A Refined ERR-based Method for Nonlinear System Identification. Application to Epilepsy.

Marc Greige Ahmad Karfoul Isabelle Merlet Régine Le Bouquin Jeannès
Univ Rennes, INSERM, LTSI - UMR 1099, F-35000 Rennes, France
email: marc.greige@univ-rennes.fr ahmad.karfoul@univ-rennes.fr
isabelle.merlet@univ-rennes.fr regine.le-bouquin-jeannes@univ-rennes.fr

***Abstract*—The goal of this paper is to refine the solution of the Error Reduction Ratio (ERR)-based method for nonlinear system identification in the context of epilepsy. Based on a predefined dictionary, the ERR-based method is composed of two main steps: (i) identifying the most relevant candidates that are required to fit the signal at hand, and (ii) estimating their respective weights in a least squares sense. However, the used candidate selection criterion, which is based on a fixed threshold, often leads to an overestimation of the number of retained candidates. This consequently affects the quality of the system identification. This point is of particular interest in epilepsy especially for the identification of brain networks involved in the seizure onset. To deal with this issue, a refined ERR-based solution is proposed in this paper. It relies on the assumption that a few number of the retained candidates using the ERR-based method are really the most significant ones. This leads to consider a sparse representation of the associated estimated coefficient vector. The well-known Proximal Alternating Linearized Minimization (PALM) is used in this paper to solve the proposed optimization problem. To guarantee good estimation results, the used regularization parameter is, at each iteration, optimally computed using the discrepancy principle. Results on simulated and real iEEG data confirm the efficacy of the proposed method.**

Keywords*—*Error Reduction Ratio; Orthogonal Least Squares; proximal optimization; epilepsy; effective connectivity

I. INTRODUCTION

Epilepsy is a group of neurological disorders that cause temporary dysfunctions of the brain electrical activity. It is characterized by repetitive seizures - called ictal periods - whose frequency and duration may vary. Epileptic seizures are induced by abnormal excessive or synchronous neuronal activity in certain regions of the brain, known as epileptogenic [1]. Around 30% of epileptic patients are drug-resistant, for whom alternative therapies, such as surgery or neural stimulation, must be considered. Satisfactory outcomes of these therapies require beforehand a reliable identification of the epileptic network underlying the initiation and/or the propagation of the epileptic seizures. Identifying the epileptic network involves not only its nodes (brain regions) but also the direction of the information flow among them leading to the concept of brain effective connectivity [2]. Intracerebral electroencephalographic (iEEG) recording is a commonly used invasive technique to record brain electrical activity [3], [4]. Albeit invasive, it provides recordings with relatively high Signal-to-Noise Ratio (SNR) and free from the volume conduction effect. Neural activities are generally the result of nonlinear processes, and hence interactions between brain regions can be qualified as nonlinear. Consequently, analyzing interactions among brain regions in a linear way is sub-optimal. The Error Reduction Ratio (ERR)-based method [5]–[9] has already shown promising results in identifying nonlinear systems and inferring effective connectivity between brain regions. It is a dictionary based approach comprising two main steps: (i) identifying from a predefined dictionary the most relevant candidates that are required to fit the signal at hand; this is performed through an orthogonal least squares (OLS) scheme combined with a threshold-based candidate selection criterion [5] [6], and (ii) estimating their respective weights (model coefficients) in a least squares sense. Despite its efficiency, the ERR-based method suffers from the presence of spurious terms whose number is subject to the predefined threshold. This often leads to low system identification quality and consequently induces errors in the inference of effective connectivity. To cope with this limitation, a refined ERR-based method, denoted by rERR, is proposed in this paper. The solution relies on the assumption that, among those retained dictionary candidates using the original ERR-based method, a few number is really contributing to the signal at hand. These few but relevant candidates are found by simply prompting a sparse representation on the retained dictionary. To this end, the Proximal Alternating Linearized Minimization (PALM) [10] method is used. Besides, an optimal computation of the regularization parameter is also performed at each iteration of the proposed iterative approach leading to a more reliable identification quality. The behavior of the proposed rERR-based approach is compared to the original ERR-based one using both simulated signals and real epileptic iEEG recordings. In this contribution, Section II is devoted to the methodology before presenting the dataset in Section III. Results are given in Section IV where the rERR-based approach is compared to the original one. Some concluding remarks are given in Section V.

II. METHODOLOGY

Assume that a set $\{\boldsymbol{y}_m\}_{m\in 1,\cdots,M}$ of M epileptic iEEG signals are recorded over a T period of time. The m-th iEEG signal $\boldsymbol{y}_m$ denotes the neural activity of the m-th brain region. As brain is a complex network of distributed interconnected regions, epileptic seizures can be initiated and propagated due to a specific brain epileptic network whose nodes are the involved brain regions and edges reflecting how these

 ISBN: 978-1-68558-057-5

brain regions interact. Thus, the activity of the m-th brain region, $\boldsymbol{y}_m$, is linked to the ones of other brain regions. More precisely, assume that $\boldsymbol{y}_m$ can be decomposed as a linear combination of a set of N_m time series, denoted by $\tilde{\boldsymbol{y}}_i^{(m)}, 1 \leq i \leq N_m$. Assume also that each of these time series is a linear and/or nonlinear combination of a subset of delayed versions of the acquired iEEG signals $\{\boldsymbol{y}_k^{\tau_k}\}_{\forall k \in \Omega_i^{(m)}, \forall \tau_k \in \Phi_i^{(m)}}$ where the indices of these time series and their related time lags are defined in the sets $\Omega_i^{(m)}$ and $\Phi_i^{(m)}$, respectively. Then, we write:

$$\begin{aligned} \boldsymbol{y}_m &= \sum_{i=1}^{N_m} \alpha_i^{(m)} \tilde{\boldsymbol{y}}_i^{(m)} + \boldsymbol{w}_m \\ \boldsymbol{y}_m &= \sum_{i=1}^{N_m} \alpha_i^{(m)} f_i^{(m)}(\{\boldsymbol{y}_k^{\tau_k}\}_{\forall k \in \Omega_i^{(m)}, \forall \tau_k \in \Phi_i^{(m)}}) + \boldsymbol{w}_m \end{aligned} \quad (1)$$

where $f_i^{(m)}$ is the i-th unknown linear or nonlinear function, $\alpha_i^{(m)}$ is the i-th decomposition coefficient and $\boldsymbol{w}_m$ is the model residual related to $\boldsymbol{y}_m$. Understanding linear/nonlinear interactions among brain regions can be summed up to (i) the identification of the set of signals $\{\boldsymbol{y}_k^{\tau_k}\}_{\forall k \in \Omega_i^{(m)}, \forall \tau_k \in \Phi_i^{(m)}}$, (ii) the estimation of the N_m functions $f_i^{(m)}$ and (iii) the identification of the coefficients vector $\boldsymbol{\alpha}_m = [\alpha_1^{(m)}, \cdots, \alpha_{N_m}^{(m)}]^\intercal$ associated to $\boldsymbol{y}_m$. A compact representation of the aforementioned decomposition problem is expressed as follows:

$$\begin{aligned} \boldsymbol{y}_m &= \boldsymbol{D}_m \boldsymbol{\alpha}_m + \boldsymbol{w}_m, \quad \forall m \in \{1, \cdots, M\} \\ &= \boldsymbol{D}\boldsymbol{\Pi}\boldsymbol{\Pi}^{-1}\boldsymbol{\theta}_m + \boldsymbol{w}_m \end{aligned} \quad (2)$$

where $\boldsymbol{D}_m = \boldsymbol{D}\boldsymbol{\Pi}$ is a matrix collecting the M times series constituting the signal $\boldsymbol{y}_m$. These times series stand for the most relevant candidates that can be selected from a predefined dictionary $\boldsymbol{D} \in \mathbb{R}^{T \times N}$ using a selection matrix $\boldsymbol{\Pi}$, N being the total number of candidates. This predefined dictionary encodes all possible time series candidates (comprising possible linear and/or nonlinear functions) and $\boldsymbol{\theta}_m \in \mathbb{R}^N$ is a coefficient vector. The matrix $\boldsymbol{\Pi}$ is binary with exactly one entry of 1 in each row and each column. More particularly, as initially suggested in [6], the most relevant candidates required to fit properly the signal $\boldsymbol{y}_m$, up to an ERR criterion [6], are found using an OLS scheme combined with a threshold-based candidate selection criterion [5] [6]. To this end, the matrix $\boldsymbol{D}$ is decomposed as $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{W}$ where $\boldsymbol{U} \in \mathbb{R}^{T \times N}$ and $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ are orthogonal and upper triangular matrices, respectively. For the sake of readability, the subscript m will be dropped from now on keeping in mind that the m-th, $m \in \{1, \cdots, M\}$, signal $\boldsymbol{y}_m$ is being processed. This leads to rewrite equation (2) as follows:

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\theta} = \boldsymbol{U}\tilde{\boldsymbol{\theta}} = \sum_{n=1}^{N} \tilde{\theta}_n \boldsymbol{u}_n \quad (3)$$

where $\tilde{\boldsymbol{\theta}} = \boldsymbol{W}\boldsymbol{\theta}$, $\boldsymbol{u}_n$ is the n-th column of $\boldsymbol{U}$ and $\tilde{\theta}_n$ stands for the n-th component of the coefficients vector $\tilde{\boldsymbol{\theta}}$. Then, decomposing $\boldsymbol{y}$ requires the identification of a subset $\Gamma = \{\boldsymbol{u}_{k_\ell}\}_{k_\ell \in \{1 \cdots, N\}, \ell \in \{1, \cdots, N_m\}}$ of the most N_m relevant column vectors of $\boldsymbol{U}$ contributing to $\boldsymbol{y}$ together with their corresponding coefficients $\tilde{\theta}_\ell, 1 \leq \ell \leq N_m$. The elements of Γ are found successively according to their contribution (from the highest to the lowest) to $\boldsymbol{y}$ [5]–[9]. To this end, for the sake of convenience, let us define $\boldsymbol{D}^{-(0)} = \boldsymbol{D}$ as the initial dictionary matrix that is used to estimate the first relevant vector, $\boldsymbol{u}_{k_1}$, in Γ. Then, the matrix $\boldsymbol{D}^{-(k_i-1)} \in \mathbb{R}^{T \times N-k_i+1}$ is a reduced dictionary matrix to be used to estimate $\boldsymbol{u}_{k_i}, k_i > 1$. The matrix $\boldsymbol{D}^{-(k_i-1)}$ is obtained by excluding one column vector from $\boldsymbol{D}^{-(k_i-2)}$. The excluded column vector in $\boldsymbol{D}^{-(k_i-2)}$ stands for the most relevant candidate model defining the vector $\boldsymbol{u}_{k_i-1}$. To find this most relevant column vector in $\boldsymbol{D}^{-(k_i-1)}$, a grid search over the columns of $\boldsymbol{D}^{-(k_i-1)}$ is applied. More precisely, let $\tilde{\boldsymbol{U}}_{k_i} = [\boldsymbol{u}_{k_i}^1, \cdots, \boldsymbol{u}_{k_i}^{N-k_i+1}] \in \mathbb{R}^{T \times N-k_i+1}$ be defined as

$$\tilde{\boldsymbol{U}}_{k_i} = \boldsymbol{D}^{-(k_i-1)} - \boldsymbol{H}_{k_i}\tilde{\boldsymbol{U}}_{k_i-1} \quad (4)$$

where $\tilde{\boldsymbol{U}}_{k_i-1} = \boldsymbol{u}_{k_i-1}\boldsymbol{1}^\intercal{}_{N-k_i+1}$ and $\boldsymbol{H} \in \mathbb{R}^{N-k_i+1 \times N-k_i+1}$ is a diagonal matrix that can be obtained by solving the following optimization problem:

$$\boldsymbol{H}_{k_i}^* = \arg\min_{\boldsymbol{H}_{k_i}} ||\boldsymbol{D}^{-(k_i-1)} - \tilde{\boldsymbol{U}}_{k_i-1}\boldsymbol{H}_{k_i}||_F^2 \; s.t. H_{k_i,\, i,j \atop \forall i \neq j} = 0 \quad (5)$$

where $H_{k_i,i,j}$ is the (i,j)-th entry of $\boldsymbol{H}_{k_i}$ and $\boldsymbol{1}_N$ is a N-dimensional column vector of ones. Once the vector $\boldsymbol{u}_{k_i}$ is estimated, the vector $\tilde{\boldsymbol{\theta}}_{k_i}$ is computed also in a least squares sense:

$$\tilde{\boldsymbol{\theta}}_{k_i}^* = \arg\min_{\tilde{\boldsymbol{\theta}}_{k_i}} ||\boldsymbol{y} - \tilde{\boldsymbol{U}}_{k_i}\tilde{\boldsymbol{\theta}}_{k_i}||_2^2 \quad (6)$$

Then, the $(N-k_i+1)$-dimensional ERR vector, denoted here by $\boldsymbol{e}$, is defined by:

$$\boldsymbol{e}_{k_i} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\tilde{\boldsymbol{\theta}}_{k_i}^{\odot 2} \quad (7)$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix whose main diagonal is the vector $[||\boldsymbol{u}_{k_i}^1||_2^2, \cdots, ||\boldsymbol{u}_{N-k_i+1}^1||_2^2]$, $\boldsymbol{\Psi} = \frac{1}{||\boldsymbol{y}||_2^2}\boldsymbol{I}_{N-k_i+1}$, $\odot$ stands for the Hadamard product (element-wise matrix product, $\tilde{\boldsymbol{\theta}}_{k_i}^{\odot 2} = \tilde{\boldsymbol{\theta}} \odot \tilde{\boldsymbol{\theta}}$ and $\boldsymbol{I}_K$ is a $(K \times K)$ identity matrix. Note that the ℓ-th component, e_ℓ, $1 \leq \ell \leq N - k_i + 1$, of the vector $\boldsymbol{e}_{k_i}$ quantifies the contribution strength of the ℓ-th candidate model, $\boldsymbol{d}_\ell^{-(k_i-1)} \in \mathbb{R}^T$, in the current dictionary $\boldsymbol{D}^{-(k_i-1)}$. Once the $N - k_i + 1$ ERR values are computed, the index of the highest ERR value, $e_{max}^{(k_i)}$, in the vector $\boldsymbol{e}_{k_i}$ refers to the position of the most relevant candidate in $\boldsymbol{D}^{-(k_i-1)}$. The above mentioned steps are repeated until N_m candidate models are selected and for which the inequality $1 - \sum_{i=1}^{N_m} e_{\text{max}}^{(i)} < \epsilon$, where ϵ is a predefined threshold chosen heuristically, becomes true. Let us now define $\boldsymbol{D}_1 \in \mathbb{R}^{T \times N_m}$ as the dictionary collecting the N_m retained column vectors of the initial dictionary $\boldsymbol{D} \in \mathbb{R}^{T \times N}$. Then, in order to avoid some spurious retained models in $\boldsymbol{D}_1$ that could be raised due to the choice of the threshold ϵ, we propose to refine the obtained dictionary $\boldsymbol{D}_1$. To this end, we assume that, among all retained models, few of them are relevant to reconstruct the signal $\boldsymbol{y}$. This formally leads to consider a sparse representation of the

 ISBN: 978-1-68558-057-5

coefficient vector, $\boldsymbol{\theta}$. Then, the refined representation of $\boldsymbol{y}$ can be obtained by solving the following optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{\lambda}{2}||\boldsymbol{y}-\boldsymbol{x}||_2^2 + ||\boldsymbol{z}||_1 \text{ s.t. } \boldsymbol{x} = \boldsymbol{D}_1\boldsymbol{\theta} \text{ and } \boldsymbol{z} = \boldsymbol{\theta} \tag{8}$$

where λ is a regularization parameter and $||.||_1$ is the $L1$-norm. Such optimization problem can be solved using the PALM method [10]. The choice of the PALM method is justified by its good convergence properties [11]. PALM minimizes the augmented Lagrangian function associated to (8) given by:

$$\begin{aligned}\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{g},\lambda) = \frac{\lambda}{2}||\boldsymbol{y}-\boldsymbol{x}||_2^2 + ||\boldsymbol{z}||_1 + \frac{\rho_1}{2}||\boldsymbol{\theta}-\boldsymbol{z}||_2^2 \\ +\boldsymbol{v}^\intercal(\boldsymbol{\theta}-\boldsymbol{z}) + \frac{\rho_2}{2}||\boldsymbol{D}_1\boldsymbol{\theta}-\boldsymbol{x}||_2^2 + \boldsymbol{g}^\intercal(\boldsymbol{D}_1\boldsymbol{\theta}-\boldsymbol{x})\end{aligned} \tag{9}$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ are auxiliary variables, $\boldsymbol{v}$ and $\boldsymbol{g}$ stand for the Lagrange multipliers and $\rho_1, \rho_2 \in \mathbb{R}_+^*$. The update rules of variables $\boldsymbol{\theta}$ and $\boldsymbol{x}$ are computed by looking for the stationary points of $\mathcal{L}$ in these two variables. This leads to :

$$\boldsymbol{\theta} = (\rho_1\boldsymbol{I}_N + \rho_2\boldsymbol{D}_1{}^\intercal\boldsymbol{D}_1)^{-1}(\boldsymbol{v} + \rho_1\boldsymbol{z} + \boldsymbol{D}_1{}^\intercal(\rho_2\boldsymbol{x} - \boldsymbol{g})) \tag{10}$$

$$\boldsymbol{x} = \frac{\lambda\boldsymbol{y} + \boldsymbol{g} + \rho_2\boldsymbol{D}_1\boldsymbol{\theta}}{\lambda + \rho_2} \tag{11}$$

As far as the Lagrangian multipliers $\boldsymbol{v}$ and $\boldsymbol{g}$ are concerned, they are updated through a gradient-ascent scheme as follows:

$$\Delta\boldsymbol{v} = \rho_1(\boldsymbol{\theta}-\boldsymbol{z}), \quad \Delta\boldsymbol{g} = \rho_2(\boldsymbol{D}_1\boldsymbol{\theta}-\boldsymbol{x}) \tag{12}$$

where $\Delta\boldsymbol{v} = \boldsymbol{v}_{i+1} - \boldsymbol{v}_i$ and $\Delta\boldsymbol{g} = \boldsymbol{g}_{i+1} - \boldsymbol{g}_i$ (i represents the iteration index). Besides, the update rule of the dual variable $\boldsymbol{z}$ is performed by:

$$\boldsymbol{z} = prox_{\phi,\lambda c_{\boldsymbol{z}}}\left(\boldsymbol{z} - \frac{1}{c_{\boldsymbol{z}}}\nabla_{\boldsymbol{z}}\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{g},\lambda)\right) \tag{13}$$

where $\nabla_{\boldsymbol{z}}\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{g},\lambda) = \frac{\partial\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{g},\lambda)}{\partial\boldsymbol{z}} = (-\boldsymbol{v} - \rho_1(\boldsymbol{\theta}-\boldsymbol{z}))$, $c_{\boldsymbol{z}} \in \mathbb{R}$ is the step-size, $\text{prox}_{\phi,\lambda c_{\boldsymbol{z}}}$ is a proximal operator dealing with the non-smooth function (here $\phi = \|.\|_1$) and initially proposed in [12] and $\lambda c_{\boldsymbol{z}}$ denotes the shrinking threshold. Besides, as the proximal operator defined in equation (13) relies mainly on a gradient-descent scheme, the gradient learning step is a crucial parameter to be accounted for. According to [10], a wise choice of such parameter is $c_{\boldsymbol{z}} > L_{\boldsymbol{z}}(\boldsymbol{z})$ where $L_{\boldsymbol{z}}$ is the Lipschitz modulus verifying [10]:

$$\begin{aligned}||\nabla_{\boldsymbol{z}}\mathcal{L}(\boldsymbol{x},\boldsymbol{z}_{i-1},\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{g},\lambda) - \nabla_{\boldsymbol{z}}\mathcal{L}(\boldsymbol{x},\boldsymbol{z}_i,\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{g},\lambda)||_2 \leq \\ L_{\boldsymbol{z}}||\boldsymbol{z}_{i-1} - \boldsymbol{z}_i||_2\end{aligned} \tag{14}$$

where $\boldsymbol{z}_i$ is the estimate of the vector $\boldsymbol{z}$ at the i-th iteration. By substituting the expression of $\nabla_{\boldsymbol{z}}\mathcal{L}(\boldsymbol{x},\boldsymbol{z}_{i-1},\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{g},\lambda)$ in the above inequality, we get $L_{\boldsymbol{z}} \geq \rho_1$. This condition leads to define a lower bound on the value of the gradient learning step $c_{\boldsymbol{z}}$. More precisely, as suggested in [10], a good behavior of the PALM algorithm is guaranteed when the gradient learning step verifies $c_{\boldsymbol{z}} = \gamma_{\boldsymbol{z}}L_{\boldsymbol{z}}$ with $\gamma_{\boldsymbol{z}} > 1$. Thus, a lower bound on the gradient learning step is obtained by combining the obtained condition on the Lipschitz modulus with the given expression on $c_{\boldsymbol{z}}$. This leads to $c_{\boldsymbol{z}} \geq \gamma_{\boldsymbol{z}}\rho_1$. In the current study the equal part of the latter inequality is considered and then the parameter $\gamma_{\boldsymbol{z}}$ is tuned, while fixing the parameter ρ_1 to one, in such a way good estimation results are obtained.

As far as the regularization parameter λ is concerned, it is optimally computed, at each iteration, by means of the discrepancy principle. In fact, the latter principle states that the regularization parameter is laying in the set $\{\boldsymbol{x} : ||\boldsymbol{x}-\boldsymbol{y}||_2^2 \leq c\}$ where $c \in \mathbb{R}$ is a coefficient related to the noise variance [13] and can be obtained through the equivalent degree of freedom method [14] [15]. Then, by considering the equality part of the latter condition together with equation (11), the update rule of λ can be written as follows:

$$\lambda = \frac{||\rho_2(\boldsymbol{y} - \boldsymbol{D}_1\boldsymbol{\theta}) - \boldsymbol{g}||_2}{\sqrt{c}} - \rho_2 \tag{15}$$

At each iteration, equations (10), (11), (12), (13) and (15) are called alternatively where each variable is updated by fixing the other ones to their last estimate. The optimization process stops either when the relative estimation error on the parameter $\boldsymbol{\theta}$ exhibits a value that is smaller than (or equal to) a predefined threshold determined empirically, or when the maximum number of iterations is reached.

III. DATASET

The evaluation of the proposed approach is performed on both simulated and real iEEG signals.

A. Simulated iEEG signals

A 3-channel nonlinear model generating iEEG-like signals [16] is considered and defined hereafter:

$$\begin{aligned}y_1(k) &= 3.4y_1(k-1)(1-y_1^2(k-1))e^{-y_1^2(k-1)} + w_1(k) \\ y_2(k) &= 3.4y_2(k-1)(1-y_2^2(k-1))e^{-y_2^2(k-1)} - 0.5y_1^2(k-1) \\ &\quad + 0.25\sqrt{2}y_2(k-1) - 0.5y_3(k-3) + w_2(k) \\ y_3(k) &= 3.4y_3(k-1)(1-y_3^2(k-1))e^{-y_3^2(k-1)} - 0.5y_1^2(k-2) \\ &\quad - 0.5y_2(k-2) - 0.25\sqrt{2}y_3(k-2) + w_3(k)\end{aligned} \tag{16}$$

where $w_m \sim \mathcal{N}(0,1), 1 \leq m \leq 3$. The interest in such model is that it covers a variety of non-linearity types which is, to a large extent, in accordance with the nonlinear characteristic of the interactions between brain regions. In this study, the initial dictionary, denoted by $\boldsymbol{D}$, is defined as the collection of sixty candidates defined as follows:

- $\{f_i^{(m)}(\boldsymbol{y}_m^{\tau_m})\}_{\substack{1\leq i\leq 3\\1\leq m\leq 3}}, \forall\tau_m \in \{1,2,3\}$ is the set of their related time lags with $f_i^{(m)}(\boldsymbol{y}_m^{\tau_m}) = (\boldsymbol{y}_m^{\tau_m})^{\odot i}$.
- $\{f_i^{(m)}(e^{-\boldsymbol{y}_m^{\tau_m}})\}_{\substack{1\leq i\leq 3\\1\leq m\leq 3}}, \forall\tau_m \in \{1,2,3\}$.
- $\{f_{i_1}^{(m)}(\boldsymbol{y}_m^1)\}_{\substack{i_1\in\{1,3\}\\1\leq m\leq 3}} \times \{f_2^{(m)}(e^{-\boldsymbol{y}_m^1})\}_{\substack{1\leq i_2\leq 3\\1\leq m\leq 3}}$.

where a time period of four seconds of iEEG signals sampled at 256 Hz ($i.e.$, 1024 time samples) is simulated.

B. Real iEEG signals

Real iEEG signals were recorded in Rennes Hospital Epilepsy Unit in one female patient aged 35. In this patient who suffered from temporal lobe epilepsy, twelve intracerebral electrodes (10-15 contacts) were implanted in the left temporal, insular, inferior frontal and inferior parietal regions. From these recordings, a 64s-epoch, sampled at 256 Hz, was considered. Based on the clinician's expertise and according to preliminary clinical and electrophysiological examinations, we only kept the most interesting bipolar channels leading to a set of 12 channels. The objective is to classify these channels into three groups. The 'Onset' group (O) is a group where rapid discharges were observed by the clinician and therefore considered as the main regions responsible for the initiation of the seizure. The 'Propagation Sink' group (P_S) consists of channels that are majorly triggered by the Onset group, and considered as less involved in the triggering of the seizure. Finally, the 'Propagation Internal' group (P_I) consists of regions that can be triggered by other regions in the O group. Besides, this P_I group can be slightly involved in the seizure setting up through delayed electrical discharges with lower intensity compared to the ones of the O group. Consequently, this P_I group refers to less epileptogenic brain regions, and therefore considered as the one linking the most epileptogenic zones to those who are the less epileptogenic. According to the neuroscience expert, the most interesting time period to be considered corresponds to the onset of the ictal phase, $i.e.$, between the 18th and 22th seconds in the recording.

IV. Results

A. Simulated model

To assess the performance of the proposed approach on the estimation of the coefficients associated to the retained candidates, a mean squared error (MSE) criterion averaged over $K = 1000$ Monte-Carlo (MC) trials was computed for each simulated signal. The MSE related to the m-th channel is given by:

$$MSE^{(m)} = \frac{1}{K}\sum_{k=1}^{K} ||\boldsymbol{y}^{(m)} - \hat{\boldsymbol{y}}_k^{(m)}||_2^2 \qquad (17)$$

$\forall m \in \{1, \cdots, M\}$ where $\hat{\boldsymbol{y}}_k^{(m)}$ is the estimate of $\boldsymbol{y}^{(m)}$ at the k-th trial. Obtained MSE results for both the original ERR-based method and the proposed rERR-based one are given in Table I. A higher performance of the proposed rERR-based method over the ERR-based one can be clearly noticed from this table.

More precisely, we can state from Table I that the proposed rERR-based solution provides around 18%, 11% and 40% improvement in the nonlinear identification quality of the simulated iEEG-like signals, y_1, y_2 and y_3 (16), respectively. Furthermore, the improvement in the obtained MSE standard deviation shows that the proposed rERR-based approach provides statistically more consistent system identification results. This fact is also confirmed through Figure 1 where a clear gap in the estimation quality between the two considered methods is to be stated in favor of the proposed rERR-based solution.

TABLE I
MSE±STD COMPUTED OVER $K = 1000$ MC TRIALS. CASE OF SIMULATED IEEG-LIKE SIGNALS.

	ERR-based method	***rERR-based method***
y_1	3.39 ± 0.22	$\mathbf{2.84}$ ± 0.18
y_2	6.51 ± 0.72	$\mathbf{5.81}$ ± 0.37
y_3	11.50 ± 2.34	$\mathbf{7.70}$ ± 0.54

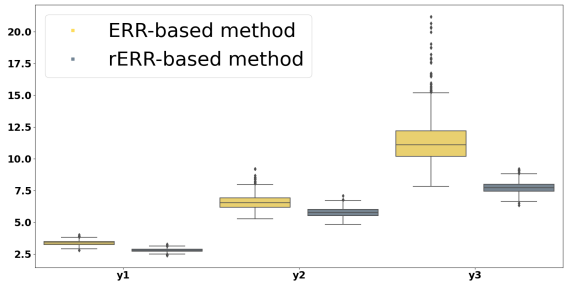

Fig. 1. MSE box-plots for the three simulated signals described in (16).

B. Real iEEG signals

In this study, each real iEEG signal, $\boldsymbol{y}_m, 1 \leq m \leq M$ (where $M = 12$) is assigned to either the O, P_I or P_S group using a defined threshold ϕ_{th}:

$$\phi_{th} = \frac{1}{4M}\sum_{m=1}^{M} |\phi_m| \qquad (18)$$

where ϕ_m is defined as follows:

$$\phi_m = \frac{OD_m - ID_m}{OD_m + ID_m} \qquad (19)$$

with OD_m and ID_m stand respectively for the outward and the inward degrees of the m-th signal (node) in the estimated brain network. More precisely, let $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_M] \in \mathbb{R}^{M\times M}$ be the adjacency matrix associated to the directed graph associated to the estimated brain network. Then, we have [17]:

$$OD_m = \sum_{i=1}^{M} \Theta_{m,i} \quad , \quad ID_m = \sum_{i=1}^{M} \Theta_{i,m} \qquad (20)$$

where $\Theta_{m,i}$ denotes the (m,i)-th entry of $\boldsymbol{\Theta}$. It is noteworthy that the adjacency matrix associated to a directed graph is a square asymmetric matrix ($i.e.$, $\Theta_{i,j} \neq \Theta_{j,i}$). Thus, the classification rule for a given signal $\boldsymbol{y}_m$ is defined by:

$$\boldsymbol{y}_m \in \begin{cases} O, & \text{if } \phi_{\geq}\phi_{th} \\ P_I, & \text{if } -\phi_{th} \leq \phi_m \leq \phi_{th} \\ P_S, & \text{if } \phi_m \leq -\phi_{th} \end{cases} \qquad (21)$$

Table II shows the expert's classification of the 12 iEEG channels. In addition, obtained classification results using both the original ERR-based method and the proposed rERR-based one are reported in Table III and Table IV respectively. Note that the two considered methods were tested on the seizure collected from the epileptic patient on the $[18s; 22s]$ time interval.

TABLE II
EXPERT'S CLASSIFICATION OF THE IEEG CHANNELS.

Expert	*Classification*	Expert	*Classification*
$Bp1$-$Bp2$	O	$Cp4$-$Cp5$	P_I
$Cp1$-$Cp2$	O	$Ap6$-$Ap7$	P_I
$Ap2$-$Ap3$	O	$Bp6$-$Bp7$	P_I
$Pp1$-$Pp2$	O	$Fp1$-$Fp2$	P_S
$Pp4$-$Pp5$	O	$Dp1$-$Dp2$	P_S
$Pp8$-$Pp9$	O	$Tp1$-$Tp2$	P_S

TABLE III
CLASSIFICATION OF REAL EPILEPTIC IEEG SIGNALS USING THE ORIGINAL ERR-BASED METHOD.

ERR-based method	*Classification*	ERR-based method	*Classification*
$Bp1$-$Bp2$	P_S	$Cp4$-$Cp5$	P_S
$Cp1$-$Cp2$	P_I	$Ap6$-$Ap7$	O
$Ap2$-$Ap3$	O	$Bp6$-$Bp7$	O
$Pp1$-$Pp2$	P_I	$Fp1$-$Fp2$	P_S
$Pp4$-$Pp5$	O	$Dp1$-$Dp2$	P_S
$Pp8$-$Pp9$	P_I	$Tp1$-$Tp2$	P_S

TABLE IV
CLASSIFICATION OF REAL EPILEPTIC IEEG SIGNALS USING THE RERR-BASED METHOD.

rERR-based method	*Classification*	rERR-based method	*Classification*
$Bp1$-$Bp2$	P_S	$Cp4$-$Cp5$	P_S
$Cp1$-$Cp2$	O	$Ap6$-$Ap7$	O
$Ap2$-$Ap3$	O	$Bp6$-$Bp7$	O
$Pp1$-$Pp2$	O	$Fp1$-$Fp2$	P_S
$Pp4$-$Pp5$	O	$Dp1$-$Dp2$	P_S
$Pp8$-$Pp9$	P_S	$Tp1$-$Tp2$	P_S

From Tables III and IV, we observe that both methods were able to correctly classify $Ap2$-$Ap3$ and $Pp4$-$Pp5$ in the O group. Besides, ERR and rERR were able to group properly all the P_S channels. Moreover, according to the expert, $Pp8-Pp9$ showed a delayed discharge, which can explain that it was classified in the P_I / P_S groups using both algorithms. As for $Ap6$-$Ap7$, it showed a rapid discharge at the onset of the seizure, which may explain its classification by the two algorithms. Now, the proposed rERR-based method outperforms the original one in the classification of $Cp1$-$Cp2$ and $Pp1$-$Pp2$ channels, in accordance with the expert's opinion. To conclude, following the expert's classification, the rERR-based approach appears attractive and more reliable in the identification of brain regions involved in the seizure onset, which is a crucial point from a therapeutic point of view.

V. CONCLUSION

In this paper, a refined ERR-based solution for nonlinear system identification problem was proposed with application to epilepsy. More precisely, the proposed solution handles the issue of the overestimation of the number of candidates required to decompose the signal at hand, which is a commonly encountered issue in the original ERR-based approach. The proposed solution relies on the assumption of a sparse representation of the model coefficient vector that the ERR-based approach provides. The defined optimization problem was solved in the proximal optimization framework using the well-known PALM algorithm combined with an optimal computation of the regularization parameter at each iteration. Numerical experiments on simulated iEEG-like and real epileptic iEEG signals showed clearly a higher system identification quality of the proposed approach compared to the original ERR-based one.

REFERENCES

[1] S. L. Moshé, E. Perucca, P. Ryvlin, and T. Tomson, "Epilepsy: new advances," *The Lancet*, vol. 385, no. 9971, pp. 884–898, 2015.

[2] K. Friston, "Functional and effective connectivity: A review," *Brain connectivity*, vol. 1, pp. 13–36, 01 2011.

[3] A. N. Almeida, V. Martinez, and W. Feindel, "The first case of invasive EEG monitoring for the surgical treatment of epilepsy: historical significance and context," *Epilepsia*, vol. 46, no. 7, pp. 1082–1085, 2005.

[4] W. Penfield and T. C. Erickson, *Epilepsy and cerebral localization*. Charles C. Thomas, 1941.

[5] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their applications to non-linear system identification," Tech. Rep. 2-30, Sheffield University, 1989.

[6] S. A. Billings, M. J. Korenberg, and S. Chen, "Identification of nonlinear output-affine systems using an orthogonal least squares algorithm," Tech. Rep. 1-11, Sheffield University, 1987.

[7] Y. Zhao, S. A. Billings, H. Wei, and P. G. Sarrigiannis, "Tracking time-varying causality and directionality of information flow using an error reduction ratio test with applications to electroencephalography data," *Physical Review E*, vol. 86, no. 5, pp. 1–11, 2012.

[8] P. G. Sarrigiannis *et al.*, "Quantitative eeg analysis using error reduction ratio-causality test; validation on simulated and real eeg data," *Clinical Neurophysiology*, vol. 125, pp. 32–46, 2014.

[9] Y. Zhao *et al.*, "Imaging of nonlinear and dynamic functional brain connectivity based on eeg recordings with the application on the diagnosis of alzheimer's disease," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1571–1581, 2020.

[10] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, pp. 459–494, 2014.

[11] R. Shefi and M. Teboulle, "On the rate of convergence of the proximal alternating linearized minimization algorithm for convex problems," *EURO Journal on Computational Optimization*, vol. 4, no. 1, pp. 27–46, 2016.

[12] L. Ding, "Reconstructing cortical current density by exploring sparseness in the transform domain," *Physics in Medicine and Biology*, vol. 54, pp. 2683 – 2697, 2009.

[13] K. El Houari *et al.*, "Investigating transmembrane current source formulation for solving the ecg inverse problem," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 371–375, 2018.

[14] N. Galatsanos and A. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.

[15] C. He, C. Hu, W. Zhang, and B. Shi, "A fast adaptive parameter estimation for total variation image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 4954–4967, 2014.

[16] C. Mahjoub, J. Bellanger, A. Kachouri, and R. Le Bouquin Jeannès, "On the performance of temporal Granger causality measurements on time series: a comparative study," *Springer Nature*, vol. 14, no. 3, pp. 1–9, 2020.

[17] N. Biggs, E. K. Lloyd, and R. J. Wilson, *Graph Theory, 1736-1936*. Oxford University Press, 1986.

 ISBN: 978-1-68558-057-5