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

Initialization  Let $\hat{\theta}^{(0)} = \{\hat{\Sigma}_1^{(0)}, \ldots, \hat{\Sigma}_q^{(0)}, \hat{\Phi}_S^{(0)}, \hat{\Phi}_T^{(0)}, \hat{\Phi}_S^{(1,0)}, \hat{\Phi}_T^{(1,0)}, \hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \hat{\mu}_1^{(0)}, \hat{\mu}_2^{(0)}\}$  be the initial parameter value. Since the MPLE objective function for LDFA-H given in Eq. (7) 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 $\{X^1[n], X^2[n]\}_{n=1, \ldots, N}$ be $N$ simultaneously recorded pairs of neural time series. We can view them as $NT$ recorded pairs of multivariate random vectors $\{X^1_{n:t}[n], X^2_{n:t}[n]\}_{(n,t)\in[N] \times [T]}$. We obtain $\hat{\beta}_1^{(0)}$ and $\hat{\beta}_2^{(0)}$ by CCA as follows:

$$\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)} = \arg\max_{\beta_1 \in \mathbb{R}^p, \beta_2 \in \mathbb{R}^q} \frac{\beta_1^{T} S_{12} \beta_2}{\sqrt{\beta_1^{T} S_{11} \beta_1} \sqrt{\beta_2^{T} S_{22} \beta_2}}$$

(A.1)

where

$$S_{11} = \frac{1}{NT} \sum_{n,t} (X^1_{n:t}[n] - \frac{1}{NT} \sum_{n,t} X^1_{n:t}[n])(X^1_{n:t}[n] - \frac{1}{NT} \sum_{n,t} X^1_{n:t}[n])^{\top}$$

$$S_{22} = \frac{1}{NT} \sum_{n,t} (X^2_{n:t}[n] - \frac{1}{NT} \sum_{n,t} X^2_{n:t}[n])(X^2_{n:t}[n] - \frac{1}{NT} \sum_{n,t} X^2_{n:t}[n])^{\top}$$

$$S_{12} = \frac{1}{NT} \sum_{n,t} (X^1_{n:t}[n] - \frac{1}{NT} \sum_{n,t} X^1_{n:t}[n])(X^2_{n:t}[n] - \frac{1}{NT} \sum_{n,t} X^2_{n:t}[n])^{\top}$$

(A.2)

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

$$\hat{Z}_1^{k,(0)}[n] = \hat{\beta}_1^{k,(0)} X^k[n]$$

(A.3)

for $n = 1, \ldots, N$ 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 $f = 1, \ldots, q$ and the matrix-variate normal estimate on $\{\hat{Z}^{k,(0)}[n] := X^k[n] - \hat{\beta}^{k,(0)} \hat{Z}^{k,(0)}[n]\}_{n\in[N]}$ 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 $\hat{\Phi}_T^{(k,0)}$ is not invertible. We instead pick one of its unidentifiable parameter sets $\hat{\theta}^{(0),(a', a^2)}$, defined in Eq. (8), with all $\hat{\Phi}_T^{(k,0)}$'s and $\hat{\Sigma}_f^{(0)}$'s invertible. Specifically, we take

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

(A.4)

for $f = 1, \ldots, q$ and $k = 1, 2$ where $\lambda_{\text{min}}(A)$ is the smallest eigenvalue of symmetric matrix $A$. Henceforth, we note $\hat{\theta}^{(0),(a', a^2)}$ by $\hat{\theta}^{(0)}$. For $t = 1, 2, \ldots$, 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 $X_{1,t}^1$ and $X_{2,s}^2$ maximizes. i.e., we initialize $\hat{\beta}_1^{1,(0)}$ and $\hat{\beta}_2^{2,(0)}$ by

$$\hat{\beta}_1^{1,(0)}, \hat{\beta}_1^{2,(0)} = \arg\max_{\beta_1 \in \mathbb{R}^p, \beta_2 \in \mathbb{R}^q} \frac{\beta_1^{T} S_{12}^{(t,s)} \beta_2}{\sqrt{\beta_1^{T} S_{11}^{(t,s)} \beta_1} \sqrt{\beta_2^{T} S_{22}^{(s,s)} \beta_2}}$$

such that $|t - s| < h_{\text{cross}}$. (A.5)
where
\[
S_{t,t}^{11} = \frac{1}{N} \sum_{n,t} (X_{t,t}^1[n] - \frac{1}{N} \sum_n X_{t,t}^1[n]) (X_{t,t}^1[n] - \frac{1}{N} \sum_n X_{t,t}^1[n])^\top
\]
\[
S_{s,s}^{22} = \frac{1}{N} \sum_{n,s} (X_{s,s}^2[n] - \frac{1}{N} \sum_n X_{s,s}^2[n]) (X_{s,s}^2[n] - \frac{1}{N} \sum_n X_{s,s}^2[n])^\top
\]
\[
S_{l,t}^{12} = \frac{1}{N} \sum_{n,t} (X_{t,t}^1[n] - \frac{1}{N} \sum_n X_{t,t}^1[n]) (X_{s,s}^2[n] - \frac{1}{N} \sum_n X_{s,s}^2[n])^\top
\]
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 \(r = 1, 2, \ldots\), we alternate an E-step and an M-step until the target parameter \(\Pi_f\) convergences.

**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 \(n = 1, \ldots, N\) follows
\[
(Z^1_1[n]; Z^1_2[n]; \ldots; Z^2_q[n]) \mid X^1[n], X^2[n] \sim \text{MVN} \left( m^{(r)}_{Z|X}[n], V^{(r)}_{Z|X}[n] \right),
\]
where
\[
V^{(r)}_{Z|X} = \begin{pmatrix} V^{(r)}_{Z_1, Z_1|X} & \cdots & V^{(r)}_{Z_1, Z_q|X} \\ \vdots & \ddots & \vdots \\ V^{(r)}_{Z_q, Z_1|X} & \cdots & V^{(r)}_{Z_q, Z_q|X} \end{pmatrix}
\]
\[
m^{(r)}_{Z|X}[n] = \left( m^{(r)}_{Z_1'|X}; m^{(r)}_{Z_2'|X}; \ldots; m^{(r)}_{Z_q'|X} \right)
\]
\[
= V^{(r)}_{Z|X} \begin{pmatrix} \hat{\beta}_f^1 \hat{\Gamma}_S^1 X^1[n] \hat{\Gamma}_T^1 f \\ \vdots \\ \hat{\beta}_q \hat{\Gamma}_S^2 X^2[n] \hat{\Gamma}_T^2 f \end{pmatrix}
\]
given
\[
W^{(r)}_{Z_f, Z_g|X}[r] = \begin{pmatrix} \hat{\beta}_f \hat{\Gamma}_S^1 \hat{\beta}_g \hat{\Gamma}_S^1 \hat{\Gamma}_T^1 + \hat{\Pi}_f, & \hat{\Pi}_f \end{pmatrix} = \begin{cases} 1, & f = g \\ 0, & \text{o.w.} \end{cases}
\]
for \(f, g = 1, \ldots, q\).

**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.
\[
\hat{\theta}^{(r)} = \arg\min \frac{1}{N} \sum_{n=1}^N \mathbb{E}_{Z[n]|X[n], \hat{\theta}^{(r-1)}} \left[ \log p(X^1[n], X^2[n], Z^1[n], Z^2[n]; \hat{\theta}^{(r-1)}) \right] \]
\[
+ \sum_{k, l=1}^2 \sum_{j=1}^2 \left\| \Lambda^k_l \circ \Pi^k_l \right\|_1 \quad \text{s.t.} \quad \hat{\Gamma}_T^k \text{ is } (2h_e^k + 1)\text{-diagonal}
\]
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. (7). Taking a block coordinate descent approach, we solve the optimization problem by alternating M1 - M4.

M1: With respect to latent precision matrices \(\Omega_f\), Eq. (A.11) reduces to a graphical Lasso problem,
\[
\hat{\Omega}_f^{(r)} = \arg\min_{\Omega_f} \left\{ -\log \det(\Omega_f) + \text{tr} \left( \Omega_f \left( V^{(r)}_{Z_f|X} + \mathbb{E}[m^{(r)}_{Z_f|X} m^{(r)}_{Z_f|X}^\top] \right) \right) + \sum_{k, l=1}^2 \left\| \Lambda^k_l \circ \Pi^k_l \right\|_1 \right\}
\]
for each $f = 1, \ldots, q$ where $\hat{E}[m^{(r)}_{Z_{ij}^k} m^{(r)\top}_{Z_{js}^k}] = \frac{1}{N} \sum_{n=1}^{N} m^{(r)}_{Z_{ij}^k[n]} m^{(r)\top}_{Z_{js}^k[n]}$. The graphical Lasso problem is solved by the P-GLASSO algorithm by Mazumder et al. (2010).

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

$$\hat{\Gamma}^{k(r)}_S = \frac{1}{T} \left( \hat{E} \left[ m^{(r)}_{e^k|x} m^{(r)\top}_{e^k|x} \right] + \sum_{f,g=1}^{q} \text{tr} \left( V^{(r)}_{z_f^k z_g^k|x} \beta_k^{(r)} \beta_k^{(r)}\top \right) \right) \quad \text{(A.13)}$$

and

$$\hat{\Gamma}^{k(r)}_T = \arg \min_{\Gamma_T} \left\{ -\log \det(\Gamma_T^k) + \frac{1}{p_k} \text{tr} \left( \Gamma_T^k \left( \sum_{f,g=1}^{q} (\beta_f^k \Gamma_S^k \beta_g^k) V^{(r)}_{z_f^k z_g^k|x} + \hat{E} \left[ m^{(r)\top}_{e^k|x} \Gamma_S^k m^{(r)}_{e^k|x} \right] \right) \right) \right\} \quad \text{s.t. } \Gamma_T^k \text{ is } (2h_k^k + 1)\text{-diagonal} \quad \text{(A.14)}$$

for each $k = 1,2$ where $m^{(r)}_{e^k|x} = X^k - \beta^k m^{(r)}_{2^k|x} - \mu^k$ and $\hat{E}[A]$ is the empirical mean of a random matrix $A$. The estimation of $\hat{\Gamma}^k_T$ under the bandedness constraint is tractable with modified Cholesky factor decomposition approach with bandwidth $h^k$ using the procedure by Bickel and Levina (2008).

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

$$\hat{\beta}^{k(r)} = \arg \max_{\beta_k} \left\{ \sum_{t,s} \Gamma^k_{T,(t,s)} \text{tr} \left( \beta^k \Gamma_S^k \beta_k \left( V^{(r)}_{z_t^k z_s^k|x} + \hat{E} \left[ m^{(r)}_{Z^k_{ts}|x} m^{(r)}_{Z^k_{ts}|x} \right] \right) \right) \right\}$$

$$- 2 \sum_{t,s} \Gamma^k_{T,(t,s)} \text{tr} \left( \Gamma_S^k \beta_k \hat{E} \left[ m^{(r)}_{Z^k_{ts}|x} m^{(r)}_{Z^k_{ts}|x} \right] \right) \quad \text{(A.15)}$$

where $\Gamma^k_{T,(t,s)}$ is the $(t, s)$ entry in $\Gamma^k_T$ and $\hat{E}(A, B)$ is the empirical covariance matrix between random vectors $A$ and $B$. The analytic form of the solution is given by

$$\beta^k = \left( \sum_{t,s} \Gamma^k_{T,(t,s)} \left( V^{(r)}_{z_t^k z_s^k|x} + \hat{E} \left[ m^{(r)}_{Z^k_{ts}|x} m^{(r)}_{Z^k_{ts}|x} \right] \right) \right)^{-1} \left( \sum_{t,s} \Gamma^k_{T,(t,s)} \hat{E} \left[ m^{(r)}_{Z^k_{ts}|x} m^{(r)}_{Z^k_{ts}|x} \right] \right)$$

$$\text{(A.16)}$$

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

$$\hat{\mu}^{k(r)} = \hat{E} \left[ X^k - \sum_{f=1}^{q} \beta_f^k m^{(r)}_{Z_f^k|x} \right] .$$

### 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 $\Pi^{12}$ by setting:

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

where $D^{1}$ and $D^{2}$ are diagonal matrices with elements $D^{1}_{(t,t)} = \sum_s \Pi^{12}_{1,ts}$ and $D^{2}_{(s,s)} = \sum_t \Pi^{12}_{ts,1}$, 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:

$$P^{kk}_{1,0} = \exp \left( -e^k (t - s)^2 \right)_{t,s} + \lambda I_T, \quad \text{(B.2)}$$

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 $\lambda I_T$, $\lambda = 1$, to render $P^{kk}$ invertible.
Figure C.1: Squared Frobenius norms of covariance matrix estimates, $\hat{\Sigma}_f$, for all factors $f = 1, \ldots, 10$. Notice that the amplitudes of the top four factors dominate the others.

We designed the true inverse cross-correlation matrix $\Pi_{12}$ 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 $\Pi_{12}$ were set:

$$\Pi_{12}(t,s) = \begin{cases} -r, & \text{where } Z_1^1, t \text{ and } Z_2^1, s \text{ partially correlate,} \\ 0, & \text{elsewhere} \end{cases}$$

(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 $\beta_k^1$ was randomly generated from standard multivariate normal distribution and then scaled to have $\|\beta_k^1\|_2 = 1$.

We generated the noise $\epsilon^k$ 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 $\{Y_1[n], Y_2[n]\}_{n=1,\ldots,N}$ 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.

$$\epsilon^k_{i,t} = Y^k_{i,t} - \mu^k_{i,t} + \eta^k_{i,t} \text{ iid } \text{MVN}(0, \lambda_{\epsilon} \text{Cov}[Y^k_{i,t}]), \text{ and } \mu^k_{i,t} = \hat{E}[Y^k_{i,t}]$$

(B.4)

where $\hat{E}[Y^k_{i,t}]$ and $\hat{\text{Cov}}[Y^k_{i,t}]$ wer the empirical mean and covariance matrix of $Y^k_{i,t}$, respectively, for $k = 1, 2, t = 1, \ldots, T$. The noise auto-correlation level was modulated by $\lambda_{\epsilon} \in \{2.78, 1.78, 0.44, 0.11\}$. We also obtained $\Sigma_1$ by scaling $P_1$ so that $\Sigma_{kk} = \beta_k^1 S_k^k \beta_k^1$. 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 $\Sigma_f$, 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, $Z_{1,1:t-1}^f$, with the reduced model using history of latent variables in V4 only, $Z_{1,1:t-1}^f$.

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

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

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

The partial $R^2$ for $Z_{1,1:t}^f$ on $Z_{2,1:t-1}^f$ given $Z_{1,1:t-1}^f$ 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.