

# Improved estimation of neural correlations suggests detailed interactions in visual cortex

Dimitri Yatsenko,<sup>1,\*</sup> Krešimir Josić<sup>2</sup>, Alexander S. Ecker<sup>1,3,4</sup>, Emmanouil Froudarakis<sup>1</sup>, R. James Cotton<sup>1</sup>, Andreas S. Tolias<sup>1,5</sup>

**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: yatsenko@cns.bcm.edu

## Abstract

Ambitious projects currently under way aim to record the activity of ever larger and denser subsets of neurons in behaving animals. It is anticipated that correlations measured in such recordings will uncover aspects of functional organization of neural circuits. However, estimation and interpretation of large correlation matrices from finite recordings can be challenging. Estimation can be improved by regularization: the biasing of the estimate toward a low-dimensional, less variable approximation. The amount of improvement depends on how closely the reduced approximation captures the dominant interactions with fewest terms. Therefore, the selection of the most efficient estimator is an empirical question dependent on the system under investigation.

We compared regularized correlation matrix estimators biased toward four respective families of low-dimensional correlation structures: independent, latent factors, sparse partial correlations, and sparse partial correlations with latent factors. The estimators' performance was evaluated on the noise correlation matrices from spatially compact groups of 91–314 neurons (average 206) in mouse visual cortex (31 sites in 24 animals) acquired by *in vivo* fast 3D random-access two-photon imaging of calcium signals. Recordings lasting between 15 and 20 minutes were deconvolved and binned at 150 ms. Each estimator was optimized and evaluated by nested cross-validation. We found that both sparse partial correlations and latent factors were required for efficient estimation. In addition to reducing estimation error, optimized estimates provided compact representations of the correlation structure consisting of a sparse network of pairwise interactions and global diffuse fluctuations. The density of positive interactions in these networks decreased rapidly with distance and with difference in orientation preference, whereas negative interactions were less selective. The physiological interpretation of these components remains in question and must be addressed by future experiments where synaptic connectivity, cell types, and local field potentials may be used to corroborate such interpretations.

## Author Summary

Correlation matrices of the activity of populations of neurons are useful descriptors of their functional organization with implications for stimulus coding and circuit architecture. Estimates of the correlation matrix can be dramatically improved by *regularization*: the biasing of the estimate toward one of several possible low-dimensional approximations. Using synthetic data, we demonstrated that different estimators performed better or worse than others depending on how well their regularization targets captured the structure of the true correlation matrix. Then, we compared the performance of several estimators on the population calcium activity of relatively large and dense groups of neurons in mouse visual cortex

under visual stimulation. We found that the best estimator for such neural data was composed of a sparse network of partial correlations between pairs of neurons combined with several latent units interacting with the entire population. We hypothesize that such succinct representations of the correlation structure may more accurately reflect the circuit’s functional organization than the conventional correlations.

## Introduction

Pearson correlations between the spiking activity of pairs of neurons, or simply *neural correlations*, are the most familiar descriptive statistics of neural population activity [1–5]. For example, *noise correlations*, *i.e.* the correlations of stimulus response variability between pairs of neurons, have been shown to have profound theoretical implications for stimulus coding [1, 2, 6–8]. In addition, neural correlations are hypothesized to reflect aspects of functional connectivity in neural circuits. Such interpretation is supported by a series of discoveries of nontrivial relationships between neural correlations and other aspects of circuit organization such as the physical distance separating the neurons [9, 10], their synaptic connectivity [11], stimulus tuning similarity [3, 11], cortical layer specificity [12, 13], progressive changes in development and in learning [14], changes due to sensory stimulation and global brain states [5, 15–17], and others.

However, neural correlations do not submit to ready or unambiguous mechanistic interpretation. Theoretical studies and simulations have shown that neural correlations on various temporal scales may arise from combinations of multiple mechanisms including direct synaptic interactions, common inputs or correlated inputs, chains of multiple synaptic connections, oscillations, top-down modulation, and background network fluctuations [18–22].

A correlation matrix provides more information than the equivalent number of pairwise correlations assessed in isolation. Yet early studies of neural correlations were based on measurements from isolated pairs of neurons; the effects of correlations on neural coding were then extrapolated in simulations and theoretical studies [2, 19]. Modern multineuronal recordings from increasingly large and dense subsets of neurons in intact circuits allow estimating the entire correlation matrix of large populations of highly interconnected neurons. Such estimates can be transformed into various representations that express different aspects of the correlation structure and may suggest different mechanistic interpretations. For example, factor analysis or principle component analysis may reveal common activity fluctuations across the entire population [23]. The correlation matrix also allows computing the *partial correlations* between specific pairs of neurons, *i.e.* correlations that remain after accounting for the linear effects of all other recorded cells. Under several assumptions (predominance of linear effects (sufficiently complete sampling of interacting neurons, and sufficient sample size), partial correlations may better reflect the direct statistical associations or interactions between pairs of neurons. In multivariate normal distributions, zero partial correlations between pairs of variables indicate conditional independence or lack of interaction between the pair and the network of non-zero interactions constitute Gaussian Graphical Models or Gauss-Markov Random Fields [24]. With other distributions, the correspondence between conditional dependence and partial correlations diminishes or breaks down [25]. Estimation of networks of partial correlations (sometimes called *association networks*) have been applied to map gene interactions [26, 27] or to establish functional connectivity between brain regions from fMRI data [28, 29].

Multineuronal recordings also present major challenges for the estimation and interpretation of correlations. As the population size increases, the amount of recorded data grows only linearly whereas the number of estimated coefficients grows quadratically. This mismatch leads to increased opportunities for spurious correlations, overestimation of shared activity (*i.e.* overestimation of large eigenvalues) [30], and poorly conditioned estimates of the partial correlations [26].

In this study, we pursued two related aims: (a) improved estimation of neural correlation matrices and (b) discovery of low-dimensional structure of correlations in recordings of multineuronal activity to facilitate their interpretation.

Estimation of the correlation matrix relies on estimating the means and variances of the activity of individual cells. To estimate the mean activity, we resort to the usual *sample mean* without attempting to improve it. Estimation of a *covariance matrix* is equivalent to estimating both variances and correlations. Therefore, we rephrase the problem as one of optimal estimation of the covariance matrix given an estimate of the mean firing rates.

Estimation can be improved through *regularization*: the deliberate biasing of the estimate toward one of several possible low-dimensional approximations referred to here as *target estimates* [26,31]. The usual estimate, the *sample covariance matrix* has the advantage of being unbiased but, on average, it falls far from the true covariance matrix due to its sensitivity to sampling noise in the data. Low-dimensional estimates of various forms are typically less susceptible to sampling noise but are also liable to introduce their respective biases away from the true covariance matrix. Regularization works by striking a favorable balance between bias and variability. Regularization can produce *some* improvement even with an arbitrary target estimate that has no relation to the true covariance structure in the data. Yet, when the target estimate is well suited for capturing the important features of the true covariance matrix with few terms, it will introduce minimal bias and outperform other estimators.

To illustrate key concepts, Figure 1 illustrates a regularization scheme based on *covariance selection* [32]. In covariance selection, the estimate is produced by fitting only an optimal subset of the coefficients of the *precision matrix* (the inverse of the covariance matrix) while setting the remaining coefficients to zero. Scaling the precision matrix so that its diagonal is composed of all  $-1$ 's produces the matrix of partial correlations. Thus zeros in the precision matrix produce zero partial correlations between the corresponding pairs of neurons. Panel 1 A depicts the sample correlation matrix of a period of somatic calcium signals of a group of 298 neurons in mouse visual cortex. Due to the low-pass filtration effect of calcium dye kinetics, correlations in unprocessed calcium signals are much higher than typical firing rate correlations on shorter temporal scales. Panel 1 B magnifies the fragment of the sample correlation matrix outlined in 1 A whereas 1 C shows the corresponding fragment of the regularized estimate. The regularized estimate is produced from the same data by zeroing the optimal set of 31501 (71.2%) of off-diagonal coefficients of the precision matrix and fitting the remaining coefficients. Panels 1 D and E show the same fragments of the respective estimates of the partial correlation matrix. The apparent similarity of the two estimates of the correlation matrix belies the utterly different partial correlation structure. The regularized estimate is less sensitive to sample noise in the data, is likely to be closer to truth in most cases than the unbiased estimate, which may be ascertained by cross-validation studies.

It may be tempting to interpret the regularized sparse estimate of partial correlations as a more accurate representation of the functional connectivity than the original correlation coefficients: after all, the regularized solution consistently produces better estimates and has fewer parameters which are easier to interpret. However, an alternative regularization scheme using 43 latent factors interacting linearly with all neurons while having about the same number of parameters, performed even better than covariance selection on this dataset.

It is therefore important to compare multiple families of regularized estimators on the system under investigation. The family that consistently outperforms the others may be proposed as candidate low-dimensional representation of the correlation structure. The greater efficiency of one estimator over another is not a mathematical necessity but an empirical finding depending on the prevalent types of correlation structure in the investigated system.

In this study, we compared four families of regularized combining partial interactions and latent units: 'shrinkage toward diagonal', 'shrinkage toward a multifactor model', 'sparse partial correlations', and 'sparse+latent partial correlations'. First, in simulation with known low-dimensional structure of correlations, we demonstrate that regularized estimators from matching families of low-dimensional targets generally outperform all other estimators. We then repeat this study on neural data from dense groups of neurons in mouse visual cortex under visual stimulation. We find that most efficient estimation was achieved by modeling correlations as a sparse network of partial correlations combined with several latent

units interacting with the entire population. Finally, we analyze the relationship between the elements of such estimates and other properties of the circuit such as physical distances between neurons and their preferred orientations.

## Results

### Covariance estimation

We aim to estimate the true covariance matrix

$$\Sigma = \mathbb{E} [(x - \mu)(x - \mu)^\top] \quad (1)$$

where  $\mathbb{E} [\cdot]$  denotes expectation;  $x$  is the  $p \times 1$  vector of real-valued instantaneous firing rates in bins of duration  $\Delta t$ . The vector of mean firing rates is  $\mu = \mathbb{E} [x]$ .

The usual estimator of the covariance matrix is the *sample covariance matrix*  $C_0$  computed from the empirical sample of observations  $x(1), \dots, x(n)$  as

$$C_0 = \frac{1}{\nu} \sum_{t=1}^n (x(t) - \mu)(x(t) - \mu)^\top, \quad (2)$$

where  $\nu$  is the number of degrees of freedom per neuron in the sample ( $\nu = n - 1$  if observations are independent). The sample covariance matrix is unbiased, *i.e.*  $\mathbb{E} [C_0] = \Sigma$ . For finite sample sizes, however,  $C_0$  is not as close to  $\Sigma$  as a number of biased estimators.

We considered four regularized estimators based on distinct families of low-dimensional target estimates: ‘independent’, ‘latent factors’, ‘sparse interactions’, and ‘sparse+latent’ (Fig. 2 row 1).

In the first regularized estimator  $C_{\text{diag}}$ , the target estimate is the diagonal matrix  $D$  containing on its diagonal estimates of the variances. The regularized estimate is obtained by linear *shrinkage* of the unbiased estimate  $C_0$  toward  $D$  controlled by the scalar *shrinkage intensity* parameter  $\lambda \in [0, 1]$ :

$$C_{\text{diag}} = (1 - \lambda)C_0 + \lambda D \quad (3)$$

The diagonal target estimate expresses the idea of a lack of dependence (or of linear association) between the activity of observed neurons (Fig. 2 A). If this assumption aptly describes recorded data, then strong shrinkage toward  $D$  will add little bias while strongly reducing the variability of the estimate. Shrinkage allows for partial commitment to the low-dimensional representation.

In the second regularized estimator  $C_{\text{factor}}$ , the target estimate is the factor model  $F = LL^\top + \Psi$  with  $d$  factors so that  $L$  is the  $p \times d$  matrix of *factor loadings* and the diagonal matrix  $\Psi$  contains the independent variances of each neuron. Then the estimate is

$$C_{\text{factor}} = (1 - \lambda)C_0 + \lambda F \quad (4)$$

This estimator has two hyperparameters: the number of factors  $d$  and shrinkage intensity  $\lambda$ . The target estimate  $F$  expresses the assumption that correlated fluctuations in the population activity are driven by a small number of latent factors that affect many cells while direct interactions between cells are insignificant (Fig. 2 B).

The third estimator  $C_{\text{sparse}}$  is based on the assumption that all correlations are the result of direct linear interactions between pairs of observed cells and that such interactions occur only between a fraction of such pairs (Fig. 2 C). This assumption is enforced by reducing to zero the majority of pairwise *partial correlations* in the recorded population. While usual correlations are calculated from the marginal distribution without conditioning on all the other neurons, the partial correlation expresses the pairwise correlation conditioned on the activity of all the other cells. If cells only exert linear effects on each other

and all neurons are recorded, partial correlations express the direct interactions between neurons. Therefore, this estimator is biased toward the assumption that correlations arise due to interactions between a subset of pairs of recorded neurons (Fig. 3C). When the partial correlation between a pair of neurons is zero, then the corresponding element of the inverse of the covariance matrix (often referred to as the *precision matrix* or *concentration matrix*) must be zero as well. Then the estimator is

$$C_{\text{sparse}} = S^{-1} \quad (5)$$

where  $S$  is a sparse matrix with a large fraction of zeros in its off-diagonal elements. The estimate has one hyperparameter to regulate the sparsity (fraction of off-diagonal zeros) in  $S$ .

Finally, we consider the fourth estimator  $C_{\text{sparse+lowrank}}$ , which provides for both common latent factors interacting with all recorded neurons and sparse interactions between the recorded neurons (Fig. 2D). This estimator has the form

$$C_{\text{sparse+latent}} = (S - LL^T)^{-1}, \quad (6)$$

where, as above,  $S$  is a sparse matrix and  $L$  is a  $d \times p$  matrix of factor loadings. The estimator has two hyperparameters: the number of latent units  $d$  and the sparsity of  $S$ .

## Simulation

To verify our approach and to illustrate the performance of the four regularized estimators, we constructed four model populations of size  $p = 50$  neurons each with correlation structures of the same type as the target estimates of the first (Fig. 2 Row 2). Panel A2 contains a diagonal correlation matrix matching the target of  $C_{\text{diag}}$ . Panel B2 is a factor model with 3 factors conforming to the family of target estimates of  $C_{\text{factor}}$ . Panel B3 is a correlation matrix with sparse partial correlations (76% off diagonal zeros in the precision matrix). Finally, Panel B4 is a correlation matrix whose inverse is composed the sum of a sparse matrix (82% sparse) and a low-rank component (rank=1). Row 3 contains sample correlation matrices calculated from samples of size  $n = 1000$  from a multivariate normal distributions with the respective correlation matrices from Row 2.

To evaluate the performance of a covariance matrix  $C$ , we define a real-valued *loss function*  $\mathcal{L}(C, \Sigma)$  such that it attains its minimum when  $C = \Sigma$ . The loss function quantifies the error of the estimate, *i.e.* the deviation of the estimate  $C$  from truth  $\Sigma$ .

For the purposes of this study, we adopted the *negative normal log-likelihood loss*<sup>1</sup> defined as:

$$\mathcal{L}(C, \Sigma) = \frac{1}{p} [\ln \det C + \text{tr}(C^{-1}\Sigma)] \quad (8)$$

We normalized the loss function by  $\frac{1}{p}$  to make its value less sensitive population size (For independent neurons, the log likelihood scales linearly with  $p$ ).

This choice is motivated by mathematical convenience. Other popular choices for the loss function are the Frobenius norm of the difference  $C - \Sigma$  [26,30], Stein's entropy loss, and quadratic loss [33,34]. We expect that the main conclusions of our study will not change qualitatively under other well behaved loss functions.

To make a loss function that assume zero at its minimum when  $C = \Sigma$ , we define *excess loss* as

$$\mathcal{L}_0(C, \Sigma) = \mathcal{L}(C, \Sigma) - \mathcal{L}(\Sigma, \Sigma) \quad (9)$$

---

<sup>1</sup>The loss function  $\mathcal{L}(\cdot|C, \Sigma)$  is related to the multivariate log-likelihood function  $L(\Sigma|C_0)$  as

$$L(\Sigma|C_0) = -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \mathcal{L}(\Sigma, C_0) \quad (7)$$

In Figure 2, row 4 shows the mean excess losses and their standard errors calculated from 30 samples of sizes  $n=250, 500, 1000, 2000$ , and  $4000$  for each estimator and each kind of ground truth. For each estimator, its hyperparameters were estimated by nested cross-validation (See Methods for more details.)

As expected, estimators with matching low-dimensional structures typically outperformed the other estimators. There were two exceptions to this observation. For small sample sizes, before the data were sufficient to reveal the true correlation structure and to allow the correct model to dominate, estimates with simpler targets often outperformed the more matching estimate.

With empirical data, the truth  $\Sigma$  is not accessible and the relative values of loss must be estimated solely from the data. This may be accomplished through *validation*. In validation, the data are split into a training subset from which the estimate is obtained and a validation subset. We then compute the *validation loss*

$$\mathcal{L}(C, C'_0) \quad (10)$$

$\mathcal{L}(\cdot, \cdot)$  is additive in its second argument so that

$$\mathcal{L}(\hat{\Sigma}, X_1) + \mathcal{L}(\hat{\Sigma}, X_2) \equiv \mathcal{L}(\hat{\Sigma}, X_1 + X_2) \quad (11)$$

Then *validation loss*

$$\hat{\ell} = \mathcal{L}(\hat{\Sigma}, \hat{\Sigma}'_0) \quad (12)$$

is an unbiased estimate of the risk:

$$\mathbb{E}_{(\hat{\Sigma}, \hat{\Sigma}'_0)} [\hat{\ell}] = \mathbb{E}_{(\hat{\Sigma}, \hat{\Sigma}'_0)} [\mathcal{L}(\hat{\Sigma}, \hat{\Sigma}'_0)] = \mathbb{E} [\mathcal{L}(\hat{\Sigma}, \mathbb{E} [\hat{\Sigma}'_0])] = \mathbb{E} [\mathcal{L}(\hat{\Sigma}, \Sigma)] = r \quad (13)$$

Therefore, estimators resulting in consistently lower validation loss can be inferred to produce estimates that are closer to truth than estimators with higher validation loss.

Other popular loss functions such as entropy loss do not comply with Eq. 11 and their validation loss is not an unbiased estimate of risk.

Since  $\mathcal{L}(\cdot, \cdot)$  is equivalent to negative normal log likelihood, the above derivation has led to the familiar criterion that the optimal covariance matrix estimator is one that consistently maximizes cross-validated normal log likelihood.

To evaluate the performance of the different estimators, we computed the excess loss for all combinations of model populations and estimators, including the sample covariance. The first striking observation is that in all cases all four regularized estimators performed substantially better than the sample covariance (Fig. 2, fourth row). In particular, this is even true for the case where the population did not have any low-dimensional structure at all (Fig. 2E). While it may appear surprising at first sight, this counter-intuitive phenomenon is known as *Stein's phenomenon* or *Stein's paradox* [35], named after its discoverer Charles Stein [36]. A common misconception about regularization is that its effect depends on accurate prior knowledge about the structure of the data. However, substantial improvement can be attained by shrinking the unbiased estimate toward an arbitrary target as long as the target is less variable than the unbiased estimator. The more accurate description of regularization is as of the optimal tradeoff between estimation and approximation error – the so-called “bias-variance tradeoff”.

Thus, if taken in isolation a regularized estimator improves the estimate we should not interpret this result to suggest that the estimator's target has the same low-dimensional structure as the data-generating process. However, when comparing multiple estimators against each other, the one whose target estimate most closely matches the true value with the smallest number of parameters will reduce the estimation error with the least increase in approximation error. Indeed, in all four toy examples with a low-dimensional structure, the estimator with the matching regularization target performed best (Fig. 3A–D, fourth row). Since the estimator combining low-rank and sparse interactions combines to low-dimensional targets and includes the simpler estimators as special cases, it performed almost as



well even when the low-dimensional structure did not include either a low-rank component or sparse interactions (Fig. 3B,C).

Since the above evaluations of excess loss require knowledge of ground truth, this analysis can be done only on simulated data. However, *validation loss* can be used as an unbiased empirical estimate of loss (see Methods, Eq. 10). Validation loss uses an independent validation sample to estimate the value of loss without access to ground truth. Indeed, this analysis revealed a pattern of results that reproduced the results obtained with access to ground truth above (Fig. 3, last row). Because validation loss is computed by comparing estimates to noisy sample covariance matrices from smaller validation sets, validation loss has higher variance than excess loss. Under our chosen loss function, validation loss does not converge to zero.

## Covariance estimation in neural data

We recorded the calcium activity of dense populations of neurons in the supragranular layers in primary visual cortex of anesthetized mice using fast random-access 3D scanning two-photon microscopy [37, 38]. We presented numerous repetitions of full-field drifting gratings (Fig. 1A and 1B) to the eye contralateral to the imaged site. This technique allowed us to record from a large number (150–350) of cells in a small volume of cortical tissue ( $200 \times 200 \times 100 \mu\text{m}^3$ ) in layers 2/3 and 4. We deconvolved somatic calcium signals using sparse nonnegative deconvolution [39] (Fig. 1C and 1D) and subtracted the average stimulus response to remove averaged linear effects of the stimulus. From this residual response we computed the sample noise covariance matrix (Fig. 1E).

In such highly localized populations both direct interactions between cells and common diffuse inputs are likely to contribute to the overall population variability. At the same time, most correlations are relatively small (Fig. 1E), suggesting that a simple shrinkage towards independence may provide a sufficiently well regularized estimate. By comparing the performance of our differently regularized estimators we can gain insights into which of these aspects are important in our data.

As expected, all regularized estimators outperformed the sample covariance estimator substantially (Fig. 4). Say something about how shrinkage, low-rank and sparse inverse relate. Finally, the combined sparse and low-rank estimator dominated all others significantly (Fig. 4), showing that both hidden units and direct interactions are important in our data. The improvement from the low-rank estimator to the combined one is much larger than that from the sparse inverse to the combined one, suggesting that in this dataset direct interactions contribute more strongly to the correlation structure than hidden units do.

## Relationship between functional covariance structure and circuit architecture

### Methods

#### Estimation of parameters and hyperparameters of covariance estimators

##### Regularization

Under many loss functions<sup>2</sup>, the risk can be decomposed as the sum

$$r = b + \varepsilon \tag{14}$$

---

<sup>2</sup>The strict equality in Eq. 14 does not hold under the loss function in Eq. 8. However, the equality does hold for its close cousin, Stein’s *entropy loss*, which only differs by the order of its arguments and a constant offset:  $\mathcal{L}_s(\hat{\Sigma}, \Sigma) \equiv \mathcal{L}(\Sigma, \hat{\Sigma}) - \mathcal{L}(\Sigma, \Sigma)$ . This deficiency presents no difficulty because we minimize the risk directly, without assessing the two error components individually. The bias-variance decomposition is presented here to motivate the use of regularization.

of *approximation error* (“bias” or systematic error)

$$b = \mathcal{L}(\bar{\Sigma}, \Sigma) \quad (15)$$

and *estimation error* (“variance”)

$$\varepsilon = \mathbb{E}_{\hat{\Sigma}} \left[ \mathcal{L}(\hat{\Sigma}, \bar{\Sigma}) \right] \quad (16)$$

where  $\bar{\Sigma} = \mathbb{E}[\hat{\Sigma}]$  is the expected value of the estimate.

An unbiased estimator, such as the sample covariance,  $\hat{\Sigma}_0$ , makes  $\bar{\Sigma} = \Sigma$  and thereby minimizes approximation error, but may be excessively susceptible to sample noise and result in high estimation error.

The estimator risk can be reduced by *regularization* – a deliberate biasing (“*shrinkage*”) of the estimate toward a low-dimensional, less variable *target estimate* [30, 31]. A regularized estimator solves the bias-variance tradeoff to produce a biased but less variable estimate to minimize the estimator’s risk. Various regularization schemes focus on the dimensionality reduction part by selecting the optimal target estimate from a family of estimates with reduced dimensionality [?]. Other estimators only shrink the sample covariance matrix toward a single target estimator [26]. Yet other regularizers effectively combine shrinkage and dimensionality reduction [?].

#### Estimator $\mathcal{A}$ : Shrinkage toward diagonal.

*Estimator  $\mathcal{A}$*  shrinks the sample covariance matrix  $\hat{\Sigma}_0$  toward that of the independent model shown in Fig. 2A. The covariance matrix of a GGM in which all units are independent will be diagonal. For example, we could leave the sample variances on the diagonal set the off-diagonal terms to zero. The target estimate is then

$$\hat{T} = \hat{\Sigma}_0 \circ I \quad (17)$$

where  $\circ$  denotes the entrywise matrix product (Hadamard product). Alternatively, the target could be a diagonal matrix with all variances equal to the average sample variance,  $v = \frac{1}{p} \text{tr}(\hat{\Sigma}_0)$ , so that

$$\hat{T} = \frac{1}{p} \text{tr}(\hat{\Sigma}_0) I \quad (18)$$

The equal-variance target in Eq. 18 has only one degree of freedom and thus lower estimation error but higher bias than the target with sample variances in Eq. 17. It is not immediately clear which target is better in a particular application. We therefore use the linear mixture of the two targets parameterized by the mixing coefficient  $\alpha \in [0, 1]$ ,

$$\hat{T}_\alpha = (1 - \alpha)(\hat{\Sigma}_0 \circ I) + \frac{\alpha}{p} \text{tr}(\hat{\Sigma}_0) I. \quad (19)$$

The estimator  $\hat{\Sigma}_{\alpha, \lambda}$  is the linear shrinkage of the sample covariance matrix  $\hat{\Sigma}_0$  toward  $\hat{T}_\alpha$  controlled by the mixing proportion  $\lambda \in [0, 1]$ ,

$$\hat{\Sigma}_{\alpha, \lambda}^{\mathcal{A}} = (1 - \lambda)\hat{\Sigma}_0 + \lambda\hat{T}_\alpha. \quad (20)$$

This covariance matrix estimator is implemented by R’s `corpcor` package [40] and is used widely [26].

We estimated the optimal values of hyperparameters  $\alpha$  and  $\lambda$  by cross-validation within the training dataset.



### Estimate $\mathcal{B}$ : Shrinkage toward a factor model

*Estimator  $\mathcal{B}$*  shrinks the sample covariance matrix  $\hat{\Sigma}_0$  toward a target estimated by *factor analysis* with  $d$  factors, which corresponds to graphical models of the kind in Fig. 2B. For multivariate normal distributions, factor analysis estimates a graphical model in which the observed units do not interact directly but are influenced by  $d$  common inputs.

The target estimate is then

$$\hat{T}_d = \hat{L}_d + \hat{\Psi} \quad (21)$$

where  $\hat{L}_d$  is a  $p \times p$  matrix of rank  $d$  and  $\hat{\Psi}$  is a  $p \times p$  diagonal matrix. The value of  $\hat{T}_d$  is found by minimizing  $\mathcal{L}(\hat{T}_d, \hat{\Sigma}_0)$ .

Just as with estimator  $\mathcal{A}$ ,  $\hat{\Sigma}_0$  is shrunk toward  $\hat{T}_d$  by linear mixing controlled by shrinkage intensity  $\lambda \in [0, 1]$ ,

$$\hat{\Sigma}_{d,\lambda}^{\mathcal{B}} = (1 - \lambda)\hat{\Sigma}_0 + \lambda\hat{T}_d, \quad (22)$$

and the optimal values of  $d$  and  $\lambda$  are found by cross-validation within the training set.

Factor analysis has long been used in natural sciences to infer the influence of a one or several latent variables exerting linear effects on a larger population of observed variables [?]. In neuroscience, the hypothesis that the population activity in small local circuits is largely driven by a few latent factors has been investigated [23, 41]. Estimation of covariance matrices by shrinkage toward a factor model has been used in finance for portfolio risk assessment [34, 42].

### Estimator $\mathcal{C}$ : Spare inverse

*Estimator  $\mathcal{C}$*  produces a low-dimensional estimate by finding an approximation of  $\hat{\Sigma}_0$  that has many zeros in its inverse. This approximation problem is known as *covariance selection* [32]. Covariance matrices of this family correspond to graphical models of the type depicted in Fig. 2C.

For Gaussian Graphical Models, zeros in the inverse covariance matrix indicate statistical independence between the corresponding pairs. Thus the inverse covariance matrix  $S = -\Sigma^{-1}$  is also the adjacency matrix of the GGM, indicating which pairs are conditionally dependent. The off-diagonal elements of  $S$  are proportional to partial pairwise correlations between corresponding pairs of neurons.

The covariance matrix estimate is then

$$\hat{\Sigma}_\rho = \hat{S}^{-1} \quad \text{with} \quad \hat{S} = \arg \min_{\substack{S \succ 0 \\ \|S\|_0 \leq \rho}} \mathcal{L}(S^{-1}, \hat{\Sigma}_0) \quad (23)$$

where  $S \succ 0$  signifies the constraint that  $S$  be positive-definite.  $\|S\|_0 \leq \rho$  signifies the constraint that matrix  $S$  have at most  $\rho$  non-zero coefficients. Solving this approximation in this form is computationally challenging. The problem can be made computationally tractable by relaxing [\(use "complex relaxation"\)](#) the  $L_0$  norm to the  $L_1$  norm, which converts the problem into one of convex optimization [43]. The resulting algorithm, known as *graphical lasso* or *GLASSO* [44–47], produces the following estimator

$$\hat{\Sigma}_\lambda^{\mathcal{C}} = \hat{S}^{-1} \quad \text{with} \quad \hat{S} = \arg \min_{S \succ 0} \mathcal{L}(S^{-1}, \hat{\Sigma}_0) + \lambda \|S\|_1 \quad (24)$$

where  $\lambda \|S\|_1$  is the  $L_1$  norm of the matrix.

Just as in the previous estimators, the optimal value of the regularization parameter  $\lambda$  is found from training data by cross-validation.

### Estimator $\mathcal{D}$ : Sparse + low-rank inverse

Estimator  $\mathcal{D}$  produces a low-dimensional estimate by approximating  $\hat{\Sigma}_0$  with a matrix whose inverse is the sum of a sparse matrix and a low-rank matrix. Covariance matrices of this form correspond to Gaussian graphical models of the form depicted in Fig. 2D.

The inverse  $S = \Sigma^{-1}$  of the covariance matrix of a GGM constitutes the graph of conditional dependencies (interactions) between the units. With the assumption of sparse organization of interactions in the network,  $S$  is assumed to be sparse. When the observed units are part of a larger GGM that includes other latent units, let the negative inverse  $S^*$  of the joint distribution of both observed and latent units be composed of four partitions

$$S^* = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^\top & S_{22} \end{pmatrix} \quad (25)$$

where partition  $S_{11}$  describes the interactions between the visible units,  $S_{12}$  describe the interactions between visible and latent units, and  $S_{22}$  describes the interactions between the latent units. Then the covariance matrix of between the observed unit is

$$\Sigma = (S_{11} - S_{12}S_{22}^{-1}S_{12}^\top)^{-1} \quad (26)$$

The rank of matrix  $S_{12}S_{22}^{-1}S_{12}^\top$  equals the number of latent units. Rather than solving for the precise dependency structure with and between latent units in matrices  $S_{12}$  and  $S_{22}$ , we simply replace the matrix  $S_{12}S_{22}^{-1}S_{12}^\top$  with a low-rank matrix  $L = S_{12}S_{22}^{-1}S_{12}^\top$ .

Then we obtain estimator  $\hat{\Sigma}_{d,\rho}$ , parameterized by the sparsity parameter  $\rho$  and the number of latent units  $d$ :

$$\hat{\Sigma}_{d,\rho} = \left( \hat{S} - \hat{L} \right)^{-1} \quad \text{where} \quad (\hat{S}, \hat{L}) = \arg \min_{\substack{S, L \\ (S-L) \succ 0 \\ \|S\|_0 \leq \rho \\ L \succeq 0 \\ \text{rank}(L) \leq d}} \mathcal{L}((S-L)^{-1}, \Sigma_0) \quad (27)$$

Again, in this formulation, the approximation problem is computationally intractable. The problem can be made convex by replacing the  $L_0$  norm with  $L_1$  norm. The resulting algorithm, known as *latent-variable graphical lasso* or *LV-GLASSO* [48, 49], produces estimator  $\hat{\Sigma}_{d,\lambda}^{\mathcal{D}}$ :

$$\hat{\Sigma}_{d,\lambda}^{\mathcal{D}} = \left( \hat{S} - \hat{L} \right)^{-1} \quad \text{where} \quad (\hat{S}, \hat{L}) = \arg \min_{\substack{S, L \\ (S-L) \succ 0 \\ L \succeq 0 \\ \text{rank}(L) \leq d}} \mathcal{L}((S-L)^{-1}, \Sigma_0) + \lambda \|S\|_1 \quad (28)$$

The optimal values of hyperparameters  $\lambda$  and  $d$  are found cross-validation within the training dataset.

## Acknowledgments

Genevra Allen  
Eftychios Pnevmatikakis

## References

1. Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366.

2. Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370: 140–143.
3. Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25: 3661–73.
4. Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area mt: time scales and relationship to behavior. *The journal of Neuroscience* 21: 1676–1697.
5. Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, et al. (2010) The asynchronous state in cortical circuits. *Science* 327: 587–90.
6. Abbott L, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural computation* 11: 91–101.
7. Berens P, Ecker AS, Gerwinn S, Tolias AS, Bethge M (2011) Reassessing optimal neural population codes with neurometric functions. *Proc Natl Acad Sci U S A* 108: 4423–8.
8. Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience* 31: 14272–14283.
9. Smith MA, Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28: 12591–12603.
10. Denman DJ, Contreras D (2013) The structure of pairwise correlation in mouse primary visual cortex reveals functional organization in the absence of an orientation map. *Cereb Cortex* .
11. Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, et al. (2011) Functional specificity of local synaptic connections in neocortical networks. *Nature* 473: 87–91.
12. Hansen BJ, Chelaru MI, Dragoi V (2012) Correlated variability in laminar cortical circuits. *Neuron* 76: 590–602.
13. Smith MA, Jia X, Zandvakili A, Kohn A (2013) Laminar dependence of neuronal correlations in visual cortex. *Journal of neurophysiology* 109: 940–947.
14. Golshani P, Gonçalves JT, Khoshkhoo S, Mostany R, Smirnakis S, et al. (2009) Internally mediated developmental desynchronization of neocortical network activity. *J Neurosci* 29: 10890–9.
15. Goard M, Dan Y (2009) Basal forebrain activation enhances cortical coding of natural scenes. *Nat Neurosci* 12: 1444–9.
16. Kohn A, Zandvakili A, Smith MA (2009) Correlations and brain states: from electrophysiology to functional imaging. *Curr Opin Neurobiol* 19: 434–8.
17. Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, et al. (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327: 584–587.
18. Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes: II. simultaneous spike trains. *Biophysical journal* 7: 419–440.
19. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18: 3870–96.
20. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience* 2: 539–550.

21. Ostojic S, Brunel N, Hakim V (2009) How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *The Journal of Neuroscience* 29: 10234–10253.
22. Rosenbaum R, Josić K (2011) Mechanisms that modulate the transfer of spiking correlations. *Neural computation* 23: 1261–1305.
23. Yu BM, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, et al. (2009) Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of neurophysiology* 102: 614–635.
24. Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. MIT press.
25. Loh PL, Wainwright MJ (2012) Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *arXiv preprint arXiv:12120478* .
26. Schäfer J, Strimmer K, et al. (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* 4: 32.
27. Peng J, Wang P, Zhou N, Zhu J (2009) Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* 104.
28. Varoquaux G, Gramfort A, Poline JB, Thirion B (2012) Markov models for fmri correlation structure: is brain functional connectivity small world, or decomposable into networks? *Journal of Physiology-Paris* 106: 212–221.
29. Ryali S, Chen T, Supekar K, Menon V (2012) Estimation of functional connectivity in fmri data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage* 59: 3852–3861.
30. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88: 365–411.
31. Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, et al. (2006) Regularization in statistics. *Test* 15: 271–344.
32. Dempster A (1972) Covariance selection. *Biometrics* : 157–175.
33. James W, Stein C (1961) Estimation with quadratic loss. In: *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*. volume 1, pp. 361–379.
34. Fan J, Fan Y, Lv J (2008) High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186–197.
35. Efron B, Morris CN (1977) *Stein’s paradox in statistics*. WH Freeman.
36. Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*. volume 1, pp. 197–206.
37. Stosiek C, Garaschuk O, Holthoff K, Konnerth A (2003) In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences* 100: 7319–7324.
38. Reddy GD, Saggau P (2005) Fast three-dimensional laser scanning scheme using acousto-optic deflectors. *J Biomed Opt* 10: 064038.

39. Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, et al. (2010) Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of neurophysiology* 104: 3691–3704.
40. Schäfer J, Opgen-Rhein R, Strimmer K (2010) corpcor: Efficient estimation of covariance and (partial) correlation. R package version 15 7.
41. Ecker AS, Berens P, Cotton RJ, Subramaniam M, George DH, et al. (2014) State dependence of noise correlations in macaque primary visual cortex. *Neuron* 0.
42. Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10: 603–621.
43. Donoho DL, et al. (2000) High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture* : 1–32.
44. Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* 34: 1436–1462.
45. Yuan M, Lin Y (2007) Model selection and estimation in the gaussian graphical model. *Biometrika* 94: 19–35.
46. Banerjee O, El Ghaoui L, d’Aspremont A (2008) Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research* 9: 485–516.
47. Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9: 432–441.
48. Chandrasekaran V, Parrilo PA, Willsky AS (2010) Latent variable graphical model selection via convex optimization. In: *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on. IEEE*, pp. 1610–1613.
49. Ma S, Xue L, Zou H (2013) Alternating direction methods for latent variable gaussian graphical model selection. *Neural computation* : 1–27.

## Figure Legends

**Figure 1. Illustration of regularized estimation of partial correlations.** **A.** The sample correlation matrix of unprocessed somatic calcium signals from a population of cells in mouse visual cortex. The outlined square fragment is magnified in **B**. **C.** The same fragment of another estimate of the correlation matrix regularized to yield sparse partial correlations. Corresponding fragments of partial correlations matrices of the unregularized and regularized estimated are shown in **D** and **E**, respectively.

**Figure 2. Estimators whose low-dimensional regularization targets can represent the structure of the true covariance matrix outperform other estimators.** **Row 1.** Graphical representations of four types of low-dimensional structures of interactions between observed neurons (green spheres) and latent units (light-shaded spheres). In the ‘independent model’ (**A**), observed neurons exert no linear effects on one another neither directly nor through interactions with common latent units. In ‘latent factors’ (**B**), the correlated activity of observed cells is driven by several latent units. In ‘sparse interactions’ (**C**), the correlation matrix is defined by a set of linear interactions between observed neurons. In ‘sparse+latent’ (**D**), correlations arise through direct linear interactions between some pairs of observed neurons and through interactions with common latent units. **Row 2.** Examples of  $50 \times 50$  correlation matrices corresponding to each type of low-dimensional structure. The factor model (**B**) has three latent units. The partial correlation matrix of the sparse model (**C**) is 73% sparse. The ‘sparse+latent’ matrix has one latent unit and its direct interactions are 78% sparse. **Row 3.** Examples of sample correlation matrices calculated from samples of 1000 observations taken from simulated random processes with corresponding correlation matrices from row 2. **Row 4.** Average normalized Gaussian log-likelihood *excess loss* (Eq. 9) attained by each of the five estimators as a function of sample size. The error bars indicate the standard error of the mean based on 30 samples. **Row 5.** Average normalized Gaussian log-likelihood *validation loss* (Eq. 10) attained by each of the five estimators. The values are relative to the validation loss of the estimator that matches the low-dimensional structure of the true covariance matrix. The error bars indicate the standard error of the mean based on 30 samples.

**Figure 3. Acquisition of neural signals for the estimation of noise correlations.** Visual stimuli comprising full-field drifting gratings interleaved with blank screens (**A**) were presented to anesthetized mice while two-photon recordings of somatic calcium signals were collected using fast 3D random-access microscopy (**B**). The visual stimulus included an initial period with 16 directions of motion for orientation tuning followed by a longer (15–20 min) period of stimulation with only 2 or 5 directions of motion for the computation of the noise correlation matrix. **C.** Representative calcium signals from eight cells out of 298 cells downsampled to 20 Hz. The inferred firing rate binned in 150 ms intervals are indicated by red ticks below each trace. **D.** The raster plot of the inferred firing rates, binned in 150 ms intervals, from the entire population from the first (left) and last (right) minute of the entire recording. The traces from **C** are highlighted in red. **E.** The spatial arrangement and orientation tuning of the 298 cells from the imaged site. **F.** The noise correlation matrix of the activity of the neural population. **G.** The histogram of the noise correlation coefficients with the mean indicated by the red line.

**Figure 4. The sparse+lowrank estimator  $C_{\text{sparse+lowrank}}$  outperforms the other estimators on neural data.** **A–D.** Histograms of average cross-validation loss differences of the respective estimators  $C_0$ ,  $C_{\text{diag}}$ ,  $C_{\text{factor}}$ , and  $C_{\text{sparse}}$  from  $C_{\text{sparse+lowrank}}$ . The histograms are based on 31 imaged sites in 24 mice. All medians (red dashed lines) were significantly greater than zero, indicating the dominance of  $C_{\text{sparse+lowrank}}$  over the other estimators. The green arrows indicate the results for the site shown in Fig. 3 and Fig. 5

**Figure 5. Example of low-dimensional correlation structure revealed by the sparse+low-rank estimator.** **A.** The regularized estimate of the correlation matrix (top-right) closely approximates the sample correlation matrix (bottom left). This close approximation is also demonstrated by the scatter plot of the correlation coefficients produced by the two estimates (**D**). However, the partial correlation matrices from the two estimates show more pronounced differences (**B** and **E**). **C.** Furthermore, the partial correlation matrix of the regularized estimate is decomposed into a sparse component with 82.2% off-diagonal zeros (bottom-left) and low-rank component of rank 15 (top-right). **F.** The sparse component of the regularized partial correlation matrix had little resemblance to the sample correlations. The gray interval indicates the range of correlations containing 82.2% of cells pairs, equal to the fraction of zeros in the sparse partial correlation matrix. This interval contained 58.9% of the partial correlations. **G.** The graphical depiction of the positive (green) and negative (magenta) partial correlations as edges between observed neurons. The line density is proportional to the magnitude of the correlation. **H.** A subset of neurons from the center of the cluster shown in **G** showing the regularized partial correlations. **I.** The same subset with sample correlations thresholded to match the sparsity of the regularized interactions.

**Figure 6. Properties of sparse+low-rank regularized estimates from all imaged sites A.** The average sample correlations vs. average partial correlations for each imaged site. In each plot, the red asterisk indicates the site shown in figures 3 and 5. **B.** The average node degree for sparse partial correlations vs. population size in each imaged site. **C.** The number of inferred latent units vs. population size in each imaged site. **D.** The number latent units vs. average node degree for sparse partial correlations for each site.

**Figure 7. Dependence of correlations and partial correlations on orientation tuning differences and physical distance between cell pairs.** **A–C.** Average sample correlations (black) and regularized partial correlations (red) between pairs of cells in the example site shown in previous figures. The correlations were normalized by the respective average correlations (6 A). **A.** Average correlations between pairs of neurons tuned to orientation with differences in preferred orientation in the intervals of 0–15°, 15–45° and 45–90°. **B.** Average correlations between pairs of neurons located at the same depth ( $\pm 25\mu\text{m}$ ) separated by lateral distances in the intervals of 0–25  $\mu\text{m}$ , 25–75  $\mu\text{m}$ , 75–150  $\mu\text{m}$ , and 150+  $\mu\text{m}$ . **C.** Average correlations between pairs of neurons displaced laterally by less than 25  $\mu\text{m}$  separated in depth by distances in the intervals of 0–25  $\mu\text{m}$ , 25–60  $\mu\text{m}$ , and 60+  $\mu\text{m}$ . **D–F.** Same measurements as **A–C** averaged across multiple sites. Only sites that had at least 20 qualifying cell pairs in each of the intervals were included in the averages. The error bars indicate the standard error of the mean (often too small to be seen). **G–I.** Normalized connectivity of positive (green) and negative (dark red) interactions from the sparse component of the regularized partial correlations in the example site shown in previous figures. Normalized connectivity was computed as the fraction of pairs connected by interactions of corresponding signs in each interval divided by fraction of non-zero interactions across the entire site. The intervals are identical to those in **A–C**. **J–L.** Same measurements as in **G–I** averaged across multiple sites. The error bars indicate the error of the mean. Only sites that had at least 20 qualifying pairs in each interval were included in the averages.