Bayesian Structure Learning for Dynamic Brain Connectivity: Supplementary material

Michael Riis Andersen, Ole Winther, Lars Kai Hansen, Russell Poldrack, & Oluwasanmi Koyejo

The purpose of this supplementary document is to provide further details for the paper "Bayesian Structure Learning for Dynamic Brain Connectivity".

Approximate Inference

Since $\boldsymbol{S}_k = \boldsymbol{v}_k \boldsymbol{v}_k^T$ is of rank one, we can rewrite eq. (5) in the paper as $\boldsymbol{\Sigma}_t = \beta^{-1}\boldsymbol{I} + \boldsymbol{V}\boldsymbol{A}_t\boldsymbol{V}^T$, where $\boldsymbol{V} = \begin{bmatrix}\boldsymbol{v}_1\, \boldsymbol{v}_2\, \dots\, \boldsymbol{v}_K\end{bmatrix}$ and $\boldsymbol{A}_t = \text{diag}\left(\alpha_{1,t}\, \alpha_{2,t}\, \dots\, \alpha_{K,t}\right)$. This form is recognized as the marginalized covariance matrix of a factor model with Gaussian latent variables, i.e. $\boldsymbol{x}_t^n \sim \mathcal{N}\left(\boldsymbol{V}\boldsymbol{z}_t^n, \beta^{-1}\boldsymbol{I}\right)$ with $\boldsymbol{z}_t^n \sim \mathcal{N}\left(0, \boldsymbol{A}_t\right)$ and using the factor model form allows us to derive a more efficient inference scheme. However, the exact posterior distribution of the parameters of interest is intractable and hence, we resort to approximate inference using a mean-field approximation. We use the family of Gaussian distributions to approximate the posterior distributions over $a_{k,t}$ and $z_{k,t}^n$ and we use the family of approximate distributions described in (Lázaro-gredilla and Titsias, 2011) to approximate the posterior of the spike and slab variables $\boldsymbol{V}$. That is,

$$Q(\boldsymbol{a}_k) = \prod_{t=1}^{T} \mathcal{N}\left(a_{k,t}|\hat{\gamma}_{k,t}, \hat{\lambda}_{k,t}\right), \qquad Q(\boldsymbol{z}_k^n) = \prod_{t=1}^{T} \mathcal{N}\left(z_{k,t}^n|\hat{\eta}_{k,t}^n, \hat{\theta}_{k,t}^n\right) \tag{1}$$

$$Q(\boldsymbol{u}_k, \boldsymbol{s}_k) = \prod_{i=1}^{D} \mathcal{N}\left(u_{i,k}|s_{i,k}\hat{\mu}_{i,k}, s_{i,k}\hat{\tau}_{i,k} + (1-s_{i,k})\tau_k\right) \prod_{i=1}^{D} \text{Bernoulli}(s_{i,k}|\hat{\pi}_{i,k}), \tag{2}$$

where $\boldsymbol{z}_k^n = (z_{k,1}^n, \dots, z_{k,T}^n) \in \mathbb{R}^T$. We minimize the KL-divergence, $\text{KL}\left[Q||P\right]$, between the approximation Q and the exact posterior distribution P by optimizing the Evidence Lower Bound (ELBO) (Blei et al., 2016) w.r.t. to the variational parameters.

Posterior expectation of the covariance components $\boldsymbol{S}_k$

The purpose of this section is to compute the expectation of the covariance components $\boldsymbol{S}_k = \boldsymbol{v}_k\boldsymbol{v}_k^T$ wrt. the variational approximation Q. Recall, the approximate posterior distribution of $\boldsymbol{v}_k = \boldsymbol{s}_k \circ \boldsymbol{u}_k$ is given by

$$\begin{aligned} Q(\boldsymbol{u}_k|\boldsymbol{s}_k) &= \prod_{i=1}^{D} \mathcal{N}\left(u_{i,k}|s_{i,k}\hat{\mu}_{i,k},\ s_{i,k}\hat{\tau}_{i,k} + (1-s_{i,k})\tau_k\right) \\ Q(\boldsymbol{s}_k) &= \prod_{i=1}^{D} \text{Bernoulli}(s_{i,k}|\hat{\pi}_{i,k}). \end{aligned} \tag{3}$$

Let $\bar{\boldsymbol{v}}_k$ be the mean of $\boldsymbol{v}_k$ wrt. Q, i.e.

$$\bar{\boldsymbol{v}}_k = \mathbb{E}_Q\left[\boldsymbol{v}_k\right] = \mathbb{E}_Q\left[\boldsymbol{s}_k \circ \boldsymbol{u}_k\right] = \hat{\boldsymbol{\pi}}_k \circ \hat{\boldsymbol{\mu}}_k. \tag{4}$$

Consider first the diagonal elements of $\mathbb{E}_Q\left[\boldsymbol{S}_k\right]$

$$\begin{aligned} \mathbb{E}_Q\left[\boldsymbol{S}_k\right]_{ii} &= \mathbb{E}_Q\left[\boldsymbol{v}_k\boldsymbol{v}_k^T\right]_{ii} && (5)\\ &= \mathbb{E}_Q\left[v_{i,k}v_{i,k}\right] && (6)\\ &= \mathbb{E}_Q\left[s_{i,k}^2 u_{i,k}^2\right] && (7)\\ &= \sum_{s_{i,k}} \int s_{i,k}^2 u_{i,k}^2 \mathcal{N}\left(u_{i,k}|s_{i,k}\hat{\mu}_{i,k}, s_{i,k}\hat{\tau}_{i,k} + (1-s_{i,k})\tau_k\right) \text{Bernoulli}(s_{i,k}|\hat{\pi}_{i,k}) \mathrm{d}u_{i,k} && (8)\\ &= \hat{\pi}_{i,k} \int u_{i,k}^2 \mathcal{N}\left(u_{i,k}|\hat{\mu}_{i,k}, \hat{\tau}_{i,k}\right) \mathrm{d}u_{i,k} && (9)\\ &= \hat{\pi}_{i,k}\left(\hat{\mu}_{i,k}^2 + \hat{\tau}_{i,k}\right). && (10) \end{aligned}$$

Next, we compute the off-diagonal elements

$$\mathbb{E}_Q\left[\boldsymbol{S}_k\right]_{ij} = \mathbb{E}_Q\left[\boldsymbol{v}_k\boldsymbol{v}_k^T\right]_{ij} = \mathbb{E}_Q\left[v_{i,k}v_{j,k}\right] = \mathbb{E}_Q\left[v_{i,k}\right]\mathbb{E}_Q\left[v_{j,k}\right] = \hat{\pi}_{i,k}\hat{\mu}_{i,k}\hat{\pi}_{j,k}\hat{\mu}_{j,k} \quad \text{for} \quad i \neq j. \tag{11}$$

Thus, the posterior expectation becomes

$$\hat{\boldsymbol{\Sigma}}_t = \bar{\boldsymbol{v}}_k \bar{\boldsymbol{v}}_k^T + \operatorname{diag}\left(\bar{\boldsymbol{v}}_k \circ \hat{\boldsymbol{\mu}}_k + \hat{\boldsymbol{\pi}}_k \circ \hat{\boldsymbol{\tau}}_k - \bar{\boldsymbol{v}}_k \circ \bar{\boldsymbol{v}}_k\right). \tag{12}$$

Posterior expectation of the mixing weights

The purpose of this section is to compute the posterior expectation of the mixing weights defined as $\alpha_{k,t} = \max(0, a_{k,t})$. Recall, the posterior distribution of $a_{k,t}$ is given by

$$Q(\boldsymbol{a}_k) = \prod_{t=1}^{T} \mathcal{N}\left(a_{k,t} | \hat{\gamma}_{k,t}, \hat{\lambda}_{k,t}\right). \tag{13}$$

Thus,

$$\begin{aligned}
\mathbb{E}_Q\left[\alpha_{k,t}\right] &= \mathbb{E}_Q\left[\max(0, a_{k,t})\right] && (14)\\
&= \int_{\mathbb{R}} \max(0, a_{k,t})\, \mathcal{N}\left(a_{k,t} | \hat{\gamma}_{k,t}, \hat{\lambda}_{k,t}\right) \mathrm{d}a_{k,t} && (15)\\
&= \int_0^\infty a_{k,t}\, \mathcal{N}\left(a_{k,t} | \hat{\gamma}_{k,t}, \hat{\lambda}_{k,t}\right) \mathrm{d}a_{k,t} && (16)\\
&\overset{(a)}{=} b \int_0^\infty x\, \mathcal{N}\left(a + bx | 0, 1\right) \mathrm{d}x \quad \text{for} \quad a = -\frac{\hat{\gamma}_{k,t}}{\sqrt{\hat{\lambda}_{k,t}}}, b = \frac{1}{\sqrt{\hat{\lambda}_{k,t}}} && (17)\\
&\overset{(b)}{=} \frac{1}{b}\left[\phi(a) + a\Phi(a) - a\right] && (18)
\end{aligned}$$

where (a) follows from the change of variable $x = \hat{\gamma}_{k,t} + \sqrt{\hat{\lambda}_{k,t}}\, a_{k,t}$ and (b) is the result of a standard Gaussian integral (Owen, 1980). The second moment evaluates to

$$\begin{aligned}
\mathbb{E}_Q\left[\alpha_{k,t}^2\right] &= \mathbb{E}_Q\left[\max(0, a_{k,t})^2\right] && (19)\\
&= \int_{\mathbb{R}} \max(0, a_{k,t})^2\, \mathcal{N}\left(a_{k,t} | \hat{\gamma}_{k,t}, \hat{\lambda}_{k,t}\right) \mathrm{d}a_{k,t} && (20)\\
&= \int_0^\infty a_{k,t}^2\, \mathcal{N}\left(a_{k,t} | \hat{\gamma}_{k,t}, \hat{\lambda}_{k,t}\right) \mathrm{d}a_{k,t} && (21)\\
&= b \int_0^\infty x^2\, \mathcal{N}\left(a + bx | 0, 1\right) \mathrm{d}x && (22)\\
&\overset{(c)}{=} \frac{1}{b^2}\left[(a^2 + 1)(1 - \Phi(a)) - a\phi(a)\right], && (23)
\end{aligned}$$

where (c) is the result of a standard Gaussian integral (Owen, 1980).

Evidence lower bound

The purpose of this section is to derive the evidence lower bound (ELBO) on the marginal likelihood of the model. The joint distribution of the model is

$$p(\mathcal{D}, \boldsymbol{U}, \boldsymbol{S}, \mathcal{A}, \boldsymbol{Z}) = \prod_{n=1}^{N}\prod_{t=1}^{T} p(\boldsymbol{x}_t^n | \boldsymbol{A}_t, \boldsymbol{U}, \boldsymbol{S}, \boldsymbol{z}_t^n) \prod_{k=1}^{K} p(\boldsymbol{a}_k) \prod_{k=1}^{K} p(\boldsymbol{u}_k) \prod_{k=1}^{K} p(\boldsymbol{s}_k) \prod_{n=1}^{N}\prod_{t=1}^{T} p(\boldsymbol{z}_t^n) \tag{24}$$

The variational distribution is given by

$$Q(\boldsymbol{U}, \boldsymbol{S}, \mathcal{A}, \boldsymbol{Z}) = \prod_{k=1}^{K} Q(\boldsymbol{u}_k | \boldsymbol{s}_k) \prod_{k=1}^{K} Q(\boldsymbol{s}_k) \prod_{k=1}^{K} Q(\boldsymbol{a}_k) \prod_{n=1}^{N}\prod_{t=1}^{T} Q(\boldsymbol{z}_t^n) \tag{25}$$

We compute the standard evidence lower bound (Blei et al., 2016; Wainwright and Jordan, 2008) as follows

$$\mathcal{L} = \mathbb{E}_Q\left[\ln p(\mathcal{D}, \boldsymbol{U}, \boldsymbol{S}, \mathcal{A}, \boldsymbol{Z})\right] - \mathbb{E}_Q\left[\ln Q(\boldsymbol{U}, \boldsymbol{S}, \mathcal{A}, \boldsymbol{Z})\right] \tag{26}$$

First, we consider the (negative) entropy of Q, which decomposes as follows

$$\mathbb{E}_Q\left[\ln Q(\boldsymbol{U},\boldsymbol{S},\mathcal{A},\boldsymbol{Z})\right] = \sum_{k=1}^{K}\mathbb{E}_Q\left[\ln Q(\boldsymbol{u}_k|\boldsymbol{s}_k)\right] + \sum_{k=1}^{K}\mathbb{E}_Q\left[\ln Q(\boldsymbol{s}_k)\right] + \sum_{k=1}^{K}\mathbb{E}_Q\left[\ln Q(\boldsymbol{a}_k)\right] + \sum_{n=1}^{N}\sum_{t=1}^{T}\mathbb{E}_Q\left[\ln Q(\boldsymbol{z}_t^n)\right]. \tag{27}$$

We can then compute the (negative) entropy of the individual components

$$\begin{aligned}
\mathbb{E}_Q\left[\ln Q(\boldsymbol{u}_k|\boldsymbol{s}_k)\right] &= \mathbb{E}_Q\left[\ln\prod_{i=1}^{D}Q(u_{i,k}|s_{i,k})\right] = \sum_{i=1}^{D}\mathbb{E}_Q\left[\ln Q(u_{i,k}|s_{i,k})\right]\\
&\overset{(d)}{=} \sum_{i=1}^{D}\left\{(1-\hat{\pi}_{i,k})\,\mathbb{E}_Q\left[\ln Q(u_{i,k}|s_{i,k}=0)\right] + \hat{\pi}_{i,k}\mathbb{E}_Q\left[\ln Q(u_{i,k}|s_{i,k}=1)\right]\right\}\\
&= -\sum_{i=1}^{D}\left\{(1-\hat{\pi}_{i,k})\,\frac{1}{2}\ln\left(2\pi\tau_k\mathrm{e}\right) + \hat{\pi}_{i,k}\frac{1}{2}\ln\left(2\pi\hat{\tau}_k\mathrm{e}\right)\right\}
\end{aligned}\tag{28}$$

where (d) is obtained by taking the expectation wrt. $s_{i,k}$ first. Furthermore,

$$\mathbb{E}_Q\left[\ln Q(\boldsymbol{s}_k)\right] = \mathbb{E}_Q\left[\ln\prod_{i=1}^{D}Q(s_{i,k})\right] = \sum_{i=1}^{D}\mathbb{E}_Q\left[\ln Q(s_{i,k})\right] = \sum_{i=1}^{D}\left[(1-\hat{\pi}_{i,k})\ln\left(1-\hat{\pi}_{i,k}\right) + \hat{\pi}_{i,k}\ln\left(\hat{\pi}_{i,k}\right)\right] \tag{29}$$

$$\mathbb{E}_Q\left[\ln Q(\boldsymbol{a}_k)\right] = \mathbb{E}_Q\left[\ln\prod_{t=1}^{T}Q(a_{k,t})\right] = \sum_{t=1}^{T}\mathbb{E}_Q\left[\ln Q(a_{k,t})\right] = -\sum_{t=1}^{T}\frac{1}{2}\ln\left(2\pi\hat{\lambda}_{k,t}\mathrm{e}\right) \tag{30}$$

$$\mathbb{E}_Q\left[\ln Q(\boldsymbol{z}_t^n)\right] = \mathbb{E}_Q\left[\ln\prod_{k=1}^{K}Q(z_{k,t}^n)\right] = \sum_{k=1}^{K}\mathbb{E}_Q\left[\ln Q(z_{k,t}^n)\right] = -\sum_{k=1}^{K}\frac{1}{2}\ln\left(2\pi\hat{\theta}_{k,t}^n\mathrm{e}\right) \tag{31}$$

Next, we consider the expectation of the logarithm of the joint distribution decomposes, which decomposes as follows

$$\mathbb{E}_Q\left[\ln p(\mathcal{D},\boldsymbol{U},\boldsymbol{S},\mathcal{A},\boldsymbol{Z})\right] = \mathbb{E}_Q\left[\ln\prod_{n=1}^{N}\prod_{t=1}^{T}p(\boldsymbol{x}_t^n|\boldsymbol{A}_t,\boldsymbol{U},\boldsymbol{S},\boldsymbol{z}_t^n)\prod_{k=1}^{K}p(\boldsymbol{a}_k)\prod_{k=1}^{K}p(\boldsymbol{u}_k)\prod_{k=1}^{K}p(\boldsymbol{s}_k)\prod_{n=1}^{N}\prod_{t=1}^{T}p(\boldsymbol{z}_t^n)\right] \tag{32}$$

$$\begin{aligned}
&= \sum_{n=1}^{N}\sum_{t=1}^{T}\mathbb{E}_Q\left[\ln p(\boldsymbol{x}_t^n|\boldsymbol{A}_t,\boldsymbol{U},\boldsymbol{S},\boldsymbol{z}_t^n)\right] + \sum_{k=1}^{K}\mathbb{E}_Q\left[\ln p(\boldsymbol{a}_k)\right] + \sum_{k=1}^{K}\mathbb{E}_Q\left[\ln p(\boldsymbol{u}_k)\right]\\
&\quad + \sum_{k=1}^{K}\mathbb{E}_Q\left[\ln p(\boldsymbol{s}_k)\right] + \sum_{n=1}^{N}\sum_{t=1}^{T}\mathbb{E}_Q\left[\ln p(\boldsymbol{z}_t^n)\right]
\end{aligned}\tag{33}$$

We then compute the expectations of the individual components

$$\begin{aligned}
\mathbb{E}_Q\left[\ln p(\boldsymbol{x}_t^n|\boldsymbol{A}_t,\boldsymbol{U},\boldsymbol{S},\boldsymbol{z}_t^n)\right] &= \mathbb{E}_Q\left[\ln\prod_{i=1}^{D}p(x_{i,t}^n|\boldsymbol{A}_t,\boldsymbol{U},\boldsymbol{S},\boldsymbol{z}_t^n)\right]\\
&= \sum_{i=1}^{D}\mathbb{E}_Q\left[\ln p(x_{i,t}^n|\boldsymbol{A}_t,\boldsymbol{U},\boldsymbol{S},\boldsymbol{z}_t^n)\right]\\
&= \sum_{i=1}^{D}\mathbb{E}_Q\left[\ln\mathcal{N}\left(x_{i,t}^n|\boldsymbol{V}_{i,\cdot}\boldsymbol{A}_t\boldsymbol{z}_t^n,\beta^{-1}\boldsymbol{I}\right)\right]\\
&= \sum_{i=1}^{D}\mathbb{E}_Q\left[\ln\frac{1}{\sqrt{2\pi\beta^{-1}}}\exp\left(-\frac{\beta}{2}\left(x_{i,t}^n-\boldsymbol{V}_{i,\cdot}\boldsymbol{A}_t\boldsymbol{z}_t^n)\right)^2\right)\right]\\
&= -\frac{D}{2}\ln\left(2\pi\beta^{-1}\right) - \sum_{i=1}^{D}\frac{\beta}{2}\mathbb{E}_Q\left[\left(x_{i,t}^n-\boldsymbol{V}_{i,\cdot}\boldsymbol{A}_t\boldsymbol{z}_t^n)\right)^2\right]
\end{aligned}\tag{34}$$

where the expectation of the square evaluates to

$$\mathbb{E}_Q\left[\left(x_{i,t}^n - \boldsymbol{V}_{i,\cdot}\boldsymbol{A}_t z_t^n\right)^2\right] = \left(x_{i,t}^n\right)^2 + \sum_{k=1}^K \left(\hat{\mu}_{i,k}^2 + \hat{\tau}_{i,k}\right)\hat{P}_{k,t}\left((\hat{\eta}_{k,t}^n)^2 + (\hat{\theta}_{k,t}^n)^2\right) + \left(\sum_{k=1}^K \hat{\mu}_{i,k}\hat{\rho}_{k,t}\hat{\eta}_{k,t}^n\right)^2$$
$$- \sum_{k=1}^K \hat{\mu}_{i,k}^2\hat{\rho}_{k,t}^2(\hat{\eta}_{k,t}^n)^2 - 2x_{i,t}^n\sum_{k=1}^K \hat{\mu}_{i,k}\hat{\rho}_{k,t}\hat{\eta}_{k,t}^n \tag{35}$$

for $\rho_{k,t} = \mathbb{E}\left[\max(0, a_{k,t})\right]$ and $P_{k,t} = \mathbb{E}\left[\max(0, a_{k,t})^2\right]$ be the first two moments of α_t.

The expectation of the logarithm of the Gaussian process prior yields

$$\begin{aligned}
\mathbb{E}_Q\left[\ln p(\boldsymbol{a}_k)\right] &= \mathbb{E}_Q\left[\ln\mathcal{N}\left(\boldsymbol{a}_k|\boldsymbol{m}_k, \boldsymbol{C}_k\right)\right] \\
&= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{C}_k| - \frac{1}{2}\mathbb{E}_Q\left[(\boldsymbol{a}_k - \boldsymbol{m}_k)^T\boldsymbol{C}_k^{-1}(\boldsymbol{a}_k - \boldsymbol{m}_k)\right] \\
&= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{C}_k| - \frac{1}{2}\left[(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{m}_k)^T\boldsymbol{C}_k^{-1}(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{m}_k) + \mathrm{Tr}\left(\boldsymbol{C}_k^{-1}\hat{\boldsymbol{\Lambda}}\right)\right].
\end{aligned} \tag{36}$$

The remaining terms yields

$$\begin{aligned}
\mathbb{E}_Q\left[\ln p(\boldsymbol{u}_k)\right] &= \mathbb{E}_Q\left[\ln\prod_{i=1}^K p(u_{i,k})\right] = \sum_{i=1}^K \mathbb{E}_Q\left[\ln p(u_{i,k})\right] = \sum_{i=1}^K \mathbb{E}_Q\left[\ln\mathcal{N}\left(u_{i,k}|0, \tau_k\right)\right] \\
&= -\frac{K}{2}\ln(2\pi\tau_k) - \frac{1}{2\tau_k}\sum_{i=1}^K \mathbb{E}_Q\left[u_{i,k}^2\right] \\
&= -\frac{K}{2}\ln(2\pi\tau_k) - \frac{1}{2\tau_k}\sum_{i=1}^K \left((1 - \hat{\pi}_{i,k})\tau_k + \hat{\pi}_{i,k}(\hat{\mu}_{i,k}^2 + \hat{\tau}_{i,k})\right) \\
\mathbb{E}_Q\left[\ln p(\boldsymbol{s}_k)\right] &= \mathbb{E}_Q\left[\ln\prod_{i=1}^D p(s_{i,k})\right] = \sum_{i=1}^D \mathbb{E}_Q\left[\ln p(s_{i,k})\right] \\
&= \sum_{i=1}^D \mathbb{E}_Q\left[\ln\left((1 - \pi_k)^{(1-s_{i,k})}\pi_k^{s_{i,k}}\right)\right] \\
&= \sum_{i=1}^D (1 - \mathbb{E}_Q\left[s_{i,k}\right])\ln(1 - \pi_k) + \mathbb{E}_Q\left[s_{i,k}\right]\ln\pi_k && (37) \\
&= \ln(1 - \pi_k)\sum_{i=1}^D (1 - \hat{\pi}_{i,k}) + \ln\pi_k\sum_{i=1}^D \hat{\pi}_{i,k} && (38) \\
\mathbb{E}_Q\left[\ln p(\boldsymbol{z}_t^n)\right] &= \mathbb{E}_Q\left[\ln\prod_{k=1}^K \mathcal{N}\left(z_{k,t}^n|0, 1\right)\right] = \sum_{k=1}^K \mathbb{E}_Q\left[\ln\mathcal{N}\left(z_{k,t}^n|0, 1\right)\right] \\
&= -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\sum_{k=1}^K \mathbb{E}_Q\left[(z_{k,t}^n)^2\right] \\
&= -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\sum_{k=1}^K ((\hat{\eta}_{k,t}^n)^2 + \hat{\theta}_{k,t}^n). && (39)
\end{aligned}$$

Log-Euclidean Riemannian Metric

As we are interested in exact recovery of the covariance matrices (and not just likelihood), we use the Log-Euclidean Riemannian Metric (LERM), which defines a metric on the manifold of symmetric positive definite matrices (Vemulapalli and Jacobs, 2015; Huang et al., 2015). Our experiments with simulated data also indicate that LERM outperformed the variational bound of the log likelihood for identifying the correct model with small samples (as in our case). For a sequence of estimated covariance matrices, we compute the time-averaged LERM-distance to the ground truth sequence as follows

$$\text{Avg. LERM}\left(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}\right) = \frac{1}{T}\sum_{t=1}^T \text{LERM}\left(\boldsymbol{\Sigma}_t, \hat{\boldsymbol{\Sigma}}_t\right),$$
$$\text{LERM}\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right) = \|\log\left(\boldsymbol{\Sigma}_1\right) - \log\left(\boldsymbol{\Sigma}_2\right)\|_F^2$$

where $\mathbf{\Sigma}_t$ and $\hat{\mathbf{\Sigma}}_t$ are the estimated and the ground truth covariance matrix at time t, respectively and $\log(\cdot)$ is the matrix logarithm.

Factor model comparison

The proposed model can be re-cast as a sparse factor model, where the variance of the factors are time-dependent. We analyzed the synthetic data set with discrete switching dynamics (described in section 3.2.1 in the paper) with a battery of factor models using the same number of factors, i.e. $K = 8$. The top-most row in Figure 1 shows the outer products of the 8 ground truth covariance components. The second row from the top shows the posterior expectations of covariance components extracted using the proposed model. Rows 3–6 show the outer products of the estimated loadings for FastICA (Hyvärinen, 2016), principle component analysis (Jolliffe, 2004), sparse principle components analysis (Zou et al., 2010) and dictionary learning (Mairal et al., 2009). The estimated components have been re-scaled to the same range and sorted to facilitate comparison. It is seen that the proposed model does indeed capture the true covariance components, while the factor models, which assume a priori temporal independence of the factors, do not capture the true covariance components.

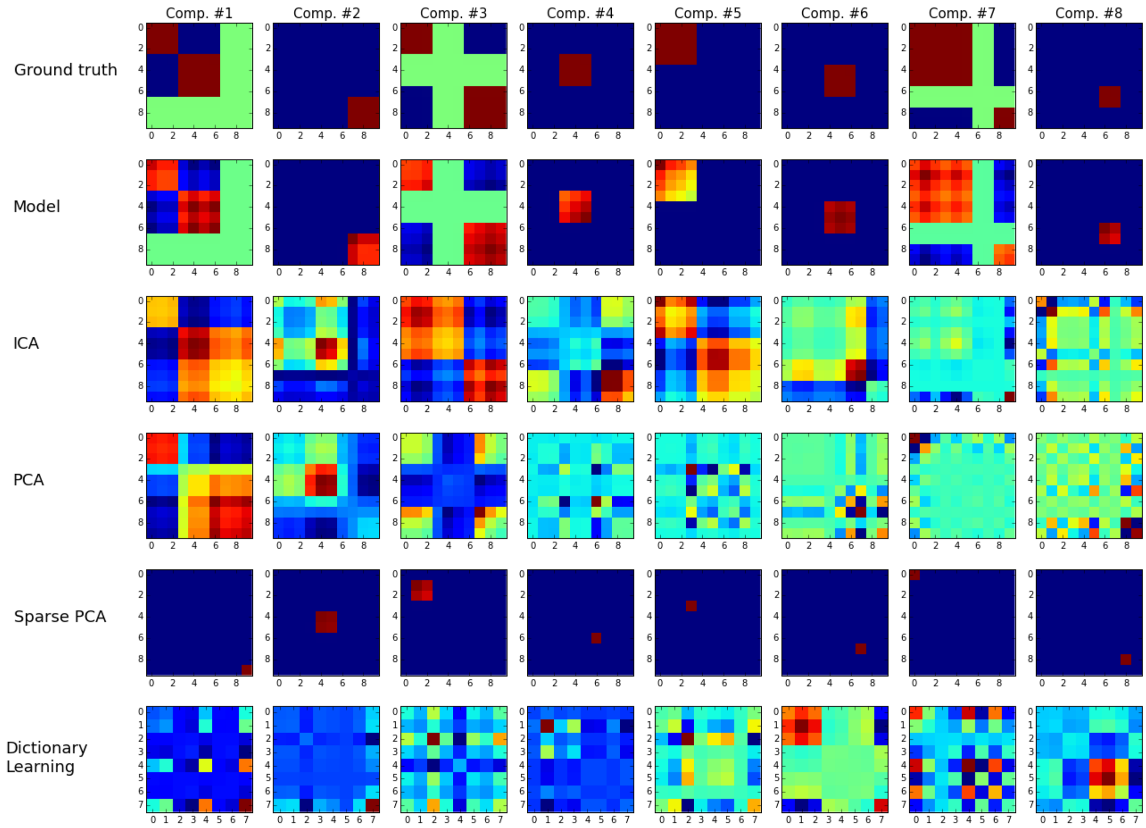
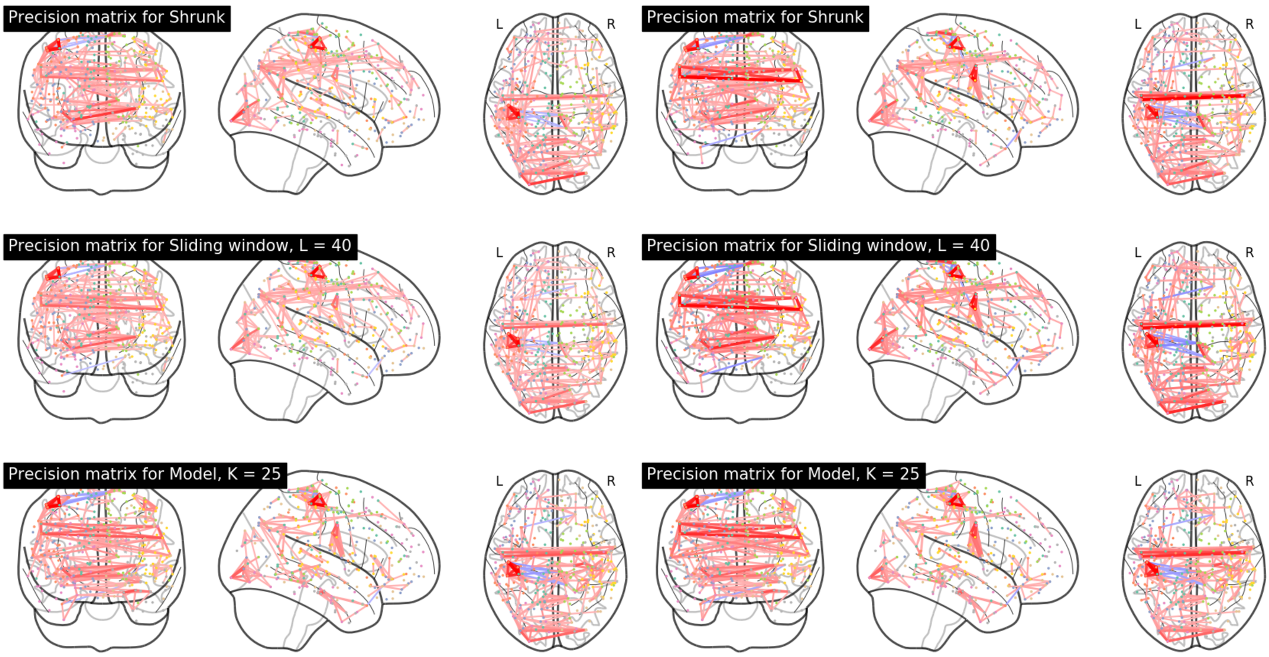


Figure 1: Comparison of estimated covariance components using a battery of different factor models.

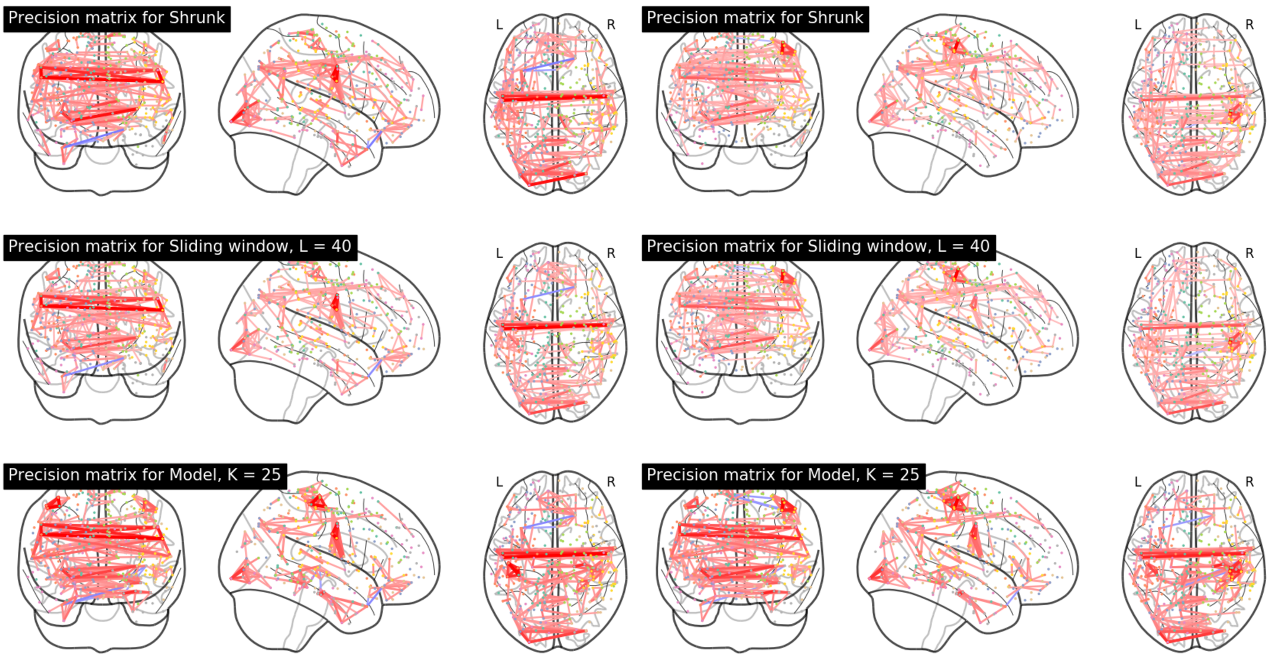
Additional figures from fMRI experiment

Figure 2 visualizes the inverses of the estimated task covariance matrices as brain networks using $N = 5$ training subjects for 4 different tasks. It is seen that the general structure of the networks is similar, although the networks from the proposed model are sparser due to the sparsity assumptions of the model. By inspecting the lower panels in Figure 2(a) and (d), it is seen that the 'left finger tapping' task induces a localized network component in the right hand side and vice versa, which is consistent with the expectations from earlier studies (Saladin, 2010).

From the confusion matrix shown in Figure 3(a), it is seen that a large proportion of the blocks belonging to the 'tongue wagging'-class is being classified as either 'left foot tapping' (28%) or 'right foot tapping' (21%). In the experimental paradigm, these two classes appear right before and right after the onset of the 'tongue wagging' task.



(a) Right hand tapping (b) Left foot tapping



(c) Tongue wagging (d) Left hand tapping

Figure 2: Visualization of estimated precision matrices for four different tasks.

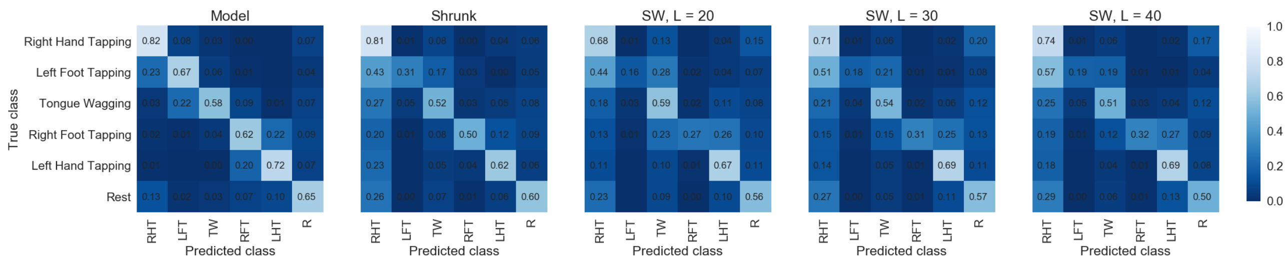


Figure 3: Confusion matrices of classification accuracies across all tasks for each method.

References

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. November 2016. URL `http://arxiv.org/abs/1601.00670`.

Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xianqiu Li, and Xilin Chen. Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 720–729, 2015.

Aapo Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. 2016. doi: 10.1.1.74.8117.

I.T. Jolliffe. *Principal component analysis.* Springer,, 2004. ISBN 0387954422.

Miguel Lázaro-gredilla and Michalis K Titsias. Spike and slab variational inference for Multi-Task and multiple kernel learning. In J Shawe-Taylor, R S Zemel, P L Bartlett, F Pereira, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 2339–2347. Curran Associates, Inc., 2011.

Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. *Acm International Conference Proceeding Series*, 382:86, 2009. doi: 10.1145/1553374.1553463.

D.B. Owen. A table of normal integrals. *Communications in Statistics Part B-simulation and Computation*, 9 (4):389–419, 1980. ISSN 15324141, 03610918. doi: 10.1080/03610918008812164.

Kenneth S. Saladin. *Anatomy & physiology : the unity of form and function.* 2010. ISBN 9780071283410. URL `http://www.worldcat.org/isbn/9780071283410`.

Raviteja Vemulapalli and David W. Jacobs. Riemannian metric learning for symmetric positive definite matrices. *CoRR*, abs/1501.02393, 2015. URL `http://arxiv.org/abs/1501.02393`.

Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. 2010. doi: 10.1.1.159.759.