Analysis of Brain States from Multi-Region LFP Time-Series Supplemental Material

Kyle Ulrich [1], **David E. Carlson** [1], **Wenzhao Lian** [1], **Jana Schaich Borg** [2], **Kafui Dzirasa** [2] and **Lawrence Carin** [1]
[1] Department of Electrical and Computer Engineering
[2] Department of Psychiatry and Behavioral Sciences
Duke University, Durham, NC 27708
{kyle.ulrich, david.carlson, wenzhao.lian, jana.borg, kafui.dzirasa, lcarin}@duke.edu

1 Reformulation as an infinite tensor mixture model

Our proposed model has similar properties to the infinite tensor mixture (ITM) model in [1], which uses a nonnegative PARAFAC tensor decomposition [2], when we look at the stationary distribution over brain states. Though similar in many ways, the PARAFAC decomposition is fundamentally different than the higher order SVD (HOSVD) decomposition [3], which is used in the simplex factor model of [4]. To reformulate our model as a nonnegative PARAFAC decomposition, we follow [1] and define probabilities over clusters $\boldsymbol{\pi}^{(aw)}$ according to an infinite tensor factorization (ITF)

$$\text{pr}(z_w^{(a1)} = \ell_1, \dots, z_w^{(aR)} = \ell_R) = \pi^{(aw)}_{\ell_1,\dots,\ell_R}, \quad \boldsymbol{\pi}^{(aw)} = \sum\nolimits_{h=1}^{\infty} \text{pr}(s_w^{(a)} = h) \bigotimes\nolimits_{r=1}^{R} \boldsymbol{\phi}_h^{(r)}, \quad (1)$$

where $\otimes$ represents the tensor product. In [1], $\text{pr}(s_w^{(a)})$ is drawn as a GEM distribution, whereas, in our work, $\text{pr}(s_w^{(a)})$ is represented by the stationary distribution of the iHMM for animal a. In this infinite tensor reformulation, each brain state is represented as a rank-1, R-way probability tensor. Given brain state h, the distribution over cluster assignments for region r is given by the probability vector $\boldsymbol{\phi}_h^{(r)}$. In our model, the prior distributions on $\boldsymbol{\phi}_h^{(r)}$ are still equivalent to their definitions in the paper, whereas in [1] they are stick-breaking processes. The resulting probability tensor $\boldsymbol{\pi}^{(aw)}$ represents the stationary joint distribution of cluster assignments for all regions.

Our clustering model differs from this ITF model of [1] in three significant ways: we place Markovian dynamics on state assignments for each animal, we model separate draws from the prior jointly for each animal, and we share cluster atoms across all regions through use of an HDP. These are significant differences that change the dynamics and interpretation of our model. For example, a primary difference is that, in our model, two regions assigned to the same index cluster are drawn *iid* from the same atom in the mixture model.

2 Parameter updates

As listed in the paper, the variational lower bound of our model is

$$\begin{aligned}\mathcal{L}(q) = \mathbb{E}[\ln p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{\Theta})] + \mathbb{E}[\ln p(\boldsymbol{Z}, \boldsymbol{S}|\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})] + \mathbb{E}[\ln p(\boldsymbol{\Phi}|\boldsymbol{\eta})] + \mathbb{E}[\ln p(\boldsymbol{\eta})] + \mathbb{E}[\ln p(\boldsymbol{\Lambda}|\boldsymbol{\beta})] \\ + \mathbb{E}[\ln p(\boldsymbol{\beta})] + \mathbb{E}[\ln p(\boldsymbol{\Psi})] + \mathbb{E}[\ln p(\boldsymbol{\Theta})] + \mathbb{H}[q(\boldsymbol{Z})] + \mathbb{H}[q(\boldsymbol{S})] + \mathbb{H}[q(\boldsymbol{\Phi})] + \mathbb{H}[q(\boldsymbol{\Lambda})], \quad (2)\end{aligned}$$

with a variational posterior factorization over model parameters

$$q(\boldsymbol{Z}) = \prod_{a,r,w} \text{Cat}(z_w^{(ar)}; \boldsymbol{\zeta}_w^{(ar)}), \qquad q(\boldsymbol{\Phi}) = \prod_{h,r} \text{Dir}(\boldsymbol{\phi}_h^{(r)}; \boldsymbol{\nu}_h^{(r)}), \qquad q(\boldsymbol{\eta}) = \delta_{\boldsymbol{\eta}^*}(\boldsymbol{\eta}),$$
$$q(\boldsymbol{S}) = \prod_a q(\{s_w^{(a)}\}_{w=1}^W), \qquad q(\boldsymbol{\Lambda}) = \prod_{g,a} \text{Dir}(\boldsymbol{\lambda}_g^{(a)}; \boldsymbol{\kappa}_g^{(a)}), \qquad q(\boldsymbol{\beta}) = \delta_{\boldsymbol{\beta}^*}(\boldsymbol{\beta}),$$
$$q(\boldsymbol{\Psi}) = \prod_a \delta_{\boldsymbol{\psi}^{(a)*}}(\boldsymbol{\psi}^{(a)}), \qquad q(\boldsymbol{\Theta}) = \prod_j \delta_{\boldsymbol{\Theta}_j^*}(\boldsymbol{\Theta}_j). \tag{3}$$

The following subsections detail the updates in this model.

2.1 Updates for cluster parameters

The variational updates for $\zeta_{w\ell}^{(ar)}$ and $\nu_{h\ell}^{(r)}$ are standard:

$$\zeta_{w\ell}^{(ar)} \propto \sum_{h=1}^{H} \rho_{wh}^{(a)} \mathbb{E}\Big[\ln \phi_{h\ell}^{(r)}\Big] - \frac{1}{2} \ln |\boldsymbol{\Sigma}_\ell| - \frac{1}{2} \boldsymbol{y}_w^{(ar)T} \boldsymbol{\Sigma}_\ell \boldsymbol{y}_w^{(ar)} \tag{4}$$

$$\nu_{h\ell}^{(r)} = \alpha_1 \boldsymbol{\eta}^* + \sum_{a=1}^{A} \sum_{w=1}^{W} \zeta_{w\ell}^{(ar)} \rho_{wh}^{(a)} \tag{5}$$

where $\mathbb{E}\Big[\ln \phi_{h\ell}^{(r)}\Big] = \psi\Big(\nu_{h\ell}^{(r)}\Big) - \psi\Big(\sum_{i=1}^{L} \nu_{hi}^{(r)}\Big)$ and $\psi(\cdot)$ is the digamma function.

2.2 Updates for state parameters

States are assigned according to a hidden Markov model (HMM). The VBM step for this procedure is:

$$\psi_h^{(a)*} \propto 1 + \mathbb{E}\Big[q(s_1^{(a)} = h)\Big] \tag{6}$$

$$\kappa_{gh}^{(a)} = \alpha_0 \boldsymbol{\beta}^* + \sum_{w=2}^{W} \mathbb{E}\Big[q(s_{w-1}^{(a)} = g, s_w^{(a)} = h)\Big] \tag{7}$$

where $\rho_{wh}^{(a)} = \mathbb{E}\Big[q(s_w^{(a)} = h)\Big]$ and $\mathbb{E}\Big[q(s_{w-1}^{(a)} = g, s_w^{(a)} = h)\Big]$ are obtained in a VBE step through the forwards backwards algorithm for variational inference. See (Beal, 2003) for more details.

2.3 Updates for global probability vectors

For both $\boldsymbol{\eta}$ and $\boldsymbol{\beta}$ we have a non-conjugate update. We generalize these two updates to be within the framework of the following update for $\boldsymbol{\beta}$:

$$\boldsymbol{\pi}_c \sim \text{DP}(\alpha_0 \boldsymbol{\beta}) \qquad\qquad \boldsymbol{\beta} \sim \text{GEM}(\gamma_0) \tag{8}$$

where $\text{GEM}(\gamma_0)$ is the stick-breaking construction for the atom weights, defined by

$$\beta_k = \beta'_k \prod_{i=1}^{k-1} (1 - \beta'_i) \qquad\qquad \beta'_k \sim \text{Beta}(1, \gamma_0) \tag{9}$$

We define an objective function $\ell(\boldsymbol{\beta}')$ for $\boldsymbol{\beta}'$ as the portions of the variational log-posterior distribution that depends on $\boldsymbol{\beta}'$. The goal is to obtain a point estimate for $\boldsymbol{\beta}'$ that maximizes the objective. This log-posterior distribution is found as:

$$\ell(\boldsymbol{\beta}') = \log\left(\prod_{k=1}^{K} \text{Beta}(\boldsymbol{\beta}'_k; 1, \gamma_0)\right) + \mathbb{E}\left[\log\left(\prod_{c=1}^{C} \text{Dir}(\boldsymbol{\pi}_c | \alpha_0 \boldsymbol{\beta})\right)\right] + const \tag{10}$$

$$= \sum_{k=1}^{K} (\gamma_0 - 1) \log(1 - \beta'_k) + \sum_{c=1}^{C} \mathbb{E}\left[-\log(\text{B}(\alpha_0 \boldsymbol{\beta})) + \sum_{k=1}^{K} (\alpha_0 \beta_k - 1) \log(\pi_{ck})\right] + const \tag{11}$$

$$= \underbrace{\sum_{k=1}^{K} (\gamma_0 - 1) \log(1 - \beta'_k)}_{\ell_1(\boldsymbol{\beta}')} + \underbrace{\sum_{k=1}^{K} (\alpha_0 \beta_k - 1) \sum_{c=1}^{C} \mathbb{E}[\log(\pi_{ck})]}_{\ell_2(\boldsymbol{\beta}')} \underbrace{- C \log(\text{B}(\alpha_0 \boldsymbol{\beta}))}_{\ell_3(\boldsymbol{\beta}')} + const \tag{12}$$

where $\mathrm{B}(\cdot)$ is the multivariate beta function. In order to maximize this log-posterior objective, we take the derivative with respect of $\boldsymbol{\beta}'$ and perform a first-order conjugate gradient optimization algorithm. The derivatives of each component (ℓ_1,ℓ_2, and ℓ_3) are found as follows:

- For $\ell_1(\boldsymbol{\beta}')$,

$$\frac{\partial \ell_1}{\partial \beta'_k} = \frac{1-\gamma_0}{1-\beta'_k} \tag{13}$$

- For $\ell_2(\boldsymbol{\beta}')$, we first begin by expanding the stick-breaking representation of β_k

$$\ell_2 = \sum_{k=1}^{K}\left(\alpha_0\beta'_k\prod_{i=1}^{k-1}(1-\beta'_i)-1\right)\sum_{c=1}^{C}\mathbb{E}[\log(\pi_{ck})] \tag{14}$$

$$\frac{\partial \ell_2}{\partial \beta'_k} = \underbrace{\alpha_0\prod_{i=1}^{k-1}(1-\beta'_i)}_{d_k}\underbrace{\sum_{c=1}^{C}\mathbb{E}[\log\pi_{ck}]}_{\zeta_k} - \sum_{l=k+1}^{K}\underbrace{\alpha_0\beta'_l\prod_{i=1;i\neq k}^{l-1}(1-\beta'_i)}_{e_{kl}}\underbrace{\sum_{c=1}^{C}\mathbb{E}[\log\pi_{cl}]}_{\zeta_l} \tag{15}$$

$$= d_k\zeta_k - \sum_{l=k+1}^{K} e_{kl}\zeta_l \tag{16}$$

- For $\ell_3(\boldsymbol{\beta}')$, we first begin by expanding the stick-breaking representation of β_k

$$\ell_3 = -C\log(\mathrm{B}(\alpha_0\boldsymbol{\beta})) \tag{17}$$

$$= -C\left[\sum_{k=1}^{K}\log\Gamma(\alpha_0\beta_k) - \log\Gamma\left(\alpha_0\sum_{k=1}^{K}\beta_k\right)\right] \tag{18}$$

$$= -C\left[\sum_{k=1}^{K}\log\Gamma(a_k) - \log\Gamma(b)\right] \tag{19}$$

where the following equalities hold:

$$a_k = \alpha_0\beta_k \tag{20}$$

$$b = \alpha_0\sum_{k=1}^{K}\beta_k \tag{21}$$

$$\frac{\partial a_\ell}{\partial \beta'_k} = \begin{cases} 0, & \text{if } k > \ell \\ d_k, & \text{if } k = \ell \\ e_{kl}, & \text{if } k < \ell \end{cases} \tag{22}$$

$$\frac{\partial b}{\partial \beta'_k} = \sum_{\ell=1}^{K}\frac{\partial a_\ell}{\partial \beta'_k} \tag{23}$$

We now take the derivative of ℓ_3 with respect to β'_k. In this derivation, we take the derivative of a gamma function where $\frac{\partial\Gamma(x)}{\partial x} = \Gamma(x)\psi_0(x)$ where $\psi_0(\cdot)$ is termed the *polygamma function*.

$$\frac{\partial \ell_3}{\partial \beta'_k} = -C\left[\sum_{\ell=1}^{K}\psi_0(a_\ell)\frac{\partial a_\ell}{\partial \beta'_k} - \psi_0(b)\frac{\partial b}{\partial \beta'_k}\right] \tag{24}$$

$$= -C\left[d_k f_k - \sum_{\ell=k+1}^{K} e_{k\ell} f_\ell\right] \tag{25}$$

where $f_\ell = \psi_0(a_\ell) - \psi_0(b)$

Combining equations 13, 16 and 24, we are able to obtain the first derivative of the objective function $\ell(\boldsymbol{\beta}')$ with respect to β'_k,

$$\frac{\partial \ell}{\partial \beta'_k} = \frac{\partial \ell_1}{\partial \beta'_k} + \frac{\partial \ell_2}{\partial \beta'_k} + \frac{\partial \ell_3}{\partial \beta'_k} \tag{26}$$

$$= \frac{1-\gamma_0}{1-\beta'_k} + \left[d_k \zeta_k - \sum_{\ell=k+1}^{K} e_{k\ell} \zeta_\ell \right] - C \left[d_k f_k - \sum_{\ell=k+1}^{K} e_{k\ell} f_\ell \right] \tag{27}$$

$$= \frac{1-\gamma_0}{1-\beta'_k} + d_k(\zeta_k - C f_k) + \sum_{\ell=k+1}^{K} e_{k\ell}(C f_\ell - \zeta_\ell) \tag{28}$$

$$= \frac{1-\gamma_0}{1-\beta'_k} + \alpha_0 \sum_{\ell=k}^{K} \frac{\partial \beta_\ell}{\partial \beta'_k} \left[\left(\sum_c \mathbb{E}[\log \pi_{c\ell}] \right) - C(\psi_0(\alpha_0 \beta_k) - \psi_0(\alpha_0)) \right] \tag{29}$$

$$= \frac{1-\gamma_0}{1-\beta'_k} + \alpha_0 \sum_{\ell=k}^{K} \frac{\partial \beta_\ell}{\partial \beta'_k} \left[\left(\sum_c \mathbb{E}[\log \pi_{c\ell}] \right) - C\mathbb{E}[\log \pi_{ck}] \right] \tag{30}$$

Most line search methods should suffice to update $\boldsymbol{\beta}$ using this derivative. We use a resilient back propagation method; see (Toussaint, 2012) for more details on this particular method.

2.4 Updates for kernel parameters

To update the parameters in $\boldsymbol{\Theta}$, we take the derivative of the variational lower bound with respect to each parameter $\boldsymbol{\Theta}_j$.

$$\frac{\partial}{\partial \boldsymbol{\Theta}_j} \mathbb{E}[\ln p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{\Theta})] = \frac{1}{2} \sum_{\ell=1}^{L} \zeta_{w\ell}^{(ar)} \left[\boldsymbol{y}_w^{(ar)T} \boldsymbol{\Sigma}_\ell^{-1} \frac{\partial \boldsymbol{\Sigma}_\ell}{\partial \boldsymbol{\Theta}_j} \boldsymbol{\Sigma}_\ell^{-1} \boldsymbol{y}_w^{(ar)} - \mathrm{tr}\left(\boldsymbol{\Sigma}_\ell^{-1} \frac{\partial \boldsymbol{\Sigma}_\ell}{\partial \boldsymbol{\Theta}_j} \right) \right] \tag{31}$$

where expectations are removed from the kernel parameters for notational simplicity and the derivative of the covariance matrix $\boldsymbol{\Sigma}_\ell$ with respect to the parameter $\boldsymbol{\Theta}_j$ is represented by the symmetric toeplitz matrix, $\frac{\partial \boldsymbol{\Sigma}_\ell}{\partial \boldsymbol{\Theta}_j} = \text{toeplitz}(\frac{\partial}{\partial \boldsymbol{\Theta}_j} k(\boldsymbol{\tau}; \boldsymbol{\theta}_\ell, \gamma))$. The derivatives of the kernel are computed to be

$$\frac{\partial}{\partial w_q} k(\tau; \boldsymbol{\theta}, \gamma) = \exp(-2\pi^2 \tau^2 \nu_q) \cos(2\pi\tau\mu_q) \tag{32}$$

$$\frac{\partial}{\partial \nu_q} k(\tau; \boldsymbol{\theta}, \gamma) = -2\pi^2 \tau^2 w_q \exp(-2\pi^2 \tau^2 \nu_q) \cos(2\pi\tau\mu_q) \tag{33}$$

$$\frac{\partial}{\partial \mu_q} k(\tau; \boldsymbol{\theta}, \gamma) = -2\pi\tau w_q \exp(-2\pi^2 \tau^2 \nu_q) \sin(2\pi\tau\mu_q) \tag{34}$$

$$\frac{\partial}{\partial \gamma} k(\tau; \boldsymbol{\theta}, \gamma) = -\frac{1}{\gamma^2} \delta_\tau \tag{35}$$

Using equation 31, any simple gradient descent algorithm can be used to optimize the marginal likelihood. We use a resilient back propagation method in order to not rely heavily on the size of the gradients.

3 Tensor method details

With the tensor factorization method, STFT coefficients are computed using 5–Hz increments in frequency and 1–minute increments in time, with half-window overlap. The tensor decomposition algorithm is applied to the 3–way tensor of brain region, frequency band and time. Each dimension of the score vectors is associated with a subset of frequency bands and brain regions. K-means is used to group the score vectors into $K = 4$ ad hoc states.

Acknowledgments

The research reported here was funded in part by ARO, DARPA, DOE, NGA and ONR.

References

[1] J Murray and D B Dunson, "Bayesian learning of joint distributions of objects," *AISTATS*, 2013.

[2] R A Harshman, "Foundations of the Parafac Procedure," *Work. Pap. Phonetics*, 1970.

[3] L D Lathauwer, B D Moor, and J Vandewalle, "On the Best Rank–1 and Rank–(R1,R2,...,RN) Approximation of Higher-Order Tensors," *SIAM J. Matrix Anal.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[4] A Bhattacharya and D B Dunson, "Simplex Factor Models for Multivariate Unordered Categorical Data," *J. Am. Stat. Assoc.*, vol. 107, no. 497, pp. 362–377, 2012.