A theory of structured noise correlations in peripheral and higher order brain areas and their significance

Jesse A. Livezey, Maximillian Dougherty, Sylvia Madhow, Pratik Sachdeva, Kristofer E. Bouchard

Summary. Neural computations take place in the presence of noise. Across repeated presentations of a stimulus, neural measurements exhibit correlated variability (noise correlations). Measuring and understanding noise correlations are important for determining fundamental limits on the fidelity of neural representations. We address two outstanding issues in the field. The first is whether predictions about the structure of noise correlations and discriminability should differ in peripheral versus higher order sensory (or motor) areas in the brain. The second is how to quantify the significance of hypotheses about the optimality of observed correlations for different measures of discriminability. Our first contribution is a prediction, based on theory, that noise correlations measured in areas of the brain that represent a set of stimuli as categorical, e.g., phonemes in STG (Chang, 2010), will have a different structure compared to those measured in areas of the brain that represent a set of stimuli as continuous, e.g., rotations of moving bars in retina (Franke, 2016, Zylberberg, 2016). Our second contribution is a new null model for testing whether observed noise correlations are optimal. Specifically, this null model tests whether there is optimal alignment between the principal axes of the observed noise correlations and the mean stimulus-responses for a given discriminability measure. By contrast, previously proposed null models only test how often a model with no inter-neuron correlations but equal per-neuron statistics generates the observed discriminability. When compared across diverse real and synthetic datasets, we observe a profound difference in statistical significance, indicating that current null models cannot be used to test the optimality of observed correlations. Together, these results provide an experimentally testable prediction that the nature of neural computation (continuous versus categorical) should determine the structure of noise correlations across the neural hierarchy and improved methods for testing hypotheses about the optimality of observed noise correlations.

Significance. Noise correlations are known to put fundamental bounds on the discriminability of neural representations (Moreno-Bote, 2014). However, whether the structure of the stimuli or behavior and the nature of neural computation (continuous versus categorical) impacts these predictions has not received sufficient investigation in the literature. We identify an alternative discriminability measure to the Linear Fisher Information (LFI) that is appropriate for categorical neural discrimination and show that it leads to different predictions for the optimal structure of noise correlations. Starting from early sensory areas such as the retina, V1, or inferior colliculus and moving to higher order areas such as STG or FFA, we expect that for categorical stimuli, the representations will change from continuous to categorical. The relationship between the nature of the computation, e.g., continuous vs. categorical, and the predicted structure of noise correlations is important for understanding brain areas away from the periphery. These predictions provide a rich set of new experimental and theoretical questions about the structure of noise correlations away from the periphery where representations can become categorical.

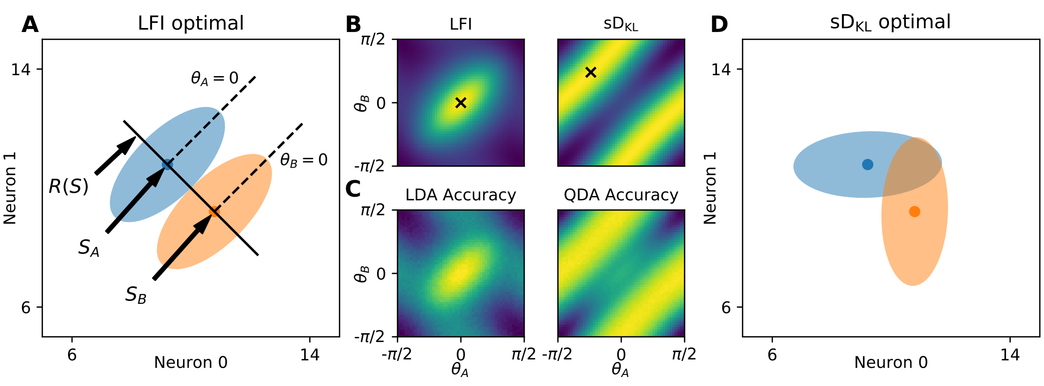

Figure 1: **(A)** Model setup and optimal LFI orientations. **(B, C)** Discriminability measures for the 4 measures as a function of the rotation angle of the two covariance matrices in **A**. Yellow is more discriminable. "x" indicates the optimal orientations shown in **A** and **D**. **(D)** Optimal orientation for the sD_{KL} measure.

Typically, to test whether observed correlations are optimal, they are compared to a null model which has no inter-neuron correlations but equal per-neuron statistics, either by trial shuffling the individual neural responses or by zeroing-out the off diagonal elements of the covariance matrix. This null model is appropriate for testing the hypothesis that the observed correlations have higher discriminability as compared to an uncorrelated distribution. However, it does not test the hypothesis that the observed noise correlations are optimal for a measure of discriminability. In order to test this stronger hypothesis about whether observed noise correlations are more generally optimal given a discriminability measure, we also provide a new null model for understanding the relationship between noise correlations, stimulus structure, and discriminability measure. To generate samples from this null, the alignment of the principal axes of the covariance matrix are rotated randomly with respect to the mean stimulus-responses. This model keeps the eigenspectra of the noise correlations fixed, but randomizes the relationship with the stimulus-response. This allows a more accurate assessment of the significance of the optimality of observed noise correlations. In previously published and new data, with this new null model, we show a reversal in statistical significance compared to the standard test in the field. This implies that, for these data, existing null models are not appropriate for testing hypotheses about the optimal alignment of noise correlations.

Results. The Linear Fisher Information quantifies how well a set of noisy neural measurements can be used to recover a stimulus parameter. In its derivation, an assumption is made that stimulus parameter is continuous. However, it has been shown that in higher-level perceptual areas, e.g., STG, perceptions and neural responses to stimuli can be categorical (e.g., phonemes, Chang, 2010). If we assume that the area of the brain being measured is responding to stimuli that are categorical and the per-stimulus responses are each normally distributed with different covariances per stimulus (see Fig. 1A for a visualization of this setup), we can use the symmetric KL Divergence ($\mathrm{sD_{KL}}$) as a measure of discriminability. We note that in the case where the covariances are constrained to be equal, the $\mathrm{sD_{KL}}$ becomes proportional to the LFI. Using $\mathrm{sD_{KL}}$ predicts that the optimal noise correlations should be asymmetrically rotated away from perpendicular (Fig. 1B right, Fig. 1D, yellow is more discriminable) by an amount that depends on the spacing of the means and the spread of the variances rather than aligned perpendicularly to the difference in means as in the LFI (Fig.1A, Fig. 1B left). This prediction is also true in classifiers that allow the two responses to have differing covariances, i.e., Quadratic Discriminant Analysis (QDA) versus Linear Discriminant Analysis (LDA) (Fig. 1C). This prediction can be tested by comparing the structure of noise correlation in peripheral versus higher order sensory areas for stimuli that are continuous and discrete.

When comparing discriminability measures on data, it is important to be able to directly test the hypothesis that the observed alignment between the noise correlations and the mean stimulus-response is statistically significant. Our proposed null model addresses this need by randomly rotating this alignment. In practice, samples from this null model can be computed by taking random rotations of the neural responses centered at their fixed mean. The null distribution is then the discriminability measure taken over these samples. Here, we compare the typically used (trial shuffle) null model and rotation null model on 2 datasets: multi-unit activity (MUA) recorded in rat A1 using micro-ECoG in response to tone-pips with varying frequency (Fig. 2A) and single unit spiking data recorded in monkey (*macaca fascicularis*) V1 in response to drifting gratings with varying angles (Fig. 2B, data from Kohn, 2016, CRCNS.org). For these early sensory datasets, we use the LFI as the measure of discriminability. For all datasets shown here (Fig. 2), there are many points that are significant under the shuffle model but not under the rotation model (points in the purple areas), few where the opposite is true (points in the brown areas), and a modest number that are significant for both models (points in the blue areas). These distributions match the distributions seen in a synthetic dataset drawn from distributions like Fig. 1A with varying mean-distances and rotations. For each point, 1e4 samples from each null model were drawn which determines the smallest possible p-values (1e-4 here). This shows that as more refined hypotheses about the structure of noise correlations are tested, the null models used will also need to be updated.

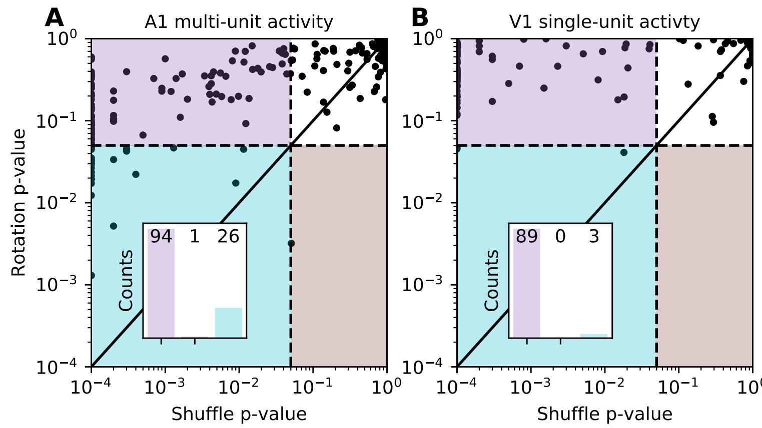

Figure 2: (A, B) The p-values for the observed LFI under the shuffle and rotation null models for the A1 multi-unit activity and V1 single-unit activity, respectively. Purple region: false positives, Brown region: false negatives, Blue region: true positives. Dashed lines indicate $p = 0.5$.