

Accuracy and learning in neuronal populations

Kechen Zhang^{1,*} and Terrence J. Sejnowski^{1,2}

¹ Howard Hughes Medical Institute, Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

² Department of Biology, University of California, San Diego, La Jolla, CA 92093, USA

Introduction

The information about various sensory and motor variables contained in neuronal spike trains may be quantified by either Shannon mutual information or Fisher information. Although they are related, the Fisher information measure is more convenient for dealing with continuous variables which are more common in lower level sensory and motor representations. The accuracy of encoding and decoding by a population of neurons as described by Fisher information has some general properties, including a universal scaling law with respect to the width of the tuning functions. The theoretical accuracy for reading out information from population activity can be reached, in principle, by Bayesian reconstruction method, which can be simplified by exploiting Poisson spike statistics. The Bayesian method can be implemented by a feedforward network, where the desired synaptic strength can be established by a Hebbian learning rule that is proportional to the logarithm of the presynaptic firing rate, suggesting that the method might be potentially relevant to biological systems.

* Corresponding author: Kechen Zhang, Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. Tel.: +1-858-453-4100, ext. 1420; Fax: +1-858-587-0417; E-mail: zhang@salk.edu

Accuracy of neural coding: Fisher information vs. Shannon information

Neuronal spike trains carry information about various sensory and motor variables. To quantify the amount of information, Shannon's mutual information has often been used (Rieke et al., 1997). Fisher information was introduced in the 1920s, two decades earlier than the introduction of Shannon information, but the application of Fisher information to neural coding is more recent (Paradiso, 1988; Lehky and Sejnowski, 1990).

To compare the two information measures, consider a neuron that encodes a one-dimensional variable x by the number of spikes n evoked within a certain time interval. Shannon mutual information between the encoded parameter x and the number of spikes n is defined as

$$I = \sum_{n,x} p(n, x) \ln \frac{p(n, x)}{p(n)p(x)} \quad (1)$$

$$= \sum_{n,x} p(n | x)p(x) \ln \frac{p(n | x)}{p(n)}, \quad (2)$$

where the second step is an identity based on the definition of conditional probability $p(n, x) = p(n | x)p(x)$, together with $p(n) = \sum_x p(n | x)p(x)$. The sums here are over all possible values of the encoded variable x and the number of spikes n . If x is continuous, the sum is understood as an approximation that approaches an integral. Fisher information is defined

as

$$J(x) = \left\langle \left(\frac{\partial}{\partial x} \ln p(n | x) \right)^2 \right\rangle \quad (3)$$

$$= \sum_n \frac{p'(n | x)^2}{p(n | x)}, \quad (4)$$

where the average $\langle \rangle$ is over n with respect to the conditional probability distribution $p(n | x)$, and the second step is the result of the average, with $p'(n | x) = \partial p(n | x) / \partial x$ being a derivative function that describes how the firing of the neuron is affected by a slight change of the encoded variable x .

Note that Fisher information $J(x)$ is defined for each value of the encoded variable x , whereas Shannon information averages over the whole range of values according to the distribution $p(x)$. Thus Shannon information is a global measure and Fisher information is a local measure for the coding of the variable of interest. Fisher information is defined for continuous variable x only, and is closely related to the minimal variance for estimating the variable from the spikes.

These two information measures are related in several ways. When a probability distribution is perturbed by adding gaussian noise, the rate of change of Shannon entropy is proportional to the Fisher information according to de Bruijn's identity (Cover and Thomas, 1991). Another relation is that Shannon mutual information between a probability density and its slightly shifted version is proportional to Fisher information (Frieden, 1998). The two measures are also related by an inequality because given the Fisher information or given the variance of parameter estimation, the entropy cannot exceed that of a gaussian distribution with the same variance (Brunel and Nadal, 1998).

Examples of Fisher information

Sensory and motor variables are typically represented by a population of broadly tuned neurons. We consider several commonly encountered tuning functions and give the total Fisher information for a population of these neurons. Fisher information J can characterize how accurately a variable x is encoded by the population because its inverse is the

Cramér-Rao lower bound on the variance or mean square error:

$$E[\varepsilon^2] \geq \frac{1}{J}, \quad (5)$$

which applies to all possible unbiased estimation methods that can read out the variable x from the population activity without systematic error (Paradiso, 1988; Seung and Sompolinsky, 1993; Snippe, 1996). Here ε^2 is the square error for estimating variable x in a single trial, and $E[\varepsilon^2]$ is the average error across repeated trials with a fixed x . The error is caused by the randomness of spikes so that the true value of x can never be completely determined, regardless of the method for reading out the information. If the population of neurons represent a continuous D -dimensional vector variable $\mathbf{x} = (x_1, x_2, \dots, x_D)$, the square error in a single trial is given by

$$\varepsilon^2 = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_D^2, \quad (6)$$

with ε_i the error for estimating x_i . Eq. 5 is still valid assuming that different representations in different dimensions are uncorrelated.

In the following, we always assume that there are N neurons with identical tuning functions whose peak positions are scattered uniformly across the range of the encoded variable. Because of the uniform distribution, the Fisher information is the same for all value of the variable. The spikes from different neurons are assumed to be independent so that the total Fisher information for the whole population is the sum of the Fisher information for each of the neurons. The spike statistics are assumed to be Poisson, a reasonable first approximation. The time interval in which the spikes are collected is always indicated by τ .

Example 1: cosine tuning

A cosine tuning function is given by

$$f = A \cos \theta + B, \quad (7)$$

where f is the mean firing rate, θ is the encoded variable, and A and B are two constants. Here θ is a one-dimensional variable, corresponding to a two-dimensional workspace. The total Fisher information for N neurons with preferred directions distributed

uniformly around the circle is

$$J = \tau N (B - \sqrt{B^2 - A^2}). \quad (8)$$

Example 2: circular normal tuning

The mean firing rate of a circular normal tuning function is given by

$$f = C \exp(K \cos \theta), \quad (9)$$

where θ is a one-dimensional circular variable, C and K are constants. The total Fisher information for N neurons is

$$J = \tau N C K I_1(K), \quad (10)$$

where $I_1(K) = i J_{-1}(iK)$ is the modified Bessel function of imaginary argument, with J_{-1} being Bessel function of the first kind of the order -1 . For small K , the Bessel function becomes

$$I_1(K) \approx K^{-1} - \sqrt{K^{-2} - 1}, \quad (11)$$

and the circular normal tuning function in Eq. 9 approaches the cosine tuning function in Eq. 7, with $B \approx C$ and $A \approx CK$. Now the Fisher information formula in Eq. 8 is recovered. For large K , the Bessel function becomes

$$I_1(K) \approx \exp(K) / \sqrt{2\pi K}, \quad (12)$$

and the circular normal tuning function in Eq. 9 approaches the gaussian tuning function in Eq. 13 ($D = 1$), with $\sigma^2 \approx 1/K$, $F \approx C \exp(K)$ and $\eta \approx N/(2\pi)$, so that the Fisher information formula in Eq. 14 ($D = 1$) is recovered.

Example 3: gaussian tuning

The mean firing rate of a gaussian tuning function in a space of dimension D is given by

$$f = F \exp\left(-\frac{x_1^2 + \dots + x_D^2}{2\sigma^2}\right), \quad (13)$$

where (x_1, x_2, \dots, x_D) are the encoded variables, F is the peak firing rate and σ is the tuning width parameter. The centers of the tuning functions for different neurons are assumed to be uniformly distributed. Total Fisher information is

$$J = \frac{(2\pi)^{D/2}}{D} \eta \tau F \sigma^{D-2}, \quad (14)$$

where η is the number of neurons per unit volume in the D -dimensional space of the encoded variables (Zhang et al., 1998).

Correlated noise

In the examples above, the total Fisher information is always proportional to the total number of neurons as well as the total number of spikes from the whole population of neurons. This is a consequence of the assumptions of the independence of different cells and the Poisson spike statistics, which implies that the information for each cell is proportional to f'^2/f , which in turn is proportional to the peak firing rate. Here f is the tuning function and f' is its derivative with respect to the encoded variable.

The property that Fisher information is proportional to the total number of neurons and total number of spikes still holds even when the neurons are not independent but have pairwise correlations of their noise. The initial observation (Zohary et al., 1994) that even weak correlated noise may destroy this proportionality depends on how this information is read out (Abbott and Dayan, 1999). The exact results also depends on the form of the correlations (Yoon and Sompolinsky, 1999). Sometimes even when the correlated noise is deliberately ignored, the decoding error may still be the same as in the independent cases (Wu et al., 2000). In particular, as shown below, the universal scaling law for tuning width is insensitive to some common forms of noise correlation.

Universal scaling law for tuning width

In Example 3 in the preceding section, the Fisher information scales with the tuning width σ according to $\eta \sigma^{D-2}$, where D is the dimension of the encoded variable, and η is the density of neuron for the encoded variable. This specific example assumes gaussian tuning function, Poisson spike statistics and independence of different neurons. In general, the accuracy of population coding by tuned neurons as a function of tuning width follows the same universal scaling law regardless of the exact shape of the tuning function and the exact probability distribution of spikes, and allows some correlated noise between neurons.

The general results are described below, followed by an intuitive explanation of this universal scaling

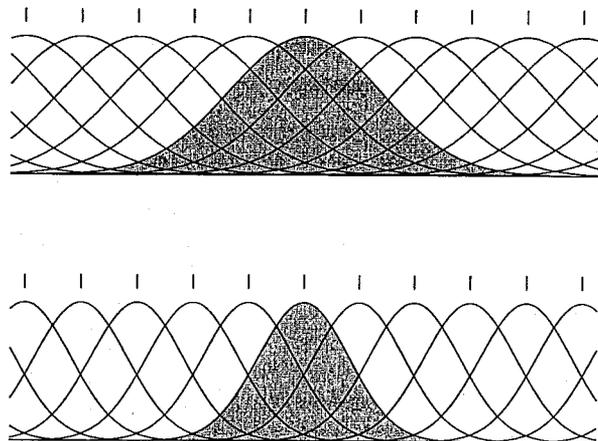


Fig. 1. Schematic diagrams showing the tuning curves of a population of neurons, with each tuning curve describing the mean firing rate of a neuron as a function of the value of the encoded variable. The accuracy of the encoding by the whole population depends on the width of the tuning curves in a universal manner regardless of the exact shape of the curve and the exact spike statistics, provided that the tuning width is not too small compared with the spacing of the tuning curve centers (vertical ticks), and also not too large compared with the full range of the variable.

(Fig. 1). Formal treatment of the result is given in Zhang and Sejnowski (1999a), following earlier works on related issues (Hinton et al., 1986; Baldi and Heiligenberg, 1988; Snippe and Koenderink, 1992; Zohary, 1992; Zhang et al., 1998).

The tuning function can be an arbitrary radially symmetric function that describes the dependence of the mean firing rate $f(\mathbf{x})$ of a neuron on the variable of interest $\mathbf{x} = (x_1, x_2, \dots, x_D)$:

$$f(\mathbf{x}) = F\phi\left(\frac{|\mathbf{x} - \mathbf{c}|^2}{\sigma^2}\right), \quad (15)$$

which depends only on the Euclidean distance to the center \mathbf{c} . Here σ is a parameter that describes the tuning width, which scales the tuning function without changing its shape. This general formula includes all radial symmetric functions, of which gaussian tuning is a special case. The probability $P(n | \mathbf{x}, \tau)$ for n spikes to occur within a time interval of length τ can be an arbitrary function of the mean firing rate $f(\mathbf{x})$.

These general assumptions are strong enough to prove that the total Fisher information has the form:

$$J \propto \eta \sigma^{D-2}, \quad (16)$$

where η is the number of neurons whose tuning function centers fall into a unit volume in the D -dimensional space of encoded variable, assuming that these neurons fire independently, and the centers of the tuning functions are uniformly distributed in the space of the encoded variable. The proportional constant (not shown) may depend on the time interval τ and the exact shape of the tuning function ϕ .

The universal scaling law with the factor σ^{D-2} implies that for one-dimensional feature ($D = 1$), more information can be coded per neuron for a sharper tuning curve, provided that all other factors are fixed, such as peak firing rate and noise correlation. For two-dimensional features ($D = 2$), such as the spatial representation by hippocampal place cells, coding accuracy should be insensitive to the tuning width (Zhang et al., 1998). In three and higher dimensions ($D \geq 3$), such as the multiple visual features represented concurrently in the ventral stream of primate visual system, more information can be coded per neuron by broader tuning. Although sharpening makes individual neurons appear more informative, it reduces the number of simultaneously active neurons, a factor that dominates in higher dimensions where neighboring tuning functions overlap more substantially. On the other hand, sharpening can always improve the Fisher information coded per spike and thus energy efficiency for spike generation (Zhang and Sejnowski, 1999a).

The universal scaling law in Eq. 16 still holds when the firing rate fluctuations of different neurons are weakly correlated. The result is

$$J = \eta \sigma^{D-2} \left(\frac{A + B}{1 - q} + B \right), \quad (17)$$

ignoring contributions from terms slower than linear with respect to the population size, where A and B are constants independent of the correlation strength q (Zhang and Sejnowski, 1999a). Here the noise covariance between neurons i and j is given by the average:

$$C_{ij} = E[(n_i - \mu_i)(n_j - \mu_j)] \\ = \begin{cases} C_i^2 & \text{if } i = j, \\ C_i C_j & \text{otherwise,} \end{cases} \quad (18)$$

where $C_i = \psi(\mu_i)$ is an arbitrary function of the mean number of spikes $\mu_i = E[n_i]$ within certain

time interval. The spike statistics are assumed to follow a multivariate gaussian distribution. Thus the noise correlation does not affect the scaling factor $\eta\sigma^{D-2}$, and the proportional constant is affected by the factor $1/(1-q)$, which slightly increases the Fisher information when there is positive correlation ($q > 0$).

An intuitive derivation of the universal scaling factor $\eta\sigma^{D-2}$ is as follows. The average square error for estimating the encoded variable from the activity of a population of neurons should scale as:

$$E[\varepsilon^2] \propto \frac{\sigma^2}{N'}, \quad (19)$$

where N' is the total number of activated neurons for each fixed value of the encoded variable. The factor $1/N'$ may be justified by the square root law for using a large number of independent neurons. The factor σ^2 arises because of the dimensionality requirement: If the length scale of the encoded variable is changed, both the value of the tuning width and the square error should change accordingly, leading to the factor σ^2 . Because each neuron is tuned with width σ in each of the D dimensions, the number of activated neurons in response to a fixed value of the encoded variable should be proportional to:

$$N' \propto \eta\sigma^D, \quad (20)$$

where η is the density of neuron as before. Substituting Eq. 20 into Eq. 19 yields the average square error:

$$E[\varepsilon^2] \propto \frac{1}{\eta\sigma^{D-2}}. \quad (21)$$

Using Eq. 5, we see that the Fisher information should scale as $\eta\sigma^{D-2}$, which is the same as Eq. 16.

More generally, if the tuning function is not radially symmetric but has different widths for different dimensions, then Eq. 20 should be replaced by

$$N' \propto \eta\sigma_1\sigma_2 \cdots \sigma_D, \quad (22)$$

so that the average square error for the i -th dimension should be:

$$E[\varepsilon_i^2] \propto \frac{\sigma_i^2}{N'} \propto \frac{\sigma_i^2}{\eta\sigma_1\sigma_2 \cdots \sigma_D}. \quad (23)$$

A formal derivation of this result using Fisher information is given in Eurich and Wilke (2000). If

the width σ_i in dimension i is sharpened while the widths of all other dimensions are kept constant, the error should decrease in proportion to σ_i , consistent with the previous result for a one-dimensional problem. When the tuning widths of different dimensions are identical: $\sigma_1 = \sigma_2 = \cdots = \sigma$, the factor $\eta\sigma^{D-2}$ is recovered.

Origin of cosine tuning function

In the above, different shapes of the tuning functions are directly used without considering why a particular shape exists. There is a general argument for why the cosine tuning curves are so widespread in the sensory and motor systems (Zhang and Sejnowski, 1999b). A cosine tuning function implies a dot product between a fixed preferred direction and the actual movement direction (Georgopoulos et al., 1986; Schwartz et al., 1988). This suggests a linear relation with the movement direction, although the actual coding may have various forms (Mussa-Ivaldi, 1988; Sanger, 1994).

The general argument below exploits the inherently low dimensionality of natural movements and shows that neuronal activity tuned to movement often obeys simple generic rules as a first approximation, insensitive to the exact sensory or motor variables that are encoded and the exact computational interpretation (Zhang and Sejnowski, 1999b). Consider the reaching problem and assume that the mean firing rate of a neuron relative to its baseline is proportional to the time derivative of an arbitrary unknown function of hand position in space during stereotyped reaching movement. As a linear approximation, the firing rate can always be written as

$$f = f_0 + pv \cos \alpha, \quad (24)$$

where f_0 is the baseline rate, p is a constant, v is the instantaneous reaching speed of the hand, and α is the angle between the preferred direction and the instantaneous reaching direction.

A similar linear approximation applies to the visual system. In response to a three-dimensional object moving at translational speed v and angular speed ω , the mean firing rate of a motion-sensitive neuron should be given by

$$f = f_0 + pv \cos \alpha + q\omega \cos \beta, \quad (25)$$

where f_0 is the baseline rate, p and q are constants, α is the angle between the instantaneous translational velocity and a fixed preferred direction, and β is the angle between the instantaneous angular velocity and a fixed preferred rotation direction. The assumption here is that the firing rate is proportional to the time derivative of an arbitrary function of the position and orientation of the object in three-dimensional space (Zhang and Sejnowski, 1999b). Thus, given a particular view of a particular object, the response above baseline is predicted to be the sum of two components, one translational and one rotational, each with cosine tuning and multiplicative modulation by speed and angular speed, respectively.

In the motor system, broad cosine-like tuning curves have been observed in many brain areas, including the primary motor cortex, premotor cortex, parietal cortex, cerebellum, basal ganglia, and somatosensory cortex. The tuning rule in Eq. 24 can account for various experimental facts besides the cosine shape, including multiplicative speed modulation, trajectory reconstruction by the population vector, curvature power law, elbow position, and the linear relation between the spike count and reaching distance.

In the visual system, the predicted tuning rule in Eq. 25 has not been tested with realistic moving three-dimensional objects. A partial confirmation of a special case of this tuning rule is provided by neurons selective to wide-field spiral visual motion with broad cosine tuning in monkey medial superior temporal area (MST), the ventral intraparietal area (VIP), and the parietal area 7a, because such motions may be generated plausibly by a large moving planar object facing the observer. A population of such neurons tuned to three-dimensional object movement might be useful for updating static-view representations in the ventral visual stream (Zhang et al., 1998).

Optimal decoding by Bayesian method

Various methods have been used to 'decode' or read out information from the spike trains of a population of neurons. Within a probabilistic framework, the theoretically optimal methods are based on the Bayes rule (Földiák, 1993; Sanger, 1996; Zhang et al., 1998; Brown et al., 1998).

A popular reconstruction method is called the population vector scheme, first used in motor cortex, where the average firing rate of a given neuron is maximal when the arm movement is in a particular direction, known as the preferred direction for that neuron (Georgopoulos et al., 1986). The population vector method estimates the direction of arm movement by summing the preferred direction vectors weighted by the firing rate of each neuron. A more general approach to reconstruction is to allow the neurons to represent more general basis functions of the physical variables (Girosi and Poggio, 1990; Pouget and Sejnowski, 1997). Each neuron contributes a basis function in this space of variables whenever it fires, and the best estimate of the physical variables can be computed from the sum of these functions weighted by the number of spikes occurring in each neuron during a time interval.

An alternative method for decoding a population code is based on Bayesian reconstruction or maximum likelihood estimation. These optimal probabilistic methods take into account prior probabilities and attempt to reconstruct the entire probability distribution. Instead of adding together the kernels, as in the basis function method, the probabilistic approach multiplies them, assuming that the spikes in each neuron are independent.

An example of how the timing of spikes in a population of neurons can be used to reconstruct a physical variable is the reconstruction of the location of a rat in its environment from the place cells in the hippocampus of the rat. The place cells are silent most of the time, and they fire maximally only when the animal's head is within a restricted region in the environment called its place field (Muller et al., 1987; Wilson and McNaughton, 1993). The reconstruction problem is to determine the rat's position based on the spike firing times of a few dozen simultaneously recorded place cells. Examples of the Bayesian method and the direct basis function method applied to the place cell data are shown in Fig. 2.

The Bayesian reconstruction method works as follows. Assume that a population of N neurons encodes several variables (x_1, x_2, \dots) , which will be written as vector \mathbf{x} . Here \mathbf{x} is the position of the animal in the maze. From the number of spikes

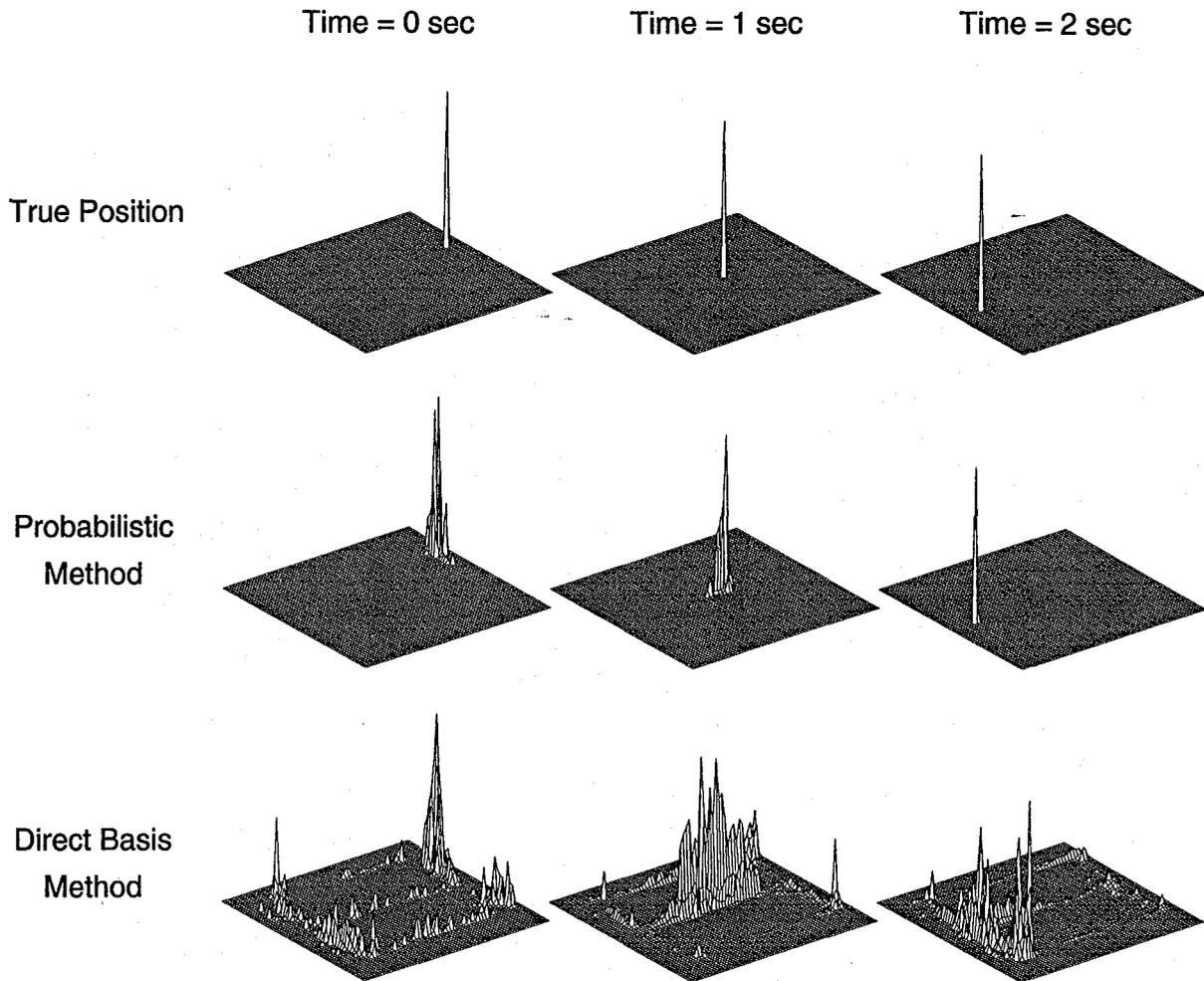


Fig. 2. Predicting the position of a freely moving rat from the spike trains of 25 simultaneously recorded hippocampal place cells. Snapshots of the reconstructed distribution density for the position of the rat are compared with its true position, which occupied a single pixel on the 64×64 grid, corresponding to 111×111 cm in real space. The probabilistic method or Bayesian method often yielded a sharp distribution with a single peak by combining the place fields multiplicatively, whereas the direct basis method typically led to a broader distribution with multiple peaks by combining the place fields additively. Taken from Zhang et al. (1998) with premission.

$\mathbf{n} = (n_1, n_2, \dots, n_N)$ fired by the N neurons within a time interval τ , we want to estimate the value of \mathbf{x} using the Bayes rule for conditional probability:

$$P(\mathbf{x} | \mathbf{n}) = P(\mathbf{n} | \mathbf{x})P(\mathbf{x})/P(\mathbf{n}), \quad (26)$$

assuming independent Poisson spike statistics of different neurons. The final Bayesian estimate is:

$$P(\mathbf{x} | \mathbf{n}) = kP(\mathbf{x}) \left(\prod_{i=1}^N f_i(\mathbf{x})^{n_i} \right) \exp \left(-\tau \sum_{i=1}^N f_i(\mathbf{x}) \right), \quad (27)$$

where k is a normalization constant, $P(\mathbf{x})$ is the

prior probability for the animal to visit position \mathbf{x} , which can be estimated from the data, and $f_i(\mathbf{x})$ is the empirical tuning function, i.e. the average firing rate of neuron i for each position \mathbf{x} . Examples of the probability $P(\mathbf{x} | \mathbf{n})$ for the probable position of the rat is shown in Fig. 2. The most probable value of \mathbf{x} can thus be obtained by finding the \mathbf{x} that maximizes $P(\mathbf{x} | \mathbf{n})$, namely,

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{n}). \quad (28)$$

By sliding the time window forward, the entire time course of \mathbf{x} can be reconstructed from the time varying-activity of the neural population.

A comparison of different reconstruction methods for this problem shows that the Bayesian reconstruction method was the most accurate (Zhang et al., 1998). As the number of neurons included in the reconstruction is increased, the accuracy of all the methods increased. The best mean error using 25–30 simultaneously recorded cells was about 5 cm, in the range of the intrinsic error of the infrared position tracking system. Alternative formulas were derived from Bayes rule assuming a gaussian model of place field and updating the estimated position only when a spike occurs (Brown et al., 1998), although its accuracy seemed to be slightly lower than that of the Bayesian method above (Chan et al., 1999). There are thousands of place cells in the hippocampus of a rat that respond in any given environment. However, it is not known how this information is used by the rat in solving spatial and memory problems.

Synaptic learning for Bayesian decoding

Although these reconstruction methods may be useful for telling us what information could be available in a population of neurons, it does not tell us what information is actually used by the brain. In particular, it is not clear whether these reconstruction methods could be implemented with neurons. Pouget et al. (1998) show how maximum likelihood decoding can be performed using the highly recurrent architecture of cortical circuits, and thus demonstrate that the theoretical limit corresponding to the Fisher information is achievable. Zhang and Sejnowski (1999b) show how a feedforward network with one layer of weights could in principle read out a Bayesian code. Thus, optimal decoding is well within the capability of the network mechanisms known to exist in the cortex.

How a feedforward network can implement the Bayesian decoding method is shown in Fig. 3. The first layer contains N cells tuned arbitrarily to a variable of interest x , and the tuning function $f_i(x)$ of cell i describes its mean firing rate. The cells in the second layer represent the value of the encoded variable x (discretized if it is continuous) by their locations in the layer. Let function $\phi_i(x)$ be the synaptic connection weight from cell i in the first layer to cell x in the second layer. Given the numbers of spikes n_1, n_2, \dots, n_N from the N cells in the first

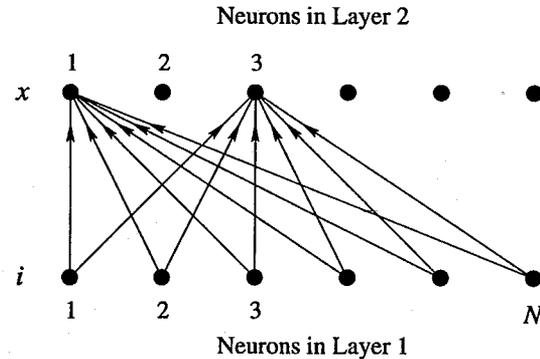


Fig. 3. The Bayesian decoding method can be implemented by a feedforward network whose synaptic strength can be learned by a Hebbian learning rule that is proportional to the presynaptic firing rate on a log scale, suggesting that the biological system might possibly be able to learn to use the theoretically optimal Bayes rule for reading out information contained in a neuronal population.

layer within a unit time interval, the distribution of activity in the second layer computes the sum

$$\sum_{i=1}^N n_i \phi_i(x). \quad (29)$$

To reconstruct the true value of x , the cell with the highest activity in the second layer should be chosen with a winner-take-all circuit. Different reconstruction methods correspond to different basis functions $\phi_i(x)$, as shown in more details in Zhang et al. (1998). A constant bias term $A(x)$ that is independent of the activity may be added to implement the Bayesian rule in Eq. 27. We only need the basis function:

$$\phi_i(x) \propto \log f_i(x), \quad (30)$$

because by taking logarithm of Eq. 27, the only term depending on the spikes n_i is the sum $\sum_{i=1}^N n_i \log f_i(x)$. Thus, the synaptic weight should be proportional to the logarithm of the tuning function.

Here we show that a simple Hebb rule is sufficient to establish the weights needed for the Bayesian method. We propose that the synaptic weight should change according to:

$$\Delta W \propto \text{Post} \times \log(\text{Pre}) \quad (31)$$

where for the presynaptic cell, Pre = firing rate, and for the postsynaptic cell, Post may be taken as

a binary variable. During training, the activation at the second layer is confined to a single unit corresponding to the true value of the encoded variable, which should vary, sampling all possible values uniformly. It can be shown that the final outcome of this learning is the synaptic weight patterns that are proportional to the average firing rate of the presynaptic cell in log units. That is, the requirement in Eq. 30 is satisfied after the training.

Thus, the Bayesian decoding method considered in the preceding section can be implemented by a feedforward network, and the synaptic weights can be learned with a synaptic weight plasticity proportional to presynaptic firing rate at logarithm scale. The bias or prior in the Bayesian method should be implemented as a tonic input independent of the activation of the cells in the first layer, so that when the context changes, the tonic input should switch accordingly. The learning rule shows how to pool a bank of existing feature detectors with Poisson spikes to quickly develop new detectors at the next stage that are optimal. In particular, this may be used to develop detectors for complex patterns, such as a 'grandmother cell'. Overall it does not seem to take much to implement Bayesian formula to achieve optimal performance in the special cases considered above, even though an explicit readout of a population code may not be needed until the final common pathway of the motor system since projections between cortical areas may simply perform transformations between different population codes.

References

- Abbott, L.F. and Dayan, P. (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput.*, 11: 91–101.
- Baldi, P. and Heiligenberg, W. (1988) How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biol. Cybernet.*, 59: 313–318.
- Brown, E.N., Frank, L.M., Tang, D., Quirk, M.C. and Wilson, M.A. (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.*, 18: 7411–7425.
- Brunel, N. and Nadal, J.-P. (1998) Mutual information, Fisher information, and population coding. *Neural Comput.*, 10: 1731–1757.
- Chan, K.-L., Zhang, K.-C., Knierim, J.J., McNaughton, B.L. and Sejnowski, T.J. (1999) Comparison of different methods for position reconstruction from hippocampal place cell recordings. *Soc. Neurosci. Abstr.*, 25: 2166.
- Cover, T.M. and Thomas, J.A. (1991) *Elements of Information Theory*. Wiley, New York.
- Eurich, C.W. and Wilke, S.D. (2000) Multi-dimensional encoding strategy of spiking neurons. *Neural Comput.*, 12: 1519–1529.
- Földiák, P. (1993) The 'ideal humuncūlus': Statistical inference from neural population responses. In: F. Eeckman and J. Bower (Eds.), *Computation and Neural Systems 1992*. Kluwer Academic, Norwell, MA.
- Frieden, B.R. (1998) *Physics from Fisher Information: A Unification*. Cambridge University Press, Cambridge.
- Georgopoulos, A.P., Schwartz, A.B. and Kettner, R.E. (1986) Neuronal population coding of movement direction. *Science*, 233: 1416–1419.
- Girosi, F. and Poggio, T. (1990) Networks and the best approximation property. *Biol. Cybernet.*, 63: 169–176.
- Hinton, G.E., McClelland, J.L. and Rumelhart, D.E. (1986) Distributed representations. In: D.E. Rumelhart and J.L. McClelland (Eds.), *Parallel Distributed Processing, Vol. 1*, MIT Press, Cambridge, MA, pp. 77–109.
- Lehky, S.R. and Sejnowski, T.J. (1990) Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *J. Neurosci.*, 10: 2281–2299.
- Muller, R.U., Kubie, J.L. and Ranck Jr., J.B. (1987) Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *J. Neurosci.*, 7: 1935–1950.
- Mussa-Ivaldi, F.A. (1988) Do neurons in the motor cortex encode movement direction? An alternative hypothesis. *Neurosci. Lett.*, 91: 106–111.
- Paradiso, M.A. (1988) A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol. Cybernet.*, 58: 35–49.
- Pouget, A. and Sejnowski, T.J. (1997) Spatial transformations in the parietal cortex using basis functions. *J. Cognit. Neurosci.*, 9: 222–237.
- Pouget, A., Zhang, K.-C., Deneve, S. and Latham, P.E. (1998) Statistically efficient estimation using population code. *Neural Comput.*, 10: 373–401.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. and Bialek, W. (1997) *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA.
- Sanger, T.D. (1994) Theoretical considerations for the analysis of population coding in motor cortex. *Neural Comput.*, 6: 12–21.
- Sanger, T.D. (1996) Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.*, 76: 2790–2793.
- Schwartz, A.B., Kettner, R.E. and Georgopoulos, A.P. (1988) Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement. *J. Neurosci.*, 8: 2913–2927.
- Seung, H.S. and Sompolinsky, H. (1993) Simple models for reading neuronal population codes. *Proc. Natl. Acad. Sci. USA*, 90: 10749–10753.

- Snippe, H.P. (1996) Parameter extraction from population codes: a critical assessment. *Neural Comput.*, 8: 511–529.
- Snippe, H.P. and Koenderink, J.J. (1992) Discrimination thresholds for channel-coded systems. *Biol. Cybernet.*, 66: 543–551.
- Wilson, M.A. and McNaughton, B.L. (1993) Dynamics of the hippocampal ensemble code for space. *Science*, 261: 1055–1058. (Corrections in Vol. 264, p. 16).
- Wu, S., Nakahara, H., Murata, N. and Amari, S. (2000) Population decoding based on an unfaithful model. In: S.A. Solla, T.K. Leen and K.R. Muller (Eds.), *Advances in Neural Information Processing Systems, Vol. 12*. MIT Press, Cambridge, MA, pp. 192–198.
- Yoon, H. and Sompolinsky, H. (1999) The effect of correlations on the Fisher information of population codes. In: M.S. Kearns, S.A. Solla, and D.A. Cohn (Eds.), *Advances in Neural Information Processing Systems, Vol. 11*, MIT Press, Cambridge, MA, pp. 167–173.
- Zhang, K.-C. and Sejnowski, T.J. (1999a) Neuronal tuning: to sharpen or broaden? *Neural Comput.*, 11: 75–84.
- Zhang, K.-C. and Sejnowski, T.J. (1999b) A theory of geometric constraints on neural activity for natural three-dimensional movement. *J. Neurosci.*, 19: 3122–3145.
- Zhang, K.-C., Ginzburg, I., McNaughton, B.L. and Sejnowski, T.J. (1998) Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.*, 79: 1017–1044.
- Zohary, E. (1992) Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biol. Cybernet.*, 66: 265–272.
- Zohary, E., Shadlen, M.N. and Newsome, W.T. (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370: 140–143.