
Latent Dynamic Factor Analysis of High-Dimensional Neural Recordings

Heejong Bong*, Zongge Liu*
Carnegie Mellon University
hbong, zonggel@andrew.cmu.edu

Zhao Ren
University of Pittsburgh
zren@pitt.edu

Matthew A. Smith, Valerie Ventura, Robert E. Kass
Carnegie Mellon University
msmith, vventura, kass@andrew.cmu.edu

Abstract

High-dimensional neural recordings across multiple brain regions can be used to establish functional connectivity with good spatial and temporal resolution. We designed and implemented a novel method, Latent Dynamic Factor Analysis of High-dimensional time series (LDFA-H), which combines (a) a new approach to estimating the covariance structure among high-dimensional time series (for the observed variables) and (b) a new extension of probabilistic CCA to dynamic time series (for the latent variables). Our interest is in the cross-correlations among the latent variables which, in neural recordings, may capture the flow of information from one brain region to another. Simulations show that LDFA-H outperforms existing methods in the sense that it captures target factors even when within-region correlation due to noise dominates cross-region correlation. We applied our method to local field potential (LFP) recordings from 192 electrodes in Prefrontal Cortex (PFC) and visual area V4 during a memory-guided saccade task. The results capture time-varying lead-lag dependencies between PFC and V4, and display the associated spatial distribution of the signals.

1 Introduction

New electrode arrays for recording electrical activity generated by large networks of neurons have created great opportunities, but also great challenges for statistical machine learning (e.g., Steinmetz et al., 2018). For example, Local Field Potentials (LFPs) are signals that represent the bulk activity in relatively small volumes of tissue (Buzsáki et al., 2012; Einevoll et al., 2013), and they have been shown to correlate substantially with the BOLD fMRI brain imaging signal (Logothetis et al., 2001; Magri et al., 2012). Typical LFP data sets may have dozens to hundreds of time series in each of two or more brain regions, recorded simultaneously across many experimental trials. A motivating example in this paper is LFP recordings from a prefrontal cortex (PFC) and visual area V4 during a visual working memory task. V4 has been reported to retain higher-order information (e.g., color and shape) and attention in visual processing (Fries et al., 2001; Orban, 2008), while PFC is considered to exert cognitive control in working memory (Miller and Cohen, 2001). Despite their spatial distance and functional difference, these regions have been presumed to cooperate during visual working memory tasks. Various approaches have been used to track the interaction among brain regions Adhikari et al. (2010); Buesing et al. (2014); Gallagher et al. (2017); Hultman et al. (2018); Jiang et al. (2015). In particular, delay-specific theta synchrony led by PFC has been discovered during visual memory tasks (Liebe et al., 2012; Sarnthein et al., 1998).

*Equal contribution.

Because many functional interactions among brain regions are transient, it is highly desirable to have methods that accommodate non-stationary behavior in the multivariate time series recorded in each region. We report here an extension of Gaussian process factor analysis (GPFA, Yu et al., 2009) to two or more groups of time series, where the main interest is non-stationary cross-group interaction; furthermore, the multivariate noise within groups can have both spatial covariation and non-stationary temporal covariation. Here, spatial covariation refers to dependence among the time series and, in the neural context, this results from the spatial arrangement of the electrodes, each of which records one of the time series. Our approach uses probabilistic CCA, but the framework allows rich spatiotemporal dependencies. These generalizations come at a cost: we now have a high-dimensional time series problem within each brain region together with a high-dimensional covariance structure. We solve these high-dimensional problems by imposing sparsity of the dominant effects, building on Bong et al. (2020), which treats the high-dimensional covariance structure in the context of observational white noise, and by incorporating banded covariance structure as in Bickel and Levina (2008). We thus call our method Latent Dynamic Factor Analysis of High-dimensional time series, LDFA-H.

In a simulation study, based on realistic synthetic time series, we verify the recovery of cross-region structure even when some of our assumptions are violated, and even in the presence of high noise. We then apply the method to 192 LFP time series recorded simultaneously from both Prefrontal Cortex (PFC) and visual area V4, during a memory task, and find time-varying cross-region dependencies.

2 Latent Dynamic Factor Analysis of High-dimensional time series

We treat the case of two groups of time series observed, repeatedly, N times. Let $X_{:,t}^1 \in \mathbb{R}^{p_1}$ and $X_{:,t}^2 \in \mathbb{R}^{p_2}$ be p_1 and p_2 recordings at time t in each of the two groups, for $t = 1, \dots, T$. As in Yu et al. (2009), we assume that a q -dimensional latent factor $Z_{:,t}^k \in \mathbb{R}^q$ drives each group, here, each brain region, according to the linear relationship

$$X_{:,t}^k | Z_{:,t}^k = \mu_{:,t}^k + \beta^k \cdot Z_{:,t}^k + \epsilon_{:,t}^k, \quad (1)$$

for brain region $k = 1, 2$, where $\mu_{:,t}^k \in \mathbb{R}^{p_k}$ are mean vectors, $\beta^k \in \mathbb{R}^{p_k \times q}$ are matrices of constant factor loadings, and $\epsilon_{:,t}^k \in \mathbb{R}^{p_k}$ are errors centered at zero (independently of the latent vectors Z). We are interested in the pairwise cross-group dependencies of the latent vectors $Z_{f,:}^1$ and $Z_{f,:}^2$, for $f = 1, \dots, q$. As in (Bong et al., 2020), we assume that the time series of these latent vectors follows a multivariate normal distribution

$$\begin{pmatrix} Z_{f,:}^1 \\ Z_{f,:}^2 \end{pmatrix} \sim \text{MVN}(0, \Sigma_f), \quad f = 1, \dots, q, \quad (2)$$

where Σ_f describes all of their simultaneous and lagged dependencies, both within and between the two vectors. We assume the N sets of random vectors (ϵ, Z) are independent and identically distributed. Fig. 1a illustrates the dependence structure of this model. We let P_f be the correlation matrix corresponding to Σ_f , and write its inverse as

$$\Pi_f = P_f^{-1} = \left(\begin{array}{c|c} \Pi_f^{11} & \Pi_f^{12} \\ \hline \Pi_f^{12\top} & \Pi_f^{22} \end{array} \right) \quad (3)$$

where Π_f^{11} and Π_f^{22} are the scaled auto-precision matrices and Π_f^{12} is the scaled cross-precision matrix. We now assume finite-range partial auto-correlation and cross-correlation for $(Z_{f,t}^1, Z_{f,t}^2)$, so that Π_f^{11} , Π_f^{22} and Π_f^{12} in Equation (3) have a banded structure. Specifically, for $k, l = 1, 2$, we assume there is a value h_f^{kl} such that Π_f^{kl} is a $(2h_f^{kl} + 1)$ -diagonal matrix. Because our goal is to address the cross-region connectivity and lead-lag relationship, we are particularly interested in the estimation of Π_f^{12} for each latent factor $f = 1, \dots, q$. Note that the non-zero elements $\Pi_{f,(t,s)}^{12}$, depicted as the red star in the expanded display within Fig. 1b, determine associations between the latent pair $Z_{f,:}^1$ and $Z_{f,:}^2$, which are simultaneous when $t = s$ and lagged when $t \neq s$.

Finally, we model the noise in Eq. (1) as a Gaussian random vector

$$\text{Vec}(\epsilon^k) = (\epsilon_{:,1}^k; \epsilon_{:,2}^k; \dots; \epsilon_{:,T}^k) \sim \text{MVN}(0, \Phi^k), \quad k = 1, 2, \quad (4)$$

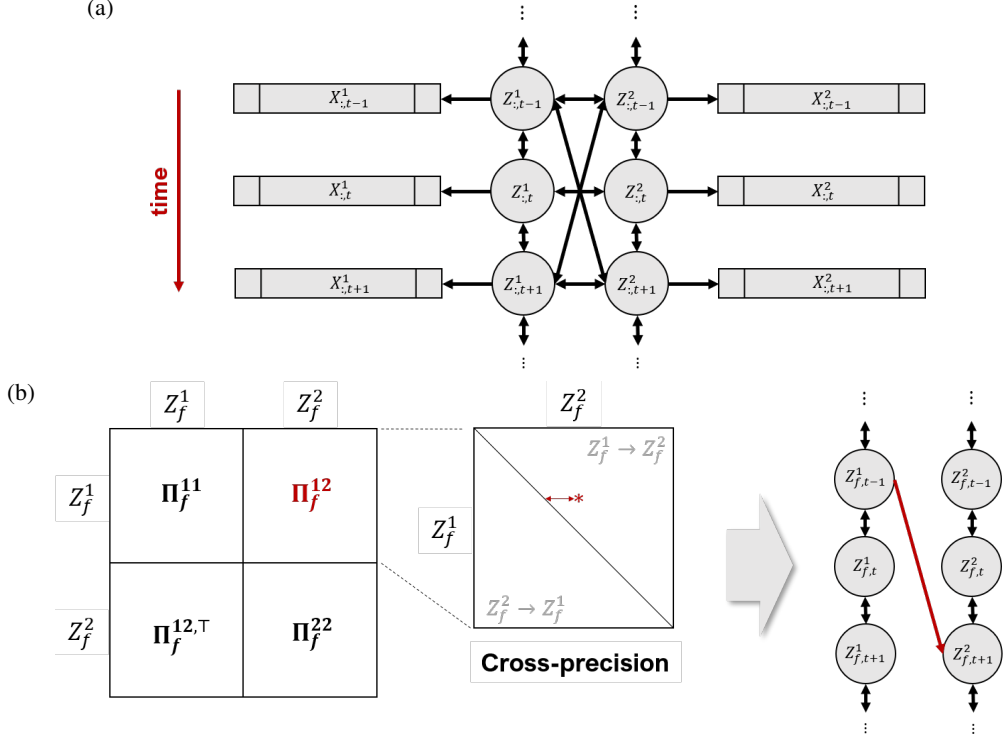


Figure 1: **LDFA-H model.** (a) Dynamic associations between vectors $X_{:,t}^1$ and $X_{:,t}^2$ are summarized by the dynamic associations between their associated 1D latent variables $Z_{:,t}^1$ and $Z_{:,t}^2$. (b) When a significant cross-precision entry is identified, e.g., the red star in the expanded view of Π_f^{12} , its coordinates and distance from the diagonal indicate at what time in the experiment connectivity between two brain areas occurs, and at what lead or lag. Here the red star is in the upper diagonal of Π_f^{12} , which means that, at this particular time, region 1 leads region 2, or $Z_f^1 \rightarrow Z_f^2$ in short (a non-zero entry in the lower diagonal would mean $Z_f^2 \rightarrow Z_f^1$). We represent this association by the red arrow on the right-most plot, with a lag of two units of time for illustration.

where we allow Φ^k to have non-zero off-diagonal elements to account for within-group spatiotemporal dependence. We assume Φ^k can be written in Kronecker product form

$$\Phi^k = \Phi_{\mathcal{T}}^k \otimes \Phi_S^k, \quad k = 1, 2, \quad (5)$$

where $\Phi_{\mathcal{T}}^k$ and Φ_S^k are the temporal and spatial components of Φ^k , as is often assumed for spatiotemporal matrix-normal distributions, e.g., (Dawid, 1981). Although this is a strong assumption, implying, for instance, that the auto-correlation of every $X_{i,:}^k$ is proportional to $\Phi_{\mathcal{T}}^k$, we regard Φ_k as a nuisance parameter: our primary interest is Σ_f in Eq. (2). We also assume an auto-regressive order at most h_ϵ^k , so that $\Gamma_{\mathcal{T}}^k = (\Phi_{\mathcal{T}}^k)^{-1}$ is a $(2h_\epsilon^k + 1)$ -diagonal matrix. In our simulation we show that we can recover Σ_f accurately even when the Kronecker product and bandedness assumptions fail to hold.

The model in Equations (1)-(5) generalizes other known models. First, when $q = 1$, and $Z^1 = Z^2$ remains constant over time, in the noiseless case ($\epsilon_k = 0$), it reduces to the probabilistic CCA model of Bach and Jordan (2005); see Theorem 2.2 of Bong et al. (2020). Thus, model (1)-(5) can be viewed as a denoising, multi-level and dynamic version of probabilistic CCA. Second, when $k = 1$, the Gaussian processes are stationary, and the ϵ vectors are white noise, (1)-(5) reduces to GPFA (Yu et al., 2009). Thus, (1)-(5) is a two-group, nonstationary extension of GPFA that allows for within-group spatio-temporal dependence.

Identifiability and sparsity constraints Despite the structure imposed on Φ_k in Eq. (5), parameter identifiability issues remain. Our model in Eqs. (1), (2) and (4) induces the marginal distribution of

the observed data (X^1, X^2) :

$$(X_{:,1}^1; X_{:,2}^1; \dots; X_{:,T}^1; X_{:,1}^2; X_{:,2}^2; \dots; X_{:,T}^2) \sim \mathcal{N}((\mu_{:,1}^1; \mu_{:,2}^1; \dots; \mu_{:,T}^1; \mu_{:,1}^2; \mu_{:,2}^2; \dots; \mu_{:,T}^2), S) \quad (6)$$

where S is the marginal covariance matrix given by:

$$S = \begin{bmatrix} \Phi_{\mathcal{T}}^1 \otimes \Phi_{\mathcal{S}}^1 & 0 \\ 0 & \Phi_{\mathcal{T}}^2 \otimes \Phi_{\mathcal{S}}^2 \end{bmatrix} + \sum_{f=1}^q \begin{bmatrix} \Sigma_f^{11} \otimes (\beta_f^1 \beta_f^{1\top}) & \Sigma_f^{12} \otimes (\beta_f^1 \beta_f^{2\top}) \\ \Sigma_f^{12\top} \otimes (\beta_f^2 \beta_f^{1\top}) & \Sigma_f^{22} \otimes (\beta_f^2 \beta_f^{2\top}) \end{bmatrix}. \quad (7)$$

The family of parameters

$$\theta^{\{\alpha^1, \alpha^2\}} = \left\{ \Sigma_1^{\{\alpha_1^1, \alpha_1^2\}}, \dots, \Sigma_q^{\{\alpha_q^1, \alpha_q^2\}}, \Phi_{\mathcal{S}}^1 - \sum_{f=1}^q \alpha_f^1 \beta_f^1 \beta_f^{1\top}, \Phi_{\mathcal{S}}^2 - \sum_{f=1}^q \alpha_f^2 \beta_f^2 \beta_f^{2\top}, \Phi_{\mathcal{T}}^1, \Phi_{\mathcal{T}}^2, \beta^1, \beta^2, \mu^1, \mu^2 \right\}, \quad (8)$$

where $\Sigma_f^{\{\alpha_f^1, \alpha_f^2\}} = \left\{ \Sigma_f + \begin{bmatrix} \alpha_f^1 \Phi_{\mathcal{T}}^1 & 0 \\ 0 & \alpha_f^2 \Phi_{\mathcal{T}}^2 \end{bmatrix} \right\}$, induce the same marginal distribution in Eq. (6),

for all $\alpha^1, \alpha^2 \in \mathbb{R}^q$ (notice that $\theta = \theta^{\{0,0\}} = \{\Sigma_1, \dots, \Sigma_q, \Phi_{\mathcal{S}}^1, \Phi_{\mathcal{S}}^2, \Phi_{\mathcal{T}}^1, \Phi_{\mathcal{T}}^2, \beta^1, \beta^2, \mu^1, \mu^2\}$ is the original parameter). Preliminary analysis of LFP data indicated that strong cross-region dependence occurs relatively rarely. We therefore resolve this lack of identifiability by choosing the solution given by maximizing the likelihood with an L1 penalty, under the assumption that the inverse cross-correlation matrix Π_f^{12} is a sparse $(2h_f^{12} + 1)$ -diagonal matrix.

Latent Dynamic Factor Analysis of High-dimensional time series (LDFA-H) Given N simultaneously recorded pairs of neural time series $\{X^1[n], X^2[n]\}_{n=1, \dots, N}$, the maximum penalized likelihood estimator (MPLE) of the inverse correlation matrix of the latent variables solves

$$\begin{aligned} (\hat{\Pi}_1, \dots, \hat{\Pi}_q) &= \operatorname{argmin} -\frac{1}{N} \sum_{n=1}^N l(\theta; X^1[n], X^2[n]) + \sum_{f=1}^q \sum_{k,l=1}^2 \|\Lambda_f^{kl} \odot \Pi_f^{kl}\|_1 \\ &\text{s.t. } \Gamma_{\mathcal{T}}^k \text{ is } (2h_{\epsilon}^k + 1)\text{-diagonal,} \end{aligned} \quad (9)$$

where the log-likelihood is

$$l(\theta; X^1, X^2) = -\log \det S - (X_{:,1}^1 - \mu_{:,1}^1; \dots; X_{:,T}^1 - \mu_{:,T}^1; X_{:,1}^2 - \mu_{:,1}^2; \dots; X_{:,T}^2 - \mu_{:,T}^2)^{\top} S^{-1} (X_{:,1}^1 - \mu_{:,1}^1; \dots; X_{:,T}^1 - \mu_{:,T}^1; X_{:,1}^2 - \mu_{:,1}^2; \dots; X_{:,T}^2 - \mu_{:,T}^2), \quad (10)$$

with S defined in Eq. (7), and the constraints are

$$\Lambda_{f,(t,s)}^{kl} = \begin{cases} \infty, & (t, s) : |t - s| > h_f^{kl}, \\ \lambda_f, & (t, s) : 0 < |t - s| \leq h_f^{kl}, \quad k \neq l, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

for factor $f = 1, \dots, q$ and brain region $k = 1, 2$. The first constraint forces the corresponding $\Pi_{f,(t,s)}^{kl}$ to zero and thus imposes a banded structure for Π_f^{kl} , and the second assigns the same sparsity constraint λ_f on the off-diagonal elements of Π_f^{12} . Finally, to make calibration of tuning parameters computationally feasible, we set the bandwidth for the latent precisions and the noise precisions within each region to a single value h_{auto} , we set the bandwidth for the latent precisions across regions to a value h_{cross} , and we set the sparsity parameters to a value λ_{cross} , i.e.,

$$h_f^{kk} = h_{\epsilon}^k = h_{\text{auto}}, \quad h_f^{12} = h_{\text{cross}} \quad \text{and} \quad \lambda_f = \lambda_{\text{cross}},$$

for each factor $f = 1, \dots, q$ and region $k = 1, 2$. The bandwidths are chosen using domain knowledge and preliminary data analyses. We determine the remaining parameters λ_{cross} and q by 5-fold cross-validation (CV).

Solving Eq. (9) requires S^{-1} . Because it is not available analytically and a numerical approximation is computationally prohibitive, we solve Eq. (9) using an EM algorithm (Dempster et al., 1977). Let $\theta^{(r)}$ be the parameter estimate at the r -th iteration. We consider the data $\{X^1[n], X^2[n]\}_{n=1, \dots, N}$ to be incomplete observations of $\{X^1[n], Z^1[n], X^2[n], Z^2[n]\}_{n=1, \dots, N}$. In the E-step, we estimate the conditional mean and covariance matrix of each $\{Z^1[n], Z^2[n]\}$ with respect to $\{X^1[n], X^2[n]\}$ and $\theta^{(r)}$. Given these sufficient statistics, the problem of computing the MPLE decomposes into two separate minimizations of

1. the negative log-likelihood of Σ_f , w.r.t. the latent factor model (Eq. (2)) and
2. the negative log-likelihood of $\Phi_S^1, \Phi_S^2, \Phi_T^1, \Phi_T^2, \beta^1, \beta^2, \mu^1, \mu^2$ w.r.t. the observation model (Eqs. (1) and (4)).

With the noise correlation and latent factor correlation disentangled, the M-step reduces to easy sub-problems. For example, the minimization with respect to Σ_f is a graphical Lasso problem (Friedman et al., 2007) and the minimization with respect to Φ_S^k and Φ_T^k is a maximum likelihood estimation of a matrix-variate distribution (Dawid, 1981). We thus obtain an affordable M-step, and alternating E and M-steps produces a solution to the MPLLE problem.

We derive the full formulations in Appendix A. Its computational cost is inexpensive: a single iteration of E and M-steps on our cluster server (with 11 Intel(R) Xeon(R) CPU 2.90GHz processors) took in average less than 45 seconds, applied to the experimental data in Section 3.2. A single fit on the same data took 42 iterations for around 30 minutes until P and $\{\beta^1, \beta^2\}$ converged under threshold $1e-3$ and $1e-5$, respectively. The code is provided at <https://github.com/HeejongBong/ldfa>.

3 Results

One major novelty of our method is its accounting for auto-correlated noise in neural time series to better estimate cross-region associations in CCA type analysis. This is illustrated in Section 3.1 based on simulated data. In Section 3.2, we apply LDFA-H to experimental data to examine the lead-lag relationships across two brain areas and the spatial distribution of factor loadings.

3.1 LDFA-H retrieves cross-correlations even when noise auto-correlations dominate

We simulated $N = 1000$ i.i.d. neural time series X^k of duration $T = 50$ from Eq. (1) for brain regions $k = 1, 2$. The latent time series Z^k were generated from Eq. (2) with $q = 1$ pair of factors and correlation matrix P_1 depicted in Fig. 2a. The noise ϵ^k was taken to be the $N = 1000$ trials of the experimental data analyzed in Section 3.2, first permuted to remove cross-region correlations, then contaminated with white noise to modulate the strength of noise correlation relative to cross-region correlations. The resulting temporal noise correlation matrices, found by averaging correlations over all pairs of simulated time series, are shown in Fig. 2b, for four levels of white noise contamination. The magnitudes of cross-region correlation and within-region noise auto-correlation are quantified by the determinant of each matrix, known as the generalized variance (Wilks, 1932); their logarithms are provided atop the panels in Fig. 2a and Fig. 2b. Generalized variance ranges from 0 (identical signals) to 1 (independent signals). Thus, larger negative values indicate stronger within-region noise correlation (see B). Other simulation details are in B.

We note that the simulation does not satisfy some of the model assumptions in Section 2. The noise vectors ϵ^k are not matrix-variate distributed as in Eqs. (4) and (5) and the derived Γ_T^k does not satisfy a banded structure as in Eq. (9). Also, the latent partial auto-correlations (Fig. 2) are not banded as assumed in Eq. (9).

We applied LDFA-H with $q = 1$ factor, $h_{\text{cross}} = 10$, h_{auto} equal to the maximum order of the auto-correlations in the 2000 observed simulated time series, and λ_{cross} determined by 5-fold CV. Fig. 3 shows LDFA-H cross-precision matrix estimates corresponding to the four level of noise correlation shown in Fig. 2b. They closely match the true Π_1^{12} shown in the right most panel of Fig. 2a.

We also applied five other methods to estimate cross-region connections in the simulated data. They include the popular averaged pairwise correlation (APC); correlation of averaged signals (CAS); and CCA (Hotelling, 1936), applied to the NT observed pairs of multivariate random vectors $\{X_{:,t}^1, X_{:,t}^2\}_{n,t \in [N] \times [T]}$ to estimate the cross-correlation matrix between the canonical variables; as well as DKCCA (Rodu et al., 2018) and LaDynS (Bong et al. (2020)). The first four methods do not explicitly provide cross-precision matrix estimates, so we display their cross-correlation matrix estimates in Fig. 4, along with LDFA-H cross-correlation estimates in the last row. It is clear that only LDFA-H successfully recovered the true cross-correlations shown in the second panel of Fig. 2a, at all auto-correlated noise levels.

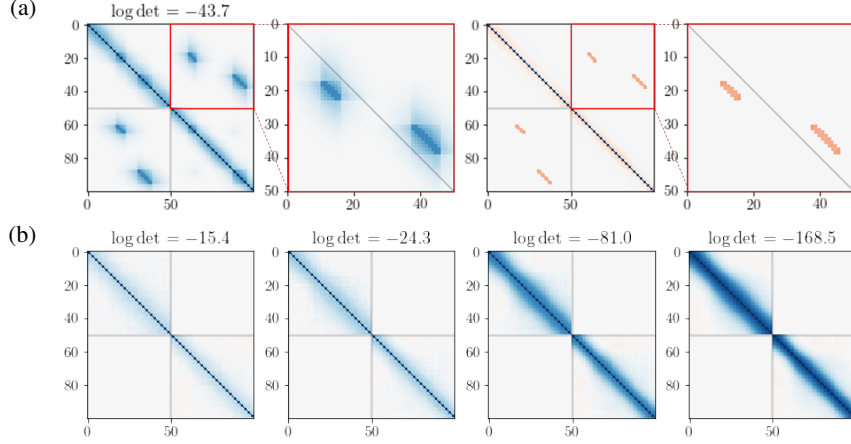


Figure 2: **Simulation settings.** (a) (Left to right panels) True correlation matrix P_1 for latent factors Z_1^1 , and Z_2^2 , from model in Eq. (2); close-up of the cross-correlation matrix; corresponding precision matrix $\Pi_1 = P_1^{-1}$; and close-up of cross-precision matrix Π_1^{12} (Eq. (3)). Matrix axes represent the duration, $T = 50$ ms, of the time series. Factors Z^1 and Z^2 are associated in two epochs: Z^2 precedes Z^1 by 7ms from $t = 13$ to 19ms, and Z^1 precedes Z^2 by 7ms from $t = 33$ to 42ms. (b) Noise auto-correlation matrices (Eq. (5)) for pairs of simulated time series at four strength levels. log det in (a) and (b) measure correlation strengths.

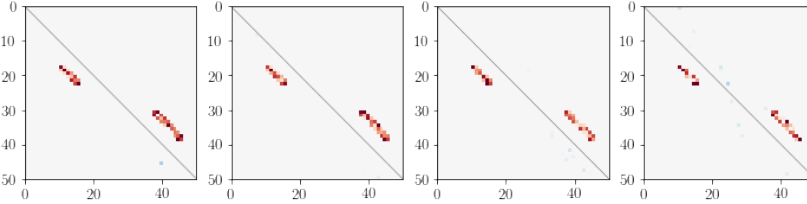


Figure 3: **Simulation results: LDFA-H cross-precision matrix estimates.** Estimates of Π_1^{12} , shown in the right-most panel of Fig. 2a, using LDFA-H, for the four noise auto-correlation strengths shown in Fig. 2b. LDFA-H identified the true cross-area connections at all noise strengths.

3.2 Experimental Data Analysis from Memory-Guided Saccade Task

We now report the analysis of LFP data in areas PFC and V4 of a monkey during a saccade task, provided by Khanna et al. (2020). One trial of the experiment consisted of four stages: (i) fixation: the animal fixated at the center of the screen; (ii) cue: a cue appeared on the screen randomly at one of eight locations; (iii) delay: the animal had to remember the cue location while maintaining eye fixation; (iv) choice: the monkey made a saccade to the remembered cue location. We focused our analysis on the 500 ms delay period, when the animal both processed cue information and prepared a saccade. LFP data were recorded for $N = 1000$ trials by two 96-electrode Utah arrays implanted in PFC and V4, β band-passed filtered, down-sampled from 1 kHz to 100 Hz.

We applied LDFA-H using $h_{\text{auto}} = h_{\text{cross}} = 10$, corresponding to 100 ms (at 100 Hz); the LFP β -power envelopes have frequencies between 12.5 Hz to 30 Hz, and $h_{\text{auto}} = 10$ enables the slowest filtered signal to complete one full oscillation period. The other tuning parameters were determined by 5-fold CV over $\lambda_{\text{cross}} \in \{0.0002, 0.002, 0.02, 0.2\}$ and $q \in \{5, 10, 15, 20, 25, 30\}$, yielding optimal values $\lambda_{\text{cross}} = 0.02$ and $q = 10$. We also regularize the diagonal elements, due to the otherwise excessively smooth β -power envelopes (see our code or Bong et al. (2020) for details). The fitted factors were ranked based on the Frobenius norms of their covariance matrices $\|\Sigma_f\|_F^2$; norms are plotted versus f in decreasing order in Fig. C.1, and $\log_{10} \|\Sigma_f\|_F^2$ of the top three factors are provided above each panel in Fig. 5a. The estimated cross-precision matrices between two brain regions corresponding to the top three factors are shown in Fig. 5a. Note that a positive entry in the precision matrix represents negative association between two regions. We also summarized, for

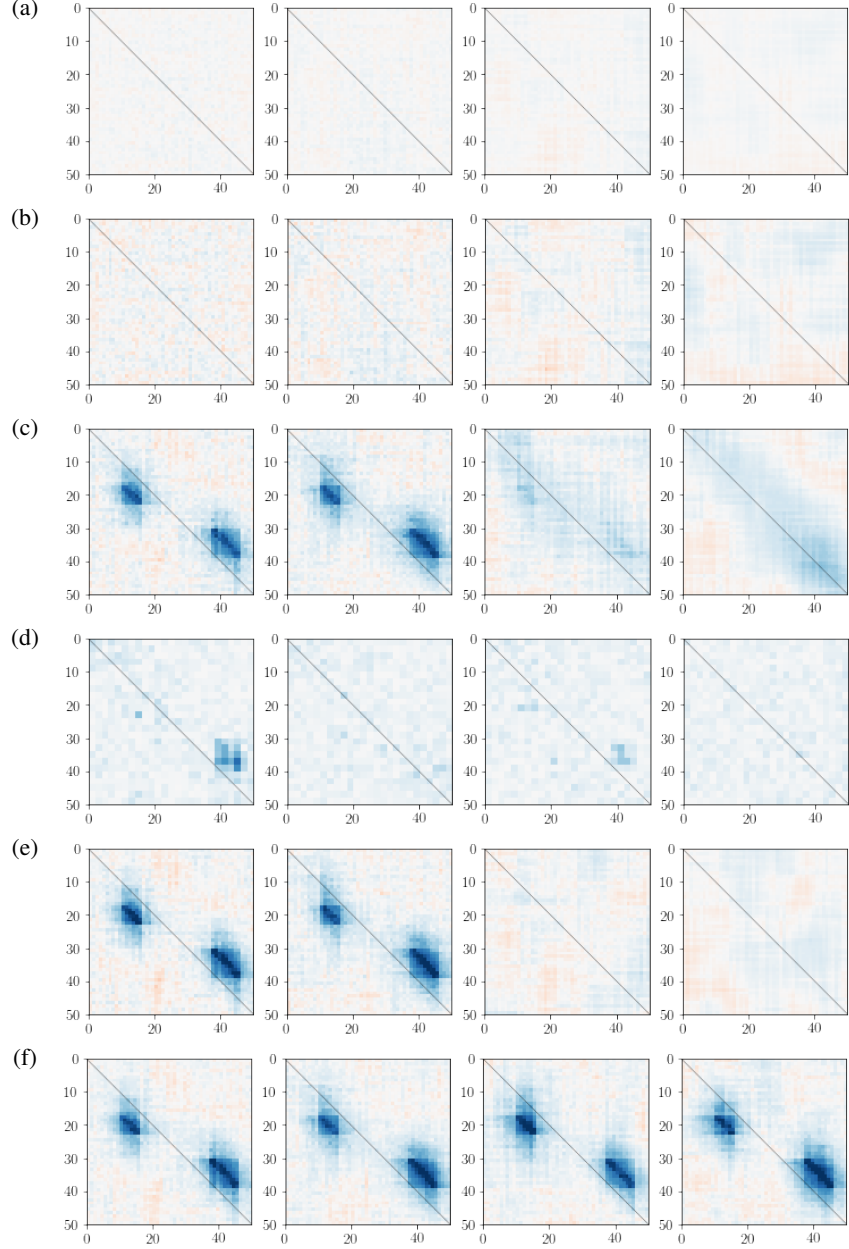


Figure 4: **Simulation results: cross-correlation matrix estimates.** Estimates of Σ_1^{12} under four noise correlation levels using (a) averaged pairwise correlation (APC), (b) correlation of averaged signal (CAS), (c) canonical correlation analysis (CCA, Hotelling (1936)), (d) dynamic kernel CCA (DKCCA, Rodu et al. (2018)), (e) LaDynS (Bong et al. (2020)), and (f) LDFA-H. Only LDFA-H successfully recovered the true cross-correlation at all noise auto-correlation strengths.

each factor f , the temporal information flow at time t from V4 to PFC and to V4 from PFC with $I_{f,PFC \rightarrow V4}(t) = \sum_{t' > t} \left| \widehat{\Pi}_{f,(t,t')}^{12} \right|$ and $I_{f,V4 \rightarrow PFC}(t) = -\sum_{t' < t} \left| \widehat{\Pi}_{f,(t,t')}^{12} \right|$, respectively, where $\widehat{\Pi}_f$ is the inverse correlation matrix estimate in Eq. (9). Fig. 5d displays smoothed $I_{f,PFC \rightarrow V4}(t)$ and $I_{f,V4 \rightarrow PFC}(t)$ as functions of t for the top three factors. Lead-lag relationships between V4 and PFC change dynamically over time, and the information flow tends to peak either early in the delay period, when the animal must remember the cue, or later, when it must make a saccade decision. The dominant first factor captures a flow from V4 to PFC centered around 200 milliseconds into the

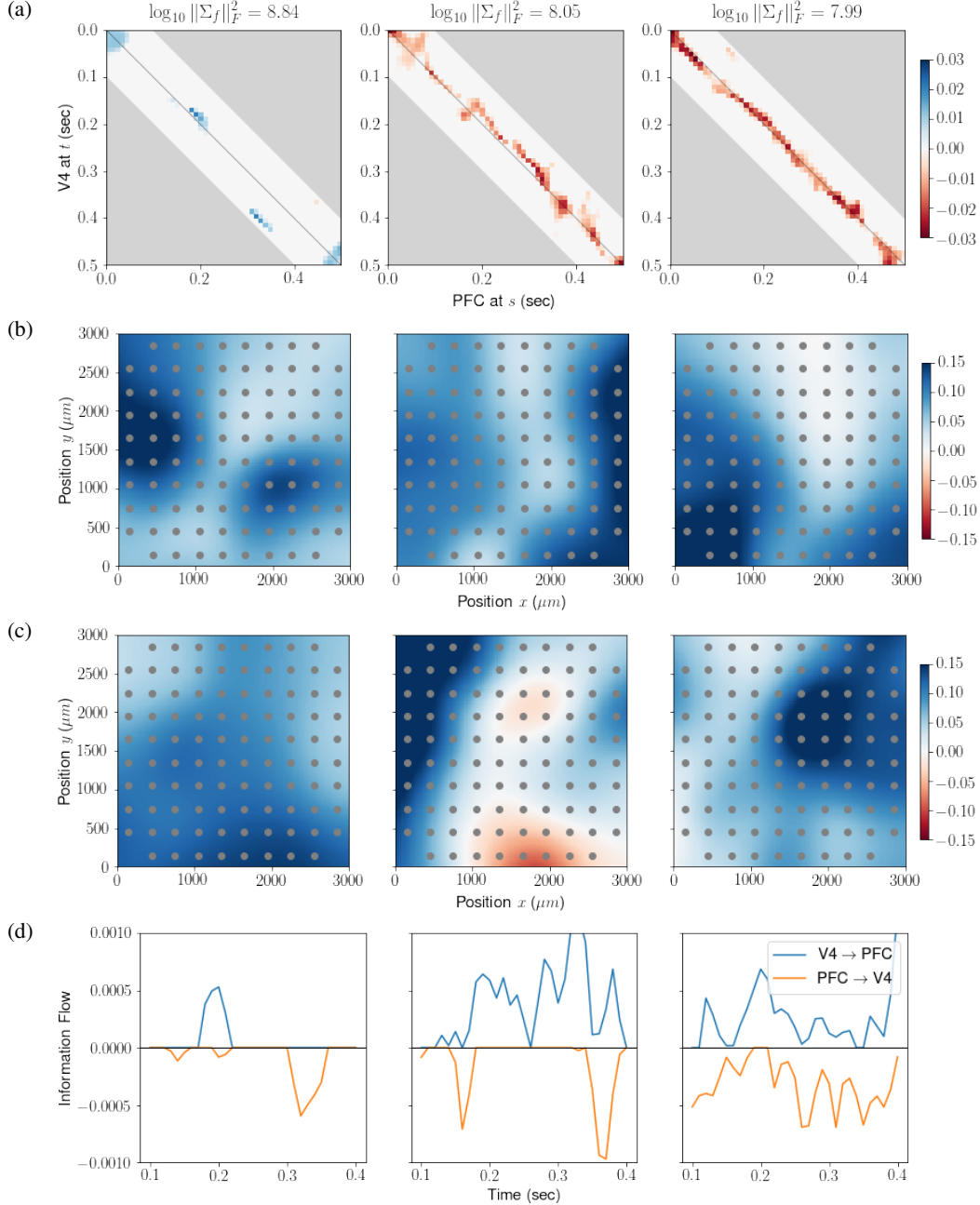


Figure 5: Experimental data results for the top 3 factors. (a) Cross-precision matrices. Blue represents positive precision matrix entries, corresponding to negative association. Factors have different connectivity patterns over the experimental trials. $\log_{10} \|\Sigma_f\|_F^2$, written atop the panels, measures the strength of each factor. The first factor is more than 6 times larger than the second and third, and displays activity in V4 leading PFC centered around 200 milliseconds and activity in PFC leading V4 centered around 320 milliseconds post cue disappearance. This is also shown in panel (d). (b,c) Factor loadings, smoothed and color coded, plotted on the electrode coordinates (μm). Here, positivity is arbitrary, due to identifiability. Panels (b) and (c) display loadings for the V4 and PFC arrays, respectively. The first factor has activity in V4 centered in two distinct subregions of the array, while activity in PFC is more broadly distributed. (d) Dynamic information flow in the directions $V4 \rightarrow PFC$ (blue) and $PFC \rightarrow V4$ (orange).

task and a flow from PFC to V4 centered around 320 milliseconds. Factor loadings (subsampling over space) for the 96 V4 and PFC electrodes are shown in Fig. 5b and Fig. 5c, respectively, for the top three factors (first three columns of the estimate of β^k in Eq. (9), with area $k = 1$ being V4 and $k = 2$ being PFC), arranged spatially according to electrode positions on the Utah array. The factors have different spatial modes over the physical space of the Utah array. Confirmation of these patterns would require additional data and analyses.

4 Conclusion

To identify dynamic interactions across brain regions we have developed LDFA-H, a nonstationary, multi-group extension of GPFA that allows for within-group spatio-temporal dependence among high-dimensional neural recordings. We applied the method to data during a memory task and found interesting, intuitive results. Although we treated the two-group case, and applied it to interactions across two brain regions, several groups can be handled with straightforward modifications. The approach could, in principle, be applied to many different types of time series, but it has some special features: first, like all methods based on sparsity, it assumes a small number of large effects are of primary interest; second, it uses repetitions, here, repeated trials, to identify time-varying dependence; third, because the within-group spatio-temporal structure is not of interest, the method can remain useful even with some modest within-group model misspecification.

Several restrictive assumptions of LDFA-H, as defined, were helpful here but could be modified for other applications. One is the Kronecker-product form of the noise process. In our simulation study, using a realistic scenario, we showed that LDFA-H can be effective even when the Kronecker-product assumption is violated, but in other cases it may be problematic. In some problems, space and/or time can be decomposed into windows within which the assumption is more reasonable (see Leng and Tang, 2012; Zhou, 2014). Another potentially bothersome assumption is independence between latent factors. It would be possible to include covariance matrix parameters between the factors, but then the model will get computationally prohibitive even with a moderate factor size. State-space models (Buesing et al., 2014; Linderman et al., 2019; Yang et al., 2016) have potential but, to be comparable to LDFA-H, they would have to accommodate nonstationary lead-lag behavior. Computationally efficient methods for identifying time-varying relationships is a vital goal in the analysis of neural data from multiple brain regions.

We applied LDFA-H to LFP data. In contrast, GPFA has been applied mainly to neural spike count data, and it is of course possible to apply LDFA-H to spike counts, as well. However, we have been struck by the strong attenuation of effects due to Poisson-like noise, as discussed in Vinci et al. (2018) and references therein. A version of LDFA-H built for Poisson-like counts, or for point processes, could be the subject of additional research. It may also be advantageous to model spatial dependence explicitly, perhaps based on physical distance between electrodes, analogously to what was done in Vinci et al. (2018), and there may be, in addition, important simplifications available in the temporal structure. It would also be helpful to have additional statistical inference procedures for assessing effects. In the future, we hope to pursue these possible directions, and refine the application of this promising approach to the analysis of high-dimensional neural data.

Broader Impact

While progress in understanding the brain is improving life through research, especially in mental health and addiction, in no case is any brain disorder well understood mechanistically. Faced with the reality that each promising discovery inevitably reveals new subtleties, one reasonable goal is to be able to change behavior in desirable ways by modifying specific brain circuits and, in animals, technologies exist for circuit disruptions that are precise in both space and time. However, to determine the best location and time for such disruptions to occur, with minimal off-target effects, will require far greater knowledge of circuits than currently exists: we need good characterizations of interactions among brain regions, including their timing relative to behavior. The over-arching aim of our research is to provide methods for describing the flow of information, based on evolving neural activity, among multiple regions of the brain during behavioral tasks. Such methods can lead to major advances in experimental design and, ultimately, to far better treatments than currently exist.

Acknowledgments and Disclosure of Funding

Bong, Liu, Ventura, and Kass are supported in part by NIMH grant R01 MH064537. Smith is supported in part by NIMH grant R01 MH118929. Ren is supported in part by NSF grant DMS 1812030.

References

- Adhikari, A., Sigurdsson, T., Topiwala, M. A., and Gordon, J. A. (2010). Cross-correlation of instantaneous amplitudes of field potential oscillations: a straightforward method to estimate the directionality and lag between brain areas. *Journal of neuroscience methods*, 191:191–200.
- Bach, F. R. and Jordan, M. I. (2005). A probabilistic interpretation of canonical correlation analysis. Technical Report 688, Department of Statistics, University of California, Berkeley, Berkeley, CA.
- Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.*, 36:199–227.
- Bong, H., Ventura, V., Smith, M., and Kass, R. (2020). Latent cross-population dynamic time-series analysis of high-dimensional neural recordings. *Manuscript submitted for publication*.
- Buesing, L., Machado, T. A., Cunningham, J. P., and Paninski, L. (2014). Clustered factor analysis of multineuronal spike data. In *Advances in Neural Information Processing Systems*, pages 3500–3508.
- Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13:407–420.
- Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, 68:265–274.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39:1–22.
- Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14:770–785.
- Friedman, J., Hastie, T., and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9:432–441.
- Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291:1560–1563.
- Gallagher, N., Ulrich, K. R., Talbot, A., Dzirasa, K., Carin, L., and Carlson, D. E. (2017). Cross-spectral factor analysis. In *Advances in Neural Information Processing Systems*, pages 6842–6852.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28:321–377.
- Hultman, R., Ulrich, K., Sachs, B. D., Blount, C., Carlson, D. E., Ndubuizu, N., Bagot, R. C., Parise, E. M., Vu, M.-A. T., Gallagher, N. M., et al. (2018). Brain-wide electrical spatiotemporal dynamics encode depression vulnerability. *Cell*, 173:166–180.
- Jiang, H., Bahramisharif, A., van Gerven, M. A., and Jensen, O. (2015). Measuring directionality between neuronal oscillations of different frequencies. *Neuroimage*, 118:359–367.
- Khanna, S. B., Scott, J. A., and Smith, M. A. (2020). Dynamic shifts of visual and saccadic signals in prefrontal cortical regions 8Ar and FEF. *Journal of Neurophysiology*. In press.
- Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association*, 107:1187–1200.
- Liebe, S., Hoerzer, G. M., Logothetis, N. K., and Rainer, G. (2012). Theta coupling between v4 and prefrontal cortex predicts visual short-term memory performance. *Nature neuroscience*, 15:456.
- Linderman, S., Nichols, A., Blei, D., Zimmer, M., and Paninski, L. (2019). Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in *c. elegans*. *bioRxiv*, page 621540.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157.
- Magri, C., Schridde, U., Murayama, Y., Panzeri, S., and Logothetis, N. K. (2012). The amplitude and timing of the bold signal reflects the relationship between local field potential power at different frequencies. *Journal of Neuroscience*, 32:1395–1407.

- Mazumder, R., Hastie, T., and Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11:2287–2322.
- Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24:167–202.
- Orban, G. A. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiological reviews*, 88:59–89.
- Rodu, J., Klein, N., Brincat, S. L., Miller, E. K., and Kass, R. E. (2018). Detecting multivariate cross-correlation between brain regions. *Journal of Neurophysiology*, 120:1962–1972. PMID: 29947591.
- Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G., and Von Stein, A. (1998). Synchronization between prefrontal and posterior association cortex during human working memory. *Proceedings of the National Academy of Sciences*, 95:7092–7096.
- Steinmetz, N. A., Koch, C., Harris, K. D., and Carandini, M. (2018). Challenges and opportunities for large-scale electrophysiology with neuropixels probes. *Current opinion in neurobiology*, 50:92–100.
- Vinci, G., Ventura, V., Smith, M. A., and Kass, R. E. (2018). Adjusted regularization of cortical covariance. *Journal of Computational Neuroscience*, 45:83–101.
- Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, pages 471–494.
- Yang, Y., Aminoff, E., Tarr, M., and Robert, K. E. (2016). A state-space model of cross-region dynamic connectivity in meg/eeg. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 1234–1242. Curran Associates, Inc.
- Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, 102:614–635. PMID: 19357332.
- Zhou, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Ann. Statist.*, 42:532–562.

A EM-algorithm to fit LDFA-H (Section 2)

Initialization Let $\hat{\theta}^{(0)} = \{\hat{\Sigma}_1^{(0)}, \dots, \hat{\Sigma}_q^{(0)}, \hat{\Phi}_S^{1,(0)}, \hat{\Phi}_S^{2,(0)}, \hat{\Phi}_T^{1,(0)}, \hat{\Phi}_T^{2,(0)}, \hat{\beta}_1^{1,(0)}, \hat{\beta}_1^{2,(0)}, \hat{\mu}^{1,(0)}, \hat{\mu}^{2,(0)}\}$ be the initial parameter value. Since the MPLLE objective function for LDFA-H given in Eq. (9) is not guaranteed convex, an EM-algorithm may find a local minimum according to a choice of the initial value. Hence a good initialization is crucial to a successful estimation. Here we suggest an initialization by a canonical correlation analysis (CCA).

Let $\{X^1[n], X^2[n]\}_{n=1, \dots, N}$ be N simultaneously recorded pairs of neural time series. We can view them as NT recorded pairs of multivariate random vectors $\{X_{:,t}^1[n], X_{:,t}^2[n]\}_{(n,t) \in [N] \times [T]}$. We obtain $\hat{\beta}_1^{1,(0)}$ and $\hat{\beta}_1^{2,(0)}$ by CCA as follows:

$$\hat{\beta}_1^{1,(0)}, \hat{\beta}_1^{2,(0)} = \underset{\beta_1^1 \in \mathbb{R}^{p_1}, \beta_1^2 \in \mathbb{R}^{p_2}}{\operatorname{argmax}} \frac{\beta_1^{1\top} S^{12} \beta_1^2}{\sqrt{\beta_1^{1\top} S^{11} \beta_1^1} \sqrt{\beta_1^{2\top} S^{22} \beta_1^2}} \quad (\text{A.1})$$

where

$$\begin{aligned} S^{11} &= \frac{1}{NT} \sum_{n,t} (X_{:,t}^1[n] - \frac{1}{NT} \sum_{n,t} X_{:,t}^1[n]) (X_{:,t}^1[n] - \frac{1}{NT} \sum_{n,t} X_{:,t}^1[n])^\top \\ S^{22} &= \frac{1}{NT} \sum_{n,t} (X_{:,t}^2[n] - \frac{1}{NT} \sum_{n,t} X_{:,t}^2[n]) (X_{:,t}^2[n] - \frac{1}{NT} \sum_{n,t} X_{:,t}^2[n])^\top \\ S^{12} &= \frac{1}{NT} \sum_{n,t} (X_{:,t}^1[n] - \frac{1}{NT} \sum_{n,t} X_{:,t}^1[n]) (X_{:,t}^2[n] - \frac{1}{NT} \sum_{n,t} X_{:,t}^2[n])^\top. \end{aligned} \quad (\text{A.2})$$

According to the equivalence between CCA and probabilistic CCA shown by A. Anonymous, it gives an estimate of the first latent factors

$$\hat{Z}_{1,:}^{k,(0)}[n] = \hat{\beta}_1^{k,(0)} X^k[n] \quad (\text{A.3})$$

for $n = 1, \dots, N$ and $k = 1, 2$. The initial second latent factors $\hat{Z}_2^{k,(0)}$ and the corresponding factor loading $\hat{\beta}_2^{k,(0)}$ is similarly set by the second pair of canonical variables, and so on. Then we assign the empirical covariance matrix of $\{\hat{Z}_f^{1,(0)}[n], \hat{Z}_f^{2,(0)}[n]\}_{n \in [N]}$ to the initial latent covariance matrix $\hat{\Sigma}_f^{(0)}$ for $f = 1, \dots, q$ and the matrix-variate normal estimate (Zhou, 2014) on $\{\hat{\epsilon}^{k,(0)}[n] := X^k[n] - \hat{\beta}_1^{k,(0)} \hat{Z}_{1,:}^{k,(0)}[n]\}_{n \in [N]}$ to $\hat{\Phi}_T^{k,(0)}$ and $\hat{\Phi}_S^{k,(0)}$ for $k = 1, 2$. Along $\hat{\mu}^{k,(0)} := \frac{1}{N} \sum_{n=1}^N X^k[n]$, the above parameters comprises the initial parameter set $\hat{\theta}^{(0)}$.

However, we cannot run an E-step on the above parameter set because $\hat{\Phi}^{k,(0)}$ is not invertible. We instead pick one of its unidentifiable parameter sets $\hat{\theta}^{(0), \{\alpha^1, \alpha^2\}}$, defined in Eq. (8), with all $\hat{\Phi}^{k,(0)}$'s and $\hat{\Sigma}_f^{(0)}$'s invertible. Specifically, we take

$$\alpha_f^k = \frac{1}{2} \lambda_{\min} \left(\Sigma_f^{1/2} \begin{bmatrix} \Phi_T^1 & 0 \\ 0 & \Phi_T^2 \end{bmatrix}^{-1} \Sigma_f^{1/2} \right) \quad (\text{A.4})$$

for $f = 1, \dots, q$ and $k = 1, 2$ where $\lambda_{\min}(A)$ is the smallest eigenvalue of symmetric matrix A . Henceforth, we notate $\hat{\theta}^{(0), \{\alpha^1, \alpha^2\}}$ by $\hat{\theta}^{(0)}$. For $t = 1, 2, \dots$, we iterate the following E-step and M-step until convergence.

Another promising initialization is by finding time (t, s) on which the canonical correlation between $X_{:,t}^1$ and $X_{:,s}^2$ maximizes. i.e., we initialize $\hat{\beta}_1^{1,(0)}$ and $\hat{\beta}_1^{2,(0)}$ by

$$\hat{\beta}_1^{1,(0)}, \hat{\beta}_1^{2,(0)} = \underset{\beta_1^1 \in \mathbb{R}^{p_1}, \beta_1^2 \in \mathbb{R}^{p_2}}{\operatorname{argmax}} \frac{\beta_1^{1\top} S_{(t,s)}^{12} \beta_1^2}{\sqrt{\beta_1^{1\top} S_{(t,t)}^{11} \beta_1^1} \sqrt{\beta_1^{2\top} S_{(s,s)}^{22} \beta_1^2}} \text{ such that } |t - s| < h_{\text{cross}}. \quad (\text{A.5})$$

where

$$\begin{aligned}
S_{(t,t)}^{11} &= \frac{1}{N} \sum_{n,t} (X_{:,t}^1[n] - \frac{1}{N} \sum_n X_{:,t}^1[n]) (X_{:,t}^1[n] - \frac{1}{N} \sum_n X_{:,t}^1[n])^\top \\
S_{(s,s)}^{22} &= \frac{1}{N} \sum_{n,s} (X_{:,s}^2[n] - \frac{1}{N} \sum_n X_{:,s}^2[n]) (X_{:,s}^2[n] - \frac{1}{N} \sum_n X_{:,s}^2[n])^\top \\
S_{(t,s)}^{12} &= \frac{1}{N} \sum_{n,t} (X_{:,t}^1[n] - \frac{1}{N} \sum_n X_{:,t}^1[n]) (X_{:,s}^2[n] - \frac{1}{N} \sum_n X_{:,s}^2[n])^\top.
\end{aligned} \tag{A.6}$$

for $(t, s) \in [T] \times [T]$. Then the other parameters are initialized as above. We can even take an ensemble approach in which we fit LDFA-H on different initialized values and pick the estimate with the minimum cost function (Eq. (9)).

Now, for $r = 1, 2, \dots$, we alternate an E-step and an M-step until the target parameter Π_f converges.

E-step Given $\hat{\theta} := \hat{\theta}^{(r-1)}$ from the previous iteration, the conditional distribution of latent factors $Z^1[n]$ and $Z^2[n]$ with respect to observed data $X^1[n]$ and $X^2[n]$ on trial $n = 1, \dots, N$ follows

$$(Z_{1,:}^1[n]; Z_{1,:}^2[n]; \dots; Z_{q,:}^2[n]) \mid X^1[n], X^2[n] \sim \text{MVN} \left(m_{\bar{Z}|X}^{(r)}[n], V_{\bar{Z}|X}^{(r)} \right), \tag{A.7}$$

where

$$V_{\bar{Z}|X}^{(r)} = \begin{pmatrix} V_{Z_1, Z_1|X}^{(r)} & \cdots & V_{Z_1, Z_q|X}^{(r)} \\ \vdots & \ddots & \vdots \\ V_{Z_q, Z_1|X}^{(r)} & \cdots & V_{Z_q, Z_q|X}^{(r)} \end{pmatrix} = \begin{pmatrix} W_{Z_1, Z_1|X}^{(r)} & \cdots & W_{Z_1, Z_q|X}^{(r)} \\ \vdots & \ddots & \vdots \\ W_{Z_q, Z_1|X}^{(r)} & \cdots & W_{Z_q, Z_q|X}^{(r)} \end{pmatrix}^{-1} \tag{A.8}$$

and

$$\begin{aligned}
m_{\bar{Z}|X}^{(r)}[n] &= \left(m_{Z_1^1|X}^{(r)}; m_{Z_2^1|X}^{(r)}; \dots; m_{Z_q^2|X}^{(r)} \right) \\
&= V_{\bar{Z}|X}^{(r)} \left(\hat{\beta}_1^\top \hat{\Gamma}_S^1 X^1[n] \hat{\Gamma}_T^1; \hat{\beta}_1^{2\top} \hat{\Gamma}_S^2 X^2[n] \hat{\Gamma}_T^2; \dots; \hat{\beta}_q^{2\top} \hat{\Gamma}_S^2 X^2[n] \hat{\Gamma}_T^2 \right)
\end{aligned} \tag{A.9}$$

given

$$W_{Z_f, Z_g|X}^{(r)} = \begin{pmatrix} (\hat{\beta}_f^{1\top} \hat{\Gamma}_S^1 \hat{\beta}_g^1) \hat{\Gamma}_T^1 & 0 \\ 0 & (\hat{\beta}_f^{2\top} \hat{\Gamma}_S^2 \hat{\beta}_g^2) \hat{\Gamma}_T^2 \end{pmatrix} + \mathbb{I}_{\{f=g\}} \hat{\Omega}_f, \quad \mathbb{I}_{\{f=g\}} = \begin{cases} 1, & f = g \\ 0, & \text{o.w.} \end{cases} \tag{A.10}$$

for $f, g = 1, \dots, q$.

M-step We find $\hat{\theta}^{(r)}$ which maximize the conditional expectation of the penalized likelihood under the same constraints in Eq. (9), i.e.

$$\begin{aligned}
\hat{\theta}^{(r)} &= \underset{\theta}{\text{argmin}} \frac{1}{N} \sum_{n=1}^N \mathbb{E}_{Z[n]|X[n], \hat{\theta}^{(r-1)}} \left[\log p(X^1[n], X^2[n], Z^1[n], Z^2[n]; \hat{\theta}^{(r-1)}) \right] \\
&\quad + \sum_{f=1}^q \sum_{k,l=1}^2 \|\Lambda_f^{kl} \odot \Pi_f^{kl}\|_1 \quad \text{s.t. } \hat{\Gamma}_T^k \text{ is } (2h_\epsilon^k + 1)\text{-diagonal}
\end{aligned} \tag{A.11}$$

where p is the probability density function of our model in Eqs. (1), (4) and (5) and the expectation $\mathbb{E}_{Z[n]|X[n], \hat{\theta}^{(r-1)}}$ follows the conditional distribution in Eq. (A.7). Taking a block coordinate descent approach, we solve the optimization problem by alternating M1 - M4.

M1: With respect to latent precision matrices Ω_f , Eq. (A.11) reduces to a graphical Lasso problem,

$$\hat{\Omega}_f^{(r)} = \underset{\Omega_f}{\text{argmin}} \left\{ -\log \det(\Omega_f) + \text{tr} \left(\Omega_f \left(V_{Z_f|X}^{(r)} + \hat{\mathbb{E}}[m_{Z_f|X}^{(r)} m_{Z_f|X}^{(r)\top}] \right) \right) + \sum_{k,l=1}^2 \|\Lambda_f^{kl} \odot \Pi_f^{kl}\|_1 \right\} \tag{A.12}$$

for each $f = 1, \dots, q$ where $\widehat{\mathbb{E}}[m_{Z_f|X}^{(r)} m_{Z_f|X}^{(r)\top}] = \frac{1}{N} \sum_{n=1}^N m_{Z_f|X}^{(r)}[n] m_{Z_f|X}^{(r)\top}[n]$. The graphical Lasso problem is solved by the P-GLASSO algorithm by Mazumder et al. (2010).

M2: With respect to Γ^k , Eq. (A.11) reduces to an estimation of matrix-variate normal model (Zhou, 2014). The estimation problem can be formulated as

$$\widehat{\Gamma}_{\mathcal{S}}^{k(r)} = \frac{1}{T} \left(\widehat{\mathbb{E}} \left[m_{\epsilon^k|X}^{(r)} m_{\epsilon^k|X}^{(r)\top} \right] + \sum_{f,g=1}^q \text{tr}(V_{Z_f^k, Z_g^k|X}^{(r)}) \beta_f^k \beta_g^{k\top} \right) \quad (\text{A.13})$$

and

$$\widehat{\Gamma}_{\mathcal{T}}^{k(r)} = \underset{\Gamma_{\mathcal{T}}^k}{\text{argmin}} \left\{ \begin{array}{l} -\log \det(\Gamma_{\mathcal{T}}^k) \\ + \frac{1}{p_k} \text{tr} \left(\Gamma_{\mathcal{T}}^k \left(\sum_{f,g=1}^q (\beta_f^{k\top} \Gamma_{\mathcal{S}}^k \beta_g^k) V_{Z_f^k, Z_g^k|X}^{(r)} + \widehat{\mathbb{E}} \left[m_{\epsilon^k|X}^{(r)\top} \Gamma_{\mathcal{S}}^k m_{\epsilon^k|X}^{(r)} \right] \right) \right) \end{array} \right\} \quad (\text{A.14})$$

s.t. $\widehat{\Gamma}_{\mathcal{T}}^k$ is $(2h_{\epsilon}^k + 1)$ -diagonal

for each $k = 1, 2$ where $m_{\epsilon^k|X}^{(r)} = X^k - \beta^k m_{Z^k|X}^{(r)} - \mu^k$ and $\widehat{\mathbb{E}}[A]$ is the empirical mean of a random matrix A . The estimation of $\Gamma_{\mathcal{T}}^k$ under the bandedness constraint is tractable with modified Cholesky factor decomposition approach with bandwidth h_{ϵ}^k using the procedure by Bickel and Levina (2008).

M3: With respect to β^k , Eq. (A.11) reduces to a quadratic program

$$\widehat{\beta}^{k(r)} = \underset{\beta^k}{\text{arg max}} \left\{ \begin{array}{l} \sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k \text{tr} \left(\beta^{k\top} \Gamma_{\mathcal{S}}^k \beta^k (V_{Z_{:,t}^k, Z_{:,s}^k|X}^{(r)} + \widehat{\text{Cov}}[m_{Z_{:,t}^k|X}^{(r)}, m_{Z_{:,s}^k|X}^{(r)}]) \right) \\ - 2 \sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k \text{tr} \left(\Gamma_{\mathcal{S}}^k \beta^k \widehat{\text{Cov}}[X_{:,t}^k, m_{Z_{:,s}^k|X}^{(r)}] \right) \end{array} \right\} \quad (\text{A.15})$$

where $\Gamma_{\mathcal{T},(t,s)}^k$ is the (t, s) entry in $\Gamma_{\mathcal{T}}^k$ and $\widehat{\text{Cov}}(A, B)$ is the empirical covariance matrix between random vectors A and B . The analytic form of the solution is given by

$$\beta^k = \left(\sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k (V_{Z_{:,t}^k, Z_{:,s}^k|X}^{(r)} + \widehat{\text{Cov}}[m_{Z_{:,t}^k|X}^{(r)}, m_{Z_{:,s}^k|X}^{(r)}]) \right)^{-1} \left(\sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k \widehat{\text{Cov}}[m_{Z_{:,s}^k|X}^{(r)}, X_{:,t}^k] \right) \quad (\text{A.16})$$

M4: With respect to μ^k , it is straight-forward that Eq. (A.11) yields

$$\widehat{\mu}^{k(r)} = \widehat{\mathbb{E}} \left[X^k - \sum_{f=1}^q \beta_f^k m_{Z_f^k|X}^{(r)\top} \right].$$

B Simulation details (Section 3)

We simulated realistic data with known cross-region connectivity as follows. Simulating $q = 1$ pair of latent time-series Z^k from Equation (2), we introduced an exact ground-truth for the inverse cross-correlation matrix Π_1^{12} by setting:

$$\Pi_1 = \begin{bmatrix} (P_{1,0}^{11})^{-1} & 0 \\ 0 & (P_{1,0}^{22})^{-1} \end{bmatrix} + \begin{bmatrix} D^1 & \Pi_1^{12} \\ \Pi_1^{12\top} & D^2 \end{bmatrix} \quad (\text{B.1})$$

where D^1 and D^2 are diagonal matrices with elements $D_{(t,t)}^1 = \sum_s \Pi_{1,(t,s)}^{12}$ and $D_{(s,s)}^2 = \sum_t \Pi_{1,(t,s)}^{12}$, which ensures that the matrix on the right hand side is positive definite. The matrix on the left hand side contains the auto-precision matrices of the two latent time series, with elements simulated from the squared exponential function:

$$P_{1,0}^{kk} = [\exp(-c^k(t-s)^2)]_{t,s} + \lambda I_T, \quad (\text{B.2})$$

with $c^1 = 0.105$ and $c^2 = 0.142$, chosen to match the observed LFPs auto-correlations in the experimental dataset (Section 3.2). We added the regularizer λI_T , $\lambda = 1$, to render P^{kk} invertible.

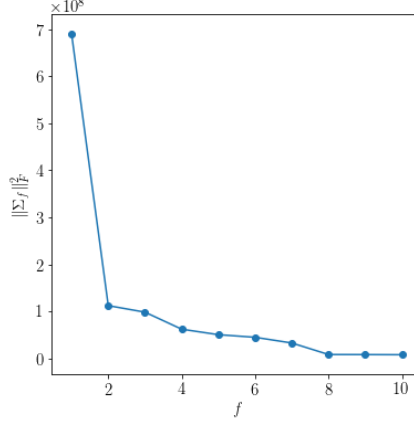


Figure C.1: Squared Frobenius norms of covariance matrix estimates, $\hat{\Sigma}_f$, for all factors $f = 1, \dots, 10$. Notice that the amplitudes of the top four factors dominate the others.

We designed the true inverse cross-correlation matrix Π^{12} to induce lead-lag relationship between Z^1 and Z^2 in two epochs as depicted in the right-most panel of Fig. 2a. Specifically, the elements of Π^{12} were set:

$$\Pi_{(t,s)}^{12} = \begin{cases} -r, & \text{where } Z_{1,t}^1 \text{ and } Z_{1,s}^2 \text{ partially correlate,} \\ 0, & \text{elsewhere,} \end{cases} \quad (\text{B.3})$$

where the association intensity $r = 0.6$ was chosen to match our cross-correlation estimate in the experimental data (Section 3.2). Finally, we rescaled $P_1 = \Pi_1^{-1}$ to have diagonal elements equal to one. The corresponding factor loading vector β_1^k was randomly generated from standard multivariate normal distribution and then scaled to have $\|\beta_1^k\|_2 = 1$.

We generated the noise ϵ^k from the $N = 1000$ trials of the experimental data analyzed in Section 3.2. First, we permuted the trials in one region to remove cross-region correlations. Let $\{Y^1[n], Y^2[n]\}_{n=1, \dots, N}$ be the permuted dataset. Then we contaminated the dataset with white noise to modulate the strength of noise correlation relative to cross-region correlations. i.e.

$$\epsilon_{:,t}^k = Y_{:,t}^k - \mu_{:,t}^k + \eta_{:,t}^k, \quad \eta_{:,t}^k \stackrel{\text{indep}}{\sim} \text{MVN}\left(0, \lambda_\epsilon \widehat{\text{Cov}}[Y_{:,t}^k]\right), \quad \text{and } \mu_{:,t}^k = \widehat{\mathbb{E}}[Y_{:,t}^k] \quad (\text{B.4})$$

where $\widehat{\mathbb{E}}[Y_{:,t}^k]$ and $\widehat{\text{Cov}}[Y_{:,t}^k]$ were the empirical mean and covariance matrix of $Y_{:,t}^k$, respectively, for $k = 1, 2, t = 1, \dots, T$. The noise auto-correlation level was modulated by $\lambda_\epsilon \in \{2.78, 1.78, 0.44, 0.11\}$. We also obtained Σ_1 by scaling P_1 so that $\Sigma_{1,(t,s)}^{kk} = \beta_1^{k\top} S_t^k \beta_1^k$. Putting all the pieces together, we generated observed time series by Eq. (1).

C Experimental data analysis details (Section 3.2)

The strength of each factor, which is characterized by Σ_f , is shown in Fig. C.1.

We also examined an alternative definition of information flow, using non-stationary regression in the spirit of Granger causality. For the latent factor f in V4 at time t , we use partial R^2 , effectively comparing the full regression model using the full history of latent variables in both area,

$$Z_{f,t}^1 \sim Z_{f,1:t-1}^1 + Z_{f,1:t-1}^2$$

with the reduced model using history of latent variables in V4 only,

$$Z_{f,t}^1 \sim Z_{f,1:t-1}^1.$$

The partial R^2 for $Z_{f,t}^1$ on $Z_{f,1:t-1}^2$ given $Z_{f,1:t-1}^1$ summarizes the contribution of PFC history to V4, after taking account of the autocorrelation in V4, and thus can be viewed as information flow from V4 to PFC at time t . Dynamic information flow from V4 to PFC is defined similarly. The results shown in Fig. C.2 are consistent with those in Fig. 5d.

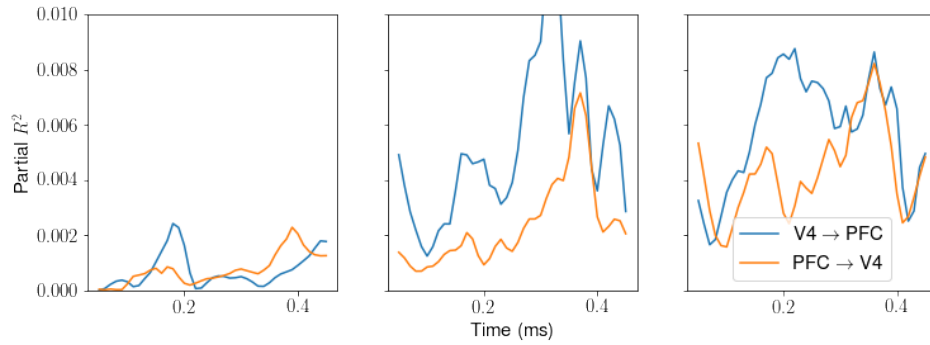


Figure C.2: **Information flow by partial R^2 for the top three factors.** In this figure, we characterize dynamic information flow in terms of partial R^2 . We show dynamic information flow from V4 \rightarrow PFC (blue) and PFC \rightarrow V4 (orange). The results in the first panel are consistent with those in the first panel of Fig. 5d.