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

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

[1]Electrical Engineering and Computer Sciences, 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

Abstract

To capture the slowly time-varying spectral content of real-world time-series, a common paradigm is to partition the data into approximately stationary intervals and perform inference in the time-frequency domain. However, this approach lacks a corresponding nonstationary time-domain generative model for the entire data and thus, time-domain inference occurs in each interval separately. This results in distortion/discontinuity around interval boundaries and can consequently lead to erroneous inferences based on any quantities derived from the posterior, such as the phase. To address these shortcomings, we propose the Piecewise Locally Stationary Oscillation (PLSO) model for decomposing time-series data with slowly time-varying spectra into several oscillatory, piecewise-stationary processes. PLSO, as a nonstationary time-domain generative model, enables inference on the entire time-series without boundary effects and simultaneously provides a characterization of its time-varying spectral properties. We also propose a novel two-stage inference algorithm that combines Kalman theory and an accelerated proximal gradient algorithm. We demonstrate these points through experiments on simulated data and real neural data from the rat and the human brain.

1 INTRODUCTION

With the collection of long time-series now common, in areas such as neuroscience and geophysics, it is important to develop an *inference* framework for data where the stationarity assumption is too restrictive. We restrict our attention to data 1) with spectral properties that change slowly over time and 2) for which decomposition into several oscillatory components is warranted for interpretation, often the case in electroencephalogram (EEG) or electrophysiology recordings. One can use bandpass filtering [Oppenheim et al., 2009] or methods such as the empirical mode decomposition [Huang et al., 1998, Daubechies et al., 2011] for these purposes. However, due to the *absence* of a generative model, these methods lack a framework for performing inference. Another popular approach is to perform inference in the time-frequency (TF) domain on the short-time Fourier transform (STFT) of the data, assuming stationarity within small intervals. This has led to a rich literature on inference in the TF domain, such as Wilson et al. [2008]. A drawback is that most of these methods focus on estimates for the power spectral density (PSD) and lose important phase information. To recover the time-domain estimates, additional algorithms are required [Griffin and J. Lim, 1984].

This motivates us to explore time-domain generative models that allow time-domain inference and decomposition into oscillatory components. We can find examples based on the stationarity assumption in the signal processing/Gaussian process (GP) communities. A superposition of stochastic harmonic oscillators, where each oscillator corresponds to a frequency band, is used in the processing of speech [Cemgil and Godsill, 2005] and neuroscience data [Matsuda and Komaki, 2017, Beck et al., 2018]. In GP literature [Rasmussen and Williams, 2005], the spectral mixture (SM) kernel [Wilson and Adams, 2013, Wilkinson et al., 2019] models the data as samples from a GP, whose kernel consists of the superposition of localized and frequency-modulated kernels.

These time-domain models can be applied to nonstationary data by partitioning them into stationary intervals and performing time-domain inference within each interval. However, we are faced with a different kind of challenge. As the inference is localized within each interval, the time-domain estimates in different intervals are independent conditioned on the data and do not reflect the dependence across intervals. This also causes discontinuity/distortion of the time-domain estimates near the interval boundaries, and consequently any quantities derived from these estimates.

Preliminary Version

Accepted for the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021).

To address these shortcomings, we propose a generative framework for data with slow time-varying spectra, termed the Piecewise Locally Stationary Oscillation (PLSO) framework[1]. The main contributions are:

Generative model for piecewise stationary, oscillatory components PLSO models time-series as the superposition of piecewise stationary, oscillatory components. This allows time-domain inference on each component and estimation of the time-varying spectra.

Continuity across stationary intervals The state-space model that underlies PLSO strikes a balance between ensuring time-domain continuity across piecewise stationary intervals and stationarity within each interval. Moreover, by imposing stochastic continuity on the interval-level, PLSO learns underlying smooth time-varying spectra accurately.

Inference procedure We propose a two-stage inference procedure for the time-varying spectra and the time-series. By leveraging the Markovian dynamics, the algorithm combines Kalman filter theory [Kalman, 1960] and inexact accelerated proximal gradient approach [Li and Lin, 2015].

In Section 2 we introduce necessary background, followed by the PLSO framework in Section 3. In Section 4, we discuss inference for PLSO. In Section 5, we discuss how PLSO relates to other frameworks. In Section 6, we present experimental results and conclude in Section 7.

2 BACKGROUND

2.1 NOTATION

We use $j \in \{1, \ldots, J\}$ and $k \in \{1, \ldots, K\}$ to denote frequency and discrete-time sample index, respectively. We use $\omega \in [-\pi, \pi]$ for normalized frequency. The j^{th} latent process centered at $\omega = \omega_j$ is denoted as $\mathbf{z}_j \in \mathbb{C}^K$, with $\mathbf{z}_{j,k} \in \mathbb{C}$ denoting the k^{th} sample of $\mathbf{z}_j$ and $\mathbf{z}_{j,k}^{\Re}$, $\mathbf{z}_{j,k}^{\Im}$ its real and imaginary parts. We also represent $\mathbf{z}_{j,k}$ as a $\mathbb{R}^2$ vector, $\widetilde{\mathbf{z}}_{j,k} = [\mathbf{z}_{j,k}^{\Re}, \mathbf{z}_{j,k}^{\Im}]^{\mathrm{T}}$. The elements of $\mathbf{z}_j$ are denoted as $\mathbf{z}_{j,k:k'} = [\mathbf{z}_{j,k}, \ldots, \mathbf{z}_{j,k'}]^{\mathrm{T}}$. The state covariance matrix for $\mathbf{z}_{j,k}$ is defined as $\mathbf{P}_k^j = \mathbb{E}[\widetilde{\mathbf{z}}_{j,k} (\widetilde{\mathbf{z}}_{j,k})^{\mathrm{T}}]$. To express an enumeration of variables, we use $\{\cdot\}$ and drop first/last index for simplicity, e.g. $\{\mathbf{z}_j\}_j$ instead of $\{\mathbf{z}_j\}_{j=1}^J$.

We use $\mathbf{y}_k$ and $\mathbf{z}_{j,k}$ for the discrete-time counterpart of the continuous observation and latent process, $y(t)$ and $z_j(t)$. With the sampling frequency $f_s = 1/\Delta$, we have $\mathbf{y}_k = y(k\Delta)$, $\mathbf{z}_{j,k} = z_j(k\Delta)$, and $T = K\Delta$.

2.2 PIECEWISE LOCAL STATIONARITY

The concept of *piecewise local stationarity* (PLS) for nonstationary time-series with slowly time-varying spectra [Adak, 1998] plays an important role in PLSO. A stationary process has a constant mean and a covariance function which depends only on the difference between two time points.

For our purposes, it suffices to understand the following on PLS: 1) It includes local stationary [Priestley, 1965, Dahlhaus, 1997] and amplitude-modulated stationary processes. 2) A PLS process can be approximated as a piecewise stationary (PS) process (Theorem 1 of [Adak, 1998])

$$z(t) = \sum_{m=1}^{M} \mathbf{1}(u_m \leq t < u_{m+1}) \cdot z^m(t), \qquad (1)$$

where $z^m(t)$ is a continuous stationary process and the boundaries are $0 = u_1 < \cdots < u_{M+1} = T$. Note that Eq. 1 does not guarantee continuity across different PS intervals,

$$\lim_{t \to u_m^-} z(t) = \lim_{t \to u_m^-} z^{m-1}(t) \neq \lim_{t \to u_m^+} z^m(t) = \lim_{t \to u_m^+} z(t).$$

3 THE PLSO MODEL AND ITS MATHEMATICAL PROPERTIES

Building on the Theorem 1 of [Adak, 1998], PLSO models nonstationary data as PS processes. It is a superposition of J different PS processes $\{\mathbf{z}_j\}_j$, with $\mathbf{z}_j$ corresponding to an oscillatory process centered at frequency ω_j. PLSO also guarantees stochastic continuity across PS intervals. We show that piecewise stationarity and continuity across PS intervals are two competing objectives and that PLSO strikes a balance between them, as discussed in Section 3.2.

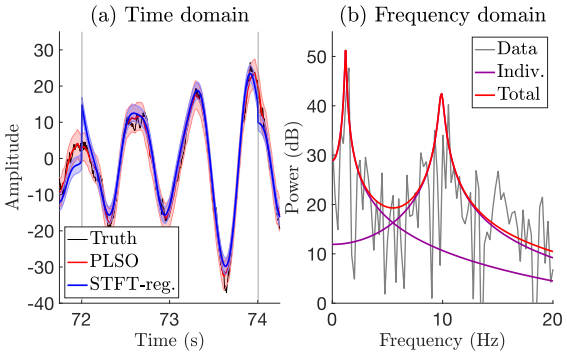

Figure 1: A simulated example. (a) Time domain. Data (black) around boundaries (gray) and inferred oscillatory components using PLSO (red) and regularized STFT (blue). (b) Frequency domain. Spectrum of the data (gray), PLSO components for $J = 2$ (purple) and their sum (red).

Fig. 1 shows an example of the PLSO framework applied to simulated data. In the time domain, the oscillation inferred using the regularized STFT (blue) [Kim et al., 2018], which imposes stochastic continuity on the STFT coefficients, suffers from discontinuity/distortion near window boundaries,

[1]Code is available at https://github.com/andrewsong90/plso.git

Preliminary version

whereas that inferred by PLSO (red) does not. In the frequency domain, each PLSO component corresponds to a localized spectrum $S_j(\omega)$, the sum of which is the PSD $\gamma(\omega)$, and is fit to the data STFT (or periodogram) $I(\omega)$. We start by introducing the PLSO model for a single window.

3.1 PLSO FOR STATIONARY DATA

As a building block for PLSO, we use the discrete stochastic harmonic oscillator for a stationary time series [Qi et al., 2002, Cemgil and Godsill, 2005, Matsuda and Komaki, 2017]. The data $\mathbf{y}$ are assumed to be a superposition of J *independent* zero-mean components

$$\begin{aligned} \widetilde{\mathbf{z}}_{j,k} &= \rho_j \mathbf{R}(\omega_j)\widetilde{\mathbf{z}}_{j,k-1} + \varepsilon_{j,k} \\ \mathbf{y}_k &= \sum_{j=1}^{J} \mathbf{z}^{\Re}_{j,k} + \nu_k, \end{aligned} \tag{2}$$

where $\mathbf{R}(\omega_j) = \begin{pmatrix} \cos(\omega_j) & -\sin(\omega_j) \\ \sin(\omega_j) & \cos(\omega_j) \end{pmatrix}$, $\varepsilon_{j,k} \sim \mathcal{N}(0, \alpha_j \mathbf{I}_{2\times 2})$, and $\nu_k \sim \mathcal{N}(0, \sigma_\nu^2)$, correspond to the rotation matrix, the state noise, and the observation noise, respectively. The imaginary component $\mathbf{z}^{\Im}_{j,k}$, which does not directly contribute to $\mathbf{y}_k$, can be seen as the auxiliary variable to write $\mathbf{z}_j$ in recursive form using $\mathbf{R}(\omega_j)$ [Koopman and Lee, 2009]. We assume $\mathbf{P}_1^j = \sigma_j^2 \cdot \mathbf{I}_{2\times 2} \ \forall j$.

We reparameterize ρ_j and α_j, using lengthscale l_j and power σ_j^2, such that $\rho_j = \exp(-\Delta/l_j)$ and $\alpha_j = \sigma_j^2(1-\rho_j^2)$. Theoretically, this establishes a connection to 1) the stochastic differential equation [Brown et al., 2004, Solin and Särkkä, 2014], detailed in **Appendix A**, and 2) a superposition of frequency-modulated and localized GP kernels, similar to SM kernel [Wilson and Adams, 2013]. Practically, this ensures that $\rho_j < 1$, and thus stability of the process.

Given Eq. 2, we can readily express the frequency spectra of PLSO in each interval, through the autocovariance function. The autocovariance of $\mathbf{z}_j$ is given as $Q_j(n') = \mathbb{E}[\mathbf{z}^{\Re}_{j,k}\mathbf{z}^{\Re}_{j,k+n'}] = \sigma_j^2 \cos(\omega_j n') \exp(-n'\Delta/l_j)$. It can also be thought of as an *exponential* kernel, frequency-modulated by ω_j. The spectra for $\mathbf{z}_j$, denoted as $S_j(\omega)$, is obtained by taking the Fourier transform (FT) of $Q_j(n')$

$$S_j(\omega) = \sum_{n'=-\infty}^{\infty} Q_j(n') \exp(-i\omega n') = \varphi_j(\omega) + \varphi_j(-\omega)$$

$$\varphi_j(\omega) = \frac{\sigma_j^2 \left(1 - \exp(-2\Delta/l_j)\right)}{1 + \exp(-2\Delta/l_j) - 2\exp(-\Delta/l_j)\cos(\omega - \omega_j)},$$

with the detailed derivation in **Appendix B**.

Given $S_j(\omega)$, we can show that PSD $\gamma(\omega)$ of the entire process $\sum_{j=1}^J \mathbf{z}_j$ is simply $\gamma(\omega) = \sum_{j=1}^J S_j(\omega)$. First, since $\mathbf{z}_j$ is independent across j, the autocovariance can be simplified, i.e., $\mathbb{E}[\sum_j \mathbf{z}^{\Re}_{j,k} \sum_j \mathbf{z}^{\Re}_{j,k+n'}] = \sum_j \mathbb{E}[\mathbf{z}^{\Re}_{j,k}\mathbf{z}^{\Re}_{j,k+n'}]$. Next, using the linearity of FT, we can conclude that $\gamma(\omega)$ is a superposition of individual spectra.

3.2 PLSO FOR NONSTATIONARY DATA

If $\mathbf{y}$ is nonstationary, we can still apply stationary PLSO of Eq. 2 for the time-domain inference. However, this implies constant spectra for the entire data ($S_j(\omega)$ and $\gamma(\omega)$ do not depend on k), which is not suitable for nonstationary time-series for which we want to track spectral dynamics. This point is further illustrated in Section 6.

We therefore segment $\mathbf{y}$ into M non-overlapping PS intervals, indexed by $m \in \{1, \ldots, M\}$, of length N, indexed by $n \in \{1, \ldots, N\}$, such that $K = MN$. We then apply the stationary PLSO to each interval, with additional Markovian dynamics imposed on $\sigma^2_{j,m}$,

$$\begin{aligned} \log(\sigma^2_{j,m}) &= \log(\sigma^2_{j,m-1}) + \eta_{j,m} \\ \widetilde{\mathbf{z}}_{j,mN+n} &= \rho_j \mathbf{R}(\omega_j)\widetilde{\mathbf{z}}_{j,mN+(n-1)} + \varepsilon_{j,mN+n} \\ \mathbf{y}_{mN+n} &= \sum_{j=1}^{J} \mathbf{z}^{\Re}_{j,mN+n} + \nu_{mN+n}, \end{aligned} \tag{3}$$

where $\varepsilon_{j,mN+n} \sim \mathcal{N}(0, \sigma^2_{j,m}(1-\rho_j^2)\mathbf{I}_{2\times 2})$, $\eta_{j,m} \sim \mathcal{N}(0, 1/\sqrt{\lambda})$ and $\nu_{mN+n} \sim \mathcal{N}(0, \sigma_\nu^2)$. We define $\mathbf{P}^j_{m,n}$ as the covariance of $\widetilde{\mathbf{z}}_{j,mN+n}$, with $\mathbf{P}^j_{1,1} = \sigma^2_{j,1}\mathbf{I}_{2\times 2}, \forall j$. The graphical model is shown in Fig. 2.

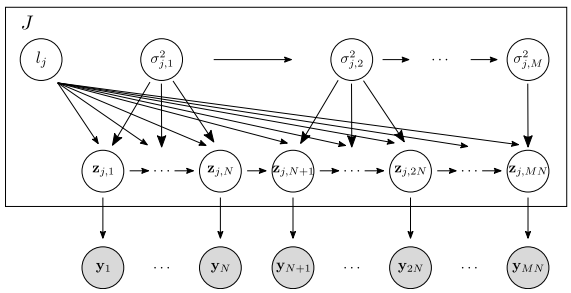

Figure 2: The graphical model for PLSO.

We can understand PLSO as providing a *parameterized* spectrogram defined by $\theta = \{\lambda, \sigma_\nu^2, \{l_j\}_j, \{\omega_j\}_j\}$ and $\{\sigma^2_{j,m}\}_{j,m}$ of the time-domain generative model. The lengthscale l_j controls the bandwidth of the j^{th} process, with *larger* l_j corresponding to *narrower* bandwidth. The variance $\sigma^2_{j,m}$ controls the power of $\mathbf{z}_j$ and changes across different intervals, resulting in time-varying spectra $S_j^{(m)}(\omega)$ and PSD $\gamma^{(m)}(\omega)$. The center frequency ω_j, and $-\omega_j$, at which $S_j^{(m)}(\omega)$ is maximized, controls the modulation frequency.

As discussed previously, the segmentation approach for nonstationary time-series produces distortion/discontinuity artifacts around interval boundaries - The PLSO, as described by Eq. 3, resolves these issues gracefully. We now analyze two mathematical properties, *stochastic continuity* and

Preliminary version

piecewise stationarity, to gain more insights on how PLSO accomplishes this.

3.2.1 Stochastic continuity

We discuss two types of stochastic continuity, 1) across the interval boundaries and 2) on $\{\sigma^2_{j,m}\}$.

Continuity across the interval boundaries In PLSO, the state-space model (Eq. 3) provides stochastic continuity across different PS intervals. The following proposition rigorously explains stochastic continuity for PLSO.

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,*

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

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. We use the connection between PLSO, which is a discrete-time model, and its continuous-time counterpart. Details are in the **Appendix C**. □

This matches our intuition that as $\Delta \to 0$, the adjacent samples from the same process should coverge to each other. For PS approaches without the sample-level continuity, even with the interval-level constraint [Rosen et al., 2009, Kim et al., 2018, Song et al., 2018, Das and Babadi, 2018, Soulat et al., 2019], convergence is not guaranteed.

We can interpret the continuity in the context of posterior for $\mathbf{z}_j$. For PS approaches without continuity, we have

$$p(\{\mathbf{z}_j\}_j \,|\mathbf{y}) \propto \prod_{m=1}^{M} p(\{\mathbf{z}_{j,(m-1)N+1:mN}\}_j |\mathbf{y}_{(m-1)N+1:mN}), \tag{4}$$

where $\{\sigma^2_{j,m}\}_{j,m}$ and θ are omitted for notational ease. This is due to $p(\{\mathbf{z}_j\}_j) = \prod_{m=1}^{M} p(\{\mathbf{z}_{j,(m-1)N+1:mN}\}_j)$, as a result of *absence* of continuity across the intervals. Consequently, the inferred time-domain estimates are conditionally independent across the intervals. On the contrary, in PLSO, the time-domain estimates depend on the entire $\mathbf{y}$, not just on a subset.

Continuity on $\sigma^2_{j,m}$ For a given j, we impose stochastic continuity on $\log \sigma^2_{j,m}$. Effectively, this pools together estimates of $\{\sigma^2_{j,m}\}_m$ to 1) prevent overfitting to the noisy data spectra and 2) estimate smooth dynamics of $\{\sigma^2_{j,m}\}_m$. The use of $\log \sigma^2_{j,m}$ ensures that $\sigma^2_{j,m}$ is non-negative.

The choice of λ dictates the smoothness of $\{\sigma^2_{j,m}\}_m$, with the two extremes corresponding to the familiar dynamics. If $\lambda \to 0$, we treat each window independently. If $\lambda \to \infty$, we treat the data as stationary, as the constraint forces $\sigma^2_{j,m} = \sigma^2_j, \forall m$. Practically, the smooth regularization prevents artifacts in the spectral analysis, arising from sudden motion or missing data, as demonstrated in Section 6.

3.2.2 Piecewise stationarity

For the m^{th} window to be piecewise stationary, the initial state covariance matrix $\mathbf{P}^j_{m,1}$ should be the *steady-state* covariance matrix for the window, denoted as $\mathbf{P}^j_{m,\infty}$.

The challenge is transitioning from $\mathbf{P}^j_{m,\infty}$ to $\mathbf{P}^j_{m+1,\infty}$. Specifically, to ensure $\mathbf{P}^j_{m+1,1} = \mathbf{P}^j_{m+1,\infty}$, given that $\mathbf{P}^j_{m,N} = \mathbf{P}^j_{m,\infty} \neq \mathbf{P}^j_{m+1,\infty}$, the *variance* of the process noise between the two samples, $\varepsilon_{j,(m+1)N+1}$, has to equal $\mathbf{P}^j_{m+1,\infty} - \exp(-2\Delta/l_j)\,\mathbf{P}^j_{m,\infty}$. However, this is infeasible. If $\mathbf{P}^j_{m+1,\infty} < \mathbf{P}^j_{m,\infty}$, the variance is negative. Even if it were positive, the limit as $\Delta \to 0$ does not equal zero, i.e., $\mathbf{P}^j_{m+1,\infty} - \mathbf{P}^j_{m,\infty}$. As a result, the Proposition 1 no longer holds and the trajectory is discontinuous.

In summary, there exists a *trade-off* between maintaining piecewise stationarity and continuity across intervals. PLSO maintains continuity across the intervals while ensuring that the state covariance quickly transitions to the steady-state covariance. We quantify the speed of transition in the following proposition.

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

$$\mathbf{P}^j_{m+1,n} = \mathbf{P}^j_{m+1,\infty} + \exp(-\frac{2n\Delta}{l_j})(\mathbf{P}^j_{m,\infty} - \mathbf{P}^j_{m+1,\infty}).$$

Proof. In **Appendix D**, we prove this result by induction. □

This implies that, except for the transition portion at the beginning of each window, we can assume stationarity. In practice, we additionally impose an upper bound on l_j during estimation and also use a reasonably-large N. Empirically, we observe that the transition period has little impact.

4 INFERENCE

Given the generative model in Eq. 3, our goal is to perform inference on the posterior distribution

$$p(\{\mathbf{z}_j\}_j, \{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y}, \theta) = \underbrace{p(\{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y}, \theta)}_{\text{window-level posterior}} \cdot \underbrace{p(\{\mathbf{z}_j\}_j \mid \{\sigma^2_{j,m}\}_{j,m}, \mathbf{y}, \theta)}_{\text{sample-level posterior}}. \tag{5}$$

We can learn θ or fix the parameters to specific values informed by domain knowledge, such as the center frequency

Preliminary version

or bandwidth of the processes. Since the posterior distribution factorizes into two terms as in Eq. 5, the *window-level* posterior $p(\{\sigma^2_{j,m}\}_{j,m}|\mathbf{y},\theta)$ and the *sample-level* posterior $p(\{\mathbf{z}_j\}_j|\{\sigma^2_{j,m}\}_{j,m},\mathbf{y},\theta)$, we break the inference into two stages.

Stage 1 We minimize the *window-level* negative log-posterior, with respect to θ and $\{\sigma^2_{j,m}\}_{j,m}$. Specifically, we obtain maximum likelihood (ML) estimate $\hat{\theta}_{\text{ML}}$ and maximum a posteriori (MAP) estimate $\{\widehat{\sigma}^2_{j,m,\text{MAP}}\}_{j,m}$. We drop subscripts ML and MAP for notational simplicity.

Stage 2 Given $\{\widehat{\sigma}^2_{j,m}\}_{j,m}$ and $\hat{\theta}$, we perform inference on the *sample-level* posterior. This includes computing the mean $\widehat{\mathbf{z}}_j = \mathbb{E}[\mathbf{z}_j|\{\widehat{\sigma}^2_{j,m}\}_{j,m},\mathbf{y},\hat{\theta}]$ and credible intervals, which quantifies the uncertainty of the estimates, or any statistical quantity derived from the posterior distribution.

4.1 OPTIMIZATION OF $\{\sigma^2_{j,m}\}_{j,m}$ AND θ

We factorize the posterior, $p(\{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y},\theta) \propto p(\mathbf{y} \mid \{\sigma^2_{j,m}\}_{j,m},\theta)\cdot p(\{\sigma^2_{j,m}\}_{j,m} \mid \theta)$. As the exact inference is intractable, we instead minimize the negative log-posterior, $-\log p(\{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y},\theta)$. This is an *empirical Bayes* approach [Casella, 1985], since we estimate $\{\sigma^2_{j,m}\}_{j,m}$ using the marginal likelihood $p(\mathbf{y} \mid \{\sigma^2_{j,m}\}_{j,m},\theta)$. The smooth hyperprior provides the MAP estimate for $\{\sigma^2_{j,m}\}_{j,m}$.

We use the *Whittle likelihood* [Whittle, 1953], defined for stationary time-series in the frequency domain, for the log-likelihood $f(\{\sigma^2_{j,m}\}_{j,m};\theta) = \log p(\mathbf{y} \mid \{\sigma^2_{j,m}\}_{j,m},\theta)$,

$$f(\{\sigma^2_{j,m}\}_{j,m};\theta) = -\frac{1}{2}\sum_{m,n=1}^{M,N}\log(\gamma^{(m)}(\omega_n)+\sigma^2_\nu)+\frac{I^{(m)}(\omega_n)}{\gamma^{(m)}(\omega_n)+\sigma^2_\nu}, \tag{6}$$

where the log-likelihood is the sum of the Whittle likelihood computed for each interval, with discrete frequency $\omega_n = 2\pi n/N$, and data STFT $I^{(m)}(\omega_n) = \left|\sum_{n'=1}^{N}\exp\left(-2\pi i(n'-1)(n-1)/N\right)\mathbf{y}_{mN+n'}\right|^2$. The Whittle likelihood, which is nonconvex, enables frequency-domain parameter estimation as a computationally more efficient alternative to the time domain estimation [Turner and Sahani, 2014]. The concave log-prior $g(\{\sigma^2_{j,m}\}_{j,m};\theta) = \log p(\{\sigma^2_{j,m}\}_{j,m} \mid \theta)$, which arises from the continuity on $\{\sigma^2_{j,m}\}_{j,m}$, is given as

$$g(\{\sigma^2_{j,m}\}_{j,m};\theta) = -\frac{\lambda}{2}\sum_{j=1}^{J}\sum_{m=1}^{M}\left(\log\sigma^2_{j,m}-\log\sigma^2_{j,m-1}\right)^2. \tag{7}$$

This yields the following nonconvex problem

$$\begin{aligned}&\min_{\{\sigma^2_{j,m}\}_{j,m},\theta} -\log p\left(\{\sigma^2_{j,m}\}_{j,m} \mid \mathbf{y},\theta\right)\\ &= \min_{\{\sigma^2_{j,m}\}_{j,m},\theta} -f(\{\sigma^2_{j,m}\}_{j,m};\theta) - g(\{\sigma^2_{j,m}\}_{j,m};\theta).\end{aligned} \tag{8}$$

We optimize Eq. 8 by block coordinate descent [Wright, 2015] on $\{\sigma^2_{j,m}\}_{j,m}$ and $\{\sigma^2_\nu,\{l_j\}_j,\{\omega_j\}_j\}$. For σ^2_ν, $\{l_j\}_j$, and $\{\omega_j\}_j$, we minimize $-f(\{\sigma^2_{j,m}\}_{j,m};\theta)$, since $g(\{\sigma^2_{j,m}\}_{j,m};\theta)$ does not affect them. We perform conjugate gradient descent on $\{l_j\}_j$ and $\{\omega_j\}_j$. We discuss initialization and how to estimate σ^2_ν in the **Appendix E**.

4.1.1 Optimization of $\{\sigma^2_{j,m}\}_{j,m}$

We introduce an algorithm to compute a local optimal solution of $\{\sigma^2_{j,m}\}_{j,m}$ for the nonconvex optimization problem in Eq. 8, by leveraging the regularized temporal structure of $\{\sigma^2_{j,m}\}_{j,m}$. It extends the inexact accelerated proximal gradient (APG) method [Li and Lin, 2015], by solving the proximal step with a Kalman filter/smoother [Kalman, 1960]. This follows since computing the proximal operator for $g(\{\sigma^2_{j,m}\}_{j,m};\theta)$ is equivalent to MAP estimation for J independent 1-dimensional linear Gaussian state-space models

$$\begin{aligned}\{\log\sigma^{(l+1),2}_{j,m}\}_{j,m} &= \text{prox}_{-\alpha^{(l)}g}(\mathbf{v}^{(l)})\\ &= \underset{\{\sigma^2_{j,m}\}_{j,m}}{\arg\min}\ \underbrace{\frac{\sum_{j,m}^{J,M}(\mathbf{v}^{(l)}_{j,m}-\log\sigma^2_{j,m})^2}{2\alpha^{(l)}}-g(\{\sigma^2_{j,m}\}_{j,m})}_{\sum_{j=1}^{J}q_j}\end{aligned} \tag{9}$$

where $\alpha^{(l)} > 0$ is a step-size for the l^{th} iteration, $q_j = \sum_{m=1}^{M}\frac{(\mathbf{v}^{(l)}_{j,m}-\log\sigma^2_{j,m})^2}{2\alpha^{(l)}}+\frac{\lambda}{2}(\log\sigma^2_{j,m}-\log\sigma^2_{j,m-1})^2$, and $\mathbf{v}^{(l)}_{j,m} = \log\sigma^{(l),2}_{j,m}+\alpha^{(l)}\frac{\partial f(\{\sigma^2_{j,m}\}_{j,m})}{\partial\log\sigma^2_{j,m}}\Big|_{\{\sigma^{(l),2}_{j,m}\}_{j,m}}$.

The j^{th} optimization problem, $\min_{\{\sigma^2_{j,m}\}_m} q_j$, is equivalent to estimating the mean of the posterior for $\{\log\sigma^2_{j,m}\}_m$ in a linear Gaussian state-space model with observations $\{\mathbf{v}^{(l)}_{j,m}\}_m$, observation noise variance $\alpha^{(l)}$, and state variance $1/\lambda$. Therefore, the solution can efficiently be computed with J 1-dimensional, Kalman filters/smoothers, with the computational complexity of $O(JM)$.

Note that Eq. 9 holds for all non-negative λ. If $\lambda = 0$, the proximal operator is an identity operator, as $\log\sigma^{(l+1),2}_{j,m} = \mathbf{v}^{(l)}_{j,m}$. This is a gradient descent with a step-size rule. If $\lambda \to \infty$, we have $\log\sigma^2_{j,m} = \log\sigma^2_{j,m-1}$, $\forall m$, which leads to $\log\sigma^{(l+1),2}_{j,m} = (1/M)\sum_{m=1}^{M}\mathbf{v}^{(l)}_{j,m}$. The algorithm is guaranteed to converge when $\alpha^{(l)} < 1/C$, where C is the Lipschitz constant for $f(\{\sigma^2_{j,m}\}_{j,m};\theta)$. In practice, we

Preliminary Version

select $\alpha^{(l)}$ according to the step-size rule [Barzilai and Borwein, 1988]. In **Appendix F & G**, we present the full algorithm for optimizing $\{\sigma^2_{j,m}\}_{j,m}$ and a derivation for C.

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

We perform inference on the posterior distribution $p(\{\mathbf{z}_j\}_j \mid \{\widehat{\sigma}^2_{j,m}\}_{j,m}, \mathbf{y}, \hat{\theta})$. Since this is a Gaussian distribution, the mean trajectories $\{\widehat{\mathbf{z}}_j\}_j$ and the credible intervals can be computed analytically. Moreover, Eq. 3 is a linear Gaussian state-space model, we can use Kalman filter/smoother for efficient computation with computational complexity $O(J^2K)$, further discussed In **Appendix H**. Since we use the point estimate $\{\widehat{\sigma}^2_{j,m}\}_{j,m}$, the credible interval for $\{\widehat{\mathbf{z}}_j\}_j$ will be *narrower* compared to the full Bayesian setting which accounts for all values of $\{\sigma^2_{j,m}\}_{j,m}$.

4.2.1 Monte Carlo Inference

We can also obtain posterior samples and perform Monte Carlo (MC) inference on any posterior-derived quantity. To generate the MC trajectory samples, we use the forward-filter backward-sampling (FFBS) algorithm [Carter and Kohn, 1994]. Assuming S number of MC samples, the computational complexity for FFBS is $O(SJ^2K)$, since for each sample, the algorithm uses Kalman filter/smoother for sampling. This is different from generating samples with the interval-specific posterior in Eq. 4. In the latter case, the FFBS algorithm is run M times, the samples of which have to be concatenated to form an entire trajectory. With PLSO, the trajectory sample is conditioned on the entire observation and is continuous across the intervals.

One quantity of interest is the *phase*. We obtain the phase as $\phi_{j,k} = \tan^{-1}(\mathbf{z}^{\Im}_{j,k}/\mathbf{z}^{\Re}_{j,k})$. Since $\tan^{-1}(\cdot)$ is a nonlinear operation, we compute the mean and credible interval with MC samples through the FFBS algorithm. Given the posterior-sampled trajectories $\{\mathbf{z}^{(s)}_j\}_{j,s}$, where $s \in \{1, \ldots, S\}$ denotes MC sample index, we estimate $\widehat{\phi}_{j,k} = (1/S)\sum_{s=1}^{S} \tan^{-1}(\mathbf{z}^{\Im,(s)}_{j,k}/\mathbf{z}^{\Re,(s)}_{j,k})$, and use empirical quantiles for the associated credible interval.

4.3 CHOICE OF J AND λ

We choose J that minimizes the Akaike Information Criterion (AIC) [Akaike, 1981], defined as

$$\mathrm{AIC}(J) = -(2/M) \cdot \log p(\mathbf{y} \mid \{\widehat{\sigma}^2_{j,m}\}_{j,m}, \hat{\theta}) + 2 \cdot 3 \cdot J, \tag{10}$$

where $3 \cdot J$ corresponds to the number of parameters $(\{l_j\}_j, \{\omega_j\}_j, \{\sigma^2_{j,m}\}_j)$. The regularization parameter λ is determined through a two-fold cross-validation, where each fold is generated by aggregating even/odd sample indices [Ba et al., 2014].

4.4 CHOICE OF WINDOW LENGTH N

The choice of window length N presents the tradeoff between 1) spectral resolution and 2) the temporal resolution of the spectral dynamics [Oppenheim et al., 2009]. For a shorter window, the estimated spectral dynamics have a *finer* temporal resolution, *coarser* spectral resolution, and *higher* variance. For a longer window, these trends are reversed. This suggests that the choice is application-dependent. For electrophysiology data, a window on the order of seconds is used, as scientific interpretations are made on the basis of fine spectral resolution (< 1Hz). For audio signal processing [Gold et al., 2011], short windows ($10 \sim 100$ ms) are used, since audio data is highly nonstationary and thus requires fine temporal resolution for processing. A survey of window lengths used in different applications can be found in the **supporting information** of Kim et al. [2018].

5 RELATED WORKS

We examine how PLSO relates to other nonstationary frameworks.

STFT/Regularized STFT In STFT, the harmonic basis is used, whereas quasi-periodic components are used for PLSO, which allows capturing of broader spectral content. Recent works regularize STFT coefficients with stochastic continuity across the windows, to infer smooth spectral dynamics [Ba et al., 2014, Kim et al., 2018]. However, this regularization leads to discontinuities at window boundaries.

Piecewise stationary GP GP regression and parameter estimation are performed within each interval [Gramacy and Lee, 2008, Solin et al., 2018]. Consequently, the recovered trajectories are discontinuous. Also, the inversion of covariance matrix leads to an expensive inference. For example, the time-complexity of mean trajectory estimation is $O(MN^3) = O(N^2K)$, whereas the time-complexity for PLSO is $O(J^2K)$. Considering that the typical sampling frequency for electrophysiology data is $\sim 10^3$ (Hz) and windows are several seconds, which leads to $N \sim 10^3$, PLSO is computationally more efficient. In the **Appendix I**, we confirm this through an experiment.

Time-varying Autoregressive model (TVAR) The TVAR model [Kitagawa and Gersch, 1985] is given as

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

with the time-varying coefficients $\{a_{p,k}\}_p$. Consequently, it does not suffer from discontinuity issues. TVAR can also be *approximately* decomposed into oscillatory components via eigen-decomposition of the transition matrix [West et al., 1999]. However, since the eigen-decomposition changes at every sample, this leads to an ambiguous interpretation of the oscillations in the data, as we discuss in Section 6.

Preliminary version

RNN frameworks Despite the popularity of recurrent neural networks (RNN) for time-series applications [Goodfellow et al., 2016], we believe PLSO is more appropriate for time-frequency analysis for two reasons.

1. RNNs operate in the time-domain with the goal of prediction/denoising and consequently less emphasis is placed on local stationarity or estimation of second-order statistics. Performing time-frequency analysis requires segmenting the RNN outputs and applying the STFT, which yields noisy spectral estimates.
2. RNN is not a generative framework. Although variational framework can be combined with RNN [Chung et al., 2015, Krishnan et al., 2017], the use of variational lower bound objective could lead to suboptimal results. On the other hand, PLSO is a generative framework that maximizes the true log-posterior.

6 EXPERIMENTS

We apply PLSO to three settings: 1) A simulated dataset 2) local-field potential (LFP) data from the rat hippocampus, and 3) EEG data from a subject under anesthesia.

We use PLSO with $\lambda = 0$, $\lambda \to \infty$, and λ determined by cross-validation, λ_{CV}. We use interval lengths chosen by domain experts. As baselines, we use 1) regularized STFT (STFT-reg.) and 2) Piecewise stationary GP (GP-PS). For GP-PS, we use the same $\{\widehat{\sigma}^2_{j,m}\}_{j,m}$ and $\widehat{\theta}$ as PLSO with $\lambda = 0$, so that the estimated PSD of GP-PS and PLSO are identical. This lets us explain differences in time-domain estimates by the fact that PLSO operates in the time-domain.

6.1 SIMULATED DATASET

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

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

where $\mathbf{z}_{1,k}$ and $\mathbf{z}_{2,k}$ are as in Eq. 2, with $(\omega_0, \omega_1, \omega_2) = (0.04, 1, 10)$ 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 sinusoid and the other a linearly-increasing signal [Ba et al., 2014]. We simulate 20 realizations and train on each realization, assuming 2-second PS intervals. For PLSO, we use $J = 2$. Additional details are provided in the **Appendix J**.

Results We use two metrics: 1) Mean squared error (MSE) between the mean estimate $\hat{\mathbf{z}}_j$ and the ground truth $\mathbf{z}^{\text{True}}_j$ and 2) $\text{jump}(\mathbf{z}_j)$. The averaged results are shown in Table 1. We define $\text{jump}(\mathbf{z}_j) = \frac{1}{M-1}\sum_{m=1}^{M-1}|\hat{\mathbf{z}}_{j,mN+1} - \hat{\mathbf{z}}_{j,mN}|$ to be the level of discontinuity at the interval boundaries. If $\text{jump}(\mathbf{z}_j)$ greatly exceeds $\text{jump}(\mathbf{z}^{\text{True}}_j)$, this implies the existence of large discontinuities at the boundaries.

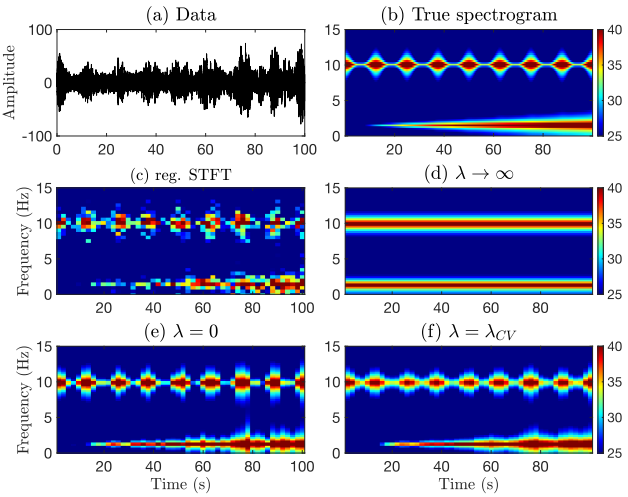

Figure 3: Spectrograms for simulation (in dB). (a) True data (b) True spectrogram (c) regularized STFT (d) PLSO with $\lambda \to \infty$ (e) PLSO with $\lambda = 0$ (f) PLSO with $\lambda = \lambda_{\text{CV}}$.

Table 1: Simulation results. For $\text{jump}(\mathbf{z}_j)$ and MSE, left/right metrics correspond to $\mathbf{z}_1$/$\mathbf{z}_2$, respectively.

	$\text{jump}(\mathbf{z}_j)$	MSE	IS div.
Truth	0.95/12.11	0/0	0
$\lambda = 0$	0.26/10.15	2.90/3.92	4.08
$\lambda \to \infty$	0.22/10.32	3.26/4.53	13.78
$\lambda = \lambda_{\text{CV}}$	0.25/10.21	**2.88/3.91**	**3.93**
STFT-reg.	49.59/81.00	6.89/10.68	N/A
GP-PS	16.99/23.28	3.00/4.04	4.08

Fig. 3 shows the true data in the time domain and spectrogram results. Fig. 3(c) shows that although the regularized STFT detects activities around 1 and 10 Hz, it fails to delineate the time-varying spectral pattern. Fig. 3(d) shows that PLSO with stationarity ($\lambda \to \infty$) assumption is too restrictive. Fig. 3(e), (f) show that both PLSO with independent window assumption ($\lambda = 0$) and PLSO with cross-validated λ ($\lambda = \lambda_{\text{CV}}$) are able to capture the dynamic pattern, with the latter being more effective in recovering the smooth dynamics across different PS intervals.

For GP-PS and STFT-reg., $\text{jump}(\mathbf{z}_j)$ exceeds $\text{jump}(\mathbf{z}^{\text{True}}_j)$, indicating discontinuities at the boundaries. An example is in Fig. 1. PLSO produces a similar jump metric as the ground truth metric, indicating the absence of discontinuities. We attribute the lower value to Kalman smoothing. For the TF domain, we use Itakura-Saito (IS) divergence [Itakura and Saito, 1970] as a distance measure between the ground truth spectra and the PLSO estimates. That the highest di-

Preliminary

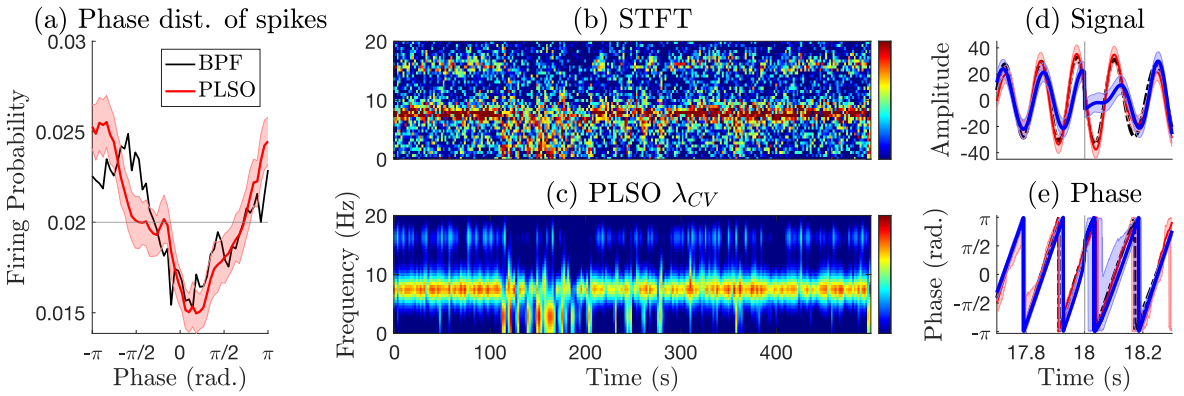

Figure 4: Result of analyses of hippocampal data. (a) Theta phase distribution of population neuron spikes, computed with bandpass-filtered LFP (black), PLSO estimate of $\hat{\mathbf{z}}_2$ with credible intervals estimated from 200 posterior samples (red). Horizontal gray line indicates the uniform distribution. (b-c) Spectrogram (in dB) for 500 seconds (b) STFT (c) PLSO with λ_{CV}. Learned frequencies are $(\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3) = (2.99, 7.62, 15.92)$ Hz, with $\widehat{\omega}_4 \sim \widehat{\omega}_5 > 25$ Hz. (d-e) Time-domain results. (d) Reconstructed signal (e) phase for $\widehat{\mathbf{z}}_2$ and interval boundary (vertical gray), with bandpass-filtered data (dotted black), STFT-reg. (blue), and PLSO (red). Shaded area represents 95% credible interval from $S = 200$ sample trajectories.

vergence is given by $\lambda \to \infty$ indicates the inaccuracy of the stationarity assumption.

6.2 LFP DATA FROM THE RAT HIPPOCAMPUS

We use LFP data collected from the rat hippocampus during open field tasks [Mizuseki et al., 2009], with $T = 1,600$ seconds and $f_s = 1,250$ Hz[2]. The theta neural oscillation band ($5 \sim 10$ Hz) is believed to play a role in coordinating the firing of neurons in the entorhinal-hippocampal system and is important for understanding the local circuit computation.

We fit PLSO with $J = 5$, which minimizes AIC as shown in Table 2, with 2-second PS interval. The estimated $\widehat{\omega}_2$ is 7.62 Hz in the theta band. To obtain the phase for non-PLSO methods, we perform the Hilbert transform on the theta-band reconstructed signal. With no ground truth, we bandpass-filter (BPF) the data in the theta band for reference.

Table 2: AIC as a function of J for Hippocampus data

J	1	2	3	4	**5**	6
AIC	2882	2593	2566	2522	**2503**	2505

Spike-phase coupling Fig. 4(a) shows the theta phase distribution of population neuron spikes in the hippocampus. The PLSO-estimated distribution (red) confirms the original results analyzed with bandpass-filtered signal (black) [Mizuseki et al., 2009]–the hippocampal spikes show a strong preference for a specific phase, π for this dataset, of the theta band. Since PLSO provides posterior sample-trajectories for the entire time-series, we can compute as many realizations of the phase distribution as the number of MC samples. The resulting credible interval excludes the uniform distribution (horizontal gray), suggesting the statistical significance of strong phase preference.

Denoised spectrogram Fig. 4(b-c) shows the estimated spectrogram. We observe that PLSO denoises the spectrogram, while retaining sustained power at $\widehat{\omega}_2 = 7.62$ Hz and weaker bursts at $(\widehat{\omega}_1, \widehat{\omega}_3) = (2.99, 15.92)$ Hz.

Time domain discontinuity Fig. 4(d-e) show a segment of the estimated signal and phase near a boundary for $\widehat{\omega}_2$. While the estimates from STFT-reg. (blue) and PLSO (red) follow the BPF result closely, the STFT-reg. estimates exhibit discontiunity/distortion near the boundary. In Fig. 4(e), the phase jump at the boundary is 38.4 degrees. We also computed $\mathrm{jump}(\phi_2)$ in degrees/sample. Considering that the theta band roughly progresses 2.16 ($= 7.5(\mathrm{Hz}) \times 360/1250\,(\mathrm{Hz})$) degrees/sample, we observe that BPF (2.23), as expected, and PLSO (λ_{CV} : 2.40, $\lambda \to \infty$: 2.66) are not affected by the boundary effect. This is not the case for STFT-reg. (26.83) and GP-PS (25.91).

Comparison with TVAR Fig. 5(a-b) shows a segment of TVAR inference results[3]. Specifically, Fig. 5(a) and (b) shows a time-varying center frequency $\widehat{\omega}_1$ and the corresponding reconstruction, for the lowest frequency component. Note that the eigenvalues, which correspond to $\{\omega_j\}_j$, and the eigenvectors, which are used for oscillatory decomposition, are derived from the *estimated* TVAR transition matrix. Consequently, we cannot explicitly control $\{\omega_j\}_j$, as shown in Fig. 5(a), the bandwidth of each component, as

[2]We use channel 1 of mouse ec013.528 for the LFP. The population spikes were simultaneously recorded.

[3]The details for TVAR is in the **Appendix K**.

Preliminary ver.

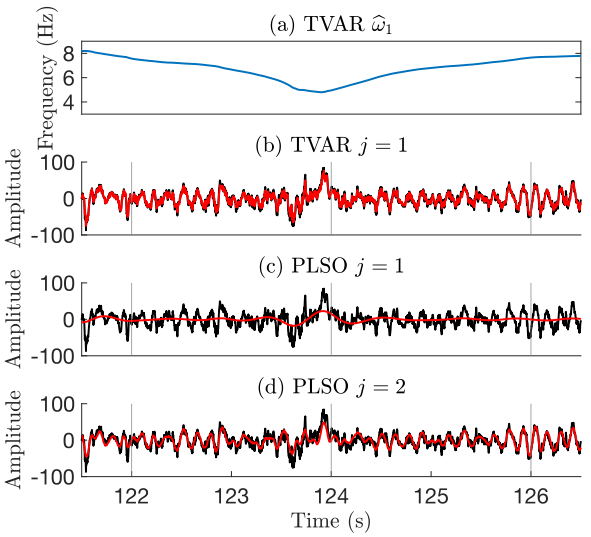

Figure 5: Hippocampus data. (a) time-varying $\widehat{\omega}_1$ for TVAR. (b-d) Inferred mean trajectory (red) for (b) TVAR $j = 1$, (c) PLSO $j = 1$, and (d) PLSO $j = 2$, with raw data (black).

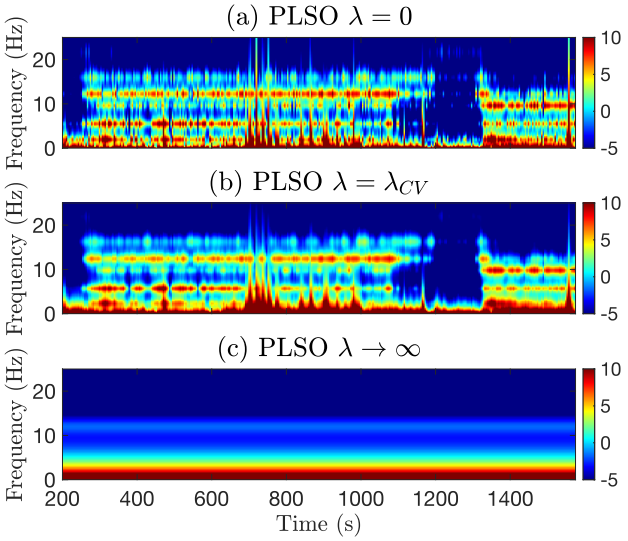

Figure 6: Spectrogram (in dB) under propofol anesthesia. PLSO with (a) $\lambda = 0$ (b) $\lambda = \lambda_{\mathrm{CV}}$ (c) $\lambda \to \infty$.

well as the number of components J. This is further complicated by the fact that the transition matrix changes every sample. Fig. 5(b) shows that this ambiguity results in the lowest-frequency component of TVAR explaining *both* the slow ($0.1 \sim 2$ Hz) and theta components. With PLSO, on the contrary, we can explicitly specify or learn the parameters. Fig. 5(c-d) demonstrates that PLSO is able to delineate the slow/theta components without any discontinuity.

6.3 EEG DATA FROM THE HUMAN BRAIN UNDER PROPOFOL ANESTHESIA

We apply PLSO to the EEG data from a subject anesthetized with propofol anesthetic drug, to assess whether PLSO can leverage regularization to recover smooth spectral dynamics, which is widely-observed during propofol-induced unconsciousness[4] [Purdon et al., 2013]. The data last $T = 2{,}300$ seconds, sampled at $f_s = 250$ Hz. The drug infusion starts at $t = 0$ and the subject loses consciousness around $t = 260$ seconds. We use $J = 6$ and assume a 4-second PS interval.

Smooth spectrogram Fig. 6(a-b) shows a segment of the PLSO-estimated spectrogram with $\lambda = 0$ and $\lambda = \lambda_{\mathrm{CV}}$. They identify strong slow ($0.1 \sim 2$ Hz) and alpha oscillations ($8 \sim 15$ Hz), both well-known signatures of propofol-induced unconsciousness. We also observe that the alpha band power diminishes between 1,200 and 1,350 seconds, suggesting that the subject regained consciousness before becoming unconscious again. PLSO with $\lambda = 0$ exhibits PSD fluctuation across windows, since $\{\sigma^2_{j,m}\}_{j,m}$ are estimated independently. The stationary PLSO ($\lambda \to \infty$) is restrictive and fails to capture spectral dynamics (Fig. 6(c)). In contrast, PLSO with λ_{CV} exhibits smooth dynamics by pooling together estimates from the neighboring windows. The regularization also helps remove movement-related artifacts, shown as vertical lines in Fig. 6(a), around $700 \sim 800$/1,200 seconds, and spurious power in $20 \sim 25$ Hz band. In summary, PLSO with regularization enables smooth spectral dynamics estimation and spurious noise removal.

[4]The EEG recording is part of de-identified data collected from patients at Massachusetts General Hospital (MGH) as a part of a MGH Human Research Committee-approved protocol.

7 CONCLUSION

We presented the Piecewise Locally Stationary Oscillatory (PLSO) framework to model nonstationary time-series data with slowly time-varying spectra, as the superposition of piecewise stationary (PS) oscillatory components. PLSO strikes a balance between stochastic continuity of the data across PS intervals and stationarity within each interval. For inference, we introduce an algorithm that combines Kalman theory and nonconvex optimization algorithms. Applications to simulated/real data show that PLSO preserves time-domain continuity and captures time-varying spectra. Future directions include 1) the automatic identification of PS intervals and 2) the expansion to higher-order autoregressive models and diverse priors on the parameters that enforce continuity across intervals.

Acknowledgements

We thank Abiy Tasissa for helpful discussion on the algorithm and the derivations. This research is supported by NIH grant P01 GM118629 and support from the JPB Foundation. AHS acknowledges Samsung scholarship.

Preliminary Version

References

S. Adak. Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, 93(444):1488–1501, 1998.

H. Akaike. Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1):3–14, 1981.

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.

A. M. Beck, E. P. Stephen, and P. L. Purdon. State space oscillator models for neural data analysis. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4740–4743, 2018.

E. N. Brown, V. Solo, Y. Choe, and Z. Zhang. Measuring period of human biological clock: infill asymptotic analysis of harmonic regression parameter estimates. *Methods in enzymology*, 383:382–405, 2004.

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

G. Casella. An introduction to empirical bayes data analysis. *The American Statistician*, 39(2):83–87, 1985.

A. T. Cemgil and S. J. Godsill. Probabilistic phase vocoder and its application to interpolation of missing values in audio signals. In *2005 13th European Signal Processing Conference*, pages 1–4, 2005.

J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio. A recurrent latent variable model for sequential data. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1):1–37, 02 1997.

P. Das and B. Babadi. Dynamic bayesian multitaper spectral analysis. *IEEE Transactions on Signal Processing*, 66(6):1394–1409, 2018.

I. Daubechies, J. Lu, and H. Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational Harmonic Analysis*, 30(2):243 – 261, 2011.

B. Gold, N. Morgan, and D. Ellis. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. Wiley-Interscience, USA, 2nd edition, 2011. ISBN 0470195363.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

R. B. Gramacy and H. K. H. Lee. Bayesian treed gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483):1119–1130, 2008.

D. Griffin and J. Lim. Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):236–243, 1984.

N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.

F. Itakura and S. Saito. A statistical method for estimation of speech spectral density and formant frequencies. 1970.

R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 1960.

S. Kim, M. K. Behr, D. Ba, and E. N. Brown. State-space multitaper time-frequency analysis. *Proceedings of the National Academy of Sciences*, 115(1):E5–E14, 2018.

G. Kitagawa and W. Gersch. A smoothness priors time-varying ar coefficient modeling of nonstationary covariance time series. *IEEE Transactions on Automatic Control*, 30(1):48–56, 1985.

S. J. Koopman and K. M. Lee. Seasonality with trend and cycle interactions in unobserved components models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(4):427–448, 2009.

R. G. Krishnan, U. Shalit, and D. Sontag. Structured inference networks for nonlinear state space models. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 2101–2109. AAAI Press, 2017.

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.

T. Matsuda and F. Komaki. Time series decomposition into oscillation components and phase estimation. *Neural Computation*, 29(2):332–367, 2017.

Preliminary version

K. Mizuseki, A. Sirota, E. Pastalkova, and G. Buzsaki. Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron*, 64:267–280, 2009.

A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-time Signal Processing (3rd Ed.)*. Prentice-Hall, Inc., 2009.

M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2):204–237, 1965.

P. L. Purdon, E. T. Pierce, E. A. Mukamel, M. J. Prerau, J. L. Walsh, K. F. K. Wong, A. F. Salazar-Gomez, P. G. Harrell, A. L. Sampson, A. Cimenser, S. Ching, N. J. Kopell, C. Tavares-Stoeckel, K. Habeeb, R. Merhar, and E. N. Brown. Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proceedings of the National Academy of Sciences*, 110(12):E1142–E1151, 2013.

Y. Qi, T. P. Minka, and R. W. Picard. Bayesian spectrum estimation of unevenly sampled nonstationary data. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II–1473–II–1476, 2002.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

O. Rosen, D. S. Stoffer, and S. Wood. Local spectral analysis via a bayesian mixture of smoothing splines. *Journal of the American Statistical Association*, 104(485):249–262, 2009.

A. Solin and S. Särkkä. Explicit Link Between Periodic Covariance Functions and State Space Models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 904–912. PMLR, 2014.

A. Solin, J. Hensman, and R. E Turner. Infinite-horizon gaussian processes. In *Advances in Neural Information Processing Systems 31*, pages 3486–3495. 2018.

A. H. Song, S. Chakravarty, and E. N. Brown. A smoother state space multitaper spectrogram. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 33–36, July 2018.

H. Soulat, E. P. Stephen, A. M. Beck, and P. L. Purdon. State space methods for phase amplitude coupling analysis. *bioRxiv*, 2019.

R. E. Turner and M. Sahani. Time-frequency analysis as probabilistic inference. *IEEE Transactions on Signal Processing*, 62(23):6171–6183, 2014.

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.

P. Whittle. Estimation and information in stationary time series. *Ark. Mat.*, 2(5):423–434, 08 1953.

W. J. Wilkinson, M. Riis Andersen, J. D. Reiss, D. Stowell, and A. Solin. Unifying probabilistic models for time-frequency analysis. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3352–3356, May 2019.

A. G. Wilson and R. P. Adams. Gaussian process kernels for pattern discovery and extrapolation. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, page III–1067–III–1075, 2013.

K. W. Wilson, B. Raj, P. Smaragdis, and A. Divakaran. Speech denoising using nonnegative matrix factorization with priors. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4029–4032, 2008.

S. J. Wright. Coordinate descent algorithms. *Mathematical Programming*, 151:3–34, 2015.

Preliminary version