SPARSE SPECTRAL ESTIMATION FROM POINT PROCESS OBSERVATIONS

Sina Miran[1] *Patrick L. Purdon*[2] *Emery N. Brown*[2,3,4] *Behtash Babadi*[1]

[1]Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, US
[2]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, US
[3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, US
[4]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, US
smiran@umd.edu, patrickp@nmr.mgh.harvard.edu, enb@neurostat.mit.edu, behtash@umd.edu

ABSTRACT

We consider the problem of estimating the power spectral density of the neural covariates underlying the spiking of a neuronal population. We assume the spiking of the neuronal ensemble to be described by Bernoulli statistics. Furthermore, we consider the conditional intensity function to be the logistic map of a second-order stationary process with sparse frequency content. Using the binary spiking data recorded from the population, we calculate the maximum *a posteriori* estimate of the power spectral density of the process while enforcing sparsity-promoting priors on the estimate. Using both simulated and clinically recorded data, we show that our method outperforms the existing methods for extracting a frequency domain representation from the spiking data of a neuronal population.

***Index Terms*—** point process models; power spectral density; spectral estimation; neural signal processing.

1. INTRODUCTION

Spectral analysis of electroencephalography (EEG) recordings has long been used for research and diagnosis purposes such as the identification of sleep disorders [1, 2] and epileptic seizures [3, 4]. With the development of invasive recording procedures, single- and multi- unit recordings have increasingly become popular in neural sciences [5, 6]. These data do not suffer from the low spatial resolution of non-invasive measurements such as EEG, as they capture the spiking activity of a localized population neurons; however, their binary nature introduces various signal processing challenges [7].

Recently, the theory of point processes has provided a mathematical framework to model and analyze neuronal data such as single and multi unit recordings in the time domain [8, 9, 10]. Following the common frequency-domain analysis of neural data such as EEG, existing methods for point process spectral estimation often compute a continuous estimate of the spiking rate and analyze the power spectral density (PSD) of this estimate. The spiking rate estimation is either done by simply smoothing the spiking histogram [11, 12, 13] or using generalized linear Gaussian state-space models to estimate the conditional intensity function (CIF) of the point process [14, 15]. However, these approaches suffer from the following shortcomings: Firstly, they are limited in terms of their spectral resolution as smoothing in the time domain for spiking rate estimation results in distortion in the frequency domain. Secondly, second-order statistics of the underlying time-series are needed for spectral estimation, and even if the spiking rate is estimated correctly, the second-order statistics need not be. Thirdly, current methods do not account for spectral sparsity which is often observed in the frequency domain analysis of biological data such as EEG.

In this work, we address the above shortcomings by casting the point process spectral estimation problem as a discrete-parameter harmonic spectral estimation in which the the second moments of the harmonic process driving the point process are estimated. To this end, we consider a conditional Bernoulli point process framework and model its CIF as the logistic map of a harmonic process. Then, we use an Expectation-Maximization (EM) algorithm to compute maximum *a posteriori* (MAP) estimate of the PSD of the harmonic process as the spectral representation of point process using the observed spiking data and sparsity-promoting priors. Simulation results and application to real multi-unit data recorded under general anesthesia verifies the superior performance of our method compared to existing techniques.

The rest of the paper is organized as follows: Section 2 introduces our model for the spiking activity of a population of neurons driven by a harmonic process. In Section 3, we derive the sparse MAP estimator of the PSD associated with the harmonic process. Section 4 provides results for simulated data and real multi-unit recordings under general anesthesia, comparing our sparse PSD estimates with existing methods. This is followed by our concluding remarks in Section 5.

2. PROBLEM FORMULATION

A point process can be fully characterized by its Conditional Intensity Function (CIF). For $t \in (0, T]$, the CIF of a point

This work was supported by the National Science Foundation under Grant No. 1552946.

process is defined as [16]

$$\lambda(t|H_t) := \lim_{\Delta\to 0} \frac{\mathrm{P}(N(t+\Delta)-N(t)=1|H_t)}{\Delta}, \quad (1)$$

where H_t and $N(t)$ denote respectively the spiking history in $(0,t)$ and the number of spikes in $(0,t]$. If the point process is discretized by small enough bins of size Δ, the data in each bin (n_k) can be approximated by a Bernoulli random variable with the success probability of $\lambda_k := \lambda\big(k\Delta|H_{k\Delta}\big)\Delta$, for $0 \le k \le K$, where $K = \lceil \frac{T}{\Delta} \rceil$. Due to the absolute refractory period of neurons, the sampling period of $\Delta \sim 1$ ms is typically sufficient to ensure that at most one spike occurs in any bin [8].

The oscillatory patterns in neuronal spiking can be directly attributed to its CIF. Thus, a spectral analysis of the CIF would provide an informative frequency domain representation of the point process. We consider an ensemble of L neurons, whose spiking data is denoted by $\mathcal{D} := \{n_k^{(\ell)}\}_{\ell=1,k=1}^{L,K}$, with $n_k^{(\ell)}$ denoting the data of neuron ℓ in bin k, and assume that they are driven by the same CIF. This assumption is plausible for multi-unit recordings as they are obtained from neighboring neurons [5]. To associate a power spectral density with the mutual CIF, we consider the CIF to be the logistic map of a second-order stationary signal with mean μ. Denoting the samples of the second-order stationary process by x_k for $0 \le k \le K$, the complete discretized model can be expressed as:

$$\begin{cases} \lambda_k = \dfrac{1}{1+e^{-x_k}}, \quad 1 \le k \le K \\ n_k^{(\ell)} \sim \text{Bernoulli}(\lambda_k), \quad 1 \le k \le K, 1 \le \ell \le L \end{cases}. \quad (2)$$

The objective is then to calculate a sparse estimate of the PSD of x_k as the spectral representation of the neural spiking data having observed $\mathcal{D}$.

3. BAYESIAN ESTIMATION OF THE PSD

According to the *Spectral Representation Theorem* [17], for the zero-mean and second-order stationary signal $x_k - \mu$ with PSD $S(\omega)$, there exists a continuous, orthogonal increment, and complex process $Z(\omega)$ such that

$$x_k - \mu = \int_0^{2\pi} e^{j\omega k} dZ(\omega) \quad (3)$$

where the integral is in the Riemann-Stieltjes sense and $E\{|dZ(\omega)|^2\} = S(\omega)d\omega$.

We consider a discrete approximation to the PSD $S(\omega)$ which results in $Z(\omega)$ being constant over intervals of length $\frac{\pi}{N}$ for large enough N [17]. Thus, $Z(\omega)$ can be replaced by a jump process in $[0,\pi)$ with jumps of $\frac{\pi}{N}(a_i + jb_i)$ at $\omega_i = \frac{i\pi}{N}$ for $i = 1,2,\ldots,N-1$, where a_i and b_i are random variables and N controls the degree of approximation. Following the *Spectral Representation Theorem* and the discrete PSD approximation, the PSD at ω_i would be $S(\omega_i) = \frac{\pi^2}{N^2}\mathbb{E}\{a_i^2 + b_i^2\}$. Furthermore, considering x_k to be *real*, Eq. (3) can be simplified as:

$$x_k - \mu = \frac{2\pi}{N} \sum_{i=1}^{N-1} \Big(a_i \cos(\omega_i k) - b_i \sin(\omega_i k) \Big). \quad (4)$$

Letting $\mathbf{x} = [x_1, x_2, \ldots, x_K]^T \in \mathbb{R}^K$, $\mathbf{v} = [\frac{N}{2\pi}\mu, a_1, b_1, \ldots, a_N, b_N]^T \in \mathbb{R}^{2N-1}$, and defining $\mathbf{A} \in \mathbb{R}^{K\times(2N-1)}$ as

$$\mathbf{A} := \frac{2\pi}{N} \begin{bmatrix} 1 & \cos(\frac{\pi}{N}) & -\sin(\frac{\pi}{N}) & \ldots & \cos\Big(\frac{(N-1)\pi}{N}\Big) & -\sin\Big(\frac{(N-1)\pi}{N}\Big) \\ 1 & \cos(\frac{2\pi}{N}) & -\sin(\frac{2\pi}{N}) & \ldots & \cos\Big(\frac{2(N-1)\pi}{N}\Big) & -\sin\Big(\frac{2(N-1)\pi}{N}\Big) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cos(\frac{K\pi}{N}) & -\sin(\frac{K\pi}{N}) & \ldots & \cos\Big(\frac{K(N-1)\pi}{N}\Big) & -\sin\Big(\frac{K(N-1)\pi}{N}\Big) \end{bmatrix}, \quad (5)$$

we can write Eq. (4) as $\mathbf{x} = \mathbf{A}\mathbf{v}$. We further assume that the process $Z(\omega)$ is Gaussian. Therefore, based on the definition of random vector $\mathbf{v}$, we have $v_i \sim \mathcal{N}(0, \sigma_i^2)$, for $i = 2,\ldots,2N-1$, and due to the orthogonality of the increments of process $Z(\omega)$, v_i's are independent. This results in $S(\omega_i) = \frac{\pi^2}{N^2}(\sigma_{2i}^2 + \sigma_{2i+1}^2)$, for $\omega_i = \frac{i\pi}{N}$ and $i = 1,2,\ldots,N-1$. The estimation of μ is not of particular importance as our goal is to infer the oscillatory patterns in the CIF. However, in order to have a consistent prior on all the elements of $\mathbf{v}$, we assume an independent Gaussian prior on μ such that $v_1 \sim \mathcal{N}(0,\sigma_1^2)$. Therefore, an estimate of the parameter vector $\boldsymbol{\theta} := [\sigma_1^2, \sigma_2^2, ..., \sigma_{2N-1}^2]^T$ results in an estimate of the approximated PSD of x_k.

To obtain an sparse estimate of $\boldsymbol{\theta}$ in the Bayesian estimation setting, we enforce independent exponential priors with parameter γ on the elements of $\boldsymbol{\theta}$, i.e. $\log f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = (2N-1)\log\gamma - \gamma\sum_{i=1}^{2N-1}\sigma_i^2$, where we estimate γ through two-fold cross validation [18]. This log-prior is equivalent to the ℓ_1-norm penalty of $\boldsymbol{\theta}$ which is known to promote sparsity. Thus, the maximum *a posterior* (MAP) estimate of $\boldsymbol{\theta}$ is defined as:

$$\widehat{\boldsymbol{\theta}}_{\mathsf{MAP}} = \arg\max_{\boldsymbol{\theta}} \Big(\log P(\mathcal{D}|\boldsymbol{\theta}) + \log f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \Big) \quad (6)$$

As $P(\mathcal{D}|\boldsymbol{\theta})$ is an intractable function of $\boldsymbol{\theta}$ and the vector $\mathbf{v}$ acts as a latent variable in the Bayesian estimation of $\boldsymbol{\theta}$, we employ the Expectation-Maximization (EM) algorithm to calculate the MAP estimate in Eq. (6) [19, 20].

The E Step: Denote the parameter vector estimate at the r^{th} iteration of the EM algorithm by $\widehat{\boldsymbol{\theta}}^{(r)} = \big[\sigma_1^{2^{(r)}}, \sigma_2^{2^{(r)}}, \cdots, \sigma_{2N-1}^{2\ (r)}\big]^T$. In the E step, we calculate the expectation of the complete data log-likelihood as:

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)}\right) &:= \mathbb{E}_{\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}}\big\{\log f(\mathcal{D},\mathbf{v}|\boldsymbol{\theta})\big\} + \log f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\ &= \mathbb{E}_{\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}}\big\{\log f_{\mathbf{v}|\boldsymbol{\theta}}(\mathbf{v}|\boldsymbol{\theta})\big\} + \log f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) + \mathsf{const.} \\ &= \sum_{i=2}^{2N-1} \Big(-\frac{1}{2}\log\sigma_i^2 - \frac{1}{2\sigma_i^2}\mathbb{E}_{\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}}\{v_i^2\} - \gamma\sigma_i^2 \Big) \\ &\quad + \mathsf{const.} \end{aligned} \quad (7)$$

where we have used the conditional independence relation $P(\mathcal{D}|\mathbf{v},\boldsymbol{\theta}) = P(\mathcal{D}|\mathbf{v})$. In (7), the term cnst. represents all terms which are not functions of $\boldsymbol{\theta}$. Calculating the distribution of $\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}$ involves intractable integrals and Monte Carlo methods are not efficient for calculating the term $\mathbb{E}_{\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}}\{v_i^2\}$ in (7) as they have to be performed in every EM iteration. However, the density of $\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}$ can be approximated by a multivariate Gaussian density $\mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{v}}^{(r)},\boldsymbol{\Sigma}_{\mathbf{v}}^{(r)}\right)$ [8, 10]. Similar to [8, 10], we calculate $\boldsymbol{\mu}_{\mathbf{v}}^{(r)}$ as the mode of the log-density $\log f_{\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}}(\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)})$ in (8) given by:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{v}}^{(r)} &= \arg\max_{\mathbf{v}}\left(\log P(\mathcal{D}|\mathbf{v}) + \log f_{\mathbf{v}|\widehat{\boldsymbol{\theta}}^{(r)}}(\mathbf{v}|\widehat{\boldsymbol{\theta}}^{(r)})\right)\\ &= \arg\max_{\mathbf{v}}\left(\sum_{k=1}^{K}\sum_{\ell=1}^{L} n_k^{(\ell)}(\mathbf{A}\mathbf{v})_k - \log\left(1+e^{(\mathbf{A}\mathbf{v})_k}\right)\right.\\ &\qquad \left. - \sum_{i=2}^{2N-1}\frac{v_i^2}{2\sigma_i^{2\,(r)}}\right),\end{aligned} \tag{8}$$

and evaluate the Hessian of this log-density at the calculated $\boldsymbol{\mu}_{\mathbf{v}}^{(r)}$ as $-\left(\boldsymbol{\Sigma}_{\mathbf{v}}^{(r)}\right)^{-1}$ [19].

The maximization problem in (8) is concave, and the Hessian is negative definite due to the diagonal terms of the form $\frac{v_i^2}{2\sigma_i^{2\,(r)}}$. Therefore, we use the standard Newton's method to efficiently compute $\boldsymbol{\mu}_{\mathbf{v}}^{(r)}$. It is worth noting that the ensemble average spiking signal $\bar{n}_k := \frac{1}{L}\sum_{\ell=1}^{L} n_k^{(\ell)}$ for $k = 1, 2, \cdots, K$, which is often referred to as the Peri-stimulus Time Histogram (PSTH), includes all the spiking information we need to calculate $\boldsymbol{\mu}_{\mathbf{v}}^{(r)}$ due to our mutual CIF assumption. The multivariate Gaussian estimate results in $E_{\mathbf{v}|\mathcal{D},\widehat{\boldsymbol{\theta}}^{(r)}}\{v_i^2\} = \left((\boldsymbol{\mu}_{\mathbf{v}}^{(r)})_i\right)^2 + (\boldsymbol{\Sigma}_{\mathbf{v}}^{(r)})_{i,i}$ for the E step, which we denote by $E_i^{(r)}$ for notational simplicity.

The M Step: Having calculated the positive $E_i^{(r)}$'s, we maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)})$ in (7) with respect to $\boldsymbol{\theta}$ in the M step. $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)})$ is a quasi-concave function over the positive orthant with a unique maximizer which is the non-negative solution of a quadratic equation. Hence, the update rule of the EM algorithm becomes:

$$(\widehat{\boldsymbol{\theta}}^{(r+1)})_i = \sigma_i^{2\,(r+1)} = \frac{-1+\sqrt{1+8\gamma E_i^{(r)}}}{4\gamma}, \quad 1 \le i \le 2N-1. \tag{9}$$

4. APPLICATION TO SIMULATED AND REAL DATA

As for our simulated data, we consider the dual-tone signal of the form

$$x(t) = 1.48\cos(2\pi f_0 t) + 0.685\cos(2\pi f_1 t) + 0.17n(t) - 5.7 \tag{10}$$

with $f_0 = 1$ Hz and $f_1 = 10$ Hz sampled at $f_s = 300$ Hz as the underlying harmonic process in our model, and $n(t)$ representing a zero-mean white Gaussian noise with unit variance. A total of $K = 1000$ samples are considered for $L = 10$ neurons. The bias term of -5.7 is chosen to make sure that the resulting spiking rate of 0.056 in PSTH is low enough and consistent with real-world neuronal spiking rates. Figure 1 shows the harmonic process as well as the resulting raster plot of the ensemble according to our model.

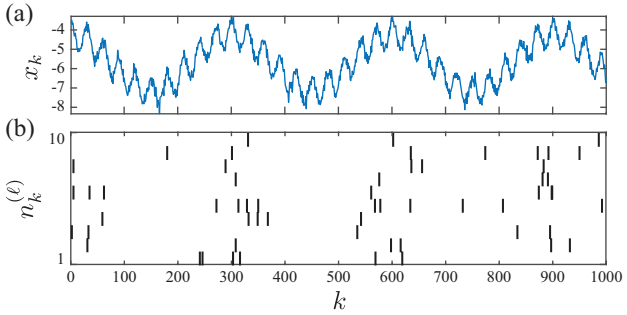

Fig. 1: (a) Dual-tone signal $x_{1:K}$ (b) Raster plot of the ensemble

Figures 2-(a) and 2-(b) respectively show the resulting PSTH and its smoothed versions using wide and narrow Gaussian kernels. Figure 2-(c) illustrates the normalized multitaper estimate [21] of the PSD corresponding to the smoothed PSTH for both the narrow and wide kernels. The multitaper estimate is among the most reliable nonparametric spectral estimation methods.

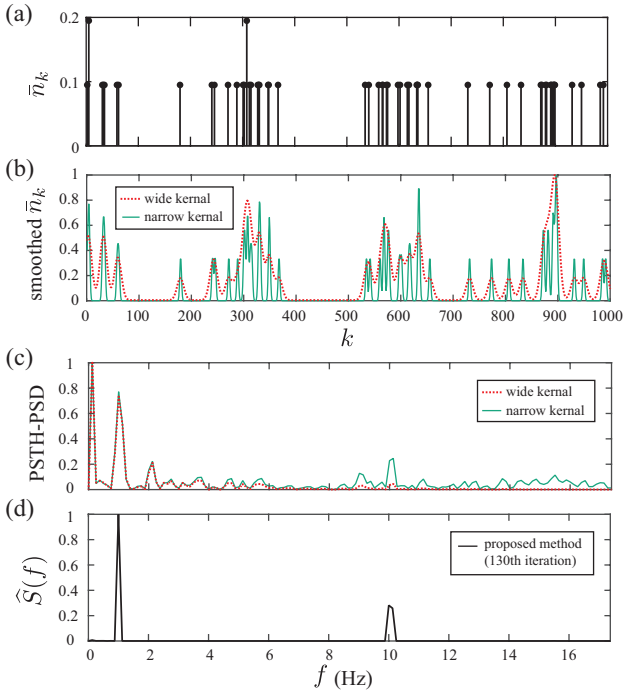

Fig. 2: Noisy dual-tone CIF model: (a) Raw PSTH of the data $\bar{n}_k$ with 0.056 spiking rate (b) Normalized smoothed PSTH using narrow and wide Gaussian kernels (c) Normalized multitaper PSD estimate of the smoothed PSTHs (d) Normalized PSD estimate of the proposed method after 130 EM iterations

The foregoing method of estimating the PSD of the smoothed PSTH, which we denote by PSTH-PSD, is one of the most commonly-used methods for extracting the spectral representation of the spiking data [11, 12, 13]. We have considered a frequency spacing of 0.125Hz for the PSD estimates here, which corresponds to $N = 1200$ in our proposed method, and DC components of the normalized PSD's are eliminated in the figures. The PSD estimate corresponding to the narrow smoothing kernel (Figure 2-(c), green trace) detects the two peaks at 1 Hz and 10 Hz, but has a high variability in higher frequencies. Also, two spurious peaks at 2 Hz and 9 Hz are detected in the PSD. On the other hand, the estimate corresponding to the wide smoothing kernel (Figure 2-(c), dotted red trace) has a smaller variability but misses the 10 Hz peak. These results show the high sensitivity of the PSTH-PSD approach to the choice of the smoothing kernel and the absence of a criterion for choosing one. The result of our proposed method in Figure 2-(d) with $\gamma_{\text{opt}} = 10^{-4}$ (obtained by two-fold cross-validation) fully recovers the two frequency components after 130 EM iterations.

Finally, we apply our algorithm to multi-unit recordings in [22] from a human subject under Propofol-induced general anesthesia. The original sampling rate of 1 kHz was reduced by the factor of 40 to reduce computational complexity, and a time window of 50s was considered ($K = 1250$). Out of the 41 neurons in the data set, we have considered the ones with at least two spikes in the 50s time frame ($L = 27$). Figure 3 shows the raster plot of the neuronal ensemble with the spiking rate of 0.1064 in the PSTH.

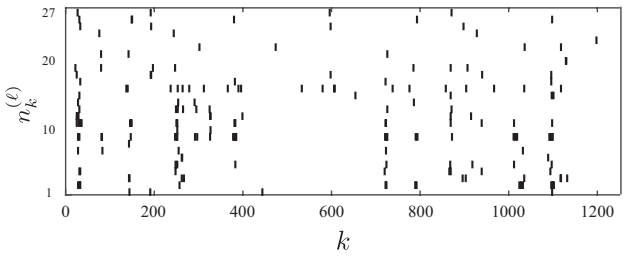

Fig. 3: Raster plot of the real multi-unit recordings in [22]

Figure 4 shows the PSD estimates using our proposed method as well as the PSTH-PSD estimates. All PSD's have the frequency spacing of 0.02Hz corresponding to $N = 625$ in our method. Similar to the foregoing simulation results, Figure 4-(c) shows the considerable variability of the narrow kernel PSTH-PSD in higher frequencies while the wide kernel PSTH-PSD suppresses the PSD components above 0.4Hz, implying the high sensitivity of the estimates to the choice of the kernel width. However, our method in Fig. 4-(d), after 100 EM iterations, calculates a sparse PSD with the regularization parameter obtained by two-fold cross-validation. It is worth noting that the results of the state-space model of [10, 14, 15] were very similar to the wide Gaussian kernel PSTH-PSD [19], and thus have not been shown in Figures 2 and 4 for brevity.

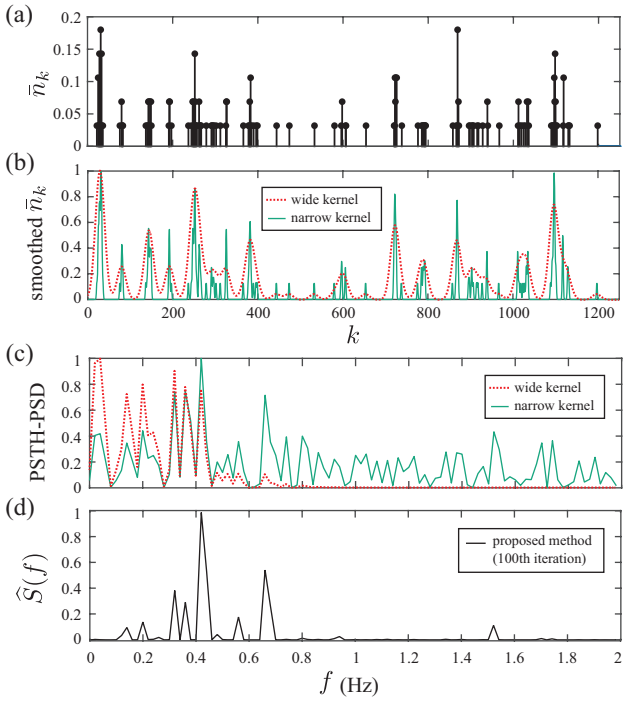

Fig. 4: Neuronal spiking data from anesthesia: (a) Raw PSTH of the data $\bar{n}_k$ with 0.1064 spiking rate (b) Normalized smoothed PSTH using narrow and wide Gaussian kernels (c) Normalized multitaper PSD estimate of the smoothed PSTHs (d) Normalized PSD estimate of the proposed method after 100 EM iterations

5. CONCLUSION

Although the theory of point process provides a concrete mathematical framework for analyzing binary data in the time domain, there has not been much work on extending this theory to incorporate a spectral representation for binary data. Existing approaches calculate an estimate of the spiking rate or the CIF driving the point process by smoothing the PSTH in the time domain or fitting a state-space model to the data. These methods have drawbacks such as limiting the resolution in frequency domain due to smoothing in the time domain and not taking into account the underlying sparse spectral structure often observed in biological signals.

In this paper, we have addressed these issues by introducing a model for spectral estimation from spiking data which directly estimates the PSD as the second-order moments of the underlying process from the observed binary data and employs sparsity-promoting priors for estimation. Application to simulated and real data reveals that our proposed method outperforms existing methods for calculating a point process spectral representation. Finally, although we have only focused on the application of our spectral estimation method to neuronal multi-unit recordings, our techniques can be similarly applied to any binary data exhibiting oscillatory behavior.

6. REFERENCES

[1] J. Fell, J. Röschke, K. Mann, and C. Schäffner, "Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures," *Electroencephalography and clinical Neurophysiology*, vol. 98, no. 5, pp. 401–410, 1996.

[2] A. D. Krystal, J. D. Edinger, W. K. Wohlgemuth, and G. R. Marsh, "Nrem sleep eeg frequency spectral correlates of sleep complaints in primary insomnia subtypes," *Sleep*, vol. 25, no. 6, pp. 630–640, 2002.

[3] S. Arroyo and S. Uematsu, "High-frequency eeg activity at the start of seizures.," *Journal of Clinical Neurophysiology*, vol. 9, no. 3, pp. 441–448, 1992.

[4] G. Alarcon, C. D. Binnie, R.D. C. Elwes, and C. E. Polkey, "Power spectrum and intracranial eeg patterns at seizure onset in partial epilepsy," *Electroencephalography and clinical neurophysiology*, vol. 94, no. 5, pp. 326–337, 1995.

[5] R. P. Vertes and R. W. Stackman, *Electrophysiological recording techniques*, vol. 54, Humana Press, 2011.

[6] W. Truccolo, J. A. Donoghue, L. R. Hochberg, E. N. Eskandar, J. R. Madsen, W. S. Anderson, E. N. Brown, E. Halgren, and S. S. Cash, "Single-neuron dynamics in human focal epilepsy," *Nature neuroscience*, vol. 14, no. 5, pp. 635–641, 2011.

[7] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature neuroscience*, vol. 7, no. 5, pp. 456–461, 2004.

[8] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects," *Journal of neurophysiology*, vol. 93, no. 2, pp. 1074–1089, 2005.

[9] L. Paninski, "Maximum likelihood estimation of cascade point-process neural encoding models," *Network: Computation in Neural Systems*, vol. 15, no. 4, pp. 243–262, 2004.

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

[11] M. Chalk, J. L. Herrero, M. A. Gieselmann, L. S. Delicato, S. Gotthardt, and A. Thiele, "Attention reduces stimulus-driven gamma frequency oscillations and spike field coherence in v1," *Neuron*, vol. 66, no. 1, pp. 114–125, 2010.

[12] B. C. Lewandowski and M. Schmidt, "Short bouts of vocalization induce long-lasting fast gamma oscillations in a sensorimotor nucleus," *The Journal of Neuroscience*, vol. 31, no. 39, pp. 13936–13948, 2011.

[13] A. Nini, A. Feingold, H. Slovin, and H. Bergman, "Neurons in the globus pallidus do not show correlated activity in the normal monkey, but phase-locked oscillations appear in the mptp model of parkinsonism," *Journal of neurophysiology*, vol. 74, no. 4, pp. 1800–1805, 1995.

[14] L. D. Lewis, V. S. Weiner, E. A. Mukamel, J. A. Donoghue, E. N. Eskandar, J. R. Madsen, W. S. Anderson, L. R. Hochberg, S. S. Cash, E. N. Brown, and P. L. Purdon, "Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness," *Proceedings of the National Academy of Sciences*, vol. 109, no. 49, pp. E3377–E3386, 2012.

[15] D. Ba, B. , P. L. Purdon, and E. N. Brown, "Robust spectrotemporal decomposition by iteratively reweighted least squares," *Proceedings of the National Academy of Sciences*, vol. 111, no. 50, pp. E5336–E5345, 2014.

[16] D. Vere-Jones, "An introduction to the theory of point processes," *Springer Ser. Statist., Springer, New York*, 1988.

[17] D. B. Percival, *Spectral analysis for physical applications*, Cambridge University Press, 1993.

[18] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning*, Springer, 2009.

[19] S. Miran, P. L. Purdon, E. N. Brown, and B. Babadi, "Robust spectral estimation from binary neuronal spiking data," *in press in IEEE Transactions on Biomedical Engineering*, 2017.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[21] B. Babadi and E. N. Brown, "A review of multitaper spectral analysis.," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1555–1564, 2014.

[22] L. D. Lewis, S. Ching, V. S. Weiner, R. A. Peterfreund, E. N. Eskandar, S. S. Cash, E. N. Brown, and P. L. Purdon, "Local cortical dynamics of burst suppression in the anaesthetized brain," *Brain*, vol. 136, no. 9, pp. 2727–2737, 2013.