DLGC: Dictionary Learning based Granger Causal Discovery for Cortico-muscular Coupling

Farwa Abbas
Imperial College London
f.abbas20@imperial.ac.uk

Verity McClelland
King's College London
verity.mcclelland@kcl.ac.uk

Zoran Cvetkovic
King's College London
zoran.cvetkovic@kcl.ac.uk

Wei Dai
Imperial College London
wei.dai1@imperial.ac.uk

***Abstract*—Investigating causal pathways between the brain and muscles is crucial for identifying biomarkers associated with movement disorders such as Parkinson's disease, multiple sclerosis, and dystonia. The transmission of information from the brain to muscles is complicated by the background activities and various forms of noise and interference. Thus, extracting meaningful causal patterns from recorded sensor data presents a significant challenge. This paper presents a novel approach to disentangling causal information from physiological signals while addressing measurement noise and other forms of perturbations. We propose a dictionary learning based autoregressive model capable of extracting meaningful features from physiological signals while simultaneously eliminating noise. To efficiently solve the non-convex, non-smooth optimization problem inherent in our approach, we employ a recent second-order proximal algorithm that leverages a local surrogate function of the objective function to converge to a potentially better local minimum. Our experimental results, conducted on real physiological signals, demonstrate the effectiveness of our proposed method in disentangling causal information and mitigating noise, thereby advancing our understanding of brain-muscle interactions in movement control.**

***Index Terms*—Dictionary learning, autoregression, causality**

I. Introduction

The analysis of causal interactions between simultaneously recorded physiological signals like EEG (electroencephalography) and sEMG (surface electromyography) is pivotal to understanding sensorimotor control. The neurological disorders affecting movement, primarily stem from impairments or disruptions in certain neural pathways linking the brain and the muscles. The primary motor cortex, a central brain region essential for acquiring and executing precise motor skills, holds valuable insights in this regard. Therefore, it can act as a predictor for forecasting motor behavior by examining past patterns of brain activity. Numerous studies have examined this, by analyzing concurrent EEG recordings from the primary motor cortex and sEMG signals from limb muscles to uncover insights into corticomuscular coupling. This can help reveal neural-muscular coordination and motor control mechanisms in both healthy and pathological conditions [1]–[5].

Researchers investigating corticomuscular coupling commonly employ various metrics, some widely used among them are coherence [4], [5] and Granger causality (GC) [6], [7]. Corticomuscular coherence (CMC) measures synchronization across different frequency bands within signals. The authors in [3] proposed a dictionary learning approach for sparse signal representation to estimate CMC. Although dictionary learning can efficiently reconstruct signals using a few high-amplitude vectors to depict sensory-motor components, presumably extracting coherent components, it turns out that sparse representation of signals alone remains insufficient to achieve a substantial increase in CMC levels. Hence, the authors proposed a hand-crafted greedy algorithm for further coherent component selection [3]. Moreover, while CMC serves as a valuable metric for extracting interactions in the frequency domain, it does not indicate the directionality of these interactions. To account for the directional causal information missed by CMC, Granger causality offers a simple yet effective measure, based on an AutoRegressive (AR) model. While Granger causality can provide valuable insights into temporal relationships based on statistical associations, further research and analysis are often necessary to establish causality definitively. Besides, there are several stringent assumptions, outlined in detail by [8], that must be met for the autoregressive models to be considered an appropriate framework for identifying Granger causal relationships.

One of the assumptions made by autoregressive models is *perfect observation* of all variables without any measurement errors [8]. This assumption does not hold true for physiological signals like EEG and sEMG, due to several experimental and environmental factors [8], [9]. Moreover, standard time-series denoising methods may not effectively handle the unique correlation structure of the autoregressive processes [10], due to which denoising such processes is a challenging problem.

In order to deal with this challenge, one line of research suggests introducing structural assumptions on the innovation process within the autoregressive model to yield robust estimates [11]. This allows for the integration of theoretical or empirical knowledge about the system, facilitating noise identification and filtration from the underlying signal. Inspired by these advances, we propose a novel method to handle measurement noise in the signal by exploiting the representative power of a learned dictionary in an autoregressive framework. Typically, the AR model utilizes an autoregressive coefficient matrix to capture the temporal dependencies between the current value of a signal and its past values. Instead of using a prefixed structure, we propose to find a sparse representation of autoregressive model parameters, in the form of coefficients with respect to a learned dictionary. To solve this problem, traditionally, an alternating minimization for the dictionary learning and sparse coding is employed based on greedy

algorithms such as Orthogonal Matching Pursuit (OMP) or their convex relaxation such as Basis Pursuit (BP). The convex version of sparse coding can be efficiently solved using the Proximal Gradient (PG) method [12]. However, these methods are sensitive to noise, as they make decisions based solely on the observed signal without considering the underlying noise structure. This can result in poor performance and inaccurate recovery of sparse signals, especially in noisy environments [13]. To overcome these limitations, we propose to use a recently proposed modified dogleg path algorithm called Proximal Dogleg Opportunistic Majorization (PDOM) [14], that enables a more comprehensive exploration of the solution space by dynamically adjusting the step size and direction based on local curvature information of the objective function. This allows for a more efficient and effective exploration of the optimization landscape compared to the sequential decision-making process of OMP and BP. Moreover, the local surrogate function can provide an accurate approximation of the optimization landscape, making the optimization process more robust to noise in the data [15]. This, in turn, will improve the accuracy and reliability of sparse signal recovery and feature selection tasks.

In the AR model setting, the innovation process term embodies the process noise linked with the inherent dynamics of the system under study. However, if our data contains significant measurement noise, the problem becomes ill-posed, as distinguishing between these two becomes challenging. To address this issue, we propose leveraging empirical knowledge about the recorded signals to create a realistic model that incorporates measurement noise. In controlled experimental setups with high-quality EEG equipment, where noise characteristics remain relatively stable over time, assuming stationary Gaussian noise to model measurement noise is reasonable. This type of Gaussian noise can be separated using various filtering techniques such as Kalman filtering [16], Wiener filtering [17], etc. Conversely, the innovation process modeling the physiological noise may exhibit non-Gaussian trends such as skewness or sudden bursts of very large amplitudes at random intervals. Therefore, to accommodate this, rather than assuming Gaussianity for the innovation process sequence, we suggest assuming Gaussianity for the recorded measurement noise. Meanwhile, the innovation process term containing non-Gaussian trends is presumed to be sparse in the Wavelet basis. This assumption provides a practical method for modeling real signals, as supported by previous research in [18] and [19], and is expected to yield enhanced causality estimates.

In essence, we propose to take into account the effect of measurement noise on recorded signals by employing realistic structural assumptions to regularize the problem. Our proposed algorithm, Dictionary Learning based Granger Causality (DLGC), alternatively minimizes the objective function to learn the feature representations while separating noise. A gradient-descent based dictionary learning method is employed to successively update the dictionary atoms, while sparse coefficients are estimated efficiently using PDOM [14]. We propose to use a Wiener filter for noise reduction since it is optimal in the mean squared error sense, meaning that it minimizes the expected value of the squared error signal. This results in a potentially high-quality output that preserves the signal features and reduces the noise effects [20]. The eventual goal of this research is to disentangle noise from the underlying signal to accurately model the recorded data, consequently improving the sensitivity of the inferred Granger causality. We make the following contributions, in particular:

1) We propose a novel dictionary learning based autoregressive model, aiming to enhance the estimation of the underlying feature representation of the signals.
2) A unique approach to handle the measurement noise in autoregressive models is developed, to disentangle different types of perturbations from the recorded signal.
3) A unified algorithm is developed to efficiently solve the proposed optimization problem by estimating the model parameters and the underlying dictionary while separating the unwanted noisy components in the signals.

II. Problem Formulation

In autoregressive modeling, the EEG signal can be represented as a linear combination of past sEMG activity, previous brain activity, and innovation process driven by random factors. Similarly, the sEMG signal can be expressed in a comparable manner, with the EEG signal used as a reference to establish bidirectional causality. The innovation process (ϵ) in the Vector AutoRegressive (VAR) model typically accounts for unexplained variability in the signal, encompassing various sources of interference and physiological noise.

Consider $y(t)$ is the EEG signal and $x(t)$ is the corresponding sEMG signal at time t, assuming the observed data does not involve measurement noise. They can be represented as:

$$y(t) = \sum_{j=1}^{m} a_{yy}^{(j)} y(t-j) + \sum_{j=1}^{m} a_{yx}^{(j)} x(t-j) + \epsilon_y(t), \quad (1)$$

$$x(t) = \sum_{j=1}^{m} a_{xy}^{(j)} y(t-j) + \sum_{j=1}^{m} a_{xx}^{(j)} x(t-j) + \epsilon_x(t), \quad (2)$$

where $\epsilon_y(t)$ and $\epsilon_x(t)$ are time-dependent innovation processes and m indicates the dependency on past instants or order of the autoregressive model. The objective is to fit an autoregressive model to estimate parameters $a_{yy}^{(j)}, a_{yx}^{(j)}, a_{xy}^{(j)}, a_{xx}^{(j)}$ $\forall j = 1, \cdots, m$ where m is predetermined. After estimating the parameters, Granger causality can be estimated using an F-test [8] comparing the *restricted model* (without cross-coupling) and *unrestricted model* (with cross-coupling) as:

$$\mathrm{GC}_{x \to y} = \frac{\mathrm{RSS}_{y_{\mathrm{res}}} - \mathrm{RSS}_{y_{\mathrm{unr}}} / (p - p')}{\mathrm{RSS}_{y_{\mathrm{unr}}} / (T - p)}, \quad (3)$$

$$\mathrm{GC}_{y \to x} = \frac{\mathrm{RSS}_{x_{\mathrm{res}}} - \mathrm{RSS}_{x_{\mathrm{unr}}} / (p - p')}{\mathrm{RSS}_{x_{\mathrm{unr}}} / (T - p)}, \quad (4)$$

where RSS means *residual sum of squares*, p and p' are the number of parameters in unrestricted and restricted models, respectively. This test shows that if the observed statistic F

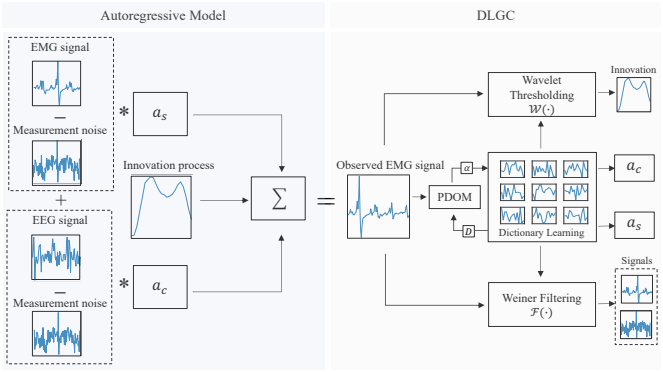

Fig. 1: Block diagram for our proposed method DLGC

exceeds nth quantile, say 95th, of an F-distribution with $p-p'$ and $T-p$ degrees of freedom then Granger causality exists.

III. Proposed Method

Consider $\boldsymbol{y} \in \mathbb{R}^T$ is the clean EEG signal and $\boldsymbol{x} \in \mathbb{R}^T$ is the corresponding clean sEMG signal, then the model in Equations (1), (2) can be rewritten in matrix form as follows:

$$\boldsymbol{y} = \boldsymbol{Y}\boldsymbol{a}_{yy} + \boldsymbol{X}\boldsymbol{a}_{yx} + \boldsymbol{\epsilon}_y, \tag{5}$$

$$\boldsymbol{x} = \boldsymbol{Y}\boldsymbol{a}_{xy} + \boldsymbol{X}\boldsymbol{a}_{xx} + \boldsymbol{\epsilon}_x, \tag{6}$$

where $\boldsymbol{\epsilon}_y, \boldsymbol{\epsilon}_x \in \mathbb{R}^{T\times 1}$, $\boldsymbol{a}_i \in \mathbb{R}^{m\times 1}$, $\forall i \in \{yy, yx, xy, xx\}$, $\boldsymbol{Y}, \boldsymbol{X} \in \mathbb{R}^{T\times m}$ represent the Toeplitz matrices associated with past values of $\boldsymbol{y}, \boldsymbol{x}$, respectively. To generalize the two equations above, we define the future values of the signal as $\boldsymbol{z} \in \mathbb{R}^T$ that can either be $\boldsymbol{y}$ or $\boldsymbol{x}$ i.e. $\boldsymbol{z} \in \{\boldsymbol{y}, \boldsymbol{x}\}$, and the above equations can be written, generically, as:

$$\boldsymbol{z} = \boldsymbol{Z}_s\boldsymbol{a}_s + \boldsymbol{Z}_c\boldsymbol{a}_c + \boldsymbol{\epsilon}_z, \tag{7}$$

where we refer to self-coupling coefficients $\boldsymbol{a}_{yy}, \boldsymbol{a}_{xx}$ as $\boldsymbol{a}_s$, cross-coupling coefficients $\boldsymbol{a}_{yx}, \boldsymbol{a}_{xy}$ as $\boldsymbol{a}_c$, and their corresponding Toeplitz matrices as $\boldsymbol{Z}_s, \boldsymbol{Z}_c$, respectively.

Now, we consider the measurement noise $\boldsymbol{n}_z \in \mathbb{R}^T$ in the observed noisy signal $\tilde{\boldsymbol{z}} \in \mathbb{R}^T$ as follows:

$$\boldsymbol{n}_z = \tilde{\boldsymbol{z}} - \boldsymbol{z} = \tilde{\boldsymbol{z}} - (\boldsymbol{Z}_s\boldsymbol{a}_s + \boldsymbol{Z}_c\boldsymbol{a}_c + \boldsymbol{\epsilon}_z),$$

where $\boldsymbol{Z}_s, \boldsymbol{Z}_c$ are constructed from past values of clean signals $\boldsymbol{z}_s, \boldsymbol{z}_c$ (self and cross signals based on the value of $\boldsymbol{z} \in \{\boldsymbol{y}, \boldsymbol{x}\}$), transformed by Toeplitz operator $\mathcal{T}(\cdot)$. Since we only have access to noisy observations, to get an estimate of the past values of the clean signals we employ Wiener filter $\mathcal{F}(\cdot)$ as follows:

$$\boldsymbol{Z}_s = \mathcal{T}(\boldsymbol{z}_s) \approx \mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_s)), \tag{8}$$

$$\boldsymbol{Z}_c = \mathcal{T}(\boldsymbol{z}_c) \approx \mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_c)). \tag{9}$$

With the notation introduced above, we formulate the optimization problem in the subsequent section. The overview of the optimization process using DLGC is shown in Figure 1.

A. Optimization Problem

Consider $\boldsymbol{z} \in \mathbb{R}^T$ is the EEG/sEMG signal, then we propose to minimize the non-convex, non-smooth optimization:

$$\begin{aligned}&\min \frac{1}{2}\|\tilde{\boldsymbol{z}} - (\mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_s))\boldsymbol{a}_s + \mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_c))\boldsymbol{a}_c + \boldsymbol{\epsilon}_z)\|_2^2 + \lambda\|\boldsymbol{\alpha}_c\|_0 \\ &+ \lambda\|\boldsymbol{\alpha}_s\|_0 + \gamma\|\mathcal{W}(\boldsymbol{\epsilon}_z)\|_1 \\ &\text{s.t. } \boldsymbol{a}_s = \boldsymbol{D}\boldsymbol{\alpha}_s, \ \ \boldsymbol{a}_c = \boldsymbol{D}\boldsymbol{\alpha}_c. \end{aligned} \tag{10}$$

where $\mathcal{W}(\cdot)$ indicates Wavelet transform operator to enforce Wavelet sparsity assumption on innovation process, $\boldsymbol{D}$ is the Dictionary used to represent the autoregressive process whereas $\boldsymbol{\alpha}_{s,(c)}$ denotes the corresponding sparse coefficients vector. The first term in the objective function is data fidelity term to enforce the autoregressive structure while simultaneously ensuring that the measurement noise in data is small. The optimization problem leverages ℓ_0 norm minimization for feature selection and model interpretability by identifying a small subset of significant components in $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_c$, while ℓ_1 norm minimization controls noise and prevents overfitting by shrinking small values to zero and penalizing large deviations. In unconstrained form using quadratic penalty terms, the objective function can be written as follows:

$$\begin{aligned}&\min \frac{1}{2}\|\tilde{\boldsymbol{z}} - (\mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_s))\boldsymbol{a}_s + \mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_c))\boldsymbol{a}_c + \boldsymbol{\epsilon}_z)\|_2^2 + \lambda\|\boldsymbol{\alpha}_c\|_0 \\ &+ \lambda\|\boldsymbol{\alpha}_s\|_0 + \frac{\rho}{2}\|\boldsymbol{a}_s - \boldsymbol{D}\boldsymbol{\alpha}_s\|_2^2 + \frac{\rho}{2}\|\boldsymbol{a}_c - \boldsymbol{D}\boldsymbol{\alpha}_c\|_2^2 + \gamma\|\mathcal{W}(\boldsymbol{\epsilon}_z)\|_1. \end{aligned}$$

To compute the update for $\boldsymbol{a}_s, \boldsymbol{a}_c$, the least square solution can be employed by using $\boldsymbol{Z}_s, \boldsymbol{Z}_c$ from equations (8) and (9).

$$\boldsymbol{a}_s = (\boldsymbol{Z}_s^T\boldsymbol{Z}_s + \rho\boldsymbol{I})^{-1}[\boldsymbol{Z}_s^T(\tilde{\boldsymbol{z}} - \boldsymbol{Z}_c\boldsymbol{a}_c - \boldsymbol{\epsilon}_z) + \rho\boldsymbol{D}\boldsymbol{\alpha}_s], \tag{11}$$

$$\boldsymbol{a}_c = (\boldsymbol{Z}_c^T\boldsymbol{Z}_c + \rho\boldsymbol{I})^{-1}[\boldsymbol{Z}_c^T(\tilde{\boldsymbol{z}} - \boldsymbol{Z}_s\boldsymbol{a}_s - \boldsymbol{\epsilon}_z) + \rho\boldsymbol{D}\boldsymbol{\alpha}_c]. \tag{12}$$

For the sake of brevity, we use a generic notation to represent either $\boldsymbol{a}_s$ or $\boldsymbol{a}_c$ i.e. $\boldsymbol{a}_{s,(c)} \in \{\boldsymbol{a}_s, \boldsymbol{a}_c\}$. The complete DLGC algorithm is shown in Algorithm 1. After estimating the parameters, the next step is to find the sparse coefficients for the dictionary $\boldsymbol{D}$ using PDOM. In Algorithm 1, L_q refers to the Lipschitz constant of $\nabla q(\boldsymbol{\alpha}_{s,(c)})$ and is lower bounded by the maximum eigenvalue of the Hessian of $q(\boldsymbol{\alpha}_{s,(c)})$. The proximal operator for ℓ_0 norm is employed to efficiently handle the non-convex part of the objective function.

$$\begin{aligned}\boldsymbol{\alpha}^k &= \text{prox}_{\tau h}(\boldsymbol{\alpha}^{k-1} - \tau\nabla q(\boldsymbol{\alpha}^{k-1})), \\ &:= \arg\min_{\boldsymbol{\alpha}} h(\boldsymbol{\alpha}) + \frac{1}{2\tau}\|\boldsymbol{\alpha} - (\boldsymbol{\alpha}^{k-1} - \tau\nabla q(\boldsymbol{\alpha}^{k-1}))\|_2^2.\end{aligned}$$

The modified dogleg path $\boldsymbol{p}(\cdot)$ defined in [14] efficiently explores the optimization landscape by dynamically adjusting the step size τ_β and direction based on the local curvature and gradient information provided by the surrogate function $m(\cdot,\cdot)$ as in [14]. The surrogate function using second order information, in turn, enables the optimization algorithm to make informed decisions about the direction and magnitude of each step. The estimate obtained by this search path is then compared against the Proximal Gradient (PG) [12] estimate to select the better one. The technical details regarding the complete PDOM algorithm and modified dogleg path can be found in [14]. The dictionary is updated next, by backpropagating the loss gradients by taking a gradient descent step to

minimize the loss $f(\boldsymbol{\alpha}^k_{s,(c)})$. Following that, the estimate for the innovation process is then computed by employing Wavelet

Algorithm 1 DLGC: Dictionary Learning for GC Esimation

Input: $\lambda, \gamma, \eta, \beta \in (1, 2], \tau \in (0, 1/L_q), \boldsymbol{z}, \boldsymbol{Z}_c, \boldsymbol{Z}_s$

Initialize $\boldsymbol{a}_c, \boldsymbol{a}_s, \boldsymbol{\alpha}_c, \boldsymbol{\alpha}_s, \boldsymbol{D}$

while stopping criterion not met **do**

Compute least square update for $\boldsymbol{a}^k_s, \boldsymbol{a}^k_c$ by (11), (12)

Compute updates for $\boldsymbol{\alpha}^k_s, \boldsymbol{\alpha}^k_c$ using PDOM as follows:

Set $h(\boldsymbol{\alpha}^k_{s,(c)}) = \lambda \|\boldsymbol{\alpha}^k_{s,(c)}\|_0$

Set $q(\boldsymbol{\alpha}^k_{s,(c)}) = \frac{\rho}{2}\|\boldsymbol{a}^k_{s,(c)} - \boldsymbol{D}^k\boldsymbol{\alpha}^k_{s,(c)}\|^2_2$

Set $f(\boldsymbol{\alpha}^k_{s,(c)}) = q(\boldsymbol{\alpha}^k_{s,(c)}) + h(\boldsymbol{\alpha}^k_{s,(c)})$

Set $\tau_\beta = -\|\boldsymbol{p}(\beta)\|^2 / \langle \nabla q(\boldsymbol{\alpha}^k_{s,(c)}), \boldsymbol{p}(\beta)\rangle$

Compute $\boldsymbol{\alpha}^k_{s,(c)} = \text{prox}_{\tau_\beta h}(\boldsymbol{\alpha}^{k-1}_{s,(c)} + \boldsymbol{p}(\beta^{k-1}))$ for β_{max} such that $m_\beta(\boldsymbol{\alpha}^k_{s,(c)}, \boldsymbol{\alpha}^{k-1}_{s,(c)}) \geq q(\boldsymbol{\alpha}^k_{s,(c)})$

Compute $\boldsymbol{\alpha}^{\text{PG}}_s = \text{prox}_{\tau h}(\boldsymbol{\alpha}^{k-1}_{s,(c)} - \tau \nabla q(\boldsymbol{\alpha}^{k-1}_{s,(c)}))$

if $f(\boldsymbol{\alpha}^k_{s,(c)}) \geq f(\boldsymbol{\alpha}^{\text{PG}}_{s,(c)})$ **then**

Set $\boldsymbol{\alpha}^k_{s,(c)} = \boldsymbol{\alpha}^{\text{PG}}_{s,(c)}$

end if

Backpropagate the loss gradients $\Delta \boldsymbol{d}^k = -\frac{\delta f(\boldsymbol{\alpha}^k_{s,(c)})}{\delta \boldsymbol{d}^k}$

Update dictionary atoms in $\boldsymbol{D}^k$ by $\boldsymbol{d}^k \leftarrow \boldsymbol{d}^{k-1} + \eta \Delta \boldsymbol{d}^k$

Compute $\boldsymbol{Z}_s = \mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_s))$, $\boldsymbol{Z}_c = \mathcal{T}(\mathcal{F}(\tilde{\boldsymbol{z}}_c))$

Compute $\boldsymbol{\epsilon}^k_z = \mathcal{W}^{-1}(\mathcal{S}_\gamma(\mathcal{W}(\tilde{\boldsymbol{z}} - \boldsymbol{Z}_s\boldsymbol{a}^k_s - \boldsymbol{Z}_c\boldsymbol{a}^k_c)))$

Compute denoised signal $\boldsymbol{z}^k = \boldsymbol{Z}_s\boldsymbol{a}^k_s + \boldsymbol{Z}_c\boldsymbol{a}^k_c + \boldsymbol{\epsilon}^k_z$

Compute noise $\boldsymbol{n}^k_z = \tilde{\boldsymbol{z}} - \boldsymbol{z}^k$

end while

Output: $\boldsymbol{D}, \boldsymbol{a}_c, \boldsymbol{a}_s, \boldsymbol{n}_z, \boldsymbol{\epsilon}_z$

thresholding operator. Eventually, the measurement noise is removed from the signal. All the above steps are repeated iteratively until convergence.

IV. Experimental Results

In this section, we perform a comparative study to investigate the performance of our proposed method DLGC against existing methods on real data. All experiments were performed in PyTorch 2.1.2 with CUDA version 11.2, GPU Tesla T4, 15 GB RAM, and Windows 11 operating system.

A. Results on Real Data

The proposed method is tested on physiological data collected from nine healthy subjects in a previously published study [1]. Subjects performed a controlled motor task, grasping a ruler between right thumb and index finger. An electromechanical tapper provided mechanical perturbations of lateral displacement to the ruler at pseudorandom intervals of 5.6 - 8.4 s. The experiment comprised 8 blocks of 25 trials. Bipolar scalp EEG was recorded over left sensorimotor cortex and sEMG was recorded from first dorsal interosseous. Data were sampled at 1024 Hz, amplified and band-pass filtered (0.5-100 Hz for EEG; 5-500 Hz for sEMG). Offline, data were divided into 5 s epochs (1.1 s pre- and 3.9s poststimulus).

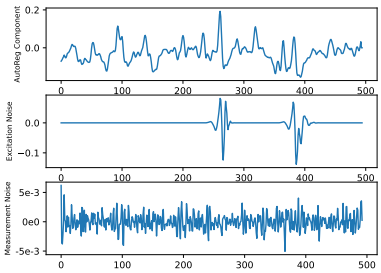

Fig. 2: EEG signal components for subject D

In our experimental settings, a Daubechies wavelet dictionary 'db4' is used to extract the innovation process, and γ is set to 70%. The level of wavelet decomposition is set at 3.

Fig. 3: ***First col:*** Blockwise GC, ***Second col:*** SS-ADMM, ***Third col:*** DLGC. ***Top to bottom:*** subjects B,D,G,K,L,M,N

We set the model order $m = 15$ and $T = 495$. The learning rate for dictionary $\boldsymbol{D}$ is $\eta = 10^{-3}$ and sparsity level $\lambda = 0.1$. It is important to acknowledge that our method comprises several parameters, and the adjustment of hyperparameters can be accomplished through cross-validation or alternative techniques to achieve optimal estimates. Figure 2 shows that our method successfully disentangles the observed signal into autoregressive component, the innovation process, and the measurement noise.

In Figure 3, we compare the performance of our proposed algorithm DLGC against the traditional Blockwise Granger causality [21], and recently proposed SS-ADMM [7]: Granger causality computed using all three algorithms is shown across the 5 second epoch for seven healthy subjects. An F-test applied with 95% significance level tests the null hypothesis that the first time series does not jointly Granger-cause the second. In each case, the DLGC algorithm detects higher levels of significant causality than the other methods.

We further compare the performance of the sparse coding using PDOM and traditional Proximal Gradient (PG) in Figure 4. With the same initialization and hyperparameter values, PDOM achieves a lower objective function value and a more sparse solution, hence a potentially better optimal point.

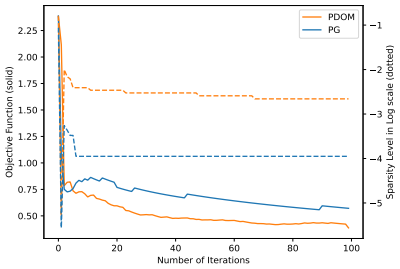

Fig. 4: Comparative performance analysis of PDOM and PG

V. Discussion

Our findings demonstrate that Granger causality detection, above critical value, has been substantially improved by the DLGC algorithm. In this active task we expect causality between brain and muscles in healthy subjects, both pre- and post-stimulus, but causality measured with blockwise GC or SS-ADMM in this study was often below or only transiently above the significance level. In contrast, DLGC enhanced the detection of bidirectional causality above the critical value both pre- and post-stimulus for most subjects. We propose that this improvement reflects the method's unique handling of noise in the signals. However, we lack ground truth Granger causality for recorded EEG and EMG data, limiting definitive validation of inferred causal relationships. This challenge is common in studies of biological signals due to complex underlying physiological interactions. Applying our approach to other physiological signals could also broaden its impact.

Acknowledgement

We thank Yiming Zhou for sharing the PDOM implementation and offering valuable suggestions for this project.

VI. Conclusion

In summary, we propose a novel approach that extends the autoregressive model by representing the model parameters as sparse coefficients with respect to a learned dictionary. The resulting algorithm employs an alternative minimization technique to effectively separate physiological effects, modeled through an innovation process, from measurement noise inherent in observed signals. Our findings demonstrate notable performance improvements in Granger causality estimation when compared to existing methodologies.

References

[1] V. M. McClelland, Z. Cvetkovic, and K. R. Mills, "Modulation of corticomuscular coherence by peripheral stimuli," *Experimental brain research*, vol. 219, pp. 275–292, 2012.

[2] V. M. McClelland and J.-P. Lin, "Sensorimotor integration in childhood dystonia and dystonic cerebral palsy—a developmental perspective," *Frontiers in Neurology*, vol. 12, p. 668081, 2021.

[3] S. Du, Q. Yu, W. Dai, V. McClelland, and Z. Cvetkovic, "Dictionary learning strategies for cortico-muscular coherence detection and estimation," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 240–244.

[4] C. L. Witham, C. N. Riddle, M. R. Baker, and S. N. Baker, "Contributions of descending and ascending pathways to corticomuscular coherence in humans," *The Journal of physiology*, vol. 589, no. 15, pp. 3789–3800, 2011.

[5] C. S. Zandvoort, J. H. van Dieën, N. Dominici, and A. Daffertshofer, "The human sensorimotor cortex fosters muscle synergies through cortico-synergy coherence," *Neuroimage*, vol. 199, pp. 30–37, 2019.

[6] R. Goebel, A. Roebroeck, D.-S. Kim, and E. Formisano, "Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and Granger causality mapping," *Magnetic resonance imaging*, vol. 21, no. 10, pp. 1251–1261, 2003.

[7] F. Abbas, V. McClelland, Z. Cvetkovic, and W. Dai, "Ss-admm: Stationary and sparse granger causal discovery for cortico-muscular coupling," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[8] A. Shojaie and E. B. Fox, "Granger causality: A review and recent advances," *Annual Review of Statistics and Its Application*, vol. 9, pp. 289–319, 2022.

[9] A. G. Patriota, J. R. Sato, and B. G. B. Achic, "Vector autoregressive models with measurement errors for testing Granger causality," *Statistical Methodology*, vol. 7, no. 4, pp. 478–497, 2010.

[10] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, A. K. Abasi, and S. N. Makhadmeh, "EEG signal denoising using hybridizing method between wavelet transform with genetic algorithm," in *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019: NUSYS'19*. Springer, 2021, pp. 449–469.

[11] J. Christmas and R. Everson, "Robust autoregression: Student-t innovations using variational bayes," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 48–57, 2010.

[12] L. Xiao and T. Zhang, "A proximal-gradient homotopy method for the sparse least-squares problem," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1062–1091, 2013.

[13] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "An iterative bayesian algorithm for sparse component analysis in presence of noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4378–4390, 2009.

[14] Y. Zhou and W. Dai, "Proximal dogleg opportunistic majorization for nonconvex and nonsmooth optimization," *arXiv preprint arXiv:2402.19176*, 2024.

[15] J. D. Lee, Y. Sun, and M. A. Saunders, "Proximal newton-type methods for minimizing composite functions," *SIAM Journal on Optimization*, vol. 24, no. 3, pp. 1420–1443, 2014.

[16] B. Manju and M. Sneha, "Ecg denoising using wiener filter and kalman filter," *Procedia Computer Science*, vol. 171, pp. 273–281, 2020.

[17] S. Yadav, S. K. Saha, and R. Kar, "Evolutionary algorithm-based optimal wiener-adaptive filter design: An application on eeg noise mitigation," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[18] R. Averkamp and C. Houdré, "Wavelet thresholding for non-necessarily gaussian noise: idealism," *The Annals of Statistics*, vol. 31, no. 1, pp. 110–151, 2003.

[19] A. Antoniadis, D. Leporini, and J.-C. Pesquet, "Wavelet thresholding for some classes of non–gaussian noise," *Statistica neerlandica*, vol. 56, no. 4, pp. 434–453, 2002.

[20] L.-M. Dogariu, J. Benesty, C. Paleologu, and S. Ciochină, "An insightful overview of the wiener filter for system identification," *Applied Sciences*, vol. 11, no. 17, p. 7774, 2021.

[21] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer Science & Business Media, 2007.