PLSO: A generative framework for decomposing nonstationary time-series into piecewise stationary oscillatory components (Supplementary material)

Andrew H. Song[1] **Demba Ba**[2] **Emery N. Brown**[3,4,5]

[1]Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[2]School of Engineering and Applied Sciences, Havard University, Cambridge, Massachusetts, USA
[3] Picower Institute of Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[4] Institute of Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[5] Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

APPENDIX

References to the sections in the section headings are made with respect to the sections in the main text. Below is the table of contents for the Appendix.

A. CONTINUOUS MODEL INTERPRETATION OF PLSO

We can establish the equivalent continuous model of the PLSO in Eq. 2, using stochastic different equation

$$\frac{d\tilde{z}_j(t)}{dt} = \underbrace{\left(\left(-\frac{1}{l_j}\right) \oplus \begin{pmatrix} 0 & -\omega_j \\ \omega_j & 0 \end{pmatrix}\right)}_{\mathbf{F}} \tilde{z}_j(t) + \varepsilon(t), \quad \text{(A.1)}$$

where $\tilde{z}_j(t) : \mathbb{R} \to \mathbb{R}^2$, $\oplus$ denotes the Kronecker sum and $\varepsilon(t) \sim \mathcal{N}(0, \sigma_j^2 \mathbf{I}_{2\times 2})$. Discretizing the solution of Eq. A.1 at Δ, such that $\widetilde{\mathbf{z}}_{j,k} = \tilde{z}_j(k\Delta)$, yields Eq. 2. Consequently, we obtain the following for $\Delta > 0$

$$\begin{aligned} &\exp(\mathbf{F}\Delta) = \exp(-\Delta/l_j)\mathbf{R}(\omega_j), \\ &\sigma_j^2 \int_0^\Delta \exp(\mathbf{F}(\Delta-\tau)) \exp(\mathbf{F}(\Delta-\tau))^{\mathrm{T}} d\tau \\ &\quad = \sigma_j^2 (1 - \exp(-2\Delta/l_j)) \mathbf{I}_{2\times 2}. \end{aligned}$$

This interpretation extends to the nonstationary PLSO. The corresponding continuous model for $\widetilde{\mathbf{z}}_{j,mN+n}$ in Eq. 3 is the same as Eq. A.1, with different variance $\mathbb{E}[\varepsilon_j(t)\varepsilon_j^{\mathrm{T}}(t)] = \sum_{m=1}^M \sigma^2_{j,m} \cdot \mathbf{1}\left(\left(\frac{m-1}{M}\right)T \le t < \left(\frac{m}{M}\right)T\right)\mathbf{I}_{2\times 2}$.

B. PSD FOR COMPLEX AR(1) PROCESS

We compute the steady-state covariance denoted as $\mathbf{P}^j_\infty$. Since we assume $\mathbf{P}^j_1 = \sigma_j^2 \mathbf{I}_{2\times 2}$, it is easy to show that $\mathbf{P}^j_k$ is a diagonal matrix from $\mathbf{R}(\omega_j)\mathbf{R}^{\mathrm{T}}(\omega_j) = \mathbf{I}_{2\times 2}$. Denoting $\mathbf{P}^j_\infty = \alpha \mathbf{I}_{2\times 2}$, we use the discrete Lyapunov equation

$$\begin{aligned} &\mathbf{P}^j_\infty \\ &= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}^j_\infty \mathbf{R}^{\mathrm{T}}(\omega_j) \\ &\quad + \sigma_j^2 (1 - \exp(-2\Delta/l_j)) \mathbf{I}_{2\times 2} \\ &\Rightarrow \alpha = \exp(-2\Delta/l_j)\alpha + \sigma_j^2 (1 - \exp(-2\Delta/l_j)) \\ &\Rightarrow \mathbf{P}^j_\infty = \sigma_j^2 \mathbf{I}_{2\times 2}, \end{aligned}$$

which implies that under the assumption $\mathbf{P}^j_1 = \sigma_j^2 \mathbf{I}_{2\times 2}$, we are guaranteed $\mathbf{P}^j_k = \sigma_j^2 \mathbf{I}_{2\times 2}, \forall k$. To compute the PSD of the stationary process $\mathbf{z}_j$, we now need to compute the autocovariance. Since only $\mathbf{z}^{\Re}_{j,k}$ contributes to $\mathbf{y}_k$, we compute the autocovariance of $\mathbb{E}[\mathbf{z}^{\Re}_{j,k}\mathbf{z}^{\Re}_{j,k+n}]$ as

$$\begin{aligned} \mathbb{E}[\mathbf{z}^{\Re}_{j,k}\mathbf{z}^{\Re}_{j,k+n}] &= \mathbb{E}[\mathbf{z}^{\Re}_{j,k} \cdot \Re(\rho_j^n \exp(i\omega_j n)\mathbf{z}_{j,k})] \\ &= \rho_j^n \mathbb{E}[\mathbf{z}^{\Re}_{j,k}\mathbf{z}^{\Re}_{j,k} \cos\omega_j n] \\ &= \rho_j^n \cos\omega_j n \cdot \mathbb{E}[\{\mathbf{z}^{\Re}_{j,k}\}^2] \\ &= \rho_j^n \sigma_j^2 \cos w_j n, \end{aligned}$$

where $\Re(\cdot)$ denotes the operator that extracts the real part of the complex argument and we used the fact that $\mathbb{E}[\mathbf{z}_{j,k}^{\Re}\mathbf{z}_{j,k}^{\Im}] = 0$. The spectra for the j^{th} component, $S_j(\omega)$ can be written as

$$\begin{aligned}
S_j(\omega) &= \sum_{n=-\infty}^{\infty} \mathbb{E}\left[\mathbf{z}_{j,k}^{\Re}\mathbf{z}_{j,k+n}^{\Re}\right]\exp\left(-i\omega n\right) \\
&= \sum_{n=-\infty}^{\infty} \rho_j^n \sigma_j^2 \cos w_j n \exp\left(-i\omega n\right) \\
&= \sigma_j^2 \sum_{n=-\infty}^{\infty} \rho_j^n \left\{\exp(i\omega_j n) + \exp(-i\omega_j n)\right\}\exp\left(-i\omega n\right) \\
&= \sigma_j^2 \sum_{n=-\infty}^{\infty} \rho_j^n \exp(-i(\omega \pm \omega_j)n).
\end{aligned}$$

Unpacking the infinite sum for one of the terms,

$$\begin{aligned}
&\sum_{n=-\infty}^{\infty} \rho_j^n \exp(-i(\omega-\omega_j)n) \\
&= 1 + \sum_{n=1}^{\infty} \rho_j^n \exp(-i(\omega-\omega_j)n) + \rho_j^n \exp(i(\omega-\omega_j)n) \\
&= 1 + \frac{\rho_j \exp(-i(\omega-\omega_j))}{1-\rho_j\exp(-i(\omega-\omega_j))} + \frac{\rho_j \exp(i(\omega-\omega_j))}{1-\rho_j\exp(i(\omega-\omega_j))} \\
&= 1 + \frac{2\rho_j\cos(\omega-\omega_j) - 2\rho_j^2}{\left(1-\rho_j\exp(-i(\omega-\omega_j))\right)\left(1-\rho_j\exp(i(\omega-\omega_j))\right)} \\
&= \frac{1-\rho_j^2}{1+\rho_j^2-2\rho_j\cos(\omega-\omega_j)}.
\end{aligned}$$

Using the relation $\rho_j = \exp(-\Delta/l_j)$ and unpacking the infinite sum for the other term, we have

$$\begin{aligned}
&S_j(w) \\
&= \frac{\sigma_j^2(1-\exp(-2\Delta/l_j))}{1+\exp(-2\Delta/l_j) - 2\exp(-\Delta/l_j)\cos(\omega-\omega_j)} \\
&\quad + \frac{\sigma_j^2(1-\exp(-2\Delta/l_j))}{1+\exp(-2\Delta/l_j) - 2\exp(-\Delta/l_j)\cos(\omega+\omega_j)}.
\end{aligned}$$

Since Fourier transform is a linear operator, we can conclude that $\gamma(\omega) = \sigma_\nu^2 + \sum_{j=1}^{J} S_j(\omega)$.

C. PROOF FOR PROPOSITION 1 *(SECTION 3.2.1)*

Proposition 1. For a given m, as $\Delta \to 0$, the samples on either side of the interval boundary, which are $\widetilde{\mathbf{z}}_{j,(m+1)N}$ and $\widetilde{\mathbf{z}}_{j,(m+1)N+1}$, converge to each other in mean square,

$$\lim_{\Delta\to 0} \mathbb{E}[\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}^{\mathrm{T}}] = 0,$$

where we use $\Delta\widetilde{\mathbf{z}}_{j,(m+1)N} = \widetilde{\mathbf{z}}_{j,(m+1)N+1} - \widetilde{\mathbf{z}}_{j,(m+1)N}$.

Proof. To analyze Eq. 3 in the limit of $\Delta \to 0$, we use the equivalent continuous model. It suffices to show that $\lim_{\Delta\to 0}\exp(\mathbf{F}\Delta) = \mathbf{I}_{2\times 2}$ and $\lim_{\Delta\to 0}\mathbb{E}[\varepsilon_{j,(m+1)N+1}\varepsilon_{j,(m+1)N+1}^{\mathrm{T}}] = \mathbf{0}$. We have,

$$\begin{aligned}
&\lim_{\Delta\to 0}\exp(\mathbf{F}\Delta) = \mathbf{I}_{2\times 2} + \lim_{\Delta\to 0}\sum_{k=1}^{\infty}\frac{\Delta^k}{k!}\mathbf{F}^k = \mathbf{I}_{2\times 2} \\
&\lim_{\Delta\to 0}\mathbb{E}[\varepsilon_{j,(m+1)N+1}\varepsilon_{j,(m+1)N+1}^{\mathrm{T}}]/\sigma_{j,m+1}^2 \\
&= \lim_{\Delta\to 0}\int_0^{\Delta}\exp\left(\mathbf{F}(\Delta-\tau)\right)\exp\left(\mathbf{F}(\Delta-\tau)\right)^{\mathrm{T}}d\tau = \mathbf{0}.
\end{aligned}$$

Since this implies $\lim_{\Delta\to 0}\mathbb{E}[\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}\Delta\widetilde{\mathbf{z}}_{j,(m+1)N}^{\mathrm{T}}] = 0$, we have convergence in mean square. □

D. PROOF FOR PROPOSITION 2 *(SECTION 3.2.2)*

Proposition 2. Assume $l_j \ll N\Delta$, such that $\mathbf{P}_{m,N}^j = \mathbf{P}_{m,\infty}^j$. In Eq. 3, the difference between $\mathbf{P}_{m,\infty}^j = \sigma_{j,m}^2\mathbf{I}_{2\times 2}$ and $\mathbf{P}_{m+1,\infty}^j = \sigma_{j,m+1}^2\mathbf{I}_{2\times 2}$ decays exponentially fast as a function of n, for $1 \le n \le N$,

$$\mathbf{P}_{m+1,n}^j = \mathbf{P}_{m+1,\infty}^j + \exp\left(-2n\Delta/l_j\right)\left(\mathbf{P}_{m,\infty}^j - \mathbf{P}_{m+1,\infty}^j\right).$$

Proof. We first obtain the steady-state covariance $\mathbf{P}_{m,\infty}^j$, similar to **Appendix B**. Since we assume $\mathbf{P}_{1,1}^j = \sigma_{j,1}^2\mathbf{I}_{2\times 2}$, we can show that $\forall m, n$, $\mathbf{P}_{m,n}^j$ is a diagonal matrix, noting that $\mathbf{R}(\omega_j)\mathbf{R}^{\mathrm{T}}(\omega_j) = \mathbf{I}_{2\times 2}$. Denoting $\mathbf{P}_{m,\infty}^j = \alpha\mathbf{I}_{2\times 2}$, we now use the discrete Lyapunov equation

$$\begin{aligned}
\mathbf{P}_{m,\infty}^j &= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}_{m,\infty}^j\mathbf{R}^{\mathrm{T}}(\omega_j) \\
&\quad + \sigma_{j,m}^2\left(1-\exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\
&\Rightarrow \alpha = \exp(-2\Delta/l_j)\alpha + \sigma_{j,m}^2\left(1-\exp\left(-2\Delta/l_j\right)\right) \\
&\Rightarrow \mathbf{P}_{m,\infty}^j = \sigma_{j,m}^2\mathbf{I}_{2\times 2}.
\end{aligned}$$

We now prove the proposition by induction. For fixed j and m, and for $n = 1$,

$$\begin{aligned}
&\mathbf{P}_{m+1,1}^j \\
&= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}_{m,N}^j\mathbf{R}^{\mathrm{T}}(\omega_j) \\
&\quad + \sigma_{j,m+1}^2\left(1-\exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\
&= \left\{\sigma_{j,m+1}^2 + \exp\left(-2\Delta/l_j\right)\left(\sigma_{j,m}^2 - \sigma_{j,m+1}^2\right)\right\}\mathbf{I}_{2\times 2}.
\end{aligned}$$

Assuming the same holds for $n = n' - 1$, we have for

$n = n'$,

$$
\begin{aligned}
&\mathbf{P}^{j}_{m+1,n'} \\
&= \exp(-2\Delta/l_j)\mathbf{R}(\omega_j)\mathbf{P}^{j}_{m,n'-1}\mathbf{R}^{\mathrm{T}}(\omega_j) \\
&\quad + \sigma^2_{j,m+1}\left(1 - \exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\
&= \exp(-2\Delta/l_j)\sigma^2_{j,m+1}\mathbf{I}_{2\times 2} \\
&\quad + \exp\left(-2n'\Delta/l_j\right)\left(\sigma^2_{j,m} - \sigma^2_{j,m+1}\right)\mathbf{I}_{2\times 2} \\
&\quad + \sigma^2_{j,m+1}\left(1 - \exp\left(-2\Delta/l_j\right)\right)\mathbf{I}_{2\times 2} \\
&= \left\{\sigma^2_{j,m+1} + \exp\left(-2n'\Delta/l_j\right)\left(\sigma^2_{j,m} - \sigma^2_{j,m+1}\right)\right\}\mathbf{I}_{2\times 2}.
\end{aligned}
$$

By the principle of induction, Eq. holds for $1 \leq n \leq N$. □

E. INITIALIZATION & ESTIMATION FOR PLSO *(SECTION 4)*

E.1 INITIALIZATION

As noted in the main text, we use AIC to determine the optimal number of J. For a given number of components J, we first construct the spectrogram of the data using STFT and identify the frequency bands with prominent power, i.e., frequency bands whose average power exceeds pre-determined threshold. The center frequencies of these bands serve as the initial center frequencies $\{\omega_j^{\text{init}}\}_j$, which are either fixed throughout the algorithm or further refined through the estimation algorithm in the main text. If J exceeds the number of identified frequency bands from the spectrogram, 1) we first place $\{\omega_j^{\text{init}}\}_j$ in the prominent frequency bands and 2) we then place the remaining components uniformly spread out in $[0, \omega_c]$, where ω_c is a cutoff frequency to be further determined in the next section. As for $\{l_j^{\text{init}}\}_j$, we set it to be a certain fraction of the corresponding $\{\omega_j^{\text{init}}\}_j$. We then fit $\{\sigma^2_{j,m}\}_{j,m}$ and θ with $\lambda = 0$, through the procedure explained in Stage 1. We finally use these estimates as initial values for other values of λ.

E.2 ESTIMATION FOR σ^2_ν

There are two possible ways to estimate the observation noise variance σ^2_ν. The first approach is to perform maximum likelihood estimation of $f(\{\sigma^2_{j,m}\}_{j,m}; \theta)$ with respect to σ^2_ν. The second approach, which we found to work *better* in practice and use throughout the manuscript, is to directly estimate it from the Fourier transform of the data. Given a cutoff frequency ω_c, informed by domain knowledge, we take the average power of the Fourier transform of $\mathbf{y}$ in $[\omega_c, f_s/2]$. For instance, it is widely known that the spectral content below 40 Hz in anesthesia EEG dataset is physiologically relevant and we use $\omega_c \simeq 40$ Hz.

F. OPTIMIZATION FOR $\{\sigma^2_{j,m}\}_{j,m}$ VIA PROXIMAL GRADIENT UPDATE *(SECTION 4.1.1)*

We discuss the algorithm to obtain a local optimal solution of $\{\sigma^2_{j,m}\}_{j,m}$ to the MAP optimization problem in Eq. 8. We define $\psi_{j,m} = \log \sigma^2_{j,m}$ and $\Psi = [\psi_{1,1}, \ldots, \psi_{1,M}, \ldots, \psi_{J,M}] \in \mathbb{R}^{JM}$ for notational decluttering, to be used in this section.

We rewrite Eq. 8 as

$$
\begin{aligned}
&\min_{\Psi} \underbrace{-\log p(\Psi|\mathbf{y}, \theta)}_{h(\Psi;\theta)} \\
&= \min_{\Psi} \underbrace{\frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\{\log \gamma^{(m)}(\omega_n) + \frac{I^{(m)}(\omega_n)}{\gamma^{(m)}(\omega_n)}\right\}}_{-f(\Psi;\theta)} \\
&\qquad + \underbrace{\frac{\lambda}{2}\sum_{j=1}^{J}\sum_{m=1}^{M}\left(\psi_{j,m} - \psi_{j,m-1}\right)^2}_{-g(\Psi;\theta)} \\
&= \min_{\Psi} -f(\Psi;\theta) - g(\Psi;\theta).
\end{aligned}
$$

The algorithm is described in Algorithm 1. It follows the steps outlined in the inexact accelerated proximal gradient algorithm [Li and Lin, 2015]. For faster convergence, we use larger step sizes with the Barzilai-Borwein (BB) step size initialization rule [Barzilai and Borwein, 1988]. For rest of this section, we drop dependence on θ. The main novelty of our algorithm is the proximal gradient update

$$
\begin{aligned}
&\mathbf{u}^{(l+1)} \\
&= \operatorname{prox}_{-\alpha^{(l)}_{\mathbf{w}} g}(\mathbf{w}^{(l)} + \alpha^{(l)}_{\mathbf{w}}\nabla f(\mathbf{w}^{(l)})) \\
&= \arg\min_{\Psi} \frac{1}{2\alpha^{(l)}_{\mathbf{w}}}\|\Psi - (\mathbf{w}^{(l)} + \alpha^{(l)}_{\mathbf{w}}\nabla f(\mathbf{w}^{(l)}))\|^2 - g(\Psi) \\
&= \arg\min_{\Psi} \sum_{j=1}^{J}\sum_{m=1}^{M} \frac{\left(\left(\mathbf{w}^{(l)}_{j,m} + \alpha^{(l)}_{\mathbf{w}} \cdot \frac{\partial f(\mathbf{w}^{(l)})}{\partial \mathbf{w}_{j,m}}\right) - \psi_{j,m}\right)^2}{2\alpha^{(l)}_{\mathbf{w}}} \\
&\qquad + \frac{\lambda}{2}(\psi_{j,m} - \psi_{j,m-1})^2,
\end{aligned}
$$

where the same holds for $\mathbf{x}^{(l+1)} = \operatorname{prox}_{-\alpha^{(l)}_{\Psi} g}(\Psi^{(l)} + \alpha^{(l)}_{\Psi}\nabla f(\Psi^{(l)}))$. The auxiliary variables $\mathbf{w}, \mathbf{u}, \mathbf{x} \in \mathbb{R}^{JM}$ ensure convergence of Ψ. We use $\mathbf{w}^{(l)}_{j,m}$ to denote $((m-1)J + j)^{\text{th}}$ entry of $\mathbf{w}^{(l)}$. As mentioned in the main text, this can be solved in a computationally efficient manner by using Kalman filter/smoother.

Algorithm 1: Inference for Ψ via inexact APG

Result: $\widehat{\Psi}$
Initialize $\Psi^{(0)} = \Psi^{(1)} = \mathbf{u}^{(1)}$,
$\beta^{(0)} = 0$, $\beta^{(1)} = 1$, $\delta > 0$, $\rho < 1$
for $l \leftarrow 1$ **to** L **do**
$\mathbf{w}^{(l)} = \Psi^{(l)} + \frac{\beta^{(l-1)}}{\beta^{(l)}}(\mathbf{u}^{(l)} - \Psi^{(l)}) + \frac{\beta^{(l-1)}-1}{\beta^{(l)}}(\Psi^{(l)} - \Psi^{(l-1)})$
(BB step size initialization rule)
$\mathbf{s}^{(l)} = \mathbf{u}^{(l)} - \mathbf{w}^{(l-1)}$,
$\mathbf{r}^{(l)} = -\nabla f(\mathbf{u}^{(l)}) + \nabla f(\mathbf{w}^{(l-1)})$
$\alpha_{\mathbf{w}}^{(l)} = ((\mathbf{s}^{(l)})^{\mathrm{T}}\mathbf{s}^{(l)})/((\mathbf{s}^{(l)})^{\mathrm{T}}\mathbf{r}^{(l)})$
$\mathbf{s}^{(l)} = \mathbf{x}^{(l)} - \Psi^{(l-1)}$,
$\mathbf{r}^{(l)} = -\nabla f(\mathbf{x}^{(l)}) + \nabla f(\Psi^{(l-1)})$
$\alpha_{\Psi}^{(l)} = ((\mathbf{s}^{(l)})^{\mathrm{T}}\mathbf{s}^{(l)})/((\mathbf{s}^{(l)})^{\mathrm{T}}\mathbf{r}^{(l)})$
(Proximal update step)
repeat
$\mathbf{u}^{(l+1)} = \mathrm{prox}_{-\alpha_{\mathbf{w}}^{(l)} g}(\mathbf{w}^{(l)} + \alpha_{\mathbf{w}}^{(l)} \nabla f(\mathbf{w}^{(l)}))$
$\alpha_{\mathbf{w}}^{(l)} = \rho \cdot \alpha_{\mathbf{w}}^{(l)}$
until $h(\mathbf{u}^{(l+1)}) \leq h(\mathbf{w}^{(l)}) - \delta\|\mathbf{u}^{(l+1)} - \mathbf{w}^{(l)}\|^2$;
repeat
$\mathbf{x}^{(l+1)} = \mathrm{prox}_{-\alpha_{\Psi}^{(l)} g}(\Psi^{(l)} + \alpha_{\Psi}^{(l)} \nabla f(\Psi^{(l)}))$
$\alpha_{\Psi}^{(l)} = \rho \cdot \alpha_{\Psi}^{(l)}$
until $h(\mathbf{x}^{(l+1)}) \leq h(\Psi^{(l)}) - \delta\|\mathbf{x}^{(l+1)} - \Psi^{(l)}\|^2$;
$\beta^{(l+1)} = \frac{1+\sqrt{4(\beta^{(l)})^2+1}}{2}$
$\Psi^{(l+1)} = \begin{cases} \mathbf{u}^{(l+1)} & \text{if } h(\mathbf{u}^{(l+1)}) \leq h(\mathbf{x}^{(l+1)}) \\ \mathbf{x}^{(l+1)} & \text{otherwise} \end{cases}$
end
$\widehat{\Psi} = \Psi^{(L)}$

G. LIPSCHITZ CONSTANT FOR $\nabla f(\{\sigma_{j,m}^2\}_{j,m}; \theta)$ *(SECTION 4.1.1)*

In this section, we prove that under some assumptions, we can show that the log-likelihood $f(\{\sigma_{j,m}^2\}_{j,m}; \theta)$ has Lipschitz continuous gradient with the Lipschitz constant C. As in the previous section, we use $\psi_{j,m} = \log \sigma_{j,m}^2$ and $\Psi = [\psi_{1,1}, \ldots, \psi_{1,M}, \ldots, \psi_{J,M}] \in \mathbb{R}^{JM}$.

Let us start by restating the definition of Lipschitz continuous gradient.

Definition A continuously differentiable function $f : \mathcal{S} \to \mathbb{R}$ is Lipschitz continuous gradient if

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq C\|\mathbf{x} - \mathbf{y}\|_2 \quad \text{for every } \mathbf{x}, \mathbf{y} \in \mathcal{S},$$

where $\mathcal{S}$ is a compact subset of $\mathbb{R}^{JM}$ and $C > 0$ is the Lipschitz constant.

Our goal is to find the constant C for the Whittle likelihood $f(\Psi; \theta)$

$$\begin{aligned} f(\Psi;\theta) &= -\frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\{\log\gamma_{m,n} + \frac{I_{m,n}}{\gamma_{m,n}}\right\} \\ &= \underbrace{-\frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\log\left(\sigma_\nu^2 + \sum_{j=1}^{J}\exp(\psi_{j,m})\,\alpha_{j,n}\right)}_{f_1(\Psi;\theta)} \\ &\quad \underbrace{-\frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{I_{m,n}}{\left(\sigma_\nu^2 + \sum_{j=1}^{J}\exp(\psi_{j,m})\,\alpha_{j,n}\right)}}_{f_2(\Psi;\theta)}, \end{aligned}$$

where we use $\gamma_{m,n} = \gamma^{(m)}(\omega_n)$ and $I_{m,n} = I^{(m)}(\omega_n)$ for notational simplicity and

$$\alpha_{j,n} = \frac{(1 - \exp(-2\Delta/l_j))}{1 + \exp(-2\Delta/l_j) - 2\exp(-\Delta/l_j)\cos(\omega_n - \omega_j)}.$$

We make the following assumptions

1. $I_{m,n}$ is bounded, i.e., $I_{m,n} \leq C_I$. With the real-world signal, we can reasonably assume that $I_{m,n}$ or energy of the signal is bounded.
2. $\forall j, m, \psi_{j,m}$ is bounded, i.e., $|\psi_{j,m}| \leq \log C_\psi$ for some $C_\psi > 1$. This implies $1/C_\psi \leq \exp(\psi_{j,m}) \leq C_\psi$.

In addition, we have the following facts

1. $I_{m,n}, \alpha_{j,n}$, and $\gamma_{m,n}$ are nonnegative.
2. $I_{m,n}$ and $\gamma_{m,n}$ are bounded. This follows from the aforementioned assumptions.
3. For given $\{l_j\}_j$, we have bounded $\alpha_{j,n}$. To see this,

note that the maximum of $\alpha_{j,n}$ is acheived at $\omega_n = \omega_j$,

$$\begin{aligned}\max \alpha_{j,n} &= \frac{(1-\exp(-2\Delta/l_j))}{1+\exp(-2\Delta/l_j)-2\exp(-\Delta/l_j)} \\ &= \frac{(1+\exp(-\Delta/l_j))}{(1-\exp(-\Delta/l_j))}.\end{aligned}$$

Therefore, denoting $l_{\max} = \max_j\{l_j\}_j$,

$$\max \alpha_{j,n} \leq \frac{(1+\exp(-\Delta/l_{\max}))}{(1-\exp(-\Delta/l_{\max}))} = C_\alpha.$$

Finally, we define $\mathcal{S} = [-\log C_\psi, \log C_\psi] \subset \mathbb{R}^{JM}$.

We want to compute the Lipschitz constant for $\nabla f_1(\Psi)$ and $\nabla f_2(\Psi)$ for $\Psi, \overline{\Psi} \in \mathcal{S}$, i.e.,

$$\begin{cases}\|\nabla f_1(\Psi) - \nabla f_1(\overline{\Psi})\|_2 \leq C_1\|\Psi - \overline{\Psi}\|_2 \\ \|\nabla f_2(\Psi) - \nabla f_2(\overline{\Psi})\|_2 \leq C_2\|\Psi - \overline{\Psi}\|_2,\end{cases}$$

where we dropped dependence on θ for notational simplicity. Consequently, the triangle inequality yields

$$\begin{aligned}&\|\nabla f(\Psi) - \nabla f(\overline{\Psi})\|_2 \\ &\leq \|\nabla f_1(\Psi) - \nabla f_1(\overline{\Psi})\|_2 + \|\nabla f_2(\Psi) - \nabla f_2(\overline{\Psi})\|_2 \\ &\leq (C_1+C_2)\|\Psi - \overline{\Psi}\|_2.\end{aligned}$$

LIPSCHITZ CONSTANT C_1 FOR $\nabla f_1(\Psi)$

Let us examine $f_1(\Psi)$ first. The derivative with respect to $\psi_{j,m}$ is given as

$$\begin{aligned}\frac{\partial f_1(\Psi)}{\partial \psi_{j,m}} &= \sum_{n=1}^N \alpha_{j,n} \cdot \underbrace{\frac{\exp(\psi_{j,m})}{\sigma_\nu^2 + \sum_{j'=1}^J \exp(\psi_{j',m})\alpha_{j',n}}}_{\widetilde{f}(\psi_{j,m})} \\ &= \sum_{n=1}^N \alpha_{j,n}\widetilde{f}(\psi_{j,m}).\end{aligned}$$

We now have

$$\left|\frac{\partial f_1(\Psi)}{\partial \psi_{j,m}} - \frac{\partial f_1(\overline{\Psi})}{\partial \psi_{j,m}}\right| = \sum_{n=1}^N |\alpha_{j,n}| \cdot \left|\widetilde{f}(\psi_{j,m}) - \widetilde{f}(\overline{\psi}_{j,m})\right|.$$

Without loss of generality, we assume $\psi_{j,m} \geq \overline{\psi}_{j,m}$. We now apply the mean value theorem (MVT) to $\widetilde{f}(\psi_{j,m})$

$$\widetilde{f}(\psi_{j,m}) - \widetilde{f}(\overline{\psi}_{j,m}) = \widetilde{f}'(\psi'_{j,m})(\psi_{j,m} - \overline{\psi}_{j,m}),$$

where $\psi'_{j,m} \in [\overline{\psi}_{j,m}, \psi_{j,m}]$.We can compute and bound $\widetilde{f}'(\psi'_{j,m}) = d\widetilde{f}(\psi'_{j,m})/d\psi'_{j,m}$ as follows

$$\begin{aligned}\widetilde{f}'(\psi'_{j,m}) &= \frac{\exp(\psi'_{j,m})}{\sigma_\nu^2 + \sum_{j'=1}^J \exp(\psi'_{j',m})\alpha_{j',n}} \\ &\quad - \frac{\alpha_{j,n}\exp^2(\psi'_{j,m})}{(\sigma_\nu^2 + \sum_{j'=1}^J \exp(\psi'_{j',m})\alpha_{j',n})^2} \\ &= \frac{\exp(\psi'_{j,m})}{\sigma_\nu^2 + \sum_{j'=1}^J \exp(\psi'_{j',m})\alpha_{j',n}} \\ &\quad \times \underbrace{\left(1 - \frac{\alpha_{j,n}\exp(\psi'_{j,m})}{\sigma_\nu^2 + \sum_{j'=1}^J \exp(\psi'_{j',m})\alpha_{j',n}}\right)}_{\leq 1} \\ &\leq \frac{\exp(\psi'_{j,m})}{\sigma_\nu^2 + \sum_{j'=1}^J \exp(\psi'_{j',m})\alpha_{j',n}} \\ &\leq \frac{C_\psi}{\sigma_\nu^2}.\end{aligned}$$

Combining both, we have

$$\begin{aligned}&\sum_{n=1}^N |\alpha_{j,n}| \cdot \left|\widetilde{f}(\psi_{j,m}) - \widetilde{f}(\overline{\psi}_{j,m})\right| \\ &= \sum_{n=1}^N |\alpha_{j,n}| \cdot \left|\widetilde{f}'(\psi'_{j,m})(\psi_{j,m} - \overline{\psi}_{j,m})\right| \\ &\leq \frac{NC_\alpha C_\psi}{\sigma_\nu^2}\left|\psi_{j,m} - \overline{\psi}_{j,m}\right|.\end{aligned}$$

We thus have,

$$\begin{aligned}\|\nabla f_1(\Psi) - \nabla f_1(\overline{\Psi})\|_2^2 &= \sum_{j=1}^J \sum_{m=1}^M \left(\frac{\partial f_1(\Psi)}{\partial \psi_{j,m}} - \frac{\partial f_1(\overline{\Psi})}{\partial \psi_{j,m}}\right)^2 \\ &\leq \left(\frac{JMNC_\alpha C_\psi}{\sigma_\nu^2}\right)^2 \|\Psi - \overline{\Psi}\|_2^2.\end{aligned}$$

LIPSCHITZ CONSTANT C_2 FOR $\nabla f_2(\Psi)$

Computing C_2 proceeds in a similar manner to computing C_1. The derivative with respect to $\psi_{j,m}$ is given as

$$\begin{aligned}&\frac{\partial f_2(\Psi)}{\partial \psi_{j,m}} \\ &= -\sum_{n=1}^N I_{m,n}\alpha_{j,n} \cdot \underbrace{\frac{\exp(\psi_{j,m})}{(\sigma_\nu^2 + \sum_{j'=1}^J \exp(\psi_{j',m})\alpha_{j',n})^2}}_{\widetilde{f}(\psi_{j,m})}.\end{aligned}$$

We now have

$$\begin{aligned}&\left|\frac{\partial f_2(\Psi)}{\partial \psi_{j,m}} - \frac{\partial f_2(\overline{\Psi})}{\partial \psi_{j,m}}\right| \\ &= \sum_{n=1}^N |I_{m,n}\alpha_{j,n}| \cdot \left|-\widetilde{f}(\psi_{j,m}) + \widetilde{f}(\overline{\psi}_{j,m})\right|.\end{aligned}$$

Without loss of generality, assume $\psi_{j,m} \geq \overline{\psi}_{j,m}$. To apply MVT, we need to compute and bound $\widetilde{f}'(\psi'_{j,m})$

$$\begin{aligned}
\widetilde{f}'(\psi'_{j,m}) &= \frac{\exp(\psi'_{j,m})}{(\sigma_\nu^2 + \sum_{j'=1}^{J} \exp(\psi'_{j',m})\alpha_{j',n})^2} \\
&\quad - \frac{2\alpha_{j,n}\exp^2(\psi'_{j,m})}{(\sigma_\nu^2 + \sum_{j'=1}^{J} \exp(\psi'_{j',m})\alpha_{j',n})^3} \\
&= \frac{\exp(\psi'_{j,m})}{(\sigma_\nu^2 + \sum_{j'=1}^{J} \exp(\psi'_{j',m})\alpha_{j',n})^2} \\
&\quad \times \underbrace{\left(1 - \frac{2\alpha_{j,n}\exp(\psi'_{j,m})}{\sigma_\nu^2 + \sum_{j'=1}^{J} \exp(\psi'_{j',m})\alpha_{j',n}}\right)}_{\leq 1} \\
&\leq \frac{\exp(\psi'_{j,m})}{(\sigma_\nu^2 + \sum_{j'=1}^{J} \exp(\psi'_{j',m})\alpha_{j',n})^2} \\
&\leq \frac{C_\psi}{\sigma_\nu^4}.
\end{aligned}$$

Applying MVT,

$$\begin{aligned}
&\sum_{n=1}^{N} |I_{m,n}\alpha_{j,n}| \cdot \left| -\widetilde{f}\left(\psi_{j,m}\right) + \widetilde{f}\left(\overline{\psi}_{j,m}\right) \right| \\
&= \sum_{n=1}^{N} |I_{m,n}\alpha_{j,n}| \cdot \left| \widetilde{f}'(\psi'_{j,m})(\psi_{j,m} - \overline{\psi}_{j,m}) \right| \\
&\leq \frac{NC_I C_\alpha C_\psi}{\sigma_\nu^4} \left| \psi_{j,m} - \overline{\psi}_{j,m} \right|.
\end{aligned}$$

We thus have,

$$\begin{aligned}
&\|\nabla f_2(\Psi) - \nabla f_2(\overline{\Psi})\|_2^2 \\
&= \sum_{j=1}^{J}\sum_{m=1}^{M} \left(\frac{\partial f_2(\Psi)}{\partial \psi_{j,m}} - \frac{\partial f_2(\overline{\Psi})}{\partial \psi_{j,m}} \right)^2 \\
&\leq \left(\frac{JMNC_\alpha C_\psi C_I}{\sigma_\nu^4} \right)^2 \|\Psi - \overline{\Psi}\|_2^2.
\end{aligned}$$

Collecting the Lipschitz constants C_1 and C_2, we finally have

$$\|\nabla f(\Psi) - \nabla f(\overline{\Psi})\|_2 \leq \underbrace{\frac{JMNC_\alpha C_\psi}{\sigma_\nu^2}\left(1 + \frac{C_I}{\sigma_\nu^2}\right)}_{C} \|\Psi - \overline{\Psi}\|_2.$$

H. INFERENCE WITH $p\left(\{\mathbf{z}_j\}_j \mid \{\sigma_{j,m}^2\}_{j,m}, \mathbf{y}, \theta\right)$ *(SECTION 4.2)*

We present the details for performing inference with $p(\{\mathbf{z}_j\}_j \mid \{\widehat{\sigma}_{j,m}^2\}_{j,m}, \mathbf{y}, \hat{\theta})$, given the estimates $\{\widehat{\sigma}_{j,m}^2\}_{j,m}$ and $\hat{\theta}$ from window-level inference. First, we present the Kalman filter/smoother algorithm to compute the mean posterior trajectory, and the credible interval. Next, we present the forward filtering backward sampling (FFBS) algorithm Carter and Kohn [1994], Lindsten and Schön [2013] to generate Monte Carlo (MC) sample trajectories.

First, we define additional notations.

1) $\widetilde{\mathbf{z}}_{j,k|k'} = \mathbb{E}[\widetilde{\mathbf{z}}_{j,k} \mid \{\widehat{\sigma}_{j,m}^2\}_{j,m}, \mathbf{y}_{1:k'}, \hat{\theta}] \in \mathbb{R}^2$

Posterior mean of $\widetilde{\mathbf{z}}_{j,k}$. We are primarily concerned with the following three types: 1) $\widetilde{\mathbf{z}}_{j,k|k-1}$, the one-step prediction estimate, 2) $\widetilde{\mathbf{z}}_{j,k|k}$, the Kalman filter estimate, and 3) $\widetilde{\mathbf{z}}_{j,k|MN}$, the Kalman smoother estimate.

2) $\widetilde{\mathbf{z}}_{k|k'} = [(\widetilde{\mathbf{z}}_{1,k|k'})^{\mathrm{T}}, \ldots, (\widetilde{\mathbf{z}}_{J,k|k'})^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2J}$

A collection of $\{\widetilde{\mathbf{z}}_{j,k|k'}\}_j$ in a single vector.

3) $\mathbf{P}_{j,k|k'} = \mathbb{E}[\left(\widetilde{\mathbf{z}}_{j,k} - \widetilde{\mathbf{z}}_{j,k|k'}\right)\left(\widetilde{\mathbf{z}}_{j,k} - \widetilde{\mathbf{z}}_{j,k|k'}\right)^{\mathrm{T}} \mid \{\widehat{\sigma}_{j,m}^2\}_{j,m}, \mathbf{y}_{1:k'}, \hat{\theta}] \in \mathbb{R}^{2\times 2}$

Posterior covariance of $\widetilde{\mathbf{z}}_{j,k}$. Just as in $\widetilde{\mathbf{z}}_{j,k|k'}$, we are interested in three types, i.e., $\mathbf{P}_{j,k|k-1}$, $\mathbf{P}_{j,k|k}$, and $\mathbf{P}_{j,k|MN}$.

4) $\mathbf{P}_{k|k'} = \mathrm{blkdiag}\left(\mathbf{P}_{j,k|k'}\right) \in \mathbb{R}^{2J\times 2J}$

A block diagonal matrix of J posterior covariance matrices.

5) $\mathbf{A} = \mathrm{blkdiag}\left(\exp\left(-\Delta/l_j\right)\mathbf{R}(\omega_j)\right) \in \mathbb{R}^{2J\times 2J}$

A block digonal transition matrix.

6) $\mathbf{H} = (1, 0, \ldots, 1, 0)$ The observation gain.

KALMAN FILTER/SMOOTHER

The Kalman filter equations are given as

$$\begin{aligned}
&\widetilde{\mathbf{z}}_{j,mN+n|mN+(n-1)} \\
&= \exp\left(-\Delta/l_j\right)\mathbf{R}(\omega_j)\widetilde{\mathbf{z}}_{j,mN+(n-1)|mN+(n-1)} \\
&\mathbf{P}_{j,mN+n|mN+(n-1)} \\
&= \exp\left(-2\Delta/l_j\right)\mathbf{R}(\omega_j)\mathbf{P}_{j,mN+(n-1)|mN+(n-1)}\mathbf{R}^{\mathrm{T}}(\omega_j) \\
&\quad + \sigma_{j,m}^2\left(1 - \exp\left(-2\Delta/l_j\right)\right) \\
&\mathbf{K}_{mN+n} \\
&= \mathbf{P}_{mN+n|mN+(n-1)}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}_{mN+n|mN+(n-1)}\mathbf{H}^{\mathrm{T}} + \sigma_\nu^2\right)^{-1} \\
&\widetilde{\mathbf{z}}_{mN+n|mN+n} \\
&= \widetilde{\mathbf{z}}_{mN+n|mN+(n-1)} \\
&\quad + \mathbf{K}_{mN+n}\left(\mathbf{y}_{mN+n} - \mathbf{H}\widetilde{\mathbf{z}}_{mN+n|mN+(n-1)}\right) \\
&\mathbf{P}_{mN+n|mN+n} = \left(\mathbf{I}_{2J\times 2J} - \mathbf{K}_{mN+n}\mathbf{H}\right)\mathbf{P}_{mN+n|mN+(n-1)}.
\end{aligned}$$

Subsequently, the Kalman smoother equations are given as

$$
\begin{aligned}
&\mathbf{C}_{mN+n} \\
&= \mathbf{P}_{mN+n|mN+n}\mathbf{A}^{\mathrm{T}}\mathbf{P}^{-1}_{mN+(n+1)|mN+n} \in \mathbb{R}^{2J\times 2J} \\
&\widetilde{\mathbf{z}}_{mN+n|MN} \\
&= \widetilde{\mathbf{z}}_{mN+n|mN+n} \\
&\quad + \mathbf{C}_{mN+n}\left(\widetilde{\mathbf{z}}_{mN+(n+1)|MN} - \widetilde{\mathbf{z}}_{mN+(n+1)|mN+n}\right) \\
&\mathbf{P}_{mN+n|MN} \\
&= \mathbf{P}_{mN+n|mN+n} \\
&\quad + \mathbf{C}_{mN+n}\mathbf{P}_{mN+(n+1)|MN}\mathbf{C}^{\mathrm{T}}_{mN+n}. \\
&\quad - \mathbf{C}_{mN+n}\mathbf{P}_{mN+(n+1)|mN+n}\mathbf{C}^{\mathrm{T}}_{mN+n}
\end{aligned}
$$

To obtain the mean reconstructed trajectory for the j^{th} oscillatory component, $\{\widehat{\mathbf{y}}_{j,k}\}_k$, we take the real part of the j^{th} component of the smoothed mean, $\widehat{\mathbf{y}}_{j,k} = \mathbf{e}^{\mathrm{T}}_{2j-1}\cdot\widetilde{\mathbf{z}}_{k|MN}$, where $\mathbf{e}_{2j-1}\in\mathbb{R}^{2J}$ is a unit vector with the only non-zero value, equal to 1, at the entry $2j-1$.

The 95% credible interval for $\widehat{\mathbf{y}}_{j,k}$, denoted as $\mathrm{CI}^{\text{lower}}_{j,k}$/$\mathrm{CI}^{\text{upper}}_{j,k}$ for the upper/lower end, respectively, is given as

$$
\begin{aligned}
\mathrm{CI}^{\text{upper}}_{j,k} &= \mathbf{e}^{\mathrm{T}}_{2j-1}\cdot\widetilde{\mathbf{z}}_{k|MN} + 1.96\cdot\sqrt{\mathbf{e}^{\mathrm{T}}_{2j-1}\mathbf{P}_{k|MN}\mathbf{e}_{2j-1}} \\
\mathrm{CI}^{\text{lower}}_{j,k} &= \mathbf{e}^{\mathrm{T}}_{2j-1}\cdot\widetilde{\mathbf{z}}_{k|MN} - 1.96\cdot\sqrt{\mathbf{e}^{\mathrm{T}}_{2j-1}\mathbf{P}_{k|MN}\mathbf{e}_{2j-1}}.
\end{aligned}
$$

FFBS ALGORITHM FOR $p(\{\mathbf{z}_j\}_j \mid \{\sigma^2_{j,m}\}_{j,m}, \mathbf{y}, \theta)$

To generate the MC trajectory samples from the posterior distribution $p(\{\mathbf{z}_j\}_j \mid \{\widehat{\sigma}^2_{j,m}\}_{j,m}, \mathbf{y}, \hat{\theta})$, we use the FFBS algorithm. The steps are summarized in Algorithm 2, which uses the Kalman estimates derived in the previous section. We denote $s = 1, \ldots, S$ as the MC sample index.

Algorithm 2: FFBS algorithm

Result: $\left\{\widetilde{\mathbf{z}}^{(s)}_k\right\}^{MN,S}_{k,s}$

for $s \leftarrow 1$ **to** S **do**

  Sample $\widetilde{\mathbf{z}}^{(s)}_{MN}$ from $\mathcal{N}\left(\widetilde{\mathbf{z}}_{MN|MN}, \mathbf{P}_{MN|MN}\right)$.

  for $k \leftarrow MN-1$ **to** 1 **do**

    Sample $\widetilde{\mathbf{z}}^{(s)}_k$ from $\mathcal{N}(\widetilde{\mu}_k, \widetilde{\mathbf{P}}_k)$, where

$$
\begin{aligned}
\widetilde{\mu}_k &= \widetilde{\mathbf{z}}_{k|k} + \mathbf{P}_{k|k}\mathbf{A}^{\mathrm{T}}\mathbf{P}^{-1}_{k+1|k}(\widetilde{\mathbf{z}}^{(s)}_{k+1} - \mathbf{A}\widetilde{\mathbf{z}}_{k|k}) \\
\widetilde{\mathbf{P}}_k &= \mathbf{P}_{k|k} - \mathbf{P}_{k|k}\mathbf{A}^{\mathrm{T}}\mathbf{P}^{-1}_{k+1|k}\mathbf{A}\mathbf{P}_{k|k}.
\end{aligned}
$$

  end

end

I. COMPUTATIONAL EFFICIENCY OF PLSO VS. GP-PS

We show the runtime of PLSO and piecewise stationary GP (GP-PS) for inference of the mean trajectory of the hippocampus data ($fs = 1,250$ Hz, $J = 5$, 2-second window) for varying data lengths (50, 100, 200 seconds corresponding to $K = 6.25\times10^4,\ 1.25\times10^5,\ 2.5\times10^5$ sample points, respectively).

As noted in Section 5, the computational complexity of PLSO is $O(J^2K)$, where as the computational complexity of GP-PS is $O(N^2K)$. Since N, the number of samples per window, is fixed ($2,500$ samples), we expect both PLSO and GP-PS to be linear in terms of the number of samples K. Table 1 indeed confirms that this is the case. However, we observe that PLSO is much more efficient than GP-PS.

Table 1: Runtime (s) for PLSO and GP-PS for varying length

	PLSO	**GP-PS**
T =50	1.7	346.8
T =100	3.1	700.6
T =200	6.5	1334.0

J. SIMULATION EXPERIMENT *(SECTION 5.1)*

We simulate from the following model for $1 \le k \le K$

$$
\mathbf{y}_k = 10\left(\frac{K-k}{K}\right)\mathbf{z}^{\Re}_{1,k} + 10\cos^4(2\pi\omega_0 k)\mathbf{z}^{\Re}_{2,k} + \nu_k,
$$

where $\mathbf{z}_{1,k}$ and $\mathbf{z}_{2,k}$ are from the PLSO stationary generative model, i.e., $\sigma^2_{j,m} = \sigma^2_j$. The parameters are $\omega_0/\omega_1/\omega_2 = 0.04/1/10$ Hz, $f_s = 200$ Hz, $T = 100$ seconds, $l_1 = l_2 = 1$, and $\nu_k \sim \mathcal{N}(0, 25)$. This stationary process comprises two amplitude-modulated oscillations, namely one modulated by a slow-frequency ($\omega_0 = 0.04$ Hz) sinusoid and the other a linearly-increasing signal Ba et al. [2014]. We assume a 2-second PS interval. For PLSO, we use $J = 2$ components and 5 block coordinate descent iterations for optimizing θ and $\{\sigma^2_{j,m}\}_{j,m}$.

K. DETAILS OF THE TVAR MODEL

As explained in Section 5, the TVAR model is defined as

$$
\mathbf{y}_k = \sum_{p=1}^{P} a_{p,k}\mathbf{y}_{k-p} + \varepsilon_k,
$$

which can alternatively be written as

$$\begin{pmatrix} \mathbf{y}_k \\ \vdots \\ \mathbf{y}_{k-P+1} \end{pmatrix} = \underbrace{\begin{pmatrix} a_{1,k} & a_{2,k} & \cdots & a_{P-1,k} & a_{P,k} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & \vdots \\ & \vdots & & & \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}}_{A_k} \begin{pmatrix} \mathbf{y}_{k-1} \\ \vdots \\ \mathbf{y}_{k-P} \end{pmatrix} + \varepsilon_k.$$

It is the transition matrix A_k that determines the oscillatory component profile at time k, such as the number of components and the center frequencies. Specifically, $\{A_k\}_k$ are first fit to the data $\mathbf{y}$ and then eigen-decomposition is performed on each of the estimated $\{A_k\}_k$. More in-depth technical details can be found in West et al. [1999].

We use publicly available code for the TVAR implementation[1]. We use TVAR order of $p = 70$, as the models with lower orders than this value did not capture the theta-band signal - The lowest frequency band in these cases were the gamma band (> 30 Hz). Even with the higher orders of p, and various hyperparameter combinations, we observed that the slow and theta frequency band was still explained by a single oscillatory component. For the discount factor, we used $\beta = 0.999$ to ensure that the TVAR coefficients and, consequently, the decomposed oscillatory components do not fluctuate much.

L. ANESTHESIA EEG DATASET *(SECTION 5.3)*

We show spectral analysis results for the EEG data of a subject anesthetized with propofol (This is a different subject from the main text.) The data last $T = 2{,}500$ seconds, sampled at $f_s = 250$ Hz. We assume a 4-second PS interval, use $J = 9$ components and 5 block coordinate descent iterations for optimizing θ and $\{\sigma^2_{j,m}\}_{j,m}$.

Fig. 1 shows the STFT and the PLSO-estimated spectrograms. As noted in the main text, PLSO with stationarity assumption is too restrictive and fails to capture the time-varying spectral pattern. Both PLSO with $\lambda = 0$ and $\lambda = \lambda_{\mathrm{CV}}$ are more effective in capturing such patterns, with the latter able to remove the artifacts and better recover the smoother dynamics.

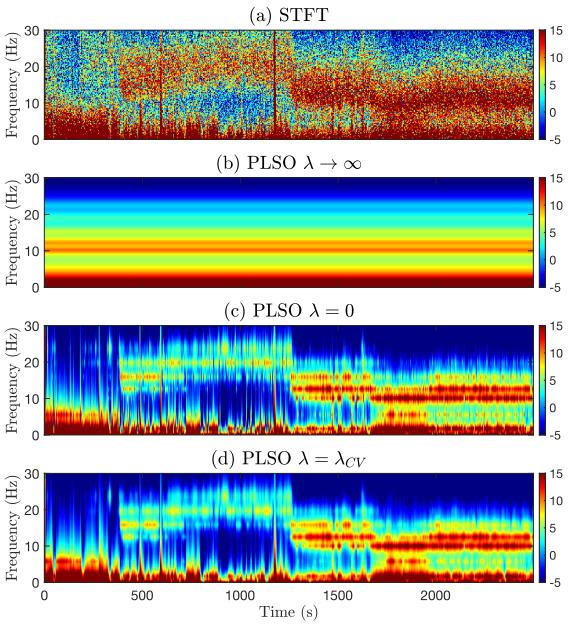

Figure 1: Spectrogram (in dB) under propofol anesthesia. (a) STFT of the data (b) PLSO with $\lambda \to \infty$ (c) PLSO with $\lambda = 0$ (d) PLSO with $\lambda = \lambda_{\mathrm{CV}}$.

References

D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown. Robust spectrotemporal decomposition by iteratively reweighted least squares. *Proceedings of the National Academy of Sciences*, 111(50):E5336–E5345, 2014.

J. Barzilai and J. M. Borwein. Two-Point Step Size Gradient Methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 01 1988.

C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

H. Li and Z. Lin. Accelerated proximal gradient methods for nonconvex programming. In *Advances in Neural Information Processing Systems 28*, pages 379–387. Curran Associates, Inc., 2015.

F. Lindsten and T. B. Schön. *Backward Simulation Methods for Monte Carlo Statistical Inference*. 2013.

M. West, R. Prado, and A. D. Krystal. Evaluation and Comparison of EEG Traces: Latent Structure in Nonstationary Time Series. *Journal of the American Statistical Association*, 94(448):1083–1095, 1999.

[1] https://www2.stat.duke.edu/ mw/mwsoftware/TVAR/index.html