Robust recovery of time-varying functional connectivity in MEG

Rahul Nadkarni Nicholas J. Foti Emily B. Fox

Abstract

A central problem in modern neuroscience is inferring interactions between brain regions, as they can provide insight into neurological disorders and how the brain processes information. Magnetoencephalography (MEG) provides a rich data source for analyzing the dynamics of brain interactions. However, one major challenge with analyzing this data is sensitivity to the definition of brain regions between which interactions are sought. We develop an approach combining state-space and factor analysis models that learns the intrinsic rank of the underlying dynamics. We develop an expectation-maximization algorithm to perform parameter estimation from data, and show in simulated MEG experiments that our method exhibits robustness to the regions included in the analysis. Our method is able to identify the true inter-regional interactions while also learning the underlying rank of the data.

1 Introduction

Much of modern neuroscience research involves collecting large amounts of data from a variety of neuroimaging modalities. While most of this data records some form of brain activation, activity alone does not capture the underlying complex processes. In particular, interactions between regions are thought to underlie many cognitive behaviors and neurological disorders, thus understanding them is paramount. Though many methods exist to quantify brain interactions, we are interested in *time-varying* and *directed functional connectivity*.

Magnetoencephalography (MEG) provides a unique perspective to study the dynamics of brain interactions due to its exceptional temporal resolution ($\sim$ 1000 Hz) and good spatial resolution. However, raw MEG sensor recordings do not correspond to locations on the cortical surface and thus are not directly amenable to analyzing brain interactions. Much work has been devoted to addressing this issue. The most popular approach is the minimum-norm estimate that solves an ill-posed inverse problem and projects the MEG sensor data onto a high-dimensional tessellation of the cortical surface [1]. The resulting *source-space* signals can then either be used directly for connectivity, or they can be aggregated into *regions of interest* (ROIs) for which connectivity can be estimated. For instance, the resulting aggregate ROI signals can be used to fit a time-varying VAR [2], such that the autoregressive coefficients can be used as an estimate of interactions between regions.

The two-stage MNE approach above has some series disadvantages. In particular, spatial and temporal artifacts arise due to the ill-posedness of the problem - projecting from hundreds of sensors to thousands of source points - as well as the fact that most methods perform the first step while ignoring dynamics and correlation structure across the cortical surface. Recently, Yang et al. [3] developed a method to estimate time-varying directed functional connectivity between two ROIs directly from MEG sensor data. The method is based on a linear dynamical system that *integrates out the high-dimensional source-space*, providing a direct map from aggregate ROI signals to MEG sensor data. This method takes an important step away from two-step analyses of MEG data, and is able to learn dynamic connectivity while also allowing for better estimation of cortical signals. The time-varying property of this model allows for learning interesting connectivity structure that would be missed by other models, such as stationary or switching processes. Yang et al. also include a penalty to encourage smoothness in their dynamics, which models the physical constraint that interactions can be expected not to spike unexpectedly, but instead be relatively consistent for timepoints that are close.

We address these issues by incorporating *low-rank* structure into the latent ROI series enabling us to learn the *intrinsic dimensionality* of the dynamics of the observed MEG data. This additional structure creates a hierarchical linear dynamical system for which we develop an EM algorithm to learn the model parametrs and infer ROI activity. We demonstrate our approach on synthetic examples where we vary the

underlying rank structure of the data as well as the number of ROIs. Our method demonstrates the ability to robustly recover underlying low-rank structure when it is present, while maintaining comparable performance to methods that don't account for this structure otherwise.

2 Time-varying linear dynamical systems for MEG data

Magnetoencephalography (MEG) data consists of recordings of the magnetic field emitted by the brain. This magnetic field is the result of an electric current flowing through tens of thousands of aligned neurons close to the cortical surface and is recorded using 306 sensors, 204 gradiometers and 102 magnetometers. Often electroencephalography (EEG) sensors recorded the electrical field emitted by the brain (from neurons orthogonal to those that MEG picks up), in our case using 50 electrodes. Thus, an MEG + EEG sensor recording is denoted $\boldsymbol{y} \in \mathbb{R}^{366}$. We observe MEG+EEG sensor recordings for a single subject performing N independent trials each of which contains $T+1$ time points so that we denote the whole set of observations as $\{\boldsymbol{y}_t^{(n)}\}_{t=0,n=1}^{T,N}$. We also assume that we are interested in inferring the time-varying, directed interactions between a fixed set of p ROIs. We next summarize the approach of [3] to set notation.

Define the *ROI activation series* $\boldsymbol{u}_t^{(n)} \in \mathbb{R}^p$ as the latent activation of each ROI at time t for trial n which we do not observe. Assume that $\boldsymbol{u}_t^{(n)}$ follows a lag-1 vector autoregressive process (VAR):

$$\boldsymbol{u}_0^{(n)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_0) \tag{1}$$

$$\boldsymbol{u}_t^{(n)} = \boldsymbol{A}_t \boldsymbol{u}_{t-1}^{(n)} + \boldsymbol{\epsilon}_t^{(n)}, \;\; \boldsymbol{\epsilon}_t^{(n)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \tag{2}$$

where the matrices $\boldsymbol{A}_t$ encode the time-varying directed interactions between the ROIs. The matrices $\boldsymbol{Q}_0$ and $\boldsymbol{Q}$ are the idiosyncratic covariance matrices between the ROIs.

The key insight of [3] was to directly map this ROI process to the MEG sensor data through a linear map by marginalizing out the high-dimensional source-space represented by a tessellation of the cortical surface and consisting of n vertices (we use $n \approx 5$K in this work). The resulting linear dynamical system mitigates the artifacts introduced by the MNE method described above. Specifically, the MEG sensor data for each trial at each time t is modeled as

$$\boldsymbol{y}_t^{(n)} = \boldsymbol{C}\boldsymbol{u}_t^{(n)} + \boldsymbol{\eta}_t^{(r)}, \;\; \boldsymbol{\eta}_t^{(n)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}), \tag{3}$$

where $\boldsymbol{C} \in \mathbb{R}^{366 \times p}$ maps the ROI signals to the MEG recordings and is given by $\boldsymbol{C} = \boldsymbol{G}\boldsymbol{L}$, where $\boldsymbol{G} \in \mathbb{R}^{366 \times n}$ is the so-called *forward-operator* that is a linearization of Maxwell's equations that describe magnetic fields [4, 5] and is computed a priori using structural MRI information about the subject. The known matrix $\boldsymbol{L} \in \{0,1\}^{n \times p}$ indicates which vertices make up each ROI, each row of which will have a single 1 as we assume ROIs do not overlap. The symmetric matrix $\boldsymbol{R} \in \mathbb{S}^{366}$ models the covariance of the sensor data and consists of three components: $\boldsymbol{R} = \boldsymbol{Q}_{\text{sensor}} + \boldsymbol{G}\boldsymbol{Q}_{\text{source}}\boldsymbol{G}^\intercal$. The matrix $\boldsymbol{Q}_{\text{sensor}}$ models the measurement covariance between the MEG sensors and is estimated from subject baseline data where they were sitting in the MEG machine doing nothing. The matrix $\boldsymbol{Q}_q$ is a $n \times n$ diagonal matrix where each entry takes on one of $p+q$ values, $\{\sigma_j^2\}_{j=1}^{p+1}$. Let $\mathcal{R}(v) = r$ if the rth column of the vth row of L is 1, then $\boldsymbol{Q}_{\text{source}}[m,m] = \sigma^2_{\mathcal{R}(m)}$. In other words, σ_j^2 is shared over all entries of $\boldsymbol{Q}_{\text{source}}$ that correspond to vertices in region j. [3] developed an EM algorithm to infer the parameters of interest: $\boldsymbol{A}_t$, $\boldsymbol{Q}$, $\boldsymbol{Q}_0$, and $\{\sigma_j^2\}_{j=1}^{p+1}$.

[3] evaluated the model using two ROIs and analyzed MEG data from a visual task. Our interest in MEG connectivity differs in that we want to infer interactions from more ROIs, such as ten, simultaneously. This raises the natural questions of i) whether the model can reliably determine interactions when more than two ROIs are present, and ii) how robust the inferred connectivity is to the definitions of the ROIS and the exact set of ROIs chosen? We have empirically observed that the model can indeed address i). But we have observed that the inferred connectivity is not robust. We address ii) in the next section.

3 Latent factor augmentation for robust network recovery

An issue with the MEG-LDS model that contributes to the sensitivity of the learned interaction matrices $\boldsymbol{A}_t$ is that groups of ROIs may actually interact with other ROIs or groups of ROIs. Though the MEG-LDS

model can infer this situation, it often instead infers edges between the ROIs in the group and then a single edge between one of these ROIs and the other ROI. This problem is compounded since the geometry of the brain can cause signal to leak between ROIs when they are close and non-convex.

We address this problem by learning the intrinsic rank of the underlying dynamics and learning an effective number of ROIs. Specifically, we introduce a set of K latent factors $\{\boldsymbol{f}_t^{(n)} \in \mathbb{R}^K\}_{t=0,n=1}^{T,N}$ that will represent the low-dimensional space that the MEG dynamics live on. The latent factors will now follow time-varying dynamics rather than the ROI series:

$$\begin{aligned} \boldsymbol{f}_0^{(n)} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \\ \boldsymbol{f}_t^{(n)} &= \boldsymbol{A}_t \boldsymbol{f}_{t-1}^{(n)} + \boldsymbol{w}_t^{(n)}, \;\; \boldsymbol{w}_t^{(n)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \end{aligned} \tag{4}$$

where $\boldsymbol{A}_t \in \mathbb{R}^K$ encodes the time-varying, directed interactions between the components of the low-dimensional latent factors.

The latent factors are then mapped to the ROI space according to a factor loadings matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times K}$ which are then mapped to the MEG sensors as Eq. (3):

$$\begin{aligned} \boldsymbol{u}_t^{(n)} &= \boldsymbol{\Lambda} \boldsymbol{f}_t^{(n)} \\ \boldsymbol{y}_t^{(n)} &= \boldsymbol{C} \boldsymbol{u}_t^{(n)} + \boldsymbol{\eta}_t^{(n)}, \;\; \boldsymbol{\eta}_t^{(n)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}), \end{aligned} \tag{5}$$

where $\boldsymbol{C}$ and $\boldsymbol{R}$ are defined as in Sec. 2. Low-rank representations have been previously considered for vector autoregressions in order to account for unobserved latent processes [6]. In our case, we wish that each latent factor influences a small set of the specified ROIs. Note that our proposed *factor-LDS* (FLDS) model for MEG converges to the MEG-LDS model when $K = p$, $\boldsymbol{\Lambda} = \boldsymbol{I}$, and the factor noise covariances are learned rather than fixed.

Interpreting the inferred factor connectivity: The inferred $\boldsymbol{A}_t$ matrices encode the time-varying, directed connectivity between the latent factors. However, these do not directly correspond to regions of the brain, and the connections of interest to neuroscientists are between the ROIs. Marginalizing the factor processes out of the model results in the following VAR(1) process for the ROI signals:

$$\boldsymbol{u}_t = \boldsymbol{\Lambda} \boldsymbol{A}_t \boldsymbol{\Lambda}^\dagger \boldsymbol{u}_{t-1} + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top) \tag{6}$$

where $\boldsymbol{\Lambda}^\dagger$ denotes the pseudo-inverse. The quantity $\boldsymbol{\Lambda} \boldsymbol{A}_t \boldsymbol{\Lambda}^\dagger$ encodes the inter-ROI connectivity encoded by the $\boldsymbol{A}_t$ matrices. Due to unidentifiability properties of state-space models with unknown parameters [7], we note that we cannot necessarily recover an interpretable factor loadings or set of dynamics matrices in factor space. However, learned rotations in these matrices will be cancelled out by calculating the ROI-space dynamics as described above, so we can still hope to recover connectivity across ROIs even if the exact connectivity across factors, and the identity of the factors themselves, is only recoverable up to orthogonal transformations.

3.1 Expectation-maximization for learning parameters

We develop an expectation-maximization (EM) algorithm to learn the model parameters for FLDS, $\Theta = \{\{\boldsymbol{A}_t\}_{t=1}^T, \boldsymbol{\Lambda}, \{\sigma_j^2\}_{j=1}^{p+1}\}$. The EM algorithm maximizes the log-likelihood of the data in situations where there is missing data, such as the latent factor processes in our FLDS model. In order to encourage temporal smoothness of the inferred $\{\boldsymbol{A}_t\}$ matrices, we incorporate a temporal regularizer, $\Omega(\{\boldsymbol{A}_t\}_{t=1}^T) = \sum_{t=2}^T ||\boldsymbol{A}_t - \boldsymbol{A}_{t-1}||_F^2$, to obtain *penalized maximum likelihood* estimates of Θ. Specifically, we optimize:

$$\max_{\Theta} \mathbb{E}_{p(\boldsymbol{f}_{1:T}^{1:N} | \boldsymbol{y}_{1:T}^{1:N}, \Theta)} \left[\log p(\boldsymbol{y}_{1:T}^{1:N}, \boldsymbol{f}_{1:T}^{1:N}; \Theta)\right] + \lambda \Omega(\{\boldsymbol{A}_t\}_{t=1}^T), \tag{7}$$

where λ is a regularization parameter allowing us to adjust the amount of temporal smoothness introduced. The EM algorithm alternates between the E-step which computes $\log p(\boldsymbol{f}_{1:T}^{1:N} \mid \boldsymbol{y}_{1:T}^{1:N}, \Theta)$, and the M-step which maximizes Eq. (7).

E-Step: Since our FLDS model consists of linear-Gaussian dynamics and emissions we use the Rauch-Tung-Striebel smoother [8] modified to account for our time-varying dynamics matrix. The quantities of interest are the means and covariances of the smoothed state distributions

$$\begin{aligned} \boldsymbol{f}_{t|T}^{(n)} &= \mathbb{E}\left[\boldsymbol{f}_t^{(n)} \mid \boldsymbol{y}_{0:T}^{(n)}\right] \\ \boldsymbol{F}_{t|T}^{(n)} &= \mathrm{Cov}\left(\boldsymbol{f}_t^{(n)} \mid \boldsymbol{y}_{0:T}^{(n)}\right) \\ \boldsymbol{F}_{t,t-1|T}^{(n)} &= \mathrm{Cov}\left(\boldsymbol{f}_t^{(n)}, \boldsymbol{f}_{t-1}^{(n)} \mid \boldsymbol{y}_{0:T}^{(n)}\right). \end{aligned} \tag{8}$$

which are recursively computed by the RTS smoother.

M-Step: We then use the expected sufficient statistics computed in the E-step to maximize the following expected joint log-probability

$$\begin{aligned} &\min_{\Theta} -\mathbb{E}_{\boldsymbol{f}|\boldsymbol{y}}\left[\log p(\boldsymbol{y}_{0:T}^{1:N}, \boldsymbol{f}_{0:T}^{1:N}; \Theta)\right] + \lambda\Omega(\{\boldsymbol{A}_t\}_{t=1}^T) \\ =&\min_{\Theta} \mathcal{L}_1(\{\boldsymbol{A}_t\}_{t=1}^T) + \mathcal{L}_2(\boldsymbol{\Lambda}, \{\sigma_j^2\}_{j=1}^{p+1}) + \lambda\Omega(\{\boldsymbol{A}_t\}_{t=1}^T) \end{aligned} \tag{9}$$

The constituent objective functions are defined as:

$$\mathcal{L}_1(\{\boldsymbol{A}_t\}_{t=1}^T) = NT\log\det(\boldsymbol{S}_0) + \mathrm{tr}\left(\sum_{t=1}^T \boldsymbol{S}_{1t} - \boldsymbol{A}_t\boldsymbol{S}_{2t}^\top - \boldsymbol{S}_{2t}\boldsymbol{A}_t^\top + \boldsymbol{A}_t\boldsymbol{S}_{3t}\boldsymbol{A}_t^\top\right) \tag{10}$$

$$\mathcal{L}_2(\boldsymbol{\Lambda}, \{\sigma_j^2\}_{j=1}^{p+1}) = NT\log\det(\boldsymbol{R}^{-1}) + \mathrm{tr}\left(\boldsymbol{R}^{-1}\left(\boldsymbol{S}_4 - \boldsymbol{H}\boldsymbol{S}_5^\top - \boldsymbol{S}_5\boldsymbol{H}^\top + \boldsymbol{H}\boldsymbol{S}_6\boldsymbol{H}^\top\right)\right) \tag{11}$$

$$\text{where } \boldsymbol{H} = \boldsymbol{C}\boldsymbol{\Lambda} \tag{12}$$

Where the $\boldsymbol{S}$ matrices are functions of the data and statistics computed in the E-step, defined as

$$\boldsymbol{S}_0 = \sum_{n=1}^N \boldsymbol{f}_0^{(n)}(\boldsymbol{f}_0^{(n)})^\top \tag{13}$$

$$\boldsymbol{S}_{1t} = \sum_{n=1}^N \left(\boldsymbol{F}_{t|T}^{(n)} + \boldsymbol{f}_{t|T}^{(n)}(\boldsymbol{f}_{t|T}^{(n)})^\top\right) \qquad \boldsymbol{S}_4 = \sum_{n=1}^N\sum_{t=0}^T \boldsymbol{y}_t^{(n)}(\boldsymbol{y}_t^{(n)})^\top \tag{14}$$

$$\boldsymbol{S}_{2t} = \sum_{n=1}^N \left(\boldsymbol{F}_{t,t-1|T}^{(n)} + \boldsymbol{f}_{t|T}^{(n)}(\boldsymbol{f}_{t-1|T}^{(n)})^\top\right) \qquad \boldsymbol{S}_5 = \sum_{n=1}^N\sum_{t=0}^T \boldsymbol{y}_t^{(n)}(\boldsymbol{f}_{t|T}^{(n)})^\top \tag{15}$$

$$\boldsymbol{S}_{3t} = \sum_{n=1}^N \left(\boldsymbol{F}_{t-1|T}^{(n)} + \boldsymbol{f}_{t-1|T}^{(n)}(\boldsymbol{f}_{t-1|T}^{(n)})^\top\right) \qquad \boldsymbol{S}_6 = \sum_{t=0}^T \boldsymbol{S}_{1t} \tag{16}$$

We optimize $\{\boldsymbol{A}_t\}$, $\boldsymbol{\Lambda}$, and $\{\sigma_j^2\}_{j=1}^{p+1}$ by iterating over gradient updates for each parameter, using `autograd` [9] to compute gradients. Specifically, we first perform gradient descent with backtracking line search to update $\{\boldsymbol{A}_t\}$, then perform an alternating update on $\{\boldsymbol{\Lambda}, \{\sigma_j^2\}_{j=1}^{p+1}\}$ by: 1) performing up to 20 iterations of gradient descent with backtracking line search on $\boldsymbol{\Lambda}$, then 2) performing up to 20 iterations of L-BFGS to update $\{\sigma_j^2\}_{j=1}^{p+1}$.

4 Related work

Various approaches have been considered to analyze directed functional connectivity from source-localized MEG data. There are two main approaches. The first is that connectivity is estimated directly on the high-dimensional source-space by fitting a VAR model [10]. Due to the dimensionality of the data, this approach used a time-invariant connectivity matrix missing interesting dynamic interactions. Additionally, often vertex-level connectivity is not interesting for neuroscientists. Instead, alternative approaches have

spatially averaged the high-dimensional source-space data into regions of interest for the data at hand and then connectivity is estimated between the ROIs. For instance, switching VAR processes have recently been considered in this setting to infer transient brain networks [11], though other models such as time-varying VAR processes could be considered.

A well-known approach to inferring connectivity from neuroimaging data is the *dynamic causal model* (DCM) which has been applied to MEG data [12]. Unlike the method proposed here, the DCM assumes point sources in each ROI and so does not propagate the signals across the cortical surface. Additionally, the DCM requires a physiologically accurate model describing the dipole dynamics which is parameterized by a time-invariant interaction matrix.

This work is heavily inspired by the state-space model of Yang, et al [3], as well as other similar state-space modeling approaches applied to neuroimaging data [13, 14, 15]. In particular, this method is inspired by the approach of Yang et al. of inferring time-varying, directed functional connectivity directly from MEG sensor data and bypassing source-space entirely. Incorporating the estimation of the ROI signals with learning the connectivity coherently accounts for the uncertainty in the mapping of MEG sensor data to ROI activity.

Low-rank structure has previously been incorporated into VAR processes in order to account for unobserved confounding signals [6]. In that work the dynamics matrix of a VAR process was described as the sum of a sparse and low-rank matrix, $\boldsymbol{A} + \boldsymbol{L}$. A lasso penalty was placed on $\boldsymbol{A}$ to promote sparsity and a nuclear norm penalty was placed on $\boldsymbol{L}$ to promote low-rank to capture the effect of the confounding series. This is different than our proposed approach since we assume that the ROI signals are themselves low-rank, not the dynamics.

5 Experiments

5.1 Synthetic data

To demonstrate the ability of our model to learn intrinsic low-rank structure, we compare to the method described in [3] in our experiments. For all experiments, we used a forward transformation $\boldsymbol{G}$, sensor noise covariance $\boldsymbol{Q}_{\text{sensor}}$, and source-space vertex-to-region mapping $\boldsymbol{L}$ from a real subject who performed an auditory attention task while being recorded by MEG. These matrices accurately capture the physics of the MEG measurement process, and using these in our experiments allows us to model the source- to sensor-space transformation in a way that accurately simulates MEG signals while still maintaining knowledge of the true underlying dynamics to validate our results. This is significantly different from using purely synthetic series, as we can demonstrate that our model performs well on data that exhibits properties similar to real MEG recordings.

In cases where we simulate low-rank data, we simulate series by generating a set of $\{\boldsymbol{A}_t\}_{t=0}^{T}$ matrices, a random factor loadings matrix $\boldsymbol{\Lambda}$, and a set of source vertex noise covariances $\{\sigma_j^2\}_{j=1}^{p+1}$. We then simulate series $\{\boldsymbol{f}_t^{(r)}\}_{t=0,r=1}^{T,N}$ for N trials and $T+1$ timepoints as given by Eq. 4, and use these to generate $\{\boldsymbol{y}_t^{(r)}\}_{t=0,r=1}^{T,N}$ using the process described in Eq. 5. To generate full-rank data, we simulate from the process described in Section 2, generating a random positive-definite matrix scaled to have eigenvalues between 0.1 and 1 to use as both $\boldsymbol{Q}$ and $\boldsymbol{Q}_0$ as well as generating $\{\boldsymbol{A}_t\}_{t=0}^{T}$ and $\{\sigma_j^2\}_{j=1}^{p+1}$, where these parameters are defined as in Section 2.

For each experiment, we apply principal component analysis on the generated $\{\boldsymbol{y}_t^{(r)}\}_{t=0,r=1}^{T,N}$, projecting $\{\boldsymbol{y}_t^{(r)}\}_{t=0,r=1}^{T,N}$, $\boldsymbol{G}$, $\boldsymbol{GL}$, and $\boldsymbol{Q}_{\text{sensor}}$ onto the number of top principal components given by the rank of $\boldsymbol{Q}_{\text{sensor}}$. While this is not necessary, we perform this operation to reduce the dimensionality of the sensor-space structures and allow for faster analyses. We use a value of $\lambda = 100$ for the smoothing parameter in all experiments, determined by inspection as a parameter that induces smooth dynamics for all models without sacrificing performance as compared to $\lambda = 0$.

5.1.1 Comparing performance on low-rank and full-rank series

We simulate MEG data with both low-rank and full-rank structure, and compare the performance of FLDS and MEG-LDS in recovering the true ROI-space dynamics as well as the ability of FLDS to uncover the true intrinsic rank of the data. We use a set of 7 predefined cortical regions, and simulate a set of 366 MEG and EEG sensors. For this experiment, we simulated $\{\boldsymbol{A}_t\}_{t=0}^{T}$ as having a constant value of 0.5 for

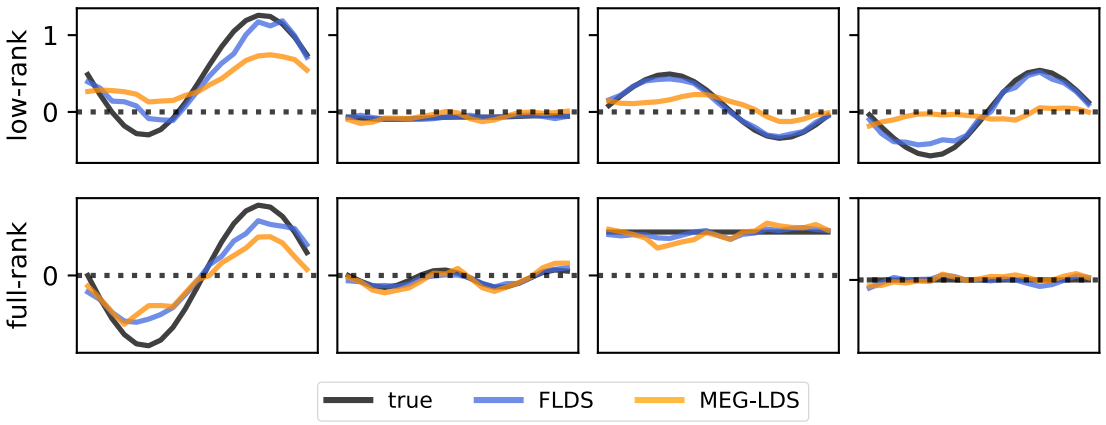

Figure 1: Example traces for selected entries of the dynamics matrices, showing true dynamic connectivity (black) as well as connectivity recovered by FLDS (blue) and MEG-LDS (orange) in the low-rank data setting (top row) as well as the full-rank data setting (bottom row). While both methods perform comparably in the setting where the dynamics are truly full-rank, FLDS outperforms MEG-LDS in the low-rank setting. In particular, FLDS discovers interesting time-varying structure that MEG-LDS is either only able to capture slightly (top row, 3rd from left) or misses entirely (top row, rightmost).

the diagonal entries, a directed connection from the first to the second region as given by a sinusoid with frequency 2π and scaled by a random value in $[-1, 1]$, a directed connection from the second to the first region as given by a sinusoid generated in the same way and pointwise multiplied by the first sinusoid, and no other directed connections between the remaining regions. In the low-rank setting, we simulate $\boldsymbol{\Lambda}$ as a 7×4 matrix with entries as $\boldsymbol{\Lambda}_{ij} \sim \mathcal{N}(0, 1)$, whereas in the high-rank setting we generate and simulate with random positive-definite $\boldsymbol{Q}$ and $\boldsymbol{Q}_0$ as described above. For both the low-rank and high-rank settings, we generated $\{\sigma_j^2\}_{j=1}^{p+1}$ by drawing from a Gamma distribution with shape parameter 2 and scale parameter 0.1. We simulated $N = 200$ trials and $T = 20$ timepoints. For the FLDS model, we fit the model using ranks in the range $k \in \{3, 4, 5, 6, 7\}$. We run 5 restarts with random initializations for the MEG-LDS model and for each value of k for FLDS. To choose the best solution for both models, we perform model selection by choosing the model that minimizes the Bayesian Information Criterion (BIC), defined as

$$\text{BIC} = -2 \log \mathcal{L} + \Phi(\Theta) \log(N) \tag{17}$$

Where $\mathcal{L}$ is the likelihood of the model, N is the number of trials, and $\Phi(\Theta)$ is the number of free parameters, given by the entries of $\{\{\boldsymbol{A}_t\}_{t=0}^{T}, \boldsymbol{Q}, \boldsymbol{Q}_0, \{\sigma_j^2\}_{j=1}^{p+1}\}$ for MEG-LDS and $\{\{\boldsymbol{A}_t\}_{t=0}^{T}, \boldsymbol{\Lambda}, \{\sigma_j^2\}_{j=1}^{p+1}\}$ for FLDS. Note that since the dimensionalities of the estimated parameters for MEG-LDS do not change between random restarts, we are effectively choosing the best result returned by MEG-LDS based on the negative log-likelihood under the model.

We show example traces of the true dynamic connectivity across ROIs along with the connectivity recovered by FLDS as well as MEG-LDS in Fig. 5.1.1, for data with true low-rank structure as well as full-rank structure. These results demonstrate that while both methods perform comparably when the underlying dynamics are full-rank, FLDS is able to better capture the true time-varying connectivity structure in the low-rank setting. Table 1 shows the resulting BIC and root mean-squared error (RMSE) in the $\boldsymbol{\Lambda}\boldsymbol{A}_t\boldsymbol{\Lambda}^\dagger$ matrices for the best random initialization for MEG-LDS and for each value of k used to fit FLDS. We observe that model selection is able to select the correct rank when comparing models with different ranks learned by FLDS.

5.1.2 Demonstrating robustness of network and rank recovery

We also perform a synthetic experiment that demonstrates the ability of FLDS to uncover low-rank structure that would otherwise lead to spurious interactions between regions. We consider a setting with three regions,

Table 1: BIC and RMSE values for FLDS vs. MEG-LDS applied to low-rank and full-rank data. Bolded values indicate best BIC result learned by FLDS, which corresponds to the true rank of $k = 4$ in the low-rank setting and the true rank of $k = 7$ (corresponding to the number of ROIs) in the full-rank setting.

	FLDS					
	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	MEG-LDS
BIC, low-rank	1,695,538	**1,673,580**	1,674,723	1,679,155	1,673,929	1,673,929
RMSE, low-rank	0.253208	**0.066729**	13.827892	47.081412	62.673791	0.190908
BIC, full-rank	1,584,881	1,539,533	1,481,391	1,470,276	**1,458,534**	1478883
RMSE, full-rank	0.164295	0.138631	0.110360	0.084820	**0.056323**	0.094678

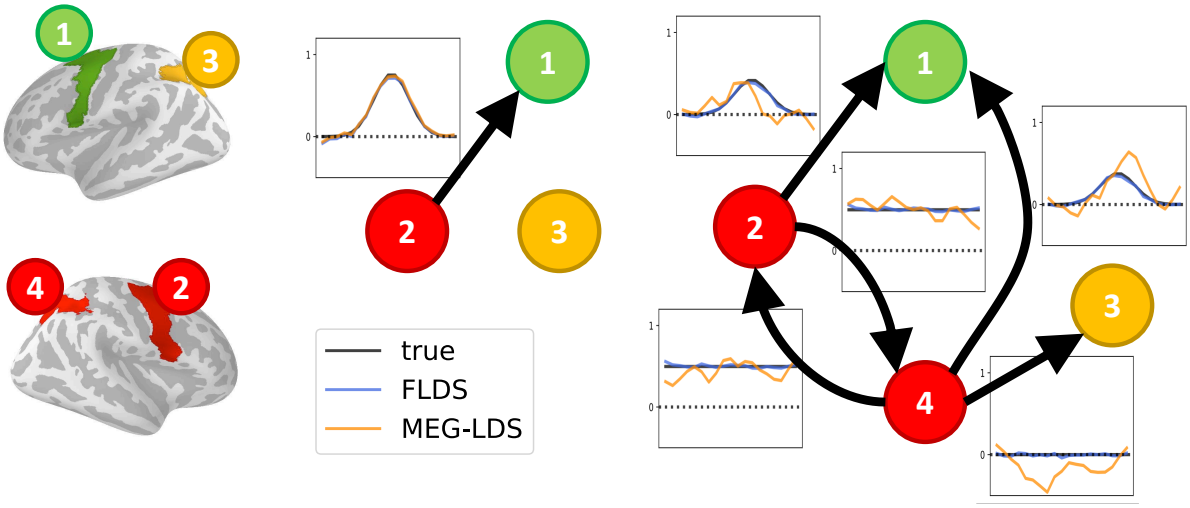

Figure 2: Dynamic connectivity learned by MEG-LDS and FLDS before and after adding in a region with highly-correlated activity to a region in the smaller subset. Full-rank structure with three regions (middle) shows that MEG-LDS and FLDS both reliably recover the true dynamic connectivity structure. The addition of a fourth region that is highly correlated with region two results in a new inferred connectivity graph (right), where MEG-LDS loses performance in recovering the exact connectivity dynamics, but FLDS remains robust. Brain regions used were four regions chosen from the Human Connectome Project Multi-modal Parcellation (left).

Table 2: BIC and RMSE values for FLDS vs. MEG-LDS on recovering robust connectivity. Bolded values indicate best result for FLDS as chosen by BIC, both of which match the true underlying rank in their respective setup. This reflects the ability of FLDS to recognize that the underlying low-rank structure has not changed even when a new ROI is added.

	FLDS			
	$k = 2$	$k = 3$	$k = 4$	MEG-LDS
BIC, low-rank	1,343,528	**1,146,739**	1,150,357	1,134,034
RMSE, low-rank	0.245526	**0.022482**	1,551.136	0.119302
BIC, full-rank	1,313,713	**1,214,488**	–	1,209,032
RMSE, full-rank	0.334446	**0.034920**	–	0.031079

simulating full-rank data by constructing $\boldsymbol{A}_t$ with constant values of 0.9 on the diagonals and a dynamic interaction from the second to the first region in the form of a standard Gaussian kernel evaluated between -4 and 4 and scaled to have a maximum value of 0.75. We set $\boldsymbol{\Lambda}$ to the identity, and all other settings match the previous experiment. We apply both MEG-LDS and FLDS to this model, running FLDS for $k \in \{2, 3\}$ with 5 random restarts for MEG-LDS and for each value of k for FLDS, choosing the best learned results for each by BIC. Both models demonstrate the ability to learn the directed connection as shown on the left-hand side of Fig. 5.1.2, as well as the constant self-connections (omitted to conserve space).

We then convert the setting to the low-rank case by adding a row to $\boldsymbol{\Lambda}$ that is a scaled version of the second row with scale factor 0.9. This models the inclusion of an additional region whose activity is highly correlated with that of region two. This models the inclusion of a new region in the analysis that is highly coupled with a region in the original set, either due to being a neighboring region or having strong functional similarity. We generate data from this model and apply both models in the same fashion as in the full-rank case, this time using ranks $k \in \{2, 3, 4\}$ for FLDS. The results demonstrate that FLDS is able to capture the true connectivity that results from the inclusion of the additional correlated region while learning the correct intrinsic low rank. In contrast, MEG-LDS not only captures these interactions imperfectly but also learns spurious connectivity between region 3 and the other regions. In effect, FLDS is able to recognize that regions 2 and 4 are highly correlated and captures interactions that demonstrate this, while the inclusion of this additional region upsets the ability of MEG-LDS to learn the true underlying low-rank connectivity structure. Table 2 shows BIC and RMSE values that again demonstrate the ability of FLDS to learn the correct rank, and reinforces the fact that while the two models have comparable performance on data with true full-rank connectivity structure, the inclusion of additional regions that are highly correlated with the original set causes the performance of MEG-LDS to drop while FLDS remains robust to the modified structure. Overall, this emphasize the fact that MEG-LDS is highly sensitive to the regions included in the analysis, and the outputted results can be highly untrue to the underlying process.

6 Discussion

We propose a method for learning time-varying directed connectivity from MEG data while modeling brain activity as having a low-rank structure to be more robust to changes in definitions of brain regions as well as the inclusion or exclusion of certain regions from the analysis. We do so by extending previous state-space modeling approaches, modeling activity across ROIs as being a mapping from a lower-dimensional set of latent factors, with the autoregressive dynamics that represent directed connectivity acting in the factor space rather than across brain regions. We demonstrate through experiments on data using real MEG structural information that we achieve comparable performance on full- rank data to a state-space modeling approach that does not explicitly model low-rank structure, and outperform such a model in recovering the true connectivity across ROIs when such structure actually exists. We show specifically that we are able to maintain recovery of interpretable dynamics across ROIs even when regions with redundant signal are incorporated into the analysis. This motivates the use of our method as an exploratory tool for discovering robust connectivity across brain regions when the regions are not clearly defined, when the analysis requires examining connectivity over multiple different sets of ROIs, or when the region signals may be highly correlated.

There are a number of extensions to this approach that would allow for better modeling of the underlying data. In particular, we restrict our analysis to connectivity with first-order autoregressive dynamics. Incorporating a higher lag order would better allow the model to capture long-range interactions. We also allow the model to learn arbitrary dynamics and factor loadings matrices, capturing low-rank structure but without enabling the recovery of an interpretable mapping from latent factors to ROIs. Imposing structured constraints on the factor loadings matrix, such as encouraging a sparse mapping or enforcing that ROI signals be additive combinations of factor signals, may enable the model to learn a more interpretable mapping. We would then not only be able to recognize when a lower-dimensional grouping of regions exists but also learn which ROIs belong to which groups. Finally, while we demonstrate that the approach of learning models with different dimensionalities and using model selection to choose the best one works well in practice, a more adaptive method that enables the identification of the number of factors while jointly learning the mapping from factors to ROIs would greatly simplify the process of using this method for data analysis.

References

[1] M. S. Hamalainen and R. J. Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. 1993.

[2] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer-Verlag, Berlin, Heidelberg, 2005.

[3] Y. Yang, E. A. Aminoff, M. J. Tarr, and R. E. Kass. A state-space model of cross-region dynamic connectivity in MEG/EEG. In *Advances in Neural Information Processing Systems*, 2016.

[4] J. C. Mosher, R. M. Leahy, and P. S. Lewis. EEG and MEG: forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering*, 46:245–259, 1999.

[5] P. C. Hansen, M. L. Kringelbach, and R. Salmelin. *MEG: An Introduction to Methods*. Oxford University Press, New York, New York, 2010.

[6] A. Jalali and S. Sanghavi. Learning the dependence graph of time series with latent factors. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[7] S. Roweis and Z. Ghahramani. A Unifying Review of Linear Gaussian Models. *Neural Computation*, 11(2):305–345, 1999.

[8] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *American Institute of Aeronautics and Astronautics Journal*, 3(8):1445–1450, 1965.

[9] D. Maclaurin, D. Duvenaud, M. Johnson, and R. P. Adams. Autograd: Reverse-mode differentiation of native Python. 2015.

[10] M. Fukushima, O. Yamashita, T. R. Knosche, and M.-a. Sato. MEG source reconstruction based on identification of directed source interactions on whole-brain anatomical networks. *NeuroImage*, 105:408–427, 2015.

[11] D. Vidaurre, A. J. Quinn, A. P. Baker, D. Dupret, A. Tejero-Cantero, and M. W. Woolrich. Spectrally resolved fast transient brain states in selectrophysiological data. *NeuroImage*, 126:81—-95, 2016.

[12] O. David, S. J. Kiebel, L. M. Harrison, J. Mattout, J. M. Kilner, and K. J. Friston. Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, 30(4):1255–1272, 2006.

[13] L. Paninski, Y. Ahmadian, D. G. Ferreira, S. Koyama, K. R. Rad, M. Vidne, J. T. Vogelstein, and W. Wu. A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29:107–126, 2009.

[14] B. Leung, P. Cheung, and B. D. Van Veen. Estimation of cortical multivariate autoregressive models for EEG/MEG using an expectation-maximization algorithm. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2010.

[15] B. Leung, P. Cheung, B. A. Riedner, G Tononi, and B. D. Van Veen. Estimation of cortical connectivity from eeg using state-space models. *IEEE Transactions on Biomedical Engineering*, 57:2122–2134, 2010.