OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Spike Triggered Covariance in Strongly Correlated Gaussian Stimuli

Johnatan Aljadeff[1,2], Ronen Segev[3], Michael J. Berry II[4], Tatyana O. Sharpee[1,2]*

1 Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California, United States of America, **2** Center for Theoretical Biological Physics and Department of Physics, University of California, San Diego, La Jolla, California, United States of America, **3** Department of Life Sciences and The Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel, **4** Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, United States of America

Abstract

Many biological systems perform computations on inputs that have very large dimensionality. Determining the relevant input combinations for a particular computation is often key to understanding its function. A common way to find the relevant input dimensions is to examine the difference in variance between the input distribution and the distribution of inputs associated with certain outputs. In systems neuroscience, the corresponding method is known as spike-triggered covariance (STC). This method has been highly successful in characterizing relevant input dimensions for neurons in a variety of sensory systems. So far, most studies used the STC method with weakly correlated Gaussian inputs. However, it is also important to use this method with inputs that have long range correlations typical of the natural sensory environment. In such cases, the stimulus covariance matrix has one (or more) outstanding eigenvalues that cannot be easily equalized because of sampling variability. Such outstanding modes interfere with analyses of statistical significance of candidate input dimensions that modulate neuronal outputs. In many cases, these modes obscure the significant dimensions. We show that the sensitivity of the STC method in the regime of strongly correlated inputs can be improved by an order of magnitude or more. This can be done by evaluating the significance of dimensions in the subspace orthogonal to the outstanding mode(s). Analyzing the responses of retinal ganglion cells probed with $1/f$ Gaussian noise, we find that taking into account outstanding modes is crucial for recovering relevant input dimensions for these neurons.

Citation: Aljadeff J, Segev R, Berry MJ II, Sharpee TO (2013) Spike Triggered Covariance in Strongly Correlated Gaussian Stimuli. PLoS Comput Biol 9(9): e1003206. doi:10.1371/journal.pcbi.1003206

Editor: Olaf Sporns, Indiana University, United States of America

Received December 28, 2012; **Accepted** July 17, 2013; **Published** September 5, 2013

Funding: This work was supported by NIH grants EY019493, EY014196 and EY017934, National Science Foundation (NSF) grant IIS-0712852, the McKnight Scholarship, the Keck and Ray Thomas Edwards Foundations, and the Center for Theoretical Biological Physics (NSF PHY-0822283). This work was also supported in part by the National Science Foundation under Grant No. PHYS-1066293 and the hospitality of the Aspen Center for Physics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sharpee@salk.edu

Introduction

How do neurons encode sensory stimuli? One of the primary difficulties in answering this long-standing problem is the fact that sensory stimuli have high dimensionality. For example, responses of many visual neurons are affected by image patterns that require at least a 10×10 pixel grid for their description as well as a temporal history spanning multiple time bins or basis functions. Determining what input combinations affect the neural responses is a key step in characterizing the neural computation. Broadly speaking, to detect the presence of certain features in the environment over a range of distances and light conditions, one needs to disambiguate the presence of this feature at a weak contrast from the presence of a similar, but different feature presented at a higher contrast. This can only be achieved with nonlinear functions that depend on multiple input components, such as the presence of an edge of correct orientation and the absence of the edge orthogonal to it [1]. In support of these arguments, the responses of neurons in different sensory modalities are found to be sensitive to multiple input combinations. Examples include vision [2–7], audition [8–10], olfaction [11], somatosensation [12] and mechanosensation [13]. Neurons respond with all-or-none responses termed spikes. The goal of different methods for characterizing neural feature selectivity is to determine how the probability of eliciting a spike from a neuron depends on its inputs. The underlying assumption is that this dependence of spike probability on input parameters will have a low-dimensional structure. Finding either the linear input dimensions that modulate the spike probability (we will refer to these dimensions as relevant) or quadratic forms of inputs [14–16] is the focus of much of the current research in the field.

Much of the analysis of neural selectivity for multiple input combinations has been carried out using uncorrelated (“white noise”) or weakly correlated inputs. With such inputs, the relevant input dimensions can be found using a computationally inexpensive method known as spike-triggered covariance (STC) [6,7,17–22]. The STC method works by comparing the change in variance along different dimensions in the input space across all stimuli and across stimuli that elicited a spike. The dimensions along which the variance is found to be significantly different represent the relevant input dimensions for the response of a particular neuron. The method is not limited to strictly Gaussian inputs provided that the inputs are still circularly symmetric [23], which is another example of an input distribution without correlations.

Author Summary

In many areas of computational biology, including the analyses of genetic mutations, protein stability and neural coding, as well as in economics, one of the most basic and important steps of data analysis is to find the relevant input dimensions for a particular task. In neural coding problems, the spike-triggered covariance (STC) method identifies relevant input dimensions by comparing the variance of the input distribution along different dimensions to the variance of inputs that elicited a neural response. While in theory the method can be applied to Gaussian stimuli with or without correlations, it has so far been used in studies with only weakly correlated stimuli. Here we show that to use STC with strongly correlated, $1/f$-type inputs, one has to take into account that the covariance matrix of random samples from this distribution has a complex structure, with one or more outstanding modes. We use simulations on model neurons as well as an analysis of the responses of retinal neurons to demonstrate that taking the presence of these outstanding modes into account improves the sensitivity of the STC method by more than an order of magnitude.

In principle the STC method can also be used with correlated Gaussian stimuli [7,20]. The case of correlated stimuli - especially with strong correlations, where the second moment of the covariance spectrum may be infinite - is important for neural coding. This is because signals in the sensory environment possess such correlations in both the second and higher orders [24–30]. Because the properties of a cell's relevant subspace may change depending on the stimulus statistics as a result of adaptation [31,32], it may not be sufficient to study neural coding using uncorrelated stimuli. Here we show that with strongly correlated inputs, the significance analysis for determining which of the dimensions obtained by the STC method are relevant for neural spiking needs to be modified to take into account a rather complicated covariance structure of randomly selected inputs drawn from such input ensembles. The nonuniform covariance structure, which has properties akin to the graph laplacian in small-world networks [33], breaks the symmetry in the input space, and therefore may obscure many significant dimensions.

The most prominent aspect of the natural scenes covariance structure is the presence of the so-called "coherent" mode [34]. This stimulus dimension approximately corresponds to the zero frequency input component and has a corresponding eigenvalue that is at least 10 times larger than the mean eigenvalue of the input covariance matrix. Even in datasets of fairly large size, the extremely large variance along the coherent mode obscures many of the truly relevant dimensions for neural spiking (Fig. 1). These effects are also reproduced in our analysis of the responses of ganglion cells from the salamander retina probed with $1/f$-type naturalistic Gaussian stimuli. We identify a close relationship between the covariance structure derived from natural scenes to that defined by the Spiked-Wishart matrix model [35,36]. This allows us to explain the effects in the context of the STC method using results from random matrix theory, and suggest ways to bypass sampling variability along the outstanding modes.

Results

Spike triggered covariance

Mathematically, the first step in the STC method is to compute the covariance matrix of stimuli that lead to a spike C^{spike} and the covariance matrix of all stimuli C^{stim}:

$$C_{ij}^{\mathrm{spike}} = \frac{1}{N_{\mathrm{spike}}-1}\sum_{t=1}^{N_{spike}}\left(\hat{s}_i^t - \langle s_i\rangle_{spike}\right)\left(\hat{s}_j^t - \langle s_j\rangle_{spike}\right) \quad (1)$$

$$C_{ij}^{\mathrm{stim}} = \frac{1}{N-1}\sum_{t=1}^{N}\left(s_i^t - \bar{s}_i\right)\left(s_j^t - \bar{s}_j\right). \quad (2)$$

Here, N_{spike} is the number of recorded spikes, N is the number of stimulus frames, s_i^t is the value of the stimulus along the ith dimension at time t, the *hat* denotes that this stimulus triggered a spike, the *bar* denotes the average across the input distribution and $\langle s_j\rangle_{spike}$ is the average across the distribution of inputs that triggered a spike (the so called "spike-triggered-average").

As the second step, one computes the difference between these covariance matrices:

$$\Delta C = C^{\mathrm{spike}} - C^{\mathrm{stim}}, \quad (3)$$

and finds the eigenvalues that are significantly different from zero. The corresponding eigenvectors span the neuron's relevant subspace.

To determine statistical significance of the eigenvalues, they need to be compared to the null distribution, which is the distribution of eigenvalues of the matrices $\Delta C^{\mathrm{null}} = C^{\mathrm{null}} - C^{\mathrm{stim}}$. The matrices $\Delta C^{\mathrm{null}}, C^{\mathrm{null}}$ are formed assuming no association between the stimulus and the neural response, i.e. by using random spike times chosen at the same rate found for real neurons. If the spike train has particular temporal structure (e.g. bursting, a refractory period), the C^{null} is obtained by random shifts of the spike train with periodic boundary conditions [20]. Significant eigenvalues of ΔC can be positive or negative. The procedures for determining statistical significance are detailed in Materials and Methods.

The final step of the STC method is to remove stimulus correlations from the estimate of dimensions found to be significant. This can be done by multiplying them with the (pseudo)inverse of C^{stim} (see Materials and Methods). The method which we use to find the optimal rank of the pseudoinverse matrix is detailed in [22,37] and for completeness described in Materials and Methods.

We note that this approach, Eqs. (1)–(3), of finding the relevant stimulus dimensions by diagonalizing ΔC is equivalent to seeking eigenvectors of the following matrix [20]:

$$\Delta\tilde{C}_{ij} = \frac{1}{N_{\mathrm{spike}}-1}\sum_{t=1}^{N_{spike}}\left(\hat{s}_i^t - \bar{s}_i\right)\left(\hat{s}_j^t - \bar{s}_j\right) - \frac{1}{N-1}\sum_{t=1}^{N}\left(s_i^t - \bar{s}_i\right)\left(s_j^t - \bar{s}_j\right). \quad (4)$$

This matrix describes a change in the second moment between the distributions stimuli that elicit a spike and that of all stimuli, after subtracting the mean stimulus $\bar{s}$. Despite the fact that $\Delta C \neq \Delta\tilde{C}$, their eigenvectors coincide.

In another formulation, instead of subtracting the matrix C^{stim} in Eq. (3), the stimulus is decorrelated ("whitened") prior to its spike triggered characterization [7]. For completeness, the details of this method are brought in Materials and Methods. Throughout

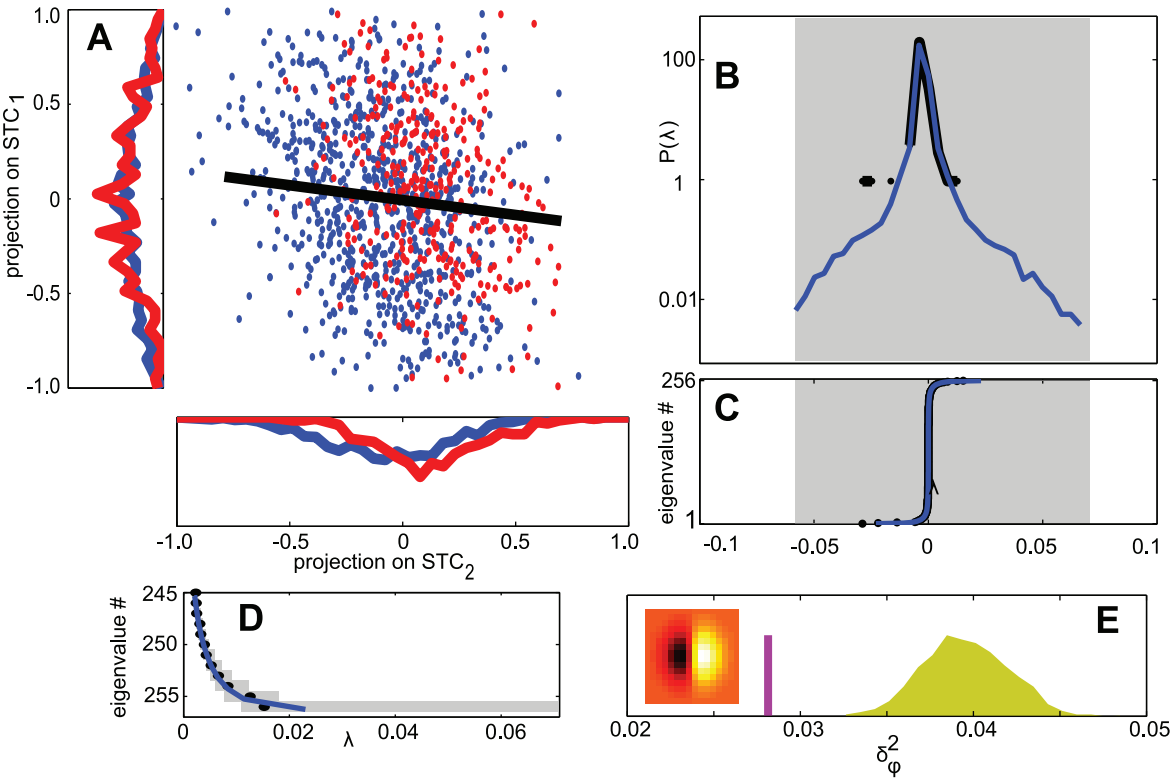

Figure 1. Spike triggered covariance analysis of simulated spike trains of a model with a single feature orthogonal to the coherent mode. (A) Inputs are plotted when projected on the eigenvectors corresponding to the largest eigenvalues of ΔC (in units of pixel illumination). Marginal distributions are plotted for each dimension. Inputs that elicited a spike are shown in red, and those that did not in blue. By construction, the change in variance is larger along the first dimension. (B) The empirical eigenvalue distribution of ΔC (black) compared to the null distribution (blue). No eigenvalues of ΔC are found to be significant (shaded area indicates 98% confidence intervals for the support of the null distribution) (C) Rank ordered eigenvalues (black) plotted with the null distribution (blue). (D) Nested rank-wise significance testing. The highest ranked eigenvalues of ΔC are within the 98% confidence intervals derived from the null distribution constructed for each rank separately (see Materials and Methods for details). (E) For each random spike train we computed δ_ϕ^2, the variance of the projection of the spike-triggered stimulus on the relevant feature ϕ (distribution shown in yellow). The purple line indicates δ_ϕ^2 for the real spike train, suggesting the spike train contains enough signal to determine the relevant feature as significant. Inset shows the relevant feature, a 16×16 image patch ($p=256$). Simulation details: $N=1189$, $N_{\text{spike}}=329$, 250 repetitions to find ΔC^{null}.
doi:10.1371/journal.pcbi.1003206.g001

the manuscript, we will refer to this method as the "one centered" method, because the null distribution is centered around the identity matrix, rather than a matrix of zeros, as in Eq. (3). Correspondingly, we will refer to the version of the STC method obtained by diagonalizing Eq. (3) as the "zero-centered" method. In essence, both the one-centered and the zero-centered versions are similarly affected by inhomogeneous sampling variability.

The authors of [7] proposed a slightly different definition of the null distribution and a nested hypothesis technique for significance testing. For the model cell simulations we used both significance analysis methods, in both the "zero-centered" and "one-centered" STC formulations, and obtained similar results. For the rest of this paper we will refer to our significance testing method as the "global" one, and focus mainly on the "zero-centered" formulation of the STC method. Using this combination the important effects of the strong stimulus correlations on the analysis are more easily understood.

Model cells presented with strongly correlated noise

We begin with an illustration of the problems that arise when the STC method is used to analyze neural responses to strongly correlated Gaussian noise (Fig. 1). We simulated a model neuron where the neuronal responses were modulated by stimulus projections onto a single dimension (termed here the relevant feature). The stimuli were constructed to match the second-order statistics from the set of images in the van Hateren dataset [38] (see Materials and Methods). In this example obtained for dataset of a moderate size, no eigenvalues fell outside of the 98% confidence intervals (1% significant bounds for the largest and smallest rank-ordered eigenvalues). Yet, the spike train contains enough signal about the cell's input-output function to identify the relevant feature for this level of significance. Specifically, the variance along the relevant dimension in the spike-triggered stimulus ($\delta_\phi^2=\text{var}(\hat{s}\cdot\phi)$) is much smaller than can be explained by random spike times (Fig. 1E).

Outstanding modes in covariance matrices derived from natural stimuli

To understand the origin of such masking of the relevant feature(s), we consider the eigenstructure of covariance matrices for stimulus ensembles with strong pairwise correlations. For example, in the case of natural scenes that exhibit long range correlations over a very wide range of spatial scales [27,39], principal component analysis (PCA) yields one outstanding

eigenvalue (for example, see eigenvalue marked λ_1 in Fig. 2A). The corresponding eigenvector has all positive components [28,30] and is often referred to as the "coherent mode" [34]. To understand why such a coherent mode appears, one can consider the case where the correlations decrease only slightly over the range of image patches used to compute the covariance matrix. In this case, the correlation values in different image patches will be approximately the same. Such a matrix will have one outstanding eigenvalue with a corresponding eigenvector that has equal weights for all stimulus dimensions [40]. Small differences in the amount of covariation for pixel pairs with different spatial separation will lead to deviations in components of the coherent mode from each other, but the basic structure will remain the same as long as the mean of the correlation values exceeds the standard deviation of their fluctuations [40]. In fact, shuffling entries in the sample covariance matrices of natural stimuli yields matrices whose spectra follow the analytical predictions exactly [40,41]. These analytical predictions generalize the Wigner semicircle law [42] for matrices whose elements have a non-zero mean:

$$g_{\mathrm{TKW}}(\lambda)=\frac{\sqrt{4\sigma^2-\lambda^2}}{2\pi\sigma^2}+\begin{cases}0 & |\mu|<\sigma\\ \frac{1}{N}\delta\left[\lambda-\left(\mu+\frac{\sigma^2}{\mu}\right)\right] & |\mu|>\sigma\end{cases}, \quad (5)$$

where μ and σ^2 are the mean and variance of matrix elements. The distribution $g_{\mathrm{TKW}}(\lambda)$ follows the semicircle law with the addition of one outstanding mode that appears once the mean of matrix elements exceeds their standard deviation. The eigenvector corresponding to the outstanding eigenvalue is $\frac{1}{\sqrt{p}}(1,\ldots,1)$. The semicircle law appears because matrices are no longer positive-definite after shuffling. However, the outstanding eigenvalue is located at exactly the same value as the outstanding eigenvalue of the natural scenes covariance matrix C^{pop} (see Fig. 2C).

In our analysis of the van Hateren database, the largest eigenvalue tends to be at least $3-4$ times larger than the second largest eigenvalue. This shows how strong the coherent mode is compared to other modes. The principal components ranked below the coherent mode form a collection of orthogonal "edge detectors", some of which correspond to an eigenvalue still much larger than the mean eigenvalue of C^{stim}, a signature of the stimulus' heavy-tailed covariance spectrum. Such large disparities in variance along the different dimensions in the stimulus space make it problematic to directly compare changes in variance induced by the observation of spikes along these different dimensions.

The detailed structure of sampling variability in the estimation of eigenvectors and eigenvalues can be understood in terms of the Spiked Wishart ensemble [35,36]. In the Spiked Wishart matrix model, the true (population) covariance eigenvalues are all equal to one, except for a small number $v \ll p$ of outstanding modes with eigenvalues larger than one $(1+\tau_1,1+\tau_2,\ldots,1+\tau_v,1,\ldots,1)$, where p is the stimulus dimensionality. The distribution of sample covariance eigenvalues $g_{sw}(\lambda)$ for a finite number of inputs has a positive bias, with the following analytical expressions [36]:

$$g_{SW}(\lambda)\propto(p-v)g_{MP}(\lambda)+\sum_{i=1}^{v}\rho_i \quad (6)$$

$$\rho_i\sim\mathcal{N}\left[(1+\tau_i)\left(1+\frac{1}{\gamma\tau_i}\right),\frac{\hat{\sigma}_i^2}{N}\right] \quad (7)$$

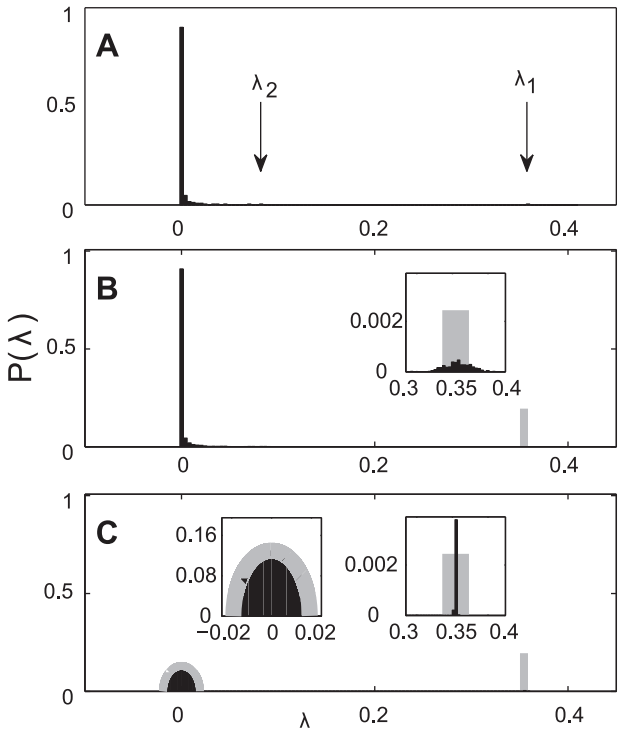

Figure 2. Spectra of the population covariance, sample covariance, and symmetric random matrices with matched element distribution. (A) Eigenvalue distribution of an example population covariance matrix C^{pop} ($p=256$) computed from the van Hateren data set. The largest eigenvalue (marked with an arrow) corresponds to an eigenvector with only positive components, and is ≈ 10 times larger than the second largest eigenvalue (also marked) and ≈ 100 times larger than the mean eigenvalue. (B) Eigenvalue distribution of a collection of sample covariance matrices computed from stimuli randomly drawn from a multivariate Gaussian distribution $\mathcal{N}(0,C^{\mathrm{pop}})$. In gray is the analytic prediction for the outstanding eigenvalue. The spread of these eigenvalues (black, inset) is in agreement with the prediction in Eq. (8). (C) Eigenvalue distribution of symmetric random matrices with elements randomly drawn from a distribution given by the elements in the sample covariance matrices. In gray is the complete prediction (semicircle and outstanding eigenvalue) given by Eq. (5). Diagonal and off-diagonal elements are drawn separately from the distribution of matrix elements in panel B.
doi:10.1371/journal.pcbi.1003206.g002

$$\hat{\sigma}_i^2=2(1+\tau_i)^2\left(1-\frac{1}{\gamma\tau_i^2}\right) \quad (8)$$

where $\gamma\equiv N/p$ and N is the number of samples. The distribution g_{MP} representing the "bulk" of eigenvalues is the so called Marčenko-Pastur distribution given by:

$$g_{MP}(\lambda)=\frac{\sqrt{(\lambda_+-\lambda)(\lambda-\lambda_-)}}{2\pi\lambda/\gamma} \quad (9)$$

$$\lambda_\pm=\left(1\pm\frac{1}{\sqrt{\gamma}}\right)^2. \quad (10)$$

This distribution corresponds to the sample covariance eigenvalues obtained when the true covariance is the identity matrix. Using

numerical simulations we verified that, although the Spiked Wishart ensemble is only an approximation to the covariance matrices derived from natural stimuli, Eqs. (7) and (8) accurately describe the scaling of the variance and the mean of sample eigenvalues as N increases.

In addition to biases in eigenvalue estimates, there are also biases in the estimation of eigenvectors. The dot product between the true (population) ith eigenvector $\hat{f}^{(i)}$ and the ith eigenvector $f^{(i)}$ of the sample covariance approaches

$$\left|f^{(i)}\cdot\hat{f}^{(i)}\right| \xrightarrow[N\to\infty]{} \sqrt{\frac{1-1/(\gamma\tau_i^2)}{1+1/(\gamma\tau_i)}}. \tag{11}$$

In other words, the "mixing" of the outstanding sample eigenvectors seen in Eq. (11) (note the dependence of this mixing on N through γ) as well as the variance and bias in the sample eigenvalues seen in Eq. (7) means that whitening cannot be exact.

In the context of the spike triggered covariance the consequences of such properties of the distribution of sample eigenvalues are twofold. First, Eq. (8) indicates that the variance of the outstanding eigenvalues around their mean increases with the square of their value and is inversely proportional to the number of samples. Thus, for sample sizes that are not much larger than the stimulus dimensionality ($1<\gamma_{\text{spike}}\lesssim 10$ in the simulation results presented in Fig. 3A), the increased variance of the outstanding sample eigenvalue means that C^{null} and C^{stim} will not cancel each other exactly along that vector. Second, the mean estimate contains a positive bias relative to the population values, cf. Eq. (7). The combination of these two effects widens the null-distribution used to test the significance of the resulting eigenvectors, effectively masking features that should otherwise be identified as being relevant.

Pre-whitening

One way to compensate for the symmetry breaking effects caused by strong correlations in the input space is to equalize variances before applying the STC method. This is the essence of the "one-centered" formulation of the STC method [7]. In principle, this "whitening" should work with Gaussian stimuli with any covariance structure. However, as discussed above, in the case of strongly correlated stimuli, the estimation of eigenvalues (i.e variances along different dimensions in the input space) possesses strong variability, cf. Eq. (7). As a consequence, normalization by a variance estimated from one part of the dataset does not fully remove correlations in a different subset of the data. With increasing dataset size, the estimate of the variance along the coherent mode improves. However, because the absolute value of variance is not relevant in the pre-whitening method, dimensions with smaller variance can cause just as much contamination as the coherent mode. In addition, the estimation of variance along dimensions corresponding to τ_i just larger than $1/\sqrt{\gamma}$ remains poor for large N. If $\tau_i=1/\sqrt{\gamma}+\varepsilon$ the sample eigenvalue estimation error diverges as $\sqrt{\gamma}\varepsilon$, as follows from Eq. (8). In other words, as the number of samples N and γ increase, the bulk of the distribution narrows, and new eigenvalues separate from the bulk. It is these eigenvalues with intermediate values that are poorly determined and make it problematic to equalize variance along different dimensions. Another signature of this phenomenon is that $f^{(i)}\cdot\hat{f}^{(i)}\approx 0$ for these dimensions, as follows from Eq. (11). Thus, these dimensions are poorly estimated from the sample covariance and, as a consequence, the variance along one stimulus dimension in the training set will be inappropriately used to normalize variance along a different stimulus dimension in the test set. Altogether, we observed that pre-whitening stimuli did not improve the estimation of relevant stimulus features compared to the zero-centered method, compare panels A and B in Fig. 3. Intuitively, in the zero-centered method the dimensions with the largest variance provide the largest uncertainty in variance estimation, whereas in the one-centered version the problematic dimensions change depending on the dataset size, and are not easily identified *a priori*.

We have also explored the possibility of using a pseudoinverse of the covariance matrix instead of the full inverse to normalize variance along different dimensions (see Materials and Methods for details). When using the pseudoinverse (instead of the inverse), stimulus dimensions with small variance in the stimulus ensemble are removed to avoid noise amplification along these dimensions (see Materials and Methods for details). However, an immediate consequence of choosing a small pseudoinverse order k ($\approx p/10$) is that the stimulus dimensionality is reduced to k. This implies that the effective γ of the problem is now $\gamma_k=N/k$, i.e. p/k times larger than γ. This could work well in some cases as illustrated in Fig. 3I. Here, in simulations based on a small number of spikes, the use of pseudoinverse can help recover one or two significant features while the standard zero-centered method fails to find any. However, the use of pseudoinverse only helps within a very narrow band of small pseudoinverse orders. This band may be difficult to determine when analyzing real neural data. In addition, this procedure limits the reconstruction to a linear combination of only a few leading stimulus dimensions. In many cases, the relevant features do include components along stimulus dimensions with smaller variance, and in those cases, the effective increase in γ will not improve the performance of the STC method. Indeed, one observes that in cases where two significant dimensions are obtained by using substantial reduction in dimensionality of the pseudoinverse, the resulting dimensions have the subspace projection onto the model features of ≈ 0.6 whereas this value is ≈ 0.8 when using the full inverse and a larger number of spikes to obtain for a comparable effective γ (Fig. 3I).

Finally, in the regime where $k\approx p$ (i.e. "almost full" inverse), the prewhitening approach works just as well as the "zero-centered" formulation, and a relatively high value of the signal-to-noise ratio parameter γ_{spike} is required for recovery of the full relevant subspace.

Correction scheme

As another way to compensate for the symmetry breaking effects caused by strong correlations in the input space, we propose to modify the "zero-centered" formulation of the STC method in the following way. Because the largest drop in variance is between the coherent mode and other dimensions, we propose here to test the significance of changes in variance separately along the coherent mode and in the subspace orthogonal to it. Explicitly, to do the analysis in the $p-1$ dimensional subspace, the coherent mode $f^{(1)}$ is projected out of all stimuli. If s is a stimulus vector and $f^{(1)}$ normalized to length 1, one can perform the STC analysis using s instead of s where:

$$\mathrm{s}_i = s_i - f_i^{(1)}\sum_{j=1}^{p} f_j^{(1)} s_j. \tag{12}$$

In this approach the correct number of relevant dimensions is determined by evaluating significance in the subspace orthogonal

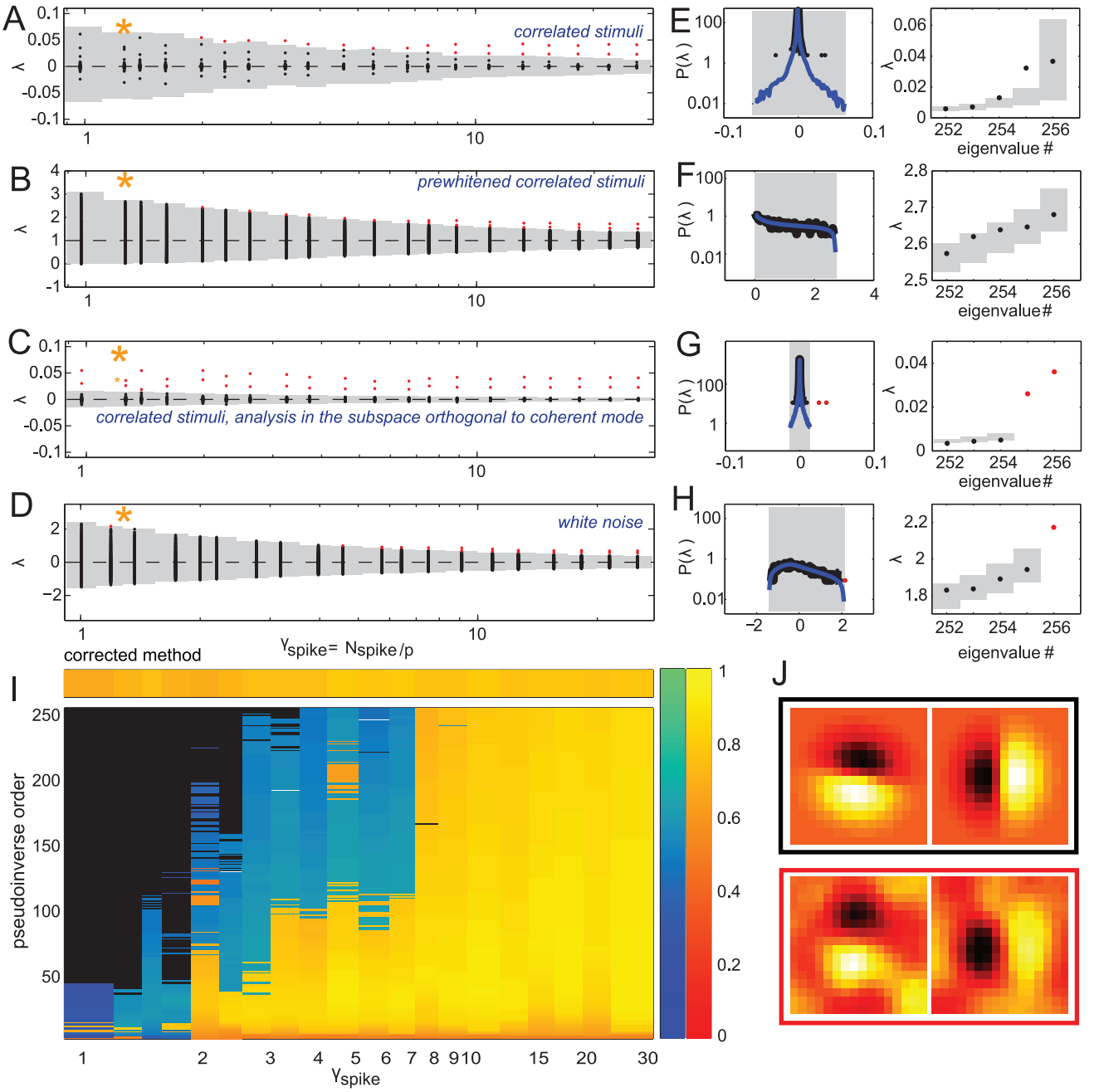

Figure 3. Spike triggered covariance analysis of a model neuron with two relevant features orthogonal to the coherent mode. Spectra of ΔC for increasing dataset size in the case of strongly correlated Gaussian noise (A–C) and white noise (D). (A–D) The range of spikes covered is from $N_{\mathrm{spike}} \approx p$ to $N_{\mathrm{spike}} \approx 22p$. Panel B shows the results using the pre-whitening ("one-centered") method, and panel C shows the results after evaluating significance in the subspace orthogonal to the coherent mode. Each vertical line shows the result of a single simulation. Significant (insignificant) eigenvalues are shown in red (black), and the range of the null distribution (1000 evaluations of ΔC^{null}, 98% confidence interval) is shown in gray. (E–H, left) Gray shaded area is the support of the null distribution, which itself is plotted in blue. The significant (insignificant) portion of the spectrum of ΔC is plotted in red (black). These example spectra, with the corresponding significant vectors, are for conditions with a small number of spikes (indicated by an orange star in A–D) for which both of the formulations find no significant dimensions. $N = 1189$, $N_{\mathrm{spike}} = 300$ for the correlated stimulus condition in panels E–G, $N = 1189$, $N_{\mathrm{spike}} = 280$ for the white stimulus condition in panel H. (E–H, right) Results of the nested significance testing. We note that the second ranked eigenvalue in panel E is outside of its confidence interval, but still cannot be found to be significant. This happens because of the noise along the coherent mode. (I) STC analysis using all pseudoinverse orders using the nested significance testing with 99% confidence intervals (large box) compared to the analysis using our proposed correction scheme. Black means no significant features were found for that combination of γ_{spike} and pseudoinverse order. Cold (hot) colors indicate that one (two) features were found to be significant. The corresponding color bars on the right indicate the geometric average of the feature projections on the two model dimensions (cold colors) or the subspace overlap with the model, cf. Eq. (36) (hot colors). (J) Results when STC is performed using the proposed correction scheme. The two relevant dimensions (black frame) and the decorrelated significant features (red frame) have subspace overlap of 0.82. The models were defined such that the mean firing rate remained unchanged between the two stimulus conditions.
doi:10.1371/journal.pcbi.1003206.g003

to the coherent mode and then adding back their projections on the coherent mode from the corresponding eigenvectors evaluated in the full input space (see below).

We find that considering the coherent mode separately from the rest of stimulus dimensions reduces the value of γ_{spike} for which the full relevant subspace is found to be significant by a factor of 10 (Fig. 3C). This improvement can be approximated from Eqs. (7) and (8). Assuming the cell's relevant subspace is exactly orthogonal to the coherent mode, the extremal values of the null distribution are distributed as $\Omega_\lambda(\gamma,\gamma_{\text{spike}},\tau_1)$. The variance of Ω_λ is:

$$\langle \Omega_\lambda^2 - \langle \Omega_\lambda \rangle^2 \rangle \approx 2(1+\tau_1)^2\left(\frac{1}{N}+\frac{1}{N_{\text{spike}}}\right). \tag{13}$$

This implies that the number of stimuli N,N_{spike} sufficient for identifying the relevant features as significant increases with τ_i as:

$$N,N_{\text{spike}} \propto (1+\tau_1)^2. \tag{14}$$

Upon removal of the coherent mode, the minimum value of N for which the signal to noise ratio will be high enough to identify the relevant dimensions scales as $(1+\tau_2)^2$ corresponding to the stimulus' second principal component. Therefore the improvement is proportional to $\left(\frac{1+\tau_1}{1+\tau_2}\right)^2 \approx \left(\frac{\tau_1}{\tau_2}\right)^2$. In our simulations (Fig. 3A,C) this corresponds to a predicted 17.8 fold improvement. Given that our model features were not exactly orthogonal to the coherent mode, and that the spectrum obtained from the van Hateren dataset has a heavy tail and does not conform exactly to the Spiked Wishart ensemble, an approximate 10 fold improvement represents a good agreement with the prediction.

It is noteworthy that the minimum requirement on the dataset size for obtaining the correct number of relevant dimensions is actually smaller with correlated stimuli than it is for white noise stimuli for the same neuron (compare Fig. 3 panels A–D) when the model parameters were matched such that the firing rate remains constant across different stimuli statistics. Another important point is that considering the coherent mode separately is different from simply discarding a "DC-like" component that could be found to be significant by the STC. This is because when N is small, no dimensions are found to be significant with the coherent mode as part of the stimulus ensemble (Fig. 1).

An important consideration is that the final analysis can include the components of the relevant dimensions onto the coherent mode. This is possible for two reasons. First, the coherent mode does not represent an arbitrary dimension in the input space but is one of the eigenvectors of the sample covariance matrix. Second, the significant eigenvectors of ΔC have a form $\sum_{n=1}^{p} \lambda_n f_i^{(n)} f_j^{(n)} \phi_j$, where λ_n is the nth eigenvalue corresponding to the nth eigenvector $f^{(n)}$ of the sample covariance matrix, and ϕ describes one of the relevant features [20]. Because of these two properties, eigenvectors evaluated in the full input space and in the subspace orthogonal to the coherent mode differ only in their components along the coherent mode (see Materials and Methods for the details of the derivation). This makes it possible to analyze cells with features that have nonzero components along the coherent mode.

We have verified that this approach also works in a large number of cases where the relevant stimulus dimensions have a large projection on the coherent mode (Fig. 4). One concern is that when such neurons are probed with a relatively small number of stimuli, then projecting the coherent mode out may "push" the relevant feature into the null eigenvalue distribution. This does not appear to be a problem in our simulations for $\gamma_{\text{spike}} > 1.3$ (Fig. 4B). If this does happen, the relevant subspace should be the one spanned by both the eigenvectors found to be significant in the full stimulus space and those found to be significant in the subspace orthogonal to the coherent mode.

Application to retinal ganglion cells

We now demonstrate the importance of this correction scheme by analyzing recordings of 22 salamander retinal ganglion cells (RGCs). These neurons were probed with a correlated noise stimulus whose covariance matrix was matched to that of natural visual stimuli. Without correcting for the presence of the coherent mode, the STC analysis yielded no significant dimensions for a third of the cells, and very few for the rest (Fig. 5). This happens because the eigenvalue corresponding to the coherent mode injects large eigenvalues into the null eigenvalue distribution (as seen in Eq. (8)), thus masking the cell's true relevant features. Following the correction, the number of significant dimensions per cell increased from 1.1 ± 0.2 to 4.6 ± 0.5 (see Fig. 5A for the full population values). The dimensionality of the relevant subspace increased for 21 out of 22 cells. For one cell, we were unable to find a significant dimension either before or after the correction of the method. The distributions of null eigenvalues used to determine which of the eigenvectors of ΔC are significant (Fig. 5B,C) became much more narrow when evaluated in the subspace orthogonal to the coherent mode.

Discussion

The goal of this work was to extend the range of applicability of a computationally simple method of spike-triggered covariance to strongly correlated stimuli. While the STC method in principle can be used with strongly correlated Gaussian stimuli, our results show that the inhomogeneous sampling variability can in practice make it difficult to recover the correct relevant subspace. We have characterized the effects generated by strong Gaussian correlations using simulations of two model neurons in a wide range of dataset sizes (which could also be viewed as an inverse measure of the neuron's level of internal noise). Results from random matrix theory, and specifically the Wigner and Spiked Wishart ensembles, suggest that the origin of these issues can be traced to the estimation bias and variance of covariance matrices with vastly different eigenvalues. We demonstrate that by considering the coherent mode, which corresponds to the largest eigenvalue, separately from the rest of stimulus dimensions, one can improve the method's sensitivity by $(\tau_1/\tau_2)^2$.

One qualitative lesson offered by these analyses is that while the bulk of the eigenvalues of ΔC is a good proxy for the width of the null distribution in the case of white noise inputs, but not in the case of strongly correlated inputs.

Furthermore, our analysis suggests that sampling variability along the secondary outstanding modes corresponding to the next few principal components may have similar masking effects to the ones reported here for the coherent mode. Possible solutions to the full problem may include performing a sequence of analyses in subspaces of decreasing dimensionality, orthogonal to several leading principal components. However, the payoff from this procedure is (at most) of order $(\tau_2/\tau_3)^2$ which in our case is 1.4. At the same time, one runs the risk of losing the ability to resolve the remaining dimensions because of the reduced signal to noise ratio.

Another potential solution is to correct for the estimation bias and variance in eigenvalues and eigenvectors, described by Eqs. (7) and (8). However, this procedure is difficult computationally and

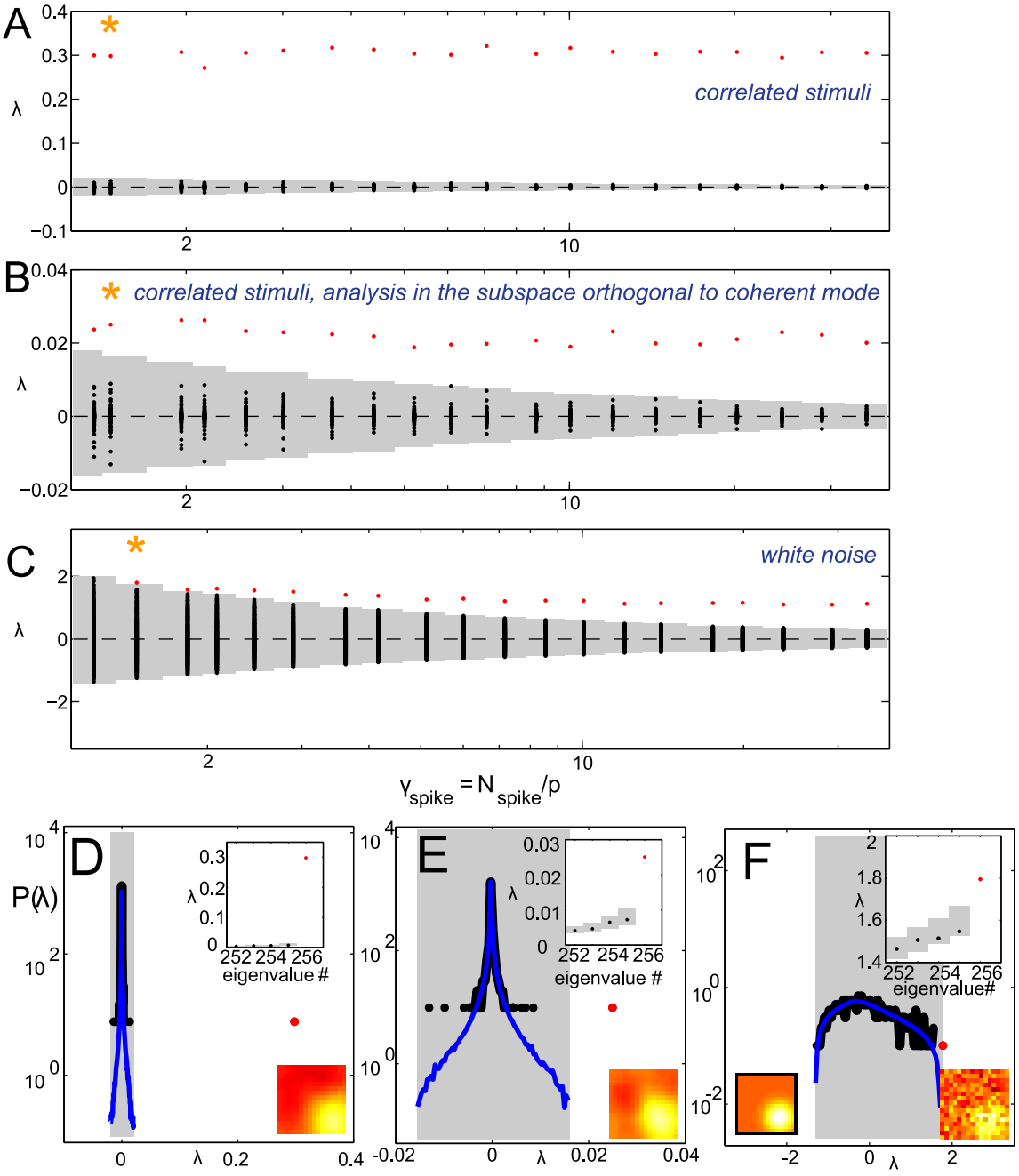

Figure 4. Spike triggered covariance analysis of a model neuron with one relevant feature that has a large component along the coherent mode. Spectra of ΔC for increasing dataset size in the case of strongly correlated Gaussian noise (A,B) and white noise (C). (A–C) The range of spikes covered is from $N_{\text{spike}} \approx 1.3p$ to $N_{\text{spike}} \approx 34p$. Panel B shows the results after evaluating significance in the subspace orthogonal to the coherent mode. Each vertical line shows the result of a single simulation. Significant (insignificant) eigenvalues are shown in red (black), and the [1,99] percentile range of the null distribution (1000 evaluations of ΔC^{null}) is shown in gray. (D–F) Gray shaded area is the support of the null distribution, which itself is plotted in blue. The correct feature is found using both stimulus conditions and both the original formulation of the STC method and our proposed correction. $N = 1995$, $N_{\text{spike}} = 655$ for the correlated stimulus condition in panels D and E, $N = 1995$, $N_{\text{spike}} = 630$ for the white stimulus condition in panel F (indicated by an orange star in panels A–C). Insets to panels D–F show the recovered feature (decorrelated in panels D,E) and the nested significance testing which does not affect the results. Black framed inset to panel D shows the model feature. The models were defined such that the mean firing rate remained unchanged between the two stimulus conditions.
doi:10.1371/journal.pcbi.1003206.g004

in most cases can only be done for simple eigenvalue distributions [43].

The treatment of the artifacts caused by a large coherent mode present in the data has been previously discussed in analyses of stock-markets [44,45], evolution of proteins [46], and Human Immunodeficiency Virus (HIV) mutations [34]. In these cases, the extra dimension was removed and the resulting covariance structure was compared against the Marčenko-Pastur eigenvalue distribution that assumes no correlation between the variables and uniform variable variances. The case of reverse correlation

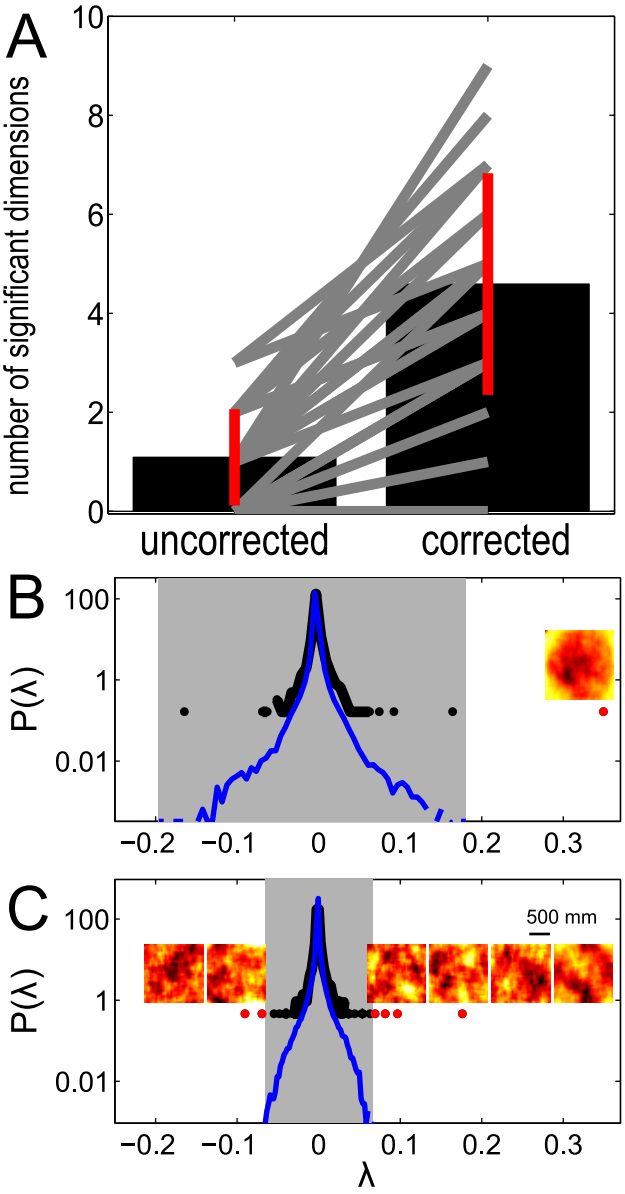

Figure 5. Analysis of salamander retinal ganglion cells. (A) Number of significant dimensions for 22 RGCs presented with a strongly correlated stimulus, before (*left*) and after (*right*) using our correction scheme. Gray lines represent single cells, and red lines represent the standard deviation. (B) Significant (insignificant) eigenvalues of ΔC in red (black) and the null eigenvalue distribution (blue) for an example cell. Null distribution were constructed using the global approach with 500 shuffled spike trains. Without the correction scheme there is only one significant dimension. The corresponding visual feature is shown in the inset before decorrelation. (C) Same for the corrected covariance matrix $\Delta\tilde{C}$. Here there are six significant dimensions with spots likely representing the subunits within the cell's receptive field. The gray shade indicates the range of the null distribution used to determine significance.
doi:10.1371/journal.pcbi.1003206.g005

experiments discussed here is different from these analyses because the spike triggered ensemble is compared to the full stimulus distribution. In addition, our analyses provide two important novel contributions. First, we show there is a crucial difference between discarding the coherent mode and projecting it out. This is because of the way the coherent mode injects noise into the null distribution. Second, the approach described here also permits the inclusion of the components of the relevant dimensions along the coherent mode in the final results. We hope that the ideas for treating the coherent mode presented here will also be relevant in other areas of computational biology.

Materials and Methods

Ethics statement

Experimental data were collected using procedures approved by the Institutional Animal Care and Use Committee of Princeton University, and in accordance with National Institutes of Health guidelines. Experimental and surgical procedures have been described previously [47].

Stimulus

Each stimulus frame s^t was randomly drawn from a multivariate Gaussian distribution with zero mean and covariance matrix C^{pop}, $s^t \sim \mathcal{N}(0, C^{\mathrm{pop}}) \in \mathbb{R}^p$. In the correlated stimulus case, the population covariance C^{pop} was computed from the covariance of 16×16 pixels patches from the van Hateren image database [38] (with no downsampling). In the uncorrelated ("white") case, C^{pop} was the identity matrix.

Testing for significance

We describe two approaches for determining significance of candidate features that were previously described in the literature: global and nested. When applied to our datasets, both of the approaches yielded similar results.

Global null distribution. Eigenvalues of ΔC are identified as significant if they lie outside the $[\alpha, 100-\alpha]$ percentile interval of the null distribution, where parameter α specifies the level of significance. The null distribution is constructed by computing many realizations of the matrix $\Delta C^{\mathrm{null}} = C^{\mathrm{null}} - C^{\mathrm{stim}}$.

If $\alpha = 0$ is used, the number of randomized spike trains is inversely proportional to the confidence interval with which significance is determined.

To describe how matrix C^{null} is computed, we recall that

$$C^{\mathrm{spike}} = \frac{1}{N_{\mathrm{spike}} - 1} \sum_{\kappa=1}^{N_{\mathrm{spike}}} \left(s_i^{\hat{t}_\kappa} - \langle s_i \rangle_{\hat{t}} \right) \left(s_j^{\hat{t}_\kappa} - \langle s_j \rangle_{\hat{t}} \right), \quad (15)$$

where $\langle s_i \rangle_{\hat{t}}$ is the spike-triggered average of the ith stimulus component and spike times are denoted as $\{\hat{t}_\kappa\}_{\kappa=1}^{N_{\mathrm{spike}}} = \left\{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{N_{\mathrm{spike}}}\right\}$.

The matrix C^{null} is computed in the exact same manner, but instead of the spike train $\{\hat{t}_\kappa\}_{\kappa=1}^{N_{\mathrm{spike}}}$, we use random spike trains $\{\tilde{t}_\kappa\}_{\kappa=1}^{N_{\mathrm{spike}}}$:

$$C^{\mathrm{null}} = \frac{1}{N_{\mathrm{spike}} - 1} \sum_{\kappa=1}^{N_{\mathrm{spike}}} \left(s_i^{\tilde{t}_\kappa} - \langle s_i \rangle_{\tilde{t}} \right) \left(s_j^{\tilde{t}_\kappa} - \langle s_j \rangle_{\tilde{t}} \right). \quad (16)$$

Note that there are just as many random spike times $\tilde{t}$ as real spike times $\hat{t}$. Moreover, when the spike train has a meaningful temporal structure, the random spike train can be obtained by a random shift of $\{\hat{t}_\kappa\}_{\kappa=1}^{N_{\mathrm{spike}}}$, defined for all $\kappa = 1 \ldots N_{\mathrm{spike}}$ and for a random integer m chosen uniformly between 1 and N [20]:

$$\tilde{t}_\kappa = (\hat{t}_\kappa + m) \bmod N. \quad (17)$$

Using all the realizations of C^{null} computed (i.e. all the randomly chosen m's), the eigenvalues of ΔC^{null} compose the null distribution.

Nested significance testing. Significance can also be tested in a nested fashion using p rank-ordered null distributions. This method is detailed in [7]. For completeness we briefly describe it here. For each of the randomized spike trains the eigenvalues of ΔC^{null} (or C^{Δ}_{null} for the pre-whitened formulation, see Eq. (25) below) are rank ordered. The eigenvalues of each order from all the randomized spike train compose a separate null distribution. Then, for each of these p distributions we found a confidence interval. If the smallest eigenvalue of ΔC (or C^{Δ}) is smaller than the lower bound of its relevant confidence interval (or if the largest eigenvalue is larger than the upper bound) the corresponding eigenvectors are determined to be significant. These eigenvectors are then projected out of the stimulus and the analysis is repeated until no eigenvalues are found to be significant. Note that according to this method some eigenvalues of ΔC (or C^{Δ}) can identified as insignificant but still be outside of the confidence interval computed for their rank if the largest eigenvalue lies within its confidence intervals (see Fig. 3E, right).

Decorrelation within the STC method

Within the STC method, stimulus correlations need to be removed from the estimates of eigenvectors obtained by diagonalizing matrix δC. This correction is needed, because the eigenvectors of ΔC have a form $\sum_{j=1}^{p} C_{ij}\phi_j$, where ϕ_j describe components of one of the relevant features [20]. As described above, one may wish to use a pseudoinverse, instead of the full inverse of the matrix C to minimize noise amplification at higher spatial frequencies. Assuming that the eigenvalues are ordered to be monotonically decreasing, the pseudoinverse of order k is given by

$$C^{-1}_{ij}(k) = \sum_{n=1}^{k} \frac{1}{\lambda_n} f_i^{(n)} f_j^{(n)}. \tag{18}$$

In the analysis of data from retinal ganglion cells, the optimal order of the pseudoinverse was determined in the following way. The dataset was divided into the training and test sets. The features were computed by diagonalizing the matrix ΔC, cf. Eq. (3), in either the full input space or in the space orthogonal to the coherent mode using the training set. Following that, the optimal pseudoinverse order $k^{\star}$ was selected as the one that yielded decorrelated features that convey the most information about, or give the largest predictive power for, the neural response. Explicitly,

$$k^{\star} = \underset{k\in\{1,\ldots,p\}}{\operatorname{argmax}}\, I(k), \tag{19}$$

$$I(k) = \int d\vec{x} P_k(\vec{x}|\text{spike}) \log \frac{P_k(\vec{x}|\text{spike})}{P_k(\vec{x})}, \tag{20}$$

where $P_m(\vec{x})$ is the probability distribution of the projections of stimuli onto the d significant eigenvectors (q_i), decorrelated by $C^{-1}(k^{\star})$. $\bar{q}_i = C^{-1}(k^{\star})q_i$ are the decorrelated significant features, and:

$$\vec{x} = (x_1, \ldots, x_d), \tag{21}$$

$$x_i = \bar{q}_i \cdot s. \tag{22}$$

Pre-whitening

As an alternative to removing stimulus correlations from the eigenvectors of ΔC, one can remove stimulus correlations from each of the stimulus vectors, prior to the diagonalization of ΔC, a procedure that is known as pre-whitening [7].

The sample stimulus covariance matrix from Eq. (2) can be written in terms of eigenvalues λ_n and eigenvectors $f^{(n)}$ as

$$C^{\text{stim}}_{ij} = \sum_{n=1}^{p} \lambda_n f_i^{(n)} f_j^{(n)}. \tag{23}$$

We can now define a matrix $C^{w}_{ij} = \sum_{n=1}^{p} \lambda_n^{-1/2} f_i^{(n)} f_j^{(n)}$. Then, the analogue of ΔC in the "one-centered" formulation is given by:

$$C^{\Delta} = C^{w} C^{\text{spike}} C^{w}. \tag{24}$$

This procedure is equivalent to whitening each of the stimulus frames independently (by multiplying it with C^{w}) and then computing the spike-triggered covariance.

In the limit of infinite data, the null hypothesis corresponds to $C^{\text{spike}} = C^{\text{stim}}$. In this case $C^{\Delta} = \mathbb{I}$. For a dataset of finite size, the null distribution is computed from many realizations of the matrix

$$C^{\Delta}_{\text{null}} = C_w C^{\text{null}} C_w \tag{25}$$

where C^{null} is defined by Eq. (16). The eigenvalues of C^{Δ} (most of which are close to 1) can then be compared to the null eigenvalue distribution, using either the nested or global comparison tests described above.

In Fig. 3 we analyzed the simulated spike trains using every pseudoinverse order of C^{stim}. The prewhitening is then done using this matrix $C^{-1}(k^{\star})$ Eq. (18) instead of the full rank matrix $C^{-1}(p)$.

Performing the pre-whitened STC analysis using all pseudoinverse orders is equivalent to testing p models. Therefore, the confidence interval of the null distribution should be adjusted from the $[\alpha, 100-\alpha]$ percentile range to $[\alpha_{DS}, 100-\alpha_{DS}]$, where α_{DS} is the Dunn-Šidák correction:

$$\alpha_{DS} = 100\left[1 - \left(1 - \frac{\alpha}{100}\right)^{\frac{1}{p}}\right] \approx \frac{\alpha}{p} \tag{26}$$

Relevant features in the full stimulus space

We recall that according to Ref. [20], the significant eigenvectors of ΔC can be written as

$$e_i = \sum_{n=1}^{p} \lambda_n f_i^{(n)} f_j^{(n)} \phi_j. \tag{27}$$

Thus, the eigenvectors of ΔC represent a sum of projection operators onto the principal components of the stimulus ensemble.

When we perform the STC method in the subspace orthogonal to the first principal component of the stimulus, the eigenvectors of ΔC can be written as

$$\tilde{e}_i = \sum_{n=2}^{p} \lambda_n f_i^{(n)} f_j^{(n)} \phi_j \tag{28}$$

(the coherent mode is exactly the vector $f^{(1)}$).

Comparing expressions for the eigenvectors of ΔC and $\Delta\tilde{C}$, one observes that there is a one-to-one correspondence between them. This correspondence can be identified based on proportionality in components along second, third, and other principal components:

$$\frac{\tilde{e}\cdot f^{(2)}}{e\cdot f^{(2)}} = \frac{\tilde{e}\cdot f^{(3)}}{e\cdot f^{(3)}} = \ldots = \frac{\tilde{e}\cdot f^{(i)}}{e\cdot f^{(i)}} \tag{29}$$

for any $i>1$. In sum, once the eigenvector $\tilde{e}$ is found to be significant in the subspace orthogonal to f_1, the eigenvector that should be identified as significant in the full stimulus space is e that satisfies the condition of Eq. (29).

Model neurons

The nonlinearity was chosen to be a logistic function because such functions maximize the cell's noise entropy and thus minimize the assumptions imposed on the cell's response [48]. Using the models, we generated simulated spike trains in response to either a white or a correlated noise stimulus.

The model used in Fig. 3 (model "A") had a two dimensional relevant subspace with features orthogonal to the coherent mode. The probability of spiking was modeled to increase when the projection of the stimulus on either of the preferred features was large in absolute value (representing a logical OR function). If $\left[\phi_1^A, \phi_2^A\right]$ are the preferred model features and s^t is the stimulus presented at time t (here, the ϕ's and s^t are $p=256$ dimensional vectors) then the probability of a spike at time t is:

$$p_t^A(\text{spike}|s^t) = 1 - \prod_{i=1,2}\left[1 - \frac{1}{1+e^{\left(\theta^A - \left|\phi_i^A\cdot s^t\right|\right)/\delta^A}}\right] \tag{30}$$

where δ^A and θ^A are parameters that determine the width and (soft) thresholds of the sigmoid nonlinearities for the model.

We have also considered the case where the projection of the stimulus on the features was not taken in absolute value, corresponding to a monotonic nonlinearity. In that case (model "A_1", used in Fig. 1) the model was one dimensional, so the probability of a spike is

$$p_t^{A_1}(\text{spike}|s^t) = \frac{1}{1+e^{\left(\theta^{A_1} - \phi_1^A\cdot s^t\right)/\delta^{A_1}}}. \tag{31}$$

The effects described above were observed for both symmetric (Fig. 3) and monotonic (Fig. 1) nonlinearities.

The second model had one relevant input feature ϕ^B with a large component along the coherent mode. In this case, the probability of a spike was modeled as:

$$p_t^B(\text{spike}|s^t) = \frac{1}{1+e^{\left(\theta^B - \phi^B\cdot s^t\right)\delta^B}} \tag{32}$$

where δ^B and θ^B are the width and the threshold of the sigmoid nonlinearity of this model.

In units of the standard deviation of the projection of the stimulus on the model features (σ_{white}, σ_{corr}) the model parameters were chosen to be:

$$\frac{\theta^A_{\text{corr}}}{\sigma^A_{\text{corr}}} = \frac{\theta^A_{\text{white}}}{\sigma^A_{\text{white}}} = 2.3, \frac{\delta^A_{\text{corr}}}{\sigma^A_{\text{corr}}} = \frac{\delta^A_{\text{white}}}{\sigma^A_{\text{white}}} = 0.73 \tag{33}$$

$$\frac{\theta^{A_1}_{\text{corr}}}{\sigma^{A_1}_{\text{corr}}} = 1.02, \frac{\delta^{A_1}_{\text{corr}}}{\sigma^{A_1}_{\text{corr}}} = 0.85 \tag{34}$$

$$\frac{\theta^B_{\text{corr}}}{\sigma^B_{\text{corr}}} = \frac{\theta^B_{\text{white}}}{\sigma^B_{\text{white}}} = 0.89, \frac{\delta^B_{\text{corr}}}{\sigma^B_{\text{corr}}} = \frac{\delta^B_{\text{white}}}{\sigma^B_{\text{white}}} = 0.28 \tag{35}$$

Subspace overlap

The overlap measure we use when the dimensionality of the relevant subspace is greater than one is given by [49]:

$$\mathcal{O}(\Phi,\Psi) = \left(\frac{\left|\det \Phi^T\Psi\right|}{\sqrt{\left|\left(\det \Phi^T\Phi\right)\left(\det \Psi^T\Psi\right)\right|}}\right)^{\frac{1}{d}}, \tag{36}$$

where Φ and Ψ are $p\times d$ matrices that hold the model and computed features, respectively, p is the input dimensionality, and d is the number of relevant features in the model.

Acknowledgments

We thank William Bialek for helpful discussions, Joel Kaardal, Sven Boemer, Jacob Maskiewicz for help with preliminary analyses of the retinal ganglion cell responses, and Andrew T. Briggs for comments on the manuscript.

Author Contributions

Conceived and designed the experiments: JA RS MJB TOS. Performed the experiments: JA RS. Analyzed the data: JA RS. Wrote the paper: JA RS MJB TOS.

References

1. Marr D, Hildreth E (1980) Theory of edge detection. Proc R Soc Lond 207: 187–217.
2. Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. The Journal of Neuroscience 22: 10811–10818.
3. Felsen G, Touryan J, Han F, Dan Y (2005) Cortical sensitivity to visual features in natural scenes. PLoS biology 3: e342.
4. Touryan J, Felsen G, Dan Y (2005) Spatial structure of complex cell receptive fields measured with natural images. Neuron 45: 781–791.
5. Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque v1 receptive fields. Neuron 46: 945–956.
6. Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, et al. (2006) Selectivity for multiple stimulus features in retinal ganglion cells. Journal of Neurophysiology 96: 2724–2738.

7. Schwartz O, Pillow JW, Rust NC, Simoncelli EP (2006) Spike-triggered neural characterization. Journal of Vision 6: 484–507.
8. Atencio CA, Sharpee TO, Schreiner CE (2008) Cooperative nonlinearities in auditory cortical neurons. Neuron 58: 956–966.
9. Atencio CA, Sharpee TO, Schreiner CE (2009) Hierarchical computation in the canonical auditory cortical circuit. Proceedings of the National Academy of Sciences 106: 21894–21899.
10. Sharpee TO, Nagel KI, Doupe AJ (2011) Two-dimensional adaptation in the auditory forebrain. Journal of Neurophysiology 106: 1841–1861.
11. Kim AJ, Lazar AA, Slutskiy YB (2011) System identification of drosophila olfactory sensory neurons. J Comput Neurosci 30: 143–161.
12. Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME (2007) Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. PLoS Biol5: e19.
13. Fox JL, Fairhall AL, Daniel TL (2010) Encoding properties of haltere neurons enable motion feature detection in a biological gyroscope. Proceedings of the National Academy of Sciences 107: 3840–3845.
14. Fitzgerald JD, Rowekamp RJ, Sincich LC, Sharpee TO (2011) Second-order dimensionality reduction using minimum and maximum mutual information models. PLoS Comput Biol 7: e1002249.
15. Rajan K, Bialek W (2012) Maximally informative 'stimulus energies' in the analysis of neural responses to natural signals. URL http://arxiv.org/abs/1201.0321. ArXiv:1201.0321 [q-bio.NC].
16. Park IMM, Pillow JW (2011) Bayesian spike-triggered covariance analysis. In: Shawe-Taylor J, Zemel R, Bartlett P, Pereira F, Weinberger K, editors, Advances in Neural Information Processing Systems 24, MIT Press. pp. 1692–1700.
17. de Ruyter van Steveninck RR, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc Lond B 265: 259–265.
18. Paninski L (2003) Convergence properties of three spike-triggered analysis techniques. Network 14: 437–464.
19. Ringach DL (2004) Mapping receptive fields in primary visual cortex. The Journal of Physiology 558: 717–728.
20. Bialek W, de Ruyter van Steveninck RR (2005) Features and dimensions: Motion estimation in fly vision. URL http://arxiv.org/abs/q-bio/0505003. ArXiv:q-bio/0505003 [q-bio.NC].
21. Pillow JW, Simoncelli EP (2006) Dimensionality reduction in neural models: An informationtheoretic generalization of spike-triggered average and covariance analysis. Journal of Vision 6: 414–428.
22. Kouh M, Sharpee TO (2009) Estimating linear-nonlinear models using Rényi divergences. Network 20: 49–68.
23. Samengo I, Gollisch T (2012) Spike-triggered covariance: geometric proof, symmetry properties, and extension beyond gaussian stimuli. Journal of Computational Neuroscience : 1–25.
24. Ruderman DL, Bialek W (1994) Statistics of natural images: scaling in the woods. Phys Rev Lett 73: 814–817.
25. van Hateren JH, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proc R Soc lond B Biol Sci 265: 2315–2320.
26. Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. J Acoust Soc Am A 114: 3394–3411.
27. Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A 4: 2379–2394.
28. Ruderman DL, Cronin TW, Chiao CC (1998) Statistics of cone responses to natural images: implications for visual coding. J Opt Soc Am A 15: 2036–2045.
29. Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193–216.
30. Lee TW,Wachtler T, Sejnowski TJ (2002) Color opponency is an efficient representation of spectral properties in natural scenes. Vision Research 42: 2095–2103.
31. Hosoya T, Baccus S, Meister M (2005) Dynamic predictive coding by the retina. Nature 436: 71–77.
32. Sharpee TO, Sugihara H, Kurgansky A, Rebrik S, Stryker M, et al. (2006) Adaptive filtering enhances information transmission in visual cortex. Nature 439: 936–942.
33. Grabow C, Grosskinsky S, Timme M (2012) Small-world network spectra in mean-field theory.Physical Review Letters 108: 218701.
34. Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, et al. (2011) Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proceedings of the National Academy of Sciences 108: 11530–11535.
35. Hoyle DC, Rattray M (2004) Principal-component-analysis eigenvalue spectra from data with symmetry-breaking structure. Physical Review E 69: 026124.
36. Paul D (2007) Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Statistica Sinica 17: 1617–1642.
37. Theunissen F, David S, Singh N, Hsu A, VinjeW, et al. (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network 12: 289–316.
38. van Hateren J, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proceedings of the Royal Society of London Series B: Biological Sciences 265: 2315–2320.
39. Ruderman DL, Bialek W (1994) Statistics of natural images: Scaling in the woods. Phys Rev Lett 73: 814–817.
40. Edwards SF, Jones RC (1976) The eigenvalues spectrum of a larg symmetric random matrix. J Phys A: Math Gen 9: 1595–1603.
41. Jones RC, Kosterlitz JM, Thouless DJ (1978) The eigenvalue spectrum of a large symmetric random matrix with a finite mean. Journal of Physics A: Mathematical and General 11: L45.
42. Wigner EP (1958) On the distribution of the roots of certain symmetric matrices. The Annals of Mathematics 67: 325–327.
43. El Karoui N (2008) Spectrum estimation for large dimensional covariance matrices using random matrix theory. The Annals of Statistics 36: 2757–2790.
44. Laloux L, Cizeau P, Bouchaud JP, Potters M (1999) Noise dressing of financial correlation matrices. Phys Rev Lett 83: 1467–1470.
45. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, et al. (2002) Random matrix approach to cross correlations in financial data. Phys Rev E 65: 066126.
46. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: Evolutionary units of three-dimensional structure. Cell 138: 774–786.
47. Segev R, Goodhouse J, Puchalla J, Berry II MJ (2004) Recording spikes from a large fraction of the ganglion cells in a retinal patch. Nat Neurosci 7: 1155–1162.
48. Fitzgerald JD, Rowekamp RJ, Sincich LC, Sharpee TO (2011) Second order dimensionality reduction using minimum and maximum mutual information models. PLoS Comput Biol 7: e1002249.
49. Rowekamp RJ, Sharpee TO (2011) Analyzing multicomponent receptive fields from neural responses to natural stimuli. Network 22: 45–73.