Latent Variable Models to Uncover Neural Population Dynamics

Tze Hui Koh (tzehuik)[1] Hillary Wehry (hawehry)[1] Cathy Su (qsu1)[1] Darby Losey (dlosey)[1]

Abstract

We present variations of the group latent auto regressive analysis (gLARA) method to model neural activity collected from interacting brain areas and propose an estimate of Granger causality between the two neural populations. Rather than identifying pairwise interactions between neurons within and across brain areas, we aimed to study interactions at the population level by using gLARA, a method that could extract a smaller number of temporally dynamic and interacting latent variables. We first implemented the factor analysis version of gLARA, termed gLAFA, which limits correlations between neurons that cannot be described by the dynamical latent time series, a common assumption in neural data analysis. Perhaps surprisingly, gLAFA presented an improvement over gLARA when applied to real data which achieves a comparable log-likelihood upon real data with faster convergence. We then extended gLAFA to capture Granger causal relationships between brain regions by constraining the direction of influence between neural populations. Since the ground truth relationship between real neural populations is unknown, we simulate situations where we know the underlying relationships and find that we are able to recover these relationships with our constrained gLAFA models using a likelihood ratio test of models with and without the influence of the other brain region. Finally, with the observation that communication between neural populations may change over time, we derived a switching version of gLAFA, which switches between model parameters depending upon the direction of information flow at each timepoint. We show that we can correctly predict the underlying state based upon simulated data.

Introduction

A critical challenge in the field of neurophysiology is to move beyond describing the activity patterns of individual neurons in the brain towards examining how populations of neurons collectively realize human behavior. In recent years, advances in recording technology have enabled the collection of neural data from multiple brain areas simultaneously and have lead to a shift towards the study of directed information flow at the population level.

Traditionally neuroscientists have tried to model pairwise interactions. With this approach, however, the number of interactions grows exponentially with the number of neurons. Moreover, recent work has highlighted that interactions between neural populations is far lower dimensional than the number of neurons within each population (Semedo et al., 2019).

To leverage this, we modified the state-of-the-art method for estimating the influence of latent variables within and among neural-populations using group latent autoregressive dimensionality reduction. This is done by representing the activity of each neural population in terms of latent variables. We provide an comparison to demonstrate that our model shows comparable performance on real neural data. Our model also provides increased interpretablility, an aspect especially important in neuroscience, because correlations between neurons can only be captured by the dynamic latent time series. We validated our approach on both simulated and real neural data. We discuss the relationship between our methods and Granger Causality and compare our results to a state-of-the-art attention-based convolutional neural network. Finally, we combine these extensions in to a switching model which allows us to capture changing communications between neural populations over time. This is made possible through a switching linear dynamical system with a new discrete state variable.

Background and Related Work

Latent Variable Models of Neural Activity

A standard method for analyzing neural population data is to consider the high-dimensional population space, such that each axis represents the firing rate of an individual neuron. When viewed in this light, neural activity traces out a trajectory that spans a low-dimensional subspace (Churchland et al., 2012). The brain is not able to readily generate neural activity patterns outside this subspace, even when learning new tasks (Sadtler et al., 2014). This finding emphasizes the significance of viewing neural data in this manner. Recently, neural population activity has been found to be even more constrained (Golub et al., 2018; Hennig et al., 2018), perhaps by both their internal dynam-

[1] Carnegie Mellon University, Pittsburgh, PA 15213, USA.

10708 Class Project, Spring, 2019.

ics and the influence of other neural areas (Russo et al., 2018; Semedo et al., 2014).

Factor Analysis

Factor Analysis (FA) has been shown to be particularly effective for describing the variability of neural activity in terms of a lower dimensional set of latent variables (Santhanam et al., 2009). The classical FA model assumes latent variable $\mathbf{z}$ gives rise to data $\mathbf{x}$ according to the following generative process:

$$\begin{aligned} \mathbf{z} &\sim \mathcal{N}(\mathbf{0}, I) \\ \mathbf{x}|\mathbf{z} &\sim \mathcal{N}(\mu + \Lambda\mathbf{z}, \psi) \end{aligned} \tag{1}$$

$\mu \in \mathbb{R}^q$, $\mathbf{z} \in \mathbb{R}^p$, $\psi \in \mathbb{R}^{q \times q}$ representing the diagonal noise covariance, and $\Lambda \in \mathbb{R}^{q \times p}$ represents the loading matrix. q and p are the dimensionality of the neural data and latent variables respectively, where $p << q$.

Granger Causality (GC)

Pairwise GC

Granger causality is an approach to detect temporally predictive relationships in time series data and is highly related to mutual information under simplifying assumptions. It is a widely used tool in neuroscience, despite the issues with parameter estimation and interpretation as the number of pairwise interactions grows exponentially with the number of neurons under study (Seth et al., 2015). A neural signal X_1 is said to "Granger-cause" a signal X_2 if past values of X_1 help to predict future values of X_2. In pairwise GC, one fits autoregressive processes to two models of order p:

$$X_1(t) = \sum_{j=1}^{p} A_{11}(j)X_1(t-j) + \sum_{j=1}^{p} A_{12}(j)X_2(t-j) + \epsilon_1(t)$$

$$X_2(t) = \sum_{j=1}^{p} A_{21}(j)X_1(t-j) + \sum_{j=1}^{p} A_{22}(j)X_2(t-j) + \epsilon_2(t)$$

If the inclusion of X_1 reduces the prediction error ϵ_2, then we can conclude that X_1 is a Granger cause of X_2 (Cadotte et al., 2008). This is statistically equivalent to assessing the significance of the linear fit including X_2 , $||A_{12}|| > 0$.

While Granger causality can be directly applied to pairwise neural signals, consider the case where neuron X synapses onto neuron Y and neuron Y synapses onto neuron Z. Classic GC analysis would erroneously infer that neuron X is causal to neuron Z, despite this influence occurring only through mediator Y. Conditional GC was proposed to eliminate the problem of erroneous GC estimates by including the history of other time series in the autoregressive model, such that that the GC influence of interest must provide predictive power above and beyond the conditioned series (Cadotte et al., 2008). However, one can imagine the difficulty in assessing the pairwise conditional influences when both are embedded in a super high-dimensional network with hidden variables. In real neural data, one might need to condition on hundreds or even hundreds of thousands of noisy neural signals to estimate the influence of a single neuron on another, and even then, the interpretation is still dependent on additional assumptions due to the confounding variable problem. Under appropriate hypothesis testing procedures each single pairwise interaction is likely insignificant, yet considering the brain areas as a whole may have revealed network-wide interactions.

GC for General Probabilistic Models

The formulation of linear GC defined above requires two key assumptions: linearity and Gaussian random variables. (Kim & Brown, 2010) proposed a general statistical framework for assessing GC interactions, where by the GC estimate from a time series X_1 to a time series X_2 is estimated by the relative reduction of the likelihood of X_2 obtained by the exclusion of X_1 compared to the likelihood obtained using the joint time series. Let us define x_i and x_j, where the past values of x_i including the contribution of x_j is $\underline{\mathbf{x}}$, and the past values obtained after excluding x_j is $\underline{\mathbf{x}}^{-\mathbf{j}}$. For simplicity x_i and x_j can be considered two univariate time series, but such an estimate can be interpreted as sets of time series as well. Thus, the following is true only if and only if x_i and x_j are independent:

$$p(x_i|\underline{\mathbf{x}}(t)) = p(x_i|\underline{\mathbf{x}}^{-j}(t)) \tag{2}$$

We can therefore assess Granger Causality from x_j to x_i using the log-likelihood ratio (Kim & Brown, 2010), with likelihood $L_i(\theta_i^{-j})$ calculated using $\underline{\mathbf{x}}^{-j}(t)$).

$$\Gamma_{ij} = \log \frac{L_i(\theta_i^{-j})}{L_i(\theta_i)} \tag{3}$$

If x_j is Granger Causal to x_i then $L_i(\theta_i) < L_i(\theta_i^{-j})$, (i.e. $\Gamma_{ij} < 0$) (Kim & Brown, 2010). The test statistic is thus $S = -2\Gamma_{ij}$,$S \sim \chi^2_M$ where the degree of freedom M is the difference in dimensionality of two models. The instantaneous causality of x_i and x_j:

$$\Gamma_{i \cdot j} = \log \frac{L_i(\theta_i)L_j(\theta_j)}{L_{ij}(\theta_{ij})} \tag{4}$$

where $L_{ij}(\theta_{ij})$ is the joint likelihood function of x_i and x_j. If $\Gamma_{i \cdot j} \neq 0$, then x_i and x_j instantaneously cause one another.

This framework extends the pairwise definition of the study of interactions to general probabilistic models. With this new metric, one can consider estimating the information flow between two neural populations as one-directional temporal prediction, for example by comparing with the data likelihood of a latent variable model of population 1 with information about population 2 to the likelihood of population 1 without information about population 2.

Auto-regressive latent variable models

Recent work has shown single neuron estimates of interaction may be insignificant, but at a population level one may discern a notable relationship (Semedo et al., 2019). Thus we were interested in combining the idea of network-level Granger causality with a dimensionality reduction technique appropriate for interacting neural populations, and in doing so more effectively estimate the directional influence of one neural population on another.

Two widely known autoregressive models, AR-principal component analysis (AR-PCA) and AR- probabilistic canonical correlation analysis (AR-pCCA), both feature an autoregressive process over latent variables. However, neither method distinguishes between groups of neural populations in the low-dimensional latent space, applying a common latent variable to all recorded neurons. If we were to directly apply these more standard dimensionality reduction technique to our data, which do not incorporate class or population labels, the estimated latent variables would capture modes of covariability across the neurons without distinguishing between-population interaction and across-population interaction. Applying factor analysis and its dynamic variants to each population individually is also not appropriate, since the across-population interaction would not necessarily be preserved.

Group latent auto-regressive analysis (gLARA)

Since previously published autoregressive methods such as AR-pCCA and AR-pPCA do not distinguish between groups of neural populations in the low-dimensional space, applying a common latent variable to all recorded neurons, (Semedo et al., 2014) developed an autoregressive model of neural activity (gLARA) such that each neural population is represented by a separate set of latent variables, but these latent variables are permitted to interact linearly under an autoregressive model.

In this model, the latent variables for each neural population interact over time. This represents the state-of-the-art autoregressive latent variable model of neural activity across multiple brain regions. Here we consider 2 neural populations, conceptualized as being recorded from different brain areas $\mathbf{x_i}$, jointly driven by p_i-dimensional latent states $\mathbf{z}_t$, where $\mathbf{x_i}$ has q_i neurons (i.e. $\mathbf{x_i} \in \mathbb{R}^{q_i}$). The equations for gLARA follow for the $M = 2$ case using the notation in the Box :

$$\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, I) \text{ if } 1 \leq t \leq \tau \tag{5}$$

$$\mathbf{z}_t | \mathbf{z}_{t-1}, \dots \mathbf{z}_{t-\tau} \sim \mathcal{N}\left(\sum_{m=1}^{2} \sum_{s=1}^{\tau} A_{i,j}^{m} \mathbf{z}_{t-s}, Q^m \right) \tag{6}$$

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_t^1 \\ \boldsymbol{z}_t^2 \end{bmatrix} + \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}, \begin{bmatrix} R^1 & \mathbf{0} \\ \mathbf{0} & R^2 \end{bmatrix} \right) \tag{7}$$

The covariance of the observed signal x_i is thus:

$$cov\left(\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right) = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix}^T + \begin{bmatrix} R^1 & \mathbf{0} \\ \mathbf{0} & R^2 \end{bmatrix}$$

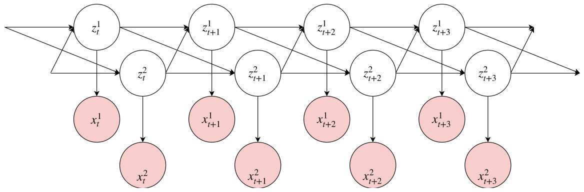


Figure 1. **Setup schematic for gLAFA and gLARA on two brain areas.** All arrows represent causal connections. All of the causal connections (arrows) indicate active and directed communication. Over time the two latent states communicate with each other and each latent communicates to the neurons in the same brain area x_t^i (see Table 1 for parameters).

The observation model incorporates a block diagonal structure that allows shared variance among neurons within a single population, but not across populations. In our work we implement and extend gLARA to assess Granger causal-like interactions.

Methods

Box 1: Notation

T: timepoints per trial
q_i: number of neurons in region i
p_i: latent state dimension in region i
$x_t^i \in \mathcal{R}^{q_i \times T}$: observations for region i at time t
$z_t^i \in \mathcal{R}^{p_i \times T}$: latent for region i at time t.
τ: order of autoregressive model
C_i: coefficient of latent in region i
S_t: hidden discrete state variable for time t.

gLAFA

We present a group latent autoregressive factor analysis (gLAFA), a more constrained version of the gLARA model. The aim of this algorithm is to extract directional influences from one population's latent variables on another. We derived and implemented the expectation maximization algorithm for the gLAFA model and implemented gLARA model from scratch (without using any existing code) in order to compare the methods. With the appropriate augmented models as mentioned in (Semedo et al., 2014), one can use a standard E-M to fit the model to data. We propose additional extensions in the sections below.

The intuition behind developing this model was that latent variable models of neural activity have the underlying assumption that covariance shared between neurons is signal we would like to preserve, whereas any variation that is independent for each neuron is treated as noise. Thus the full within-population R noise covariance presented in gLARA loses the central benefit of FA, which we have previously shown to be widely effective for capturing the low-dimensional structure of neural population activity (Sadtler et al., 2014).

We derived the M-step of gLAFA to constrain the observation covariance matrix:

$$\begin{aligned} R^i = \frac{1}{T}\sum_{t=1}^{T}\{&((\boldsymbol{x}_t^i - \boldsymbol{d}^i)(\boldsymbol{x}_t^{i'} - \boldsymbol{d}^i)) \circ I \\ &-(C^i E(\boldsymbol{z}_t^i)(\boldsymbol{x}_t^i - \boldsymbol{d}^i)') \circ I \\ &-((\boldsymbol{x}_t^i - \boldsymbol{d}^i)E(\boldsymbol{z}_t^{i'})C^{i'}) \circ I \\ &+(C^i E(\boldsymbol{z}_t^i \boldsymbol{z}_t^{i'})C^{i'}) \circ I\} \end{aligned} \tag{8}$$

Probabilistic Granger Causality for gLAFA

As described above, gLARA and gLAFA both allow instantaneous interactions between the two sets of latent variables. The model fit returned from the E-M algorithm alone cannot determine if the influence of the other population offers profound improvement over the first population's history– especially if activity in the two networks is correlated. Furthermore, for high-dimensional neural data sets, we desire a valid statistical hypothesis testing framework in order to make scientific statements about the interactions between neural networks. Under the probabilistic GC metrics defined above 3, we then considered how to estimate the data likelihood of each population under the gLAFA model with and without information from the other population's latent variables.

Thus we considered additional constraints over the autoregressive variables (matrix A, above) to satisfy gLAFA models with either one-directional flow, bidirectional flow, or independence. For example, under the one-directional Area 1 → Area 2 gLAFA model, the A matrix must be projected to minimize $||A_{12}||_1$. The data likelihoods for the observations from each population were also separated.

Referring to 3, the Granger causal estimate from population X_2 to X_1 (Γ_{12}), one first computes two data likelihoods: (a) the log likelihood of X_1 under the model in which directional influences from $2->1$ is unconstrained but X_1 cannot influence X_2, and (b) the log likelihood of X_1 under the model in which the two populations are not interacting. The GC12 estimate is simply the second quantity minus the first. Following (Kim et al., 2011), this estimate is then multiplied by -2 to find a χ^2r distributed statistic that can be used for hypothesis testing.

Temporal Causal Discovery Framework (TCDF)

The assumed generative model for our data has latent states that dynamically change across time and influence the latent states of other neural populations. We wanted to compare causal relationships detected by gLAFA to some other benchmark. We therefore used TCDF, a neural network which has shown state-of-the-art performance on the task of learning causal structure from time series data despite confounders (Nauta et al., 2019).

TCDF can detect causal relationships between input time series, including hidden confounders. It does this by learning the dependence of each input time series upon the others with a distinct convolutional neural network with an attention mechanism. From this, many potential causal graphs are produced. To distinguish between the graphs, causal validation is performed by checking that causes have temporal precedence as well as that manipulation of putative causal variables will indeed cause the predicted effects. TCDF also outputs the number of timesteps between cause and effect.

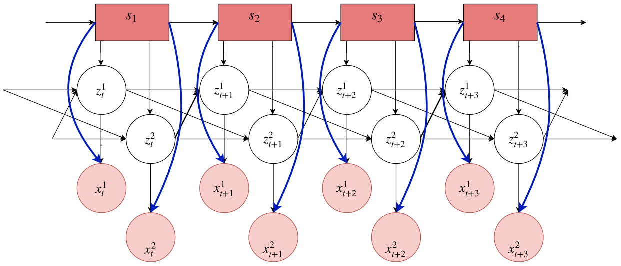


Figure 2. **Setup schematic for switching gLARA.** Adding to the system described by Figure 1, we additionally have hidden state variables S_t which determine which set of parameters to use at a specific time. each latent to the corresponding observations x_t^i (see Table 1 for parameters).

Switching gLAFA

In the gLARA/gLAFA models, the parameters $\theta = \{A, R, Q, d, C\}$ are static across time. However in neural populations, the strength and direction of communication between brain areas may change over time. For example, brain region 1 may be sending input to brain region 2 at time point 1, but subsequently brain region 2 may send input to brain region 1 at time point 2. From our work on constraining information flow in the previous section on Granger Causality, one might imagine a mixture of 4 models which model the different combinations of relationships between neural populations (e.g. both populations influence each other, population 1 influences population 2 etc.). This is the intuition behind why we were interested in extending gLAFA to model changing communications.

One approach for this as outlined in (Murphy, 1998) is to use a mixture of linear dynamical systems. To do so we introduce a new hidden discrete variable S_t which determines the set of parameters to be used at each time point. S_t is governed by Markovian dynamics

$$S_t = f(S_{t-1}) \tag{9}$$

The graphical model for the switching gLAFA is illustrated in Figure 2. As shown the switching variable functions to select the subprocess that is passed to the output variable. Equations 2, 3, and 4 remain as the generative model of switching gLAFA, except that the choice of model parameters are now governed by the discrete state variable.

Switching gLAFA derivations

We follow the model presented by (Murphy, 1998) with markovian dynamics for the state matrix. Following from the setup introduced by equations (6) to (7) we introduce

the state variable S_t. Since we want to know the maximum likelihood estimate and the S_t are not observed, the EM algorithm is used to determine the parameters $\theta_t := \{A_t, R_t, Q_t, d_t, C_t\}$.
This represents a modified system than from the gLAFA case. By augmenting our generative model in the same manner as that found in (Semedo et al., 2014), we can then use the switching Kalman smoother equations found in (Murphy, 1998) for the expectation step. Thus the joint log probability would be for augmented parameters $\theta = \{\bar{A}, \bar{R}, \bar{Q}, \bar{d}, \bar{C}\}$:

$$\begin{aligned}\log P(\bar{\boldsymbol{z}}_{1:T}, \bar{\boldsymbol{x}}_{1:T}, s_{1:T}) = -\frac{1}{2}\sum_{t=1}^{T}[\bar{\boldsymbol{x}}_t - \bar{C}\bar{\boldsymbol{z}}_t]^T \bar{R}^{-1}[\bar{\boldsymbol{x}}_t - \bar{C}\boldsymbol{z}_t]) \\ - \frac{T}{2}\log|\bar{R}| \\ - \frac{1}{2}\sum_{t=2}^{T}[\bar{\boldsymbol{z}}_t - \bar{A}\bar{\boldsymbol{z}}_{t-1}]^T \bar{Q}^{-1}[\bar{\boldsymbol{z}}_t - \bar{A}\bar{\boldsymbol{z}}_{t-1}] \\ - \frac{1}{2}\sum_{t=1}^{T}\bar{\boldsymbol{x}}_t^T I \bar{\boldsymbol{x}}_t \\ - \frac{T(p+q)}{2}\log(2\pi) + \log(\pi_1) \\ + \sum_{t=2}^{T}\log Z(s_{t-1}, s_t)\end{aligned}$$

For the M step, we take the derivative of the joint log-likelihood with respect to each individual parameter and set it to 0 to obtain the maximum likelihood estimate of the model parameters, where the quantity we maximize is

$$\tilde{L} = E_{\log P(\bar{\boldsymbol{z}}_{1:T}, \bar{\boldsymbol{x}}_{1:T}, \bar{s}_{1:T})}(L)$$

Where we have computed $E[\mathbf{x}_t^j], E[\mathbf{z}_t^j \mathbf{z}_t^{j}{}'], E[\bar{\mathbf{z}}_t \bar{\mathbf{z}}'_{t-1}]$ and $W_t^j = Pr(S_t = j|\mathbf{x}_{1:T})$ in the expectation step.

M STEP

$$\begin{bmatrix} A_1^{11} \dots A_K^{11} & A_1^{12} \dots A_K^{12} \\ A_1^{21} \dots A_K^{21} & A_1^{22} \dots A_K^{22} \end{bmatrix} = A_i$$

$$A_i = (\sum_{t=2}^{T} W_t^i E[\bar{\mathbf{z}}_t \bar{\mathbf{z}}'_t])(\sum_{t=2}^{T} W_t^i E[\bar{\mathbf{z}}_t \bar{\mathbf{z}}_t^T])^{-1}$$

$$\begin{aligned}\begin{bmatrix} C_1^j & \mathbf{d}_i^j \end{bmatrix} =& \left(\sum_{t=1}^{T} W_t^i \mathbf{x}_t^j \left[\begin{bmatrix} E[\mathbf{z}_t^{j}{}^T & 1 \end{bmatrix} \right] \right) \\ &\cdot \left(\sum_{t=1}^{T} W_t^i \begin{bmatrix} E[\mathbf{z_t^j z_t^j}{}'] & E[\mathbf{z_t^j}] \\ E[\mathbf{z_t^j}] & 1 \end{bmatrix} \right)^{-1}\end{aligned}$$

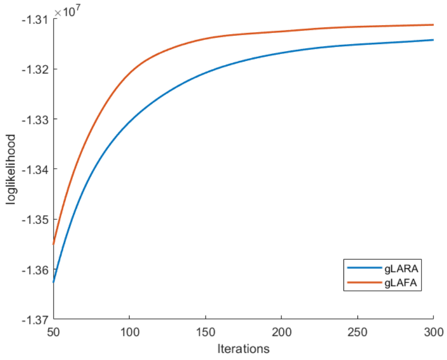


Figure 3. **gLAFA performs slightly better then state-of-the-art method gLARA on real neural data.** gLAFA takes fewer iterations to converge, and (not shown) converges to the same parameters as gLAFA.

$$\begin{aligned}R_i^j =& \frac{1}{\sum_{t=1}^{T} W_t^i}\sum_{t=1}^{T}\Big(W_t^i ((\mathbf{x_t^j} - \mathbf{d_i^j})(\mathbf{x_t^j} - \mathbf{d_i^j})' \\ &- \mathbf{C_i^j} E[\mathbf{z_t^j}](\mathbf{x_t^j} - \mathbf{d_i^j})' - (\mathbf{x_t^j} - \mathbf{d_i^j}) E[\mathbf{z_t^j}{}'] \mathbf{C_i^j}{}' \\ &+ \mathbf{C_i^j} E[\mathbf{z_t^j z_t^j}] \mathbf{C_i^j}{}') \Big)\end{aligned}$$

$$Q_i = \frac{1}{\sum_{t=2}^{T} W_t^i}\sum_{t=1}^{T} W_t^i E[\bar{\boldsymbol{z}}_t \bar{\boldsymbol{z}}_t{}'] - A_i \sum_{i=2}^{T} W_t^i E[\bar{\boldsymbol{z}}_t \bar{\boldsymbol{z}}'_{t-1}]$$

$$Z_{i,j} = \frac{\sum_{t=2}^{T} P(S_{t-1} = i, S_t = j|\boldsymbol{x}_{1:T})}{\sum_{t=1}^{T} W_t^i}$$

$$\boldsymbol{\pi}_i = W_1^i$$

Experiments

gLAFA method verification

Using the generative framework in Equation 7 to simulate data with known “ground truth” latent variables and parameters, we performed a series of sanity checks to confirm that our implementation of the EM algorithm for parameter estimation was returning expected results. This is possible because gLARA/gLAFA are generative models.

DESCRIPTION OF TESTBED

We pick our parameter values from uniform random distributions, while also ensuring that the eigenvalues of the A matrices are less than 1 for stability. Here we set $p_1 = 1, p_2 = 1, q_1 = 2, q_2 = 2$.

DESCRIPTION OF EXPERIMENT

Figure 10 illustrates the log-likelihood is non-decreasing, which confirms that the E-M algorithm is converging to a local optimum. The estimated estimated latent variable and model parameters also grow closer to the ground truth across iterations; this is shown in Figure 10.

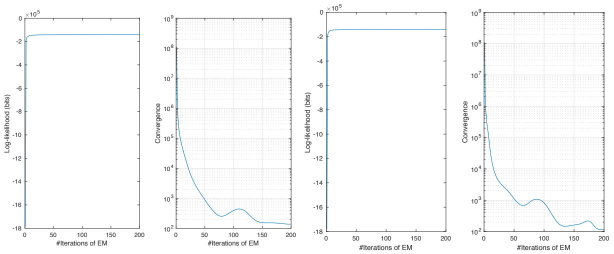


Figure 4. **Convergence over EM iterations for gLAFA and gLARA.** Here the convergence and log likelihood are plotted which correspond to results in figure 5 for gLARA (top) and gLAFA (bottom). Convergence is measured as the difference of log likelihood from zero.

Comparison of gLAFA and gLARA on real data

To answer whether our method gLAFA could be used to model real data with as much efficacy as gLARA, which is the state of the art existing approach, we compared the performance of the two methods on the real data provided by (Semedo et al., 2014).

Description of testbed

The data for comparison contained recordings from the V1 population of an anaesthetised monkey. In this data there are 200 bins of 5ms each (1s of data total) for each of the V1 and V2 brain regions. We have $q = 111$ and $q = 37$ neurons for the two regions, respectively. In addition to concur with (Semedo et al., 2014) we chose $p = 10$ for both V1 and V2.

Description of Experiment

To compare gLARA and gLAFA, we divided the data such that three quarters was used to learn the model parameters, then performed leave one out (LOO) neuron prediction on the remaining quarter of neurons following (Semedo et al., 2014). The LOO procedure entails estimating the latent states $E(z|y)$ in the absence of one neuron, followed by prediction of its activity using the learned parameters and estimated latent state.

As shown in figure 5, the performance of gLAFA is comparable to gLARA, and produces a smoother estimate of the mean firing rate, reaching a similar log-likelihood to gLARA and converging faster. The faster convergence makes sense intuitively, since the constraint on the noise covariance reduces the number of parameters to be fit for gLAFA. Despite having less parameters, it was surprising that the log-likelihoods of both models converged to similar values. This could suggest that the additional parameters in gLARA do not contribute much to the data fit. Additionally with this new constraint, we have preserved the shared variance between neurons, treated as firing rate variability, and discarded independent variance, which is thought of spiking variability (Cunningham & Yu, 2014). This lends to the interpretability of the model.

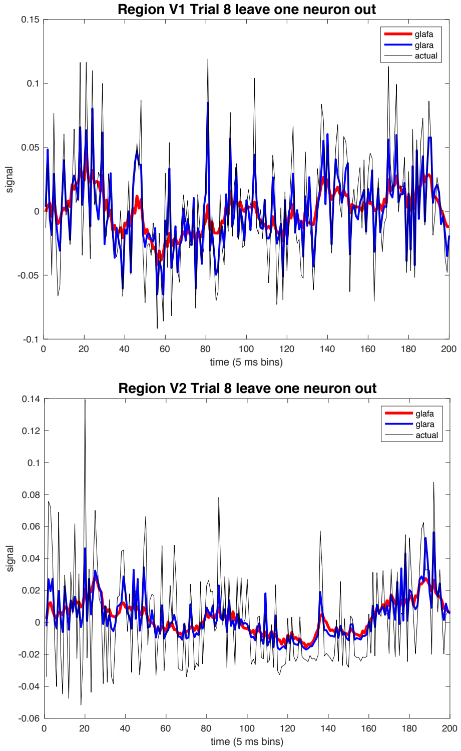


Figure 5. **Leave one out neuron prediction using gLAFA and gLARA.** Observed activity (black) and the leave one out neuron prediction of gLARA (blue) and gLAFA (red) for a representative held out trial averaged over the V1 (top) and V2 (bottom) population.

Performance of Granger Causal test on data generated from gLAFA with simulated causal relationships

To further our goal of revealing causal relationships in real neural data, we first simulated different Granger causal-type scenarios to test whether they could be captured by gLAFA.

Description of testbed

We generated a toy dataset under the model defined by equation 7 but with different constraints on A. We define the predictive relationships between the simulated neural populations *a priori*, but the algorithm does not have access to the ground truth. Here we simulate 4 neurons ($q_1 = 4, q_2 = 4$) in each population, with 2 underlying latent variables each ($p_1 = 2, p_2 = 2$). We pick Q, R, d, C from standard normal distributions. For the A matrix, we constructed three different scenarios. First, where the latent variables of population 1 and population 2 do not interact with each other, only themselves, which corresponds to an A matrix with the off diagonal quadrants set to 0 ('Area 1 $\not\leftrightarrow$ Area 2 scenario). In the second case, both populations cause one another, which corresponds to a full A matrix (Area 1 $\leftrightarrow$ Area 2 scenario). The third case, where population 1 causes population 2, corresponds to the lower left quadrant of the A matrix set to 0.

Description of experiment

Constructing such A matrices helps us to validate the potential of applying Granger causality to a constrained version of gLAFA in scenarios where we know the ground truth. This is illustrated in the graphical models shown in figure 6.

After generating the data, we ran modified versions of gLAFA under the above constraints to detect the different types of relationships between populations and then computed the Granger causality statistic described in the methods section in order to estimate the Granger causal predictability of the two time series simulated in the system. To estimate the directional influences within a single dataset, the 4 modifications of the gLAFA and gLARA EM algorithms described abovewere implemented and applied to the toy dataset.

The 5-fold cross-validated Granger causality estimates for each direction of information flow are shown in the table 6. Γ_{12} is the estimate for the directed influence of Area 2 on Area 1 and should be a large negative number if this influence is significant. Any positive GC estimate is by definition non-significant. This estimate is then evaluated under a $^{2}_{p^2}$ distribution, and significant results are indicated in red. We then binarized the results in the form of 2×2 tables (see Figure 6), in which indicate that our measure was capable of recovering this predictive relationship between the populations. The framework was thus able to reliably return the ground truth simulated influences under the setting of a moderate signal to noise ratio and fully observed network, which is expected since the data was generated under a fairly strong auto-regressive assumption over the latent variables. One concern now is that the framework described here does not allow for comparison of the magnitude of directional influence, only whether or not such an influence is significant. In such cases where the ratio of directional influence may be changing in a context dependent manner, we cannot currently compare such differences.

Performance of TCDF on data generated from gLAFA with simulated causal relationships

We aimed to benchmark the performance of the granger causality statistic described in the 'Granger causality on gLAFA' section against the TCDF neural network.

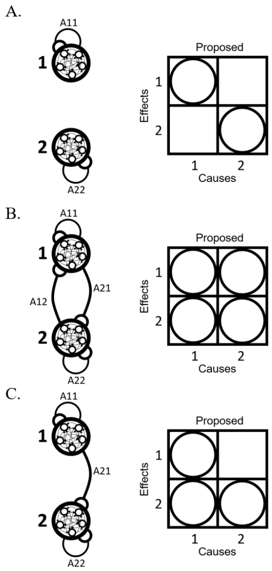
Description of testbed

For a fair comparison of the two methods, we evaluated the performance of TCDF using the same toy dataset that was used to test the 'granger causality on gLAFA'.

Description of experiment

We summarized the results in Table 1. We observed that TCDF always recovered relationships between the correct populations. For 4 of 5 datasets we profiled, when using 1 hidden layer, TCDF was able to categorize each of one component of the latent variables r1 and r2 as a cause. However there are mistakes as TCDF sometimes identifies neurons within the same region instead of the latent variable as causal. We speculate that since the strength of the signal is evolving over time, the root cause can be easy to misidentify. Another shortcoming is that the delay discovered for the causal relationships is not constant even though the data was consistently simulated with $\tau = 1$.

Thus, compared to the results of the Granger Causality statistic, TCDF does not perform as well at capturing causal relationships.



Simulation Ground Truth

	Area 1 → Area 2		**Area 1 ↔ Area 2**		**Area 1 $\not\leftrightarrow$ Area2**	
	gLARA	gLAFA	gLARA	gLAFA	gLARA	gLAFA
Γ_{12}	0.8924	1.1381	-7.59e4, $p \ll 0.01$	-1.09e4, $p \ll 0.01$	-0.362	-0.3264
Γ_{21}	-1.01e4, $p \ll 0.01$	-9.01e4, $p \ll 0.01$	-2.53e4, $p \ll 0.01$	-2.02e3, $p \ll 0.01$	52.0807	1.0785
$\Gamma_{1 \cdot 2}$	435.74	2.79e4	-3.36e3, $p \ll 0.01$	-3.51e3, $p \ll 0.01$	1.15e3	508.4504

Figure 6. **Visualization of causal relationships within the simulated neural populations 1 and 2 (left), and results from the probabilistic Granger Causality statistic run on gLAFA results (right).** On the left outer circles represent distinct populations while the inner circles represent neurons, which are connected to show temporally causal communication. The figure illustrates cases A. where there is no causal relationship between populations 1 and 2. B. both populations cause each other C. one population causes the other. The A labels on the edges are representative of the A matrices described in the gLAFA model.

Dataset	**Regions intra-connected?**	**Regions interconnected?**	**Discovered delay**
Area 1 $\rightarrow$ Area 2	Y	Y: z1_2 causes x2_2	0 for all, except: z1_2 to x2_2, z2_1 to x2_2, x1_3 to x1_1
Area 2 $\leftarrow$ Area 1	Y	Y: z2_1 causes x1_1	0 for all, except: z2_1 to x1_1
Area 1 $\leftrightarrow$ Area 2	Y	N	0 for all except: z2_2 to x2_4 and x2_1, x1_3 to x2_4
Area 1 $\nleftrightarrow$ Area 2	Y	N	0 for all

Table 1. **Learned causal connections from TCDF on simulated data.** For each dataset that was generated from the directional gLARA, we ran TCDF and discovered intra region causal connections for each cell population. TCDF also successfully recovered one inter region causal connection of the correct direction for the Area 2 $\rightarrow$ 1 and Area 1 $\rightarrow$ Area 2 datasets. Importantly, and perhaps surprisingly, it does not seem to distinguish the Area 1 $\leftrightarrow$ Area 2 versus the Area 1 $\nleftrightarrow$ Area 2 case.

Results from switching gLAFA

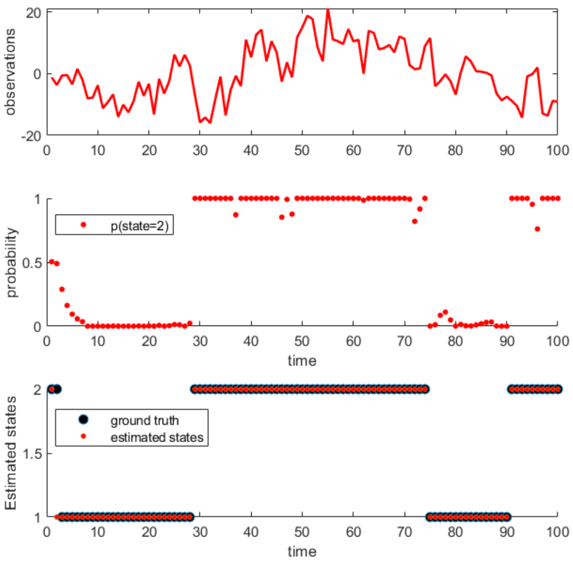


Figure 7. **Probabilities of the states extracted from switching gLARA.** Top: Observed data, middle: estimated probability of system being in state 2 from the switching Kalman smoother, bottom: predicted state of the system as compared to the ground truth states

Having derived the necessary equations for the switching gLAFA, we show a preliminary implementation of the algorithm given the parameters (i.e. without the need to do maximum likelihood over the log-likelihood) in figure 7.

Description of testbed

We simulated data from the generative model of the switching linear dynamical system, with $q = 2$ and $p = 1$. We again drew the parameters from a random distribution while ensuring that the system is stable (i.e. the A matrix has eigenvalues < 1). We simulated 1 trial and 100 time points.

Description of experiment

Here we are able to estimate the state of the system, and predict the corresponding latent variables. To obtain an estimate of the ground truth latent variable, we then collapse the latent variables per state as a weighted sum of the individual latent variables, as can be seen in figure 8.

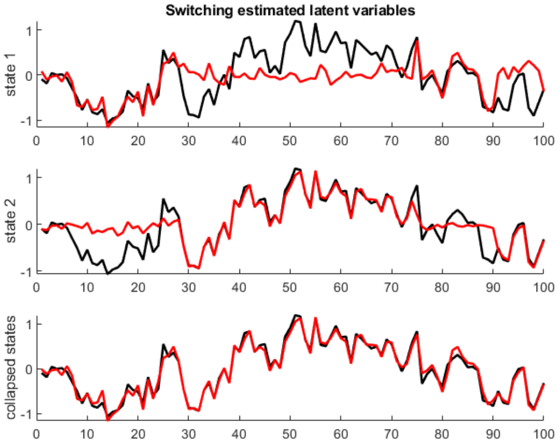


Figure 8. **Estimates of the hidden latent variables for switching gLAFA.** Estimates for each of the hidden latent variables are shown (top and middle), with which a weighted sum (bottom) would provide an estimate of the generated latent variable.

Conclusion

Discussion

Here we presented gLAFA, an autoregressive model of neural signal across multiple populations. We also provide extensions on gLAFA to find causal connections and to allow state switching.

We compared the performance of gLAFA to the the state of the art model, gLARA on real neural recordings from the brain regions V1 and V2. We found gLAFA converges faster and by visual inspection offers comparable reconstruction of the signal. This makes sense as gLAFA treats region independent variance as noise, offering more interpretability. Furthermore since this results in fewer parameters, gLAFA is more robust to overfitting.

Additionally, we developed a likelihood ratio test to determine Granger causality between different brain regions. We found that this Granger causal statistic performs better at identifying causal relationships than a state of the art neural network developed for the same task (TCDF).

Future Work

We have developed a method to model neuron signals across brain regions, gLAFA, and shown that it has good performance on both simulated and on real neural recording data.

While these initial results were promising, we also propose some future directions that will be useful to determine the full utility of the model and its extensions:

- We have shown gLAFA is able to reconstruct the averaged signal across neurons for two brain regions. It is natural to extend the model to more regions in order to better simulate the complexity in the brain.
- Although we have developed the granger causal metric and shown that it works well on gLAFA- generated data, we also need to look whether it will also be applicable on a real dataset.
- Full implementation of the switching gLARA model: this would enable us to fully estimate model states from the current gLAFA simulation data, but also upon data simulated upon the switching model and upon real data. The switching model also has the potential to help us evaluate the relative strength of causal connections between different brain regions.

Appendix

References

Cadotte, A. J., DeMarse, T. B., He, P., and Ding, M. Causal measures of structure and plasticity in simulated and living neural networks. *PloS one*, 3(10):e3355, 2008.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. Neural population dynamics during reaching. *Nature*, 487(7405):51, 2012.

Cunningham, J. P. and Yu. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17 (11):1500, 2014.

Golub, M. D., Sadtler, P. T., Oby, E. R., Quick, K. M., Ryu, S. I., Tyler-Kabara, E. C., Batista, A. P., Chase, S. M., and Yu, B. M. Learning by neural reassociation. *Nature neuroscience*, 21(4):607–616, 2018.

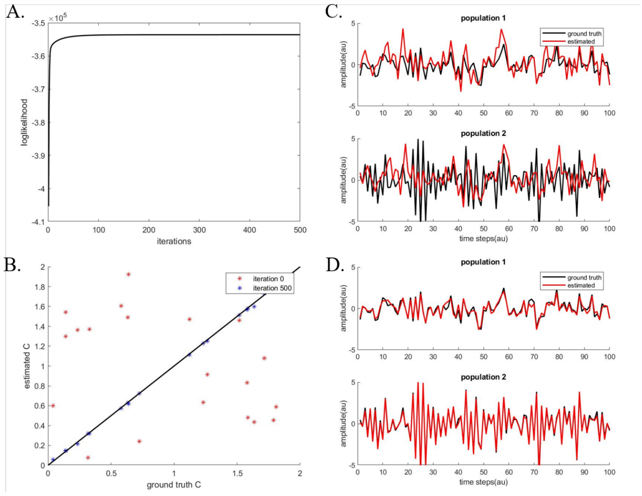


Figure 9. **gLAFA E-M algorithm passes sanity checks.** A. Non-decreasing log-likelihood; B. Convergence of parameter C to ground truth from iteration 0 to 500. gLAFA estimate of simulated latents improves over multiple iterations. Sample estimated latent variables over time steps at C. the first iteration and D. 500th iteration (bottom) of EM. The fit visibly improves.

S0.52S

Figure 10. **Visualization of causal relationships learned by TCDF for the Area 1 $\not\leftrightarrow$ Area 2 dataset**. All dimensions of the latent variables were included as input time series. The nodes with name pattern n1 and n2 belong to brain regions 1 and 2 respectively. The latent variable components (r1, r2) are also shown. Edges indicate causal relationship. In this case, causal relationship between the latent variables and neurons are not all learned; however only variables from the same brain region are found to interact. Additionally TCDF successfully finds that there are no causal connections found between the two populations.

Hennig, J. A., Golub, M. D., Lund, P. J., Sadtler, P. T., Oby, E. R., Quick, K. M., Ryu, S. I., Tyler-Kabara, E. C., Batista, A. P., Byron, M. Y., et al. Constraints on neural redundancy. *eLife*, 7:e36774, 2018.

Kim, S. and Brown, E. N. A general statistical framework for assessing granger causality. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2222–2225. IEEE, 2010.

Kim, S., Putrino, D., Ghosh, S., and Brown, E. N. A granger causality measure for point process models of ensemble neural spiking activity. *PLoS computational biology*, 7(3):e1001110, 2011.

Murphy, K. P. Switching kalman filters. Technical report, 1998.

Nauta, M., Bucur, D., and Seifert, C. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):312–340, 2019. ISSN 2504-4990. doi:

10.3390/make1010019. URL http://www.mdpi.com/2504-4990/1/1/19.

Russo, A. A., Bittner, S. R., Perkins, S. M., Seely, J. S., London, B. M., Lara, A. H., Miri, A., Marshall, N. J., Kohn, A., Jessell, T. M., et al. Motor cortex embeds muscle-like commands in an untangled population response. *Neuron*, 97(4):953–966, 2018.

Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., Byron, M. Y., and Batista, A. P. Neural constraints on learning. *Nature*, 512(7515):423, 2014.

Santhanam, G., Byron, M. Y., Gilja, V., Ryu, S. I., Afshar, A., Sahani, M., and Shenoy, K. V. Factor-analysis methods for higher-performance neural prostheses. *Journal of neurophysiology*, 2009.

Semedo, J., Zandvakili, A., Kohn, A., Machens, C. K., and Byron, M. Y. Extracting latent structure from multiple interacting neural populations. In *Advances in neural information processing systems*, pp. 2942–2950, 2014.

Semedo, J. D., Zandvakili, A., Machens, C. K., Byron, M. Y., and Kohn, A. Cortical areas interact through a communication subspace. *Neuron*, 2019.

Seth, A. K., Barrett, A. B., and Barnett, L. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015.