Reconstruction of sparse connectivity in neural networks from spike train covariances

J. Stat. Mech. (2013) P03008

(http://iopscience.iop.org/1742-5468/2013/03/P03008)

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Reconstruction of sparse connectivity in neural networks from spike train covariances

Volker Pernice and Stefan Rotter

Bernstein Center Freiburg and Faculty of Biology, University of Freiburg, Hansastraße 9a, 79104 Freiburg, Germany
E-mail: pernice@bcf.uni-freiburg.de
and stefan.rotter@biologie.uni-freiburg.de

Received 15 June 2012
Accepted 24 September 2012
Published 12 March 2013

Online at stacks.iop.org/JSTAT/2013/P03008
doi:10.1088/1742-5468/2013/03/P03008

Abstract. The inference of causation from correlation is in general highly problematic. Correspondingly, it is difficult to infer the existence of physical synaptic connections between neurons from correlations in their activity. Covariances in neural spike trains and their relation to network structure have been the subject of intense research, both experimentally and theoretically. The influence of recurrent connections on covariances can be characterized directly in linear models, where connectivity in the network is described by a matrix of linear coupling kernels. However, as indirect connections also give rise to covariances, the inverse problem of inferring network structure from covariances can generally not be solved unambiguously.

Here we study to what degree this ambiguity can be resolved if the sparseness of neural networks is taken into account. To reconstruct a sparse network, we determine the minimal set of linear couplings consistent with the measured covariances by minimizing the L_1 norm of the coupling matrix under appropriate constraints. Contrary to intuition, after stochastic optimization of the coupling matrix, the resulting estimate of the underlying network is directed, despite the fact that a symmetric matrix of count covariances is used for inference.

The performance of the new method is best if connections are neither exceedingly sparse, nor too dense, and it is easily applicable for networks of a few hundred nodes. Full coupling kernels can be obtained from the matrix of full covariance functions. We apply our method to networks of leaky integrate-and-fire neurons in an asynchronous–irregular state, where spike train covariances are well described by a linear model.

1742-5468/13/P03008+15$33.00

Keywords: neuronal networks (theory), network dynamics, network reconstruction, computational neuroscience

Contents

1. Introduction

Technological advances have made it possible to record the activity of an increasingly large number of neurons, with a temporal and spatial resolution depending on the experimental technique. At the same time, information about the synaptic layout of networks becomes more and more available, thanks to modern immunostaining and optogenetic methods. An important step in interpreting such data in a functional context is to understand the relation between the underlying connectivity among neurons and the neuronal dynamics. An equally active area of research deals with the inverse problem of inferring physical connections from the observed dynamics.

In contrast to the physical connection between neurons, functional connectivity is often defined in terms of correlations or coherences in their activity. Relations between properties of individual cells, network structure and the resulting correlations on various levels of resolution have been the subject of a large number of studies. Linear models present a practical way to explore the relationship between observable dynamics and the synaptic network [1]–[6]. These models should be interpreted as a linear approximation about a given operating point of the dynamics of a complex nonlinear system, and they offer useful theoretical insights into non-trivial effects of the network structure. In particular, the issue of correlation transfer [7] as well as decorrelation in balanced networks of excitatory and inhibitory neurons has been studied on this level [8]. Furthermore, it turns out that linear models are remarkably accurate under a variety of conditions in a

J. Stat. Mech. (2013) P03008

state of weakly correlated neurons, akin to the one experimentally observed, for example, in networks of integrate-and-fire neurons [3, 6, 4]. These results suggest that a linear framework constitutes a suitable starting point to approach the inverse problem of finding networks consistent with the observed correlated activity.

A host of methods are in use for connectivity reconstruction in neural networks. In particular, to deal with the problem of correlations due to indirect connections, the following approaches have been suggested: Granger causality [9], dynamic Bayesian networks [10], transfer entropy [11, 12] and, in the frequency domain, partial coherence [13] which can also be generalized to nonlinear systems [14, 15]. The coupling parameters of maximum entropy models are also used to infer functional interactions [16]. A maximum entropy model is consistent with measured rates and correlations but assumes no constraints beyond that. Sparse models can be constructed by using only a selected set of interactions [17]. Alternatively, dynamical models are directly fitted to the data, and connections are inferred from the resulting parameters. Examples include autoregressive processes [18], Ising models [19], nonlinear dynamical systems [20], coupled Markov chains [21] and networks of leaky integrate-and-fire neurons [22, 23]. In another widely used approach generalized linear models are employed [24]–[26].

It is commonly assumed that the direct use of measured correlations has two prominent disadvantages. First, indirect connections and shared input induce spurious correlations in addition to the ones caused by direct connections. This is the case even if the activity of all nodes of a network is observed and correlations induced from correlated external inputs are excluded, which is the situation we examine in this study. Second, the covariance matrix is symmetric and, therefore, cannot reflect directed interactions. In this work we want to argue that both assumed pitfalls can be overcome, and that an estimation of directed and weighted connections can be achieved under certain conditions, even if only a matrix of covariances without temporal information is available.

The basic idea is to search for a matrix with a minimal number of connections consistent with the observed covariances. In this way, spurious links that do not correspond to a synaptic connection are correctly interpreted. The procedure will give an adequate estimate of the directed connectivity, because the covariance matrix depends distinctly on the direction of the individual connections, and precisely accounts for their indirect contributions. As an example, figure 1, the covariance matrix of a network with very few isolated connections (here two links between four neurons) does not depend on their directions. If, however, further connections are present, various contributions of the covariances are introduced by indirect connections that depend on the precise circuit structure (here the direction of the two initial connections). Only one configuration of these specific connections is consistent with a given covariance matrix and a fixed number of connections.

2. Covariances in linear models

In certain linear models of spiking activity covariance functions between individual pairs of neurons can be calculated explicitly.

In a framework of linearly interacting point processes [27], the spike train of neuron i, $s_i(t) = \sum_j \delta(t - t_j^i)$, is regarded as a realization of an inhomogeneous Poisson process with time dependent rate, such that $\langle s_i(t) \rangle = y_i(t)$. The vector of firing rates, y, in turn, depends

J. Stat. Mech. (2013) P03008

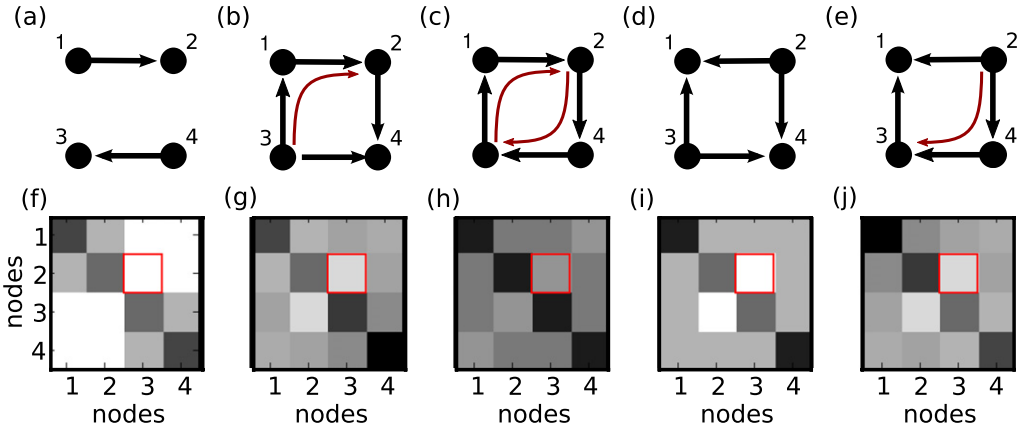

Figure 1. In general, the covariance matrix contains information about the direction of connections that can be exploited for connectivity reconstruction. (a) In this network with only two isolated connections, the covariance matrix based on equation (3), (f), does not depend on the direction of the connections (darker colors correspond to stronger covariances). (b)–(e) If additional connections $3 \to 1$ and $2 \to 4$ are inserted, the covariance matrix depends on the direction of the initial two connections, (g)–(j). The reason is distinct combinations of indirect connections and shared inputs. For example, the differences in indirect connections between nodes 2 and 3 (red arrows) cause different covariances C_{23} (red boxes).

on the external input x and the recurrent input due to synaptic connections between nodes. The influence of connections is captured by a matrix $G(t)$ of linear coupling kernels. To ensure causality, $G(t) = 0$ for negative time lags $t < 0$. The dynamics of the network is then governed by the equation

$$y(t) = x + G * s(t), \tag{1}$$

where $*$ denotes the convolution product of functions. The input x is assumed to be constant and positive. All vectors are to be interpreted as column vectors, and the elements of the coupling matrix $G_{ij}(t)$ describe the influence of node j on node i.

Assuming stationarity, time-averaged rates are given by $\langle y \rangle = (\mathbb{1} - \int G(t)\,\mathrm{d}t)^{-1}x$. With the ensemble expectation $\langle\cdot\rangle$, covariance functions are defined as $C(\tau) = \langle s(t)s^T(t-\tau)\rangle$. It can be shown that they can, in the Fourier domain, be expressed as

$$\hat{C}(\omega) = [\mathbb{1} - \hat{G}(\omega)]^{-1}\hat{Y}(\omega)[\mathbb{1} - \hat{G}^*(\omega)]^{-1}, \tag{2}$$

where $\mathbb{1}$ denotes the identity matrix, * is the conjugate transpose of a complex matrix and $\hat{f}(\omega)$ indicates the Fourier transform of a function $f(t)$. The elements of the diagonal matrix $\hat{Y}$ are given by the time-averaged rates of the network neurons. The term $(\mathbb{1} - \hat{G})^{-1}$ reflects the effect of the recurrent network. A rigorous discussion of this equation and its properties can be found in [27, 5].

Analogous expressions also arise in a model involving fluctuating firing rates. In this case, rates are considered to be continuous signals that obey $y(t) = x(t) + G * y(t)$, instead of equation (1). For simplicity, it is assumed that x is stationary white noise with $\langle x \rangle = 0$ and the firing rates y fluctuate about some stationary value. After a Fourier

J. Stat. Mech. (2013) P03008

transformation, $\hat{y}(\omega) = \hat{x}(\omega) + \hat{G}(\omega)\hat{y}(\omega)$, and it follows that

$$\langle \hat{y}\hat{y}^* \rangle = \hat{C}(\omega) = [\mathbb{1} - \hat{G}(\omega)]^{-1}\hat{Y}(\omega)[\mathbb{1} - \hat{G}^*(\omega)]^{-1}. \tag{3}$$

Now, the matrix $\hat{Y} = \langle \hat{x}\hat{x}^* \rangle$ depends on the properties of the external input. If it is uncorrelated for different neurons, $\hat{Y}$ is again diagonal. The same equation is also derived in [4], where $y(t)$ corresponds to a mixed point and continuous process that approximates the output of an integrate-and-fire neuron.

For both models, for each frequency ω, some information on the direction of the connection between nodes is contained in the imaginary part of the covariances, as it reflects the asymmetry of the covariance functions in the time domain. Only for $\omega = 0$ is this not the case. For point processes, the matrix $\hat{C}(\omega = 0)$ of integrated cross-covariance functions defines the count covariances between neuron pairs for infinitely large time bins.

Linear models can be used to approximate the nonlinear dynamics of a recurrent network at some operating point in the framework of linear response theory [28, 4, 6]. Even if many different terms resulting from multisynaptic pathways or shared inputs contribute, this approximation is remarkably accurate for networks of more or less asynchronously spiking neurons. This in particular applies to the cancelation of some excitatory and inhibitory interactions in balanced networks. The nonlinear properties of single neurons are of minor importance under these conditions.

3. Reconstruction of the network from covariances

The objective of this paper is to find an estimate of the coupling matrix G describing the linear effects of the true synaptic connectivity from the known covariance matrix, using equation (3). The inverse covariance matrix can be expressed as

$$\hat{C}^{-1}(\omega) = \hat{B}(\omega)^*\hat{B}(\omega), \tag{4}$$

with $\hat{B} = \sqrt{\hat{Y}}(\mathbb{1} - \hat{G})$. Based on the covariances, only the elements of $\hat{B}$ instead of the linear couplings can be estimated. Nevertheless, this measure reflects the strength and sign of the couplings. Note that, strictly speaking, self-connections, that is diagonal couplings, cannot be inferred based on $\hat{B}$, as the true value of the diagonal term $\hat{Y}$ is unknown. The elements of $\hat{G}$ describe the linear influences nodes exert upon each other. For example, $\hat{G}_{ij}(0)$ denotes the expected total number of spikes of i that are evoked by an extra spike of neuron j. As $\hat{B}_{ij} = -\hat{G}_{ij}\sqrt{\hat{Y}_{ii}}$ for $i \neq j$, the estimated linear couplings are distorted by a positive factor. For the Hawkes model, $\hat{Y}_{ii}$ is simply the rate of neuron i. In the rate model, $\hat{Y}_{ii}(\omega)$ is the power spectrum of the external input to i, and therefore not directly accessible to measurement. Its order of magnitude can still be estimated from the covariance matrix, if couplings are weak, as $\hat{C}_{ii}^{-1} = \hat{Y}_{ii}(1 - \hat{G}_{ii} - \hat{G}_{ii}^*) + \sum_k |\hat{G}_{ki}|^2\hat{Y}_{kk}$, from equation (3). The partial spectral coherence, a related measure that has been proposed for the inference of interactions, is given by $\hat{C}_{ij}^{-1}/\sqrt{\hat{C}_{ii}^{-1}\hat{C}_{jj}^{-1}}$. In comparison to raw covariances, some linear effects of neurons $k \neq i, j$ are removed in this measure [13]. However, as $\hat{C}^{-1} = \hat{Y} - \hat{G}^*\hat{Y} - \hat{Y}\hat{G} + \hat{G}^*\hat{Y}\hat{G}$ spurious interactions are still present. For example, in the last term, the elements $\sum_k \bar{\hat{G}}_{ki}\hat{Y}_{kk}\hat{G}_{kj}$, with $\bar{X}$ denoting the complex conjugate of

J. Stat. Mech. (2013) P03008

X, introduce connections between neurons that share a common postsynaptic neuron ('marrying parents of joint children'). This effect is avoided if $\hat{B}$ can be estimated.

However, one has to face a fundamental ambiguity here. For any unitary transformation U with $UU^* = U^*U = \mathbb{1}$, we have

$$\hat{B}^* U^* U \hat{B} = \hat{B}^* \hat{B}.$$

As a consequence, $\hat{B}$ is only constrained by $\hat{C}$ up to a unitary transformation. This expresses the fact that a multitude of different connectivity matrices give rise to an identical covariance matrix. The U include, for instance, simultaneous rotations of the column vectors or permutations of the rows of $\hat{B}$. For example, since $\hat{C}^{-1}_{ij} = \sum_k \bar{\hat{B}}_{ki} \hat{B}_{kj}$, a permutation of the summation indices k does not affect the covariances. This corresponds to a permutation of the targets of the nodes i, j in the network given by $\hat{B}$. Note that, due to the contribution of the identity matrix to $\hat{B}$, these transformations do not exactly correspond to permutations of the neuron output given by $\hat{G}$. What is also not allowed, of course, is permutation of the neuron identities, because this would shuffle entries in the covariance matrix.

A complex matrix $\hat{B}$ has $2n^2$ real degrees of freedom, if n is the number of neurons in the network. The fact that $\hat{C}$ is self-adjoint, which follows from $C(t) = C^T(-t)$, which holds for any covariance matrix, imposes n^2 constraints on $\hat{B}$, and the missing n^2 degrees of freedom correspond exactly to the real dimension of the unitary group $U(n)$.

Note that in general the matrix $\hat{B}$ is non-normal, that is $\hat{B}^*\hat{B} \neq \hat{B}\hat{B}^*$. Specifically, for $\omega = 0$, this matrix is real. Even in this case, the transpose of the connectivity matrix, corresponding to a reversal of all arrows, yields generally a different covariance matrix.

4. Sparse connectivity can resolve ambiguity

Often networks have sparse connectivity. This fact can be exploited to disambiguate between connectivity matrices consistent with a given covariance matrix. As $\hat{G}$ cannot be accessed directly, a sparse solution for $\hat{B}$ is desired. We use the entry-wise L_1 matrix norm

$$\Gamma(\hat{B}) = \sum_{i \neq j} |\hat{B}_{ij}| \tag{5}$$

of the off-diagonal elements of $\hat{B}$ to define a cost function Γ [29]. Minimizing this function tends to favor sparse matrices. L_1-minimization has been used to find sparse regression estimates in the Lasso method [30], as well as in compressed sensing [31, 32], although commonly for real matrices. There, it is exploited that, under certain conditions, the minimization of the L_1 norm is equivalent to minimizing the number of non-zero entries of a matrix [33]. For a recent review on compressed sensing in neuroscience see also [34]. Note that minimization of the L_2 norm is of no use in this context, as a unitary transformation does not affect the L_2 norm.

As an example, consider the matrix

$$\hat{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

J. Stat. Mech. (2013) P03008

With respect to 'small' orthogonal transformations (all Euler angles are small), this matrix has an optimal $\Gamma(\hat{B})$. Therefore, the matrix with no connections, $\hat{G} = 0$, has a local minimum in the cost function. If just one connection of weight $g \ll 1$ is present,

$$\hat{B} = \begin{pmatrix} 1 & g & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

one finds that the matrix $\hat{B}'$ obtained after an orthogonal transformation

$$\hat{B}' = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \hat{B} = \begin{pmatrix} \cos(\phi) & 0 & 0 \\ -\sin(\phi) & \cos(\phi) - g\sin(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with angle $\phi = \arctan(-g)$ has a smaller cost attached than $\hat{B}$. In this matrix, the direction of the connection is reversed, with a slightly smaller weight. The change in the diagonal elements ensures that covariances remain unchanged. Rotations mixing all three dimensions will alter connection weights to the third neuron as well.

However, if $\hat{B}$ has two connections,

$$\hat{B} = \begin{pmatrix} 1 & g & g \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$\hat{B}$ is again the local minimum with respect to Γ. The intuition from these examples is thus that a correct local minimum can be found if the matrix is not exceedingly sparse, and if a reasonable initial guess is available that in particular rules out permutations of the rows.

As a first guess for $\hat{B}$ from $\hat{C}$, we will use the Cholesky decomposition of $\hat{C}^{-1}$. For Hermitian and positive definite $\hat{C}$ it returns an upper triangular matrix A with $\hat{C}^{-1} = A^*A$ that has strictly positive diagonal entries. This initial guess is advantageous for several reasons.

- As the diagonal elements of A are positive, as are the ones from $\hat{B}$, it is not necessary to consider unitary transformations with $\det(U) = -1$.
- If $\hat{C} = \hat{C}^{-1} = \mathbb{1}$, then $A = \mathbb{1}$. Therefore, if covariances are small, the initial guess A can be expected to be not far from the original $\hat{B}$, in the sense that the off-diagonal entries of the unitary transformation U are small. This rules out permutations of rows of $\hat{B}$.
- The decomposition $\hat{B} = UA$ can be carried out using a sequence of unitary transformations in a planar subspace (Givens rotations) where elements of the lower triangular part of $\hat{B}$ are sequentially eliminated (Givens reduction method [35]).

5. Minimization of the cost function

The Fourier transformed matrix of covariance functions is considered independently for each frequency. A simple way to find an estimate of the matrix $\hat{B}(\omega)$ consistent with a

J. Stat. Mech. (2013) P03008

given covariance matrix $\hat{C}(\omega)$ that is optimized with respect to the cost function (5) is a random search. A search step consists of a random unitary transformation. For real matrices one can use Givens rotations about a small angle ϕ. This is equivalent to a two-dimensional orthogonal transformation of the $2 \times n$ matrix consisting of two randomly chosen rows of the current estimate for $\hat{B}$. The rotation angle can be drawn from a suitably chosen distribution depending on the expectation of the entries of the coupling matrix. In simulated neural networks, we choose angles uniformly at random with a maximum size given by the mean absolute value of off-diagonal entries of A. This value seems reasonable, as for small coupling weights angles with $\sin(\phi) \approx \phi$ can be used to reverse the directions of connections.

For $\omega > 0$, complex Givens rotations have to be used. We parametrize random unitary transformations u in a planar subspace as

$$u = \begin{pmatrix} \cos(\phi) & \sin(\phi)\mathrm{e}^{\mathrm{i}\beta} \\ -\sin(\phi)\mathrm{e}^{-\mathrm{i}\beta} & \cos(\phi) \end{pmatrix},$$

where the small angle parameter ϕ is chosen as above and the phase β uniformly from $(-\pi/2, \pi/2]$ to cover values between $+\mathrm{i}$ and $-\mathrm{i}$.

After a sufficiently large number of steps an estimate for a sparse $\hat{B}(\omega) = \sqrt{\hat{Y}(\omega)}[\mathbb{1} - \hat{G}(\omega)]$ is obtained from A that is consistent with the given covariance matrix. Off-diagonal linear couplings then correspond to the elements of $\sqrt{\hat{Y}} - \hat{B} = \sqrt{\hat{Y}}\hat{G}$, up to the positive normalization factor depending on the elements of the diagonal matrix $\hat{Y}$. If only the existence of connections is of interest, data for different frequencies can be combined to

$$\bar{B} = \int |\hat{B}(\omega)|^2 \, \mathrm{d}\omega, \tag{6}$$

and a connection is assumed if the elements of $\bar{B}$ exceed a threshold. The integration in equation (6) is performed element-wise. The rationale is that zero-valued elements should coincide for different frequencies, as missing connections do not contribute for any frequency, even if the cost functions (5) are optimized for each $\hat{B}(\omega)$ independently. This is useful for noisy covariance functions estimated from real data. It can be expected that a combination of different frequencies reduces the noise, if it is independent for different frequencies.

The approach is summarized in figure 2. Panel (a) shows an example of a randomly generated real matrix representing the matrix of integrated coupling kernels $\hat{G}(\omega = 0)$ of a linear system. The expected count covariances $\hat{C}(0)$ of such a system are given by equation (3). For simplicity, we set $\hat{Y} = \mathbb{1}$ in panel (b). Generally, covariance functions would be obtained from recorded data, or from numerical simulations of the dynamics of a model network. The Cholesky decomposition (c) of $\hat{C}^{-1}$ yields, after optimization with respect to the cost function Γ, an estimate of $\hat{B}$, and eventually, up to normalization, of $\hat{G}$, panel (d), that can be compared to the original matrix of direct connections (e)–(f). The stepwise procedure used for optimization suggests the use of simulated annealing to avoid local minima and improve the results.

J. Stat. Mech. (2013) P03008

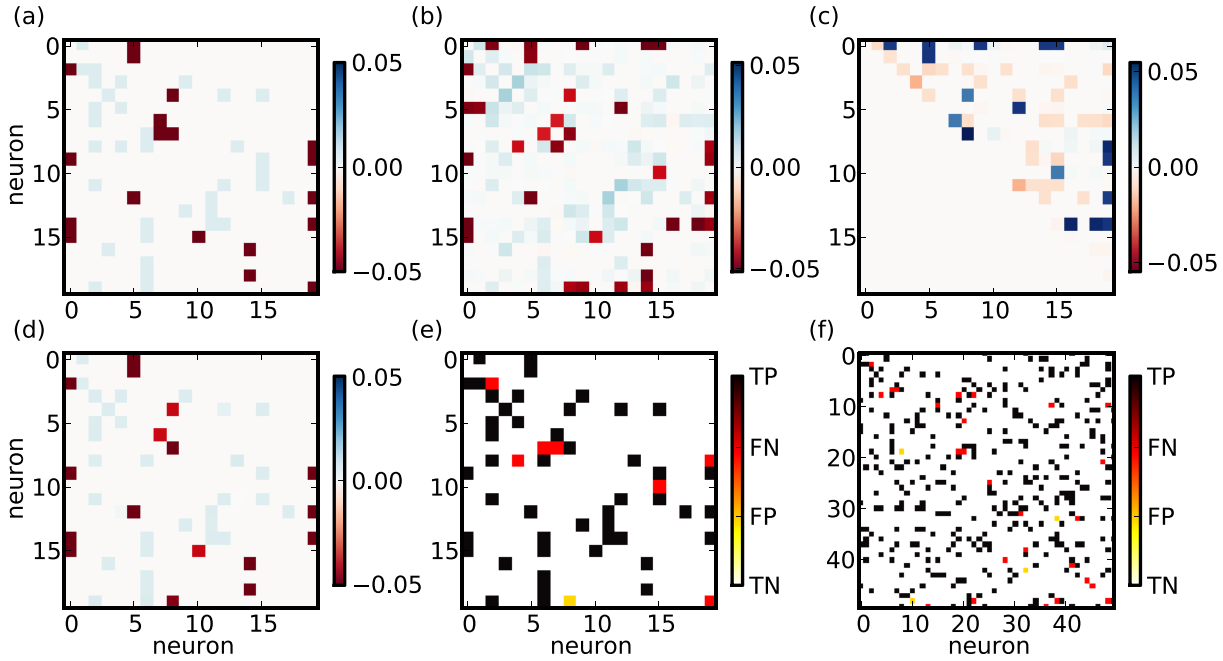

Figure 2. Estimation of connectivity for an example of a real matrix of count covariances. (a) Connectivity matrix of linear couplings $\hat{G}(0)$ with excitatory (positive, blue) and inhibitory (negative, red) connections (only part of a network of 50 neurons is shown for better visibility). (b) The resulting covariance matrix $\hat{C}(0)$, and (c) Cholesky decomposition of the covariance matrix (diagonal elements not shown for better visibility). (d) Estimated couplings, consistent with $\hat{C}(0)$, but with minimized L_1 norm. (e) Estimated adjacency matrix using a threshold. White: true negatives, yellow: false positives, red: false negatives, black: true positives. (f) Full estimated adjacency matrix.

6. Numerical results

We conducted a series of numerical simulations to test the applicability of the presented algorithm. In section 6.1 we study to what degree directed connections can be inferred, if no temporal information but the exact covariance matrix $\hat{C}(0)$ of a linear system is available. In section 6.2 we demonstrate that additional information from the matrix of full covariance functions can be used to make up for the degrading effects of noise, which invariably exists in measured data.

6.1. Dependence of the performance on network parameters

The success of the method depends on the density of connections, as well as on the strength of the couplings. For parameter scans, we generated random connectivity matrices with varying connection probability p and real entries. Neurons were chosen with 80 % probability to be excitatory (connection weights g_E) and with 20 % probability to be inhibitory (connection weights $g_I = -5g_E$). Covariance matrices were then calculated from equation (3) with $\hat{Y} = \mathbb{1}$ (the dynamics of the system was not modeled explicitly). A number of $n/2 \times 4 \times 10^6$ unitary transformations were tried during the optimization of

J. Stat. Mech. (2013) P03008

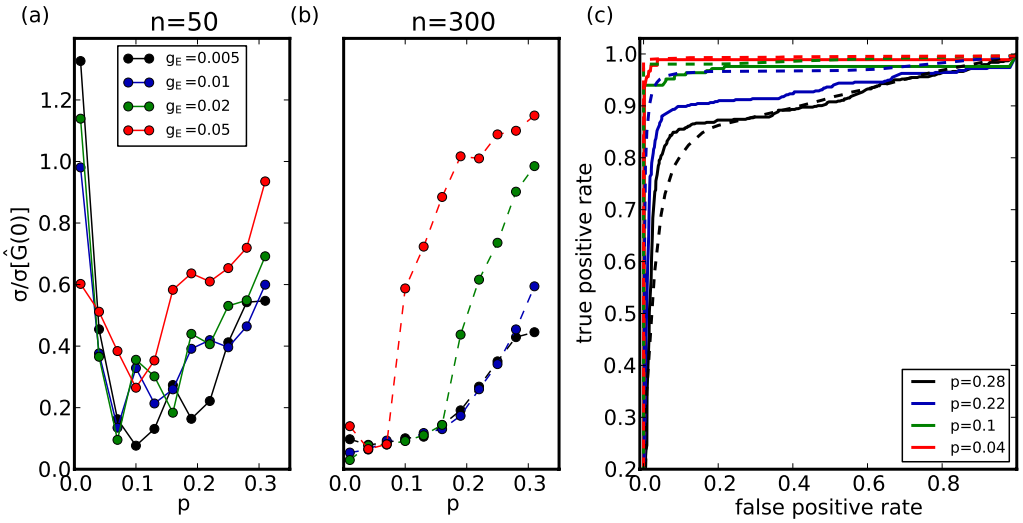

Figure 3. Performance of the reconstruction method depending on the connection probability p and different coupling strengths g_E for (a) $n = 50$ and (b) $n = 300$. The standard error of reconstructed couplings normalized by the standard deviation of the original couplings in the network is plotted. Reconstruction works best for small couplings and relatively small connection probability. (c) The ROC curves ($g_E = 0.01$) also demonstrate a good performance for small p. Solid lines: $n = 50$, dashed lines: $n = 300$.

Γ. After this number of steps, the cost decrease was very slow and the process was ended even if full convergence had not been reached. The number of steps therefore constitutes a compromise between precision of the final result and computational cost. It was kept fixed to limit the computation time of the parameter scans. To implement simulated annealing, transformations resulting in cost increase were also accepted with probability $e^{-\Delta\Gamma/T}$, where $\Delta\Gamma$ was the resulting cost difference and T a 'temperature' parameter. We set $T = 10^{-4}$, and T was decreased by 1% after $n/2 \times 10^4$ steps. The parameter T was chosen such that the number of accepted transformations increased by a small fraction, and the remaining cooling parameters to ensure a sufficiently slow speed of cooling with respect to the convergence of the cost function, without increasing the rate of convergence too much.

In figure 3, panels (a) and (b) show the standardized error of estimated connection strengths for a range of values for the parameters p, n and g_E. The performance is best for low, but not too low, connection probabilities. As expected, the connectivity can neither be estimated well in very sparse networks with few connections, nor in dense networks. For very sparse networks, not enough indirect contributions to the covariances exist in order to infer the direction of the connections. For dense networks, the cost function favoring sparse connectivity cannot be expected to lead to the original connectivity matrix. The performance is better for the larger network, as long as the connection weights are not too high. Qualitatively similar results emerge when using other measures for the performance, such as, for example, the area under the receiver-operating characteristic (ROC) curve.

In figure 3(c) ROC curves are plotted. By comparison of a thresholded version of the estimated weight matrix with the original adjacency matrix, the rates of correctly

classified connections were calculated. The depicted curves correspond to the error rates for thresholds between 0 and the maximum coupling value that was observed. These curves confirm that connectivity matrices with small (but not too small) connection probabilities are reconstructed best. Additionally, the reconstruction tends to be somewhat better in the larger network. Note that the covariance matrices in these examples are real and symmetric, so that the direction of connections cannot be observed directly.

6.2. Connectivity in simulated neural networks

We tested our method on simulated data from networks of leaky integrate-and-fire (LIF) neurons. In this model, the membrane potential V_k of neuron k is described by

$$\tau_{\mathrm{m}}\dot{V}_k(t) = -V_k(t) - \tau_{\mathrm{m}}\theta s_k(t) + \tau_{\mathrm{m}}\sum_j g_{kj}s_j(t-d) + RI_{\mathrm{ext}}.$$

A spike is evoked if the membrane potential reaches a threshold at $\theta = 20$ mV, after which the potential is reset to 0 mV, and the neuron remains refractory for 2 ms. Spiking input from presynaptic neurons results in δ-shaped currents, $s_k(t) = \sum_j \delta(t - t_j^k)$, where the t_j^k are the spike times of neuron k. A synaptic transmission delay of $d = 2$ ms was used. The connection weights g_{kj} from neuron j to neuron k were 0.2 mV for excitatory connections and -1.3 mV for inhibitory neurons. This resulted in an approximate ratio of 1:5 between effective excitatory and inhibitory couplings, as was used in the simulations before. The membrane time constant was set to $\tau_{\mathrm{m}} = 20$ ms. Neurons received additional external input from a Poisson process I_{ext} with rate $\theta/(g_E\tau_{\mathrm{m}})$ and weight resulting in membrane potential jumps of the size g_E for the input resistance R. Thus, the input was just strong enough to bring the mean membrane potential to threshold. 80% of the neurons were excitatory, 20% inhibitory. Due to its strong inhibitory connections, the network was in an asynchronous–irregular state [36]. Covariance functions were computed with a maximum time lag of 200 ms and a temporal resolution of 10 ms. A discrete Fourier transform was applied to the resulting histograms to obtain the measured covariance functions $\hat{C}(\omega)$.

Figure 4 shows the performance of the numerical procedure described in section 6.1, measured by ROC curves, corresponding to the scenario of low connection probability and weak couplings, cf figure 3. In this slightly more realistic scenario, information about directed connections can likewise be obtained from the matrix of count covariances $\hat{C}(0)$. Nonetheless, the amount of random fluctuations in the measured covariances compromises the performance significantly when compared to the idealized case discussed in the previous section. The noise in the estimated covariances is due to both deviations from the linear approximation and the limited amount of data in the finite-length simulations. The performance is, however, much better if information from all frequency components is used, cf equation (6). This shows that noise in different frequency bands is sufficiently independent to be averaged out.

7. Discussion

We showed that, contrary to intuition, it is possible to infer directed connections directly from a matrix of pairwise covariance functions.

A simple random search algorithm minimizing the L_1 norm already delivers quite reliable results. The disadvantage of this algorithm is that the convergence of the

J. Stat. Mech. (2013) P03008

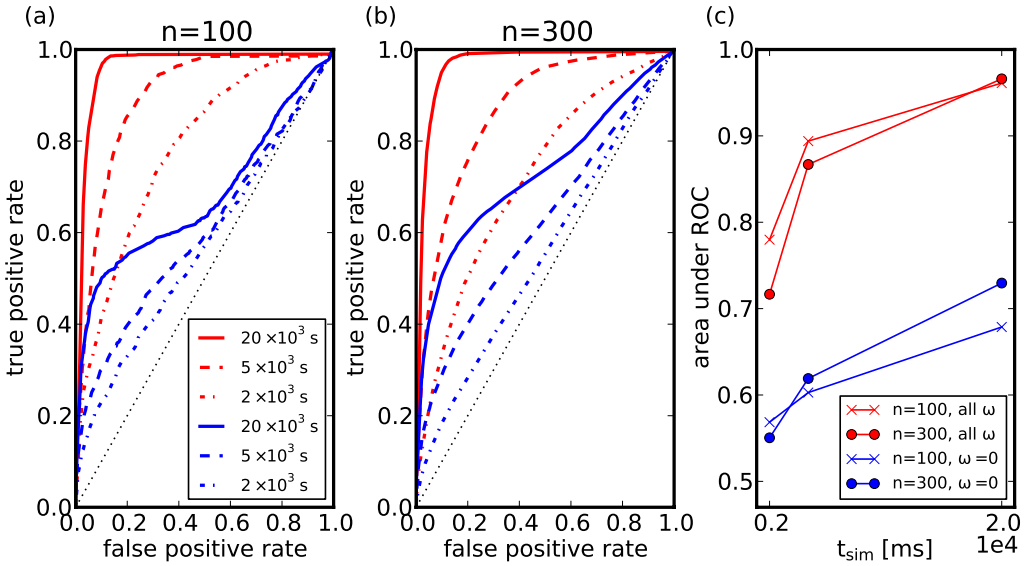

Figure 4. The quality of the reconstruction from simulated LIF networks of size (a) $n = 100$ and (b) $n = 300$, as measured by ROC curves. Covariances were estimated from spike trains recorded over different simulation times (see the legend). The quality of the reconstruction increases with the simulation time as the noise decreases in the covariance estimates. Blue: estimation on the basis of $C(\omega = 0)$ (count covariances) results in better than random predictions (dotted line). Red: combined estimation on the basis of all $C(\omega)$ leads to a performance comparable to the noise-free case in figure 3. A maximum time lag of 200 ms and 10 ms resolution for the covariance functions resulted in 21 different values of ω. (c) Direct comparison of the areas under the ROC curves.

optimization is very slow for large networks. Furthermore, due to the non-convexity of the problem, convergence to a global optimum cannot be guaranteed.

As is to be expected for a method relying on sparseness, the performance degrades for dense networks. Because the number of indirect connections then increases, estimation of direct connectivity becomes more difficult. This might, in fact, not be a problem specific to the approach described here. Minimization of the L_1 norm is an established method that favors sparse coefficients. Applications include the estimation of sparse regression coefficients [30], sparse principal component analysis [37] and compressive sensing [32, 31], where signals can be recovered from a seemingly insufficient amount of data. In these studies the constraints are linear, while in our work the constraints are quadratic, which leads in the case of the decomposed covariance matrix to solutions connected by unitary transformations. The non-differentiability of the L_1 norm makes analytic results regarding the performance difficult, but note for example [38], where the dependence of the reconstruction performance for compressed sensing on the degree of sampling and signal sparsity has been studied.

In order to meaningfully interpret inferred connections from a linearized model as physical connections, it is necessary that a linear approximation is applicable. In networks of integrate-and-fire neurons in an asynchronous and irregular state, for instance, this seems to be a reasonable assumption [6, 4].

J. Stat. Mech. (2013) P03008

Methods based on covariances in the form of partial coherences have been studied intensively [13, 18] and have also been generalized to nonlinear systems [14, 15]. An important difference in our contribution is that, by virtue of a minimization of the L_1 norm, spurious connections due to shared input and direct connections within the observed network are suppressed, such that the estimate is directly related to the matrix of linear couplings. This is not the case for the partial coherence, where in particular the 'marrying parents to joint children' effect occurs [13, 39]. The measure for connection strength that we obtain is related to the partial directed coherence [18]. However, in the current approach it is not necessary to explicitly fit an autoregressive process to the data. In strictly linear models, non-zero couplings are equivalent to the corresponding nodes being linearly Granger-causally related [40]. Granger causality is closely related to the concept of transfer entropy, an approach based on information theory [11, 9, 41]. It does not rely on any specific model, and has been shown to perform well in practice [12]. On the other hand, it is more difficult to interpret the resulting connection strength, and to differentiate between excitatory and inhibitory connections. Pairwise interactions derived from maximum entropy models [23] can equally be adapted to be sparse for greater learning efficiency and can be related to integrated covariances [17]. Connections resulting from indirect interactions are avoided, but the obtained couplings are undirected.

Methods that fit explicit models of the dynamics to the data can also deal with non-stationary activity [20, 42]. In [21], a generalized linear model (GLM) was employed, and in addition to a prior on the L_1-norm, Dale's law was imposed. Direct use of compressed sensing has also been proposed in [43] for incomplete anatomical data. The maximum likelihood optimization in the connectivity estimation from parameters in GLMs involves only a convex optimization [24, 25] and can therefore be computed efficiently, making it possible to decipher detailed spatio-temporal response properties. As the full information present in spike trains is used, some of the indeterminacy stemming from indirect recurrent connections is avoided; additionally regularization can be used to promote sparse models. In comparison, the relative simplicity of the linear model makes the ambiguity more explicit and provides a way to evaluate the information contained in the common measure of covariances.

An advantage of our approach is that precise spike times are not needed. Moreover, there are no constraints on the time resolution of the signal, up to the point that only covariances integrated over very long times can be used. Nonetheless, a higher time resolution of the covariance function improves the results, as averaging over reconstructed coupling matrices at different frequencies reduces the noise. An explicit measure of connection strength is returned, making it, in particular, possible to differentiate between excitatory and inhibitory connections. Finally, since covariance functions result from time averages over pairs of spike trains, it is not necessary to handle large amounts of data simultaneously. The method is easily applicable to networks of a few hundred nodes. For larger networks, though, the slow convergence of the algorithm results in high computational costs. Another drawback is that relatively long observation times are needed to ensure a low noise level in the covariance functions.

In the numerical examples, we used homogeneous connection weights. Small heterogeneities in the weights will not change the errors in the prediction of the couplings. Larger heterogeneities, like for example a distribution of weights between 0 and some larger value, will make it more difficult to distinguish connections of small weights from non-

J. Stat. Mech. (2013) P03008

existing connections, especially if the measurements are noisy. The same can be expected for heterogeneities in the dynamics of neurons, as long as the approximation of linear couplings can still be justified. The method will probably fail if connections are sparse only in the sense that a few strong connections and a dense connectivity of very weak connections exist.

We have not approached the problem of incompletely sampled networks in this work. Unobserved nodes can be expected to give rise to spurious connections and are difficult to infer [44], but preliminary results show that strong connections can still be detected with reasonable acuity for some degree of undersampling. If it cannot be assumed that a large part of the network is observed and the external input is uncorrelated, an additional degree of ambiguity to the one studied in this work arises, as all correlations that are used to infer connections can potentially be induced from the outside, see for example [26].

We applied our technique to simulated spike times, but the method is not restricted to this kind of data and could, for example, also be tested on covariance functions obtained from fMRI data in a resting state. In principle, the theory can be applied to data from all systems with sparse and weak interactions where a linear approximation to the dynamics seems reasonable.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF; grant 01GQ0420 'BCCN Freiburg' and grant 01GQ0830 'BFNT Freiburg∗Tübingen'), and the German Research Foundation (DFG; CRC 780, subproject C4).

References

[1] Brunel N and Hakim V, *Fast global oscillations in networks of integrate-and-fire neurons with low firing rates*, 1999 *Neural Comput.* **11** 1621–71
[2] Galán R F, *On how network architecture determines the dominant patterns of spontaneous neural activity*, 2008 *PLoS One* **3** e2148
[3] Ostojic S, Brunel N and Hakim V, *How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains*, 2009 *J. Neurosci.* **29** 10234–53
[4] Trousdale J, Hu Y, Shea-Brown E and Josić K, *Impact of network structure and cellular response on spike time correlations*, 2012 *PLoS Comput. Biol.* **8** e1002408
[5] Pernice V, Staude B, Cardanobile S and Rotter S, *How structure determines correlations in neuronal networks*, 2011 *PLoS Comput. Biol.* **7** e1002059
[6] Pernice V, Staude B, Cardanobile S and Rotter S, *Recurrent interactions in spiking networks with arbitrary topology*, 2012 *Phys. Rev.* E **85** 031916
[7] de La Rocha J and Doiron B, *Correlation between neural spike trains increases with firing rate*, 2007 *Nature* **448** 802–6
[8] Tetzlaff T, Helias M, Einevoll G T and Diesmann M, *Decorrelation of neural-network activity by inhibitory feedback*, 2012 *PLoS Comput. Biol.* **8** e1002596
[9] Bressler S L and Seth A K, *Wiener-Granger causality: a well established methodology*, 2011 *Neuroimage* **58** (2) 323–9
[10] Eldawlatly S, Zhou Y, Jin R and Oweiss K G, *On the use of dynamic Bayesian networks in reconstructing functional neuronal networks from spike train ensembles*, 2010 *Neural Comput.* **22** 158–89
[11] Schreiber T, *Measuring information transfer*, 2000 *Phys. Rev. Lett.* **85** 461–4
[12] Stetter O, Battaglia D, Soriano J and Geisel T, *Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals*, 2012 *PLoS Comput. Biol.* **8** e1002653

J. Stat. Mech. (2013) P03008

[13] Dahlhaus R, Eichler M and Sandkühler J, *Identification of synaptic connections in neural ensembles by graphical models*, 1997 *J. Neurosci. Methods* **77** 93–107
[14] Schelter B, Winterhalder M, Dahlhaus R, Kurths J and Timmer J, *Partial phase synchronization for multivariate synchronizing systems*, 2006 *Phys. Rev. Lett.* **96** 208103
[15] Nawrath J, Romano M C, Thiel M, Kiss I Z, Wickramasinghe M, Timmer J, Kurths J and Schelter B, *Distinguishing direct from indirect interactions in oscillatory networks with multiple time scales*, 2010 *Phys. Rev. Lett.* **104** 038701
[16] Cocco S, Leibler S and Monasson R, *Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods*, 2009 *Proc. Nat. Acad. Sci.* **106** 14058–62
[17] Ganmor E, Segev R and Schneidman E, *The architecture of functional interaction networks in the retina*, 2011 *J. Neurosci.* **31** 3044–54
[18] Baccalá L A and Sameshima K, *Partial directed coherence: a new concept in neural structure determination*, 2001 *Biol. Cybern.* **84** 463–74
[19] Roudi Y, Tyrcha J and Hertz J, *Ising model for neural data: Model quality and approximate methods for extracting functional connectivity*, 2009 *Phys. Rev.* E **79** 051915
[20] Shandilya S G and Timme M, *Inferring network topology from complex dynamics*, 2011 *New J. Phys.* **13** 013004
[21] Mishchenko Y, Vogelstein J T and Paninski L, *A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data*, 2011 *Ann. Appl. Stat.* **5** 1229–61
[22] Van Bussel F, Kriener B and Timme M, *Inferring synaptic connectivity from spatio-temporal spike patterns*, 2011 *Front. Comput. Neurosci.* **5** 3
[23] Monasson R and Cocco S, *Fast inference of interactions in assemblies of stochastic integrate-and-fire neurons from spike recordings*, 2011 *J. Comput. Neurosci.* **31** 199–227
[24] Paninski L, *Maximum likelihood estimation of cascade point-process neural encoding models*, 2004 *Network* **15** 243–62
[25] Pillow J W, Shlens J, Paninski L, Sher A, Litke A M, Chichilnisky E J and Simoncelli E P, *Spatio-temporal correlations and visual signalling in a complete neuronal population*, 2008 *Nature* **454** 995–9
[26] Vidne M, Ahmadian Y, Shlens J, Pillow J W, Kulkarni J, Litke A M, Chichilnisky E J, Simoncelli E and Paninski L, *Modeling the impact of common noise inputs on the network activity of retinal ganglion cells*, 2012 *J. Comput. Neurosci.* **33** 97–121
[27] Hawkes A G, *Point spectra of some mutually exciting point processes*, 1971 *J. R. Stat. Soc.* B **33** 438–43
[28] Lindner B, Doiron B and Longtin A, *Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback*, 2005 *Phys. Rev.* E **72** 061919
[29] Yuan M and Lin Y, *Model selection and estimation in the Gaussian graphical model*, 2007 *Biometrika* **94** 19–35
[30] Tibshirani R, *Regression shrinkage and selection via the lasso*, 1996 *J. R. Stat. Soc.* B **58** 267–88
[31] Donoho D L, *Compressed sensing*, 2006 *IEEE Trans. Inform. Theory* **52** 1289–306
[32] Candes E J and Romberg J K, *Stable signal recovery from incomplete and inaccurate measurements*, 2006 *Commun. Pure Appl. Math.* **59** 1207–23
[33] Candes E J and Tao T, *Decoding by linear programming*, 2005 *IEEE Trans. Inform. Theory* **51** 4203–15
[34] Ganguli S and Sompolinsky H, *Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis*, 2012 *Annu. Rev. Neurosci.* **35** 485–508
[35] Meyer C D, 2000 *Matrix Analysis and Applied Linear Algebra* (Philadelphia: Society for Industrial and Applied Mathematics)
[36] Brunel N, *Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons*, 2000 *J. Comput. Neurosci.* **8** 183–208
[37] Zou H and Hastie T, *Sparse principal component analysis*, 2006 *J. Comput. Graph. Stat.* **15** 265–86
[38] Ganguli S, *Statistical mechanics of compressed sensing*, 2010 *Phys. Rev. Lett.* **104** 188701
[39] Dahlhaus R, *Graphical interaction models for multivariate time series*, 2000 *Metrika* **51** 157–72
[40] Schelter B, Timmer J and Eichler M, *Assessing the strength of directed influences among neural signals using renormalized partial directed coherence*, 2009 *J. Neurosci. Methods* **179** 121–30
[41] Barnett L, *Granger causality and transfer entropy are equivalent for gaussian variables*, 2009 *Phys. Rev. Lett.* **103** 238701
[42] Roudi Y and Hertz J, *Mean field theory for nonequilibrium network reconstruction*, 2011 *Phys. Rev. Lett.* **106** 048702
[43] Mishchenko Y and Paninski L, *A bayesian compressed-sensing approach for reconstructing neural connectivity from subsampled anatomical data*, 2012 *J. Comput. Neurosci.* **33** 371–88
[44] Nykamp D, *Pinpointing connectivity despite hidden nodes within stimulus-driven networks*, 2008 *Phys. Rev.* E **78** 021902

J. Stat. Mech. (2013) P03008