Optimize the Distribution of Preferred Stimulus in a Population Code

Si Wu and Hiroyuki Nakahara
RIKEN Brain Science Institute
The Institute for Physical and Chemical Research
Hirosawa 2-1, Wako-shi, Saitama, Japan
{phwusi,hiro}@brain.riken.go.jp

Abstract

We consider two methods to optimize the distribution of preferred stimulus in a population code based on the knowledge of the distribution of stimulus. One method is to maximize the mean Fisher information of the population with respect to the stimulus ensemble. The other is to minimize the lower bound of the mean decoding error. The implication of the two methods is discussed.

1 Introduction

In a population code information is processed distributely by the joint activities of a group of neurons [2, 7, 11, 15, 16, 17]. Neurons, for example, in early sensory areas, respond to stimuli with their tuning functions and, as a group, should cover the whole range of potentially valuable stimuli. A stimulus induces a dynamic state of the population from which the brain would infer the external world. A good property of the population code is that the noise in the coding of a single neuron can be averaged out. Some biological evidences support the population code[12]. For example, in the medial temporal region (MT), the joint activities of neurons may encode the direction of visual motion [8, 9].

In this study, we consider the problem of optimizing the distribution of preferred stimulus (DPS) in a population code with respect to the probabilistic nature of stimuli. Stimulus ensembles often possess some statistical structures, for example, the translation- and scale-invariant property in the ensemble of natural images[1, 5, 6]. This regularity is believed to be utilized in early vision processing to construct an efficient coding [1]. Hence it is interesting to investigate how the structure of the stimulus ensemble affects the optimal population coding. Brunel and Nadal [3] discussed the relationship between Fisher information and mutual information in the context of population code. The former is more related to the decoding accuracy and the latter to the encoding accuracy. They obtained optimal DPS through maximizing the mutual information between the stimulus and the response for the one dimensional case. In this work we solve the same problem based on the Fisher information. To consider optimization in a coding scheme, it is important to specify the nominal criterion for optimization (i.e., the cost function). We discuss two different optimization criterion. One is to maximize the mean Fisher information of the population with respect to the stimulus ensemble, and the other is to minimize the lower bound of the mean decoding error.

2 Maximizing the Mean Fisher Information

Consider a population of N neurons coding a variable x. Denote $\mathbf{r} = \{r_i\}$ the activities of neurons, i.e., the number of spikes emitted by neurons during a fixed time interval. The mean value of r_i after many trials is given by

$$\langle r_i \rangle = f_i(x), \tag{1}$$

where $f_i(x)$ is the tuning function of the ith neuron, here assumed to be a normalized Gaussian,

$$f_i(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-c_i)^2/2\sigma^2}, \tag{2}$$

where c_i is the preferred stimulus of the ith neuron.

In a population code, encoding can be considered from the conditional probability $P(\mathbf{r}|x)$ and decoding from $P(x|\mathbf{r})$. For simplicity we assume all neurons' activities are uncorrelated, i.e.,

$$P(\mathbf{r}|x) = \prod_i P(r_i|x). \tag{3}$$

There are two common noise models to determine $P(r_i|x)$. One model assumes that r_i is Poisson distributed, i.e.,

$$P(r_i|x) = e^{-f_i(x)} \frac{f_i(x)^{r_i}}{r_i!}. \tag{4}$$

The other one assumes r_i is Gaussian distributed, i.e.,

$$P(r_i|x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(r_i - f_i(x))^2/2\sigma_0^2}, \tag{5}$$

where σ_0 is the variance.

For any unbiased estimate $\hat{x}$, the Fisher information provides a useful measure for the decoding accuracy, that is, the averaged square decoding error $\langle(\hat{x}-x)^2\rangle$ is greater than or equal to the inverse of the Fisher information J_x. This is called the Cramér-Rao bound [4]. In a population code, the Fisher information is given by

$$\begin{aligned} J_x &= \int P(\mathbf{r}|x)\left[\frac{d\ln P(\mathbf{r}|x)}{dx}\right]^2 d\mathbf{r} \\ &= \sum_i J_{c_i,x}, \end{aligned} \tag{6}$$

where $J_{c_i,x}$ is the Fisher information of the ith neuron,

$$J_{c_i,x} = \int P(r_i|x)\left[\frac{d\ln P(r_i|x)}{dx}\right]^2 dr_i. \tag{7}$$

Maximum likelihood (ML) inference is a decoding scheme to asymptotically achieve the Cramér-Rao bound when there is no correlation between neurons' responses (If correlation is considered, the asymptotically efficiency of ML inference needs to be carefully checked [17].). Recently, the work of Pouget et al. [13] shows that a biologically reasonable recurrent network could perform ML inference in the case without correlations. In this work, since we only consider the case that all neurons' activities are uncorrelated, we may regard the Cramér-Rao bound as a realizable limit.

In the large N limit, we can write

$$\begin{aligned} \langle J_x\rangle &= \frac{J_x}{N} \\ &= \int J_{c,x}P(c)dc, \end{aligned} \tag{8}$$

where $\langle J_x\rangle$ is the mean Fisher information of the population with respect to the stimulus x, $P(c)$ is the distribution of preferred stimulus.

Suppose the distribution of the stimulus is $P(x)$, the mean Fisher information of the population with respect to the stimulus ensemble is given by

$$\langle J\rangle = \int J_{c,x}P(c)P(x)dcdx. \tag{9}$$

We can also write $\langle J\rangle$ as

$$\langle J\rangle = \int J_c P(c)dc, \tag{10}$$

where

$$J_c = \int J_{c,x}P(x)dx, \tag{11}$$

is the Fisher information of the neuron having preferred stimulus c. Larger J_c means that this neuron is more informative with respect to the stimulus ensemble.

It is natural to maximize the mean Fisher information (10) to get an optimal distribution of $P(c)$. However, this is an ill-posed problem [3]. The solution turns out to be a δ function, $P(c)=\delta(c-c^*)$, where c^* denoting the point where J_c has the maximum value. It is easy to see that this solution is unreasonable. Actually, when there is only one neuron, it is impossible to construct an unbiased estimator due to the symmetry of the tuning function, that is, $f(x)=f(2c-x)$. The stimuli x and $2c-x$ are indistinguishable. Therefore, the Cramér-Rao bound for unbiased estimators by which the solution is derived does not hold.

A way to make an ill-posed problem well-defined is to impose a regularization constraint. Intuitively, we may want $P(c)$ to be distributed and to have a large value at points where J_c is large. This can be achieved by applying a regularization term corresponding to the Kullback divergence between the normalized J_c and $P(c)$, i.e., $\int \tilde{J}_c \ln \tilde{J}_c/P(c)dc$, where $\tilde{J}_c = J_c/\int J_c dc$. In practice, this regularization term is simplified to be $\int J_c \ln P(c)dc$, since the difference between them is irrelevant to $P(c)$.

In summary, the problem of optimizing DPS is converted to

$$\text{Maximizing} \quad \langle J\rangle = \int J_c P(c)dc + \gamma\int J_c \ln P(c)dc, \tag{12}$$

$$\text{Subject to} \quad \int P(c)dc = 1, \tag{13}$$

$$P(c) > 0, \tag{14}$$

where γ is a positive number.

The solution of the above optimization problem is

$$P(c) = \frac{\gamma J_c}{\lambda - J_c}, \tag{15}$$

where λ is a constant to be determined by the normalization condition (13). It is easy to check that $P(c) > P(c')$ when $J_c > J_{c'}$, which is the property we want. The parameter γ controls the balance between information maximization and broad distribution.

Obviously, when $P(x)$ is uniform, the optimal $P(c)$ is also uniform. As an another example, we consider $P(x)$ to be a Gaussian distribution,

$$P(x) = \frac{1}{\sqrt{2\pi}\mu}e^{-x^2/2\mu^2}. \tag{16}$$

By using the Poisson noise model (4), we get

$$J_c = \frac{1}{2\pi\sigma(\mu^2+\sigma^2)^{3/2}}\left[\frac{\mu^2}{\sigma^2}+\frac{c^2}{\mu^2+\sigma^2}\right]e^{-c^2/2(\mu^2+\sigma^2)} \tag{17}$$

It is interesting to look into the property of J_c, which essentially has two kinds of shape (see Fig.1). When $2\sigma^2 > \mu^2$, it has one peak centered at the original point (Fig.1(a)); when $2\sigma^2 < \mu^2$, it has two peaks (Fig.1(b)). The result for the Gaussian noise model is similar except that the value $2\sigma^2$ is replaced by σ^2. Fig.1 shows examples of the optimal $P(c)$ in two different parameter settings. Intuitively, the result shows that if the stimulus ensemble is broadly distributed, then the distribution of preferred stimuli $P(c)$ is also distributed.

An alternative choice of a regularization constraint is a smoothness criterion, for example, $\int (dP(c)/dc)^2 dc$. The shortcoming of this criterion is that we need an assumption on the form of $P(c)$ to make sure that the constraint (14) is satisfied.

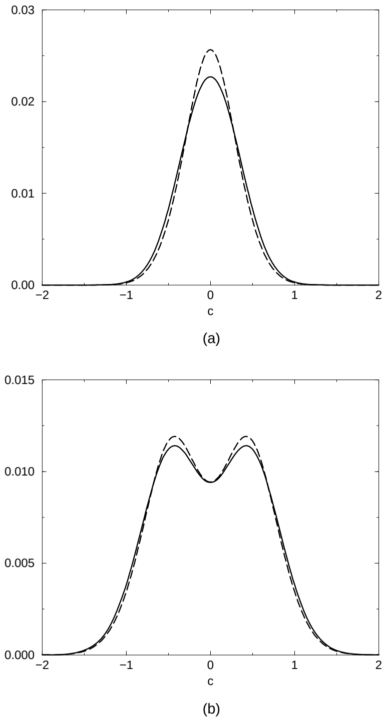

Figure 1: The optimal $P(c)$ in two different parameter settings. The noise model is the Poisson distribution one. The full line denotes the normalized J_c. The dashed line denotes $P(c)$. (a) $\mu = 0.3, \sigma = 0.1, \gamma = 0.05$; (b) $\mu = 0.3, \sigma = 0.3, \gamma = 0.05$.

3 Minimizing the Lower Bound of the Mean Decoding Error

We can also solve the above optimization problem by using a different criterion, that is, to minimize the lower bound of the mean decoding error. According to the Cramér-Rao bound, we have

$$\langle\langle(\hat{x} - x)^2\rangle\rangle \geq \int \frac{P(x)}{N \int J_{c,x} P(c) dc} dx \qquad (18)$$

where $\langle\langle\cdot\rangle\rangle$ denoting the average over sampling and the distribution of the stimulus.

Minimizing the right side of the above inequality with respect to $P(c)$ leads to

$$\int J_{c,x} P(c) dc = \lambda P(x)^{1/2}, \qquad (19)$$

where λ is a constant to be determined by the normalization condition $\int P(c) dc = 1$. To get the result (19), we have used the condition that $\int J_{c,x} dx$ is a constant, which is generally true in the context of the present study. Thus the optimal $P(c)$ is obtained when the convolution of $P(c)$ with the Fisher information $J_{c,x}$ matches the square root of the distribution of stimulus. Note that the solution is different from the one obtained by maximizing the mutual information, where the convolution of $P(c)$ is to match $P(x)$ instead of $P(x)^{1/2}$ [3]. The difference comes from the use of different criterion. Recall that maximizing the mutual information is to enhance the encoding accuracy, whereas minimizing the lower bound of the mean decoding error via the Cramér-Rao bound is more related to the decoding accuracy.

4 Conclusion

In this study two approaches are employed to optimize DPS in a population code based on the knowledge of the distribution of stimulus. One is to maximize the mean Fisher information with respect to the stimulus ensemble. To make the problem solvable, we use a regularization term imitating the Kullback divergence between DPS and the normalized Fisher information with respect to the stimulus ensemble. We found that depending upon the distribution of stimulus ensemble, double-peaked DPS is possible. The connection of this finding with biological neurons is a future research. The other is to minimize the lower bound of the mean decoding error. The present study aimed to briefly sketch the two approaches by focusing solvable cases and to illuminate their difference. Certainly, the feature research should be extended to investigate a more biologically plausible situation of stimulus ensemble[10], higher dimension case[19], and encoding-decoding relationship[18].

Acknowledgments

We thank Prof. Shun-ichi Amari for constructive suggestions. The authors acknowledge the support of the RIKEN Brain Science Institute for funding this project.

References

[1] Atick, J. J. & Redlich, A. N., What does the retina know about natural scenes? *Neural Computation, 4*, 196-210, 1992.

[2] Baldi, P. & Heiligenberg, W., How sensory maps could enhance resolution through

ordered arrangements of broadly tuned receivers. *Biological Cybernetics, 59*, 313-318, 1988.

[3] Brunel, N. & Nadal, J.-P., Mutual information, Fisher information, and Population coding. *Neural Computation, 10*, 1731-1757, 1998.

[4] Cover, T. M. & Thomas, J. A., *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[5] Dong, D. W. & Atick, J. J., Statistics of natural time-varying images. *Network: Comput. Neural Syst. 6*, 345-58, 1995.

[6] Field, D. J., Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A 4*, 2379-94, 1987.

[7] Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E., Neuronal population coding of movement direction. *Science, 243*, 1416-1419, 1986.

[8] Maunsell, J. H. R. & Van Essen, D. C., Functional properties of neurons in middle temporal visual area of the Macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology, 49(5)*, 1127-1147, 1983.

[9] Newsome, W. T. & Pare, E. B., A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience, 8*, 2201-2211, 1988.

[10] Olshausen, B. A. & Field, D. J., Natural image statistics and efficient coding. *Network: Computation in Neural Systems, 7*, 333-339, 1996.

[11] Paradiso, M. A., A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics, 58*, 35-49, 1988.

[12] Parker, A. J. & Newsome, W. T., Sense and the single neuron: Probing the physiology of perception. *Annu. Rev. Neurosci. 21*, 227-277, 1998.

[13] Pouget, A., Zhang, K., Deneve, S. & Latham, P. E., Statistically efficient estimation using population coding. *Neural Computation, 10*, 373-401, 1998.

[14] Sanger, T. D., Probability density estimation for the interpretation of neural population codes, *Journal of Neurophysiology, 76(4)*, 2790-2793, 1996.

[15] Salinas, E. & Abbott, L. F., Vector reconstruction from firing rates. *Journal of Computational Neuroscience, 1*, 89-107, 1994.

[16] Seung, H. S. & Sompolinsky, H., Simple models for reading neuronal population codes. *Proceedings of the National Academy of Science, USA, 90*, 10749-10753, 1993.

[17] Wu, S., Nakahara, N., Murata, N. & Amari, S., Population decoding based on an unfaithful model. (preprint).

[18] Zemel, R. S., Dayan, P. & Pouget, A., Probabilistic interpretation of population codes. *Neural Computation, 10*, 403-430, 1998.

[19] Zhang, K. & Sejnowski, T. J., Neuronal tuning: to sharpen or broaden? *Neural Computation, 11*, 75-84, 1999.