

Infer Functional Connectivity from Partially Observed Neural Signals

Xincheng Tan
Harvard University
Cambridge, MA
xinchengtan@g.harvard.edu

Abstract—Functional connectivity is an imperative aspect in understanding the brain networks. However, due to technology or budget limitations, researchers may not be able to obtain the simultaneous recording of all neuron pairs in their experiments, which hinders the neuron connectivity analysis under the traditional statistical frameworks. In this study, we investigate a data-driven approach to infer functional connectivity when only a subset of neurons are simultaneously recorded. Our method is able to recover the sample covariance with over 60% correlation to the true sample covariance matrix in off-diagonal entries, and the derived connectivity graph has an average F1 score of over 0.45 compared to the graph estimated under full observations.

Index Terms—Functional connectivity, Graphical models, Matrix completion, Kernel methods, Computational neuroscience

I. INTRODUCTION

The connectivity pattern among neuron populations is a fundamental topic in understanding the brain. There are two major types of connectivity: structural and functional connectivity. Structural connectivity is defined as the existence of the white matter tracts physically interconnecting the brain regions, while functional connectivity is the statistical correlation between neural signals, such as fMRI signals [14]. Although structural connectivity provides anatomical basis for how the neurons may interact with each other, it does not necessarily contain functional connectivity – researchers have discovered that certain brain regions without physical connections can exhibit strong functional connectivity [14]. In fact, functional connectivity provides extra insight in the clinical setting. For example, studies have found a strong correlation between static functional connectivity and certain neurological diseases such as depression, schizophrenia, and Alzheimer’s disease [8].

Recently, with the development in functional neuroimaging, researchers are able to quantify and record the activities of a large neuron population in various animals. In particular, two-photon calcium imaging technology enables simultaneous monitoring of thousands of neurons in mice without cortical invasion. To derive meaningful insights from these video datasets, a powerful calcium imaging analysis pipeline, which includes neuron detection, motion correction, spike inference etc., has been developed to extract more computationally tractable neural signals [10].

However, the cost of acquiring and maintaining the two-photon calcium imaging technology could be prohibitively high, while more affordable options are only able to record a small set (on a scale of tens and hundreds) of neurons in each cortex simultaneously such that some neuron blocks are never observed at the same time. These constraints pose significant challenges for estimating the functional connectivity with the off-the-shelf statistical models, which require at least some periods of full observations of the entire neuron population.

In this study, we use Graphical Lasso estimator as the primary framework for modeling the neuronal functional connectivity [5]. To recover the missing entries in the connectivity matrix, we develop a novel data-driven approach that incorporates pairwise auxiliary information, such as nuclei distance and tuning curve similarity, into inferring the connectivity network.

II. RELATED WORKS

A. Graphical Models

Graphical models combine graph theory and probability theory to infer complex network structure within large systems. They aim to uncover an underlying graph $G(V, E)$, where each node represents a feature of interest and each edge reflects the conditional dependence between the two features. To estimate the functional connectivity, we focus on the existence of the connection rather than the direction of such connection. Thus, we only discuss undirected graphical models for the rest of this paper.

Formally, let $X = (X_v : v \in V = \{1, 2, \dots, p\})$ be a p -dimensional random vector indexed by the vertices of an undirected graph $G = (V, E)$ where $E \subseteq V \times V$. Moreover, we say X satisfies the *pairwise Markov property* with respect to G if

$$X_i \perp X_j | X_{V \setminus \{i, j\}} \iff (i, j) \notin E \quad \forall i, j \in V, i \neq j \quad (1)$$

The pairwise Markov property translates each absence of an edge into “full” conditional independence. In other words, two random variables are conditionally independent if and only if the nodes representing them are not connected under the graphical model representation.

B. Gaussian Graphical Model

A graphical model is Gaussian when $x = (x_1, \dots, x_n)$ follows a Gaussian distribution $\mathcal{N} = (\mu, \Sigma)$ where μ is the

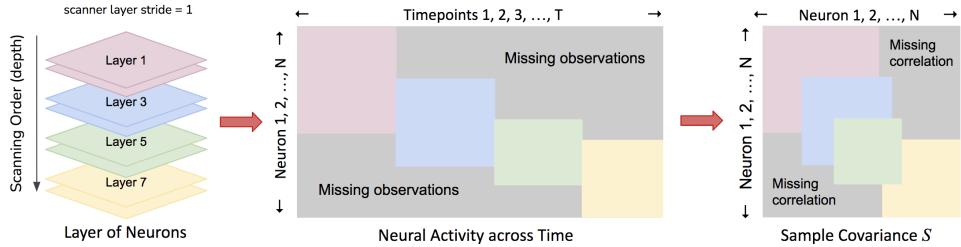


Fig. 1: Diagram of a scanning pattern that leads to missing observations and incomplete sample covariance matrix. The scanner captures two adjacent layers at a time and scans one layer deeper after a particular time period, so only every two consecutive neuron layers have simultaneous observations, leading to an incomplete sample covariance matrix.

mean vector and Σ is the covariance matrix. We let $\Theta = \Sigma^{-1}$ denote the inverse covariance matrix, which is also referred to as *precision matrix*. Gaussian graphical model (GGM) uses the maximum likelihood estimator (MLE) to estimate Θ . Let S be the sample covariance matrix computed from a set of i.i.d observations under a Gaussian distribution. The MLE maximizes the following log-likelihood function

$$L_{GGM}(\Theta) = \log \det(\Theta) - \text{tr}(S\Theta) \quad (2)$$

Given a precision matrix Θ , we can obtain the corresponding Gaussian conditional dependence graph by determining the zero entries in Θ . We say i and j are conditionally independent if and only if the precision matrix entry of the pair equals zero [3]. Formally, we have

$$\Theta_{ij} = 0 \iff (i, j) \notin E \quad (3)$$

In the context of functional connectivity estimation, x represents the activity of n neurons at a particular time, and a non-zero entry $\hat{\Theta}_{ij}$ means neuron i and neuron j are functionally connected.

C. Graphical Lasso

A popular variant of general undirected Gaussian graphical model is the *Graphical Lasso* (GLasso) proposed by Yuan and Lin in [15] and Banerjee *et al.* in [1]. Based on the general GGM, it adds an ℓ_1 regularization term in the MLE objective function:

$$L_{GLasso}(\Theta) = \log \det(\Theta) - \text{tr}(S\Theta) - \lambda \|\Theta\|_1 \quad (4)$$

where λ is a regularization parameter that controls the sparsity of the graph estimation.

Compared to traditional GGM, the formulation of GLasso has a few advantages. First, after adding the regularization term, its objective function is still convex and the convex optimization can be solved in polynomial time. Second, when the number of observations is insufficient ($n < p$), GLasso has strong statistical guarantees which include convergence in Frobenius and spectral norm [7]. Third, GLasso provides a hyperparameter λ to control the sparsity in the graph while preserving the statistical significance of the selected connections, which is very convenient if we are only interested in certain strong connections and wish to ignore the weaker ones.

D. Matrix Completion with Missing Data

In the literature of conditional dependence graph and precision matrix estimation, a few existing methods have been developed to handle noisy and missing observations. Fan *et.al.* propose an alternating direction method of multipliers (ADMM) algorithm to estimate precision matrices with indefinite input and potentially nonconvex penalties [4]. Kohler and Xing study a two-step procedure for estimating sparse precision matrices from data with missing values, which is tractable in high-dimensions and does not require imputation of missing values [9]. Städler and Bühlmann develop an efficient EM algorithm for estimating the precision matrix in the high-dimensional multivariate normal model in presence of missing data [13]. However, these methods only apply when the missing values do not lead to an incomplete sample covariance matrix. Moreover, they assume that the observations are *missing at random*, which does not hold in our problem context where neural signals are missing by blocks.

An alternative approach of leveraging graphical models for functional connectivity estimation under missing data is to recover the sample covariance matrix. Since sample covariance matrix is always positive semidefinite (PSD), matrix completion algorithms tailored for PSD matrices could potentially apply here. In [2], Bishop and Yu propose a symmetric positive semidefinite matrix completion method by stitching in order the rank-factorized components of each revealed principle submatrices. However, their method imposes hidden constraints on the matrix rank and fails to recover the PSD matrix when the rank of the full matrix is high. In our dataset, the true sample covariance matrix is almost always positive definite, and thus a low-rank approximation is not desirable in our setting.

III. PROPOSED METHOD

A. Data-driven Covariance Completion

In this study, we propose a data-driven approach to estimating functional connectivity by leveraging the auxiliary information that correlates with the connectivity. Even though certain neuron pairs are not simultaneously recorded, the behavioral features of the subject, such as running speed and pupil area, are recorded throughout the entire experiment

session. We hypothesize that neural functional connectivity strongly associates with how similar each pair of neuron act in accordance with certain behavioral events, and in identifying such associations, we are able to recover their functional connectivity.

Denote $X \in R^{N \times T}$ as a neural signal matrix where N is the number of neurons and T is the number of time points. Due to block missingness in X , its sample covariance matrix S is incomplete. Let S_{obs} denote the covariance among pairs with simultaneous recordings and \overline{S}_{obs} denote the pairs with missing values in S . Note that such missingness is not random – it has a fixed pattern determined by the scanning pattern of the calcium imaging scanner. For instance, if the scanner focuses on adjacent layers of neurons in an alternating fashion at a time, we would only have simultaneous observations in among interleaving neuron layers but not in any two adjacent layers.

Denote $P \in R^{p \times (N \times N)}$ as the pairwise similarity matrices over p different variables. Each $P_k \in R^{N \times N}$ is a complete and positive semidefinite matrix.

In our method, we aim to learn a function f that maps each $P_{:,i,j} \in R^p$ to the corresponding $(S_{obs})_{i,j} \neq \emptyset$ and use the learned function f to infer the missing values $S_{\overline{obs}}$. If the recovered sample covariance \hat{S} is not positive semidefinite, we use existing methods for nearest correlation matrix (NCM) to correct \hat{S} to its closest positive semidefinite matrix [6]. Finally, we apply GLasso on \hat{S} with the regularization parameter λ to estimate the precision matrix $\hat{\Theta}$. The pseudocode of our method is summarized in Algorithm 1 below and Fig 2 is its diagram.

Algorithm 1 Auxiliary-feature-aided Graphical Lasso

Input: Incomplete data matrix $X \in R^{N \times T}$, regularization strength λ , pairwise feature matrix $P \in R^{p \times (N \times N)}$, predictive model f

Output: Precision matrix $\hat{\Theta}$ and covariance matrix $\hat{\Sigma}$

- 1: $S \leftarrow S_{obs} + S_{\overline{obs}}$, an incomplete sample covariance matrix of X ;
- 2: Fit f which maps P_{obs} to S_{obs} entry-wise;
- 3: $\hat{S} \leftarrow S_{obs} + f(P)_{\overline{obs}}$;
- 4: **if** \hat{S} is not positive semidefinite **then**
- 5: Perturb \hat{S} to the nearest PSD matrix;
- 6: $\hat{\Sigma}, \hat{\Theta} \leftarrow \text{Graphical Lasso}(\hat{S}, \lambda)$;
- 7: **Return** $\hat{\Sigma}, \hat{\Theta}$;

B. Pairwise Feature Matrices

In our problem context, the pairwise feature matrices are constructed from two measures: physiological proximity and tuning curve similarity. For the former, we choose the Euclidean distance since the neuron position is in the 3D coordinate system and is low-dimensional. For the latter, existing literature on functional connectivity and tuning curve leverages Pearson correlation and cosine similarity to model how neuron activity is tuned to a range of the behavioral or stimuli variables [11].

We extend the current approaches by generalizing the similarity measure to the kernel function representation. The Pearson correlation corresponds to the normalized linear kernel while the cosine similarity is a kernel itself. Since the kernelized representation is able to encode a high-dimensional feature space, we expect more sophisticated kernels would enhance the predictive power of f . Specifically, we further utilize the RBF kernel to encode the tuning curve similarity.

IV. EXPERIMENTS

A. Datasets

The dataset in this study is gathered and published by Stringer *et al.* [12]. It includes nine simultaneous calcium imaging recordings of the mice primary visual cortex when they were performing spontaneous activities on a treadmill. Each recording captures over ten thousand neurons over 9 to 11 planes equally spaced with $35 \mu\text{m}$ in depth. The full recordings span from 90 to 120 minutes with an imaging frequency of either 2.5 or 3 Hz. The auxiliary variables include nuclei 3D location, pupil area, running speed, preprocessed facial and whisker image SVD etc. Fig 3 shows a subset of neuron 3D location with three example neural signals in one recording session, and to the right is the recorded running speed and pupil area series throughout the session.

B. Setup and Procedure

In our experiment, we take a subset of neuron cubes from four consecutive neuron layers in five sessions: M150824_MP019_2016-04-05, M160907_MP028_2016-09-26, M161025_MP030_2016-11-20, M161025_MP030_2017-06-16, M161025_MP030_2017-06-23. Each selected subset contains 200 to 400 neurons over a period of 30 minutes.

We first fit a GLasso model on the full observations for each recording data, with the corresponding λ selected via 5-fold cross validation. Note that when computing the sample covariance matrix, we in fact normalize the data by using the sample correlation matrix to account for the different neuron activity baseline across different layers.

Then, we simulate the missing observations by neuron plane. As in Fig 4, only neurons in the adjacent two layers have simultaneous observations. We assume the imaging tool scans the cortex periodically from cortex surface to bottom and focuses on every two adjacent layers for around 10 seconds.

Next, we fit a predictive model f on the pairwise feature matrices. We leverage two types of predictive models: linear regression and ridge regression. In the ridge regression fitting, the regularization strength is automatically selected based on 5-fold cross validation.

To construct the predictors of f , we use the tuning curve similarity of the following behavioral features: running speed, change in running speed, pupil area and change in pupil area. Since these variables are all continuous, we bucket their values between their respective range with arbitrary number of bins.

After f is learned on the pairwise feature matrices for all the observed pairs, we apply it to infer the missing entries in sample covariance. In all our experiments, the resulting \hat{S} is

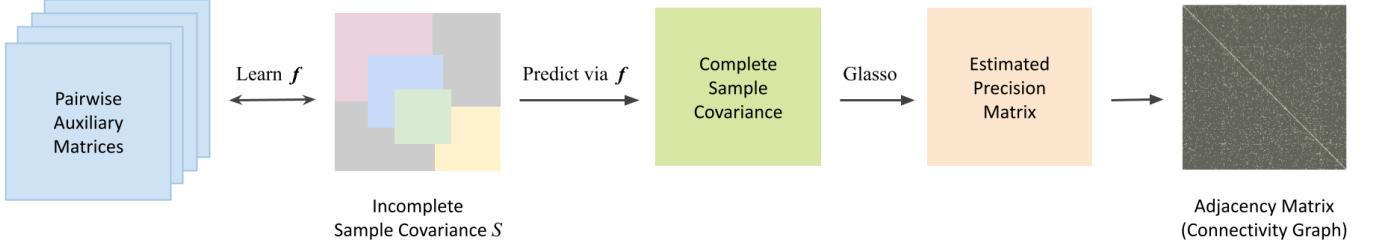


Fig. 2: Diagram of Algorithm 1

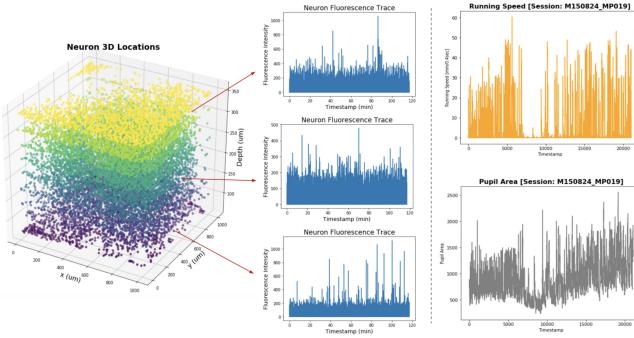


Fig. 3: Nuclei 3D location (left), three example neural signals (middle), mouse running speed and pupil area (right) in recording M150824_MP019_2016-04-05

always non-PSD, so we use the alternating projection method developed in [6] to perturb it to the nearest PSD matrix.

At last, we apply GLasso with a prespecified λ on \hat{S} to estimate the precision matrix. The connectivity graph is then derived based on its non-zero entries with an arbitrary cutoff of 10^{-8} . More details can be found in our GitHub repository: <https://github.com/XinchengTan/AuxIAFunConn.git>

C. Evaluation Metrics

We evaluate our method based on two aspects: the quality of the inferred sample covariance \hat{S} and GLasso-estimated precision matrix $\hat{\Theta}$, as well as the similarity of the connectivity graph compared to that under full observations. Below is a full list of all the metrics covered in this study:

- Rooted mean squared error (RMSE) of sample covariance recovery and the downstream precision matrix estimation: $\frac{1}{N} \|S_{full} - \hat{S}\|_{frob}$ and $\frac{1}{N} \|\Theta_{full} - \hat{\Theta}\|_{frob}$
- Off-diagonal Pearson correlation of S_{full} and \hat{S} , and Θ_{full} and $\hat{\Theta}$
- Accuracy, recall, precision and F1 score of the recovered graph with respect to the GLasso-estimated connectivity under full data.

Note that since the connectivity graph is typically very sparse, accuracy is not very indicative and we use F1 score as our

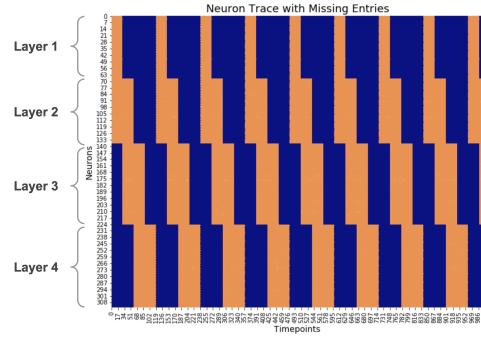


Fig. 4: Example of a partially observed data matrix. The orange blocks represent observed neural activities while the dark blue blocks are missing observations. Some neuron layers, e.g. layer 1 and 4, are never simultaneously recorded.

primary performance indicator of the functional connectivity estimation.

V. RESULTS

A. Auxiliary Predictors

We examine how different auxiliary variables in the pairwise feature matrices influence the connectivity estimation. Across all evaluation metrics as in Table I, II and III below, there is not a significant difference in their predictive power of the missing entries of the connectivity graph when the sample covariance completion function f is random forest and XGBoost regressor. However, when f is linear regression, using all similarity features actually worsen the performance. This is likely due to the limitation of its representation power.

B. Similarity Encoder

We experiment with different pairwise neuron similarity encoders, linear, cosine, RBF and polynomial ($d = 3, 4$) kernel, under different covariance completion model and similarity features. From Table I, all linear, cosine and RBF kernels have similar performance in recovering the final connectivity graph, reaching an almost uniform F1 score of 0.46. However, polynomial kernels give the worst

TABLE I: Average F1 score of connectivity graph

Covariance Completion f	Distance	Pupil Area	Pupil Area Change	Running Speed	Acceleration	All Five
Linear Kernel						
Linear Regression	0.45	0.46	0.45	0.45	0.44	0.43
Random Forest Regressor	0.46	0.46	0.46	0.46	0.46	0.46
XGBoost Regressor	0.46	0.46	0.46	0.46	0.46	0.46
Cosine Kernel						
Linear Regression	0.45	0.45	0.45	0.44	0.45	
Random Forest Regressor	0.46	0.46	0.46	0.46	0.46	
XGBoost Regressor	0.46	0.46	0.46	0.46	0.46	
RBF Kernel						
Linear Regression	0.45	0.46	0.46	0.46	0.46	0.46
Random Forest Regressor	0.46	0.46	0.46	0.46	0.46	0.46
XGBoost Regressor	0.46	0.46	0.46	0.46	0.46	0.46
Polynomial Kernel ($d = 3$)						
Linear Regression	0.45	0.46	0.38	0.46	0.42	0.41
Random Forest Regressor	0.46	0.46	0.39	0.46	0.42	0.41
XGBoost Regressor	0.46	0.46	0.39	0.46	0.42	0.41
Polynomial Kernel ($d = 4$)						
Linear Regression	0.45	0.46	0.38	0.46	0.42	0.40
Random Forest Regressor	0.46	0.46	0.38	0.46	0.42	0.40
XGBoost Regressor	0.46	0.46	0.39	0.46	0.42	0.41

TABLE III: Average Off-diagonal Matrix Correlation of S, \hat{S}

Covariance Completion f	Distance	Pupil Area	Pupil Area Change	Running Speed	Acceleration	All Five
Linear Kernel						
Linear Regression	0.59	0.56	0.54	0.57	0.55	0.50
Random Forest Regressor	0.62	0.63	0.62	0.62	0.62	0.63
XGBoost Regressor	0.62	0.62	0.62	0.62	0.62	0.62
Cosine Kernel						
Linear Regression	0.59	0.6	0.57	0.58	0.55	0.52
Random Forest Regressor	0.62	0.62	0.62	0.62	0.62	0.62
XGBoost Regressor	0.62	0.62	0.62	0.62	0.62	0.62
RBF Kernel						
Linear Regression	0.59	0.62	0.62	0.62	0.62	0.62
Random Forest Regressor	0.62	0.62	0.62	0.62	0.62	0.62
XGBoost Regressor	0.62	0.62	0.62	0.62	0.62	0.62
Polynomial Kernel ($d = 3$)						
Linear Regression	0.59	0.52	0.47	0.52	0.51	0.47
Random Forest Regressor	0.62	0.52	0.47	0.52	0.51	0.47
XGBoost Regressor	0.62	0.53	0.47	0.53	0.51	0.47
Polynomial Kernel ($d = 4$)						
Linear Regression	0.59	0.53	0.47	0.53	0.51	0.47
Random Forest Regressor	0.62	0.53	0.47	0.53	0.51	0.47
XGBoost Regressor	0.62	0.53	0.47	0.53	0.51	0.47

performance, especially when the feature set contains pupil area tuning curve.

From the perspective of completing the sample covariance as in Table II and III, RBF kernel leads to the best performance when the covariance completion model is linear regression, reaching an average RMSE of 0.025 and off-diagonal correlation of 0.62 across all combinations of the feature set and similarity encoder, which is on par with the best of linear regression. This observation applies to any individual similarity feature as well as all five together.

However, similarity encoders has no impact on the performance when we use random forest and XGBoost regressor to complete sample covariance. They both saturate at an average RMSE of 0.025 and off-diagonal correlation of 0.62 across all combinations of the feature set and similarity encoder, which is on par with the best of linear regression. This is because random forest and XGBoost already have a sufficient parameter space to regress on the existing pairwise features as opposed to a simplistic linear model, which requires a more sophisticated kernel to implicitly encode non-linear feature information to enhance its power.

TABLE II: Average RMSE of Sample Covariance S, \hat{S}

Covariance Completion f	Distance	Pupil Area	Pupil Area Change	Running Speed	Acceleration	All Five
Linear Kernel						
Linear Regression	0.45	0.03	0.03	0.029	0.03	0.035
Random Forest Regressor	0.46	0.025	0.025	0.025	0.025	0.025
XGBoost Regressor	0.46	0.025	0.025	0.025	0.026	0.025
Cosine Kernel						
Linear Regression	0.45	0.027	0.029	0.028	0.03	0.033
Random Forest Regressor	0.46	0.025	0.025	0.025	0.025	0.025
XGBoost Regressor	0.46	0.025	0.025	0.025	0.025	0.025
RBF Kernel						
Linear Regression	0.45	0.025	0.025	0.025	0.025	0.025
Random Forest Regressor	0.46	0.025	0.025	0.025	0.025	0.025
XGBoost Regressor	0.46	0.025	0.025	0.025	0.025	0.025
Polynomial Kernel ($d = 3$)						
Linear Regression	0.45	0.032	0.046	0.032	0.036	0.042
Random Forest Regressor	0.46	0.032	0.046	0.032	0.036	0.042
XGBoost Regressor	0.46	0.032	0.046	0.032	0.036	0.042
Polynomial Kernel ($d = 4$)						
Linear Regression	0.45	0.031	0.046	0.031	0.036	0.043
Random Forest Regressor	0.46	0.031	0.046	0.031	0.036	0.043
XGBoost Regressor	0.46	0.031	0.046	0.031	0.036	0.043

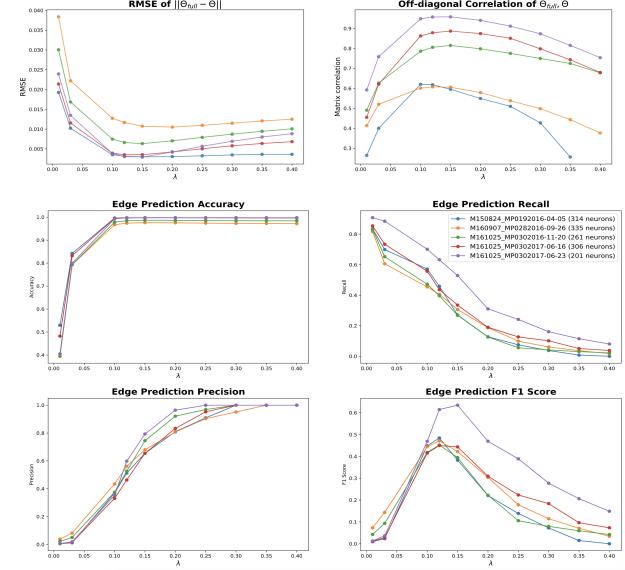


Fig. 5: Quality of GLasso-estimated Precision Matrix (top row) and Functional Connectivity (middle and bottom rows) over Different λ s. x -axis is the regularization strength λ . y -axis is the respective performance measures. Each color represents a recording session.

C. Regularization Strength

Another important parameter that influences the connectivity estimation is the regularization strength. For all five sessions, the optimal λ on the full recording is 0.1 selected by cross validation.

As shown in Fig 5, when we adjust λ in the GLasso on the partial data, the quality of both the estimated precision matrix and the connectivity graph first increases and then decreases, reaching a sweet spot at $\lambda \approx 0.13$ for all five sessions. It is slightly higher than the optimal value under full data because the inferred sample covariance matrix is not perfect and introduces noise. Therefore, it requires a stronger regularization strength to preserve the same level of sparsity. However, if λ is too large, the model suffers from a high miss rate in that the estimated connectivity is overly sparse and fails to detect a lot of edges. On the other hand, when λ is too small, the estimated graph is too dense and false discovery rate is very high.

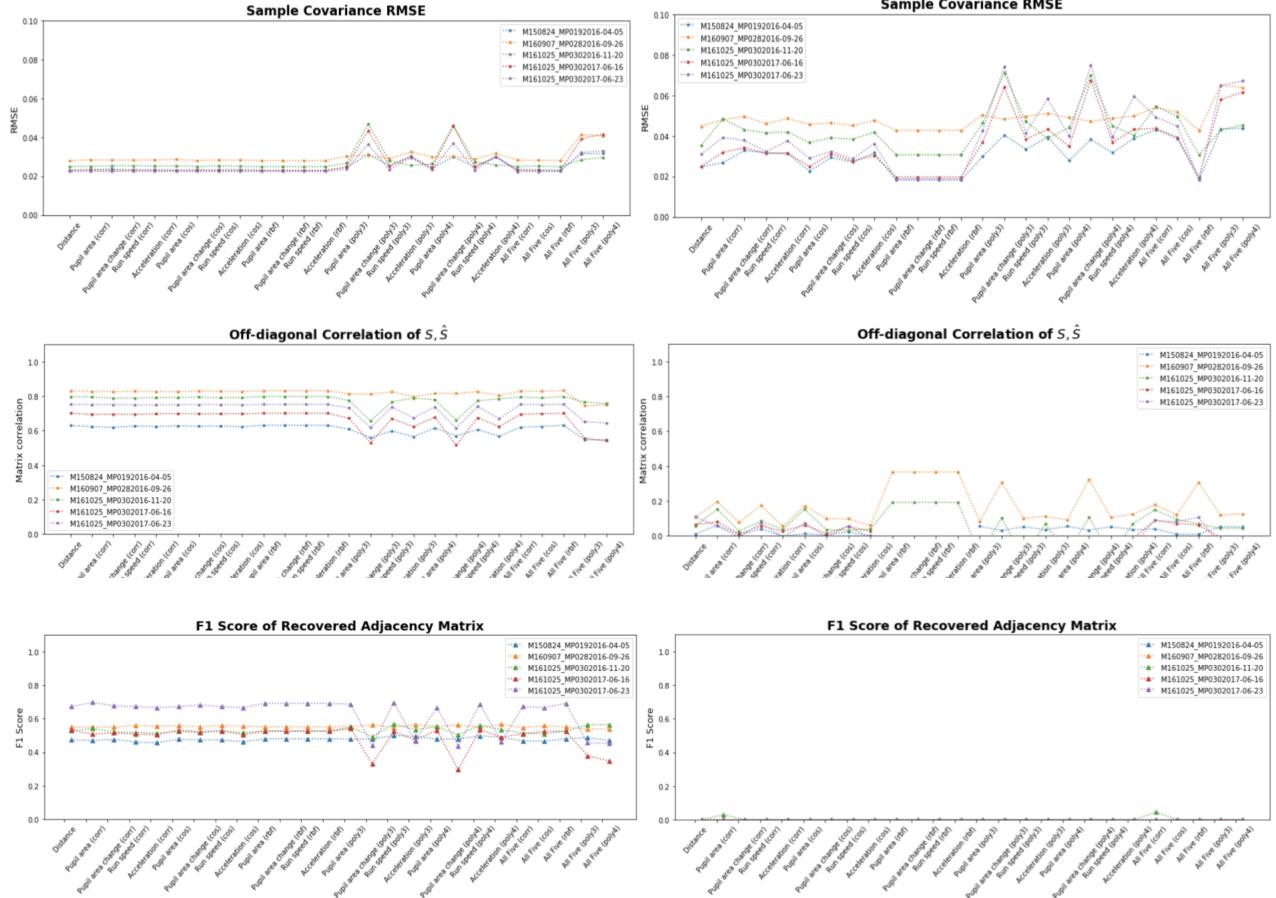


Fig. 5: Performance on neurons with (left) and without (right) simultaneous observations. For all plots, x -axis is the combination of auxiliary feature set and the similarity encoder. From top to bottom, the plots are the RMSE of recovered sample covariance, off-diagonal matrix correlation of recovered sample covariance and the F1 score of the estimated connectivity pattern. Left plots are on the neuron pairs with simultaneous observation. Right plots are on the neuron pairs without simultaneous observation. Each color represents a recording session.

D. With or without Simultaneous Observations

To gain a clearer picture of how the recovered entries contribute to the overall performance described above, we further analyze the connectivity recovery on the neuron pairs with and without simultaneous observations separately. Fig 5 shows the recovery performance over the neuron pairs with and without simultaneous observations. Both have relatively small RMSE of lower than 0.06, but the matrix correlation offers a different picture. Among neurons with simultaneous observations, the matrix correlation is quite high (> 0.6). However, among neurons without simultaneous observations, these entries are inferred by the covariance completion function and have a low correlation (< 0.4) to the true values. This implies that the covariance completion function is not generalizing the data well.

When evaluating the connectivity pattern, we see that our algorithm does not perform well in all combinations of the feature set and similarity kernel, since the F1 score is almost all 0s. This is partially due to the internal edge sparsity. Even

when precision and recall are mostly 0, the accuracy is still very high, which means that the ground truth connectivity graph is very sparse.

VI. DISCUSSION

In this study, we investigate a data-driven approach to estimate the missing entries in the sample covariance matrix of a large neuron population under severe and non-random missingness. We hypothesize that neural functional connectivity is associated with their response similarity to the behavioral features, such as pupil area and running speed. We use kernel methods to encode such similarity in the tuning curve to these behavioral events, and leverage such similarities to infer the missing entries in the sample covariance. We experiment with various kernel types and covariance completion functions, achieving an average F1 score of 0.46 in the estimated connectivity graph on the overall neuron population.

There are important limitations of our study. Further analysis suggests that such performance is mainly contributed

by the neuron pairs with simultaneous observations, which do not rely on the covariance completion function. Our results suggest that the feature set and the similarity kernels explored so far fail to accurately recover the missingness in the covariance matrix. Future work lies in the following aspects:

- 1) Our study only compares four similarity kernels: linear kernel, cosine kernel, RBF and polynomial kernel. It would be worthwhile to experiment with other kernels, such as Fourier kernel and neural tangent kernel, which might have different power of encoding the similarity information between neuron pairs. We need to further examine their performance in recovering connectivity when combined with different predictive models.
- 2) The dataset used in this study is limited in the behavioral feature set. It is important to explore other datasets which contain a richer set of features, such as visual stimuli, tail position, etc.
- 3) Based on the performance difference in varying λ s, it would be ideal to automate λ selection under partial data without knowing the optimal parameter for the full data. To achieve this, we can fit GLasso under multiple subsets of neurons with simultaneous observations. Cross validating on all subsets, we can obtain a list of optimal λ for each subset and ensemble them to obtain a proper value for GLasso on the partial data.
- 4) This study considers only a particular scanning pattern: every few *consecutive* layers are simultaneously recorded. There are other scanning patterns, such as every few *interleaving* layers or every few adjacent cuboids of neurons are recorded at a time. We need to further experiment our data-driven method on different scanning patterns.

ACKNOWLEDGMENT

We acknowledge Bahareh Tolooshams and Spandan Mandan for the helpful discussion and suggestions.

REFERENCES

- [1] O. Banerjee, L. E. Ghaoui, and A. d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, 9(15):485–516, 2008.
- [2] W. E. Bishop and B. M. Yu. Deterministic symmetric positive semidefinite matrix completion. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.
- [3] M. Drton and M. H. Maathuis. Structure learning in graphical modeling, 2016.
- [4] R. Fan, B. Jang, Y. Sun, and S. Zhou. Precision matrix estimation with noisy and missing data, 2019.
- [5] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the lasso, 2007.
- [6] N. J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002.
- [7] C.-J. Hsieh, M. A. Sustik, I. S. Dhillon, and P. Ravikumar. Quic: Quadratic approximation for sparse inverse covariance estimation. *J. Mach. Learn. Res.*, 15(1):2911–2947, Jan. 2014.
- [8] M. M. Jones DT, Vemuri P. Non-stationarity in the resting brain's modular architecture. *PLoS One*, 7, 2012.

- [9] M. Kolar and E. Xing. Estimating sparse precision matrices from data with missing values. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, 1, 01 2012.
- [10] E. A. Pnevmatikakis. Analysis pipelines for calcium imaging data. *Current Opinion in Neurobiology*, 55:15 – 21, 2019. Machine Learning, Big Data, and Neuroscience.
- [11] I. Stevenson, B. London, E. Oby, N. Sachs, J. Reimer, B. Englitz, S. David, S. Shamma, T. Blanche, K. Mizuseki, A. Zandvakili, N. Hatsopoulos, L. Miller, and K. Kording. Functional connectivity and tuning curves in populations of simultaneously recorded neurons. *PLoS Computational Biology*, 8(11), Nov. 2012.
- [12] C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, and K. D. Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437), 2019.
- [13] N. Städler and P. Bühlmann. Missing values: sparse inverse covariance estimation and an extension to sparse regression. *Statistics and Computing*, 22(1):219–235, Dec 2010.
- [14] L. Q. Uddin. Complex relationships between structural and functional brain connectivity. *Trends in Cognitive Sciences*, 17:600 – 602, 2013.
- [15] M. Yuan and Y. Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.