Information-geometric decomposition in spike analysis

Hiroyuki Nakahara, Shun-ichi Amari
Lab. for Mathematical Neuroscience, RIKEN Brain Science Institute
2-1 Hirosawa, Wako, Saitama, 351-0198 Japan
{hiro, amari}@brain.riken.go.jp

Abstract

We present an information-geometric measure to systematically investigate neuronal firing patterns, taking account not only of the second-order but also of higher-order interactions. We begin with the case of two neurons for illustration and show how to test whether or not any pairwise correlation in one period is significantly different from that in the other period. In order to test such a hypothesis of different firing rates, the correlation term needs to be singled out ‘orthogonally’ to the firing rates, where the null hypothesis might not be of independent firing. This method is also shown to directly associate neural firing with behavior via their mutual information, which is decomposed into two types of information, conveyed by mean firing rate and coincident firing, respectively. Then, we show that these results, using the ‘orthogonal’ decomposition, are naturally extended to the case of three neurons and n neurons in general.

1 Introduction

Based on the theory of hierarchical structure and related invariant decomposition of interactions by information geometry [3], the present paper briefly summarizes methods useful for systematically analyzing a population of neural firing [9].

Many researches have shown that the mean firing rate of a single neuron may carry significant information on sensory and motion signals. Information conveyed by populational firing, however, may not be only an accumulation of mean firing rates. Other statistical structure, e.g., coincident firing [13, 14], may also carry behavioral information. One obvious step to investigate this issue is to single out a contribution by coincident firing between two neurons, i.e., the pairwise correlation [2, 6].

In general, however, it is not sufficient to test a pairwise correlation of neural firing, because there can be triplewise and higher correlations. For example, three variables (neurons) are not independent in general even when they are pairwise independent.

We need to establish a systematic method of analysis, including these higher-order
correlations [1, 5, 7, 13]. We propose one approach, the information-geometric measure that uses the dual orthogonality of the natural and expectation parameters in exponential family distributions [4]. We represent a neural firing pattern by a binary random vector \( \mathbf{x} \). The probability distribution of firing patterns can be expanded by a log linear model, where the set \( \{ p(\mathbf{x}) \} \) of all the probability distributions forms a \((2^n - 1)\)-dimensional manifold \( \mathbb{S}_n \). Each \( p(\mathbf{x}) \) is given by \( 2^n \) probabilities

\[
 p_{i_1 \cdots i_n} = \text{Prob} \{ X_1 = i_1, \cdots, X_n = i_n \}, \quad i_k = 0, 1, \quad \text{subject to} \sum_{i_1, \cdots, i_n} p_{i_1 \cdots i_n} = 1
\]

and expansion in \( \log p(\mathbf{x}) \) is given by

\[
 \log p(\mathbf{x}) = \sum \theta_i x_i + \sum_{i < j} \theta_{ij} x_i x_j + \sum_{i < j < k} \theta_{ijk} x_i x_j x_k \cdots + \theta_{1 \cdots n} x_1 \cdots x_n - \psi,
\]

where indices of \( \theta_{ijk} \), etc. satisfy \( i < j < k \), etc. We can have a general theory of this \( n \) neuron case [3, 9], however, to be concrete given the limited space, we mainly discuss two and three neuron cases in the present paper. Our method shares some features with previous studies (e.g. [7]) in use of the log linear model. Yet, we make explicit use of the dual orthogonality so that the method becomes more transparent and more systematic.

In the present paper, we are interested in addressing two issues: (1) to analyze correlated firing of neurons and (2) to connect such a technique with behavioral events. In (1), previous studies often assumed independent firing as the null hypothesis. However, for example, when we compare firing patterns in two periods, as control and ‘test’ periods, there may exist a weak correlation in the control period. Hence, benefiting from the ‘orthogonal’ coordinates, we develop a method applicable to the null hypothesis of non-independent firing, irrespective of firing rates. It is equally important to relate such a method with investigation of behavioral significance as (2). We show that we can do so, using orthogonal decomposition of the mutual information (MI) between firing and behavior [11, 12].

In the following, we discuss first the case of two neurons and then the case of three neurons, demonstrating our method with artificial simulated data. The validity of our method has been shown also with experimental data [9, 10] but not shown here due to the limited space.

## 2 Information-geometric measure: case of two neurons

We denote two neurons by \( X_1 \) and \( X_2 \) (\( X_i = 1, 0 \) indicates if neuron \( i \) is silent or not in a short time bin). Its joint probability \( p(\mathbf{x}) \), \( \mathbf{x} = (x_1, x_2) \), is given by \( p_{ij} = \text{Prob} \{ x_1 = i; x_2 = j \} > 0 \), \( i, j = 0, 1 \). Among four probabilities, \( \{ p_{00}, p_{01}, p_{10}, p_{11} \} \), only three are free. The set of all such distributions of \( \mathbf{x} \) forms a three-dimensional manifold \( \mathbb{S}_2 \). Any three of \( p_{ij} \) can be used as a coordinate system of \( \mathbb{S}_2 \).

There are many different coordinate systems of \( \mathbb{S}_2 \). The coordinates of the expectation parameters, called \( \eta \)-coordinates, \( \eta = (\eta_1, \eta_2, \eta_{12}) \), is given by

\[
 \eta_i = \text{Prob} \{ x_i = 1 \} = E [x_i], \quad i = 1, 2, \quad \eta_3 = \eta_{12} = E [x_1 x_2] = p_{12},
\]

where \( E \) denotes the expectation and \( \eta_i \) and \( \eta_{12} \) correspond to the mean firing rates and the mean coincident firing, respectively.

As other coordinate systems, we can also use the triplet, \( (\eta_1, \eta_2, \text{Cov} [X_1, X_2]) \), where \( \text{Cov} [X_1, X_2] \) is the covariance and/or the triplet \((\eta_1, \eta_2, \rho)\), where \( \rho \) is the correlation coefficient (COR), \( \rho = \frac{\eta_{12} - \eta_1 \eta_2}{\sqrt{\eta_1 (1-\eta_1) \eta_2 (1-\eta_2)}} \), often called N-JPSTH [2].
Which quantity would be convenient to represent the pairwise correlational component? It is desirable to define the degree of the correlation independently from the marginals \((\eta_1, \eta_2)\). To this end, we use the ‘orthogonal’ coordinates \((\eta_1, \eta_2, \theta)\), originating from information geometry of \(S_2\), so that the coordinate curve of \(\theta\) is always orthogonal to those of \(\eta_1\) and \(\eta_2\).

The orthogonality of two directions in \(S_2\) (\(S_n\) in general) is defined by the Riemannian metric due to the Fisher information matrix \([8, 4]\). Denoting any coordinates in \(S_n\) by \(\xi = (\xi_1, \ldots, \xi_n)\), the Fisher information matrix \(G\) is given by

\[
G = (g_{ij}), \quad g_{ij}(\xi) = E \left[ \frac{\partial}{\partial \xi_i} l(\mathbf{x}; \xi) \frac{\partial}{\partial \xi_j} l(\mathbf{x}; \xi) \right].
\]

(1)

where \(l(\mathbf{x}; \xi) = \log p(\mathbf{x}; \xi)\). The orthogonality between \(\xi_i\) and \(\xi_2\) is defined by \(g_{ij}(\xi) = 0\). In case of \(S_2\), we desire to have \(E \left[ \frac{\partial}{\partial \eta_1} l(\mathbf{x}; \eta_1, \eta_2, \theta) \frac{\partial}{\partial \eta_1} l(\mathbf{x}; \eta_1, \eta_2, \theta) \right] = 0 \quad (i = 1, 2)\). When \(\theta\) is orthogonal to \((\eta_1, \eta_2)\), we say that \(\theta\) represents pure correlations independently of marginals. Such \(\theta\) is given by the following theorem.

**Theorem 1.** The coordinate

\[
\theta = \log \frac{p_{1210}}{p_{0010}}
\]

(2)

is orthogonal to the marginals \(\eta_1\) and \(\eta_2\).

We have another interpretation of \(\theta\). Let’s expand \(p(\mathbf{x})\) by \(\log p(\mathbf{x}) = \sum_{i=1}^{2} \theta_i x_i + \theta_{12} x_1 x_2 - \psi\). Simple calculation lets us get the coefficients, \(\theta_1 = \log \frac{p_{1210}}{p_{0010}}, \quad \theta_2 = \log \frac{p_{0120}}{p_{0010}}, \quad \psi = -\log p_{00}, \quad \text{and} \quad \theta = \theta_{12}\) (as Eq 2). The triplet \(\theta = (\theta_1, \theta_2, \theta_{12})\) forms another coordinate system, called the natural parameters, or \(\theta\)-coordinates. We remark that \(\theta_{12}\) is 0 when and only when \(X_1\) and \(X_2\) are independent.

The triplet

\[
\zeta \equiv (\eta_1, \eta_2, \theta_{12})
\]

forms an ‘orthogonal’ coordinate system of \(S_2\), called the mixed coordinates \([4]\).

We use the Kullback-Leibler divergence (KL) to measure the discrepancy between two probabilities \(p(\mathbf{x})\) and \(q(\mathbf{x})\), defined by \(D[p : q] = \sum \mathbf{x} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})}\). In the following, we denote any coordinates of \(p\) by \(\xi\) etc (the same for \(q\)). Using the orthogonality between \(\eta\) and \(\theta\)-coordinates, we have the decomposition in the KL.

**Theorem 2.**

\[
D[p : q] = D[p : r^*] + D[r^* : q], \quad D[q : p] = D[q : r^{**}] + D[r^{**} : p],
\]

(3)

where \(r^*\) and \(r^{**}\) are given by \(\zeta^{r^*} = (\eta_1^r, \eta_2^r, \theta_1^r)\) and \(\zeta^{r^{**}} = (\eta_1^{**}, \eta_2^{**}, \theta_1^{**})\), respectively.

The squared distance \(ds^2\) between two nearby distributions \(p(\mathbf{x}, \xi)\) and \(p(\mathbf{x}, \xi + d\xi)\) is given by the quadratic form of \(d\xi\),

\[
ds^2 = \sum_{i,j \in \{1, 2, 3\}} g_{ij}(\xi)d\xi_id\xi_j,
\]

which is approximately twice the KL, i.e., \(ds^2 \approx 2D[p(\mathbf{x}, \xi) : p(\mathbf{x}, \xi + d\xi)]\).

Now suppose \(\xi\) is the mixed coordinates \(\zeta\). Then, the Fisher information matrix is of the form \(g^\zeta_{ij}\)

\[
\begin{bmatrix}
  g_{11} & g_{12} & 0 \\
  g_{12} & g_{22} & 0 \\
  0 & 0 & g_{33}
\end{bmatrix}
\]

and we have \(ds^2 = ds_1^2 + ds_2^2\), where \(ds_1^2 = g_{11}d\eta_1 d\eta_1\), \(ds_2^2 = \sum_{i,j \in \{1, 2\}} g_{ij}d\eta_id\eta_j\), corresponding to Eq 3.
This decomposition comes from the choice of the orthogonal coordinates and gives us the merits of simple procedure in statistical inference. First, let us estimate the parameter \( \eta = (\eta_1, \eta_2) \) and \( \theta \) from \( N \) observed data \( x_1, \ldots, x_N \). The maximum likelihood estimator (mle) \( \hat{\eta} \), which is asymptotically unbiased and efficient, is easily obtained by \( \hat{\eta}_i = \frac{1}{N}\#\{x_i = 1\} \) and \( \hat{\theta} = \log \frac{\eta_1(1-B_1) - \eta_2 + \eta_1}{(\eta_1 - \eta_2) / (\eta_1 - \eta_2)} \). Using \( \hat{\eta}_2 = \frac{1}{N}\#\{x_1x_2 = 1\} \). The covariance of estimation error, \( \Delta \eta \) and \( \Delta \theta \), is given asymptotically by \( \text{Cov} \left[ \frac{\Delta \eta}{\Delta \theta} \right] = \frac{1}{N} G^{-1} \). Since the cross terms of \( G \) or \( G^{-1} \) vanish for the orthogonal coordinates, we have \( \text{Cov} [\Delta \eta, \Delta \theta] = 0 \), implying that the estimation error \( \Delta \eta \) of marginals and that of interaction are mutually independent. Such a property does not hold for other non-orthogonal parameterization such as the COR \( \rho \), the covariance etc. Second, in practice, we often like to compare many spike distributions, \( q(x(t)) \) (i.e., \( \zeta(t) \)) for \( t = 1, \ldots, T \), with a distribution in the control period \( p(x) \), or \( \tilde{\zeta}^0 \). Because the orthogonality between \( \eta \) and \( \theta \) allows us to treat them independently, these comparisons become very simple.

These properties bring a simple procedure of testing hypothesis concerning the null hypothesis

\[
H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta \neq \theta_0, \tag{4}
\]

where \( \theta_0 \) is not necessarily zero, whereas \( \theta_0 = 0 \) corresponds to the null hypothesis of independent firing, which is often used in literature in different setting. Let the log likelihood of the models \( H_0 \) and \( H_1 \) be, respectively,

\[
l_0 = \max_{\eta} \log p(x_1, \ldots, x_N; \eta, \theta_0) \quad \text{and} \quad l_1 = \max_{\eta, \theta} \log p(x_1, \ldots, x_N; \eta, \theta). \tag{5}
\]

The likelihood ratio test uses the test statistics \( \lambda = 2 \log \frac{l_0}{l_1} \). By the mle with respect to \( \eta \) and \( \theta \), which can be performed independently, we have

\[
l_0 = \log p(\tilde{x}, \tilde{\eta}, \theta_0), \quad l_1 = \log p(\tilde{x}, \tilde{\eta}, \hat{\eta}_2),
\]

where \( \tilde{\eta} \) are the same in both models. A similar situation holds in the case of testing \( \eta = \eta_0 \) against \( \eta \neq \eta_0 \) for unknown \( \theta \).

Under the hypothesis \( H_0 \), \( \lambda \) is approximated for a large \( N \) as

\[
\lambda = 2 \sum_{i=1}^{N} \log \frac{p(x_i; \tilde{\eta}, \theta_0)}{p(x_i; \tilde{\eta}, \hat{\eta}_2)} \approx N \frac{\partial^2 \zeta}{\partial \theta^2} (\theta - \theta_0)^2 \sim \chi^2(1). \tag{6}
\]

Thus, we can easily submit our data to a hypothetical testing of significant coincident firing against null hypothesis of any correlated firing, independently from the mean firing rate modulation\(^1\).

We now turn to relate the above approach with another important issue, which is to relate such a coincident firing with behavior. Let us denote by \( Y \) a variable of discrete behavioral choices. The MI between \( X = (X_1, X_2) \) and \( Y \) is written by

\[
I(X, Y) = E_{p(X, Y)} \left[ \log \frac{p(x, y)}{p(x)p(y)} \right] = E_{p(Y)} [D[p(X|y) : p(X)]].
\]

Using the mixed coordinates for \( p(X|y) \) and \( p(X) \), we have \( D[p(X|y) : p(X)] = D[\zeta(x|y) : \zeta(X)] = D[\zeta(x|y) : \zeta^*] + D[\zeta^* : \zeta(X)], \) where \( \zeta^* = \zeta(X, y) = \left( \eta_1(x, y), \eta_2(x, y), \theta_1(x) \right) \).

\(^1\) A more proper formulation in this hypothetical testing can be derived, resulting in using \( p \) value from \( \chi^2(2) \) distribution, but we omit it here due to the limited space [9].