ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095111

SS-ADMM: STATIONARY AND SPARSE GRANGER CAUSAL DISCOVERY FOR CORTICO-MUSCULAR COUPLING

Farwa Abbas[†]*, Verity McClelland*[*]*, Zoran Cvetkovic*[*]*, and Wei Dai*[†]

[†]Imperial College London, [*]King's College London

ABSTRACT

Cortico-muscular communication patterns reveal important information about motor control. However, inferring significant causal relationships between motor cortex electroencephalogram (EEG) and surface electromyogram (sEMG) of concurrently active muscles is challenging since relevant processes involved in muscle control are relatively weak compared to additive noise and background activities. In this paper, a framework for identification of cortico-muscular linear time invariant communication is proposed that simultaneously estimates model order and its parameters by enforcing sparsity and stationarity conditions in a convex optimization program. The experimental results demonstrate that our proposed algorithm outperforms existing techniques for autoregressive model estimation, in terms of computational speed and model identification for causality estimation.

***Index Terms*—** Autoregressive model, Granger causality, stationarity, sparsity

1. INTRODUCTION

Exploring temporal dependencies in multivariate time series is a long-standing problem in applications ranging from statistics, through econometrics and predictive maintenance, to neuroscience. This provides a systematic framework to perform causal inference and extract useful insights about underlying physical processes. Autoregressive models have been used to model the association between different processes to determine directional *causality* [1], [2], [3]. The notion of causality involved with autoregressive models was first introduced by Granger [4], the main idea being to quantify the extent to which knowledge of a certain process limits the uncertainty associated with another one. Causal inference on non-invasive biomedical signals collected from human subjects, such as EEG and sEMG, is important for neuroscientists and biomedical engineers since these signals indicate neurophysiological changes in the state of a subject's central nervous system or muscles. Understanding the causality of brain-muscle interactions has applications in rehabilitation engineering [5], designing brain-computer interfaces [6], and identifying biomarkers for motor disorders [7].

The primary motor cortex is responsible for planning goal-directed movements, hence, it is adequate to predict motor behaviour based on past brain activity. Several studies have explored this line of research by analysing synchronously recorded primary motor cortex EEG and limb muscle sEMG to extract cortico-muscular coupling [8], [9], [10] . In the context of Granger causality estimation, the EEG signal recorded at a particular time-point is represented by the linear combination of lagged values of muscle signals (sEMG) and past brain activity. Consider $y(t)$ is the EEG signal and $x(t)$ is the corresponding sEMG signal at time t. They can be represented mathematically as:

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

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

where $n_y(t)$ and $n_x(t)$ are time-dependent white noise signals and $\bar{m}$ indicates the maximum dependency on past instants or order of the autoregressive model. The objective is to fit an autoregressive model to estimate parameters $a_{yy}^{(k)}, a_{yx}^{(k)}, a_{xy}^{(k)}, a_{xx}^{(k)}\ \forall k = 1, \cdots, \bar{m}$ where $\bar{m}$ is known a priori. However, in practice, determining the order of the autoregressive model is a non-trivial problem.

1.1. Model Identification by Hierarchical Sparsity

Various studies propose different ways to estimate the order of the autoregressive model. The optimal model order should be small enough so that the forecasts are only based on observations, and not on intermediate forecasts. It should also be large enough to capture the underlying trend of the process. In practice, Granger causality estimated from a small model order limits the *frequency resolution* of physiological signals and hence the frequency ranges cannot be distinguished adequately [7]. In existing literature, model identification has been performed either by relying on visual inspection of the autocorrelation function or by minimizing a loss function such as negative loglikelihood or least squares loss. Akaike Information Criterion [11] and Bayesian Information Criterion [12] are the most popular methods that use this approach. However, these approaches are unable to accurately infer the true data distribution [13] and may overfit in practice [14]. Authors in [15] proposed to alternatively identify the model and estimate the parameters.

Recently, a simultaneous model identification and parameter estimation method for ARMA models has been proposed in [16] that makes use of a hierarchical sparsity-based regularization approach. The authors proposed an ADMM-based method to compute a proximal operator for Latent Overlapping Group (LOG) lasso regularizer [17] and the proximal operator is then used inside a Block Coordinate Descent method to estimate parameters. Inspired by the work in [16], we propose to extend this approach to a convex optimization problem that can guarantee global optimality.

1.2. Stationarity in VAR Model

Autoregressive models are characterized by their distinctive ability of being *self-explanatory*. However, this ability makes them prone to *instability* if the estimation process is not appropriately regulated. Estimation of Granger causality also requires the stationarity assumption for the underlying process as a prerequisite. Most existing

Authorized licensed use limited to: Imperial College London. Downloaded on July 07,2024 at 09:06:34 UTC from IEEE Xplore. Restrictions apply.

works either assume the process to be stationary [18] or use heuristics such as detrending [19] and differencing [2] to induce stationarity in the model. Recently, in [16], an improved method is proposed that computes parameter estimates and project onto a *stationary subspace*, which may not be unique. In this paper, we propose a novel way to encourage stable estimates by efficiently enforcing stationarity on the underlying process. The proposed approach improves on [16] by penalizing unstable estimates. To our knowledge, this is the first work that explicitly enforces stationary condition in a convex program to ensure stability. In this paper we make the following contributions:

1. Propose a unified approach to simultaneously determine model order and parameter estimates in a convex optimization framework solved by ADMM.
2. Efficiently evaluate the proximal operator of regularization based on stationarity condition.
3. Adaptively identify model orders for different sets of parameters in VAR model up to an upper bound.

2. PROBLEM FORMULATION

The system of equations in (1-2) can be compactly represented as a bivariate vector autoregressive model (VAR) as follows:

$$\begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} a_{yy}^{(1)} \cdots a_{yy}^{(\bar{m})} & a_{yx}^{(1)} \cdots a_{yx}^{(\bar{m})} \\ a_{xy}^{(1)} \cdots a_{xy}^{(\bar{m})} & a_{xx}^{(1)} \cdots a_{xx}^{(\bar{m})} \end{bmatrix} \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-\bar{m}) \\ x(t-1) \\ \vdots \\ x(t-\bar{m}) \end{bmatrix} + \begin{bmatrix} n_y(t) \\ n_x(t) \end{bmatrix}, \tag{3}$$

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{H} + \boldsymbol{N}, \tag{4}$$

where the matrices in Equation (3) are assigned to corresponding variables in Equation (4) and $t = \bar{m}+1, ..., \bar{m}+T$. Hence, $\boldsymbol{Y} \in \mathbb{R}^{2\times T}$, $\boldsymbol{A} \in \mathbb{R}^{2\times 2\bar{m}}$, $\boldsymbol{H} \in \mathbb{R}^{2\bar{m}\times T}$ and $\boldsymbol{N} \in \mathbb{R}^{2\times T}$. The model in Equation (1) is the termed as *unrestricted* model. On the other hand, a *restricted* model would be the one that does not involve any cross-coupling between $x(t)$ and $y(t)$ i.e. off-diagonal terms are zero. Mathematically, the *restricted* model amounts to the one below:

$$\begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} a_{yy}^{(1)'} \cdots a_{yy}^{(\bar{m})'} & 0 \cdots\cdots 0 \\ 0 \cdots\cdots 0 & a_{xx}^{(1)'} \cdots a_{xx}^{(\bar{m})'} \end{bmatrix} \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-\bar{m}) \\ x(t-1) \\ \vdots \\ x(t-\bar{m}) \end{bmatrix} + \begin{bmatrix} n_y(t) \\ n_x(t) \end{bmatrix}, \tag{5}$$

$$\boldsymbol{Y} = \boldsymbol{A'}\boldsymbol{H} + \boldsymbol{N'}, \tag{6}$$

where the matrices in Equation (5) are assigned to corresponding variables in Equation (6) and $\boldsymbol{A'} \in \mathbb{R}^{2\times 2\bar{m}}$ and $\boldsymbol{N'} \in \mathbb{R}^{2\times T}$ are the coefficients and noise matrices for restricted model. Granger causality can be tested using an F-test [20] comparing the two models as:

$$\mathrm{GC}_{x\to y} = \frac{\mathrm{RSS}_{y_{\mathrm{res}}} - \mathrm{RSS}_{y_{\mathrm{unr}}} / (p-p')}{\mathrm{RSS}_{y_{\mathrm{unr}}} / (T-p)}, \tag{7}$$

$$\mathrm{GC}_{y\to x} = \frac{\mathrm{RSS}_{x_{\mathrm{res}}} - \mathrm{RSS}_{x_{\mathrm{unr}}} / (p-p')}{\mathrm{RSS}_{x_{\mathrm{unr}}} / (T-p)}, \tag{8}$$

where RSS means *residual sum of squares*, p and p' are the number of parameters in unrestricted and restricted models, respectively. This metric used to determine Granger causality quantifies the impact of cross-coupling terms in prediction of a variable. Note that the bivariate VAR model in Equation (3) can also be written as follows:

$$\boldsymbol{\xi}_t = \boldsymbol{A}_1\boldsymbol{\xi}_{t-1} + \boldsymbol{A}_2\boldsymbol{\xi}_{t-2} + \cdots + \boldsymbol{A}_{\bar{m}}\boldsymbol{\xi}_{t-\bar{m}} + \boldsymbol{n}_t,$$

where $\boldsymbol{\xi}_t := \begin{bmatrix} y(t) \\ x(t) \end{bmatrix}$, $\boldsymbol{A}_i := \begin{bmatrix} a_{yy}^{(i)} & a_{yx}^{(i)} \\ a_{xy}^{(i)} & a_{xx}^{(i)} \end{bmatrix}$ and $\boldsymbol{n}_t := \begin{bmatrix} n_y(t) \\ n_x(t) \end{bmatrix}$. The $\bar{m}$-th order VAR process can be written as a first order VAR process VAR(1) by stacking the variables as:

$$\begin{bmatrix} \xi_t \\ \xi_{t-1} \\ \xi_{t-2} \\ \vdots \\ \xi_{t-\bar{m}+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_{\bar{m}-1} & \boldsymbol{A}_{\bar{m}} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \xi_{t-1} \\ \xi_{t-2} \\ \xi_{t-3} \\ \vdots \\ \xi_{t-\bar{m}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{n}_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix},$$

$$\boldsymbol{\xi}_t = \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t,$$

where $\boldsymbol{\Gamma}(\cdot) : \mathbb{R}^{2\times 2\bar{m}} \to \mathbb{R}^{2\bar{m}\times 2\bar{m}}$ is the transformation to extract the companion matrix corresponding to $\boldsymbol{A}$. For the VAR model to be stationary, the eigenvalues of $\boldsymbol{\Gamma}(\boldsymbol{A})$ must lie inside the unit circle.

3. PROPOSED METHOD

In this paper, we propose to simultaneously identify model order by enforcing sparsity on autoregressive coefficients in a hierarchical fashion. We assume the knowledge of the upper bound $\bar{m}$ on the model order m. The hierarchical sparsity structure is induced by regularizing the objective function by the LOG penalty discussed in [16, 17]. Let $\boldsymbol{c} = [a_{yy}^{(1)} \cdots a_{yy}^{(\bar{m})} a_{yx}^{(1)} \cdots a_{yx}^{(\bar{m})} a_{xy}^{(1)} \cdots a_{xy}^{(\bar{m})} a_{xx}^{(1)} \cdots a_{xx}^{(\bar{m})}]^{\mathrm{T}} \in \mathbb{R}^{4\bar{m}\times 1}$ be the vector containing all of the parameters of the VAR model. Then the LOG penalty function is defined as:

$$\Omega_{\mathrm{LOG}}(\boldsymbol{c}) = \min_{\boldsymbol{l}^{(g)}, g\in\mathcal{G}} \{ \sum_{g\in\mathcal{G}} w_g \|\boldsymbol{l}^{(g)}\|_2 \mid \sum_{g\in\mathcal{G}} \boldsymbol{l}^{(g)} = \boldsymbol{c}, \boldsymbol{l}_{g^c}^{(g)} = 0 \},$$

where $\mathcal{G} = \{\{1\}, \cdots \{1, .., \bar{m}\}, \{\bar{m}+1\}, \cdots \{\bar{m}+1, .., 2\bar{m}\}, \{2\bar{m}+1\}, \cdots \{2\bar{m}+1, .., 3\bar{m}\}, \{3\bar{m}+1\}, \cdots \{3\bar{m}+1, .., 4\bar{m}\}\}$ is the set of all groups $g \in \mathcal{G}$, $\boldsymbol{l}^{(g)} \in \mathbb{R}^{4\bar{m}\times 1}$ is a latent vector indexed by g, and w_g is the weight for set g. As a result of $\Omega_{\mathrm{LOG}}(\boldsymbol{c})$, model order for $a_{yy}, a_{yx}, a_{xy}, a_{xx}$ will be chosen adaptively instead of using the same order. The hierarchical sparsity term is paired with a stationarity constraint for the model to be stable. Unlike the approach in [16], that finds a Euclidean projection using an iterative minimization of the objective over a feasible set, we propose suitable constraints on the companion matrix of the parameters. We propose a spectral norm based regularization technique Ψ_{SP} to restrict the eigenvalues of the companion matrix associated with estimated parameters inside the unit circle. Spectral norm of a matrix is defined as the largest singular value of the matrix

$$\|\boldsymbol{\Gamma}(\boldsymbol{A})\|_2 := \max_i \mid \boldsymbol{\sigma}_i \mid .$$

Hence, the condition for stationarity is $\|\boldsymbol{\Gamma}(\boldsymbol{A})\|_2 < 1$. The proximal operator for spectral norm can be written as follows:

$$\Psi_{\mathrm{SP}}(\boldsymbol{A}) := \arg\min_{\boldsymbol{A}} \gamma \|\boldsymbol{\Gamma}(\boldsymbol{A})\|_2 + \frac{\rho}{2} \|\boldsymbol{X} - \boldsymbol{\Gamma}(\boldsymbol{A})\|_F^2.$$

Authorized licensed use limited to: Imperial College London. Downloaded on July 07,2024 at 09:06:34 UTC from IEEE Xplore. Restrictions apply.

The above problem amounts to finding the proximal operator for ℓ_∞ norm i.e. if $\boldsymbol{X} = \boldsymbol{U}\text{diag}(\boldsymbol{\sigma})\boldsymbol{V}^{\mathrm{T}}$ then

$$\begin{aligned}\boldsymbol{v}^* &= \arg\min_{\boldsymbol{v}} {}^{\gamma}\!/_{\rho}\|\boldsymbol{v}\|_\infty + \frac{1}{2}\|\boldsymbol{v}-\boldsymbol{\sigma}\|_2^2,\\ &= \boldsymbol{\sigma} - {}^{\gamma}\!/_{\rho}\mathcal{P}_{\|\cdot\|_1\le 1}(\frac{\boldsymbol{\sigma}}{{}^{\gamma}\!/_{\rho}}),\\ \Pi_{\gamma/\rho}(\boldsymbol{X}) &:= \boldsymbol{\Gamma}^{-1}(\boldsymbol{U}\text{diag}(\boldsymbol{v}^*)\boldsymbol{V}^{\mathrm{T}}),\end{aligned}$$

where the above operations are stacked inside $\Pi(\cdot)$ for brevity, $\boldsymbol{\Gamma}^{-1}(\cdot) : \mathbb{R}^{2\bar{m}\times 2\bar{m}} \to \mathbb{R}^{2\times 2\bar{m}}$ can be computed by inverting the transformation $\boldsymbol{\Gamma}$, and $\mathcal{P}_{\|\cdot\|_1\le 1}$ is the projection inside the unit ℓ_1 norm ball. This can be done in a non-iterative fashion and exactly by using the approach in [21]. Here, we discuss two useful properties of $\boldsymbol{\Gamma}(\cdot)$ operator that will be referenced later.

Property 1: $\|\boldsymbol{\Gamma}(\boldsymbol{X})\|_F^2 = \|\boldsymbol{X}\|_F^2 + 2(\bar{m}-1)$ for any $\boldsymbol{X} \in \mathbb{R}^{2\times 2\bar{m}}$

Property 2: $\boldsymbol{\Gamma}(\boldsymbol{X}-\boldsymbol{Y}) = 2\boldsymbol{\Gamma}(\frac{1}{2}\boldsymbol{X}) - \boldsymbol{\Gamma}(\boldsymbol{Y})$ for any $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{2\times 2\bar{m}}$.

In addition to reconstruction error, the optimization program will include the Ω_{LOG} and Ψ_{SP} terms to enforce sparsity and stationarity respectively.

$$\begin{aligned}&\min \frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{H}\|_F^2 + \lambda\Omega_{\text{LOG}}(\boldsymbol{c}) + \gamma\|\boldsymbol{\Gamma}(\boldsymbol{Z})\|_2\\ &+ \frac{1}{2}\|\boldsymbol{Y}'-\boldsymbol{A}'\boldsymbol{H}'\|_F^2 + \lambda'\Omega_{\text{LOG}}(\boldsymbol{c}') + \gamma'\|\boldsymbol{\Gamma}(\boldsymbol{Z}')\|_2\\ &\text{s.t.}\quad \boldsymbol{c} = \text{vec}(\boldsymbol{A}^{\mathrm{T}}),\quad \boldsymbol{c}' = \text{vec}(\boldsymbol{A}'^{\mathrm{T}}),\quad \boldsymbol{A} = \boldsymbol{Z},\quad \boldsymbol{A}' = \boldsymbol{Z}'.\end{aligned}$$

More precisely, by expanding $\Omega_{\text{LOG}}(\boldsymbol{c})$ it becomes:

$$\begin{aligned}&\min \frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{H}\|_F^2 + \frac{1}{2}\|\boldsymbol{Y}'-\boldsymbol{A}'\boldsymbol{H}'\|_F^2 + \gamma\|\boldsymbol{\Gamma}(\boldsymbol{Z})\|_2\\ &+ \gamma'\|\boldsymbol{\Gamma}(\boldsymbol{Z}')\|_2 + \lambda\sum_{g\in\mathcal{G}} w_g\|\boldsymbol{P}_{\cdot g}\|_2 + \frac{1}{2}\|\sum_{g\in\mathcal{G}}\boldsymbol{Q}_{\cdot g} - \boldsymbol{c}\|_2^2\\ &+ \lambda'\sum_{g\in\mathcal{G}} w_g\|\boldsymbol{P}'_{\cdot g}\|_2 + \frac{1}{2}\|\sum_{g\in\mathcal{G}}\boldsymbol{Q}'_{\cdot g} - \boldsymbol{c}'\|_2^2\\ &\text{s.t.}\\ &\boldsymbol{c} = \text{vec}(\boldsymbol{A}^{\mathrm{T}}), \boldsymbol{P} = \boldsymbol{Q}, \boldsymbol{c}' = \text{vec}(\boldsymbol{A}'^{\mathrm{T}}), \boldsymbol{P}' = \boldsymbol{Q}', \boldsymbol{A} = \boldsymbol{Z},\\ &\boldsymbol{A}' = \boldsymbol{Z}', (\boldsymbol{P}_{\cdot g})_{g^c} = 0, a_{yx}^{(j)'} = a_{xy}^{(j)'} = 0 \quad \forall j = 1,\cdots,\bar{m}.\end{aligned}$$

The update for parameter matrix $\boldsymbol{A}$ can be computed by finding the gradient of objective with respect to $\boldsymbol{A}$ as follows:

$$\boldsymbol{A} = [\rho(\boldsymbol{Z} - \boldsymbol{U}_3 + \text{vec}^{-1}(\boldsymbol{c}-\boldsymbol{u}_1)^{\mathrm{T}}) + \boldsymbol{Y}\boldsymbol{H}^{\mathrm{T}}](\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} + \rho\mathbf{I})^{-1},$$

where $\rho := \rho_1 + \rho_3$. Cholesky decomposition has beeen used to compute the inverse once. To estimate $\boldsymbol{A}'$ for restricted model, there is an additional constraint $a_{xy}^{(i)'} = a_{yx}^{(i)'} = 0 \quad \forall i = 1,\cdots,\bar{m}$. There are two approaches to address the assumption of known support. First, by intensifying hierarchical sparsity λ', particularly on cross-coupling terms such that they become zero. Second, by discarding the cross-coupling terms in each iteration after solving the ordinary least squares problem. To avoid difficulties associated with both approaches, we shift the sparsity from $\boldsymbol{A}'$ to data matrices $\boldsymbol{Y}$ and $\boldsymbol{H}$ such that cross-coupling terms do not play any role in optimization. By redefining matrices in Equation (3) we can write:

$$\boldsymbol{Y}' = \boldsymbol{A}'\boldsymbol{H}' + \boldsymbol{N}',$$

$$\boldsymbol{Y}' := \begin{bmatrix} y(t) & \mathbf{0} \\ \mathbf{0} & x(t)\end{bmatrix} \in \mathbb{R}^{2\times 2T'},\ \boldsymbol{H}' := \begin{bmatrix} y(t-1) & & \\ \vdots & & \mathbf{0} \\ y(t-\bar{m}) & & \\ & x(t-1) & \\ & \vdots & \\ \mathbf{0} & & x(t-\bar{m})\end{bmatrix} \in \mathbb{R}^{2\bar{m}\times 2T'}$$

and $\boldsymbol{A}' := \begin{bmatrix} a_{yy}^{(1)'} & \cdots & a_{yy}^{(\bar{m})'} & a_{yx}^{(1)'} & \cdots & a_{yx}^{(\bar{m})'} \\ a_{xx}^{(1)'} & \cdots & a_{xx}^{(\bar{m})'} & a_{xy}^{(1)'} & \cdots & a_{xy}^{(\bar{m})'}\end{bmatrix} \in \mathbb{R}^{2\times 2\bar{m}}$, and $\boldsymbol{N}' := \begin{bmatrix} n_y(t)' & \mathbf{0} \\ \mathbf{0} & n_x(t)'\end{bmatrix} \in \mathbb{R}^{2\times 2T'}$. Now we can update $\boldsymbol{A}'$ by the equation below.

$$\begin{aligned}\boldsymbol{A}' = &[\rho'(\boldsymbol{Z}' - \boldsymbol{U}_3' + \text{vec}^{-1}(\boldsymbol{c}'-\boldsymbol{u}_1')^{\mathrm{T}}) + \boldsymbol{Y}'\boldsymbol{H}'^{\mathrm{T}}]\\ &(\boldsymbol{H}'\boldsymbol{H}'^{\mathrm{T}} + \rho'\mathbf{I})^{-1},\end{aligned}$$

where $\rho' := \rho_1' + \rho_3'$. After computing $\boldsymbol{A}'$ we can safely discard cross-coupling terms without any loss of information. Closed-form solutions for hierarchical sparsity terms $\boldsymbol{P}, \boldsymbol{Q}$ and $\boldsymbol{P}', \boldsymbol{Q}'$ are obtained as discussed in [16]. For details, refer to Algorithm 1. To find the update for $\boldsymbol{Z}$, **Property 1** and **Property 2** are leveraged to find the closed-form solution:

$$\begin{aligned}\boldsymbol{Z} &= \arg\min_{\boldsymbol{Z}} \gamma\|\boldsymbol{\Gamma}(\boldsymbol{Z})\|_2 + \frac{\rho_3}{2}\|\boldsymbol{A}+\boldsymbol{U}_3-\boldsymbol{Z}\|_F^2,\\ &= \arg\min_{\boldsymbol{Z}} \gamma\|\boldsymbol{\Gamma}(\boldsymbol{Z})\|_2 + \frac{\rho_3}{2}\|2\boldsymbol{\Gamma}(\frac{1}{2}(\boldsymbol{A}+\boldsymbol{U}_3)) - \boldsymbol{\Gamma}(\boldsymbol{Z})\|_F^2,\\ \boldsymbol{Z} &= \Pi_{\gamma/\rho_3}(2\boldsymbol{\Gamma}(\frac{1}{2}(\boldsymbol{A}+\boldsymbol{U}_3))).\end{aligned}$$

Unrestricted and restricted models are evaluated by Algorithm 1. For initialization phase one-time cost is $\mathcal{O}(8\bar{m}^3 + 4T\bar{m})$, however, the total computational cost for the iterative process is $\mathcal{O}(8\bar{m}^2 + 8\bar{m}^3 + 2\bar{m}\log(2\bar{m}))\times$ number of iterations.

Algorithm 1 SS-ADMM for GC Estimation

Input: $\boldsymbol{Y}, \boldsymbol{H}, \bar{m}, \lambda, \gamma, \rho_1, \rho_2, \rho_3$

while stopping criterion not met **do**
- $\boldsymbol{B} := \rho(\boldsymbol{Z} - \boldsymbol{U}_3 + \text{vec}^{-1}(\boldsymbol{c}-\boldsymbol{u}_1)^{\mathrm{T}})$
- $\boldsymbol{A} = (\boldsymbol{B} + \boldsymbol{Y}\boldsymbol{H}^{\mathrm{T}})(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} + \rho\mathbf{I})^{-1}$
- **for** $g = 1, ..., 4\bar{m}$ **do**
  - $\boldsymbol{P}_{gg} = \mathbf{prox}_{\lambda w_g\|\cdot\|}(\boldsymbol{P}_{gg} + \boldsymbol{q}_g - \boldsymbol{u}_{2_g} - \boldsymbol{p}_g)$
  - $\boldsymbol{P}_{g^c g} = 0$
- **end for**
- $\boldsymbol{p} = {}^{1}\!/_{4\bar{m}}\sum_{g\in\mathcal{G}}\boldsymbol{P}$
- $\boldsymbol{q} = {}^{1}\!/_{(\rho_2+4\bar{m})}(\boldsymbol{c} + \rho_2(\boldsymbol{u}_2 + \boldsymbol{p}))$
- $\boldsymbol{c} = \text{vec}(\boldsymbol{A}^{\mathrm{T}}) + \sum_{g\in\mathcal{G}}\boldsymbol{P} + \text{vec}(\boldsymbol{U}_1)$
- $\boldsymbol{Z} = \boldsymbol{\Pi}_{\gamma/\rho_3}(2\boldsymbol{\Gamma}(\frac{1}{2}(\boldsymbol{A}+\boldsymbol{U}_3)))$
- $\boldsymbol{u}_1 = \boldsymbol{u}_1 + \text{vec}(\boldsymbol{A}^{\mathrm{T}}) - \boldsymbol{c}$
- $\boldsymbol{u}_2 = \boldsymbol{u}_2 + \boldsymbol{p} - \boldsymbol{q}$
- $\boldsymbol{U}_3 = \boldsymbol{U}_3 + \boldsymbol{A} - \boldsymbol{Z}$
- $m_i = \text{card}(\boldsymbol{P}_{(i-1)\bar{m}+1:i\bar{m}} \neq 0)\ \forall i = 1,2,3,4$

end while

Output: $\boldsymbol{A}, m_1, m_2, m_3, m_4$

4. EXPERIMENTAL RESULTS

In this section, we perform a comparative study to investigate the performance of proposed method Stationary and Sparse ADMM (SS-ADMM) on both synthetic and real data. All the experiments have been performed on MATLAB 2022b with Core i7 CPU (2.90 GHz), 8 GB RAM, and Windows 11 operating system.

Authorized licensed use limited to: Imperial College London. Downloaded on July 07,2024 at 09:06:34 UTC from IEEE Xplore. Restrictions apply.

4.1. Results on Synthetic Data

Data has been generated randomly using an underlying mathematical model governing $\boldsymbol{Y}$, $\boldsymbol{A}$, and $\boldsymbol{H}$. Results obtained by repeating the experiment several times are compared to assess performance for both stationarity and sparsity. The performance for model identification is compared with BIC [12] in Figure 1. Upper bound on order was set to $\bar{m} = 40$. It can be observed that SS-ADMM estimates model order either equal or higher than the true order whereas BIC usually underestimates model order losing significant information about the process. It is worth noting that SS-ADMM also allows flexibility by fine-tuning the hyperparameter λ to precisely estimate the model order. Proposed method SS-ADMM disentangles four different model orders instead of two (as in BIC) for a bivariate model. It must be noted that our approach is flexible. I

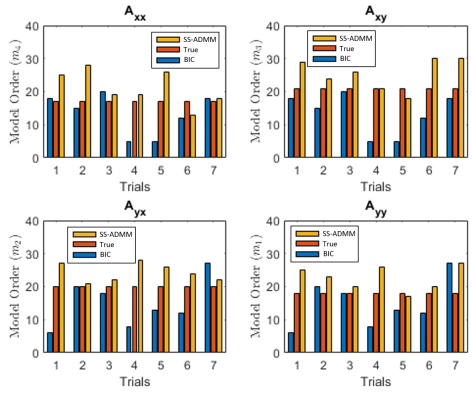


Fig. 1. Comparison of performance for model identification with BIC [12]. Blue bar graphs in a row are equal.

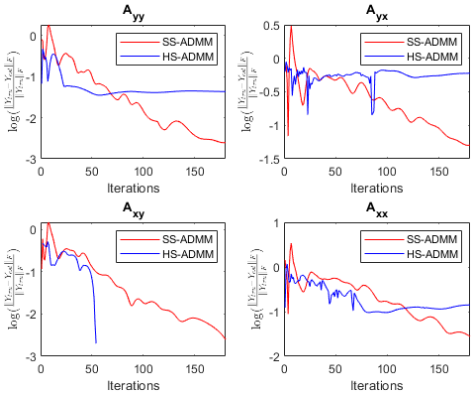


Fig. 2. Log of normalized error of HS-ADMM [16] vs SS-ADMM.

The results obtained by enforcing stationarity by finding euclidean projections as in Hierarchical Sparsity ADMM (HS-ADMM) [16] compared with our proposed approach Ψ_{SP} are depicted in Figure 2. The projection based approach takes longer to execute. On the other hand, our proposed approach SS-ADMM achieves a speedup of $2\times$ or more and the error decreases more smoothly as shown in 2 when executed for same number of iterations.

4.2. Results on Physiological Data

The proposed method is tested on physiological data collected from nine healthy subjects in a previously published study [22]. The subjects performed a controlled motor task, grasping a ruler between 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 (mean 7 s). The experiment comprised 8 blocks of 25 trials. sEMG was recorded from first dorsal interosseous and bipolar scalp EEG was recorded over left sensorimotor cortex. Both signals 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 post-stimulus). Epochs containing movement or blink artefacts were eliminated. Granger causality across the 5 second epoch is shown for 4 subjects in Figure 3. An F-test applied with 95% significance level tests the null hypothesis that the first time series does not jointly Granger-cause the second.

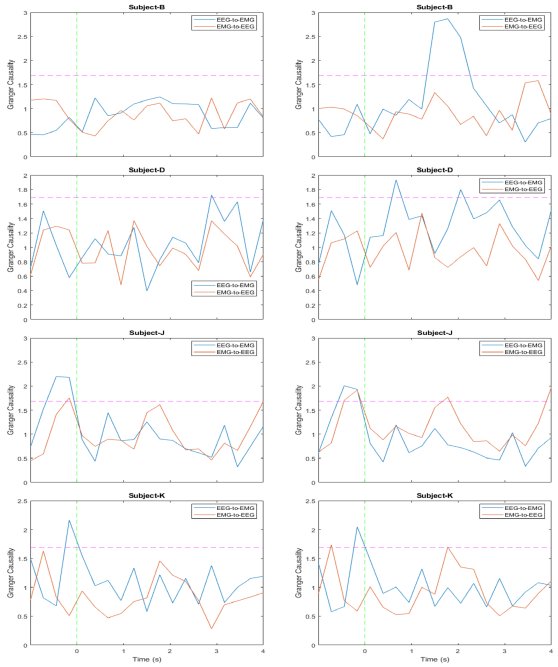


Fig. 3. ***Left:*** Blockwise GC [2], ***Right:*** SS-ADMM. Green and pink lines show stimulus onset and critical values of F-distribution respectively.

In this active task we expect causality between brain and muscles, so the observed pre-stimulus causality in some subjects (Figure 3), is not unanticipated. A post-stimulus increase in Granger causality can be observed in subjects B and D using SS-ADMM which was not detected using Blockwise GC [2]. Similar findings were observed in five out of nine subjects by tuning the hyperparameters, indicating the effectiveness of our approach.

5. CONCLUSION

To conclude, we proposed a novel way to incorporate stationarity assumption in autoregressive model identification and parameter estimation. We used a convex optimization framework to guarantee global optimality that leads to improved estimation of Granger causality in physiological data to extract cortico-muscular coupling. Experimental results demonstrate the effectiveness of our approach for both synthetic and real-world data.

Authorized licensed use limited to: Imperial College London. Downloaded on July 07,2024 at 09:06:34 UTC from IEEE Xplore. Restrictions apply.

6. REFERENCES

[1] Cheng Hsiao, "Autoregressive modeling and causal ordering of economic variables," *Journal of Economic Dynamics and Control*, vol. 4, pp. 243–259, 1982.

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

[3] Rainer Goebel, Alard Roebroeck, Dae-Shik Kim, and Elia 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.

[4] Clive WJ Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

[5] Stephen H Scott, "Optimal feedback control and the neural basis of volitional motor control," *Nature Reviews Neuroscience*, vol. 5, no. 7, pp. 532–545, 2004.

[6] Hesam Varsehi and S Mohammad P Firoozabadi, "An eeg channel selection method for motor imagery based brain–computer interface and neurofeedback using granger causality," *Neural Networks*, vol. 133, pp. 193–206, 2021.

[7] Esther Florin, Johannes Pfeifer, Veerle Visser-Vandewalle, Alfons Schnitzler, and Lars Timmermann, "Parkinson subtype-specific granger-causal coupling and coherence frequency in the subthalamic area," *Neuroscience*, vol. 332, pp. 170–180, 2016.

[8] Shengjia Du, Qi Yu, Wei Dai, Verity McClelland, and Zoran Cvetkovic, "Dictionary learning strategies for corticomuscular coherence detection and estimation," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 240–244.

[9] Zhenghao Guo, Verity M McClelland, and Zoran Cvetkovic, "Unravelling causal relationships between cortex and muscle with errors-in-variables models," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 967–970.

[10] Zhenghao Guo, Verity M McClelland, Osvaldo Simeone, Kerry R Mills, and Zoran Cvetkovic, "Multiscale wavelet transfer entropy with application to corticomuscular coupling analysis," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 2, pp. 771–782, 2021.

[11] Hirotugu Akaike, "Fitting autoregressive models for prediction," *Annals of the institute of Statistical Mathematics*, vol. 21, no. 1, pp. 243–247, 1969.

[12] Gideon Schwarz, "Estimating the dimension of a model," *The annals of statistics*, pp. 461–464, 1978.

[13] Galit Shmueli, "To explain or to predict?," *Statistical science*, vol. 25, no. 3, pp. 289–310, 2010.

[14] Nicandro Cruz-Ramírez, Héctor-Gabriel Acosta-Mesa, Rocío-Erandi Barrientos-Martínez, and Luis-Alonso Nava-Fernández, "How good are the bayesian information criterion and the minimum description length principle for model selection? a bayesian network analysis," in *Mexican International Conference on Artificial Intelligence*. Springer, 2006, pp. 494–504.

[15] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.

[16] Yin Liu and Sam Davanloo Tajbakhsh, "Fitting arma time series models without identification: A proximal approach," *arXiv preprint arXiv:2002.06777*, 2020.

[17] Laurent Jacob, Guillaume Obozinski, and Jean-Philippe Vert, "Group lasso with overlap and graph lasso," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 433–440.

[18] Jitkomut Songsiri, "Sparse autoregressive model estimation for learning granger causality in time series," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 3198–3202.

[19] Mark W Watson, "Univariate detrending methods with stochastic trends," *Journal of monetary economics*, vol. 18, no. 1, pp. 49–75, 1986.

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

[21] Weiran Wang and Miguel A Carreira-Perpinán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," *arXiv preprint arXiv:1309.1541*, 2013.

[22] Verity M McClelland, Zoran Cvetkovic, and Kerry R Mills, "Modulation of corticomuscular coherence by peripheral stimuli," *Experimental brain research*, vol. 219, pp. 275–292, 2012.

Authorized licensed use limited to: Imperial College London. Downloaded on July 07,2024 at 09:06:34 UTC from IEEE Xplore. Restrictions apply.