Tutorial - Correlations in Neural Populations

Understanding how noise correlations (correlations of the activity of different neurons across trials with the same sensory stimulus) shape the encoding of sensory information has been a major question in systems neuroscience, because it reveals the emergent properties of population codes. In this tutorial, we delve into the application of the NIT Toolbox to investigate how these correlations, ranging from pairs to tens of neurons, shape the information encoded within neural populations.

1. Generating Simulated Data

First, we generate simulated data for three scenarios to illustrate three primary ways in which correlations can influence population coding, using the NIT toolbox. The simulated data comprises correlated spike trains for a population of 20 neurons in the tonic low-rate low-information regime. To generate the data, we use the poisson\_spike\_gen function of the NIT toolbox.

In the **first scenario** spike trains were independently generated for each neuron with Poisson statistics (low rate to the first stimulus - high rate to the second stimulus), showing zero noise correlations but positive signal correlations.

In the second and third scenario correlations between neurons were created by generating responses to each stimulus as the sum of an independent Poisson process (independent outcome for each neuron) and a shared Poisson process (same outcome across neurons).

The **second scenario** introduced information-limiting noise correlations, with weakly stimulus-dependent positive noise correlations (pairwise Pearson noise correlation was 0.2 for stimulus 1 and 0.1 for stimulus 2).

The **third scenario** maintained the same trial-averaged rates for both stimuli but exhibited a strong stimulus-modulation of correlation strength (pairwise Pearson noise correlation was 0.5 for   
stimulus 1 and 0 for stimulus 2).

1. Calculating the Information Breakdown

Using NIT, it can be examined how noise correlation strength affects the stimulus information in neuronal populations. The information breakdown of neuron pairs without decoding can be calculated with the information function of the toolbox.

In the **first scenario**, with no noise correlations present, the information breakdown correctly identifies that the contribution from noise correlations is null. The information function of the toolbox reveals that the only redundancy between cells arises from the similarity of trial-averaged response profiles, leading to a negative term .

Moving to the **second scenario** with information-limiting noise correlations, the information breakdown confirms an overall decrease in information due to positive signal and noise correlations. The larger negative term reflects this reduction, outweighing the minimal information added by weak stimulus dependence .

In the **third scenario**, the information breakdown correctly reveals that individual cell firing rates convey no information (null and terms). Instead, all the information is carried by the stimulus-dependent noise correlations, highlighted by the positive and substantial term.

The information breakdown method excels in capturing pairwise coding mechanisms but is practically limited to pairs and triplets due to challenges in sampling the necessary trials for larger neural populations. Theoretical considerations suggest that the impact of noise correlations increases proportionally with population size compared to the contribution of individual cell firing rates. Thus, the overall impact of correlations on large populations can only be computed by estimating information from the full population rather than from pairs.

1. Calculating the Population Information

By employing a numerical approach, the overall effects of noise correlations in large neuronal populations can be analyzed. The first step involves predicting the stimuli based on the population activity, achieved through the application of a cross-validated Support Vector Machine (SVM).

A comparative analysis between linear and radial basis function (RBF) SVMs reveals different aspects of the effect of noise correlations. While linear SVMs focus on discerning individual cell responses, RBF SVMs excel in capturing information embedded in stimulus-dependent correlations.

The subsequent step is to calculate the Mutual Information (MI) using the information function of the toolbox to quantify the extent to which information about stimuli is obtained from the decoded population activity. To ensure a robust assessment, a comparative analysis is conducted for information derived from both actual population spike trains, which inherently contain correlations, and information sourced from pseudo-population responses. These pseudo-responses are generated by shuffling trials, effectively eliminating noise correlations. The entire procedure is executed for both shuffled and simulated data in each of the three scenarios. The calculated Mutual Information is then compared to gain insights into the effects of noise correlations on Population Information in large neuronal populations.

In the **first scenario**, the numerical approach correctly indicated that correlations held no information, as evidenced by identical information values for real and shuffled data, and for both linear and RBF SVMs.

In the **second scenario**, the analysis revealed an information-limiting impact of correlations at the population level, nearly halving the total information compared to a relatively minor effect at the pairwise level. This aligns with theoretical expectations and recent reports highlighting significant correlation effects in larger populations. Furthermore, the results suggested a more substantial positive contribution from stimulus-dependent correlations, indicated by higher information decoded with radial SVM compared to linear SVM.

In the **third scenario**, the population analysis correctly identified that individual cell firing rates contributed no information. Instead, considerable information was available from the stimulus dependence of noise correlations, particularly evident in the large information found only with radial SVM decoding in real population response.