

STATE-SPACE ANALYSIS ON TIME-VARYING CORRELATIONS IN PARALLEL SPIKE SEQUENCES

Hideaki Shimazaki¹, Shun-ichi Amari¹, Emery N. Brown^{2,3}, and Sonja Grün¹

¹ Theoretical Neuroscience Group, RIKEN Brain Science Institute, Wako-shi, Saitama, Japan

² Anesthesia and Critical Care, Massachusetts General Hospital, Boston, MA, USA

³ Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA

ABSTRACT

A state-space method for simultaneously estimating time-dependent rate and higher-order correlation underlying parallel spike sequences is proposed. Discretized parallel spike sequences are modeled by a conditionally independent multivariate Bernoulli process using a log-linear link function, which contains a state of higher-order interaction factors. A nonlinear recursive filtering formula is derived from a log-quadratic approximation to the posterior distribution of the state. Together with a fixed-interval smoothing algorithm, time-dependent log-linear parameters are estimated. The smoothed estimates are optimized via EM-algorithm such that their prior covariance matrix maximizes the expected complete data log-likelihood. In addition, we perform model selection on the hierarchical log-linear state-space models to avoid over-fitting. Application of the method to simultaneously recorded neuronal spike sequences is expected to contribute to uncover dynamic cooperative activities of neurons in relation to behavior.

Index Terms— State space methods, Point processes, Generalized linear model, Correlation, Information geometry

1. INTRODUCTION

Classical studies in neurophysiology are based on the idea that stimulus information is encoded in the firing rates of single neurons. Alternatively, precise spike coordination is discussed as an indication of coordinated network activity in form of a cell assembly [1] relevant for information processing. Activation of a cell assembly predicts higher-order correlation (HOC) between the spiking activities of its member neurons [2]. Supportive evidence for this concept was provided by existence of excess synchrony among neuronal spiking activities occurring dynamically in relation to behavioral context [3, 4, 5]. However, available approaches for correlation analysis do not allow for identifying time-dependent HOCs to trace active assemblies.

To characterize higher-order interaction, the log-linear model is a useful tool because it provides a well-defined measure of correlation based on information geometry [6].

Although its natural (canonical) parameters are not orthogonal, HOCs can be extracted in a quasi-orthogonal manner from a mixture of dually affine Riemannian coordinates [7, 8]. Former studies performed a regression analysis on parallel spike trains using either a log-linear model [7, 8, 9], or the log-linear model considering up to pairwise interaction only (maximum entropy model) [10] to characterize neuronal data. The existing approaches, however, assume stationarity, a condition that is typically not fulfilled in neuronal spike data from awake behaving animals.

The state-space method [11] was suggested as a framework to model a time-dependent system by representing its parameters (states) as a Markov process. It allows to estimate a filtered/smoothed posterior distribution of the time-dependent state conditional on observed data. The approximation method for non-Gaussian point process observation in a recursive filtering algorithm was successfully applied to neuronal spike data [12, 13, 14]. Existing state-space models incorporated ensembles of spike histories into the univariate spike response model [14, 15], however, without considering correlations between the spike sequences.

In this contribution, we provide a method for estimating the dynamics of HOCs by combining the log-linear model with a state-space analysis. For that estimation, we combined a fixed-interval smoother [11] with a nonlinear recursive filtering algorithm which we derived from a log-quadratic approximation of the posterior distribution. To obtain the most predictive model, we compared goodness-of-fit of hierarchical log-linear state-space models with different order of interactions by using the Akaike's Bayesian information criterion (ABIC) [16]. While inclusion of increasingly higher-order interaction terms improves the model accuracy, the estimation of higher-order parameters may suffer large variance due to the paucity of synchronous spikes in the data. This trade-off is optimally resolved with the model that minimizes the ABIC. The model complexity is thus selected based on the sample size as well as the prominence of higher-order structure.

An earlier version of this paper was presented in abstract form [Shimazaki, Brown, Grün, Statistical Analysis of Neuronal Data 4, Pittsburgh, May 2008].

2. METHODS

2.1. Log-linear model of higher-order interactions

We model N simultaneous spike sequences as a multivariate Bernoulli process. The processes, repeated over n trials, are discretized into T bins of bin size Δ . Let $\mathbf{X}^{t,l} = (X_1^{t,l}, X_2^{t,l}, \dots, X_N^{t,l})$ be N -tuple binary variables at t -th bin of l -th trial, where 1 denotes a bin filled with one or more spikes, and 0 denotes no spike. Let $\mathbf{x} = (x_1, x_2, \dots, x_N)$ represent 2^N spike patterns, and Ω_k be a k -subset of N elements: $\Omega_1 = \{1, 2, \dots, N\}$, $\Omega_2 = \{12, 13, \dots\}$, $\Omega_3 = \{123, 124, \dots\}$, etc. We define the interaction terms across the variables as $F_i(\mathbf{x}) = x_i$, $F_{ij}(\mathbf{x}) = x_i x_j$, $F_{ijk}(\mathbf{x}) = x_i x_j x_k$ ($1 \leq i < j < k \leq N$), etc. The log-linear model up to the r -th order interaction ($r \leq N$) is defined as [6]

$$\log p_t^{(r)}(\mathbf{x}) = -\psi_t + \sum_{i \in \{\Omega_1, \dots, \Omega_r\}} \theta_i^t F_i(\mathbf{x}) \quad (1)$$

with ψ_t being a normalization parameter. θ_i^t is a natural parameter referring to an interaction among variables indicated in the index i . Let $\boldsymbol{\theta}_t = [\theta_1^t, \dots, \theta_{12}^t, \dots, \theta_{1\dots r}^t, \dots]'$ be a $d = \sum_{k=1}^r \binom{N}{k}$ dimensional vector. Our goal is to obtain the estimate of $\boldsymbol{\theta}_t$ from the observed parallel spike sequences.

The log-linear model belongs to the exponential family which defines a dually flat space [6]. Its e-affine coordinate is given as the natural parameter $\boldsymbol{\theta}_t$, its m-affine coordinates as the expectation parameter $\boldsymbol{\eta}_t = [\eta_{11}^t, \dots, \eta_{12}^t, \dots, \eta_{1\dots r}^t, \dots]'$, where

$$\eta_i^t = E[F_i(\mathbf{x}) | \boldsymbol{\theta}_t] \quad (2)$$

for $i \in \{\Omega_1, \dots, \Omega_r\}$. An efficient estimator of η_i^t is the synchrony rate defined for the t -th bin as

$$y_i^t = \frac{1}{n} \sum_{l=1}^n F_i(\mathbf{X}^{t,l}) \quad (3)$$

for $i \in \{\Omega_1, \dots, \Omega_r\}$. Synchrony rates up to the r -th order, $\mathbf{y}_t = [y_{11}^t, \dots, y_{12}^t, \dots, y_{1\dots r}^t, \dots]'$, constitute sufficient statistic of the log-linear model up to the r -th order interaction.

2.2. State-space analysis on the log-linear model

Assuming conditionally independent observations, the likelihood of observed spike sequences is given as

$$p(\mathbf{y}_{1:T} | \boldsymbol{\theta}_{1:T}) = \prod_{t=1}^T \exp[n(\mathbf{y}_t' \boldsymbol{\theta}_t - \psi_t)], \quad (4)$$

with $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T\}$ and $\boldsymbol{\theta}_{1:T} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_T\}$.

Our prior assumption is the following state equation

$$\boldsymbol{\theta}_t = \mathbf{F} \boldsymbol{\theta}_{t-1} + \boldsymbol{\xi}_t, \quad (5)$$

for $t = 2, \dots, T$. A random vector $\boldsymbol{\xi}_t$ ($d \times 1$ matrix) is drawn from a zero-mean multivariate normal distribution with a covariance matrix \mathbf{Q} ($d \times d$ matrix). A matrix \mathbf{F} ($d \times d$ matrix)

is the first order autoregressive parameters. The initial value obeys $\boldsymbol{\theta}_1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $\boldsymbol{\mu}$ ($d \times 1$ matrix) and $\boldsymbol{\Sigma}$ ($d \times d$ matrix). We denote the set of hyper-parameters $\mathbf{w} = [\mathbf{F}, \mathbf{Q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}]$.

The joint probability distribution is obtained by combining the observation equation (Eq. 4) and the state equation (Eq. 5). The optimization principle for \mathbf{w} is chosen as the maximization of the marginal log-likelihood,

$$l(\mathbf{w}) = \log \int p(\mathbf{y}_{1:T}, \boldsymbol{\theta}_{1:T} | \mathbf{w}) d\boldsymbol{\theta}_{1:T}. \quad (6)$$

We make use of the expectation-maximization (EM) algorithm known to efficiently combining the estimation of the posterior density and the optimization of the hyper-parameter. Instead of Eq. 6, the method maximizes its lower bound,

$$\begin{aligned} Q(\mathbf{w} | \mathbf{w}^*) &= E[\log p(\mathbf{y}_{1:T}, \boldsymbol{\theta}_{1:T} | \mathbf{w}) | \mathbf{y}_{1:T}, \mathbf{w}^*] \\ &= n \sum_{t=1}^T (\mathbf{y}_t' E \boldsymbol{\theta}_t - E \psi_t) \\ &\quad - d \log \sqrt{2\pi} - \log \sqrt{\det \boldsymbol{\Sigma}} \\ &\quad - \frac{1}{2} E(\boldsymbol{\theta}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_1 - \boldsymbol{\mu}) \\ &\quad - (T-1) d \log \sqrt{2\pi} - (T-1) \log \sqrt{\det \mathbf{Q}} \\ &\quad - \frac{1}{2} \sum_{t=2}^T E(\boldsymbol{\theta}_t - \mathbf{F} \boldsymbol{\theta}_{t-1})' \mathbf{Q}^{-1} (\boldsymbol{\theta}_t - \mathbf{F} \boldsymbol{\theta}_{t-1}), \quad (7) \end{aligned}$$

by alternating expectation (E) and maximization (M) steps. In the E-step, we obtain the expected values with respect to $\boldsymbol{\theta}_t$ in Eq. 7 using \mathbf{w}^* . In the M-step, the hyper-parameter \mathbf{w} that maximizes Eq. 7 is obtained. The obtained \mathbf{w} is then used for \mathbf{w}^* in the next E-step. The details of each step are given as follows.

E-step Mean and covariance of the one-step prediction density $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})$ for the Gaussian state transition is given as [11, 12, 13]

$$\boldsymbol{\theta}_{t|t-1} = \mathbf{F} \boldsymbol{\theta}_{t-1|t-1}, \quad (8)$$

$$\mathbf{W}_{t|t-1} = \mathbf{F} \mathbf{W}'_{t-1|t-1} \mathbf{F}' + \mathbf{Q}. \quad (9)$$

Here $\boldsymbol{\theta}_{t|s}$ and $\mathbf{W}_{t|s}$ are the conditional mean $E[\boldsymbol{\theta}_t | \mathbf{y}_{1:s}]$ and covariance $E[(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|s})(\boldsymbol{\theta}_t - \boldsymbol{\theta}'_{t|s}) | \mathbf{y}_{1:s}]$. The filter distribution is given by the Bayes' theorem as

$$\begin{aligned} p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t, \mathbf{y}_{1:t-1}) p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})} \\ &\propto \exp[n(\mathbf{y}_t' \boldsymbol{\theta}_t - \psi_t)] \\ &\quad - \frac{1}{2} (\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1})' \mathbf{W}_{t|t-1}^{-1} (\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1}). \quad (10) \end{aligned}$$

Here we make a Gaussian approximation to the posterior by Laplace's method: the posterior mean, $\boldsymbol{\theta}_{t|t}$, is obtained from $\frac{\partial}{\partial \boldsymbol{\theta}_t} \log p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t}) = \mathbf{0}$; the covariance is given as $-\left[\frac{\partial^2}{\partial \boldsymbol{\theta}_t \partial \boldsymbol{\theta}_t'} \log p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t}) |_{\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t|t}}\right]^{-1}$. Note that the first

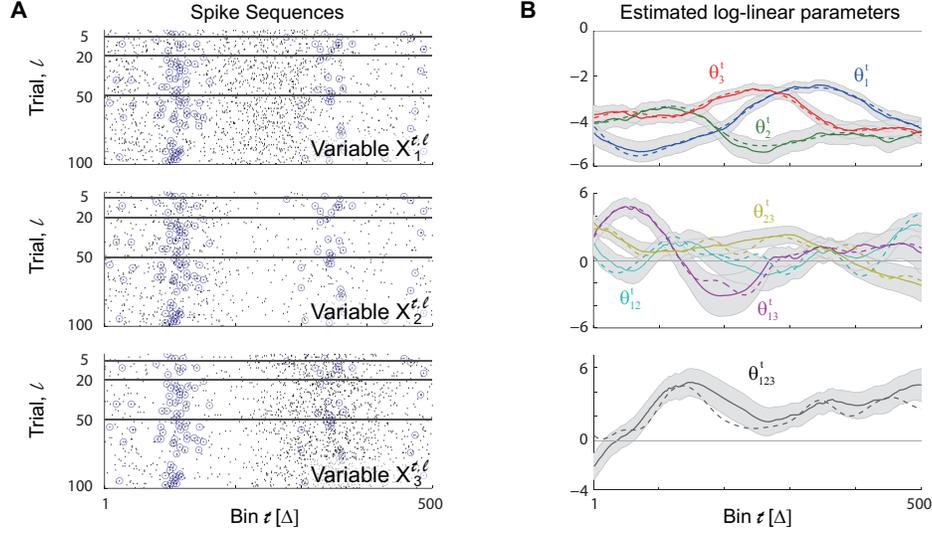


Fig. 1. (A) Simulated parallel spike sequences, $\mathbf{X}^{t,l}$. According to a time-dependent formulation of the log-linear model (see dashed lines in B for the model parameters), $N = 3$ parallel spike sequences were simulated repeatedly for $n = 100$ trials (duration: $T = 500$ bins of width $\Delta = 1$ [ms]). Each of the 3 panels show the spike events for the 3 variables $X_1^{t,l}, X_2^{t,l}, X_3^{t,l}$ ($t = 1, \dots, T$ and $l = 1, \dots, n$). Synchronous spikes across all 3 simultaneous sequences (detection in bins of 1ms width) are marked with circles. Horizontal lines indicate trial numbers $l = 5, 20, 50, 100$. (B) Smoothed estimates of the log-linear parameters, $\theta_{t|T}$. The different panels show the smoothed estimates (solid lines) of the log-linear parameters θ_i^t of different orders ($i \in \{\Omega_1, \Omega_2, \Omega_3\}$) derived from the data shown in A. The gray band indicates the 99% credible interval of the marginalized posterior density of the state (see text). The dashed lines show the true time-dependent parameters used for the generation of the spike sequences.

derivative of the moment generating function provides the expectation parameter $\frac{\partial \psi_t}{\partial \theta_t} = \boldsymbol{\eta}_t$. The second derivative is the Fisher metric $\frac{\partial^2 \psi_t}{\partial \theta_t \partial \theta_t'} = \mathbf{G}_t$. Here $(\mathbf{G}_t)_{ij} = \eta_{\omega_i \cup \omega_j}^t - \eta_{\omega_i}^t \eta_{\omega_j}^t$, where ω_i is an index of the i th element of $\boldsymbol{\eta}_t$. Hence we obtain the equations for the filtered mean and covariance as

$$\boldsymbol{\theta}_{t|t} = \boldsymbol{\theta}_{t|t-1} + n\mathbf{W}_{t|t-1}(\mathbf{y}_t - \boldsymbol{\eta}_{t|t}), \quad (11)$$

$$\mathbf{W}_{t|t}^{-1} = \mathbf{W}_{t|t-1}^{-1} + n\mathbf{G}_t. \quad (12)$$

Eq. 11 is solved by the Newton-Raphson method. Together with the one-step prediction formulae Eqs. 8 and 9, we evaluate the nonlinear recursion formulae Eqs. 11 and 12 for $t = 2, \dots, T$, by using the initial values $\boldsymbol{\theta}_{1|0} = \boldsymbol{\mu}$ and $\mathbf{W}_{1|0} = \boldsymbol{\Sigma}$.

Starting from $\boldsymbol{\theta}_{T|T}$ and $\mathbf{W}_{T|T}$ obtained by the filtering algorithm, the following fixed-interval smoothing algorithm [11, 12, 13] provides the smoothed mean and covariance. For $t = T-1, T-2, \dots, 2$, we compute

$$\boldsymbol{\theta}_{t|T} = \boldsymbol{\theta}_{t|t} + \mathbf{A}_t (\boldsymbol{\theta}_{t+1|T} - \boldsymbol{\theta}_{t+1|t}), \quad (13)$$

$$\mathbf{W}_{t|T} = \mathbf{W}_{t|t} + \mathbf{A}_t (\mathbf{W}_{t+1|T} - \mathbf{W}_{t+1|t}) \mathbf{A}_t'. \quad (14)$$

with

$$\mathbf{A}_t = \mathbf{W}_{t|t} \mathbf{F}' \mathbf{W}_{t+1|t}^{-1}. \quad (15)$$

The lag-one covariance smoother $\mathbf{W}_{t-1,t|T}$ is obtained by the

method of De Jong and Mackinnon [17]:

$$\begin{aligned} \mathbf{W}_{t-1,t|T} &= E[(\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_{t-1|T})(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|T})' | \mathbf{y}_{1:T}] \\ &= \mathbf{A}_{t-1} \mathbf{W}_{t|T}. \end{aligned} \quad (16)$$

M-step Given $\boldsymbol{\theta}_{t|T}$, $\mathbf{W}_{t|T}$, $\mathbf{W}_{t-1,t|T}$, the update of \mathbf{w} is given by the following equations. The covariance matrix \mathbf{Q} and the auto-regressive parameter \mathbf{F} are updated according to

$$\begin{aligned} \mathbf{Q} &= \frac{1}{T-1} \sum_{t=2}^T [\mathbf{W}_{t|T} - \mathbf{W}_{t-1,t|T} \mathbf{F}' \\ &\quad - \mathbf{F} \mathbf{W}_{t-1,t|T}' + \mathbf{F} \mathbf{W}_{t-1|T} \mathbf{F}'] \\ &\quad + \frac{1}{T-1} \sum_{t=2}^T (\boldsymbol{\theta}_{t|T} - \mathbf{F} \boldsymbol{\theta}_{t-1|T}) (\boldsymbol{\theta}_{t|T} - \mathbf{F} \boldsymbol{\theta}_{t-1|T})' \end{aligned} \quad (17)$$

and

$$\begin{aligned} \mathbf{F} &= \left[\sum_{t=2}^T (\mathbf{W}_{t-1,t|T} + \boldsymbol{\theta}_{t|T} \boldsymbol{\theta}_{t-1|T}') \right] \\ &\quad \cdot \left[\sum_{t=2}^T (\mathbf{W}_{t-1|T} + \boldsymbol{\theta}_{t-1|T} \boldsymbol{\theta}_{t-1|T}') \right]^{-1}. \end{aligned} \quad (18)$$

The mean of the initial distribution is updated with $\boldsymbol{\mu} = \boldsymbol{\theta}_{1|T}$. The covariance matrix $\boldsymbol{\Sigma}$ is not updated. Instead, we use a nominal diagonal matrix as $\boldsymbol{\Sigma}$.

Table 1. The ABIC of the r -th order log-linear model applied to data of different number of trials n (data shown in Fig.1-A). The asterisk indicates the model that minimizes the ABIC.

| | $r = 1$ | $r = 2$ | $r = 3$ |
|-----------|---------|---------|---------|
| $n = 5$ | 2085* | 2097 | 2144 |
| $n = 20$ | 7913 | 7698* | 7728 |
| $n = 50$ | 19263 | 18542 | 18540* |
| $n = 100$ | 38478 | 36811 | 36781* |
| $n = 200$ | 76683 | 73330 | 73238* |

2.3. Model Selection

Let us define a log-linear state-space model with up to the r -th order interaction by E_r . Since E_r is a submanifold of E_{r+1} , the log-linear models naturally form hierarchical structure, $E_1 \subset E_2 \subset \dots \subset E_N$. Comparison of the hierarchical models is performed by computing the ABIC [16]:

$$\text{ABIC} = -2l(\mathbf{w}) + 2 \dim \mathbf{w}. \quad (19)$$

The marginal log-likelihood is computed by using the one-step prediction formula for the mean Eq. 8 as $l(\mathbf{w}) = \sum_{t=1}^T \log p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{w}) \approx n \sum_{t=1}^T (\mathbf{y}_t' \boldsymbol{\theta}_{t|t-1} - \psi_{t|t-1})$.

3. RESULTS

We applied the method to $N = 3$ parallel spike sequences with $n = 100$ repetitions (Fig.1-A), generated by a time-dependent log-linear model (Fig.1-B, dashed lines). Figure 1-B (solid lines) shows the estimates of the log-linear parameters from the data in Figure 1-A. The gray band is the 99% credible interval of the marginalized posterior distribution $p(\theta_i^t | \mathbf{y}_{1:T})$ for $i \in \{\Omega_1, \Omega_2, \Omega_3\}$. Note that parameter θ_{123}^t indicates triplewise correlation, i.e. excess synchrony compared to expectation given pairwise correlations.

While the full model (E_3) is an unbiased model for the triple binary sequences, the fitted parameter for the triplewise correlation θ_{123}^t may suffer large variance (over-fitting) due to the paucity of triplet synchrony. Thus a submodel with $r < 3$ can be close to the generative model in terms of the Kullback-Leibler risk. To validate that the inclusion of the triplewise correlation improves the goodness-of-fit, we computed the ABICs for the hierarchical models ($r = 1, 2, 3$) as shown in Table 1. To test the influence of the sample size of the data upon the model selection, we varied the number of trials n used to fit the hierarchical log-linear models. For small number of trials ($n = 5$), the model without correlation structure was selected. With increasing trial numbers, models with increasing correlation orders were selected. For $n = 50, 100$, and 200, the full model predicts best.

4. CONCLUSION

We developed a method for identifying the time-varying higher-order correlation structure in parallel spike sequences. Its application to simultaneous recordings of neuronal activity is expected to provide us with new insights into the dynamics of assembly activities, their composition and behavioral relevance.

Acknowledgements This work was supported in part by JSPS Research Fellowships for Young Scientists (HS), R01 MH59733 / DP1 OD 003646 (EB), and RIKEN Strategic Programs for R&D (SG).

5. REFERENCES

- [1] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*, Wiley, New York, 1949.
- [2] S. Grün et al. "Impact of higher-order correlations on coincidence distributions of massively parallel data.," in *Lect Notes Comput Sc*, vol. 5286, pp. 96-114, Nov 2008
- [3] E. Vaadia et al. "Dynamics of neuronal interactions in monkey cortex in relation to behavioral events," *Nature*, vol. 373, no. 6514, pp. 515-518, Feb 1995.
- [4] A. Riehle et al. "Spike synchronization and rate modulation differentially involved in motor cortical function," *Science*, vol. 278, no. 5345, pp. 1950-1953, Dec 1997.
- [5] S. Grün et al. "Unitary events in multiple single-neuron spiking activity: II. nonstationary data.," *Neural Comput*, vol. 14, no. 1, pp. 81-119, Jan 2002.
- [6] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society, 2000.
- [7] S.-I. Amari, "Information geometry on hierarchy of probability distributions," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1701-1711, Jul 2001.
- [8] H. Nakahara and S.-I. Amari, "Information-geometric measure for neural spikes.," *Neural Comput*, vol. 14, no. 10, pp. 2269-2316, Oct 2002.
- [9] L. Martignon et al. "Detecting higher-order interactions among the spiking events in a group of neurons.," *Biol Cybern*, vol. 73, no. 1, pp. 69-81, Jun 1995.
- [10] E. Schneidman et al. "Weak pairwise correlations imply strongly correlated network states in a neural population.," *Nature*, vol. 440, no. 7087, pp. 1007-1012, Apr 2006.
- [11] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*, Springer-Verlag, New York, 1996.
- [12] E. N. Brown et al. "A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells.," *J Neurosci*, vol. 18, no. 18, pp. 7411-7425, Sep 1998.
- [13] A. C. Smith and E. N. Brown, "Estimating a state-space model from point process observations.," *Neural Comput*, vol. 15, no. 5, pp. 965-991, May 2003.
- [14] W. Truccolo et al. "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects.," *J Neurophysiol*, vol. 93, no. 2, pp. 1074-1089, Feb 2005.
- [15] M. Okatan et al. "Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity.," *Neural Comput*, vol. 17, no. 9, pp. 1927-1961, Sep 2005.
- [16] H. Akaike, "Likelihood and the Bayes procedure.," in *Bayesian Statistics*, University Press, pp. 143-166, 1980.
- [17] P. De Jong and M. J. Mackinnon, "Covariances for smoothed estimates in state space models.," *Biometrika*, vol. 75, pp. 601-602, Sep 1988.