State-Space Global Coherence to Estimate the Spatio-Temporal Dynamics of the Coordinated Brain Activity

Ali Yousefi, Reza Saadati Fard, Uri T. Eden, and Emery N. Brown

***Abstract*— Characterizing coordinated brain dynamics present in high-density neural recordings is critical for understanding the neurophysiology of healthy and pathological brain states and to develop principled strategies for therapeutic interventions. In this research, we propose a new modeling framework called State Space Global Coherence (SSGC), which allows us to estimate neural synchrony across distributed brain activity with fine temporal resolution. In this modeling framework, the cross-spectral matrix of neural activity at a specific frequency is defined as a function of a dynamical state variable representing a measure of Global Coherence (GC); we then combine filter-smoother and Expectation-Maximization (EM) algorithms to estimate GC and the model parameters. We demonstrate a SSGC analysis in a 64-channel EEG recording of a human subject under general anesthesia and compare the modeling result with empirical measures of GC. We show that SSGC not only attains a finer time resolution but also provides more accurate estimation of GC.**

I. INTRODUCTION

Synchronization of neural activity in specialized brain areas underlies a variety of brain functions linked to perceptual and cognitive tasks [1]. Abnormal synchronization of neural activity is associated with a variety of neurological disorders including schizophrenia, autism, epilepsy, and Parkinson's disease [2]. These discoveries highlight the importance of developing reliable techniques capable of characterizing the spatiotemporal dynamics of neural activity across multiple brain areas.

Coherence-based analyses including paired and global measures are widely utilized to characterize synchronized neural activity between brain areas [3],[4]. Paired-coherence analysis, which is a more established technique, provides a measure of synchrony across two brain nodes, and its utilization to study neural synchrony across a larger set of brain areas typically involves using pairwise coherence estimates to construct networks of associated brain activity. Evaluating uncertainty in features of such networks from pairwise measures is challenging [5]. Global Coherence (GC), which is a more recent technique adopted for neural data analysis, provides a measure of overall synchrony across multiple brain areas. GC is defined by the ratio of the largest eigenvalue to the sum of all of the eigenvalues of the cross-spectral matrix at a given frequency and time window; under this definition, GC variation over multiple time windows reflects how the cross-spectral matrix eigenvalues – as well as its basis vectors – change over time. GC provides a low-dimensional representation – or, a summary statistic – of the degree and spatial structure of synchrony of multiple channels of data and thus it can provide a better picture of the extent of neural synchrony across multiple brain areas. GC has been applied to neural data recorded during sleep-wake cycles, epilepsy, and anesthesia, and has opened doors to better understand the neural activity mechanisms across brain areas under different brain states [4],[6],[7]. However, there is significant computational challenge in estimating GC when the dimension of the neural recording or the number of channels grows. GC calculation requires robust and accurate estimation of the cross-spectral matrix; otherwise, GC estimated can become significantly biased. GC estimation uses the sample cross-spectral matrix, typically computed using the FFTs across channels over consecutive and non-overlapping time intervals. When the number of channels – or brain areas – grows, a larger number of independent samples is required for accurate cross-spectral matrix estimation, and this limits the temporal resolution of the GC estimate. When GC is dynamic, this leads to estimates that change slowly, even when the true underlying dynamics include rapid transitions, limiting their use in some applications such as closed-loop experiments. In this work, we propose a new modeling framework using a state-space model called State-Space Global Coherence (SSGC). In SSGC, the signal noise covariance matrix is defined as a function of a set of latent state variables, and thus it allows correlated changes across multivariate time-series data to be captured over time.

The state space modeling framework has been successfully applied in multiple neuroscience data analysis problems including decoding arm movement using neural activity of distributed brain activity across motor areas or decoding a rat's movement trajectory given ensemble spiking activity from hippocampus [8],[9]. The SSGC modeling framework proposed here inherits many properties of these previously developed frameworks; however, it incorporates new techniques that address the challenges of estimating dynamic rhythmic associations globally. In SSGC, we hypothesize that the temporal changes in the cross-spectral matrix arise in response to dynamical latent variables, and estimating these variables at each time helps to better capture coordinated activity present across multiple channels. We propose an iterative estimation procedure for the cross-spectral matrix, which allows us to estimate GC instant by instant, even as the number of observed signals or neural recordings grows large. For SSGC; we develop a filter-smoother solution for the latent state process and use the EM algorithm to identify the model parameters [8]. The SSGC framework provides a scalable

* This research was partially funded by R01 MH105174 and SCGB grant #320135

R.S. is with the Department of Electrical and Computer Engineering of Isfahan University of Technology, Isfahan, Iran

E.N.B. is with the Department of Brain and Cognitive Science of MIT, Cambridge, Massachusetts, USA

U.T.E and A.Y are with the Department of Mathematics and Statistics of Boston University, Boston, MA 02116 USA (e-mail: ayousefi@mgh.harvard.edu)

platform to characterize the causal relationship between different brain states and its distributed neural activity; it can be also expanded to study how stimulation factors – or, a combination of physiological and external mechanisms – will be represented in the brain synchronous activity.

In the Methods section, we describe the SSGC modeling framework and derive the GC and model parameter estimation algorithms. In the Application section, we illustrate GC and parameter estimation in 64-channel EEG recordings from a human subject undergoing anesthesia induction. We finally discuss further properties of the framework and additional applications.

II. Methods

A. State Space Global Coherence

Let C_k^f represent an estimate of the cross-spectral matrix of M data channels at frequency f and time interval k. C_k^f is a Hermitian semi-definite matrix whose ij^{th} element is given by

$$C_k^f(i,j) = < \tilde{X}_{i,k}(f)\tilde{X}_{j,k}^*(f) > \qquad (1)$$

where $\tilde{X}_{i,k}(f)$ and $\tilde{X}_{j,k}(f)$ are spectral estimates of i^{th} and j^{th} channels for the time points $[(k-1)N \quad kN]$, N is the length of the window over which the spectrum is estimated, and $\tilde{X}_{j,k}^*(f)$ is the complex conjugate of $\tilde{X}_{j,k}(f)$.

The GC measure of the time series at time interval k and frequency f is defined by

$$g_k^f = \lambda_{1,k}^f / \sum_{m=1}^{M} \lambda_{m,k}^f \qquad (2)$$

where $\lambda_{m,k}^f$ is the m^{th} largest eigenvalue of C_k^f.

Estimation of C_k^f requires multiple, independent estimates of the cross-spectrum across data channels – typically obtained either by computing FFTs across non-overlapping time intervals, using multiple orthogonal data tapers, or some combination of these [10]. As the number of time series – or signal channels – grow, more independent samples are required to provide precise estimates of the cross-spectral matrix, and this causes GC estimation to lose its temporal resolution. Using SSGC, we build parametric models for the dynamics of the cross-spectral matrix, which allows us to retain fine temporal resolution by optimally combining information across consecutive time intervals. We assume that C_k^f is a function of a dynamical latent variable, x_k^f, which carries information about rhythmic associations across time intervals.

We assume that x_k^f is a first-order Markov process defined by a random walk model

$$x_{k+1}^f = x_k^f + v_k^f, \;\; v_k^f \sim N(0, \sigma_{v,f}^2), \;\;\; x_0^f \sim N(\mu_{0,f}, \sigma_{0,f}^2) \qquad (3)$$

where, x_k^f represents the latent variable at time interval k and frequency f, v_k^f represents the process noise, and $\sigma_{v,f}^2$, $\mu_{0,f}$, and $\sigma_{0,f}^2$ are model parameters.

The observation process at each time interval, Y_k^f, is a vector of spectral estimates from each of the M data channels at frequency f. While many spectral estimators are possible, including multitaper method estimates, here, we focus on a single FFT estimate. We assume that the observation process has a complex multivariate normal distribution defined by [11]

$$Y_k^f = \mu + W_k \qquad W_k \sim N(0, Q_k^f) \qquad (4)$$

where, the mean of the observation vector, μ, is stationary and the observation noise process is defined by a time-varying complex covariance matrix, Q_k^f. We define an observation model to express Q_k^f as a function of the state variable, x_k^f. To simplify the notation, from now on, we will drop the f superscript from both the state transition and observation processes; the whole framework is described for the specific frequency f.

The Hermitian cross-spectral matrix can be decomposed into an orthonormal basis of eigenvectors and a diagonal matrix consisting of its eigenvalues [3]. Thus, we let

$$Q_k(x_k) = L_k(x_k)D_k(x_k)L_k^H(x_k)$$
$$L_k(x_k)L_k^H(x_k) = I \qquad (5)$$

where, D_k and L_k can be functions of time and x_k. Superscript H represents the conjugate transpose operation.

Our preliminary data analyses – using EEG data recorded across multiple different experiments – suggest that L_k is often relatively stationary; the eigenvectors of the cross spectral matrix do not change substantially from window to window. Instead, changes in the cross-spectral matrix are mainly reflected in the matrix eigenvalues – D_k's diagonal elements. As a result, we propose the following parametric model for the observation covariance matrix

$$Q_k(x_k) = LD_kL^H \qquad LL^H = I \qquad (6.a)$$

$$D_{k,m} = e^{a_m + b_m x_k} \qquad m = 1, \cdots, M \qquad (6.b)$$

where, L consists of the orthogonal bases of the covariance matrix. The eigenvalues, contained in the diagonal elements of D_k, are set to be an exponential function of x_k, ensuring that Q_k remains positive definite as the state process evolves.

Equations (3), (4), and (6) define the SSGC state transition and observation processes. In the next section, we explain how to estimate the distribution of the state variable GC at each time step from the observed data using these equations.

B. State and GC Estimation

Assume to start that the model parameters are known or well-estimated; our goal is to estimate the distribution of x_k at time k, given either the set of observations up to the current time, $Y_{1:k}$, or through the entire experiment, $Y_{1:K}$. $p(x_k|Y_{1\ldots k})$ is called the filter distribution and $p(x_k|Y_{1\ldots K})$ the smoother distribution [12]. We then use the filter and smoother distributions for x_k to calculate the GC distributions at each time interval.

Using Bayes rule, the filter distribution can be written as

$$p(x_k | Y_{1\ldots k}) \propto L(Y_k; x_k)p(x_k|Y_{1\ldots k-1}) \qquad (7)$$

, where $L(Y_k; x_k)$ is the likelihood of the observed data as a function of the state, $p(x_k|Y_{1\ldots k-1})$ is called the one-step prediction density, and we have neglected a normalizing constant that does not depend on the state. The one-step prediction is computed using the Chapman–Kolmogorov equation [12] defined by

$$p(x_k|Y_{1\dots k-1}) = \int p(x_k|x_{k-1})\, p(x_{k-1}|Y_{1\dots k-1}) dx_k \quad (8)$$

where, $p(x_k|x_{k-1})$ is the state transition probability defined by equation (3), and $p(x_{k-1}|Y_{1\dots k-1})$ is the filter distribution computed from the previous time step. For low dimensional state processes, this integral can be calculated numerically, for example by using a simple Riemann sum [13]. Note that at the time step 0, $p(x_0)$ is determined by Equation (3).

Unlike standard state-space models, for which the state determines the expected values of the observations, here the state controls their covariance structure. Therefore, the state appears in the likelihood term in Equation (7) within the covariance matrix,

$$L(Y_j; x_j) = \frac{e^{-(Y_j-\mu)^* Q_j(x_j)^{-1}(Y_j-\mu)}}{\det(\pi Q_j(x_j))} \quad (9)$$

where, μ is the observation mean vector defined in equation (4) and $Q_j(x_j)$ is the cross-spectral matrix for the value of x_j. Combining equations (7)-(9), we obtain an iterative expression for the filter density, which allows us to update the estimate from the prior time step by incorporating the state transition and observation models.

Similarly, the smoother distribution can be computed by applying Bayes' rule and the Chapman-Kolmogorov equation:

$$p(x_{k-1}|Y_{1\dots K}) = p(x_{k-1}|Y_{1\dots k-1}) \int \frac{p(x_k|x_{k-1})p(x_k|Y_{1\dots K})}{p(x_k|Y_{1\dots k-1})} dx_k \quad (10)$$

where, $p(x_{k-1}|Y_{1\dots k-1})$ and $p(x_k|Y_{1\dots k-1})$ are the filter and one-step prediction distributions respectively from equation (8), $p(x_k|x_{k-1})$ is the state transition probability defined by equation (3), and $p(x_k|Y_{1\dots K})$ is the smoothing distribution from the next time step. This equation is computed iteratively, starting at the final time point $k = K$, stepping down to $k = 1$. Note that the filter and smoother distribution for time $k = K$ are the same. Equation (10) provides the update rule for fixed-interval smoothing [12]. For low dimensional state processes, it can also be calculated numerically, for example by using Reimann summation [12].

C. Model Parameter Estimation

In the previous section we assumed that the model parameters were known or well estimated. One component of the parameter set includes those associates with the state transition process, which consist of its noise process variance, σ_v^2, and its initial value distribution parameters, (μ_o, σ_0^2). The other set of the model parameters are those associated with the observation process, which include the observation mean – μ, the orthonormal basis vectors – L, and parameter set a and b – note, a and b are vectors of length M. To estimate these model parameters, we use an EM algorithm. The EM algorithm and its variants, including sequential Monte Carlo EM [13,15], are established solutions for estimating the parameters of state-space models – here, any subset of $\boldsymbol{\theta} = \{\mu_o, \sigma_0^2, \sigma_v^2, \mu, L, a, b\}$. It uses a recursive update rule to estimate the maximum likelihood solution [15]. The EM algorithm has two steps:

1) Expectation step, which finds the expectation of the log-likelihood evaluated using the current estimate for the model parameters. This expectation for the SSGC model is defined by

$$\begin{aligned} H(x_k, Y_k;\, \boldsymbol{\theta}, k = 0 \cdots K) &= \mathrm{E}[\log \textstyle\prod_{k=0}^{K} L(x_k, Y_k;\, \boldsymbol{\theta})] \\ &-\frac{1}{2}\log\sigma_0^2 - \frac{\mathrm{E}[{x_o}^2] + {m_o}^2 - 2m_o\mathrm{E}[x_o]}{2\sigma_0^2} \\ &-\frac{K}{2}\log\sigma_v^2 - \textstyle\sum_{k=1}^{K}\frac{\mathrm{E}[{x_k}^2] + \mathrm{E}[{x_{k-1}}^2] - 2\mathrm{E}[x_k x_{k-1}]}{2\sigma_v^2} \\ &-\textstyle\sum_{k=1}^{K}(Y_k-\mu)^H \mathrm{E}[Q_k(x_k)^{-1}](Y_k-\mu) \\ &-\textstyle\sum_{k=1}^{K}\log \mathrm{E}[\det(Q_k(x_k))] + Z \end{aligned} \quad (11)$$

where, each expectation of is taken over a single state variable x_k, or over at most the joint distribution of two states, x_k and x_{k-1}, given all of the observations. Z includes all extra terms not linked to x_k or model parameters.

2) Maximization step, which estimates a new set of parameters that maximize the expected log-likelihood function calculated in the expectation step. The parameter update rule is defined by

$$\boldsymbol{\theta}^t = \arg\max_{\boldsymbol{\theta}} H(x_k, Y_k;\, \boldsymbol{\theta}^{t-1}, k = 0 \cdots K) \quad (12)$$

where, t denotes the present iteration step and $\boldsymbol{\theta}^{t-1}$ is the model parameter estimate from previous iteration, $t-1$.

The EM algorithm calls the expectation and maximization steps iteratively until they convergence to a local maximum [15]. Here, we use a numerical approach combined with Taylor expansions to compute each of the expectations in equation (11). In the Appendix section, we provide a detailed derivation of the solution for parameters a and b; for brevity, we have excluded the analogous mathematical derivations developed for updating the other model parameters.

III. Application

In this section, we demonstrate the application of the SSGC framework to estimate GC as a function of 64-lead EEG recordings from a subject undergoing anesthesia induction. The length of experiment is 35 minutes, during which, the subject received a computer-controlled infusion of the anesthetic propofol at time 15 minutes. The subject remained under anesthesia through the remainder of the experiment. Scalp EEG was recorded with 64-leads at a sampling rate of 250 Hz. A surface Laplacian reference was used [4]. **Figure 1.a** and **1.b** show spectrogram estimates of the EEG data for two contact electrodes. The oscillatory activity around the alpha-band substantially increases after the 15-minute mark, when propofol is injected. In contrast, oscillatory activity in higher frequencies including the beta-band are less affected by the propofol injection.

We utilize the SSGC model to estimate GC and the underlying state variable for frequencies of 12 and 25 Hz. The observations include tapered FFT measurements from 20 channels calculated using a 256-point FFT at frequencies 12 Hz and 25 Hz. This is equal to an update time of 1 second for each GC estimate. Here, we assume all SSGC model parameters except for a and b are known. For the state transition process, we assume σ_v^2 is equal to 0.001. We also assume x_0 is zero with a large variance: $\sigma_0^2 \gg 1$; this setting corresponds to a relatively flat prior over x_0. At each frequency, we use the empirical GC estimator introduced in [3],[4] over the whole experiment to find μ, L, and an initial setting for a and b. μ corresponds to the mean of FFT measurements per channel. We use cross-spectral matrix estimate and its singular value decomposition to find L and set the initial values of a and b. We assume the initial values for all the elements of b elemnts are 1, and each element of a can be initialized by

$$a_m = \log(\lambda_m) \qquad m = 1, \dots, M \tag{14}$$

where, λ_m is the m^{th} largest eigenvalue of the empirical cross-spectral matrix SVD decomposition. We assume that x_0 is zero. With this initialization, we use the EM algorithm to estimate a and b; we also the Bayes filter and smoother defined in equations (8)-(10) to estimate x_k and GC at both frequencies.

Figure 1.c shows the GC estimates computed using discrete sliding windows. Here, we use 256-point tapered FFT measurements over 8 non-overlapping time windows to estimate the cross spectral matrix and GC. The GC is measured using a sliding window of 256 points which corresponds to 1 second of the time domain signals. For 12 Hz, GC begins to increase around the 15-minute mark in response to the propofol injection. However, using this sliding window approach, this transition occurs slowly and the estimates are noisy. For 25 Hz, there is no noticeable change in GC.

Figure 1.d and **1.e** show the distribution of estimated GC smoother distribution using the SSGC approach at these two frequencies. The GC estimate for 12 Hz now rises rapidly at the 15 minute mark. This represents a sudden and clearly significant change in the distribution of 12 Hz rhythmic activity in response to Propofol. After this point, the GC estimate gradually decays for the rest of the experiment. For 25 Hz, The SSGC estimate still remains relatively constant. In addition to detecting changes in GC more rapidly, another advantage of the SSGC methodology is to provide a distribution of GC, which allows us to easily make statements about the uncertainty or statistical significance of our estimates.

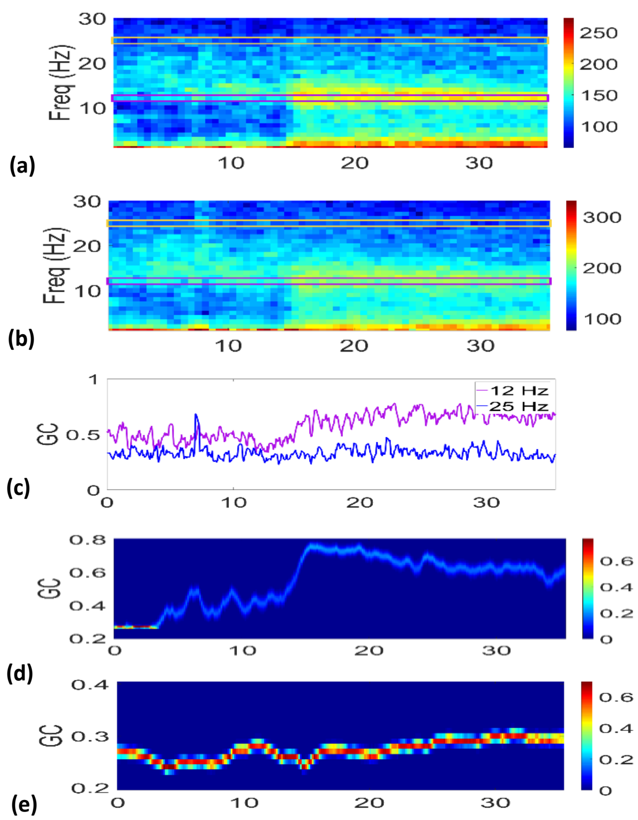

Figure 1 Sample power sepctrum of EEG recording, SSGC and empirical GC result at frequecies 12 and 25 Hz **a.** Mutlitaper spectrogram of channel 5 **b.** Multitaper spectrogram of channel 18. **c.** Emperical GC estimates at 12Hz (purple) and 25Hz (blue) **d.** GC smoother distribution using SSGC for 12 Hz **e.** GC smoothing distribution using SSGC for 25 Hz

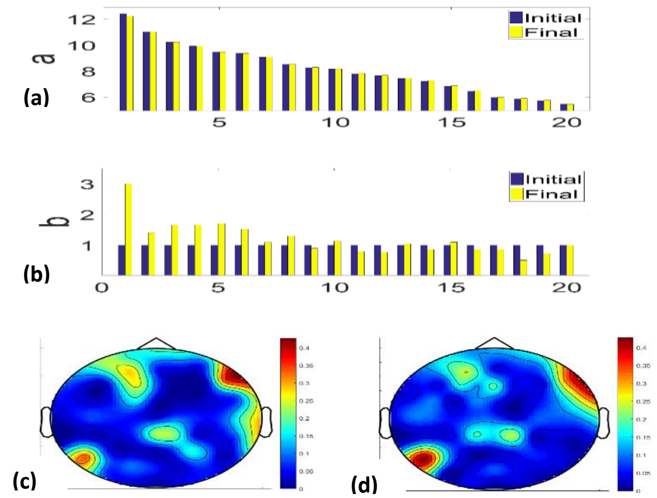

Figure 2 Estimates of a and b parameters and two largest eignevectors over the scalp at freqeuncy 12 Hz **a.** The initial and estimated values of a after 10 iteraions **b.** The initial and estimated values for b **c.** The eigenvector corresponding to the largenst eigenvalue over the scalp **d.** The eigencvtor corresponding to the second largest eigenvalue

We can also utilize SSGC to study changes in the spatial distribution of neural oscillations across different brain regions. **Figure 2.a** shows the initial and estimated values for parameter a at 12 Hz respectively. **Figure 2.b** shows the initial and estimated values for b. **Figure 2.c** and **2.d** show the eigenvectors corresponding to the first and second largest eigenvalues of cross-spectral matrix for 12 Hz, projected back onto the brain. The fact that the b parameter is dominated by its first component suggests that the rapid rise in x_k following Propofol infusion corresponds to increased alpha synchronization across the brain areas highlighted by the largest eigenvalue. This sort of analysis provides a better understanding of how widespread neural oscillations propagate over time.

IV. Conclusion

Here, we proposed a new modeling framework to estimate global coherence robustly and with fine temporal resolution and a robust estimation. We demonstrated its application in a 64-channel EEG analysis of anesthesia induction. This is an important approach as it extends common coherence measures beyond pairs of signals. The proposed framework can be applied to other modalities of neural data, including fMRI, ECoG, and depth electrode recordings. It also has potential application to closed-loop experiments relying on a global measure of the brain synchrony.

Appendix

In this section, we derive the solution that is used for estimating the a and b parameters under the EM algorithm (12). Here, we assume all other model parameters except a and b are known or well estimated.

In the E-step, we calculate the expectation of the log-likelihood with respect to $x_k,\;\; k = 1 \dots K$. To calculate the expectation, we first need to calculate the determinant of Q_k as a function of x_k. The determinant of a matrix is equal to the product of its eigenvalues;

$$\det(Q_k(x_k)) = \prod_{m=1}^{M} \exp(a_m + b_m\, x_k) \quad \text{(A.1.a)}$$

$$\log(\det(Q_k(x_k))) = \sum_{m=1}^{M} a_m + b_m\, x_k \quad \text{(A.1.b)}$$

Using equation (A.1.a), the expectation of log-likelihood function can be written as

$$H(x_k, Y_k;\, \boldsymbol{\theta}, k = 0 \cdots K) =$$
$$-\sum_{k=1}^{K} (Y_k - \mu)^H L\, \mathrm{E}[D_k(x_k)^{-1}]\, L^H (Y_k - \mu)$$
$$+\sum_{k=1}^{K}\sum_{m=1}^{M} a_m + b_m\, \mathrm{E}[x_k] + U \quad \text{(A.2)}$$

where, U includes all terms of the expectation not linked to a and b. To complete the expectation, we need to calculate the expectation of $D_k(x_k)^{-1}$. Given $D_k(x_k)$ is a diagonal matrix, its expectation is

$$\mathrm{E}[D(x_k)^{-1}] = \begin{bmatrix} e^{-a_1}\mathrm{E}[e^{-b_1 x_k}] & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{-a_M}\mathrm{E}[e^{-b_M x_k}] \end{bmatrix} \text{(A.3)}$$

(A.3)

To simplify the parameter update rule, we consider the following change of variables

$$C_k = (Y_k - \mu)^* L = [c_{1,k} \;\ldots\; c_{M,k}] \quad \text{(A.4.a)}$$

$$C_k^{\,*} = L^H (Y_k - \mu) = \begin{bmatrix} c_{1,k}^{\,*} \\ \vdots \\ c_{M,k}^{\,*} \end{bmatrix} \quad \text{(A.4.b)}$$

Now, we can write equation (A.2) by

$$H(x_k, Y_k;\, \boldsymbol{\theta}, k = 0 \cdots K) =$$
$$-\sum_{k=1}^{K}\sum_{m=1}^{M} c_{m,k} c_{m,k}^{*}\, e^{-a_m}\, \mathrm{E}[e^{-b_m x_k}] + a_m + b_m\, \mathrm{E}[x_k]`$$
$$= -\sum_{k=1}^{K}\sum_{m=1}^{M} \left|c_{m,k}\right|^2 e^{-a_m}\, \mathrm{E}[e^{-b_m x_k}] + a_m + b_m\, \mathrm{E}[x_k] \quad \text{(A.5)}$$

Note that, L is an orthonormal matrix, and thus $c_{m,i} c_{m,j}^{*} = 0$ for any $i \neq j$.

To find the update rule for a and b, we compute the derivatives of $H(\cdot)$ with respect to a and b. The derivative with respect to a_m is straightforward; however, to get the derivative with respect to b_m, we need to calculate the derivative of $\mathrm{E}[e^{-b_m x_k}]$. To address this, we use a second order Taylor expansion of $e^{-b_m x_k}$ around a known point, $b_{m,0}$. The expectation of this Taylor expansion up to order 2 is

$$\mathrm{E}[e^{-b_m x_k}] \cong e_{0,k} - e_{1,k}(b_m - b_{m,0}) + \frac{1}{2} e_{2,k}(b_m - b_{m,0})^2 \quad \text{(A.6.a)}$$

$$e_{0,k} = \mathrm{E}[e^{-b_{m,0}\, x_k}] \quad \text{(A.6.b)}$$

$$e_{1,k} = \mathrm{E}[x_k e^{-b_{m,0}\, x_k}] \quad \text{(A.6.c)}$$

$$e_{2,k} = \mathrm{E}[x_k^2\, e^{-b_{m,0}\, x_k}] \quad \text{(A.6.c)}$$

Using equations (A.5) and (A.6), we can calculate the derivative with respect to a and b. The derivative of $H(\cdot)$ with respect to a_m is

$$\partial H / \partial a_m = e^{-a_m} \sum_{k=1}^{K} \left|c_{m,k}\right|^2\, \mathrm{E}[e^{-b_m x_k}] - K \quad \text{(A.7)}$$

By setting this term to zero, we can find the update rule for a_m $m = 1, \cdots, M$

$$a_m = \log\left(\sum_{k=1}^{K} \left|c_{m,k}\right|^2\, \mathrm{E}[e^{-b_m x_k}] / K\right) \quad \text{(A.8)}$$

Note that the update for a_m is a function of b_m. The derivative of $H(\cdot)$ with respect to b_m is

$$\partial H / \partial b_m = -\sum_{k=1}^{K} \mathrm{E}[x_k]$$
$$-\sum_{k=1}^{K} \left|c_{m,k}\right|^2 e^{-a_m} \left(-e_{1,k} + e_{2,k}(b_m - b_{m,0})\right) \quad \text{(A.9)}$$

By setting this term to zero, we can find the update rule for b_m $m = 1, \cdots, M$

$$b_m = b_{m,0} + \frac{\left(K e^{a_m} \mathrm{E}[x_k] + \sum_{k=1}^{K} \left|c_{m,k}\right|^2 e_{1,k}\right)}{\left(\sum_{k=1}^{K} \left|c_{m,k}\right|^2\, e_{2,k}\right)} \quad \text{(A.10)}$$

Note that the update rule for b_m is also a function of a_m and $b_{m,0}$. We set $b_{m,0}$ to be the estimate of b_m from the previous EM step, and we then use a recursive update of a_m and b_m using equations (A.8) and (A.10). We check how much a_m and b_m changes over iterations and stop if their changes drop below a preset small threshold.

References

[1] A. M. Bastos and J.-M. Schoffelen, "A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls," *Front. Syst. Neurosci.*, vol. 9, p. 175, Jan. 2016.

[2] P. J. Uhlhaas and W. Singer, "Neural Synchrony in Brain Disorders: Relevance for Cognitive Dysfunctions and Pathophysiology," *Neuron*, vol. 52, no. 1, pp. 155–168, Oct. 2006.

[3] P. Mitra and H. Bokil, *Observed Brain Dynamics*. Oxford University Press, 2007.

[4] A. Cimenser *et al.*, "Tracking brain states under general anesthesia by using global coherence analysis.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 21, pp. 8832–7, May 2011.

[5] E. P. Stephen *et al.*, "Assessing dynamics, spatial scale, and uncertainty in task-related brain network analyses," *Front. Comput. Neurosci.*, vol. 8, p. 31, 2014.

[6] S. M. Bowyer, "Coherence a measure of the brain networks: past and present," *Neuropsychiatr. Electrophysiol.*, vol. 2, no. 1, p. 1, 2016.

[7] K. J. Friston, "Transients, metastability, and neuronal dynamics," *Neuroimage*, vol. 5, no. 2, pp. 164–171, 1997.

[8] A. C. Smith and E. N. Brown, "Estimating a state-space model from point process observations," *Neural Comput.*, vol. 15, no. 5, pp. 965–991, 2003.

[9] W. Truccolo, L. R. Hochberg, and J. P. Donoghue, "Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes," *Nat. Neurosci.*, vol. 13, no. 1, p. 105, 2010.

[10] T. W. Anderson, T. W. Anderson, T. W. Anderson, T. W. Anderson, and E.-U. Mathématicien, *An introduction to multivariate statistical analysis*, vol. 2. Wiley New York, 1958.

[11] S.-E. Kim, M. K. Behr, D. Ba, and E. N. Brown, "State-space multitaper time-frequency analysis," *Proc. Natl. Acad. Sci.*, vol. 115, no. 1, pp. E5--E14, 2018.

[12] S. Särkkä, *Bayesian filtering and smoothing*, vol. 3. Cambridge University Press, 2013.

[13] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*. Courier Corporation, 2007.

[14] R. K. S. Hankin, "The complex multivariate Gaussian distribution," *R J.*, vol. 7, no. 1, pp. 73–80, 2015.

[15] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, 1996.