Robust Nonlinear Neural Codes

Qianli Yang[1] and Xaq Pitkow[1,2]

[1]Rice University, Department of Electrical and Computer Engineering
[2]Baylor College of Medicine, Department of Neuroscience

Abstract

Most natural task-relevant variables are encoded in the early sensory cortex in a form that can only be decoded nonlinearly. Yet despite being a core function of the brain, nonlinear population codes are rarely studied and poorly understood. Interestingly, the most relevant existing quantitative model of nonlinear codes (Shamir et al., 2004) is inconsistent with known architectural features of the brain. In particular, for natural population sizes, such a code would contain more information than its sensory inputs, in violation of the data processing inequality. This is because the noise correlation structures assumed by this model provides the population with an information content that scales with the size of the cortical population, and this correlation structure could not arise in cortical populations that are much larger than their sensory input populations. Here we provide a valid theory of nonlinear population codes that obeys the data processing inequality by generalizing recent work on information-limiting correlations (Moreno et al., 2014) in linear population codes. Although these generalized, nonlinear information-limiting correlations bound the performance of any decoder, they also make decoding more robust to suboptimal computation, allowing many suboptimal decoders to achieve nearly the same efficiency as an optimal decoder. Although these correlations are extremely difficult to measure directly, particularly for nonlinear codes, we provide a simple, practical test by which one can use choice-related activity in small populations of neurons to determine whether decoding is limited by correlated noise or by downstream suboptimality. Finally, we discuss simple sensory tasks likely to require approximately quadratic decoding, to which our theory applies.

1 Supporting materials

Quadratic coding model. We consider a population of N neurons that code for a scalar stimulus s. The response of the ith neuron is given by $r_i = f_i(s) + \eta_i(s)$, where $f_i(s)$ is the tuning curve of neuron i and η_i reflects the trial-to-trial variability in the neural responses. The variability is assumed to follow a multivariate normal distribution with zero mean $\langle \boldsymbol{\eta} \rangle = 0$ and stimulus-dependent noise covariance, $\Sigma(s) = \left\langle \boldsymbol{\eta}\boldsymbol{\eta}^\top \right\rangle$.

In order to isolate the coding properties of a purely nonlinear neural code, we assume that the tuning curve is constant, while the noise covariances $\Sigma_{ij}(s)$ depend smoothly on the stimulus. We then quantify the information content of the neural population using the Fisher information J.

According to this model, the distribution of neural responses is described by the exponential family with quadratic sufficient statistics, $\boldsymbol{T}(\boldsymbol{r}) = (\ldots, r_i r_j, \ldots)$, and $p(\boldsymbol{r}|s) \sim \exp\left[\boldsymbol{H}(s)^\top \boldsymbol{T}(\boldsymbol{r})\right]$ where $\boldsymbol{H}(s)$ is a vector of parameters with the same dimension as $\boldsymbol{T}$. For response distributions drawn from this family, the stimulus can be estimated efficiently using a quadratic decoder.

Information-limiting noise. When a network inherits information from a smaller input population, noise in the input is processed by the same pathway as the signal, and this generates correlations that cannot be averaged away. Thus these noise correlations will bound the efficiency of decoding to match the information in the input layer (Moreno et al., 2014). This information-limiting noise can ultimately be referred back to the stimulus, and appears in the quadratic code as $\boldsymbol{r} \sim N(\boldsymbol{f}, \Sigma(s+ds))$ where $ds \sim N(0, \alpha)$. To a first approximation, this noise changes the covariance of the quadratic sufficient statistics $\boldsymbol{T}$ according to

$$\Gamma = \left\langle \delta\boldsymbol{T}\, \delta\boldsymbol{T}^\top \right\rangle = \Gamma_0 + \epsilon \left\langle \boldsymbol{T} \right\rangle' \left\langle \boldsymbol{T} \right\rangle'^\top \tag{1}$$

where Γ_0 is any other covariance matrix which does not limit quadratic information in the limit of large populations.

Information bound. When these information-limiting correlations are present, the Fisher information in the output will be bounded by $J = \frac{1}{\alpha + 1/J_0}$, where α is the variance of information-limiting correlation, andJ_0 is the Fisher information of the output in the absence of information-limiting correlations. When the population size grows, the unlimited information term J_0 grows rapidly, so the output Fisher information will be dominated by α. This is true for both optimal decoding and many forms of suboptimal decoding. We also show that the relative efficiency of different suboptimal quadratic decoders is higher in the presence of information-limiting noise.

Nonlinear choice correlations. Furthermore, we show that for a purely quadratic code, neurons have no linear correlation with behavioral choice. However, by computing nonlinear 'choice correlations' for a pair of neurons, defined as $C_{ij} = \text{Corr}(\hat{s}, r_i r_j)$, one can generalize past results on linear codes (Haefner et al., 2013) to identify the nonlinear strategies the brain uses to decode the population of neurons. When there are information-limiting correlations, these nonlinear choice correlations tend to their optimal value,

$$C_{ij}^{\text{opt}} = \sqrt{J_{ij}/J} \tag{2}$$

where J_{ij} is the Fisher information for the pair of neurons. This formula constitutes a simple test for optimal decoding that requires only pairwise recordings from a behaving animal.

References

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nature neuroscience*, *17*, 1410–1417.

Shamir, M., & Sompolinsky, H. (2004). Nonlinear population codes. *Neural Computation*, *16(6)*, 1105–1136.

Haefner RM, Gerwinn S, Macke JH, Bethge M (2013). Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nat. Neurosci. 16*: 235–242.