

## Statistical Mechanics for Neural Spike Data Analysis Using Log-Linear Model

Hayaru SHOUNO,<sup>1</sup> Koji WADA<sup>2</sup> and Masato OKADA<sup>3,4,5</sup>

<sup>1</sup>*Faculty of Engineering, Yamaguchi University, Ube 755-8611, Japan*

<sup>2</sup>*Kochi National College of Technology, Nangoku 783-8508, Japan*

<sup>3</sup>*RIKEN BSI, Wako 351-0198, Japan*

<sup>4</sup>*Japan Scientific Technology Corp., Kashiwa 277-8561, Japan*

<sup>5</sup>*Graduate School of Frontier Science, The University of Tokyo,  
Kashiwa 277-8561, Japan*

Recently, we can simultaneously record spike data from many neurons in the field of electrophysiology, and thus it is required to develop mathematical framework for extracting higher-order correlation of neural firings. The joint probability of neural spike can be represented using the log-linear model. From statistical-mechanical point of view, the log-linear model can be regarded as a multi-body interacted Ising spin model or the Boltzman machine with higher-order interactions. The estimation of higher-order correlation of neural firing corresponds to that of higher-order interactions in this Ising spin system, and to the hyper-parameter estimation in the Bayesian inference. In this paper, we apply maximization of marginal likelihood (MML) method to this problem, and discuss the properties of MML analytically using statistical-mechanical method.

### §1. Introduction

We treat a spike signal analysis method, which is based on a log-linear model, using statistical mechanics. In these decades, in the field of electrophysiology, simultaneous multi neurons' activities are enabled to be recorded, and growing attention has been paid to the coincidence of firing neurons as a carrier of information. To analyze the correlated firing of neurons, Nakahara and Amari proposed using a log-linear model to describe the statistical structure of firing neurons.<sup>1)</sup> The log-linear model, which describes probability of state of firing neurons, includes pairwise, triple-wise, and higher-order interactions as hyper-parameters in its form. Thus, we should estimate those hyper-parameters from observation data. Wada and Okada formulated the problem as a maximization of likelihood, and analyzed it with treating the log-linear model as a multi-body interaction Ising model. In this work, we introduce a observation noise into Wada and Okada's model, and apply EM algorithm<sup>5)</sup> for the log likelihood maximization.<sup>3),4)</sup> Since the model can also be regarded as the multi-body interacted Ising model, we analyze the averaged property of hyper-parameter estimation problem with replica method.

### §2. Model

We assume  $N$  neurons are observed and each neuron can have two states, that is, one is firing state denoted as '+1' and the other is non-firing state denoted as '-1'. We describe the state of neurons as  $\xi$ , and each component,  $\xi_i \in \{-1, +1\}$

where  $1 \leq i \leq N$ , describes the single neuron state. Assuming the neurons states reach a stationary state, the log-linear model can be described as follows:

$$\mathcal{P}(\boldsymbol{\xi}) = \frac{1}{Z_{II}(\boldsymbol{\theta})} \exp(-H(\boldsymbol{\xi})), \quad (2.1)$$

$$H(\boldsymbol{\xi}) = - \sum_{r=1} \sum_{i_1 < \dots < i_r} \theta_{i_1 \dots i_r} \xi_{i_1} \dots \xi_{i_r}, \quad (2.2)$$

$$Z_{II}(\boldsymbol{\theta}) = \text{Tr}_{\boldsymbol{\xi}} \exp(-H(\boldsymbol{\xi})). \quad (2.3)$$

Equation (2.1) implies a probability distribution for  $\boldsymbol{\xi}$ , which is a kind of Gibbs distribution. Equation (2.2) represents the Hamiltonian, and the hyper-parameters are described as  $\boldsymbol{\theta} = \{\{\theta_i\}_{1 \leq i \leq N}, \{\theta_{i_1 i_2}\}_{1 \leq i_1 < i_2 \leq N}, \dots, \{\theta_{i_1 \dots i_r}\}_{1 \leq i_1 < \dots < i_r \leq N}, \dots, \theta_{1 \dots N}\}$ , which consists of  $2^N - 1$  parameters. However, carrying out such a lot of hyper-parameter estimation is not practical, Nakahara and Amari assumed the homogeneity of same-order interactions for reducing the number of hyper-parameters. Then  $r$ th-order hyper-parameters  $\{\theta_{i_1 \dots i_r}\}_{1 \leq i_1 < \dots < i_r \leq N}$  can be rewritten by a parameter  $J_r$ :  $\theta_{i_1, \dots, i_r} = J_r \frac{r!}{N^{r-1}}$ , where  $\frac{r!}{N^{r-1}}$  is a scaling factor for analysis, and the number of hyper-parameters reduces to  $N$ . Hence, the Hamiltonian equation (2.2) is rewritten as

$$H(\boldsymbol{\xi}) = - \sum_{r=1}^R \frac{J_r r!}{N^{r-1}} \sum_{i_1 < \dots < i_r} \xi_{i_1} \dots \xi_{i_r} = -N \sum_{r=1}^R J_r \left( \frac{1}{N} \sum_i \xi_i \right)^r. \quad (2.4)$$

We introduce observation noises like binary symmetric channel (BSC), that is,  $\mathcal{P}(\tau_i = \mp 1 | \xi_i = \pm 1) = p$ , where  $\tau_i$  represents the observation for the signal  $\xi_i$ , and  $p$  represents flip probability. Thus, the conditional probability  $\mathcal{P}(\boldsymbol{\tau} | \boldsymbol{\xi}; \beta_p)$  can be described as

$$\mathcal{P}(\boldsymbol{\tau} | \boldsymbol{\xi}; \beta_p) = \prod_{i=1}^N \mathcal{P}(\tau_i | \xi_i; \beta_p) = \frac{\exp(\beta_p \sum_i \tau_i \xi_i)}{(2 \cosh(\beta_p))^N}, \quad (2.5)$$

where  $\beta_p = \frac{1}{2} \ln \frac{1-p}{p}$ .

To estimate hyper-parameter  $\mathbf{J}$  and  $\beta_p$ , we applied maximization of log-likelihood  $\ln \mathcal{P}(\boldsymbol{\tau}; \beta_p, \mathbf{J})$ :

$$(\tilde{\mathbf{J}}, \tilde{\beta}_p) = \underset{\mathbf{J}, \beta_p}{\text{argmax}} \ln \mathcal{P}(\boldsymbol{\tau}; \beta_p, \mathbf{J}). \quad (2.6)$$

The log-likelihood can be denoted as following three functions:

$$\ln \mathcal{P}(\boldsymbol{\tau}; \beta_p, \mathbf{J}) = \ln Z(\beta_p, \mathbf{J}) - \ln Z_{II}(\mathbf{J}) - \ln Z_L(\beta_p), \quad (2.7)$$

where

$$Z(\beta_p, \mathbf{J}) = \text{Tr}_{\boldsymbol{\sigma}} \exp \left( N \sum_r J_r \left( \frac{1}{N} \sum_{i=1}^N \sigma_i \right)^r + \beta_p \sum_{i=1}^N \tau_i \sigma_i \right), \quad (2.8)$$

$$Z_L(\beta_p) = \text{Tr}_{\boldsymbol{\sigma}} \exp \left( \beta_p \sum_{i=1}^N \sigma_i \tau_i \right) = (2 \cosh \beta_p)^N. \quad (2.9)$$

A configuration  $\sigma$  denotes an estimate of a particular original signal  $\xi$ .

2.1. *Mean field theory*

To evaluate the log-likelihood, we introduce mean field approximation. For  $Z_{II}$ , introducing a parameter  $m_0 = \frac{1}{N} \sum_i \sigma_i$ , we can obtain saddle point equations:<sup>2)</sup>

$$Z_{II} = \exp(-N f_0(m_0)), \tag{2.10}$$

$$f_0(m_0) = \sum_r J_r(r-1)m_0^r - \ln 2 \cosh\left(\sum_r J_r r m_0^{r-1}\right), \tag{2.11}$$

$$m_0 = \tanh\left(\sum_r J_r r m_0^{r-1}\right). \tag{2.12}$$

On the other hand, the logarithm partition function  $\ln Z$  includes randomness of observation, so that we evaluate configuration average of it with replica method  $[\ln Z] = \lim_{n \rightarrow 0} \frac{[Z^n] - 1}{n}$ . The averaged replicated partition function  $[Z^n]$  can be described as

$$[Z^n] = \prod_{\alpha=1}^n \text{Tr}_{\{\sigma^\alpha\}} \text{Tr}_{\xi, \tau} \mathcal{P}(\tau | \xi; \beta_p^*) \mathcal{P}(\xi; \mathbf{J}^*) Z^n, \tag{2.13}$$

where  $\beta_p^*$  and  $\mathbf{J}^*$  denote true hyper parameters. The conditional probability  $\mathcal{P}(\tau | \xi; \beta_p^*)$  is denoted as Eq. (2.5) with true hyper-parameter  $\beta_p^*$ . The prior probability  $\mathcal{P}(\xi; \mathbf{J}^*)$  is described as  $\mathcal{P}(\xi; \mathbf{J}^*) = \prod_{i=1}^N \frac{\exp(\xi_i \sum_r J_r^* r (m_0^*)^{r-1})}{2 \cosh(\sum_r J_r^* r (m_0^*)^{r-1})}$ , where  $m_0^*$  is the solution of Eq. (2.12) using true hyper-parameter  $\mathbf{J}^*$ . Using the replica symmetric solution  $m = m_\alpha = \sum_{i=1} \sigma_\alpha$ ,  $[\ln Z]$  can be denoted as

$$-\frac{1}{N} [\ln Z] = \sum_r J_r(r-1)m^r - \text{Tr}_\tau w^*(\tau) \ln 2 \cosh\left(\sum_r J_r r m^{r-1} + \beta_p \tau\right), \tag{2.14}$$

$$m = \text{Tr}_\tau w^*(\tau) \tanh\left(\beta_p \tau + \sum_r J_r r m^{r-1}\right), \tag{2.15}$$

where  $w^*(\tau) = \frac{\cosh(\beta_p^* \tau + \sum_r J_r^* r m_0^{*r-1})}{2 \cosh(\beta_p^*) \cosh(\sum_r J_r^* r m_0^{*r-1})}$ .

**§3. Hyper-parameter estimation**

To estimate the hyper-parameters  $\mathbf{J}$  and  $\beta_p$ , we apply EM algorithm which reduces a value of pseudo loss function  $\mathcal{Q}(\beta_p, \mathbf{J} | \beta_p^{(t)}, \mathbf{J}^{(t)})$ :

$$\mathcal{Q}(\beta_p, \mathbf{J} | \beta_p^{(t)}, \mathbf{J}^{(t)}) = \text{Tr}_\sigma \mathcal{P}(\sigma | \tau; \beta_p^{(t)}, \mathbf{J}^{(t)}) \times \ln \mathcal{P}(\tau | \sigma; \beta_p) \mathcal{P}(\sigma; \mathbf{J}) \tag{3.1}$$

$$= \beta_p \frac{\partial \ln Z^{(t)}}{\partial \beta_p^{(t)}} + \sum_r J_r \frac{\partial \ln Z^{(t)}}{\partial J_r^{(t)}} - \ln Z_L(\beta_p) - \ln Z_{II}(\mathbf{J}), \tag{3.2}$$

where  $Z^{(t)} = Z(\beta_p^{(t)}, \mathbf{J}^{(t)})$  and  $t$  represents update index. In the EM algorithm, the update rules of hyper-parameters can be denoted as:

$$\beta_p^{(t+1)} = \underset{\beta_p}{\text{argmax}} Q(\beta_p, \mathbf{J} | \beta_p^{(t)}, \mathbf{J}^{(t)}), \quad J_r^{(t+1)} = \underset{J_r}{\text{argmax}} Q(\beta_p, \mathbf{J} | \beta_p^{(t)}, \mathbf{J}^{(t)}). \tag{3.3}$$

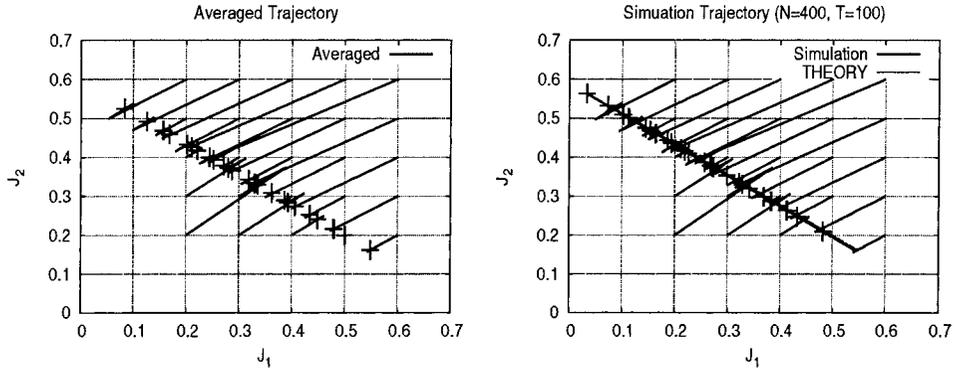


Fig. 1. Trajectories of hyper-parameter estimation for  $\mathbf{J}$ .  $\mathbf{J}^* = (0.5, 0.2)$ . The left figure shows theoretically averaged behavior of the EM algorithm. The right one shows simulation results, which is averaged  $T = 100$  trials, where  $N = 400$ .

One of our interests is in the conduct of the EM algorithm for the hyper-parameter estimation, which is denoted by the averaged update rules:

$$\beta_p^{(t+1)} = \operatorname{argmax}_{\beta_p} \beta_p \frac{\partial[\ln Z^{(t)}]}{\partial \beta_p^{(t)}} - \ln Z_L(\beta_p), \quad (3.4)$$

$$J_r^{(t+1)} = \operatorname{argmax}_{J_r} J_r \frac{\partial[\ln Z^{(t)}]}{\partial J_r^{(t)}} - \ln Z_{II}(\mathbf{J}). \quad (3.5)$$

#### §4. Conclusion

Figure 1 shows the trajectories of hyper-parameters estimation for  $\mathbf{J}^* = (0.5, 0.2)$ , where each trajectory starts with several initial states  $\mathbf{J}^0$ . The left figure shows the averaged trajectories of EM algorithm, and the right one shows the simulation results corresponding to the left. Each averaged trajectory is very similar to the corresponding simulation result, and both of them look as if converging to the line which include the true hyper-parameter point  $\mathbf{J}^*$  since the values of the log-likelihood are identical and maximized along the line. The averaged hyper-parameters estimation does not converge to the true hyper-parameters  $\mathbf{J}^*$ , so that, we conclude it is difficult to determine the true hyper parameters from the log-linear model using EM algorithm. This difficulty comes from the assumption for the hyper-parameters, that is, the hyper-parameters in each order are identical.<sup>1)</sup> Thus we should more carefully treat the assumption for the hyper-parameters.

#### References

- 1) H. Nakahara and S. Amari, *Neural Computation* **14** (2002), 2269.
- 2) K. Wada and M. Okada, *The IEICE Trans. Inf. Syst.* **85-D-II** (2002), 1582.
- 3) J. Inoue and K. Tanaka, *Phys. Rev. E* **65** (2002), 016125.
- 4) J. Inoue and K. Tanaka, *J. of Phys. A* **36** (2003), 10997.
- 5) A. P. Dempster, N. M. Larid and D. B. Rubin, *J. Roy. Statist. Soc. B* **39** (1977), 1.