Essential nonlinear properties in neural decoding

Qianli Yang[12], Xaq Pitkow[12]

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

Summary. To decode task-relevant information from sensory observations, the brain must eliminate nuisance variables that affect those observations. For natural tasks, this generally requires nonlinear computation. Here we contribute new concepts and methods to characterize behaviorally relevant nonlinear computation downstream of recorded neurons. Linear decoding weights can be inferred from correlations between neurons and behavior [1, 2]. However, these weights do not adequately describe the neural code when, due to nuisance variation, mean neural responses are poorly tuned to the task while higher-order statistics of neural responses are well tuned. The task-relevant stimulus information can then be extracted only by nonlinear operations. For example, detecting an object boundary in an image requires contrast invariance: an edge appears when the foreground object is darker *or* lighter than the background, yet any linear function will exhibit opposite responses in these two cases. We generalize past weight-inference methods to determine the brain's nonlinear neural computations from joint higher-order statistics of neural activity and behavioral choices in perceptual tasks. This method is based on a new statistical measure we call nonlinear choice correlation, defined as the correlation coefficient between behavioral choices and nonlinear functions of measured neural responses. Importantly, the exact neural transformations may not be uniquely identifiable, since many neural nonlinearities can generate the same behavioral output. This is expected when sensory signals are expanded into a larger cortical response space, creating a redundant code. We exploit this redundancy to define a new concept of equivalence classes for neural transformations. We then demonstrate how to quantify essential properties of these equivalence classes, and provide simulations that show how these properties can be extracted using neural data from behaving animals. Finally, we explain the functional importance of these nonlinearities in specific perceptual tasks.

Supporting information: We assume a feedforward model $s \to \boldsymbol{r} \to \boldsymbol{R}(\boldsymbol{r}) \to \hat{s}$ in which the stimulus s drives neural responses $\boldsymbol{r}$ (here, spike counts in a given time bin) which are transformed nonlinearly into $\boldsymbol{R}(\boldsymbol{r})$ and then decoded linearly to give perceptual estimates $\hat{s} = \boldsymbol{w} \cdot \boldsymbol{R}(\boldsymbol{r}) + c$.

In a nonlinear code, information about a stimulus s is not encoded in the mean neural responses $\boldsymbol{r}$ but instead by higher-order statistics. To quantify the information content of a nonlinear population code, we use the Fisher information. For responses $\boldsymbol{r}$ drawn from the exponential family with nonlinear sufficient statistics $\boldsymbol{T}(\boldsymbol{r})$, so that $p(\boldsymbol{r}|s) \propto e^{\boldsymbol{H}(s)\cdot\boldsymbol{T}(\boldsymbol{r})-\eta(\boldsymbol{r})}$, the Fisher information is given by $J = \boldsymbol{F}'^{\top}\Gamma^{-1}\boldsymbol{F}'$, where $\boldsymbol{F}(s) = \langle \boldsymbol{T}|s\rangle$ and $\Gamma(s) = \mathrm{Cov}(\boldsymbol{T}|s)$ are the mean and covariance of the relevant nonlinear sufficient statistics [3]. These quantities are analogous to a tuning curve (signal) and correlations (noise) in the linear case, and determine the structure of optimal decoding. For example, in a purely quadratic code with zero mean response, so that $\boldsymbol{T}(\boldsymbol{r}) = \mathrm{vec}(\boldsymbol{r}\boldsymbol{r}^{\top})$, the signal is the stimulus-dependent response covariance $\boldsymbol{F}(s) = \mathrm{vec}\left\langle \boldsymbol{r}\boldsymbol{r}^{\top}|s\right\rangle$ (which would usually be classified as noise correlations), and the relevant noise is a fourth-order quantity, $\Gamma_{ijkl} = \langle r_i r_j r_k r_l|s\rangle - F_{ij}F_{kl}$.

For such a nonlinear code, we define the **nonlinear choice correlation**,

$$C_{R_k} = \mathrm{Corr}(R_k(\boldsymbol{r}), \hat{s}) \tag{1}$$

where $R_k(\boldsymbol{r})$ is any nonlinear function of the neural responses. When the decoding nonlinearity and weights are both optimal, we prove that the nonlinear choice correlation equals the ratio of information in specific nonlinearity over information in the entire neural population, $C^{\mathrm{opt}}_{R_k} = \sqrt{J_{R_k}/J}$.

The actual neural nonlinearities the brain used is complex and unknown. A crucial question is then: how can we describe that nonlinearity? Just as linear decoding models collapse all neural computation into a single set of decoding weights, we summarize the cumulative effects of the actual nonlinear neural computation by a set of nonlinear basis functions. For concreteness, here we invoke a multivariate Taylor series, $\boldsymbol{R} = \{r_i, r_i r_j, r_i r_j r_k, \ldots\}$, as a set of 'proxy nonlinearities' (Figure 1A). We then specify a two-step decoding scheme, $s \to \boldsymbol{r} \to \boldsymbol{R} \to \hat{\mathbf{s}} \to \hat{s}$, where $\hat{\mathbf{s}} = W\boldsymbol{R} + \boldsymbol{c}$ is a vector of stimulus estimates, each derived from only one class of nonlinearities $\hat{\mathrm{s}}_i = \boldsymbol{w}_i^\top \boldsymbol{R}_i + c_i$. The final estimate $\hat{s}$ is a linear combination of estimates from all nonlinearities $\hat{s} = \boldsymbol{a}^\top \hat{\mathbf{s}} + c$. Note that, in general, this two-stage decoding scheme does not capture all possible decoders, but we prove that it is accurate when neural signals are redundant.

Many fine-scale decoder weights $\boldsymbol{w}_i$ are equivalent because many neurons are interchangeable in large redundant populations. In contrast, after the dimensionality reduction from different nonlinear classes $\boldsymbol{R}$ to the estimates $\hat{\mathbf{s}}$, this redundancy has been squeezed out. The specific scaling $\boldsymbol{a}$ of each estimate is then meaningful, and describes the importance of different nonlinear types to the brain's decoding. We can determine these scalings $\boldsymbol{a}$ from nonlinear choice correlations between behavior and R_{ik} (the kth element of nonlinearity $\boldsymbol{R}_i$) using the equation

$$C_{ik} = \mathrm{Corr}(\hat{s}, R_{ik}) \approx \frac{(\boldsymbol{E}\boldsymbol{a})_k}{\boldsymbol{a}^\top \boldsymbol{E}\boldsymbol{a}} \sqrt{\frac{J_{ik}}{J}} \tag{2}$$

where $\boldsymbol{E} = \mathrm{Cov}(\hat{\mathbf{s}}|s)$ is the measured covariance of the estimates from the proxy nonlinearities. This equation predicts that, for redundant populations, the nonlinear choice correlations should be proportional to the information content of each statistic R_{ik}, with proportionality that depends on the scaling $\boldsymbol{a}$ (Figure 1B). Simulations show that the scalings $\boldsymbol{a}$ can successfully be inferred from the resultant slopes of the nonlinear choice correlations (Figure 1C). This demonstrates the conceptual and practical advantage of modeling computation after dimensionality reduction based on task variables, rather than trying to model individual neural nonlinearities.

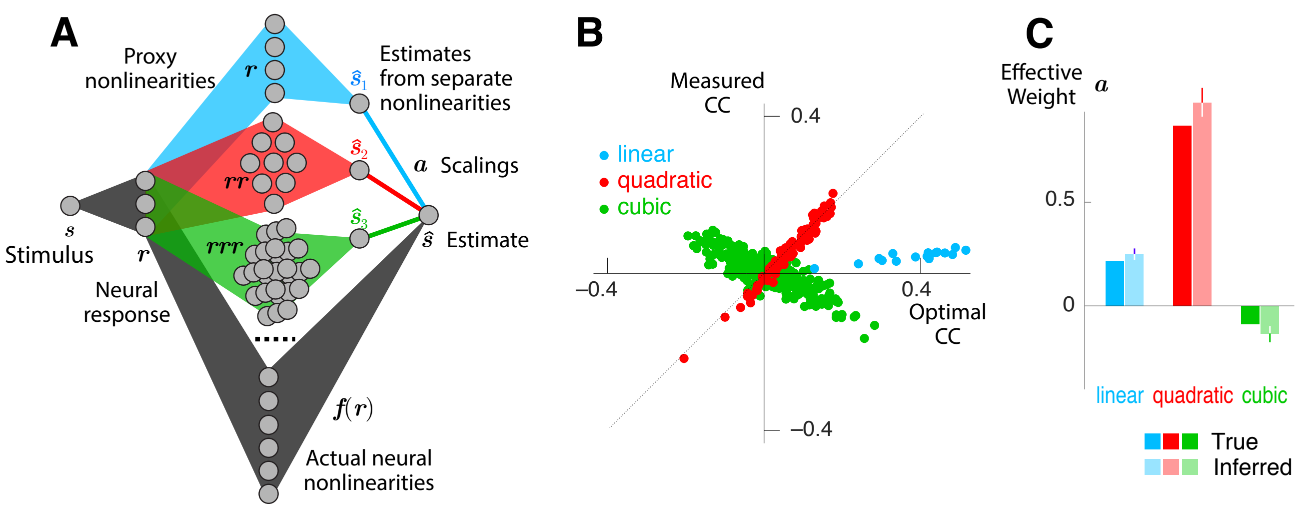

Figure 1: **A**: In a redundant code, the observed nonlinear neural transformations (grey) can be enacted in other equivalent ways (color). Since many fine details do not matter, it is valuable to characterize these equivalent nonlinearities by the relative scaling $\boldsymbol{a}$ of different classes of nonlinearity (e.g. polynomials). **B**: Simulations for this scheme show that, as predicted by Eq. 2, different nonlinearities produce different slopes when plotting measured versus optimal nonlinear choice correlations. **C**: Inferred scaling of proxy nonlinearities from nonlinear choice correlation compared to the true weights. The results prove that from the nonlinear choice correlations we can infer the importance of different classes of nonlinearities.

References

[1] Ralf M Haefner, Sebastian Gerwinn, Jakob H Macke, and Matthias Bethge. Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nature Neuroscience*, 16(2):235–242, 2013.

[2] Kaushik Lakshminarasimhan, Alexandre Pouget, and Xaq Pitkow. How brain areas can predict behavior, yet have no influence. *Cosyne abstract*, 2014.

[3] Jeffrey Beck, Vikranth R Bejjanki, and Alexandre Pouget. Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. *Neural computation*, 23(6):1484–1502, 2011.