling of time-varying systems is a challenging problem due to the nonstationarity and nonlinearity of the associated processes. In this paper, a novel parametric modeling algorithm is proposed to deal with this problem based on a time-varying nonlinear autoregressive with exogenous input (TV-NARX) model. A new class of multiple beta wavelet (MBW) basis functions is introduced to represent the time-varying coefficients of the TV-NARX model to enable the tracking of both smooth trends and sharp changes of the system behavior. To produce a parsimonious model structure, a locally regularized ultra-orthogonal forward regression (LRUOFR) algorithm aided by the adjustable prediction error sum of squares (APRESS) criterion is investigated for sparse model term selection and parameter estimation. Simulation studies and a real application to EEG data show that the proposed MBW-LRUOFR algorithm can effectively capture the global and local features of nonstationary systems and obtain an optimal model, even for signals contaminated with severe color noise.

Index Terms—EEG, locally regularized ultra-orthogonal forward regression (LRUOFR), multiple beta wavelet (MBW), parametric estimation, system identification.

I. Introduction

MOST processes in nature including biomedical signals exhibit nonstationary properties where numerous transient components are associated with the underlying psychological activities. Identification of nonstationary systems is a challenging problem and has been attracting widespread attention [1-3]. One common strategy to characterize such nonstationary processes is to establish a time-varying nonlinear autoregressive with exogenous input (TV-NARX) model [4]. The wide application and popularity of this model mainly stems from its easy-to-compute parameters [5].

Many approaches have been proposed to identify TV-NARX models, which can be broadly classified into three categories: multi-model approach [6], adaptive estimation algorithm [7], and basis function expansion method [8, 9]. In the first strategy, a global system model is divided into a set of local models by a time shifting window, then the local model can be treated as a stationary process and identified by a time-invariant modeling approach [10]. However, many nonstationary signals, e.g. EEG, cannot simply be partitioned into stationary time series since it is difficult to determine the size of the window. For example, if the window is too large then it is not appropriate to treat the segments to be stationary; if, however, the window is too small, the segments turn out to be too short that the estimates may be unreliable. In the second strategy, the TV coefficients of the model are considered as random processes with certain stochastic model structure [11, 12]. The main limitation of this scheme is the possible tracking lag presented in the estimated parameters due to the slow convergence rate, which makes these approaches inaccurate for tracking abrupt changes of the underlying signals [13, 14]. Recently, the third strategy combining basis function expansion with linear regression approaches has been proposed to identify nonlinear TV systems, where TV parameters are approximated by a set of predefined basis functions [14, 15]. In this way, the unknown TV parameters can be converted into a set of constant coefficients of the basis functions [16]. Specifically, the implementation of this strategy can be briefly described in two steps: step 1, a basis function expansion approach is used to transform the original TV model to a time-invariant regression problem [8]; step 2, a model structure selection algorithm, such as the classical orthogonal forward regression (OFR) algorithm [17] or its variants [18, 19], is applied to obtain a parsimonious

This work was supported in part by the National Natural Science Foundation of China under Grant 61403016, Grant 61671042 and Grant 61773039, in part by the Beijing Natural Science Foundation under Grant 4172037, in part by the Open Fund Project in Minjiang University under Grant MJUKF201702. (Corresponding authors: Heng Yuan; Hua-Liang Wei.)

Y. Li is with the Department of Automation Sciences and Electrical Engineering, Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing Advanced Innovation Center for Big Date-based Precision Medicine, Beihang University, 100191, China (e-mail: liyang@buaa.edu.cn).

J. B. Zhang and W. G. Cui are with the Department of Automation Sciences and Electrical Engineering, Beihang University, 100191, China (e-mail: eckert@buaa.edu.cn; cuiweigang94@foxmail.com).

H.Yuan is with the School of Instrumentation Science and Opto-electronics Engineering, Beihang University, 100191, China (e-mail: hengyuan@buaa.edu.cn).

Hua-Liang Wei is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3DJ, U.K (e-mail: w.hualiang@sheffield.ac.uk).

model which includes a relatively small number of regression terms.

In the first step, an over complete set of basis functions, with good presentation properties, is employed to approximate the TV coefficients. Hence, an appropriate selection of the basis functions is critical to guarantee the performance of the identified model if we want the model to be sparse [16]. For instance, numerical experiments showed that the Legendre polynomials are efficient for smoothly or slowly changing parameters, and Walsh functions generally work well for piecewise stationary TV parameters [20]. For a system with sharply or rapidly changing parameters, Li et al. introduced multi-wavelets formed by cardinal B-splines to approximate the TV coefficients, which has been verified in simulations and real biological signals [19, 21]. Although the multi-wavelets can be an appropriate choice in the expansion process, the simple waveform structure and few variants of cardinal B-splines limit its ability to capture local information of TV signals [22]. To overcome this limitation, a novel class of basis functions formed by multiple beta wavelets (MBWs) is proposed in this paper, where the beta wavelet is a compactly supported one-cyclic wavelet introduced in the work of Oliveira et al. [23]. Beta wavelets have been widely used in some fields due to the excellent flexibility and good approximation characteristics, such as image processing and signal compression [24, 25]. However, to the best of our knowledge, not much work has been done in the existing literature on exploiting the attractive properties of beta wavelets and applying them to TV nonlinear system identification. Particularly, considering that the beta wavelet has a waveform similar to a neural pulse signal and possesses various variations controlled by two characteristic parameters [23], this study will explore its power in capturing the local information of the abrupt positions of TV coefficients.

The main tasks of the second step in system identification are model structure detection and model reduction, which aims to remove redundant regression terms and produce a parsimonious model structure. Although the OFR algorithm is effective and commonly applied in the process of system identification, the determination of the optimal model structure is still a challenging work when the system is not persistently excited or data are severely contaminated by noise [26]. To improve the performance and accuracy of resulting model structure, Li et al. employed the advanced ultra-OFR (UOFR) algorithm to identify significant regressors and find a more accurate model compared to the classical OFR algorithm [19]. The UOFR algorithm detects the correlation among the data points of time series, and determines the model structure by using the hidden information that is not fully explored by the traditional least squares type algorithms. However, the UOFR method ignores the interference of overlapping information among candidate regressors, resulting in the inclusion of spurious or redundant model terms in some cases. In order to further improve the performance of the UOFR approach for dealing with overlapping information in signals, this paper introduces a locally regularized UOFR (LRUOFR) algorithm for system identification, which assigns an individual regularization parameter to each candidate term and iteratively updates the parameters to achieve optimal estimates [27, 28]. In fact, LRUOFR not only considers the interconnections among the sample points of the signals [19], but also evaluates an individual influence of each candidate regressor in the OFR process [28]. As illustrated in the example presented in section III-A, the proposed LRUOFR approach takes into consideration more regressor information and is capable of selecting significant terms in the model identification process.

In this paper, a novel MBW-LRUOFR algorithm is proposed for the identification of TV-NARX model, where a finite number of predefined MBW basis functions are used to approximate the TV coefficients, and the model structure is determined by using the LRUOFR algorithm together with the adjustable prediction error sum of squares (APRESS) criterion [26, 28, 29]. The MBW basis functions are locally linearly independent and have many variations [23], which is capable of providing a powerful tool for representing TV signals. The local regularization-based method has been proven to enhance the sparsity of the resulting model and effectively avoid numerical ill-conditioned problems during the selection of significant terms [28]. With the incorporation of the APRESS cross-validation criterion, the model size (i.e., the number of model terms to be included in the final model) can be well determined [29]. One of the main contributions of this study is that for the first time, the MBW basis function is adopted to approximate TV coefficients; it adds an effective choice to the existing basis function expansion approach and thus enhances the capability of the existing approach to model and track rapid changing signals. The main advantage of the proposed MBW-LRUOFR algorithm is that it is more efficient to select significant model terms under the condition that data are not persistent or highly noisy. In order to illustrate the effectiveness of this method for tracking time-varying signals, the identification performance is compared to other three methodologies: the classical recursive least squares (RLS) algorithm [30], the Bspline-UOFR method [19] and the MBW-UOFR algorithm. Simulation and application results have shown the effective identification performance of the proposed method for nonstationary systems and further illustrated that the new proposed framework is capable of tracking time-varying signals.

The remainder of this paper is organized as follows. In section II, the identification methodology is introduced. More specifically, section II-A describes the construction process of a TV-NARX model; section II-B introduces the properties of beta wavelets and the implementation of MBW basis function expansion method; section II-C elaborates the theoretical framework of the LRUOFR algorithm with the APRESS cross-validation criterion. In section III, three numerical simulations are given to illustrate the effectiveness of the proposed method. In section IV, an application based on EEG signals is implemented to verify the practicality of the proposed scheme for solving real data modelling problems. Finally, the conclusion of this paper is shown in section V.

II. METHODOLOGY

A. The Time-Varying NARX Model

A wide class of input-output nonstationary systems can be represented by a nonlinear autoregressive with exogenous inputs (NARX) model [31], which can be expressed by:

$$y(t)=f\left(y(t-1),\cdots,y(t-n_y),u(t-1),\cdots,u(t-n_u)\right)+e(t) \quad (1)$$

where $y(t)$ and $u(t)$ denote the output, input sequences, with maximum lags n_y and n_u, respectively; $f(\cdot)$ is a nonlinear function characterizing the input and output relationship; $e(t)$ denotes an error term (noise, residual, etc.) which is assumed to be independent, bounded and uncorrelated with the input $u(t)$.

The unknown nonlinear function f can be expressed in various types of model structures, such as fuzzy logic-based models, rational models, and neural networks. The most common expression is the polynomial regressions, which has been used for a wide range of nonlinear systems. The NARX model can be further expressed in a linear-in-the-parameters form [32]:

$$y(t)=\boldsymbol{\varphi}^{T}(t)\boldsymbol{\theta}+e(t) \quad (2)$$

where $\boldsymbol{\varphi}(t)$ is the regression vector which contains monomials of lagged output and input terms; $\boldsymbol{\theta}$ is the associated parameter vector, and $e(t)$ is a zero mean noise sequence.

When modelling a TV system, the parameter vector $\boldsymbol{\theta}$ in the NARX model can be replaced with a TV parameter vector $\boldsymbol{\theta}(t)$ to obtain a polynomial TV-NARX model:

$$\begin{aligned} y(t) &= \sum_{g=1}^{G} y_g(t)+e(t) \\ &= \sum_{g=1}^{G}\sum_{p=0}^{g}\sum_{k_1,k_{p+q}=1}^{K}\zeta_{p,q}(k_1,\cdots,k_{p+q},t) \\ &\quad\times\prod_{i=1}^{p}y(t-k_i)\prod_{i=p+1}^{p+q}u(t-k_i)+e(t) \\ &= \boldsymbol{\varphi}^{T}(t)\boldsymbol{\theta}(t)+e(t) \end{aligned} \quad (3)$$

where G is the degree of the nonlinearity; p and q are the numbers of output and input terms, respectively, which satisfy $p+q=g$; $\sum_{k_1,k_{p+q}}^{K}\equiv\sum_{k_1=1}^{K}\cdots\sum_{k_{p+q}=1}^{K}$ is a simple representation of multiple sums, with $k_i=1,\cdots,K$; $\boldsymbol{\theta}(t)$ indicates the TV parameter vector which can be expressed as $\boldsymbol{\theta}(t)=\left[\zeta_{0,1}(1,t),\cdots,\zeta_{0,1}(K,t),\zeta_{1,0}(1,t),\cdots,\zeta_{p,q}(K,\cdots,K,t)\right]^{T}$.

The model (3) may consist of a large number of candidate terms and the number depends on the degree (G), the order of terms (p and q), and the corresponding maximum lag (K). However, not all candidate terms are significant in general, those that are redundant or make no or little contribution can be removed from the initial model. The identification process of model (3) includes two main tasks: the selection of significant terms from a pre-specified candidate term dictionary and the estimation of corresponding parameters. However, the standard sparse model identification algorithm, such as the OFR algorithm [17, 18, 31] and principal component analysis (PCA) [33], cannot directly identify a TV model due to the assumption that the individual model parameters are constants.

In order to effectively estimate the change of TV parameters, an effective identification procedure, which makes use of a new class of MBW basis functions, is introduced in this paper. The basic idea is that each of the time-varying coefficient is approximated by using the MBW basis functions, in this way, the identification of TV model is converted to a time-invariant regression model problem which can be solved by means of a conventional model structure detection algorithm, such as the OFR algorithm or its variants.

B. Multiple Beta Wavelet Basis Functions

From the work of [23], a new continuous beta wavelet is derived from the beta distribution by using 'blur' derivatives, which is defined as:

$$\mathcal{B}_{\alpha,\beta}(t)=\frac{(t-a)^{\alpha-1}\cdot(b-t)^{\beta-1}}{\mu(\alpha,\beta)L^{\alpha+\beta-1}}\cdot\left(\frac{\alpha-1}{t-a}-\frac{\beta-1}{b-t}\right) \quad (4)$$

where $\mu(\alpha,\beta)=\Gamma(\alpha)\cdot\Gamma(\beta)/\Gamma(\alpha+\beta)$ is the normalizing factor of beta distribution, and $\Gamma(\cdot)$ denotes the generalized factorial function of Euler; $[a,b]=\left[-\sqrt{\alpha+\beta+1}/\sqrt{\beta/\alpha},\sqrt{\alpha+\beta+1}/\sqrt{\alpha/\beta}\right]$ is the support set of beta wavelet function; $L=b-a$ is the length of the support set; $\alpha\geq 2$ and $\beta\geq 2$ are the characteristic parameters of the function.

Beta wavelets generated by the function (4) have only one-cycle which includes a positive half-cycle and a negative half-cycle. In a sense, the waveforms of beta wavelets are similar to the neural active shapes, which give them good approximation characteristics and make them play a crucial role in the adaptive capacity of capturing the nonstationary signals [24]. The property of beta wavelets is determined by parameters α and β. For example, if $\alpha=\beta$, the wavelets are asymmetrical, and if $\alpha\neq\beta$, the wavelets are non-symmetrical. An example of the waveform with different parameters can be clearly observed in Fig. 1. Note that a bell-shaped half-cycle and a smooth half-cycle appear due to the difference between α and β. The wavelets with a narrow bell-shaped half-cycle perform well on the sharp or abrupt change of signals, while the wavelets with a wide bell-shaped half-cycle or a smooth half-cycle tend to track slow changes of signals [5]. Different variants allow the capability to capture the overall and local information of TV coefficients; and the combination of multiple variants can effectively identify complex nonstationary systems. Another attractive feature of beta wavelets is the great properties of complete support, regularity, and orthogonality [34], which enable the operation of the multiresolution decomposition to be much more convenient.

From the wavelet theory [21], a square integrable scalar function $h\in L^2(\square)$ can be arbitrarily approximated using the multi-resolution wavelet decomposition below:

$$h(x)=\sum_{l} c_{j_0,l}\phi_{j_0,l}(x)+\sum_{j\geq j_0}\sum_{l} d_{j,l}\psi_{j,l}(x) \quad (5)$$

where $\psi_{j,l}(x)=2^{j/2}\psi(2^j x-l)$ and $\phi_{j,l}(x)=2^{j/2}\phi(2^j x-l)$, with $j,l\in\mathbb{Z}$ ($\mathbb{Z}$ is a set consisting of whole integers), are the dilated

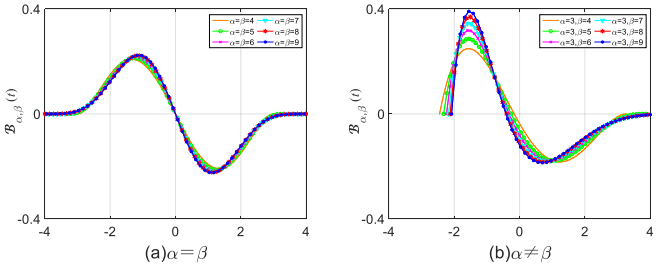

Fig. 1. The beta functions with different parameters α and β, (a) $\alpha=\beta$, (b) $\alpha\neq\beta$.

and shifted derivations of the mother wavelet ψ and the associated scale function ϕ; $c_{j_0,l}$ and $d_{j,l}$ are the wavelet decomposition coefficients; j_0 is an arbitrary integer representing the coarsest resolution or scale level. Simultaneously, based on the properties of multi-resolution analysis theory, any square integrable function h can be arbitrarily approximated using the basic scale functions $\phi_{j,l}(x)$ by setting the resolution scale level to be sufficiently large, that means, there exists an integer j, such that:

$$h(x)=\sum_{l} c_{j,l}\phi_{j,l}(x) \quad (6)$$

As the beta wavelet function $\mathcal{B}_{\alpha,\beta}$ is completely supported and defined on the section $[a,b]$, the set of the functions $\phi_{j,l}(x)=2^{j/2}\mathcal{B}_{\alpha,\beta}(2^j x-l)$, with the scale and shift indices j and l, should satisfy $a\leq 2^j x-l\leq b$. Assume that the function $h(x)$ approximated with decompositions (5) or (6) is defined within $[0,1]$, then the effective values for the shift index l are restricted to the collection $\Gamma_{\alpha,\beta}=\{l\in\mathbb{Z}\,|-b\leq l\leq 2^j-a\}$ for any given scale index j. We can obtain a set of basis functions $\{\phi_{j,l}^{(\alpha,\beta)}|\alpha,\beta,j\in\mathbb{Z},l\in\Gamma_{\alpha,\beta}\}$ by a shifted and dilated derivation of a beta wavelet function $\mathcal{B}_{\alpha,\beta}(t)$.

The MBW basis functions $\{\phi_{j,l}^{(\alpha_1,\beta_1)}\}\cup\{\phi_{j,l}^{(\alpha_2,\beta_2)}\}\cup\cdots\cup\{\phi_{j,l}^{(\alpha_n,\beta_n)}\}$ composed of various groups of beta wavelet basis functions, obtained by different parameters $\{(\alpha_1,\beta_1),(\alpha_2,\beta_2),\cdots,(\alpha_n,\beta_n)\}$, can effectively track complicated TV signals with both fast-varying and slowly-varying features. For most nonlinear dynamical modelling problems, multiple appropriate variants with a narrow bell-shaped half-cycle, a wide bell-shaped half-cycle, and a smooth half-cycle, such as the combination of different parameters α and β $\{(3,6),(3,9),(9,9)\}$, are capable of capturing both abrupt and slow changes of nonstationary signals, simultaneously [34]. Therefore, the parameters α and β with $\{(3,6),(3,9),(9,9)\}$ are adopted in the present of study. Additionally, theoretically, choosing a higher value of scale j, more basis functions will be involved in approximating the TV parameters, which may improve the resolution but would increase the computational cost. As a tradeoff, $j=3$ or 4 is generally an appropriate choice for many applications using MBW basis functions [8, 19].

Based on the wavelet theory, each TV parameter in (3) can be expanded into the following form by the MBW basis functions:

$$\zeta_{p,q}(k_1,\cdots,k_{p+q},t)=\sum_{n}\sum_{l\in\Gamma_{\alpha_n,\beta_n}} c_{p,q,l}^{\alpha_n,\beta_n}(k_1,\cdots,k_{p+q})\phi_{j,l}^{(\alpha_n,\beta_n)}\left(\frac{t}{N}\right) \quad (7)$$

where $\{\phi_{j,l}^{(\alpha_n,\beta_n)}\}$ indicates a group of beta wavelet basis functions controlled by characteristic parameters (α_n,β_n), with the wavelet scale j and the shift indices $l\in\Gamma_{\alpha_n,\beta_n}$, $\Gamma_{\alpha_n,\beta_n}=\{l\in\mathbb{Z}\,|-b\leq l\leq 2^j-a\}$; $c_{p,q,l}^{\alpha_n,\beta_n}(k_1,\cdots,k_{p+q})$ denotes the associated expansion parameter which is time-invariant; N is the number of observations of the signal.

Substituting (7) into (3) yields an expanded version of the TV-NARX model:

$$\begin{aligned} y(t)&=\sum_{g=1}^{G}\sum_{p=0}^{g}\sum_{k_1,k_{p+q}=1}^{K}\sum_{n}\sum_{l\in\Gamma_{\alpha_n,\beta_n}} c_{p,q,l}^{\alpha_n,\beta_n}(k_1,\cdots,k_{p+q}) \\ &\quad\times\left(\phi_{j,l}^{(\alpha_n,\beta_n)}\left(\frac{t}{N}\right)\prod_{i=1}^{p}y(t-k_i)\prod_{i=p+1}^{p+q}u(t-k_i)\right)+e(t) \quad (8)\\ &=\boldsymbol{\Psi}^{T}(t)\boldsymbol{\Theta}+e(t)\end{aligned}$$

where $\boldsymbol{\Psi}(t)$ is the expanded regression vector at time t; and $\boldsymbol{\Theta}=[c_{0,1,l}^{\alpha_n,\beta_n},\cdots,c_{1,0,l}^{\alpha_n,\beta_n},\cdots,c_{p,q,l}^{\alpha_n,\beta_n},\cdots]^{T}$ is the corresponding expanded time-invariant parameter vector.

The original TV-NARX model is now transformed into a time-invariant regression model which is linear-in-the-parameters. However, there might be a large number of redundant terms in the expanded regression vector $\boldsymbol{\Psi}(t)$, especially when the group number of beta wavelets (n), the maximum lag (K) and the degree (G) of the TV-NARX model are large. Therefore, reducing the number of terms in the expanded model and determining a parsimonious model structure become a crucial step in the identification of the original nonlinear TV problem.

In this paper, we propose a LRUOFR algorithm to select the significant terms from an over-complete dictionary of the expanded candidate model terms, and estimate the corresponding time-invariant parameters, so as to obtain the desired parsimonious model. In order to achieve a tradeoff between the model complexity and the value of model error, a modified generalized APRESS criterion is incorporated in the LRUOFR algorithm to determine the appropriate number of the significant terms in the parsimonious model. The novel algorithm to deal with this identification problem will be introduced in next section.

C. LRUOFR Algorithm Incorporating APRESS Criterion

The estimation of the parameters $\boldsymbol{\Theta}$ in (8) can be achieved by minimizing an ultra-least squares (ULS) criterion [26]:

$$\begin{aligned} J_{ULS} = & \left\| y - \sum_{g=1}^{G}\sum_{p=0}^{g}\sum_{k_1,k_{p+q}=1}^{K}\sum_{n} \right. \\ & \left. \sum_{l\in\Gamma_{\alpha_n,\beta_n}} c_{p,q,l}^{\alpha_n,\beta_n}\left(k_1,\cdots,k_{p+q}\right) x_{p,q,l}^{\alpha_n,\beta_n}\left(k_1,\cdots,k_{p+q}\right) \right\|_2^2 \\ & + \sum_{z=1}^{z_0} \left\| \bar{y}^z - \sum_{g=1}^{G}\sum_{p=0}^{g}\sum_{k_1,k_{p+q}=1}^{K}\sum_{n} \right. \\ & \left. \sum_{l\in\Gamma_{\alpha_n,\beta_n}} c_{p,q,l}^{\alpha_n,\beta_n}\left(k_1,\cdots,k_{p+q}\right)\left(\bar{x}_{p,q,l}^{\alpha_n,\beta_n}\left(k_1,\cdots,k_{p+q}\right)\right)^z \right\|_2^2 \end{aligned} \tag{9}$$

where $x_{p,q,l}^{\alpha_n,\beta_n}\left(k_1,\cdots,k_{p+q}\right)=\phi_{j,l}^{(\alpha_n,\beta_n)}(t/N)\prod_{i=1}^{p}y(t-k_i)\prod_{i=p+1}^{p+q}u(t-k_i)$ indicates an expanded term; $\bar{y}^z$ and $\left(\bar{x}_{p,q,l}^{\alpha_n,\beta_n}\right)^z$ represent weak derivative expressions of the signal y and model term $x_{p,q,l}^{\alpha_n,\beta_n}$, respectively; z_0 is the maximum degree of the weak derivative.

The weak derivative is a generalization of the commonly used classical derivative, which can be used to measure local correlation among the data points. Different from the derivatives defined for differentiable functions, the weak derivative can be calculated for all integrable functions. For a given sample data set, a discrete-time representation of the weak derivatives $\bar{y}^z$ and $\left(\bar{x}_{p,q,l}^{\alpha_n,\beta_n}\right)^z$ can be expressed as:

$$\begin{aligned} \bar{y}^z(\tau) &= \sum_{t=\tau}^{\tau+\tau_0} y(t)\bar{\omega}^{(z)}(t-\tau) \\ \left(\bar{x}_{p,q,l}^{\alpha_n,\beta_n}\right)^z(\tau) &= \sum_{t=\tau}^{\tau+\tau_0} x_{p,q,l}^{\alpha_n,\beta_n}(t)\bar{\omega}^{(z)}(t-\tau) \end{aligned} \tag{10}$$

where $\bar{\omega}^{(z)}(t)$ is the z-th derivative of a normalized test function, which can be calculated as $\bar{\omega}^{(z)}=\omega^{(z)}\big/\left\|\omega^{(z)}\right\|_2$; τ_0 is the support of the test function and τ=1,2,⋯,N$-$T$_0$. In this paper, the spline function is used as the test function, and the sampled data are modulated by the first- and second-order derivatives of the spline function [26].

Then the extended model (8) can be further expressed as a ULS system with weak derivative information:

$$\mathbf{Y}=\boldsymbol{\Phi}\cdot\boldsymbol{\Theta}+\mathbf{E} \tag{11}$$

where

$$\begin{aligned} \mathbf{Y}=\big[& y(1),\cdots,y(N),\bar{y}^1(1),\cdots,\bar{y}^1(N-\tau_0),\cdots, \\ & \bar{y}^{z_0}(1),\cdots,\bar{y}^{z_0}(N-\tau_0)\big]^T \end{aligned} \tag{12}$$

$$\boldsymbol{\Phi}=\begin{bmatrix} x_{0,1,l}^{\alpha_n,\beta_n}(k_1)(1) & \cdots & x_{G,0,l}^{\alpha_n,\beta_n}(k_1,\cdots,k_G)(1) \\ \vdots & \cdots & \vdots \\ x_{0,1,l}^{\alpha_n,\beta_n}(k_1)(N) & \cdots & x_{G,0,l}^{\alpha_n,\beta_n}(k_1,\cdots,k_G)(N) \\ \left(\bar{x}_{0,1,l}^{\alpha_n,\beta_n}(k_1)\right)^1(1) & \cdots & \left(\bar{x}_{G,0,l}^{\alpha_n,\beta_n}(k_1,\cdots,k_G)\right)^1(1) \\ \vdots & \cdots & \vdots \\ \left(\bar{x}_{0,1,l}^{\alpha_n,\beta_n}(k_1)\right)^{z_0}(N) & \cdots & \left(\bar{x}_{G,0,l}^{\alpha_n,\beta_n}(k_1,\cdots,k_G)\right)^{z_0}(N) \end{bmatrix} \tag{13}$$

and $\boldsymbol{\Theta}$ denotes the time-invariant parameter vector; $\mathbf{E}$ represents the noise of the system.

Assume that the regression matrix $\boldsymbol{\Phi}$ is full rank in columns and can be orthogonally decomposed as $\boldsymbol{\Phi}=\mathbf{WA}$, where $\mathbf{W}$ is a matrix with M orthogonal columns, denoted as $\mathbf{W}=[\mathbf{w}_1,\mathbf{w}_2,\cdots,\mathbf{w}_M]$, which satisfy $\mathbf{w}_i^T\mathbf{w}_j=0$, if $i\neq j$; $\mathbf{A}$ is an upper triangular matrix, expressed as follows:

$$\mathbf{A}=\begin{bmatrix} 1 & a_{1,2} & \cdots & a_{1,M} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{M-1,M} \\ 0 & \cdots & 0 & 1 \end{bmatrix} \tag{14}$$

The model (11) can alternatively be expressed as:

$$\mathbf{Y}=\mathbf{WA}\cdot\boldsymbol{\Theta}+\mathbf{E}=\mathbf{W}\cdot\boldsymbol{\upsilon}+\mathbf{E} \tag{15}$$

where the orthogonal regression weight vector $\boldsymbol{\upsilon}=[\upsilon_1,\upsilon_2,\cdots,\upsilon_M]^T$ satisfies the triangular system $\boldsymbol{\upsilon}=\mathbf{A}\cdot\boldsymbol{\Theta}$, and we can determine the time-invariant parameter vector $\boldsymbol{\Theta}$ if knowing $\boldsymbol{\upsilon}$ and $\mathbf{A}$.

The objective of model identification is to produce an optimal model that can well capture the inherent dynamics of underlying system, which can be achieved by minimizing the square of the norm (9). However, the ULS criterion ignores the interference of overlapping information which may lead to an ill-conditioned problem during forward regression selection process. Actually, there is a lot of overlapping information among the candidate terms in model (8), which makes it difficult to select a correct parsimonious model structure.

In order to avoid this problem, a stricter locally regularized ultra-least squares (LRULS) criterion is proposed in this study, which can be expressed as follow:

$$J_{LRULS}=J_{ULS}+\sum_{i=1}^{M}\lambda_i\upsilon_i^2=\mathbf{E}^T\mathbf{E}+\boldsymbol{\upsilon}^T\boldsymbol{\Lambda}\boldsymbol{\upsilon} \tag{16}$$

where $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots,\lambda_M]^T$ is the regularization parameter vector, and $\boldsymbol{\Lambda}$=diag$\{\lambda_1,\lambda_2,\cdots,\lambda_M\}$. Obviously, the LRULS criterion includes three parts: the first part is same as the standard least squares criterion that emphases the overall agreement between two time series; the second part considers the consistency of weak derivative information; and the third part is the regularization error which associates each candidate term with an individual regularization parameter to avoid the ill-conditioned problem caused by the overlapping information.

We can simplify the criterion (16) and obtain a comprehensible form [28]:

$$\frac{\mathbf{E}^T\mathbf{E}+\boldsymbol{\upsilon}^T\boldsymbol{\Lambda}\boldsymbol{\upsilon}}{\mathbf{Y}^T\mathbf{Y}}=1-\sum_{i=1}^{M}\frac{\left(\mathbf{w}_i^T\mathbf{w}_i+\lambda_i\right)\upsilon_i^2}{\mathbf{Y}^T\mathbf{Y}} \tag{17}$$

In order to measure the regularization error, the regularized error reduction ratio (RERR) is defined as:

$$RERR_i=\frac{\left(\mathbf{w}_i^T\mathbf{w}_i+\lambda_i\right)\upsilon_i^2}{\mathbf{Y}^T\mathbf{Y}} \tag{18}$$

Based on RERR, significant regressors can be selected by a forward-regression procedure. Note that, in the selection procedure, if $\mathbf{w}_i^T\mathbf{w}_i$ is too small (near zero), this term will not be selected. Thus, any ill-conditioning or singular situations can automatically be avoided.

The Bayesian evidence procedure is a practical choice to optimize the regularization parameters [28]. From the Bayesian viewpoint, the following error criterion is equivalent to the criterion (16):

$$J_B(\boldsymbol{\upsilon},\boldsymbol{\varepsilon},\varpi)=\varpi\mathbf{E}^T\mathbf{E}+\sum_{i=1}^{M}\varepsilon_i\upsilon_i^2=\varpi\mathbf{E}^T\mathbf{E}+\boldsymbol{\upsilon}^T\mathbf{H}\boldsymbol{\upsilon} \quad (19)$$

where ϖ is the noise parameter (estimate of the inverse of noise variance), $\boldsymbol{\varepsilon}=[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_M]^T$ is the hyperparameter vector, and $\mathbf{H}=\mathrm{diag}\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_M\}$. The relationship between a regularization parameter and its corresponding hyperparameter is obviously given by:

$$\lambda_i=\frac{\varepsilon_i}{\varpi} \quad (20)$$

Following Bayesian inference principle [35], it can be shown that the log evidence for $\boldsymbol{\varepsilon}$ and ϖ is:

$$\begin{aligned}\mathbf{L}_{ev}=&\sum_{i=1}^{M}\frac{1}{2}\log(\varepsilon_i)-\frac{M}{2}\log(\pi)-\frac{N_{ULS}}{2}\log(2\pi)\\&+\frac{N_{ULS}}{2}\log(\varpi)-\sum_{i=1}^{M}\frac{1}{2}\varepsilon_i\upsilon_i^2-\frac{1}{2}\varpi\mathbf{E}^T\mathbf{E}\\&-\frac{1}{2}\log(\det(\mathbf{B}_\mathbf{H}))+\frac{M}{2}\log(2\pi)\end{aligned} \quad (21)$$

where N_{ULS} denotes the length of the ULS system signal; $\mathbf{B}_\mathbf{H}$ represents the Hessian matrix which is diagonal and can be expressed as:

$$\begin{aligned}\mathbf{B}_\mathbf{H}&=\mathbf{H}+\varpi\mathbf{W}^T\mathbf{W}\\&=\mathrm{diag}\left\{\varepsilon_1+\varpi\mathbf{w}_1^T\mathbf{w}_1,\cdots,\varepsilon_M+\varpi\mathbf{w}_M^T\mathbf{w}_M\right\}\end{aligned} \quad (22)$$

Setting the derivatives of $\mathbf{L}_{ev}$ with respect to $\boldsymbol{\varepsilon}$ and ϖ to zeroes yields the updating formulas for $\boldsymbol{\varepsilon}$ and ϖ, respectively. Substituting these updating formulas into (20) results in the updating formulas for the regularization parameters:

$$\lambda_i^{new}=\frac{\gamma_i^{old}}{N_{ULS}-\gamma^{old}}\frac{\mathbf{E}^T\mathbf{E}}{\upsilon_i^2},\qquad 1\le i\le M \quad (23)$$

where γ_i and γ can be calculated by $\gamma_i=\mathbf{w}_i^T\mathbf{w}_i/(\lambda_i+\mathbf{w}_i^T\mathbf{w}_i)$ and $\gamma=\sum_{i=1}^{M}\gamma_i$, respectively. If $\boldsymbol{\lambda}$ remains sufficiently unchanged in two successive iterations or a pre-set maximum iteration number is reached, this update can be stopped.

Based on the above explanation, the implementation process of the LRUOFR algorithm is specifically presented in Appendix, where the test set $\mathbf{I}_C$ is used to avoid any ill-conditioning or singular problem. After this selection process, M_{ex} expanded model terms and the corresponding time-invariant parameters can be obtained. These parameters are then used to reconstruct the TV coefficients, and recover the selected terms from model (3). To avoid overfitting and ameliorate the effectiveness of the LRUOFR algorithm, a modified leave-one out (LOO) type cross-validation criterion, APRESS, can be employed to determine the optimal number of selected terms.

The APRESS statistic [29] expressed as follows can be used:

$$J[n_s]=p[n_s]MSE[n_s]=\frac{MSE[n_s]}{(1-C(n_s,\sigma))^2} \quad (24)$$

where $C(n_s,\sigma)=n_s\sigma/N$, with the adjustable parameter $\sigma\ge1$, is the complexity cost function; $p[n_s]=1/[1-C(n_s,\sigma)]^2$ is the penalty function; $MSE[n_s]=(1/N)\sum_{i=1}^{N}[y(i)-\hat{y}(i)]^2$ indicates the mean-squared-errors (residuals) calculated from the associated n_s-term model and $\{\hat{y}(i)\}_{i=1}^{N}$ is the one-step-ahead prediction sequence from the identified model of n_s model terms.

The criterion (24) consists of two parts: the mean-squared-error of the fit to the data, and the penalty. The optimal number $M_{optimal}$ of reconstructed terms for the desired model can be determined by minimizing the APRESS values:

$$M_{optimal}=\arg\min_{1\le n_s\le M_{re}}\{J[n_s]\} \quad (25)$$

where M_{re} is the number of recovered model terms.

Practically, a distinct point of the APRESS statistic with respect to the model length can be easily found through the change of adjustable parameter σ (see section III-A).

The new proposed algorithm for TV-NARX identification can be summarized as follows.

1) Set up the TV-NARX model (1) to be identified, and expand all TV coefficients of model terms by using MBW basis functions to obtain the model (8).
2) Based on the ULS criterion, construct a new model (11) according to (10) by using a normalized test function $\bar{\omega}$ to modulate the output vector and the regression matrix in the model (8).
3) Perform the local regularization-based OFR process with the output $\mathbf{Y}$ and regression matrix $\boldsymbol{\Phi}$ of model (11), and iteratively update the regularization parameter vector $\boldsymbol{\lambda}$ using (23).
4) Reselect significant expanded terms by returning to the OFR process with the updated $\boldsymbol{\lambda}$, and estimate corresponding time-invariant parameters according to the relationship $\boldsymbol{\Theta}=\mathbf{A}^{-1}\cdot\boldsymbol{\upsilon}$ obtained by (15).
5) Reconstruct the estimation of the TV coefficients using (7), and list the selected terms in order of the RERR values.
6) Determine the number $M_{optimal}$ of parsimonious model terms by using the APRESS criterion (24), and achieve the identification result of a nonstationary system.

III. Simulation Examples

In this section, three numerical simulations are given to illustrate the efficiency of the proposed MBW-LRUOFR algorithm. Furthermore, we compare this approach with three other methodologies: a classical adaptive method (the RLS algorithm), a latest parameter expansion method (the Bspline-UOFR algorithm), and a hybrid method (the MBW-UOFR algorithm) [19, 30, 36].

All of the following examples are performed via Monte Carlo simulations involving 100 realizations, and the results are given in terms of mean values. The first example presents a nonlinear TV system disturbed by severe colored noise. The second example is a discrete-time nonstationary system with non-continuously changing TV coefficients, and aims to verify the effectiveness of the MBW basis functions for capturing the local information around the abrupt change positions. Furthermore, the third example considers a more complex second-order TV nonlinear system with both smoothly and abruptly changing coefficients. Simultaneously, the identification accuracy of the TV coefficients at different noise levels (in term of SNRs) is given to verify the robustness and generalization property of the proposed approach.

A. Example 1: Detection of the Model Structure

Consider a TV nonlinear system of the form:

$$\begin{aligned} y(t) &= \zeta_{1,0}(t)\,y(t-2) + \zeta_{0,1}(t)\,u(t-1) + \zeta_{2,0}(t)\,y^2(t-1) \\ &\quad + \zeta_{0,2}(t)\,u^2(t-2) + \frac{1}{1-0.5z^{-1}}\varepsilon(t) \end{aligned} \tag{26}$$

where $\varepsilon(t) \sim N(0, 0.05^2)$, and the input signal is generated by an autoregressive process:

$$u(t) = \frac{0.25}{1-0.4z^{-1}+0.16z^{-2}}v(t) \tag{27}$$

where $v(t)$ is a Gaussian distributed noise $v(t) \sim N(0,1)$.

The TV coefficients in (26) are given as:

$$\begin{aligned} \zeta_{1,0}(t) &= -0.1 + 0.4\cos(4\pi t/1000), \quad 1 \le t \le 1000 \\ \zeta_{0,1}(t) &= \begin{cases} -0.6, & 1 \le t \le 300 \\ 0.5, & 300 < t \le 700 \\ -0.7, & 700 < t \le 1000 \end{cases} \\ \zeta_{2,0}(t) &= \begin{cases} -0.8, & 1 \le t \le 500 \\ 0.4, & 500 < t \le 1000 \end{cases} \\ \zeta_{0,2}(t) &= \begin{cases} 0.6, & 1 \le t \le 200 \\ -0.3, & 200 < t \le 800 \\ 0.4, & 800 < t \le 1000 \end{cases} \end{aligned} \tag{28}$$

Driven by the input signal (27), the system was simulated and a total of 1000 input-output data points were sampled. Note that the signal-to-noise ratio for the observed signal is SNR = 10 dB.

To increase the difficulty of system structure identification, the candidate model inputs are purposely chosen in an incorrect maximum lag of 7, which is much larger than the correct maximum lag 2. There are totally 120 candidate model terms included in the term dictionary when the nonlinear degree of the polynomial model is 2. As mentioned above, the parameters α and β are chosen to be $\{(3,6),(3,9),(9,9)\}$. The scale index involved in the beta wavelet (4) is $j = 3$. With these choices, the resulting MBW basis functions are used to expand TV coefficients. As a comparison, the Bspline-UOFR algorithm and the MBW-UOFR algorithm are also employed to identify the model structure, where B-spline functions of order 2 to 5 are adopted to generate basis functions.

All the significant model terms selected in the OFR process are reconstructed via (8) and listed in order of the RERR values in each Monte Carlo realization. For example, the reconstructed result produced by the MBW-LRUOFR algorithm in one simulation is presented in TABLE I. Notice that it still exists numerous redundant terms. The APRESS criterion is then used to determine the optimal number of model terms by setting the adjustable parameter σ = 1, 3, ⋯, 9, respectively. The corresponding curves of the statistic are shown in Fig. 2, where a distinct turning point suggests that 4 is the optimal model length. The model structure for this simulation can be determined as the first four terms given in TABLE I, which are highlighted in bold.

TABLE I
RECONSTRUCTED RESULTS PRODUCED BY THE MBW-LRUOFR ALGORITHM IN ONE SIMULATION FOR EXAMPLE 1

No.	Terms	$\mathbf{RERR}_i \times 100\%$
1	$\sum \boldsymbol{\phi}_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times \boldsymbol{u(t-1)}$	**64.8437**
2	$\sum \boldsymbol{\phi}_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times \boldsymbol{y(t-2)}$	**10.6934**
3	$\sum \boldsymbol{\phi}_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times \boldsymbol{u^2(t-2)}$	**7.0200**
4	$\sum \boldsymbol{\phi}_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times \boldsymbol{y^2(t-1)}$	**3.4173**
5	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times y(t-1)$	0.7955
6	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times u(t-2)$	0.3605
7	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times y(t-3)u(t-1)$	0.3136
8	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times u(t-3)$	0.2939
9	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times u^2(t-5)$	0.2075
10	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times y(t-6)$	0.1830
11	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times y(t-2)u(t-5)$	0.1528
12	$\sum \phi_{j,l}^{(\alpha_n,\beta_n)}(t/N) \times y(t-5)y(t-7)$	0.1517

Note: terms in bold indicate the correct model terms.

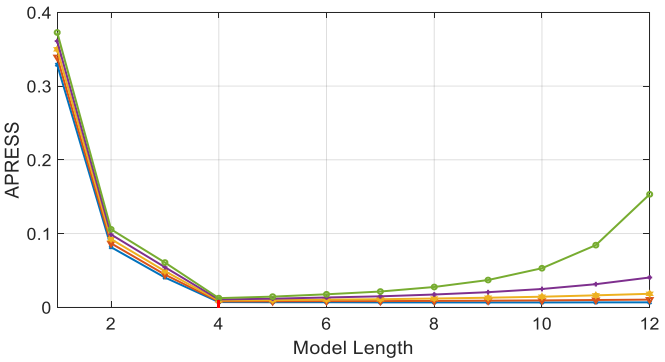

Fig. 2. The APRESS statistic versus the model length: the lines from bottom to the top correspond to σ = 1, 3, 5, 7, 9.

TABLE II
THE PERCENTAGES OF CORRECT TERMS SELECTED BY DIFFERENT ALGORITHMS FOR EXAMPLE 1

Approach	Bspline-UOFR	MBW-UOFR	MBW-LRUOFR
Percentage	78.50%	81.25%	89.50%

The percentage of the correctly selected model terms in each Monte Carlo realization is recorded, and the mean values for the three different algorithms are given in TABLE II. Obviously, the MBW-LRUOFR method with regularization parameters works better than the other two methods in determining the model structure from the given noisy simulation data. Compared to the UOFR-based methods, the proposed MBW-LRUOFR algorithm allocates an updated regularization parameter to each candidate regressor, this can effectively avoid the interference of overlapping information and assist the orthogonal regression process to produce a more accurate

model structure.

B. Example 2: Estimation of Non-Continuously Changing Time-Varying Coefficients

Consider the following TV-NARX model:

$$\begin{aligned} y(t) &= \zeta_{0,1}(t)u(t-1)+\zeta_{1,0}(t)y(t-2) \\ &+\zeta_{1,1}(t)y(t-2)u(t-2)+\frac{1}{1-0.32z^{-1}}\varepsilon(t) \end{aligned} \tag{29}$$

where $\varepsilon(t) \sim N(0,0.02^2)$ which makes the SNR to be around 30 dB; the input signal $u(t)$ is a pseudo-random binary sequence (PRBS), which is a frequency rich signal; the TV coefficients are designed to change in an abruptly varying manner as:

$$\zeta_{0,1}(t)=\begin{cases} 0.3, & 1\le t\le 200 \\ -0.7, & 200<t\le 400 \\ 0.2, & 400<t\le 600 \\ -0.3, & 600<t\le 800 \\ 0.6, & 800<t\le 1000 \end{cases}$$

$$\zeta_{1,0}(t)=\begin{cases} -0.4, & 1\le t\le 300 \\ 0.6, & 300<t\le 500 \\ -0.7, & 500<t\le 800 \\ 0.2, & 800<t\le 1000 \end{cases} \tag{30}$$

$$\zeta_{1,1}(t)=\begin{cases} 0.5, & 1\le t\le 200 \\ -0.3, & 200<t\le 500 \\ 0.7, & 500<t\le 700 \\ -0.6, & 700<t\le 1000 \end{cases}$$

The system was simulated and a total of 1000 input-output data points were sampled. Similar to Example 1, the parameters α and β are $\{(3,6),(3,9),(9,9)\}$, the scale index j equals 3; with these choices, the resulting MBW basis functions are used to expand the TV coefficients. All significant terms selected by the LRUOFR algorithm are listed in order of the RERR values, and the APRESS criterion is similarly used to determine the optimal number of model terms.

To verify the ability of the MBW basis functions in capturing the local information around the abrupt change positions, a comparison of the TV coefficients estimated by methods such as RLS (forgetting factor μ = 0.98), Bspline-UOFR, MBW-UOFR and MBW-LRUOFR is hown in Fig. 3. It can be observed that although the estimates produced by RLS algorithm can represent the actual TV coefficients to some extent, the approach cannot capture the transient properties of the jumps due to the limitation of the convergence speed. The Bspline-UOFR algorithm can estimate TV coefficients with a relatively higher accuracy than the RLS method, but the local information of the step position is missing. In contrast, the LRUOFR algorithm and UOFR algorithm, based on the MBW basis function expansion method, can not only recover the global features of the TV system, but also well capture the local information of the abrupt position of TV coefficients. In fact, Fig. 3 only shows those estimated results of the cases where all the model terms are correctly selected , this can facilitate the comparison between MBW-UOFR and MBW-LRUOFR algorithms. Fig. 3 shows that the MBW basis functions outperform these existing parametric modelling approaches for charactering local features of TV coefficients with sharp changes or jumps.

In order to further compare the identification accuracy of the above four algorithms, two error assessment criteria, namely,

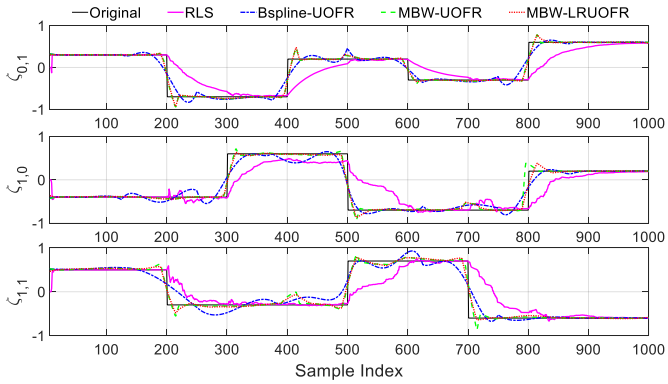

Fig. 3. Identification results of the TV coefficients using different approaches in example 2.

TABLE III
COMPARISON OF THE ESTIMATED RESULTS FOR EXAMPLE 2

Approach	Estimated coefficients	Error assessment criteria	
		MAE	RMSE
RLS (μ=0.98)	$\zeta_{0,1}$	0.1516	0.7878
	$\zeta_{1,0}$	0.1615	0.7289
	$\zeta_{1,1}$	0.1437	0.5507
Bspline-UOFR	$\zeta_{0,1}$	0.0611	0.3525
	$\zeta_{1,0}$	0.0819	0.2798
	$\zeta_{1,1}$	0.1068	0.3976
MBW-UOFR	$\zeta_{0,1}$	0.0315	0.2454
	$\zeta_{1,0}$	0.0449	0.1987
	$\zeta_{1,1}$	0.0576	0.2378
MBW-LRUOFR	$\zeta_{0,1}$	**0.0309**	**0.2435**
	$\zeta_{1,0}$	**0.0433**	**0.1966**
	$\zeta_{1,1}$	**0.0546**	**0.2337**

Note: bold values indicate the best results.

mean absolute error (MAE) and normalized root mean squared error (RMSE), are used to measure the TV coefficient estimation performance. MAE and RMSE are respectively defined as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{\zeta}(i)-\zeta(i)\right| \tag{31}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{\left[\hat{\zeta}(i)-\zeta(i)\right]^2}{\zeta(i)^2}} \tag{32}$$

where $\hat{\zeta}$ represents the estimates of TV coefficients ζ in the TV-NARX model, and N indicates the maximum sample index.

The mean values of MAE and RMSE for the three TV coefficients in Monte Carlo simulations are presented in TABLE III. It is obvious that the MAE and RMSE values for the two MBW-based methods are smaller than that for the Bspline-UOFR and RLS methods, this is consistent with the visual comparison shown in Fig. 3. This statistically indicates that even though the Bspline-based method possesses higher identification accuracy than RLS method, it still cannot achieve the performance of the MBW-based method. The comparison between Bspline-UOFR method and MBW-based methods further confirms that the MBW expansion method shows more attractive approximation characteristics than Bspline in tracking rapidly changing TV coefficients. Given the advantage of the LRUOFR algorithm in determining model structures, the

TABLE IV
COMPARISON OF THE ESTIMATED RESULTS IN DIFFERENT CASES FOR EXAMPLE 3 (SNR = 20, 15, AND 10 dB)

Approach	Estimated coefficients	SNR=20		SNR=15		SNR=10	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
RLS ($\mu = 0.95$)	$\zeta_{1,0}$	0.1241	1.8281	0.1279	1.8761	0.1331	1.9402
	$\zeta_{0,1}$	0.1888	0.5492	0.1700	0.5029	0.2006	0.5838
	$\zeta_{2,0}$	0.1658	3.0457	0.1817	2.8946	0.1740	2.6793
	$\zeta_{0,2}$	0.0305	0.1373	0.0508	0.1710	0.0726	0.2312
Bspline-UOFR	$\zeta_{1,0}$	0.0521	0.9072	0.0724	1.3676	0.0956	1.7393
	$\zeta_{0,1}$	0.0607	0.1872	0.0726	0.2037	0.0918	0.2333
	$\zeta_{2,0}$	0.0640	1.1595	0.0978	1.9009	0.1344	2.2924
	$\zeta_{0,2}$	0.0296	0.0812	0.0462	0.1255	0.0652	0.1811
MBW-UOFR	$\zeta_{1,0}$	0.0371	0.4294	0.0585	0.8044	0.0923	1.4094
	$\zeta_{0,1}$	0.0338	0.1334	0.0507	0.1586	0.0809	0.2159
	$\zeta_{2,0}$	0.0356	0.5594	0.0616	0.9914	0.1093	1.6222
	$\zeta_{0,2}$	0.0222	0.0618	0.0385	0.1048	0.0624	0.1624
MBW-LRUOFR	$\zeta_{1,0}$	**0.0362**	**0.4246**	**0.0565**	**0.7207**	**0.0871**	**1.2605**
	$\zeta_{0,1}$	**0.0322**	**0.1259**	**0.0497**	**0.1511**	**0.0768**	**0.2092**
	$\zeta_{2,0}$	**0.0346**	**0.5573**	**0.0582**	**0.9561**	**0.1090**	**1.5809**
	$\zeta_{0,2}$	**0.0216**	**0.0617**	**0.0376**	**0.1078**	**0.0613**	**0.1613**

where bold values indicate the best results.

proposed MBW-LRUOFR algorithm is more adaptive and performs better for identifying model structures and capturing local information of TV signals in the presence of colored noise.

C. Example 3: Identification of a Second-Order TV Nonlinear System

The third example is designed to test the performance of the proposed algorithm for dealing with a system with both smooth and sharp changes in system model parameters. The system is described by the model:

$$\begin{aligned} y(t) &= \zeta_{1,0}(t)\,y(t-2) + \zeta_{0,1}(t)\,u(t-1) \\ &\quad + \zeta_{2,0}(t)\,y(t-1)\,y(t-2) \\ &\quad + \zeta_{0,2}(t)\,u(t-1)\,u(t-2) + \frac{\varepsilon(t)}{1-0.25z^{-1}} \end{aligned} \tag{33}$$

where $\varepsilon(t) \sim N(0, 0.08^2)$ is zero-mean Gaussian white noise; the input $u(t)$ is a PRBS; $\zeta_{1,0}(t)$, $\zeta_{0,1}(t)$, $\zeta_{2,0}(t)$, and $\zeta_{0,2}(t)$ are TV coefficients of this system, expressed as:

$$\begin{aligned} \zeta_{1,0}(t) &= \begin{cases} 0.32\cos(1.5-\cos(4\pi t/N+\pi)), & 1\le t\le N/4 \\ 0.32\cos(3-\cos(4\pi t/N+\pi/2)), & N/4+1\le t\le 3N/4 \\ 0.32\cos(1.5-\cos(4\pi t/N+\pi)), & 3N/4+1\le t\le N \end{cases} \\ \zeta_{0,1}(t) &= \begin{cases} 0.54, & 1\le t\le N/4 \\ -0.65, & N/4+1\le t\le N/2 \\ 0.54, & N/2+1\le t\le 3N/4 \\ -0.65, & 3N/4+1\le t\le N \end{cases} \\ \zeta_{2,0}(t) &= 0.43\cos(4\pi t/N), \quad 1\le t\le N \\ \zeta_{0,2}(t) &= 0.5, \quad 1\le t\le N \end{aligned} \tag{34}$$

where N = 512 is the length of sampled data.

The parameters of MBW are the same as in Example 1. The LRUOFR algorithm is applied to select significant model terms from the candidate terms expanded by the MBW basis functions, and the APRESS criterion is similarly employed to determine the number of optimal model terms.

For a comparison, the four model coefficients reconstructed by the following 4 methods are shown in Fig. 4: RLS with forgetting factor μ = 0.95 (purple curve), Bspline-UOFR (blue curve), MBW-UOFR (green curve) and MBW-LRUOFR (red curve). Notice that the estimated results are compared based on the premise of all the model terms are correctly selected, to facilitate the comparison of the MBW-UOFR algorithm with the proposed MBW-LRUOFR algorithm. Based on this premise, it can be seen that the proposed MBW-LRUOFR algorithm performs better than the other methods in tracking the variations of the TV coefficients, especially in the abrupt positions. These results show that MBW-LRUOFR can effectively track the variation of different waveforms: the constant value, smooth changes, and abrupt changes.

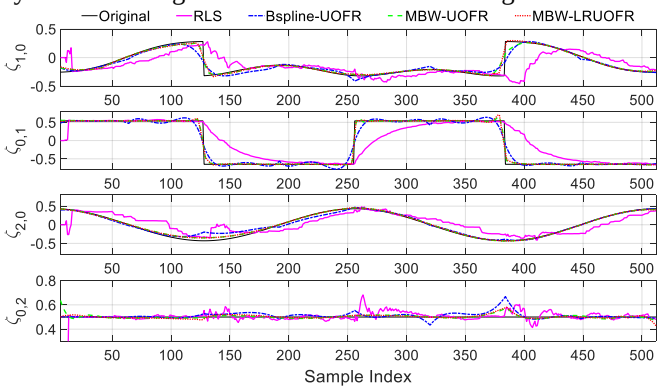

Fig. 4. Identification results of the TV coefficients using different approaches in example 3.

In order to verify the robustness and noise immunity of the proposed scheme, colored noise of the following three cases are added to the original system by adjusting the standard deviation of $\varepsilon(t)$, where the SNR is 20, 15, and 10 dB, respectively. The mean values of MAE and RMSE for estimated TV coefficients are given in TABLE IV, where it can be noted that the MAE and RMSE values given by MBW-LRUOFR are smaller than those by the RLS method and the Bspline-UOFR method for all the three cases. Specifically, the MBW-LRUOFR algorithm

based on the local regularization method can effectively capture

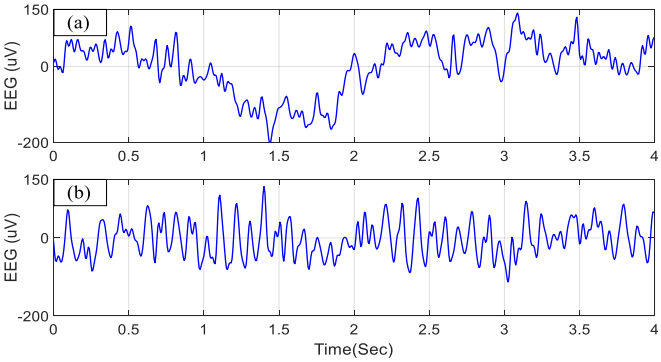

Fig. 5. EEG signals recorded during 4s with a sampling rate of 160 Hz. (a) EEG from a hand-moving MI task, (b) EEG from an eyes-closed resting state.

the major and local information of the TV coefficients when the noise level increases. These numerical results show that the MBW-LRUOFR method has better performance for noise immunity.

IV. Application To EEG Data

In this section, the proposed MBW-LRUOFR algorithm is applied to scalp EEG data to illustrate its ability for solving real-world TV modelling problem. In fact, the brain is a complicated black box system where the true model structure is unknown, thus it is necessary to identify a parsimonious model from available experimental data, and produce an accurate description of recording regions during brain activity [16]. The central objective of this section is to propose an effective data-based model for single-channel EEG recordings by using the MBW-LRUOFR algorithm.

The EEG recordings used in this study are available from Physionet [37], and created by the BCI2000 instrumentation system [38]. We choose two snapshots of EEG recordings sampled from the same channel of a same subject at different states, as shown in Fig. 5, where EEG1 (Fig. 5(a)) was recorded during a hand-moving motor imagery (MI) task and EEG2 (Fig. 5(b)) was recorded during an eyes-closed resting state. A second-order TV-NARX model without exogenous inputs is constructed with the maximum lag $K=10$, which is sufficient to reveal the underlying changes of EEG signals [19]. Thus totally 66 candidate model terms are involved in the initial full model:

$$\begin{aligned} y(t) &= \sum_{k_1=1}^{10}\sum_{k_2=1}^{10} \zeta_{2,0}(k_1,k_2,t)\, y(t-k_1)\, y(t-k_2) \\ &\quad + \sum_{k_1=1}^{10} \zeta_{1,0}(k_1,t)\, y(t-k_1) + \zeta_{0,0} + e(t) \end{aligned} \tag{35}$$

To obtain a compact model structure, the MBW-LRUOFR algorithm is used to select significant terms and estimate corresponding TV coefficients. The scale index of MBW function is chosen to be 3, and the APRESS criterion is adapted to determine the number of model terms. With the estimated TV coefficients presented in Fig. 6, the parsimonious model of EEG1 can be described as:

$$\begin{aligned} y(t) &= \sum_{k_1=1}^{3} \zeta_{1,0}(k_1,t)\, y(t-k_1) + \zeta_{2,0}(1,1,t)\, y^2(t-1) \\ &\quad + \zeta_{2,0}(2,2,t)\, y^2(t-2) + e(t) \end{aligned} \tag{36}$$

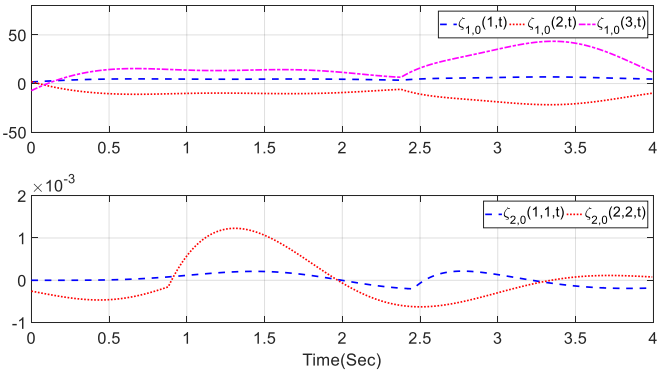

Fig. 6. The estimated TV coefficients of NARX model (36) for EEG1.

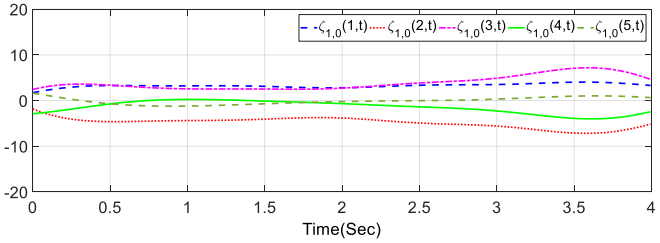

Fig. 7. The estimated TV coefficients of NARX model (37) for the EEG2.

Similarly, with the estimated TV coefficients presented in Fig. 7, the identified model of EEG2 can be described as:

$$y(t) = \sum_{k_1=1}^{5} \zeta_{1,0}(k_1,t)\, y(t-k_1) + e(t) \tag{37}$$

Note that the identified model (36) obtained from the MI EEG recordings is more complex than model (37) which contains only linear terms.

From the estimated TV coefficients depicted in Fig. 6 and Fig. 7, some interesting observations of the underlying changing behavior of EEG1 and EEG2 signals can be obtained. For example, during the MI task of EEG1, the coefficients corresponding to the first-order model terms change relatively smoothly, while the coefficients of the second-order model terms change relatively more violently, especially in the period of 1 to 3 seconds. In addition, a significant turning point occurs around 2.5 second in the estimated TV coefficients, which can be understood as the characteristic change of the sampled signal. However, all the TV coefficients of model (37) estimated from EEG2 recordings are smooth during this experimental time, which is consistent with the fact that the subject was in a resting state.

Furthermore, the recovered signals obtained by model (36) and model (37) are compared with the original EEG recordings (see Fig. 8), to verify the effectiveness of the identified models. For a clear visualization, only the data points in the period of 2.5 to 3.5 seconds are displayed. By comparing the estimated signals with the real signals, it can be seen that the models constructed by the proposed method can well follow the changing process of the scalp EEG signal. The identification performance indicates that the MBW-LRUOFR algorithm is effective for modeling the real EEG data.

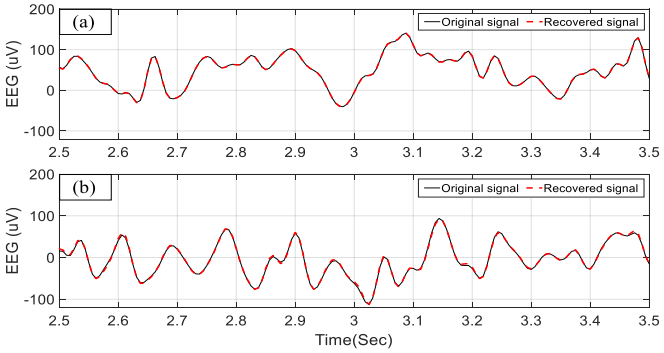

Fig. 8. A comparison of the recovered signals and the original EEG recordings. (a) EEG from a hand-moving MI task, (b) EEG from an eyes-closed resting state. For a clear visualization, only the data points in the period of 2.5 to 3.5 seconds are displayed.

V. Conclusions

A novel MBW-LRUOFR algorithm incorporating the modified generalized APRESS criterion has been proposed for the identification of nonstationary systems, where time-dependent coefficients of a TV-NARX model were approximated by a set of MBW basis functions. Three numerical simulation examples have been used to test the performance of the proposed scheme. Many typical TV coefficients, including both smooth and abrupt changes, were considered in the three simulation case studies. The identification results indicated that the proposed method can effectively determine the optimal model structure and accurately estimate the TV coefficients. Furthermore, an application to scalp EEG data showed that the proposed scheme performed well in tracking quickly changing nonstationary systems and revealing the underlying mechanism of EEG signals.

An advantage of the MBW-LRUOFR algorithm over the previous methods is that it can effectively capture the overall and local information of a nonstationary system. However, the computational load of the proposed method is much higher than existing functional series expansion methods due to the existence of an iterative process of regularization parameters. Actually, the number of regressors decreases dramatically within the first few iterations, and typically about 10 iterations in total suffice to construct desired parsimonious model [28]. So that, compared to the improvement of identification accuracy, this computational issue becomes less critical when a high performance PC is available.

A major application of the proposed method in our study is to investigate the TV model of nonstationary systems including EEG signals. Actually, the works of Li et al. have shown that an effective model can assist reveal the underlying mechanisms of biological signals, for example, the studies of the causality between signals in different channels [30, 39]. Thus, a promising research direction is the further applications in time-frequency distribution and causality detection of biomedical signals. These work will be presented in our future separate publication.

Appendix

Algorithm 1: Pseudocode for LRUOFR algorithm

Input:

ULS system output $\mathbf{Y}=\left[\mathrm{y}(1),\cdots,\bar{\mathrm{y}}^{z_0}\left(\mathrm{N}-\tau_0\right)\right]^{\mathrm{T}}$

regression matrix $\boldsymbol{\Phi}=\left[\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2,\cdots,\boldsymbol{\Phi}_{\mathrm{M}}\right]$

Initialize:

predetermined thresholds $\chi=10^{-10}$, $\rho=10^{-3}$

initial regularization parameters $\left\{\lambda_{\mathrm{i}}=10^{-3}\,\middle|\,1\le \mathrm{i}\le \mathrm{M}\right\}$

Local regularization-based OFR process:

Let $\mathbf{Y}^{(1)}=\mathbf{Y}$; $\boldsymbol{\Phi}^{(1)}=\boldsymbol{\Phi}$;

For $\kappa=1$ to M

$$\boldsymbol{\Phi}^{(\kappa)}=\left[\mathbf{w}_1,\cdots,\mathbf{w}_{\kappa-1},\boldsymbol{\Phi}_{\kappa}^{(\kappa)},\cdots,\boldsymbol{\Phi}_{\mathrm{M}}^{(\kappa)}\right];$$

$$\mathbf{Y}^{(\kappa)}=\mathbf{Y}^{(\kappa-1)}-\frac{\mathbf{w}_{\kappa-1}^{\mathrm{T}}\mathbf{Y}^{(\kappa-1)}}{\mathbf{w}_{\kappa-1}^{\mathrm{T}}\mathbf{w}_{\kappa-1}+\lambda_{\kappa-1}}\mathbf{w}_{\kappa-1};$$

$$\mathbf{I}_{\mathrm{C}}=\left\{\mathrm{i}\,\middle|\,\left(\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa-1)}\right)^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa-1)}<\chi,\ \kappa\le \mathrm{i}\le \mathrm{M}\right\};$$

For $\mathrm{i}=\kappa$ to M

$$\mathrm{a}_{\kappa,\mathrm{i}}=\frac{\left(\boldsymbol{\Phi}_{\kappa}^{(\kappa)}\right)^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa)}}{\left(\boldsymbol{\Phi}_{\kappa}^{(\kappa)}\right)^{\mathrm{T}}\boldsymbol{\Phi}_{\kappa}^{(\kappa)}};\quad \upsilon_{\mathrm{i}}^{(\kappa)}=\frac{\left(\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa)}\right)^{\mathrm{T}}\mathbf{Y}^{(\kappa)}}{\left(\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa)}\right)^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa)}+\lambda_{\mathrm{i}}};$$

$$\mathrm{rerr}_{\mathrm{i}}=\frac{\left(\upsilon_{\mathrm{i}}^{(\kappa)}\right)^2\left[\left(\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa)}\right)^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa)}+\lambda_{\mathrm{i}}\right]}{\mathbf{Y}^{\mathrm{T}}\mathbf{Y}};$$

end for

$$\iota_{\kappa}=\arg\max\left\{\mathrm{rerr}_{\mathrm{i}}\,\middle|\,\kappa\le \mathrm{i}\le \mathrm{M}\text{ and }\mathrm{i}\notin\mathbf{I}_{\mathrm{C}}\right\};$$

$$\mathrm{RERR}_{\kappa}=\mathrm{rerr}_{\iota_{\kappa}};\ \mathbf{w}_{\kappa}=\boldsymbol{\Phi}_{\iota_{\kappa}}^{(\kappa)};\ \boldsymbol{\Phi}_{\iota_{\kappa}}^{(\kappa)}=\boldsymbol{\Phi}_{\kappa}^{(\kappa)};$$

$$\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa+1)}=\boldsymbol{\Phi}_{\mathrm{i}}^{(\kappa)}-\mathrm{a}_{\kappa,\mathrm{i}}\mathbf{w}_{\kappa},\ \kappa+1\le \mathrm{i}\le \mathrm{M};$$

$$\mathrm{sum}=\sum_{\mathrm{i}=1}^{\kappa}\mathrm{RERR}_{\mathrm{i}};$$

If $1-\mathrm{sum}<\rho$

$\mathrm{M}_{\mathrm{ex}}=\kappa$; break;

end if

end for

Update regularization parameters $\boldsymbol{\lambda}$

$$\lambda_{\mathrm{i}}^{\mathrm{new}}=\frac{\gamma_{\mathrm{i}}^{\mathrm{old}}}{\mathrm{N}_{\mathrm{ULS}}-\gamma^{\mathrm{old}}}\frac{\mathbf{E}^{\mathrm{T}}\mathbf{E}}{\upsilon_{\mathrm{i}}^2};\quad \mathrm{Dev}=\frac{\sum\left|\lambda_{\mathrm{i}}^{\mathrm{new}}-\lambda_{\mathrm{i}}^{\mathrm{old}}\right|/\lambda_{\mathrm{i}}^{\mathrm{new}}}{\mathrm{M}_{\mathrm{ex}}};$$

If $\mathrm{Dev}\le 0.1$ *(for example)*

stop updating;

else

return to the OFR process with updated $\boldsymbol{\lambda}$;

end if

Estimate time-invariant parameters:

$\boldsymbol{\Theta}=\mathbf{A}^{-1}\cdot\boldsymbol{\upsilon}$;

Output:

time-invariant parameters: $\boldsymbol{\Theta}$

selected model terms: $\boldsymbol{\Upsilon}=\left[\boldsymbol{\Phi}_{\iota_1},\boldsymbol{\Phi}_{\iota_2},\cdots,\boldsymbol{\Phi}_{\iota_{\mathrm{M}_{\mathrm{ex}}}}\right]$

REFERENCES

[1] C. Huang, C. Liu, and W. Su, "Application of Cauchy wavelet transformation to identify time-variant modal parameters of structures," Mechanical Systems and Signal Processing, vol. 80, pp. 302-323, 2016.

[2] S. Raj and K. C. Ray, "ECG signal analysis using DCT-based DOST and PSO optimized SVM," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 3, pp. 470-478, 2017.

[3] E. Zhang, M. Schoukens, and J. Schoukens, "Structure detection of Wiener–Hammerstein systems with process noise," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 3, pp. 569-576, 2017.

[4] Y. F. Zhao, S. A. Billings, H. L. Wei, and P. G. Sarrigiannis, "A parametric method to measure time-varying linear and nonlinear causality with applications to EEG data," IEEE Transactions on Biomedical Engineering, vol. 60, no. 11, pp. 3141-3148, 2013.

[5] Y. Li, Q. Liu, S. R. Tan, and R. H. Chan, "High-resolution time-frequency analysis of EEG signals using multiscale radial basis functions," Neurocomputing, vol. 195, pp. 96-103, 2016.

[6] J. M. Bravo, A. Suarez, M. Vasallo, and T. Alamo, "Slide window bounded-error time-varying systems identification," IEEE Transactions on Automatic Control, vol. 61, no. 8, pp. 2282-2287, 2016.

[7] A. Klepka and T. Uhl, "Identification of modal parameters of non-stationary systems with the use of wavelet based adaptive filtering," Mechanical Systems and Signal Processing, vol. 47, no. 1-2, pp. 21-34, 2014.

[8] Y. Li, M. L. Luo, and K. Li, "A multiwavelet-based time-varying model identification approach for time–frequency analysis of EEG signals," Neurocomputing, vol. 193, pp. 106-114, 2016.

[9] Y. Li, W. G. Cui, M. L. Luo, K. Li, and L. Wang, "High-resolution time–frequency representation of EEG data using multi-scale wavelets," International Journal of Systems Science, vol. 48, no. 12, pp. 2658-2668, 2017.

[10] A. A. Adeniran and S. El Ferik, "Modeling and identification of nonlinear systems: A review of the multimodel approach—Part 1," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 7, pp. 1149-1159, 2017.

[11] R. Abdolee, B. Champagne, and A. H. Sayed, "Estimation of space-time varying parameters using a diffusion LMS algorithm," IEEE Transactions on Signal Processing, vol. 62, no. 2, pp. 403-418, 2014.

[12] S. Braun and E. A. Habets, "Online Dereverberation for Dynamic Scenarios Using a Kalman Filter With an Autoregressive Model," IEEE Signal Processing Letters, vol. 23, no. 12, pp. 1741-1745, 2016.

[13] C. P. Chen, T. Zhang, L. Chen, and S. C. Tam, "I-Ching divination evolutionary algorithm and its convergence analysis," IEEE Transactions on Cybernetics, vol. 47, no. 1, pp. 2-13, 2017.

[14] Y. Li, X. D. Wang, M. L. Luo, K. Li, X. F. Yang, and Q. Guo, "Epileptic seizure classification of EEGs using time–frequency analysis based multiscale radial basis functions," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, pp. 386-397, 2018.

[15] M. Gan, C. P. Chen, H. X. Li, and L. Chen, "Gradient radial basis function based varying-coefficient autoregressive model for nonlinear and nonstationary time series," IEEE Signal Processing Letters, vol. 22, no. 7, pp. 809-812, 2015.

[16] Y. Li, H. L. Wei, S. A. Billings, and P. Sarrigiannis, "Time-varying model identification for time–frequency feature extraction from EEG data," Journal of Neuroscience Methods, vol. 196, no. 1, pp. 151-158, 2011.

[17] S. A. Billings and H. L. Wei, "Sparse model identification using a forward orthogonal regression algorithm aided by mutual information," IEEE Transactions on Neural Networks, vol. 18, no. 1, pp. 306-310, 2007.

[18] X. Hong, S. Chen, J. B. Gao, and C. J. Harris, "Nonlinear identification using orthogonal forward regression with nested optimal regularization," IEEE Transactions on Cybernetics, vol. 45, no. 12, pp. 2925-2936, 2015.

[19] Y. Li, W. G. Cui, Y. Z. Guo, T. W. Huang, X. F. Yang, and H. L. Wei, "Time-varying system identification using an ultra-orthogonal forward regression and multiwavelet basis functions with applications to EEG," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 7, pp. 2960-2972, 2018.

[20] T. Song and D. Lin, "Hybrid time-variant frequency response function estimates using multiple sets of basis functions," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 2, pp. 263-279, 2017.

[21] Y. Li, H. L. Wei, and S. A. Billings, "Identification of time-varying systems using multi-wavelet basis functions," IEEE Transactions on Control Systems Technology, vol. 19, no. 3, pp. 656-663, 2011.

[22] P. Z. Csurcsia, J. Schoukens, I. Kollár, and J. Lataire, "Nonparametric time-domain identification of linear slowly time-variant systems using B-splines," IEEE Transactions on Instrumentation and Measurement, vol. 64, no. 1, pp. 252-262, 2015.

[23] H. De Oliveira and G. De Araújo, "Compactly Supported One-cyclic Wavelets Derived from Beta Distributions," J. Commun. Inform. Syst., vol. 20, no. 3, pp. 27–33, 2005.

[24] R. Kumar, A. Kumar, and R. K. Pandey, "Beta wavelet based ECG signal compression using lossless encoding with modified thresholding," Computers & Electrical Engineering, vol. 39, no. 1, pp. 130-140, 2013.

[25] T. Zhang, C. P. Chen, L. Chen, X. M. Xu, and B. Hu, "Design of highly nonlinear substitution boxes based on I-Ching operators," IEEE Transactions on Cybernetics, no. 99, pp. 1-10, 2018.

[26] Y. Z. Guo, L. Guo, S. Billings, and H. L. Wei, "Ultra-orthogonal forward regression algorithms for the identification of non-linear dynamic systems," Neurocomputing, vol. 173, pp. 715-723, 2016.

[27] X. Hong, C. J. Harris, and S. Chen, "Robust neurofuzzy rule base knowledge extraction and estimation using subspace decomposition combined with regularization and D-optimality," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 34, no. 1, pp. 598-608, 2004.

[28] J. Dolinský, K. Hirose, and S. Konishi, "Readouts for echo-state networks built using locally regularized orthogonal forward regression," Journal of Applied Statistics, vol. 45, no. 4, pp. 740-762, 2018.

[29] Y. L. Gu and H. L. Wei, "A robust model structure selection method for small sample size and multiple datasets problems," Information Sciences, vol. 451, pp. 195-209, 2018.

[30] Y. Li, H. L. Wei, S. A. Billings, and X. F. Liao, "Time-varying linear and nonlinear parametric model for Granger causality analysis," Physical Review E, vol. 85, no. 4, p. 041906, 2012.

[31] S. A. Billings, Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains. Hoboken, NJ, USA: Wiley, 2013.

[32] F. He, H. L. Wei, and S. A. Billings, "Identification and frequency domain analysis of non-stationary and nonlinear systems using time-varying NARMAX models," International Journal of Systems Science, vol. 46, no. 11, pp. 2087-2100, 2015.

[33] Z. H. Lai, Y. Xu, Q. C. Chen, J. Yang, and D. Zhang, "Multilinear sparse principal component analysis," IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 10, pp. 1942-1950, 2014.

[34] C. B. Amar, M. Zaied, and A. Alimi, "Beta wavelets. Synthesis and application to lossy image compression," Advances in Engineering Software, vol. 36, no. 7, pp. 459-474, 2005.

[35] M. Luessi, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian simultaneous sparse approximation with smooth signals," IEEE Transactions on Signal Processing, vol. 61, no. 22, pp. 5716-5729, 2013.

[36] L. Huang and H. Hjalmarsson, "A multi-time-scale generalization of recursive identification algorithm for ARMAX Systems," IEEE Transactions on Automatic Control, vol. 60, no. 8, pp. 2242-2247, 2015.

[37] A. L. Goldberger et al., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215-e220, 2000.

[38] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," IEEE Transactions on Biomedical Engineering, vol. 51, no. 6, pp. 1034-1043, 2004.

[39] Y. Li, M. Y. Lei, Y. Z. Guo, Z. Y. Hu, and H. L. Wei, "Time-varying nonlinear causality detection using regularized orthogonal least squares and multi-wavelets with applications to EEG," IEEE Access, vol. 6, pp. 17826-17840, 2018.