Granger Causal Inference from Indirect Low-Dimensional Measurements with Application to MEG Functional Connectivity Analysis

Behrad Soleimani
Department of ECE
University of Maryland
College Park, MD, USA
behrad@umd.edu

Proloy Das
Department of ECE
University of Maryland
College Park, MD, USA
proloy@umd.edu

Joshua Kulasingham
Department of ECE
University of Maryland
College Park, MD, USA
joshuapk@umd.edu

Jonathan Z. Simon
Departments of ECE & Biology
Inst. for Sys. Res. (ISR)
University of Maryland
College Park, MD, USA
jzsimon@umd.edu

Behtash Babadi
Department of ECE
Inst. for Sys. Res. (ISR)
University of Maryland
College Park, MD, USA
behtash@umd.edu

***Abstract*—We consider the problem of determining Granger causal influences among sources that are indirectly observed through low-dimensional and noisy linear projections. Commonly used methods proceed in a two-stage fashion, by first solving an inverse problem to localize the sources, and then inferring the Granger causal influences from the estimated sources. The inferred Granger causal links thus inherit the various biases that are used in source localization techniques, in the form of spatiotemporal priors designed in favor of spatial localization. In addition, this approach does not account for the structural properties of the underlying functional networks such as sparsity of the links. We address these issues by modeling the source dynamics as sparse vector autoregressive processes and estimate the model parameters directly from the observations, with no recourse to intermediate source localization. We evaluate the performance of the proposed methodology using both simulated and experimentally-recorded MEG data.**

***Index Terms*—Granger causality, Statistical inference, EM algorithm, MEG.**

I. Introduction

In many applications of interest, the output of a system is observed as the collective activity of latent sources, and inferring the causal relations among these sources is desired. For example, in case of the human brain, collective activation of different cortical areas results in accomplishing a single task, and extracting the causal relations among different cortical regions allows to probe the underlying circuitry.

A popular characterization of the aforementioned causal relationships is given by the notion of Granger Causality: if including knowledge of one source improves the prediction of another source, the latter is thought to be causally driven by the former. More specifically, for two processes $\{\mathbf{x}_{1,t}, \mathbf{x}_{2,t}\}_{t=1}^{\infty}$, if knowing $\{\mathbf{x}_{2,t}\}_{t=1}^{\infty}$ improves the prediction of $\{\mathbf{x}_{1,t}\}_{t=1}^{\infty}$, we say $\mathbf{x}_2$ Granger causes $\mathbf{x}_1$, or there exists a Granger Causal (GC) link, $\mathbf{x}_2 \mapsto \mathbf{x}_1$ [1]. This methodology has been utilized in several studies where the whole brain activity is measured via functional magnetic resonance imaging (fMRI) [2], [3], [4]. These measurements of the brain activity are quantified by changes in the blood oxygen level dependent (BOLD) signals as surrogates of neural activity. In these studies, Granger causality between different regions of the brain are directly obtained from these BOLD signals. Given the low temporal resolution of the BOLD signal (typically < 1 Hz), it is not possible to account for Granger causal influences that occur at the millisecond resolution, which is known to be relevant to various cognitive and sensory processes in the brain.

Magnetoencephalography (MEG) and Electroencephalography (EEG) address this challenge by providing measurements of the neural activity at the millisecond resolution. However, the M/EEG sensors capture low-dimensional linear mixtures of the underlying neural sources. As a result, inferring the causal relationships between the sources from low-dimensional measurements is not straightforward. One conventional approach is to first solve an inverse problem to estimate the source activities, often referred to as source localization. Then, the estimated sources could be used to infer connectivity measures such as GC links [5], [6], [7], [8].

Source localization has been a topic of active research over the past few decades, and has resulted in many algorithms that achieve high localization accuracy using specific spatiotemporal priors in well-designed trial based experiments. However, the desirable spatial localization of these algorithms comes at the cost of introducing spatiotemporal biases in the solution. As such, using source estimates that are biased towards accurate spatial localization for GC analysis is likely to result in a biased characterization of the underlying functional networks [9]. In addition, such inverse solutions do not account for the inherent sparsity of the functional networks, i.e., GC links.

To address this challenge, in this paper we model the source dynamics using a sparse latent vector autoregressive (VAR) parametric model and propose a methodology to *directly* estimate the model parameters from noisy and low-dimensional measurements, with no recourse to intermediate source localization. To explicitly model the sparsity of the underlying networks, we estimate the VAR parameters via sparse regularization using an instance of the Expectation-Maximization (EM) algorithm. We subsequently estimate and remove the bias introduced in the GC measure due to regularization, and characterize the strength of the GC links using Youden's J-statistics. Finally, we illustrate the performance of our proposed methodology using simulated and experimentally-recorded MEG data.

This work is supported in part by the National Science Foundation Awards No. 1552946 and 1734892.

II. Preliminaries and Problem Definition

The source activity, $\mathbf{x}_t \in \mathbb{R}^{N_x}$ and the M/EEG observations, $\mathbf{y}_t \in \mathbb{R}^{N_y}$, both at time t, are related by the following linear mapping:

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{n}_t, \quad t = 1, 2, \ldots, T \ , \tag{1}$$

where the matrix $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$ is often referred to as the lead field matrix, and $\mathbf{n}_t \in \mathbb{R}^{N_y}$ is the additive measurement noise vector at time t. Here, the N_x, N_y, and T represent the dimension of the sources, number of sensors, and number of time points, respectively. The lead field matrix can be calculated using a 3D head model obtained by magnetic resonance imaging. The measurement noise in Eq. (1), $\mathbf{n}_t$ is considered to be a zero-mean Gaussian random vector with covariance matrix $\mathbf{R}$ and i.i.d. across time.

In order to explicitly model the spatiotemporal dependencies of the sources, we use a VAR model with q lags, denoted by VAR(q), as follows:

$$\mathbf{x}_t = \sum_{k=1}^{q} \mathbf{A}_k \mathbf{x}_{t-k} + \mathbf{w}_t, \quad t = q+1, \ldots, T, \tag{2}$$

where $\mathbf{A}_k \in \mathbb{R}^{N_x \times N_x}, \ k = 1, \ldots, q$ is the coefficient matrix corresponding to the k^{th} lag, with $[\mathbf{A}_k]_{i,j} = a_{i,j,k}$ denoting the coupling coefficient from the j^{th} source to the i^{th} source at lag k. The process noise vector $\mathbf{w}_t := [w_{1,t}, w_{2,t}, \ldots, w_{N_x,t}]^\top \in \mathbb{R}^{N_x}$ is assumed to be zero-mean temporally white Gaussian noise with a spatial diagonal covariance matrix $\mathbf{Q} = \text{diag}(\sigma_1^2, \ldots, \sigma_{N_x}^2)$. We refer to the model in Eq. (2) as the *full model*.

Within this framework, to test the Granger causality from source $\tilde{i}$ to i, shown as $(\tilde{i} \mapsto i)$, we remove the contribution of the $\tilde{i}^{\text{th}}$ source from the i^{th} source in Eq. (2) by enforcing $a_{i,\tilde{i},k} = 0, \ \forall k$. We similarly consider the process noise for the i^{th} source to be temporally white and $w'_{i,t} \sim \mathcal{N}(0, \sigma^2_{i\setminus\tilde{i}})$. We refer to this model as the *reduced model*. We then define the GC statistic as follows:

$$\mathcal{F}_{(\tilde{i}\mapsto i)} = \text{sign}\Big(\sum_{k=1}^{q} \widehat{a}_{i,\tilde{i},k}\Big) \log\left(\frac{\widehat{\sigma}^2_{i\setminus\tilde{i}}}{\widehat{\sigma}^2_i}\right), \tag{3}$$

where $\widehat{a}_{i,\tilde{i},k}$, $\widehat{\sigma}^2_{i\setminus\tilde{i}}$ and $\widehat{\sigma}^2_i$ denote the estimated versions of $a_{i,\tilde{i},k}$, $\sigma^2_{i\setminus\tilde{i}}$ and σ^2_i, respectively. The sign(.) function, determines the excitatory (positive) or inhibitory (negative) nature of the GC link. The reduced model is nested in the full model, so one would generally expect $\widehat{\sigma}^2_{i\setminus\tilde{i}} \geq \widehat{\sigma}^2_i$. If $\widehat{\sigma}^2_{i\setminus\tilde{i}} \approx \widehat{\sigma}^2_i$, knowledge of the past activity of the $\tilde{i}^{\text{th}}$ source does not improve the prediction of the i^{th} source, i.e., $\mathcal{F}_{(\tilde{i}\mapsto i)} \approx 0$ and we can rule out the GC link $(\tilde{i} \mapsto i)$. On the other hand, if $\widehat{\sigma}^2_{i\setminus\tilde{i}} \gg \widehat{\sigma}^2_i$, knowledge of the past activity of the $\tilde{i}^{\text{th}}$ source improves the prediction of the i^{th} source, and we have $|\mathcal{F}_{(\tilde{i}\mapsto i)}| \gg 0$, i.e., the GC link $(\tilde{i} \mapsto i)$ exists.

Our goal is to determine which of the members of the set $\mathcal{I} := \{(\tilde{i} \mapsto i) \big| \ i, \tilde{i} = 1, 2, \ldots, N_x, \ i \neq \tilde{i}\}$, which includes all the existing GC links among the N_x sources, underlie the source dynamics. The challenge in doing so is three-fold: 1) the VAR parameters are assumed to be sparse and unknown, and need to be reliably estimated from the low-dimensional observations $\{\mathbf{y}_t\}_{t=1}^T$, 2) sparse estimation of the coefficients $\{\mathbf{A}_k\}_{k=1}^q$ results in biased estimates of the process noise variances, which in turn bias the GC statistic, and 3) the fact that $|\mathcal{F}_{(\tilde{i}\mapsto i)}| > 0$ does not imply a statistically significant GC link $(\tilde{i} \mapsto i)$.

III. Inference Algorithm

To identify the GC links in the set $\mathcal{I}$, given the observations $(\mathbf{y}_{1:T})$, the VAR model parameters need to be estimated. We assume that the observation noise covariance $\mathbf{R}$ is known. In our application of interest, i.e., MEG analysis, it can indeed be estimated based on empty room recordings [10]. We address the foregoing three challenges in the following subsections.

A. EM-based Parameter Estimation

We first note that expressing the log-likelihood of the observations $\mathbf{y}_{1:T}$ only in terms of the model parameters $\mathbf{Q}$ and $\mathbf{A}_k, \ k = 1, \ldots, q$ results in an intractable expression. We thus consider the source activities $\mathbf{x}_{1:T}$ as latent variables and proceed with an EM procedure [11]. Let the unknown parameters be denoted by $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{N_x})$ where $\boldsymbol{\theta}_i := (\sigma_i^2, \mathbf{a}_i)$, with $\mathbf{a}_i := [a_{i,j,k}, \forall j, k]^\top$.

We first consider estimation of the parameters in the full model. The *complete* likelihood parameterized by $\boldsymbol{\theta}$ is denoted by $p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}; \boldsymbol{\theta})$. Starting with an initial guess of $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(0)}$, at each iteration l, we compute the conditional expectation of $\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}; \boldsymbol{\theta})$ given the current estimate of $\widehat{\boldsymbol{\theta}}^{(l)}$ and observations $\mathbf{y}_{1:T}$ as a surrogate function, and maximize it to get an updated estimate of the parameters, $\widehat{\boldsymbol{\theta}}^{(l+1)}$. The surrogate function, often referred to as the Q-function, is thus defined as $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(l)}) := \mathbb{E}\left[\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}; \boldsymbol{\theta}) \big| \mathbf{y}_{1:T}; \widehat{\boldsymbol{\theta}}^{(l)}\right]$, and can be generically expressed in our case as:

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(l)}) = \mathcal{K}(\widehat{\boldsymbol{\theta}}^{(l)}) - \frac{T-q}{2} \sum_{i=1}^{N_x} \log(\sigma_i^2) \\ - \sum_{i=1}^{N_x} \frac{1}{2\sigma_i^2} \left(\mathbf{a}_i{}^\top \mathbf{G}^{(l)} \mathbf{a}_i - 2\mathbf{h}_i^{(l)\top} \mathbf{a}_i + f_i^{(l)} \right), \end{aligned} \tag{4}$$

where $\mathcal{K}(\widehat{\boldsymbol{\theta}}^{(l)})$ denotes terms that are independent of $\boldsymbol{\theta}$, and $\mathbf{G}^{(l)}$, $\mathbf{h}_i^{(l)}$, and $f_i^{(l)}$ are functions of first- and second-order moments of the conditional density $p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T}; \widehat{\boldsymbol{\theta}}^{(l)})$. Given the foregoing structure of the Q-function, the l^{th} iteration of the EM algorithm can be summarized as follows:

E-step: Under the Gaussian assumption on $\mathbf{n}_t$ and $\mathbf{w}_t$, the conditional density of $p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T}; \boldsymbol{\theta})$ can be shown to be Gaussian, and hence the mean and covariances can be efficiently computed via the Fixed Interval Smoothing (FIS) algorithm [12] in an efficient manner.

M-step: The low-dimensionality of the observations results in an ill-posed maximization problem. We mitigate ill-posedness by regularization and seek updated parameters via:

$$\widehat{\boldsymbol{\theta}}^{(l+1)} = \underset{\boldsymbol{\theta}}{\text{argmax}} \left\{ Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(l)}) + R_p(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right\}, \tag{5}$$

where $R_p(\boldsymbol{\lambda},\boldsymbol{\theta}) := -2\sum_{i=1}^{N_x}\left(\lambda_i \|\mathbf{a}_i\|_p^p\right)$ is the regularization function and $\boldsymbol{\lambda} = [\lambda_1,\ldots,\lambda_{N_x}]^\top \in \mathbb{R}^{N_x}$ is the vector of regularization coefficients. For $p = 2$, the solution can be obtained in closed form:

$$\widehat{\mathbf{a}}_i^{(l+1)} = \left(\mathbf{G}^{(l)} + \lambda_i \mathbf{I}\right)^{-1}\mathbf{h}_i^{(l)}, \tag{6}$$

$$\widehat{\sigma}_i^{2(l+1)} = \frac{1}{T-q}\Big(\widehat{\mathbf{a}}_i^{(l+1)\top}\mathbf{G}^{(l)}\widehat{\mathbf{a}}_i^{(l+1)} - 2\mathbf{h}_i^{(l)\top}\widehat{\mathbf{a}}_i^{(l+1)} + f_i^{(l)}\Big). \tag{7}$$

In order to enforce sparsity, we consider $p = 1$, for which closed-form solutions do not exist. However, the problem can be solved via another instance of the EM algorithm, known as Iteratively Re-weighted Least Squares (IRLS) [13], utilizing the solution in Eq. (6) with suitable re-weighting of the underlying parameters. We consider $\boldsymbol{\lambda} = \lambda\mathbf{1}$, where λ can be tuned by standard cross-validation procedures.

Using the same procedure, we also fit the parameters of the reduced model corresponding to the GC link $(\tilde{i} \mapsto i)$. The GC statistic in Eq. (3) is then computed using the process noise variance estimates from the full and reduced models.

B. Assessing the Statistical Significance of the GC Links

Due to variability of the estimated coefficient in Eq. (3), a non-zero value of $\mathcal{F}_{(\tilde{i}\mapsto i)}$ does not necessarily imply existence of the GC link $(\tilde{i} \mapsto i)$. A statistical inference scheme is thus required to assess the significance of the estimated GC links.

Consider the link $(\tilde{i} \mapsto i) \in \mathcal{I}$. We denote the corresponding parameter estimates from the full and reduced models by $\boldsymbol{\theta}_i^F := \boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_i^R := \boldsymbol{\theta}_{i\backslash\tilde{i}}$, respectively, where $\boldsymbol{\theta}_{i\backslash\tilde{i}} = (\sigma^2_{i\backslash\tilde{i}}, \mathbf{a}_{i\backslash\tilde{i}};\ a_{i,\tilde{i},k} = 0,\ \forall k)$. The dimensions of unknown parameter vectors of the full and reduced models are represented by $M^F = qN_x + N_x$ and $M^R = q(N_x-1)+N_x$, respectively. In this particular case, the null hypothesis on the existence of the GC link is defined as $H_{(\tilde{i}\mapsto i),0}:\ \boldsymbol{\theta}_i = \boldsymbol{\theta}_i^R$, i.e., there is no GC influence from source $\tilde{i}$ to i, and the alternative hypothesis as $H_{(\tilde{i}\mapsto i),1}:\ \boldsymbol{\theta}_i = \boldsymbol{\theta}_i^F$, i.e., there exists a GC link from source $\tilde{i}$ to i.

One conventional statistic for hypothesis testing in this context is the deviance difference between the estimated full and reduced model parameters:

$$\mathcal{D}_{(\tilde{i}\mapsto i)} := 2\big(\ell_i(\widehat{\boldsymbol{\theta}}_i^F) - \ell_i(\widehat{\boldsymbol{\theta}}_i^R)\big), \tag{8}$$

where $\ell_i(\boldsymbol{\theta}_i) := \log p(x_{i,q+1:T}|x_{i,1:q};\boldsymbol{\theta}_i)$. It is easy to see that $\mathcal{D}_{(\tilde{i}\mapsto i)} = \text{sign}\Big(\sum_{k=1}^q \widehat{a}_{i,\tilde{i},k}\Big)(T-q)\mathcal{F}_{(\tilde{i}\mapsto i)}$.

Due to regularization, the deviance difference estimate is biased and hence does not readily admit a well-defined asymptotic behavior. We therefore use the *de-biased* version of deviance difference which is defined as [14]:

$$\mathcal{D}^{db}_{(\tilde{i}\mapsto i)} = \mathcal{D}_{(\tilde{i}\mapsto i)} - B(\widehat{\boldsymbol{\theta}}_i^R) + B(\widehat{\boldsymbol{\theta}}_i^F), \tag{9}$$

where $B(\boldsymbol{\theta}_i) := -\dot{\boldsymbol{\ell}}_i(\boldsymbol{\theta}_i)^\top \ddot{\boldsymbol{\ell}}_i(\boldsymbol{\theta}_i)^{-1}\dot{\boldsymbol{\ell}}_i(\boldsymbol{\theta}_i)$, is the empirical bias term incurred due to regularization [15], with $\dot{\boldsymbol{\ell}}_i(.)$ and $\ddot{\boldsymbol{\ell}}_i(.)$ denoting the gradient vector and Hessian matrix of the log-likelihood function $\ell_i(.)$, respectively. Removal of the bias allows to recover known asymptotics of the deviance difference. To this end, we have:

Theorem 1. *The de-biased deviance differences converge weakly to the following distributions, under the null and alternative hypotheses (as $T \to \infty$):*

$$[\mathcal{D}^{db}_{(\tilde{i}\mapsto i)}|H_{(\tilde{i}\mapsto i),0}] \xrightarrow{d} \chi^2(M^d), \tag{10}$$

$$[\mathcal{D}^{db}_{(\tilde{i}\mapsto i)}|H_{(\tilde{i}\mapsto i),1}] \xrightarrow{d} \chi^2(M^d, \nu_{(\tilde{i}\mapsto i)}), \tag{11}$$

where $\chi^2(k)$ represents the χ^2 distribution with k degrees of freedom, and $\chi^2(k,\nu)$ represents the non-central χ^2 distribution with k degrees of freedom and non-centrality parameter ν, and $M^d := M^F - M^R = q$.

Proof. The proof is based on [14] and [16], and is omitted here for brevity. ■

The non-centrality parameter in Eq. (11) can be estimated as $\widehat{\nu}_{(\tilde{i}\mapsto i)} = \max\big\{\sum_{n=1}^N \mathcal{D}^{db,(n)}_{(\tilde{i}\mapsto i)}/N - M^d, 0\big\}$ where $\mathcal{D}^{db,(n)}_{(\tilde{i}\mapsto i)}$ is the n^{th} sample of the deviance computed from N independent trials [17]. Theorem 1 allows to both characterize the strength of the hypothesis test and control for false discovery rate (FDR), as described next.

FDR Control Procedure: Since multiple comparisons are required to identify the significant GC links among a large number of source pairs, the rate of type-I error (false discovery rate) needs to be controlled. First, to control the type-I error at a desired significance level α, the threshold for rejecting $H_{(i\mapsto\tilde{i}),0}$ should be chosen at $F^{-1}_{\chi^2(M^d)}(1-\alpha)$, where $F^{-1}_{\chi^2(M^d)}(.)$ is the inverse cumulative distribution function (CDF) of $\chi^2(M^d)$. To correct for multiple comparisons, the Benjamini-Yekutieli procedure can be utilized to control the average FDR at a rate of $\overline{\alpha} := \frac{(|\mathcal{I}|+1)\alpha}{2|\mathcal{I}|\log|\mathcal{I}|}$ [18].

Test Strength Characterization: From Eq. (11), the false rejection rate at a confidence level of $1-\alpha$ for the detected links can be quantified as $\eta_{(\tilde{i}\mapsto i)}(\alpha) := F_{\chi^2(M^d,\widehat{\nu}_{(\tilde{i}\mapsto i)})}(F^{-1}_{\chi^2(M^d)}(1-\alpha))$, where $F_{\chi^2(M^d,\widehat{\nu}_{(\tilde{i}\mapsto i)})}(.)$ represents the CDF of $\chi^2(M^d,\widehat{\nu}_{(\tilde{i}\mapsto i)})$. Then, for the given confidence level of $1-\alpha$, the overall performance of the tests can be quantified by the Youden's J-statistic as:

$$J_{(\tilde{i}\mapsto i)} = 1 - \alpha - \eta_{(\tilde{i}\mapsto i)}(\alpha). \tag{12}$$

The J-statistic takes values in the interval $[0,1]$, summarizing the confidence in the detected GC link by combining type-I and type-II errors. When $J_{(\tilde{i}\mapsto i)} \approx 0$, while the null hypothesis is rejected, there is no strong evidence to choose the alternative over null. On the other hand, when $J_{(\tilde{i}\mapsto i)} \approx 1$, there is strong evidence that the alternative is true. Therefore, $0 < J_{(\tilde{i}\mapsto i)} \leq 1$ can be thought as a normalized strength the GC link $(\tilde{i} \mapsto i)$. We finally construct the GC map $\boldsymbol{\Phi}$, with elements given by:

$$[\boldsymbol{\Phi}]_{i,\tilde{i}} := \begin{cases} \text{sign}\Big(\sum_{k=1}^q \widehat{a}_{i,\tilde{i},k}\Big) J_{(\tilde{i}\mapsto i)}, & (\tilde{i}\mapsto i) \in \mathcal{I} \\ 0, & \text{otherwise} \end{cases}, \tag{13}$$

where the sign of each element shows the excitatory (positive) or inhibitory (negative) nature of the link.

IV. Application to Simulated and Real Data

We first present a simulation study to evaluate the performance of our proposed algorithm, and then demonstrate its utility in application to experimentally-recorded MEG data.

A. Simulation Study

We consider a scenario with $N_x = 100$ sources and $N_y = 50$ sensors, where 10 sources with unknown locations are active. The sensor measurements are available for a duration of $T = 1000$ samples. The observation noise covariance matrix is assumed to be $\mathbf{R} = (0.1)\mathbf{I}_{N_y}$, where $\mathbf{I}_{N_y}$ is the identity matrix with dimension N_y. The sources are generated by a VAR(5) process. In fitting the model parameters, the VAR order (q) is determined using the AIC criterion, [19] and the regularization parameter ($\boldsymbol{\lambda}$) is obtained via two-fold cross-validation.

Fig. 1A shows the ground truth GC structure among a selected subset of the sources (for graphical convenience), and Fig. 1B shows the ground truth J-statistics, which take values of 1 (excitation) or -1 (inhibition). Fig. 1C shows the estimated $\boldsymbol{\Phi}$ matrix for all the active sources (indexed by $1, 2, \cdots, 10$) and a subset of inactive sources (indexed by $11, 12, \cdots, 20$), with FDR corrected at a confidence level of 98%. Note that active sources refer to those that explain 90% of the variance. As a comparison benchmark, the results of the common two-stage procedure that infers the GC links from localized sources via the minimum norm estimate (MNE) algorithm, are also shown in Fig. 1D. The MNE solution for $\widehat{\mathbf{x}}_{1:T}$ is obtained by solving:

$$\widehat{\mathbf{x}}_{q+1:T} = \underset{\mathbf{x}_{q+1:T}}{\operatorname{argmin}} \|\mathbf{x}_{q+1:T}\|_2^2, \text{ s.t } \|\mathbf{y}_{q+1:T} - \mathbf{C}\mathbf{x}_{q+1:T}\|_2^2 \leq \zeta, \quad (14)$$

for some $\zeta \geq 0$. As it can be seen in Fig. 1D, the two-stage procedure results in considerable spatial leakage and therefore the GC matrix significantly differs from the ground truth, with numerous spuriously detected links. Fig. 1E shows 4 selected source activities, with source 6 considered as the target. While including the contribution of source 6 improves the estimation of source 10, it does not do so for sources 7 and 12.

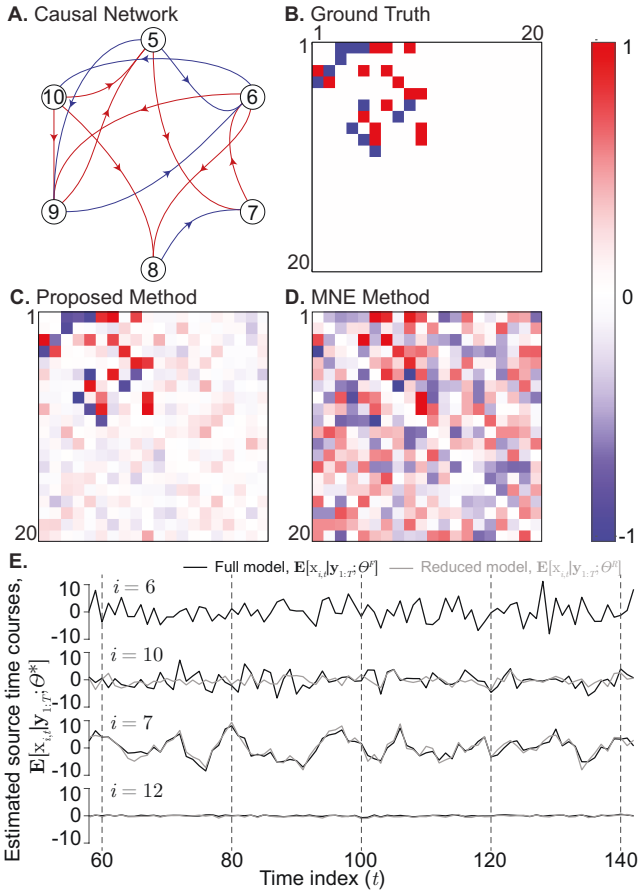

Fig. 1: Simulation results. A. The GC network corresponding to sources $\{5, 6, \ldots, 10\}$, where the blue and red arrows show the inhibitory and excitatory links, respectively. B. Ground truth GC matrix $[\boldsymbol{\Phi}]_{1:20,1:20}$ corresponding to 20 sources. C. Estimated $[\boldsymbol{\Phi}]_{1:20,1:20}$ using the proposed method. D. Estimated $[\boldsymbol{\Phi}]_{1:20,1:20}$ based on the MNE solution. E. Selected estimated source activities. Including the contribution of source 6 improves the estimation of source 10, but does not do so for sources 7 and 12.

B. Application to Experimental MEG Data

Next, we apply our proposed method to infer GC links between different brain regions from MEG recordings. The data were collected during the intervals between auditory tasks while the participant was passively listening to task instructions. The software package MNE-Python [20] is used to band-pass filter the data between $0.5 - 20$ Hz and down-sample to 100 Hz. The measurement noise covariance is estimated from empty room recordings via an automated model selection procedure [10]. The participant's head is co-registered to the 'fsaverage' brain [21] using digitized head shapes. An 'fsaverage' surface source space (namely, 'ico-4') is morphed to the participant's head and the vertices belonging to each of the $N_x = 68$ regions of interest (ROIs) are identified using a parcellation method based on the 'Desican-Killiany' cortical atlas [22]. Using this organization of vertices, we can represent the current density for each ROI by one current dipole placed at its center of mass (defined based on the norm of the lead field vector) [8]. The current dipoles are also directionally constrained to be normal to the average cortical patches, which resulted in a reduced lead field matrix of dimension 155×68. We apply our inference algorithm to two selected 4 second segments of MEG data, and model the sources as a VAR(10) process. Note that even though $N_y > N_x$, a total number of $N_x^2 q = 46240$ parameters need to be estimated from $N_y T/2 = 31000$ data samples (using two-fold cross-validation), which renders the problem ill-posed.

The inferred GC links from the two data segments are shown in Fig. 2A and B, for sources explaining 80% of the data variance. The color of the links encodes the signed J-statistics (red: excitatory, blue: inhibitory), and the black arrow-heads indicate GC link direction. Given that the neural activity pertain to "between task" conditions during these two segments, the GC links show some variability between the two windows shown in Fig. 2A and B. However, some of the main GC links are strikingly persistent. We have singled out five such sources enclosed by purple circles in Fig. 2A, which corresponds to known frontal and sensory cortical areas

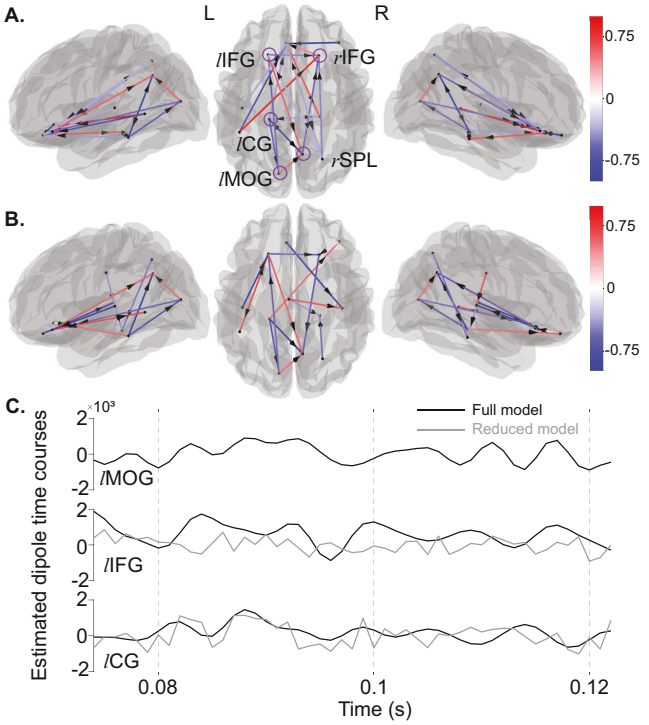

Fig. 2: Application to experimentally-recored MEG data. A and B. The detected GC links from two different 4 second windows of data. The color of the links encodes the signed J-statistics (red: excitatory, blue: inhibitory) and the black arrow-heads show the link direction. FDR is controlled at 3%. The encircled sources (rIFG: right inferior frontal gyrus, lIFG: left inferior temporal grys, lCG: left cingulate gyrus, rSPL: right superior parietal lobe, and lMOG: left medial occipital gyrus) correspond to a consistent sub-network across these two segments of the experiment, which correspond to known areas involved in sensory/cognitive processing. C. Selected source activity estimates. Estimation of lIFG changes considerably by including lMOG, which is not the case for lCG.

involved in sensory/cognitive processing [23]. Fig 2C shows estimated time courses of three selected sources under both full and reduced models. Estimation of lIFG changes considerably by including lMOG as a covariate, which is not the case for lCG. This observation is confirmed by the presence of the GC link (lMOG $\mapsto$ lIFG).

V. Conclusion

In this paper, we considered the problem of extracting Granger causal influences among a number of sources that are indirectly observed via low-dimensional and noisy measurements. Instead of adopting the common two-stage procedure of first solving the inverse problem to estimate the sources followed by GC inference, we directly inferred the GC links from the observations via two nested EM algorithms under a sparse VAR model. Moreover, we developed a statistical framework to assess the significance of the detected links. We illustrated the performance of our proposed methodology on both simulated and experimentally-recorded MEG data. Our results show that the proposed methodology is capable of revealing the GC structure of the underlying sources in a more robust fashion than the common two-stage procedure. Our future work is focused on incorporating the effect of external/sensory stimuli and speeding up the algorithmic implementation in order to efficiently analyze larger volumes of MEG data from sensory and cognitive MEG experiments.

References

[1] S. L. Bressler and A. K. Seth, "Wiener–Granger causality: a well established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.

[2] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI," *Neuroimage*, vol. 25, no. 1, pp. 230–242, 2005.

[3] G. Deshpande, S. LaConte, G. A. James, S. Peltier, and X. Hu, "Multivariate Granger causality analysis of fMRI data," *Human brain mapping*, vol. 30, no. 4, pp. 1361–1373, 2009.

[4] P. Liang, Z. Li, G. Deshpande, Z. Wang, X. Hu, and K. Li, "Altered causal connectivity of resting state brain networks in amnesic MCI," *PLoS One*, vol. 9, no. 3, p. e88476, 2014.

[5] Q. Lu, K. Bi, C. Liu, G. Luo, H. Tang, and Z. Yao, "Predicting depression based on dynamic regional connectivity: A windowed Granger causality analysis of MEG recordings," *Brain Research*, vol. 1535, pp. 52–60, 2013.

[6] M. J. Brookes, P. K. Tewarie, B. A. Hunt, S. E. Robson, L. E. Gascoyne, E. B. Liddle, P. F. Liddle, and P. G. Morris, "A multi-layer network approach to MEG connectivity analysis," *NeuroImage*, vol. 132, pp. 425–38, 2016.

[7] F. Liu, E. P. Stephen, M. J. Prerau, and P. L. Purdon, "Sparse multi-task inverse covariance estimation for connectivity analysis in EEG source space," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2019, pp. 299–302.

[8] A. Wodeyar and R. Srinivasan, "Contribution of structural connectivity to MEG functional connectivity," *bioRxiv*, 2019. [Online]. Available: https://www.biorxiv.org/content/early/2019/09/30/785600

[9] J.-M. Schoffelen and J. Gross, "Source connectivity analysis with MEG and EEG," *Human brain mapping*, vol. 30, no. 6, pp. 1857–1865, 2009.

[10] D. A. Engemann and A. Gramfort, "Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals," *NeuroImage*, vol. 108, pp. 328–342, 2015.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Stat. Soc.: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[12] B. D. Anderson and J. B. Moore, *Optimal filtering*. Courier Corporation, 2012.

[13] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 183–195, 2013.

[14] A. Sheikhattar, S. Miran, J. Liu, J. B. Fritz, S. A. Shamma, P. O. Kanold, and B. Babadi, "Extracting neuronal functional network dynamics via adaptive Granger causality analysis," *Proceedings of the National Academy of Sciences*, vol. 115, no. 17, pp. E3869–E3878, 2018.

[15] S. van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure, "On asymptotically optimal confidence regions and tests for high-dimensional models," *Ann. Stat.*, vol. 42, no. 3, pp. 1166–1202, 2014.

[16] R. R. Davidson and W. E. Lever, "The limiting distribution of the likelihood ratio statistic under a class of local alternatives," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 209–224, 1970.

[17] K. L. Saxena and K. Alam, "Estimation of the non-centrality parameter of a chi squared distribution," *Ann. Stat.*, pp. 1012–1016, 1982.

[18] Y. Benjamini, D. Yekutieli *et al.*, "The control of the false discovery rate in multiple testing under dependency," *Ann. Stat.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[19] J. Ding, V. Tarokh, and Y. Yang, "Model selection techniques: An overview," *IEEE Sig. Proc. Mag.*, vol. 35, no. 6, pp. 16–34, 2018.

[20] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, 2014.

[21] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012.

[22] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[23] J. E. Peelle, I. S. Johnsrude, and M. H. Davis, "Hierarchical processing for speech in human auditory cortex and beyond," *Frontiers in Human Neuroscience*, vol. 4, 2010.