Published in final edited form as:
Biosystems. 2007 ; 89(1-3): 16–23.

A simple measure of the coding efficiency of a neuronal population

Emilio Salinas and Nicholas M. Bentley

Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Winston-Salem, NC 27157-1010, USA

Abstract

We derive a simple measure for quantifying the average accuracy with which a neuronal population can represent a stimulus. This quantity, the basis set error, has three key properties: (1) it makes no assumptions about the form of the neuronal responses; (2) it depends only on their second order statistics, so although it is easy to compute, it does take noise correlations into account; and (3) its magnitude has an intuitive interpretation in terms of the accuracy with which information can be extracted from the population using a simple method — “simple” meaning linear. We use the basis set error to characterize the efficacy of several types of population codes generated synthetically in a computer. In general, the basis set error typically ranks different encoding schemes in a way that is qualitatively similar to Shannon's mutual information, except when nonlinear readout methods are necessary. Because this measure is concerned with signals that can be read out easily (i.e., through linear operations), it provides a lower bound on coding accuracy relative to the computational capabilities that are accessible to a neuronal population.

Keywords

Neural coding; Mutual information; Function approximation; Basis set; Neural network

1 Introduction

Neurophysiological experiments in which the activities of single neurons are recorded often result in a data base of responses that vary across a number of tested conditions. Analyses of these data typically evaluate the correlation between the evoked responses and the physical variables that change across conditions, such as stimulus location, sound frequency, or arm movement direction, for example. Thus, given a set of neuronal responses to a stimulus x , one often asks, how effectively do the neurons represent x ?

Common measures to quantify the statistical association between responses and stimuli include the (linear) correlation coefficient, the analysis of variance, or when there are only two conditions, ROC curves (Press et al., 1992; Dayan and Abbott, 2001). These techniques, however, are mostly applicable to single neurons, and may have other drawbacks. Two quantities overcome these limitations: the Fisher information, which bounds the accuracy with which any decoding scheme can extract an estimate of an encoded quantity, and Shannon's mutual information, which measures how much of a signal is theoretically possible to recover

Correspondence should be addressed to: Emilio Salinas, Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Winston-Salem, NC 27157-1010, USA, Phone: (336) 713-5176, Fax: (336) 713-4534, e-mail: esalinas@wfubmc.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

from a response that is related to it (Dayan and Abbott, 2001). Both measures are powerful because they are assumption-free; their disadvantage, however, is that they require knowledge of the full probability distribution relating the responses to the signal (or stimulus), and this requires very large numbers of trials. Furthermore, these measures represent theoretical limits, which in practice may be hard to achieve.

Here we develop an alternative method for quantifying the coding capacity of a population. The resulting quantity, which we call the basis set error, or E_{BS} , is closely related to the error that one would obtain by decoding the stimulus with an optimal linear estimator (Salinas and Abbott, 1994; Ben Hamed et al., 2003; Johnson et al., 2005). This is a natural measure of coding efficiency that is relatively easy to compute. We study its properties and illustrate potential applications.

2 Derivation of E_{BS}

The following approach is motivated by ideas in function approximation (Poggio, 1990; Poggio and Girosi, 1989; Pouget and Sejnowski, 1997; Salinas and Abbott, 2000). Consider a population of n neurons that are tested in N conditions. Each condition may represent, for instance, a different stimulus value. The n neurons under study are called the basis set, or basis neurons, and we are interested in estimating the accuracy with which they can generate an arbitrary downstream function of the condition. Thus, the basis neurons drive an additional downstream neuron.

The matrix $\mathbf{r}$ represents the collection of basis responses, such that r_{ij} is the firing rate (or some other response measure) of neuron i in condition j . These firing rates are variable, they are not identical across trials, so their mean values are denoted by $\langle r_{ij} \rangle$, where the brackets indicate an average over multiple repetitions of the same condition. The response profile or tuning curve of neuron i as a function of the condition is simply a plot of $\langle r_{ij} \rangle$ versus j . The firing rate of the downstream neuron is modeled as a weighted sum of the basis responses. Thus,

$R_k = \sum_j w_j r_{jk}$ is the downstream response in condition k , where w_j represents the strength of the connection from basis neuron j to the downstream unit. In matrix notation this is $\vec{R} = \vec{w} \mathbf{r}$ (note that vectors are considered rows). Now suppose that the desired downstream response is $\vec{F}$. That is, we know in advance how the downstream unit *should* respond in each condition. In order to make the driven response as close as possible to the desired one, we need to adjust the synaptic weights $\vec{w}$ appropriately. Thus, the first step in developing an error measure is to consider the average square difference between the driven and the desired downstream firing rates,

$$E = \sum_k c_k \left\langle \left(\sum_j w_j r_{jk} - F_k \right)^2 \right\rangle. \quad (1)$$

Here, the coefficient c_k represents the probability of condition k , so $\sum_k c_k = 1$, and the brackets indicate an average over trials, as before.

So far, this is a classic function approximation problem where a weighted sum of basis responses is used to approximate a desired output (Poggio, 1990; Poggio and Girosi, 1989; Pouget and Sejnowski, 1997; Salinas and Abbott, 2000). To find the synaptic weights that minimize the expression above, the standard procedure is to compute the partial derivatives of E with respect to w_i , equate the result to zero, and solve the resulting system of equations. This gives $w_i^{\text{opt}} = \sum_{j,k} c_k F_k \langle r_{jk} \rangle C_{ji}^{-1}$. These optimal connections depend on the inverse $\mathbf{C}^{-1}$ of the correlation matrix between basis responses, where $C_{ij} = \sum_k c_k \langle r_{ik} r_{jk} \rangle$. With the weights w_i^{opt} , the driven downstream response $\vec{R}$ approximates the target function $\vec{F}$ as

accurately as possible given the mean basis responses and their noise, which needs to be taken into account in the products $\langle r_{ik} r_{jk} \rangle$.

Although we have found synaptic weights that minimize the approximation error, we still want to know how big it is. This is simple: substitute the optimal weights of the preceding paragraph into Eq. (1) to obtain

$$E_{\min} = \sigma_F^2 - \sum_{i,j,k,q} c_k c_q \langle r_{ik} \rangle \langle r_{jq} \rangle F_k F_q C_{ij}^{-1} \quad (2)$$

where $\sigma_F^2 = \sum_k c_k F_k^2$ is the total variance of the desired downstream function. This expression gives the minimum approximation error, and depends critically on two things, the second order statistics of the basis responses and the choice of target function. Now, here is the key insight: to characterize the overall approximation capacity of the basis neurons, what we want is a quantity that depends only on their responses; hence, we need to eliminate the dependence on $\vec{F}$. This can be done by averaging over $\vec{F}$, but some knowledge about the distribution of function values F_j is needed. A simple and general condition is this: $\langle F_i F_j \rangle_F = \delta_{ij}$, where the brackets $\langle \cdot \rangle_F$ denote an average over F_j values, and $\delta_{ij} = 1$ if $i = j$ and is zero otherwise. This condition represents a worst-case scenario in which, to generate $\vec{F}$, each entry F_j is chosen randomly from a distribution with zero mean and unit variance. This is a worst-case assumption because the most difficult type of function to approximate is a random one. So, using this condition, calculate the expected value of Eq. (2), $\langle E_{\min} \rangle_F$. The result is the basis set error

$$E_{BS} = 1 - \sum_{i,j,k} c_k^2 \langle r_{ik} \rangle \langle r_{jk} \rangle C_{ij}^{-1}. \quad (3)$$

After averaging, the dependence on $\vec{F}$ has disappeared. The remaining quantity depends only on the neuronal responses and on the relative frequencies of the conditions. It reflects the efficacy with which the basis set approximates essentially arbitrary functions. Because the connection weights are optimal for each $\vec{F}$, E_{BS} represents the smallest possible error that, on average, can be achieved with the given set of basis responses and a linear readout. Note that E_{BS} varies between 0 and 1.

3 Comparing two conditions

With actual neurophysiological recordings, one is often interested in comparing only two experimental conditions ($N = 2$). To obtain some intuition about E_{BS} , we consider this simple case first, looking at two neurons at a time ($n = 2$). Fig. 1a shows a hypothetical example where the responses of two neurons in conditions 1 and 2 produce two distinct clouds of data points. To understand what E_{BS} represents, draw two vectors from the origin to the centers of the clouds. Now suppose that you want to reach an arbitrary point in the plane using a linear combination of those two vectors. The arbitrary point corresponds to the downstream function $\vec{F}$ discussed in the previous section, and E_{BS} represents the average accuracy with which a variety of such points (chosen randomly) can be reached. In Fig. 1a the error is low because the tips of the two vectors are clearly separate, so any point in the plane can be reached quite accurately by adding and subtracting multiples of the two vectors. In Fig. 1b the error is higher because, although the centers of the data clouds are the same as in panel a, their spread is much larger; the tips of the two vectors are thus much more variable (across trials). In Fig. 1c the two data clouds are again separate, yet the error is highest. This is because in this case the two vectors are parallel, so only points along one direction can be reached. E_{BS} is about one half because the two neurons span only one of the two relevant dimensions.

How does E_{BS} compare with other methods for comparing sets of responses in two conditions? A popular technique, which is based on optimal signal detection, is to calculate a receiver operating characteristic (ROC) curve (Green and Swets, 1966; Britten et al., 1992; Dayan and Abbott, 2001). The ROC curve is computed for a single neuron. It is useful because the area under it is equal to p_c , the probability with which an ideal observer can discriminate correctly which condition was presented in a given trial given the response of the neuron. When the responses of the neuron in conditions 1 and 2 have Gaussian distributions, p_c can be determined analytically. For instance, if the discrimination is based on the activity of neuron i , then optimal performance achieves a probability

$$p_c = \frac{1}{2} \operatorname{erfc} \left(-\frac{1}{2} \frac{|\langle r_{i2} \rangle - \langle r_{i1} \rangle|}{\sqrt{(\sigma_{i2}^2 + \sigma_{i1}^2)/2}} \right) \quad (4)$$

where erfc is the complementary error function. Thus defined, the minimum of p_c is 0.5, which represents chance performance. This expression depends on the difference between the mean responses relative to the variances, i.e. on the separation between the data clouds in Figs. 1a–c relative to their overlap. Therefore, we hypothesize that, when calculated for two conditions, E_{BS} may behave in a qualitatively similar way as p_c .

A comparison between these two quantities is shown in Fig. 2. Here, to put the ROC result in the same scale as E_{BS} , we plot $2 - 2p_c$, which varies from 0 to 1. Note, however, that p_c can be read directly from the plots by changing the y-axis limits from [0, 1] to [1, 0.5]. For these plots, all mean rates $\langle r_{i1} \rangle$ and $\langle r_{i2} \rangle$ were picked randomly from uniform distributions between 0–40 spikes/s, and two noise models were tested. In the first one (Figs. 2a, c), the SD of the response of neuron i in condition j was

$$\sigma_{ij} = \alpha \left(r_{\max} + 0.1 \sqrt{\langle r_{ij} \rangle} \right) \quad (5)$$

where $r_{\max} = 40$ spikes/s is the maximum of $\langle r_{ij} \rangle$. Thus, the noise had response-dependent and response-independent components, and in this case most of the variability was independent of the mean response. In contrast, in Figs. 2b, d the variability was dominated by the term that depends on the mean response:

$$\sigma_{ij} = \alpha \left(0.1 r_{\max} + \sqrt{\langle r_{ij} \rangle} \right). \quad (6)$$

As shown in Figs. 2a, b, when only two neurons are used to compute E_{BS} , the result follows very closely the discrimination performance of the ROC ideal observer. Here, note that a single neuron suffices to calculate p_c , whereas at least two neurons are needed to compute E_{BS} . This is because the function approximation procedure is limited to linear operations, whereas the ROC procedure is nonlinear (it is based on a threshold). However, in contrast to p_c , which is difficult to generalize to more than a single detector (neuron), E_{BS} can be easily computed with more neurons. Figs. 2c, d show the E_{BS} values that result from populations of 20 neurons; the error is, as expected, much lower. In summary, with two neurons, E_{BS} follows very closely the probability of correct discrimination of an ideal observer that decides which of two conditions occurred based on pairs of responses from a single neuron, but in contrast to the ideal observer performance, with E_{BS} the results from multiple neurons can be easily pooled.

4 Dependence of E_{BS} on noise and dimensionality

In some simplified situations it is possible to compute E_{BS} analytically. This is useful for understanding its meaning when different numbers of neurons and experimental conditions are involved. For this, assume that all conditions are equally probable, so $c_k = 1/N$. In addition,

suppose that the mean response variance is the same for all neurons, so $\sigma_i^2 = (1/N) \sum_k [\langle r_{ik}^2 \rangle - \langle r_{ik} \rangle^2] = \sigma^2$ for all i . In that case, the correlation matrix is

$$\mathbf{C} = \frac{1}{N} \langle \mathbf{r} \mathbf{r}^T \rangle = \frac{1}{N} (\mathbf{M} + N \sigma^2 \mathbf{I}_n) \quad (7)$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{M} = \langle \mathbf{r} \rangle \langle \mathbf{r} \rangle^T$ is the correlation between mean responses. Then, the basis set error is

$$E_{BS} = 1 - \frac{1}{N^2} \text{tr}(\mathbf{M} \mathbf{C}^{-1}) = 1 - \frac{1}{N} \sum_{i=1}^n \frac{\lambda_i}{\lambda_i + N \sigma^2} \quad (8)$$

where the λ_i are the eigenvalues of $\mathbf{M}$. The first equality is simply Eq. (3) rewritten using $c_k = 1/N$, the definition of $\mathbf{M}$, and the trace of a matrix, where $\text{tr}(\mathbf{A}) = \sum_i A_{ii}$. The second equality is obtained by using Eq. (7) and substituting $\mathbf{M}$ with its diagonalized form (its eigen-decomposition).

This expression reveals some interesting features of the basis set error. First, as the variability of the responses approaches infinity, E_{BS} tends to 1. Thus, higher noise always leads to a higher error. Second, at the other extreme, when there is no variability the error is equal to $1 - n/N$, assuming that the neurons have different response profiles (otherwise, some eigenvalues would be zero and the effective number of neurons would be less than n , as in Figs. 1c, f). Therefore, the error is zero if the number of noiseless neurons is equal to the number of conditions. This makes sense: if the number of neurons is smaller than the number of degrees of freedom in the approximated functions, even with maximum reliability the neurons can only cover, at most, a fraction n/N of the relevant dimensions. This is a consequence of the linearity of the mechanism by which the downstream responses are driven. Also, it might seem from Eq. (8) that E_{BS} could become negative if $n > N$. But this cannot happen because $\mathbf{M}$ can have at most N non-zero eigenvalues; its rank is the minimum of n and N .

In summary, then, E_{BS} has a combined dependence on the dimensionality of the problem and on the variability of the basis responses.

5 Mutual information versus E_{BS}

Because E_{BS} is concerned with signals that can be read out easily, i.e., through linear operations, it provides an indication of the information about the condition that is readily accessible from the basis responses. We now compare E_{BS} with a classic information-theoretic approach.

The mutual information (Cover and Thomas, 1991; Dayan and Abbott, 2001) between a signal s and a response r is

$$I = \sum_{r,s} P(s) P(r|s) \log_2 \left(\frac{P(r|s)}{P(r)} \right) \quad (9)$$

where $P(s)$ is the probability of observing the signal s , $P(r)$ is the probability of observing the response r , and $P(r|s)$ is the (conditional) probability of observing the response r given that the signal was s ; the sum is over all possible values of s and r . In our case, s is the condition and takes the values 1, 2, ..., N , whereas r contains the firing rates of all the neurons in a particular trial; it is thus a vector $\vec{r} = (r_1, r_2, \dots, r_n)$. To compute I reliably, we set the conditional probability density of the responses to be Gaussian (Abbott et al., 1996).

In the next example we consider the responses of n basis neurons to $N = 8$ stimuli, so each condition corresponds to a different stimulus value, and we calculate the mutual information that the basis responses provide about the stimulus. One final step is that, to facilitate the comparison with E_{BS} , we actually plot the residual information, $I_r = 1 - (I/I_{\max})$. This quantity is 0 when all the information can be recovered from the responses and is equal to 1 when no information is recovered, so it has the same range and direction as E_{BS} . We use 8 stimuli with equal probabilities, so the maximum information is $I_{\max} = \log_2(8) = 3$ bits.

Fig. 3 plots E_{BS} and I_r for populations of simulated neurons of different sizes, all tested with 8 stimuli. For each neuron i responding to stimulus k , the mean rate $\langle r_{ik} \rangle$ and SD σ_{ik} were picked randomly from uniform distributions. The rates were always between 0–40 spikes/s, and the neurons were assumed to be independent (i.e., their fluctuations were uncorrelated). In Fig. 3a, the SDs varied randomly between 0–2 spikes/s. The variability is so small that E_{BS} follows very closely the $1 - n/N$ relationship expected from Eq. (8) with zero noise. In contrast, I_r is already close to zero with one neuron. This is because the identity of the stimulus or condition can, in principle, be determined exactly with one noiseless neuron — but note, however, that nonlinear operations are required for this. With the SDs chosen randomly between 0–8 (Fig. 3b) or 0–16 spikes/s (Fig. 3c), both quantities decrease smoothly with the size of the neuronal pool. I_r does so more quickly than E_{BS} because it assumes that any conceivable function of the firing rates can be used to extract the information about the stimuli. In Fig. 3d the mean rates were the same as in Fig. 3c but the SDs were constant (8 spikes/s for all stimuli and neurons). Thus, the variability of the neurons carried no information about the stimuli in this case, although its average magnitude was the same as in Fig. 3c. This change caused a small increase in I_r (slightly lower coding efficiency) and a small decrease in E_{BS} (slightly higher coding efficiency), revealing a key difference between the two measures. The converse case, however, highlights this difference much more dramatically, as shown in Fig. 3e. Here, the mean firing rates were constant (20 spikes/s) and only the SDs varied across conditions and neurons (0–16 spikes/s). In this case, E_{BS} stays near its maximum value of 1, so the linear approximation is practically useless. In contrast, although I_r increases substantially, all the information can still be recovered with about 30 neurons. Intuitively, this is because by analyzing the spread in the firing rates one can figure out which stimulus generated a given set of responses, even if the mean rates do not change. For this, however, squaring operations or multiplications are absolutely necessary. A last comparison is shown in Fig. 3f, which plots the results when, for each neuron, means and SDs were independent only for the first four conditions; the other four were repeats of the latter. Here, I_r saturates at $\log_2(4)/\log_2(8) = 0.66$; E_{BS} saturates at 0.5 because, as in Figs. 1c, f, only half of the dimensions are covered by the basis neurons.

Results similar to those in Fig. 3 were obtained in other situations; for instance, when the SDs were proportional to the square root of the mean responses, or when the mean rates were drawn from exponential distributions.

6 Quantifying the sensitivity to two quantities

As a final application of E_{BS} , we discuss the case of neurons that are sensitive to more than one type of stimulus. A classic example is that of parietal cortex, where neurons respond to visual stimuli and are modulated by gaze angle (Andersen et al., 1989; Brothie et al., 1995). In accordance with the data, the mean response of parietal neuron i is modeled as

$$\langle r_i(x, y) \rangle = f(x - a_i) \left((1 - \delta) + \delta g(y - b_i) \right) \quad (10)$$

where x is the location of a visual stimulus and y is the gaze angle. We used 10 values of x and 8 values of y for a total of 80 conditions. Figs. 4a, b show examples of tuning curves obtained with this expression. Their shapes are given by f , which is a Gaussian curve centered at the

preferred location of the cell, a_i . The amplitude of this Gaussian is modulated by eye position, and the strength of this effect depends on δ , the modulation depth. In Fig. 4a the modulation is weak ($\delta = 0.2$), so the three curves shown for three different gaze angles have similar amplitudes; in Fig. 4b the modulation is strong ($\delta = 0.8$). The dependence on y is typically monotonic, so we use increasing and decreasing sigmoidal curves for g , with different middle points b_i across cells.

A key question that is often raised with these kinds of neurons (Pouget and Sejnowski, 1997; Ben Hamed et al., 2001) is, how accurately do they encode x and y ? Furthermore, how does the accuracy of the representation change as a function of modulation depth? As shown in Figs. 4c, d, E_{BS} provides a simple way to evaluate this quantitatively. There are two ways to do the calculations. In Fig. 4c we pretend that one of the variables is unknown. So, for instance, $E_{BS}(x)$ gives the accuracy with which functions of x can be constructed when y is completely ignored; that is, rate variations due to y are not controlled for, so they enter as an additional source of noise. Therefore, the curve for $E_{BS}(x)$ increases because an increase in modulation depth is equivalent to an increase in noise. $E_{BS}(y)$ is calculated analogously. Its curve decreases only very slightly with δ because x always causes large variations in the responses. The second way to perform the calculations is by taking into consideration the value of the second variable. In this case, E_{BS} is computed with respect to x for a fixed value of y , and the results are averaged over all values of y . This produces $E_{BS}(x|y)$, which is the error in the encoding of x given that y is known. As shown in Fig. 4d, the errors are much smaller in this case, as one would expect. Now, as the modulation depth increases, the encoding accuracy for eye position y improves quite notoriously, although the error associated with x is always much smaller. Therefore, the type of encoding described by Eq. 10 strongly favors the representation of x over that of y . E_{BS} gives this answer with a minimum of assumptions and computational resources.

7 Discussion

E_{BS} is a relatively simple quantity based on the assumption that the responses of a neuronal population are read-out or processed using linear transformations. In essence, then, E_{BS} reflects the fraction of the signal that cannot be extracted through linear combinations of the basis responses. It ranks a variety of encoding schemes in a way that is qualitatively consistent with Shannon's mutual information, except when nonlinear readout methods are essential. E_{BS} is also consistent with the performance of an ideal observer in two-alternative forced-choice discrimination tasks.

This measure of coding efficiency is convenient because it has an intuitive interpretation and is easy to compute. We showed that it can be used to characterize how coding efficiency changes with the number of neurons and to investigate which variable is encoded more accurately by a neuronal population that is sensitive to several variables. E_{BS} could also be useful for comparing the coding accuracy of neurons from different brain areas, and to assess the effects of measured noise correlations (Zohary et al., 1994; Romo et al., 2003).

Perhaps the most interesting feature of E_{BS} is that, in a way, it relies on the opposite assumptions as Shannon's mutual information. Mutual information represents a limit in which as much knowledge as possible about a signal is extracted from a response, and any means to extract this knowledge are valid. In contrast, E_{BS} represents the knowledge that can be acquired only by the simplest means—linear operations. Thus, mutual information and E_{BS} represent upper and lower bounds on coding efficiency in terms of the computational capabilities accessible to a system. In addition, E_{BS} relies only on second-order statistics. This is of practical importance, because often these are the most that can be extracted reliably from actual neurophysiological recordings, which may limit the applicability of information-based measures.

Finally, whereas information and other quantities describe the coding efficiency of an ideal system, E_{BS} describes the efficiency of one that is constrained. Since constraints likely limit information processing mechanisms in the real brain, it is crucial that we understand how to accommodate them, and this framework may provide clues as to how to construct more adequate measures of coding efficiency in the future.

Acknowledgements

Research was supported by NINDS grant NS044894.

References

- Abbott LF, Rolls ET, Tovee MJ. Representational capacity of face coding in monkeys. *Cereb Cortex* 1996;6:1047–3211.
- Andersen RA, Essick GK, Siegel RM. Encoding of spatial location by posterior parietal neurons. *Science* 1985;230:450–458. [PubMed: 17816078]
- Ben Hamed S, Page W, Duffy C, Pouget A. MSTd neuronal basis functions for the population encoding of heading direction. *J Neurophysiol* 2003;90:549–558. [PubMed: 12750416]
- Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 1992;12:4745–4765. [PubMed: 1464765]
- Brothie PR, Andersen RA, Snyder LH, Goodman SJ. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* 1995;375:232–235. [PubMed: 7746323]
- Cover, TM.; Thomas, JA. *Elements of Information Theory*. Wiley; New York: 1991.
- Dayan, P.; Abbott, LF. *Theoretical Neuroscience*. MIT Press; Cambridge, MA: 2001.
- Green, DM.; Swets, JA. *Signal Detection Theory and Psychophysics*. Wiley; New York: 1966.
- Johnson A, Seeland K, Redish AD. Reconstruction of the postsubiculum head direction signal from neural ensembles. *Hippocampus* 2005;15:86–96. [PubMed: 15390162]
- Poggio T. A theory of how the brain might work. *Cold Spring Harbor Symp Quant Biol* 1990;5:899–910. [PubMed: 2132866]
- Poggio, T.; Girosi, F. *AI Memo 1140*. MIT; 1989. *A Theory of Networks for Approximation and Learning*.
- Pouget A, Sejnowski TJ. Spatial transformations in the parietal cortex using basis functions. *J Cog Neurosci* 1997;9:222–237.
- Press, WH.; Flannery, BP.; Teukolsky, SA.; Vetterling, WT. *Numerical Recipes in C*. New York: Cambridge University Press; 1992.
- Romo R, Hernández A, Zainos A, Salinas E. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 2003;38:649–657. [PubMed: 12765615]
- Salinas E, Abbott LF. Vector reconstruction from firing rates. *J Comput Neurosci* 1994;1:89–107. [PubMed: 8792227]
- Salinas E, Abbott LF. Do simple cells in primary visual cortex form a tight frame? *Neural Comput* 2000;12:313–335. [PubMed: 10636945]
- Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 1994;370:140–143. [PubMed: 8022482]

Fig 1.

E_{BS} for populations of two neurons responding in two conditions. Each panel shows the responses of a pair of neurons in two conditions. In (a–c) the response of neuron 2 is plotted vs. the response of neuron 1 for three different pairs. Dots and crosses indicate trials for conditions 1 and 2, respectively. Bar graphs below (d–f) show the mean firing rates ± 1 SD for the same three pairs. In each pair, neuron 1 corresponds to white bars and neuron 2 to gray bars.

Fig 2.

Evaluating coding efficacy for only two conditions ($N = 2$). Open symbols show E_{BS} computed using $n = 2$ neurons (a, b) or $n = 20$ neurons (c, d). Filled symbols correspond to $2 - 2p_c$. All p_c values were computed from single neuron data using Eq. (4) and averaging over neurons. All mean responses were chosen randomly between 0 and 40 spikes/s. In (a) and (c), SDs were given by Eq. (5); in (b) and (d) they were given by Eq. (6). Note the different x-axes.

Fig 3.

Basis set error and residual information as functions of the number of neurons, n , in populations with random tuning curves. In all panels, $N = 8$ conditions were used. Except in (e), $\langle r_{ij} \rangle \in [0, 40]$, meaning that mean rate values were drawn randomly between 0 and 40. (a) $\sigma_{ij} \in [0, 2]$. (b) $\sigma_{ij} \in [0, 8]$. (c) $\sigma_{ij} \in [0, 16]$. (d) $\sigma_{ij} = 8$ for all i, j . (e) $\langle r_{ij} \rangle = 20$ and $\sigma_{ij} \in [0, 16]$. (f) As in (c), except that, for all neurons, mean and SDs for conditions $j = 5, 6, 7, 8$ were equal to those for $j = 1, 2, 3, 4$, respectively. Units of $\langle r_{ij} \rangle$ and σ_{ij} are spikes/s.

Fig 4.

Basis set error for two simultaneously-coded quantities, x and y . (a) Responses of one model neuron as a function of x . The three curves correspond to three values of y (-10 , 0 and 10) for a modulation depth $\delta = 0.2$. (b) As in (a), but for $\delta = 0.8$. (c) E_{BS} when the same set of population responses are considered functions of x alone (filled circles) or y alone (open circles), as a function of modulation depth. Here, variations due to the variable that is not being evaluated are simply considered noise. Each data point was generated based on 200 neurons that followed Eq. 10 with f a Gaussian function and g a sigmoidal function. Preferred locations a_i and b_i were uniformly distributed. All responses were evaluated at 10 points in the x dimension and 8 points in the y dimension, and noise followed Eq. 5 with $\alpha = 0.1$. (d) Results using the same data as in (c), but taking into account the value of the second variable. To compute $E_{BS}(x|y)$, $E_{BS}(x)$ was calculated separately for constant values of y and the results were averaged.