

Improved estimation and interpretation of correlations in neural circuits

Dimitri Yatsenko¹, Krešimir Josić², Alexander S. Ecker^{1,3,4}, Emmanouil Froudarakis¹, R. James Cotton¹, Andreas S. Tolias^{1,5,*}

1 Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA

2 Department of Mathematics and Department of Biology and Biochemistry, University of Houston, Houston, TX, USA

3 Werner Reichardt Center for Integrative Neuroscience and Institute for Theoretical Physics, University of Tübingen, Germany

4 Bernstein Center for Computational Neuroscience, Tübingen, Germany

5 Department of Computational and Applied Mathematics, Rice University, Houston, TX, USA

*** E-mail: atolias@cns.bcm.edu**

Abstract

Ambitious projects aim to record the activity of ever larger and denser neuronal populations *in vivo*. Correlations in neural activity measured in such recordings can reveal important aspects of neural circuit organization. However, estimating and interpreting large correlation matrices is statistically challenging. Estimation can be improved by regularization, *i.e.* by imposing a structure on the estimate. The amount of improvement depends on how closely the assumed structure represents dependencies in the data. Therefore, the selection of the most efficient correlation matrix estimator for a given neural circuit must be determined empirically. Importantly, the identity and structure of the most efficient estimator informs about the types of dominant dependencies governing the system. We sought statistically efficient estimators of neural correlation matrices in recordings from large, dense groups of cortical neurons. Using fast 3D random-access laser scanning microscopy of calcium signals, we recorded the activity of nearly every neuron in volumes 200 μm wide and 100 μm deep (150–350 cells) in mouse visual cortex. We hypothesized that in these densely sampled recordings, the correlation matrix should be best modeled as the combination of a sparse graph of pairwise partial correlations representing interactions between the observed neuronal pairs and a low-rank component representing common fluctuations and external inputs. Indeed, in cross-validation tests, the covariance matrix estimator with this structure consistently outperformed other regularized estimators. The sparse component of the estimate defined a graph of interactions. These interactions reflected the physical distances and orientation tuning properties of cells: The density of positive ‘excitatory’ interactions decreased rapidly with geometric distances and with differences in orientation preference whereas negative ‘inhibitory’ interactions were less selective. Because of its superior performance, this ‘sparse + latent’ estimator likely provides a more physiologically relevant representation of the functional connectivity in densely sampled recordings than the sample correlation matrix.

Author Summary

It is now possible to record the spiking activity of hundreds of neurons at the same time. A meaningful statistical description of the collective activity of these neural populations – their ‘functional connectivity’ – is a forefront challenge in neuroscience. We addressed this problem by identifying statistically efficient estimators of correlation matrices of the spiking activity of neural populations. Various underlying processes may reflect differently on the structure of the correlation matrix: Correlations due to common network fluctuations or external inputs are well estimated by low-rank representations, whereas correlations due to linear interactions between specific pairs of neurons are well approximated by their pairwise *partial* correlations. In our data obtained from fast 3D two-photon imaging of calcium signals

of large and dense groups of neurons in mouse visual cortex, the best estimation performance was attained by decomposing the correlation matrix into a sparse network of partial correlations (‘interactions’) combined with a low-rank component. The inferred interactions were both positive (‘excitatory’) and negative (‘inhibitory’) and reflected the spatial organization and orientation preferences of the interacting cells. We propose that the most efficient among many estimators provides a more informative picture of the functional connectivity than previous analyses of neural correlations.

Introduction

Functional connectivity is a statistical description of observed *multineuronal* activity patterns not reducible to the response properties of the individual cells. Functional connectivity reflects local synaptic connections, shared inputs from other regions, and endogenous network activity. Although functional connectivity is a phenomenological description without a strict mechanistic interpretation, it can be used to generate hypotheses about the anatomical architecture of the neural circuit and to test hypotheses about the processing of information at the population level.

Pearson correlations between the spiking activity of pairs of neurons are among the most familiar measures of functional connectivity [1–5]. In particular, *noise correlations*, *i.e.* the correlations of trial-to-trial response variability between pairs of neurons, have a profound impact on stimulus coding [1, 2, 6–11]. In addition, noise correlations and correlations in spontaneous activity have been hypothesized to reflect aspects of synaptic connectivity [12]. Interest in neural correlations has been sustained by a series of discoveries of their nontrivial relationships to various aspects of circuit organization such as the physical distances between the neurons [13, 14], their synaptic connectivity [15], stimulus response similarity [3–5, 15–22], cell types [23], cortical layer specificity [24, 25], progressive changes in development and in learning [26–28], changes due to sensory stimulation and global brain states [5, 21, 29–32].

Neural correlations do not come with ready or unambiguous mechanistic interpretations. They can arise from monosynaptic or polysynaptic interactions, common or correlated inputs, oscillations, top-down modulation, and background network fluctuations, and other mechanisms [33–38]. But multineuronal recordings do provide more information than an equivalent number of separately recorded pairs of cells. For example, the eigenvalue decomposition of the covariance matrix expresses shared correlated activity components across the population; common fluctuations of population activity may be accurately represented by only a few eigenvectors that affect all correlation coefficients. On the other hand, a correlation matrix can be specified using the *partial correlations* between pairs of the recorded neurons. The partial correlation coefficient between two neurons reflects their linear association conditioned on the activity of all the other recorded cells [39]. Under some assumptions, partial correlations measure conditional independence between variables and may more directly approximate causal effects between components of complex systems than correlations [39]. For this reason, partial correlations have been used to describe interactions between genes in functional genomics [40, 41] and between brain regions in imaging studies [42, 43]. These opportunities have not yet been explored in neurophysiological studies where most analyses have only considered the distributions of pairwise correlations [2, 4, 5, 13].

However, estimation of correlation matrices from large populations presents a number of numerical challenges. The amount of recorded data grows only linearly with population size whereas the number of estimated coefficients increases quadratically. This mismatch leads to an increase in spurious correlations, overestimation of common activity (*i.e.* overestimation of the largest eigenvalues) [44], and poorly conditioned partial correlations [40]. The *sample correlation matrix* is an unbiased estimate of the true correlations but its many free parameters make it sensitive to sampling noise. As a result, on average, the sample correlation matrix is farther from the true correlation matrix than some structured estimates.

Estimation can be improved through *regularization*, the technique of deliberately imposing a structure on an estimate in order to reduce its estimation error [40, 45]. To ‘impose a structure’ on an estimate means to bias (‘shrink’) it toward a reduced representation with fewer free parameters, the *target estimate*.

The optimal target estimate and the optimal amount of shrinkage can be obtained from the data sample either analytically [40, 44, 46] or by cross-validation [47]. An estimator that produces estimates that are, on average, closer to the truth for a given sample size is said to be more *efficient* than other estimators.

Although regularized covariance matrix estimation is commonplace in finance [46], functional genomics [40], and brain imaging [43], surprisingly little work has been done to identify optimal regularization of neural correlation matrices.

Improved estimation of the correlation matrix is beneficial in itself. For example, improved estimates can be used to optimize decoding of the population activity [47]. But reduced estimation error is not the only benefit of regularization. Finding the most efficient among many regularized estimators leads to insights about the system itself: the structure of the most efficient estimator is a parsimonious representation of the regularities in the data.

The advantages due to regularization increase with the size of the recorded population. With the advent of big neural data [48], the search for optimal regularization schemes will become increasingly relevant in any model of population activity. Since optimal regularization schemes are specific to systems under investigation, the inference of functional connectivity in large-scale neural data will entail the search for optimal regularization schemes. Such schemes may involve combinations of heuristic rules and numerical techniques specially designed for given types of neural circuits.

What structures of correlation matrices best describe the multineuronal activity in specific circuits and in specific brain states? More specifically, are correlations in the visual cortex during visual stimulation best explained by common fluctuations or by local interactions within the recorded microcircuit?

To address these questions, we evaluated four regularized covariance matrix estimators that imposed different structures on the estimate. The estimators are designated as follows:

C_{sample} – sample covariance matrix, the unbiased, unregularized estimator.

C_{diag} – linear shrinkage of covariances toward zero, *i.e.* toward a diagonal covariance matrix.

C_{factor} – a low-rank approximation of the sample covariance matrix, representing inputs from unobserved shared factors (latent units).

C_{sparse} – sparse partial correlations, *i.e.* a large fraction of the *partial* correlations between pairs of neurons are set to zero.

$C_{\text{sparse+latent}}$ – sparse partial correlations between the recorded neurons *and* linear interactions with a number of latent units.

First, we used simulated data to demonstrate that the selection of the optimal estimator indeed pointed to the true structure of the dependencies in the data.

We then performed a cross-validated evaluation to establish which of the four regularized estimators was most efficient for representing the population activity of dense groups of neurons in mouse primary visual cortex recorded with high-speed 3D random-access two-photon imaging of calcium signals. In our data, the sample correlation coefficients were largely positive and low. We found that the best estimate of the correlation matrix was $C_{\text{sparse+latent}}$. This estimator revealed a sparse network of partial correlations (‘interactions’), between the observed neurons; it also inferred latent units exerting linear effects on the observed neurons. We analyzed these networks of partial correlations and found the following: Whereas significant noise correlations were predominantly positive, the inferred interactions had a large fraction of negative values possibly reflecting inhibitory circuitry. Moreover, we found that these interactions exhibited a stronger relationship to the physical distances and to the differences in preferred orientations than noise correlations. In contrast, the inferred negative interactions were less selective.