

# Consistent estimation of complete neuronal connectivity in large neuronal populations using sparse “shotgun” neuronal activity sampling

Yuriy Mishchenko

the date of receipt and acceptance should be inserted later

**Abstract** We investigate the properties of the recently proposed “shotgun” sampling approach for the common inputs problem in the functional estimation of neuronal connectivity. We study the asymptotic correctness, the speed of convergence, and the data size requirements of such an approach. We find that the shotgun approach can be expected to allow the inference of the complete connectivity matrix in large neuronal populations under some rather general conditions. However, we find that the posterior error of the shotgun connectivity estimator may grow quickly with the size of the unobserved neuronal populations, the connectivity strength, and the square of the observations’ sparseness. This implies that the shotgun connectivity estimation will require significant amounts of neuronal activity data whenever the number of neurons in the observed populations is small. We present a numerical approach for solving the shotgun estimation problem in general settings and use it to demonstrate the shotgun connectivity inference in simulated synfire and weakly coupled cortical neuronal networks.

**Keywords** functional connectivity, neuronal circuit reconstruction, calcium imaging, neuronal population activity

## 1 Introduction

Recent advances in multi-neuronal activity recordings using calcium-sensitive fluorescence imaging have made it possible to image the activity of large neuronal populations over extended periods of time (Tsien, 1989; Yuste et al., 2006; Cossart et al., 2003; Ohki et al., 2005; Reddy et al., 2008a; Grewe et al., 2010). Bulk-loading of organic calcium-sensitive fluorescent dyes offers the fluorescent signal-to-noise ratio (SNR) sufficient for resolving individual action potentials (spikes) of neurons (Yuste et al., 2006; Stosiek et al., 2003) and genetically encoded calcium indicators are approaching the SNR levels necessary for neuronal activity imaging with single spike accuracy (Wallace et al., 2008). Modern cooled CCD-microscopy can allow imaging of calcium fluorescence in neuronal populations in-vitro with frame-rates of over 60 Hz (Djurisic et al., 2004) and 2-photon laser scanning microscopy offers similar frame-rates in-vivo (Iyer et al., 2006; Salome et al., 2006; Reddy et al., 2008b; Cotton et al., 2013; Theis et al., 2015). These advances now allow studying the single-cell structure of neuronal circuits in the brain using accurate statistical approaches (Pillow et al., 2008; Stevenson et al., 2008a; Stevenson et al., 2008b; Stevenson et al., 2009; Mishchenko et al., 2011).

One of the biggest challenges of the functional analysis of the neuronal connectivity in the brain remains the presence of unobserved or hidden inputs in the recordings of neuronal population activity (Nykamp, 2007; Pillow and Latham, 2007; Vidne et al., 2009). Because func-

---

Yuriy Mishchenko  
Toros University, Mersin, Turkey  
Tel: +90(324)3253300  
Fax: +90(324)3253301  
E-mail: yuriy.mishchenko@gmail.com

tional connectivity estimation relies on correlating the activity of different neurons in a neuronal population over extended periods of time, the presence of unobserved inputs can contribute errors to functional analysis. In particular, the well-known “common inputs” problem causes one to mistake the correlations in the activity of different neurons caused by the correlations in their unobserved inputs for a direct connection between those neurons (Nykamp, 2005; Nykamp, 2007; Kulkarni and Paninski, 2007; Pillow and Latham, 2007). Despite rapid progress in experimental neuronal population activity imaging techniques, the simultaneous observation of the activity of all neurons even in the smallest of neuronal circuits is currently not plausible, and the development of robust analytical and computational techniques for overcoming the hidden inputs problem remains one of the important open questions of computational neuroscience (Nykamp, 2005; Nykamp, 2007; Kulkarni and Paninski, 2007; Pillow and Latham, 2007; Vidne et al., 2009; Keshri et al., 2013).

Recently a promising approach for overcoming the hidden inputs problem—the shotgun sampling—had been proposed in (Turaga et al., 2013; Keshri et al., 2013). In this approach, neurons in a large neuronal population are proposed to be imaged in small random groups, whereas the connectivity matrix of the complete neuronal population is assembled statistically by combining the information about the neuronal connectivity from different such partial measurements. The shotgun approach offers the possibility for reconstructing the connectivity of large neuronal circuits by using limited imaging resources, without the need to simultaneously image the entire neuronal circuits.

In this paper, we perform a systematic analysis of certain aspects of the shotgun sampling proposal such as the asymptotic correctness of the connectivity estimation, the speed of convergence, and the necessary data sizes. It may not be clear at first if the shotgun connectivity estimation is really free from the hidden inputs problem. One may observe that, even as all neurons in a neuronal population do get imaged with this approach over different points of time, these observations still fail to provide the complete input-output relationships of even a single neuron in the population: the total number of neurons that need to be simultaneously imaged to provide all the inputs of even a single neuron in mammalian cortex can be as high as

10,000. Without having the information about the complete set of inputs and outputs of any single neuron, one may wonder if the hidden inputs problem is really resolved. Furthermore, if the shotgun approach does allow the unambiguous determination of the complete connectivity matrix, what are the trade-offs that had been made? In particular, what is the minimal imaging time required for a given accuracy of the connectivity matrix reconstruction and how does this time scale with the size of the unobserved populations, observation sparseness, and other parameters?

Here, we show that the shotgun approach can be expected to recover the complete neuronal connectivity matrix in general neuronal activity models under some rather general conditions. We calculate the speed of convergence of the shotgun connectivity estimator and show that the imaging time required to achieve a given estimation accuracy scales with the number of neurons in the unobserved neuronal populations as well as the inverse square of the fraction of neurons observed during one imaging trial. This scaling is inopportune for the reconstructions of neuronal connectivity in the situations, where the number of unobserved neurons remains large. We further discuss a numerical approach for solving the connectivity estimation problem from shotgun sampling data and use it to demonstrate the shotgun connectivity estimation in small model neuronal populations.

In this paper, we focus specifically on the problem of the inference of the underlying neuronal connectivity matrix from partial neuronal population activity observations. In the case of calcium imaging—currently the most plausible modality for collecting large scale neuronal population activity data—another important component of such inference is the deconvolution of the calcium signal into the underlying neuronal spike trains. In recent years, significant progress had been made with respect to the deconvolution of the calcium imaging signal, for example, see (Vogelstein et al., 2009) and (Vogelstein et al., 2010), and (Theis et al., 2015) for a survey and comparison of recent deconvolution approaches. Unfortunately, recent experimental work had indicated that the existing calcium signal deconvolution approaches still may need significant improvement to be practically usable (Cotton et al., 2013; Theis et al., 2015). In this paper, we do not consider

this important aspect of the analysis of functional data. Instead, we focus on the mathematical and computational problem of specifically the shotgun estimation, assuming that the deconvolved neuronal activity data is already available (Section 2.4) or that an effective neuronal activity model based on the continuous calcium fluorescence signal can be used (Section 2.2). However, in the future the incorporation of the calcium imaging deconvolution problem into and the impact of this problem onto the inference of neuronal connectivity will require more detailed investigation, including the understanding of the impact of the added uncertainties in the spike timing (as deduced from a lower frame-rate calcium imaging data), the apparently high fractions of lost spikes and the tendency to overestimate bursting and underestimate isolated neuronal spiking activity (Cotton et al., 2013; Theis et al., 2015).

In the remainder of the paper our presentation is organized as follows. In Materials and Methods, Section 2.1, we provide a general overview of the shotgun neuronal connectivity estimation problem and offer its mathematical formalization. In Section 2.2, we investigate the shotgun connectivity estimation problem in a linear neuronal population activity model. We analytically obtain the properly marginalized observations log-likelihood in this model in asymptotic limit and derive the conditions under which the maximum likelihood estimator in this model is consistent. Furthermore, we explicitly derive the sufficient conditions for the consistency of the shotgun connectivity estimation in general spiking models of neuronal activity, where the rate of neuronal spiking is described as a general function of linearly summed inputs. We also analytically demonstrate and discuss the impact of the hidden inputs problem on the connectivity estimation in the linear model. In Section 2.3, we study the shotgun connectivity estimation in general settings and propose an exact numerical algorithm for solving the associated estimation problem in general causal neuronal population activity models. In Section 2.4, we discuss the numerical simulations performed in this paper.

In Results, Section 3.1, we generalize the results of Section 2.2 to show that the shotgun maximum likelihood estimator can be proved more generally to converge to the true connectivity matrix of a neuronal population as long as certain conditions are met by the set of the

neuronal subpopulations imaged by the shotgun sampling. In Section 3.2, we use this result to propose a different organization of the shotgun neuronal activity sampling that may be more advantageous for experimental realization. In Sections 3.3 and 3.4, we discuss the results of the shotgun connectivity estimation in simulated linear and generalized linear neuronal models. Discussion and conclusions follow in Section 4.

## 2 Materials and Methods

### 2.1 The shotgun sampling approach for neuronal connectivity estimation

In the shotgun sampling, one’s objective is to recover the effective connectivity matrix of a large neuronal population using a collection of partial samples of the activity of that population’s different subpopulations. One simple way to visualize this idea is to think about the reconstruction of a large image by using small fragments of that image from the image’s different locations.

More formally, we define the effective connectivity matrix  $\mathcal{W}$  as a parameter of a statistical model of neuronal population activity,  $P(\mathcal{X}|\mathcal{W})$ , where  $\mathcal{X}$  stands for the raster of the historical activities of all neurons in a neuronal population, and  $P(\mathcal{X}|\mathcal{W})$  is the model likelihood of observing a particular raster  $\mathcal{X}$ . In network models,  $\mathcal{W}$  is typically a matrix  $\mathcal{W} = (w_{ij})$  having as many rows and columns as there are neurons in the population,  $N$ . Each element  $w_{ij}$  characterizes the impact of the past activity of one “input” neuron  $j$  on the activity of one “output” neuron  $i$ . Our objective here, therefore, is to estimate the complete matrix  $\mathcal{W}$  given a series of partial observations of the activity  $\mathcal{X}$ .

In the context of this problem we introduce the following notation. We denote the raster of the activities of the entire neuronal population (that is, both observed and hidden) over all observations by the subscript-less symbol  $\mathcal{X}$ , and we denote the activity of that population during one observation  $t$  by the symbol  $\mathcal{X}_t$ . That is,  $\mathcal{X} = \{\mathcal{X}_t, t = 1 \dots T\}$ , where  $T$  is the number of the observations. Similarly, we denote the collection of all observed neuronal activities over all observations by the subscript-less symbol  $X$ , whereas the activity of the observed neurons

during one shotgun observation  $t$  is denoted by the symbol  $X_t$ , so that  $X = \{X_t, t = 1 \dots T\}$ . Finally, the collection of all unobserved neuronal activities will be referred to by using the symbols  $Y$  and  $Y_t$ , respectively, so that  $Y = \{Y_t, t = 1 \dots T\}$ .

We formally say that the shotgun estimation problem is the problem of estimating the effective connectivity matrix  $\mathcal{W}$  of a statistical model of neuronal population activity  $P(\mathcal{X}|\mathcal{W})$  given a set of partial observations of neuronal activity  $X$ .

For the simplicity of the following discussion, in a significant part of this manuscript we will focus on the shotgun connectivity estimation formulated for a single output neuron and a randomly sampled population of neuronal inputs. In this picture, the output neuron is observed continuously in every observation, whereas the set of observed neuronal inputs changes. This assumption allows us to significantly simplify the discussion, while not leading to any significant loss of generality. The latter is because in typical network models of neuronal activity the activities of neurons are conditionally independent given the activity of the presynaptic neuronal population and the respective input connection weights. In other words, the full likelihood  $P(\mathcal{X}|\mathcal{W})$  in such models factorizes over the rows of the connectivity matrix  $\mathcal{W}$ ,

$$P(\mathcal{X}|\mathcal{W}) = \prod_{i,t} P(\mathcal{X}_{it}|\{\mathcal{X}_{t'}, t' < t\}; W_i), \quad (1)$$

where  $W_i = \{w_{ij}, j = 1 \dots N\}$ . It is possible, therefore, to perform the estimation of  $\mathcal{W}$  one row  $W_i$  at a time.

In the rest of the paper we will consistently make use of the following notation. The script symbol  $\mathcal{W}$  will always refer to the complete connectivity matrix of the entire neuronal population of  $N$  neurons, whereas the symbols  $W_i$  will refer to a single row of  $\mathcal{W}$ , corresponding to the set of all input connection weights of one neuron. The symbol  $w_{ij}$  will refer to one connection weight between an output neuron  $i$  and an input neuron  $j$ .

The script symbols  $\mathcal{X}$  and  $\mathcal{X}_t$  will be always used to refer to the neuronal activity of the entire neuronal population, whereas the plain symbols  $X$ ,  $X_t$ ,  $Y$ , and  $Y_t$  will be used to denote the observed and the unobserved parts of the neuronal activity, respectively. We shall occasionally make use of the symbols  $X_t$  and  $Y_t$

to refer to the set of neurons contained in  $X_t$  or  $Y_t$ , as opposed to their activities. This distinction will always be made clear by the context. Finally,  $\mathcal{X}_t$ ,  $X_t$  and  $Y_t$  will be understood to be column-vectors, while  $W_i$  will always be a row-vector.

## 2.2 The shotgun connectivity estimation in linear neuronal population activity model

### 2.2.1 The linear model of neuronal activity

The linear model of neuronal activity has the advantage here of allowing the analysis of the shotgun connectivity estimation approach to be carried out analytically. The primary reason warranting the inspection of this model here, thus, is its analytical tractability.

In the linear neuronal activity model, we model the activity of a neuronal population using a continuous variable  $\mathcal{X}_t$ , which can be thought, for example, to represent suitably smoothed firing rates of different neurons or the standard  $\Delta F/F$  calcium imaging measure of neuronal activity. The input-output relationships between these variables is assumed to be linear,

$$\mathcal{X}_{t+1} = \mathcal{W}\mathcal{X}_t + \epsilon_t. \quad (2)$$

Thus, here  $\mathcal{X}_t$  is an  $N$ -element column-vector representing the activity of the population of  $N$  neurons, while the index  $t$  plays the role of discrete time. The parameter  $\mathcal{W}$  is a  $N \times N$  connectivity matrix, and  $\epsilon$  is a i.i.d. normal random noise variable. Without loss of generality, we assume  $\epsilon$  to have the variances of one. The model probability density  $P(\mathcal{X}|\mathcal{W})$  then would be defined by,

$$P(\mathcal{X}|\mathcal{W}) \propto \exp\left(-\sum_t (\mathcal{X}_{t+1} - \mathcal{W}\mathcal{X}_t)^2/2\right). \quad (3)$$

The problem of estimating the connectivity matrix  $\mathcal{W}$  is simple, if the activity of the entire neuronal population is observed. In that case, we consider the observations' log-likelihood,

$$\log P(\mathcal{X}|\mathcal{W}) = \sum_t \{\mathcal{X}_{t+1}^T \mathcal{W} \mathcal{X}_t - Tr[\mathcal{W} \mathcal{X}_t \mathcal{X}_t^T \mathcal{W}^T]/2\} + const. \quad (4)$$

The maximum likelihood estimation (MLE) of  $\mathcal{W}$  is found as the maximum of Eq. (4). One

can see rather immediately that such MLE converges to the true connectivity matrix whenever the matrix  $\sum_t \mathcal{X}_t \mathcal{X}_t^T$  is of full rank.

In the partial observations case, a similar program can be pursued but one has to properly consider now the marginalized likelihood  $P(X|\mathcal{W}) = \int dY P(X, Y|\mathcal{W})$ , where the unobserved neural activity variables  $Y$  had been integrated out. For the linear neuronal activity model, we can calculate this likelihood analytically (that is, in the asymptotic limit of large numbers of observations) and study the properties of such MLE explicitly.

### 2.2.2 The calculation of the partial observations likelihood in the linear neuronal activity model

In this section, we aim to calculate the marginalized observations likelihood  $P(X|\mathcal{W})$  for the shotgun connectivity estimation in the linear neuronal activity model. We specifically focus on the input-output form of the model (2), with a scalar output variable  $Z_{it}$  and two vector input variables  $X_t$  and  $Y_t$ , defined by the relationship,

$$Z_{it} = W_{iX_t} X_t + W_{iY_t} Y_t + \epsilon_{it}. \quad (5)$$

Here,  $Z_{it}$  represents the ‘‘output’’ activity of one neuron  $i$  at time  $t+1$ , while  $X_t$  with  $Y_t$  represent the activities of the observed and the unobserved ‘‘input’’ neuronal populations at time  $t$ .  $W_i$  is the row of the complete connectivity matrix  $\mathcal{W}$  specifying the input connection weights of the output neuron  $i$ , and  $W_{iX_t}$  and  $W_{iY_t}$  are the parts of  $W_i$  corresponding to the observed and the unobserved inputs  $X_t$  and  $Y_t$ , respectively.  $\epsilon_{it}$  is an i.i.d. zero mean and unit variance normal random variable representing noise.

The inputs  $X_t$  and  $Y_t$  are assumed to be drawn jointly from a distribution  $P(X_t, Y_t)$ . Understanding that  $X_t$  and  $Y_t$  represent the activity of the neuronal population in the model given by Eq. (2), the distribution  $P(X_t, Y_t)$  can be taken as the stationary distribution of the Markov process defined by Eq. (2). This distribution is Gaussian. By suitably offsetting the variables  $\mathcal{X}_t$  in Eq.(2), it is always possible to make this distribution have vanishing first moments. However, the covariance matrix of this distribution,  $\Sigma$ , in general will be nontrivial.

It shall be noted that in the above the connectivity vector  $W_i$  and the covariance matrix

$\Sigma$  are both the parameters of the model (5). However, it can be also observed that  $\Sigma$  is not an independent parameter per-se since, by virtue of  $P(X_t, Y_t)$  being the stationary distribution of (2), it depends on the connectivity matrix  $\mathcal{W}$ . We will ignore this point for now, assuming no connection between  $\mathcal{W}$  and  $\Sigma$ , as this will have no impact on the complexity of the calculations that we will need to perform, while a more general result can be obtained by ignoring this connection. However, see Section 2.2.5 for some important repercussions.

The specific quantity of interest in this section is the average log-likelihood of the observed data  $Z_i$  and  $X$  in model (5) (in the sense of the compositional estimation of  $\mathcal{W}$  one output neuron at a time, see Section 2.1), marginal over the missing data represented by the hidden input variables  $Y$ ,

$$l(W_i, \Sigma|Z_i, X) = \frac{1}{T_i} \log P(Z_i, X|W_i, \Sigma), \quad (6)$$

where  $P(Z_i, X|W_i, \Sigma)$  is the marginal distribution of the observed variables in model (5),  $P(Z_i, X|W_i, \Sigma) = \int dY P(Z_i, X, Y|W_i, \Sigma)$ . Note that  $Z_i$  and  $X$  in Eq. (6) are the collections of all observations  $(Z_{it}, X_t)$  for all  $t \in \{t_1, t_2, \dots, t_{T_i}\}$  such that the output neuron  $i$  is observed, per the general conventions of Section 2.1. Therefore, in the RHS of Eq. (6)  $Z_i$ ,  $X$ , and  $P(Z_i, X|W_i, \Sigma)$  are all dependent on  $T_i$  in this manner. Also note that both  $Z_i$  and  $X$  are the observed data.

When the number of observations  $T_i$  is large, the RHS of Eq. (6) converges in probability to the expectation value of  $\log P(Z_{it}, X_t|W_i, \Sigma)$  under the true distribution of the observed inputs and outputs  $P(Z_{it}, X_t)$ ,

$$\begin{aligned} l(W_i, \Sigma|Z_i, X) &= \frac{1}{T_i} \log P(Z_i, X|W_i, \Sigma) \\ &= \frac{1}{T_i} \sum_{t=1}^{T_i} \log P(Z_{it}, X_t|W_i, \Sigma) \\ &\approx E_{P(Z_{it}, X_t)}[\log P(Z_{it}, X_t|W_i, \Sigma)], \end{aligned} \quad (7)$$

in which we recognize the expected log-likelihood function,

$$l(W_i, \Sigma) \approx E[\log \int dY_t P(Z_{it}, X_t, Y_t|W_i, \Sigma)], \quad (8)$$

where the expectation, again, is taken with respect to the true density of the observed input and output variables,  $X_t$  and  $Z_{it}$ , and we used

$$P(Z_{it}, X_t|W_i, \Sigma) = \int dY_t P(Z_{it}, X_t, Y_t|W_i, \Sigma). \quad (9)$$

We see that it is necessary for us now to calculate,

$$P(Z_{it}, X_t | W_i, \Sigma) = \int dY_t \exp \left( - (Z_{it} - W_{iX_t} X_t - W_{iY_t} Y_t)^2 / 2 - \mathcal{X}_t^T \Sigma^{-1} \mathcal{X}_t / 2 + const \right), \quad (10)$$

where  $\mathcal{X}_t$  is the vector of the complete input activities formed by suitably combining  $X_t$  and  $Y_t$ , that is,  $\mathcal{X}_t = [X_t; Y_t]$ . The integral in Eq. (10) can be taken explicitly, although the respective calculation is complex and we move it to Appendix B. The result of this calculation can be stated as follows,

$$\log P(Z_{it}, X_t | W_i, \Sigma) = -1/2(1 + B_{it}^2)^{-1} (Z_{it} - W_{i\Sigma_* X_t} \Sigma_{X_t X_t}^{-1} X_t)^2 - 1/2 \log(1 + B_{it}^2) - X_t^T \Sigma_{X_t X_t}^{-1} X_t / 2 - 1/2 \log \det \Sigma_{X_t X_t} + const, \quad (11)$$

where the scalars  $B_{it}^2$  are defined by

$$B_{it}^2 = W_i \Sigma W_i^T - W_{i\Sigma_* X_t} \Sigma_{X_t X_t}^{-1} \Sigma_{X_t *} W_i^T = W_{iY_t} (\Sigma_{Y_t Y_t} - \Sigma_{Y_t X_t} \Sigma_{X_t X_t}^{-1} \Sigma_{X_t Y_t}) W_{iY_t}^T,$$

and the subscripted notation for  $\Sigma$  refers to the parts of  $\Sigma$  corresponding to the neuronal inputs identified in  $X_t$  and  $Y_t$ . For example,  $\Sigma_{X_t X_t}$  refers to the submatrix of  $\Sigma$  composed of all elements of  $\Sigma$  located at the intersection of the rows and the columns identified by  $X_t$ . Similarly,  $\Sigma_* X_t$  is the rectangular submatrix of  $\Sigma$  containing all the columns corresponding to the observed inputs  $X_t$ , and  $\Sigma_{X_t *}$  is a similar rectangular submatrix of all the  $X_t$ -rows of  $\Sigma$ .

Eq. (11) allows us to obtain the final expression for the expected log-likelihood  $l(W_i, \Sigma)$ ,

$$l(W_i, \Sigma) = -1/2E \left[ \frac{(Z_{it} - W_{i\Sigma_* X_t} \Sigma_{X_t X_t}^{-1} X_t)^2}{1 + B_{it}^2} + \log(1 + B_{it}^2) + X_t^T \Sigma_{X_t X_t}^{-1} X_t + \log \det \Sigma_{X_t X_t} \right], \quad (12)$$

where the expectation is again with respect to the true distribution of the observed inputs and outputs  $X_t$  and  $Z_{it}$ . Consider now the expected log-likelihood  $l(\hat{W}_i, \hat{\Sigma})$  for an estimate of the parameters  $\hat{W}_i$  and  $\hat{\Sigma}$ . We first take the average in Eq. (12) over all  $X_t$  such that the set of

neurons contained in  $X_t$  is fixed. This allows us to rewrite  $l(\hat{W}_i, \hat{\Sigma})$  in the following form,

$$l(\hat{W}_i, \hat{\Sigma}) \approx -1/2E \left\{ \frac{1 + W_i \Sigma W_i^T - 2\hat{W}_i \mathcal{A}_{X_t} W_i^T + \hat{W}_i \mathcal{A}'_{X_t} \hat{W}_i^T}{1 + \hat{B}_{it}^2} + \log(1 + \hat{B}_{it}^2) \right\} - 1/2E \left\{ \Sigma_{X_t X_t} \hat{\Sigma}_{X_t X_t}^{-1} + \log \det \hat{\Sigma}_{X_t X_t} \right\}, \quad (13)$$

where the matrices  $\mathcal{A}$  and  $\mathcal{A}'$  are defined by,

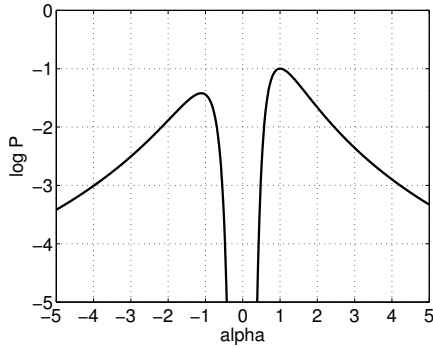
$$\mathcal{A}_{X_t} = \hat{\Sigma}_{* X_t} \hat{\Sigma}_{X_t X_t}^{-1} \Sigma_{X_t *}, \quad \mathcal{A}'_{X_t} = \hat{\Sigma}_{* X_t} \hat{\Sigma}_{X_t X_t}^{-1} \Sigma_{X_t X_t} \hat{\Sigma}_{X_t X_t}^{-1} \hat{\Sigma}_{X_t *}, \quad (14)$$

and  $W_i$  and  $\Sigma$  are the true connection weights and the true covariance matrix, respectively, and where the remaining average is over the different subsets of observed neurons  $X_t$ .

It can be verified by a direct inspection that  $\hat{W}_i = W_i$  and  $\hat{\Sigma} = \Sigma$  achieves the global maximum of Eq. (13) by separately maximizing the expressions under both expectation values for every  $X_t$ , that is,  $\hat{\Sigma}_{X_t X_t} = \Sigma_{X_t X_t}$  and  $\Sigma_{X_t *} \hat{W}_i = \Sigma_{X_t *} W_i$ . Moreover, this global maximum is unique whenever the covariance matrix  $\Sigma$  is nonsingular and the submatrices  $\Sigma_* X_t$  and  $\Sigma_{X_t X_t}$  for the sampled  $X_t$  cumulatively (but separately) tile the entirety of the matrix  $\Sigma$ . This is because the collection of individual- $X_t$  optimality conditions  $\hat{\Sigma}_{X_t X_t} = \Sigma_{X_t X_t}$  and  $\Sigma_{X_t *} \hat{W}_i = \Sigma_{X_t *} W_i$  in this case uniquely specifies  $\hat{\Sigma}$  and  $\hat{W}_i$ .

However, we also find that Eq. (13) may contain local optima different from the global maximum. For example, a second local ‘‘mirror’’ optimum can be found at  $\hat{\Sigma} = \Sigma$  and  $\hat{W}_i \approx -W_i$ , when the fraction of observed neurons in  $X_t$  is small, Figure 1. This mirror optimum can be shown to arise from the cancellation of the variations of the terms  $(1 + \hat{B}_{it}^2)^{-1}$  and  $\log(1 + \hat{B}_{it}^2)$  in Eq. (13). The difference in the expected log-likelihoods of the true and the mirror solution tends to zero with the the average fraction of the observed neurons  $X_t$ ,  $p$ , as  $\Delta l = O(p^2)$ .

The conditions above give the sufficient conditions for the consistency of the MLE in the considered neuronal activity model. We point out that the condition of tiling of  $\Sigma$  by  $\Sigma_* X_t$  and  $\Sigma_{X_t X_t}$  can be restated as the condition that the set of all observed neuronal subsets  $X_t$  cumulatively covers the entire range of possible inputs  $j = 1 \dots N$  (that is,  $\Sigma_* X_t$  tiles  $\Sigma$ ), and



**Fig. 1** An example of  $l(\hat{W}_i, \hat{\Sigma}|Z_i, X)$  for a choice of  $W_i$  and  $\Sigma = I$  along the solution ray  $\hat{W}_i = \alpha W_i$ ,  $\hat{\Sigma} = \Sigma$ . The plot shows that  $\alpha = 1$  is not a unique local maximum. The second “mirror” optimum is located at  $\alpha \approx -1$ .

the set of all  $X_t \times X_t$  covers the range of all possible input-input pairs  $\{(j, j') : j, j' = 1 \dots N\}$  (that is,  $\Sigma_{X_t X_t}$  also tiles  $\Sigma$ ). Furthermore, we can note that the fact of the full coverage of  $\Sigma$  by the collections of  $\Sigma_{*X_t}$  and  $\Sigma_{X_t X_t}$  is important and not the manner in which it is achieved. Thus, any observations organization that can provide such a coverage will be equally capable of uniquely constraining  $\mathcal{W}$ . For example, a plausible strategy for estimating the complete connectivity matrix  $\mathcal{W}$  in these settings can be to image the activity of all individual neuronal pairs  $(\mathcal{X}_{it}, \mathcal{X}_{jt})$  and  $(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$ , one pair at a time, in any order.

### 2.2.3 The hidden inputs bias in the linear neural activity model

The expected log-likelihood  $l(W_i, \Sigma)$  calculated in Section 2.2.2 can be used to provide an explicit example of the hidden inputs problem. Specifically, we consider the situation when the set of the observed input neurons  $X_t$  is kept constant throughout the experiment,  $X_t \equiv X$ . In this case, we can remove in Eq. (13) the expectation value with respect to the different sets of neurons  $X_t$ . Performing the variation with respect to  $\hat{W}_{iX}$  then yields the following equation for  $\hat{W}_{iX}$ ,

$$\hat{W}_{iX} = W_{iX} + \Sigma_{XX}^{-1} \Sigma_{XY} W_{iY} - \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY} \hat{W}_{iY}. \quad (15)$$

Eq. (15) shows that, if one has access to the incoming connection weights of the hidden neuronal population,  $\hat{W}_{iY} = W_{iY}$ , and the correct covariance matrix,  $\hat{\Sigma} = \Sigma$ , then it is possible to estimate the correct connectivity weights

$W_{iX}$  even without observing the hidden neuronal activity. However, if hidden contributions are ignored, for example, by setting  $\hat{W}_{iY} = 0$  or  $\hat{\Sigma}_{XY} = 0$ , then the hidden inputs induce a bias in the estimated neuronal connectivity, which we obtain here explicitly as,

$$E[\hat{W}_{iX} - W_{iX}] = \Sigma_{XX}^{-1} \Sigma_{XY} W_{iY}. \quad (16)$$

### 2.2.4 The variance of the shotgun connectivity estimator in the linear neuronal activity model

We can evaluate the variance of the shotgun ML connectivity estimator in the linear neuronal activity model by calculating the Laplace approximation in Eq. (13) around  $\hat{W}_i = W_i$ . In this case, denoting  $\hat{W}_i = W_i + \delta W_i$ , we find with respect to  $\delta W_i$ ,

$$l(\delta W_i) \approx -\frac{T_i}{2} \left\{ 1 + \delta W_i \frac{\mathcal{A}'}{1+B^2} \delta W_i^T \right\} + o(\delta W_i^2). \quad (17)$$

From Eq. (17), we read out the variance of the ML estimator  $\hat{W}_i$  as,

$$\text{var}(\hat{W}_i) = (1 + B^2)(\mathcal{A}'T_i)^{-1}. \quad (18)$$

For weakly correlated case,  $\Sigma \approx I$ , this can be reduced to a much more intuitive expression,

$$\text{var}(\hat{W}_i) = \frac{1 + (1-p)|W_i|_2^2}{pT_i}, \quad (19)$$

where  $p$  is the fraction of the neurons contained on average in one observation  $X_t$ , and  $|W_i|_2$  is the 2-norm of the input connectivity row-vector  $W_i$ .

The above result indicates that the posterior variance of the shotgun connectivity estimator grows with the square of neuronal connectivity as well as the size of the unobserved neuronal populations. In addition, the estimator variance grows as  $1/pT_i$  as  $p$  drops.

As we will see in Section 3, this result has a particularly simple interpretation: The statistical error in  $\hat{W}_i$  is introduced by the intrinsic noise  $\epsilon$  as well as the uncontrolled variability of the hidden inputs  $\zeta = W_{iY_t} Y_t$ . The reduction in that error for a given weight  $w_{ij}$  occurs in the proportion to the number of the observations in which the output of neuron  $i$  and the input of neuron  $j$  are simultaneously observed,  $pT_i$ . If both the output neuron and the input neuron are observed randomly with probability

$p$ , as in the original shotgun proposal, then the estimator error respectively is,

$$\text{var}(\hat{W}_i) = \frac{1 + (1-p)|W_i|_2^2}{p^2 T} \text{ (original shotgun),} \quad (20)$$

where  $T$  is now the total number of the observations and  $p^2 T$  is the number of the observations of different input-output neuronal pairs.

### 2.2.5 The general correctness of the shotgun ML connectivity estimation in the linear neuronal activity model

Theorem 1 proved in Section 3.1 can be used to put in context the analytical results of Section 2.2.2. In particular, Theorem 1 shows that, if a neuronal population activity model parameter  $\mathcal{W}$  can be brought into a unique correspondence with a set of partial neuronal activity distributions  $\mathbf{P}(\mathbf{S}) = \{P(X_{\mathcal{S}}|\mathcal{W}), \mathcal{S} \in \mathbf{S}\}$ , where  $\mathcal{S}$  stands for different subsets of observed neurons and  $P(X_{\mathcal{S}}|\mathcal{W})$  is the distribution of their activity, then such parameter can be uniquely estimated given any set of partial neuronal activity observations containing  $\mathbf{P}(\mathbf{S})$ . We can rephrase this statement to say that, if the set  $\mathbf{S}$  is such that different model parameters  $\mathcal{W}$  always identify distinct subpopulation activity distributions in  $\mathbf{P}(\mathbf{S})$ , then MLE provably converges to the true  $\mathcal{W}$  on the set of observations  $\{X_{\mathcal{S}}, \mathcal{S} \in \mathbf{S}\}$ .

We can use this theorem to put in context the result of Section 2.2.2. In particular, by multiplying Eq. (2) on the left with  $\mathcal{X}_t^T$  and averaging over time, we can obtain

$$\Sigma_1 = \mathcal{W}\Sigma, \quad (21)$$

where  $\Sigma_1 = E[\mathcal{X}_{t+1}\mathcal{X}_t^T]$  and  $\Sigma = E[\mathcal{X}_t\mathcal{X}_t^T]$ . As long as the matrix  $\Sigma$  is nonsingular, Eq. (21) establishes a unique correspondence between the neuronal connectivity matrix  $\mathcal{W}$  and the correlation matrices  $\Sigma_1$  and  $\Sigma$ . That is, if two connectivity matrices are not equal, by virtue of Eq. (21) they require two different pairs of the correlation matrices  $\Sigma_1$  and  $\Sigma$ . Since a given probability distribution implies a definite value of the correlation matrix, this establishes the desired correspondence between  $\mathcal{W}$  and the set of neuronal activity distributions  $P(X_{\mathcal{S}}|\mathcal{W})$ .

More specifically, we conclude that in the linear neuronal activity model different connectivity matrices  $\mathcal{W}$  necessarily require different

sets of neuronal activity probability distributions  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  (defining the matrix  $\Sigma_1$ ) and  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  (defining the matrix  $\Sigma$ ). By Theorem 1, any set of partial activity observations that fully specify these two sets is sufficient to uniquely identify the full  $\mathcal{W}$ . This statement is identical to the result of Section 2.2.2: providing all distributions  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  and  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  is equivalent to fully tiling  $\Sigma$  with the submatrices  $\Sigma_{*X_t}$  and  $\Sigma_{X_t X_t}$ .

Finally, we recall the note of Section 2.2.2 that, although the input-output model defined by Eq. (5) needs to be specified by two matrix parameters  $\mathcal{W}$  and  $\Sigma$ , the parameter  $\Sigma$  in fact is not independent, if considered within the scope of Eq. (2). For example, for the stationary distribution of Eq. (2)  $\Sigma(\mathcal{W}) = \sum_{n=0}^{\infty} (\mathcal{W})^n (\mathcal{W}^T)^n$ . It can be shown then in principle that the single set of the distributions  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  suffices to completely constrain the full model  $\mathcal{W}$ .

Thus we note that, either non-deterministic sampling protocols or otherwise reasonable deterministic protocols that change which subpopulation is imaged more slowly than the recording device's temporal resolution, can allow estimating all distributions  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  and  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$ , thus satisfying the sufficient conditions obtained in this section and in Section 2.2.2.

### 2.2.6 The sufficient conditions for the correctness of the shotgun ML connectivity estimation in exponential generalized linear models of spiking neural activity

Of special interest to applications in basic neuroscience is the so called spiking generalized linear model (GLM) of neuronal activity with exponential nonlinearity (Brillinger, 1988; Rigat et al., 2006; Pillow et al., 2008). In greater details this model is described in this paper in Section 2.4.2. For this model, we obtain explicitly important results regarding the correctness of the shotgun ML connectivity estimation in the practically interesting case of the large number of neurons  $N$ .

More specifically, we examine the spiking exponential-GLM described by a nonuniform Poisson spiking process with the instantaneous spiking rate

$$r_{it} = \exp \left( b_i + \sum_{j=1}^N W_{ij} \mathcal{X}_{j,t-1} \right). \quad (22)$$



Note that here the probability of neuronal spiking depends only on the previous neuronal population's state  $\mathcal{X}_{t-1}$ . This situation is simpler to analyze, while it remains sufficiently general so that the GLM in the form given by Eq. (40-41) can be reduced to Eq. (22) by concatenating into  $\mathcal{X}_t$  the neural spikes from many past times. Each  $\mathcal{X}_{it}$  is binary, either 1 or 0, corresponding respectively to spike or no-spike of neuron  $i$  at time  $t$ . The standard log-likelihood is,

$$\log P(\mathcal{X}|b, \mathcal{W}) = \sum_{i,t} (\mathcal{X}_{it} J_{it} - \exp(J_{it}) \Delta t), \quad (23)$$

where

$$J_{it} = b_i + \sum_{j=1}^N W_{ij} \mathcal{X}_{j,t-1}. \quad (24)$$

In the case when  $N$  is large, the law of large numbers states that the sum in Eq. (24) can be approximated under rather general conditions by a Normally distributed random variable with the mean  $m_i = b_i + \sum_j w_{ij} \mu_j$  and variance  $\sigma_i^2 = W_{i*} \Sigma W_{i*}^T$ , where  $\mu = E[\mathcal{X}_t]$  and  $\Sigma = E[\mathcal{X}_t \mathcal{X}_t^T]$  is the same-time covariance matrix, as before. The sum over the second term in Eq. (23) then can be replaced with

$$\sum_t \exp(J_{it}) \Delta t = \frac{T}{(2\pi W_{i*} \Sigma W_{i*}^T)^{\frac{1}{2}}} \int dJ e^J e^{-(J-m_i)^2 / (2W_{i*} \Sigma W_{i*}^T)}, \quad (25)$$

which clearly only depends on  $m_i$  and  $\Sigma$ , aside from the sought model parameters  $b$  and  $\mathcal{W}$ . The sum over the first term in Eq. (23) is explicitly

$$\sum_{i,t} \mathcal{X}_{it}^T J_{it} = \sum_i \mu_i b_i + Tr[\mathcal{W} \Sigma_1^T], \quad (26)$$

where  $\Sigma_1 = E[\mathcal{X}_{t+1} \mathcal{X}_t^T]$  is the time-shifted covariance matrix, as before. We can conclude, then, that the log-likelihood of the exponential-GLM in the limit of large number of neurons is a function of only  $\mu$ ,  $\Sigma$ , and  $\Sigma_1$ ,

$$\log P(\mathcal{X}|b, \mathcal{W}) = F(b, \mathcal{W}; \mu, \Sigma, \Sigma_1). \quad (27)$$

By Theorem 1 in Section 3.1, then, any set of partial observations completely constraining  $\mu$ ,  $\Sigma$  and  $\Sigma_1$  is sufficient to uniquely identify the full  $\mathcal{W}$ . That is, any set of observations specifying all distributions  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  and  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  suffices to uniquely constrain  $\mathcal{W}$ ,

insofar as any two exponential GLM are distinguishable on the full set of observations.

Note that this also provides an extremely handy method for estimating  $\mathcal{W}$  (Soudry et al., 2015). In particular, once  $\mu$ ,  $\Sigma$  and  $\Sigma_1$  had been calculated from any observations,  $\mathcal{W}$  can be found by maximizing

$$l(b, \mathcal{W}) = \mu^T b + Tr[\mathcal{W} \Sigma_1^T] - J_2(\mu, \Sigma, \mathcal{W}), \quad (28)$$

where

$$J_2(\mu, \Sigma, \mathcal{W}) = \sum_i e^{m_i + 2W_{i*} \Sigma W_{i*}^T}, \quad (29)$$

and the column-vector  $m = b + \mathcal{W} \mu$ .

We can also note, as before, that the parameters  $\mu$ ,  $\Sigma$  and  $\Sigma_1$  are not all independent and, in particular, knowing  $\mu$  and  $\Sigma_1$  one can expect that  $\Sigma$  can be extracted from the dynamics equations of the model. Then,  $\mathcal{W}$  can be found by using

$$\log P(\mathcal{X}|b, \mathcal{W}) = F(b, \mathcal{W}; \mu, \Sigma(\mu, \Sigma_1), \Sigma_1), \quad (30)$$

whereas the measurement of all  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  already suffice to determine  $\mathcal{W}$ . However, in that case one needs to explicitly obtain the solution for  $\Sigma(\mu, \Sigma_1)$ , which may be more challenging in practice than simply measuring  $\Sigma$  separately.

### 2.2.7 The sufficient conditions for the correctness of the shotgun ML connectivity estimation in general generalized linear models of spiking neural activity

Here, we extend the result of the previous section to general spiking generalized linear neural activity models, where the spiking rate is defined by

$$r_{it} = f(J_{it}), \quad (31)$$

where  $J_{it}$  is given by Eq. (24). The standard observations' log-likelihood now is given by,

$$\log P(\mathcal{X}|b, \mathcal{W}) = \sum_{i,t} \mathcal{X}_{it} \log f(J_{it}) - \sum_{i,t} f(J_{it}) \Delta t. \quad (32)$$

As before, in the limit of large  $N$  the distribution of  $J_{it}$  can be approximated using Gaussian. Then, we can rewrite the log-likelihood (32) as

$$\log P(\mathcal{X}|b, \mathcal{W}) = J_1(b, \mathcal{W}; \mu, m', \Sigma') - J_2(b, \mathcal{W}; m, \Sigma), \quad (33)$$

where

$$J_1(b, \mathcal{W}; \mu, m', \Sigma') = \sum_i \mu_i \left\{ \frac{1}{(2\pi W_{i*} \Sigma'(i) W_{i*}^T)^{\frac{1}{2}}} \times \int dJ \log f(J + m'(i)) e^{-J^2 / (2W_{i*} \Sigma'(i) W_{i*}^T)} \right\}, \quad (34)$$

and

$$J_2(b, \mathcal{W}; m, \Sigma) = \sum_i \left\{ \frac{1}{(2\pi W_{i*} \Sigma W_{i*}^T)^{\frac{1}{2}}} \times \int dJ f(J + m_i) e^{-J^2 / (2W_{i*} \Sigma W_{i*}^T)} \right\}. \quad (35)$$

Where  $m$ ,  $\mu$  and  $\Sigma$  were defined in Section 2.2.6 and  $m'(i)$  and  $\Sigma'(i)$  are,

$$\begin{aligned} m'(i) &= b_i + W_{i*} E[\mathcal{X}_t | \mathcal{X}_{i,t+1} = 1] \\ \Sigma'(i) &= E[\mathcal{X}_t \mathcal{X}_t^T | \mathcal{X}_{i,t+1} = 1] \end{aligned} \quad (36)$$

By Theorem 1 in Section 3.1, these imply that the set of distributions  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$ ,  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  and  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt}, \mathcal{X}_{kt})$  is sufficient to identify  $\mathcal{W}$ .

2.3 The numerical solution of the shotgun connectivity estimation problem in general causal models of neuronal population activity

### 2.3.1 The general causal model of neuronal population activity

The problem of estimating the effective connectivity matrix  $\mathcal{W}$  in realistic settings will require numerical solution. In this section, we develop a numerical approach for solving this problem using the Expectation Maximization algorithm. In the following, we assume throughout that the activity  $\mathcal{X}_t$  of the neuronal population in question can be modeled by a general Markov model defined by the relationship

$$\mathcal{X}_t \sim P(\mathcal{X}_t | \mathcal{X}_{t-1}; \mathcal{W}), \quad (37)$$

where  $\mathcal{X}_t$  is the activity of the neuronal population at time  $t$ , defined to be conditional on the state of the neuronal population at time  $t - 1$ ,  $\mathcal{X}_{t-1}$ , and some parameter  $\mathcal{W}$ . The linear neuronal activity model, Section 2.2, is evidently in this form. However, Eq. (37) is more general and covers a great variety of other important statistical approaches for neuronal activity modeling, with a particularly important case being that of the generalized linear model of neuronal activity, discussed in greater detail in Section 2.4.2.

### 2.3.2 Expectation Maximization algorithm for the shotgun connectivity estimation problem in general neuronal activity models

We formulate the problem of connectivity estimation from the shotgun sampling data as the estimation of  $\mathcal{W}$  in model (37) over a set of partial observations of the neuronal activity  $X \subset \mathcal{X}$ . The standard method for solving such a parameter estimation problem in the presence of missing data is the Expectation Maximization (EM) algorithm (Dempster et al., 1977). To recap briefly, the EM algorithm produces a sequence of parameter estimates  $\hat{\mathcal{W}}^k$  with uniformly increasing likelihoods  $P(X | \hat{\mathcal{W}}^k)$ , guaranteeing at least a locally-maximum likelihood estimate of  $\mathcal{W}$ . The sequence  $\hat{\mathcal{W}}^k$  is produced by iteratively maximizing the functions

$$Q(\mathcal{W} | \hat{\mathcal{W}}^k) = E_{P(Y|X, \hat{\mathcal{W}}^k)}[\log P(X, Y, \mathcal{W})], \quad (38)$$

where  $P(Y|X; \hat{\mathcal{W}}^k)$  is the posterior distribution of the hidden neuronal activities given the available observations  $X$  and the current estimate  $\hat{\mathcal{W}}^k$ . This maximization is typically realized by constructing  $M$  samples of the missing data  $Y^l$  from  $P(Y|X; \hat{\mathcal{W}}^k)$  and then calculating  $Q(\mathcal{W} | \hat{\mathcal{W}}^k)$  as  $Q(\mathcal{W} | \hat{\mathcal{W}}) = \frac{1}{M} \sum_l \log P(X, Y^l; \mathcal{W})$ .

To apply the EM algorithm here, we reformulate Eq. (37) as a Hidden Markov Model (HMM) in which the neuronal population activities,  $Y_t$ , are treated as hidden states and the activities  $X_t$  are treated as observations. We first construct the sample of the unobserved neuronal activities  $Y$ . Given the above definitions, this problem can be stated as the sampling of the hidden state sequences  $Y_t$  from a HMM constrained on the observations  $X$ , and solved efficiently using the standard forward-backward algorithm (Rabiner, 1989; Paninski et al., 2010).

Assuming the sample  $Y \sim P(Y|X; \hat{\mathcal{W}}^k)$  had been constructed, we evaluate  $Q(\mathcal{W} | \hat{\mathcal{W}}^k)$  by using,

$$Q(\mathcal{W} | \hat{\mathcal{W}}) = \log P(\mathcal{W}) + E[\log P(\mathcal{X}_0)] + \sum_t E[\log P(\mathcal{X}_t | \mathcal{X}_{t-1}; \mathcal{W})], \quad (39)$$

where the expectation values are with respect to the produced sample of  $Y$ . This function can be calculated straightforwardly and also can be made convex with a suitable choice of the causation in Eq. (37)—for example, by using a log-

concave rate function  $f$  in Eqs. (40-41) (Paninski, 2004; Paninski et al., 2010). If the latter is achieved, the maximum of  $Q(\mathcal{W}|\hat{\mathcal{W}})$  can be found efficiently with the help of standard gradient descent algorithms.

A detailed description of the implementation of the EM algorithm for model (37) is given in Appendix A.

## 2.4 Numerical simulations

### 2.4.1 Numerical simulations of the shotgun connectivity estimation in the linear neuronal activity model

As a check of the calculations in Section 2.2.2, we performed numerical simulations of the shotgun connectivity estimation in the linear neuronal activity model. The model was simulated using the definition given by Eq. (5) for a single output neuron and a population of neuronal inputs. The input connection weights  $W$  were chosen uniformly at random on the interval  $[0, W_{max}]$ , with a connection probability  $s$ . The activities of the input neurons  $X_t$  and  $Y_t$  were drawn from a multivariate normal distribution with zero mean and a covariance matrix  $\Sigma$ , randomly generated from uniform distribution on  $[0, 1]$  and normalized to unit variances. More specifically, the elements of  $\Sigma$  were initially independently chosen from uniform random distribution on the interval  $[0, 1]$ . Then, thus obtained matrix was symmetrized by using  $\Sigma := \Sigma + \Sigma^T$  and offset by using  $\Sigma := \Sigma + \sqrt{N} \cdot I$ , to guarantee the positive definiteness (Furedi and Komlos, 1981). Finally, each row and each column of  $\Sigma$  were divided by the square of the respective diagonal element,  $\sqrt{\Sigma_{ii}}$ , to normalize  $\Sigma$  to unit variances  $\Sigma_{ii} = 1$ . The observations were simulated by continuously observing the activity of the output neuron and sampling the activities of the input neurons using the block-wise round-robin strategy, discussed in Section 3.2. The original connectivity was numerically recovered from thus simulated data using an implementation of the EM algorithm in Matlab, following the discussion of Section 2.3.2, with the sample size of the hidden neuronal activities set at  $M = 100$ .

### 2.4.2 Numerical simulations of the shotgun connectivity estimation in the generalized linear neuronal activity model

In order to test the shotgun connectivity estimation in more realistic settings, we performed the numerical simulations of the shotgun connectivity matrix estimation using several model cortical neuronal networks, described using the generalized linear model (GLM) (Brillinger, 1988; Rigat et al., 2006; Pillow et al., 2008).

GLM is a particularly general and powerful class of statistical models of neuronal population activity, described as a nonuniform Poisson neuronal spiking process,

$$Pr(\eta_{it} = 1) = f \left( b_i + \sum_{j=1}^N w_{ij} J_{jt} \right) \quad (40)$$

$$J_{jt} = a_j \cdot (J_{j,t-1} + \eta_{j,t-1}) + \epsilon_{jt}, \quad (41)$$

Here,  $\eta_{it}$  is a binary variable specifying whether  $i$ -th neuron fired at time  $t$ ,  $f$  is a nonlinear firing rate function,  $b_i$  is a scalar offset, and  $w_{ij}$  are coupling weights. The driving currents  $J_{jt}$  are defined as autoregressive processes with decay constants  $a_j$  and a normal noise  $\epsilon_{jt} \sim N(0, \sigma_j^2)$ . More than one autoregressive current may be defined per neuron, allowing the modeling of complex neuronal responses as the combinations of  $J_{jt}$ 's with different decay constants. The diagonal weights  $w_{ii}$  model the self-dependencies in neuronal activity, such as refractory period, bursting, etc., whereas the off-diagonal weights  $w_{ij}$  model the interactions between neurons. Generalized linear models had been successfully applied in the literature to model the statistical properties of individual neurons as well as that of large neuronal populations (Brillinger, 1988; Chornoboy et al., 1988; Brillinger, 1992; Plesser and Gerstner, 2000; Paninski et al., 2004; Paninski, 2004; Rigat et al., 2006; Trucolo et al., 2005; Nykamp, 2007; Kulkarni and Paninski, 2007; Pillow et al., 2008; Vidne et al., 2009; Stevenson et al., 2009). The model given by Eqs. (40-41) falls under the general definition given by Eq. (37) after the identification  $\mathcal{X}_{it} = (\eta_{it}, J_{it})$ .

In this paper, we simulated two GLM model neuronal networks—a small toy model of a synfire neuronal circuit and a larger realistic weakly coupled network of cortical neurons.

The synfire network was simulated in order to demonstrate the resolution of the canonical common inputs situation in the shotgun

approach. The synfire model was created as an all-excitatory network of  $N = 10$  neurons with strong feed-forward connectivity, as described in Figure 4. All connection strengths were chosen as a constant  $W_{syn}$ , selected so that the probability of a post-synaptic neuron spiking conditional on a spike of a connected pre-synaptic neuron was approximately 80%. The model was then simulated using Eqs. (40-41) and the parameter definitions in Table 2. The neuronal activity was collected using the block-wise round-robin sampling, described in Section 3.2.

The model of a weakly coupled cortical neuronal network was adopted here from (Mishchenko et al., 2011), and was intended to demonstrate the shotgun connectivity estimation in a more realistic setting. The model closely reproduced the experimental data about local cortical neuronal circuits available in the literature (Braitenberg and Schuz, 1998; Gomez-Urquijo et al., 2000; Lefort et al., 2009; Sayer et al., 1990). More specifically, the neuronal population was created as 80% excitatory and 20% inhibitory neurons. The neurons were connected with each other randomly and homogeneously with the probability of 10%. The Dale's law was respected. The strength of the excitatory connections was set using the peak excitatory post-synaptic potential (PSP) values randomly chosen from an exponential distribution with mean 0.5 mV (Lefort et al., 2009; Sayer et al., 1990). The strength of the inhibitory connections was set using a similar exponential distribution with the mean chosen to balance the average excitatory and inhibitory inputs in the network (Abeles, 1991). All neurons had the refractory periods of 1 msec, enforced in Eqs. (40-41) via the self-currents  $J_{it}$  with the decay time-constants 1 msec and  $w_{ii} = -100$ . Individual PSPs were modeled using the alpha function (Koch, 1999), described as the difference of two exponentials with a rise time of 1 msec and a decay time of 10 to 20 msec (Sayer et al., 1990). Given the simulations time step of 1 msec here, such PSPs can be described more precisely as an instantaneous jump followed by an exponential decay of 10 to 20 msec, as described by Eq. (41). The spiking activity was simulated at 1 msec time step ignoring the interneuronal conduction delays, negligible for spatially compact neuronal circuits. The neuronal activities were downsampled at 100 Hz, in order to simulate the observations using a slow imag-

ing technique such as calcium imaging. Generated neuronal activities were collected using the block-wise round-robin sampling strategy, Section 3.2. The detailed parameters used in the simulations are listed in Table 3.

The connectivity matrix was estimated from the simulated data using a Matlab implementation of the EM algorithm described in Appendix A. The sample size of the hidden neuronal population activity was taken as  $M = 50$  for the cortical neuronal networks and  $M = 150$  for the synfire network. Only the connectivity matrix weights were estimated, taking the PSP time-constants as known. This was done to separate different sources of errors in the estimation, understanding that the focus here is on the problem of the shotgun connectivity inference. The PSP time-constants, more generally, can be set uniformly in the population from physiological data without a significant impact on the connectivity estimation, or also included into the EM procedure. For more in depth discussion of this issue see (Mishchenko et al., 2011).

The EM algorithm was executed on a 4 dual-core i7 processor desktop computer with 8GB of RAM memory. The numerical estimation problems could be generally solved in a reasonable amount of time, however, we found that the necessity to keep up to  $M$  samples of the complete activity of the hidden neuronal populations imposed drastic requirements on RAM memory. In particular, for the GLM cortical neuronal networks described here and using the said workstation, we could solve at most the models with  $T = 3000$  seconds of neuronal activity data and  $M = 50$  examples of the hidden activity data. The above configuration, thus, set the limits on the numerical experiments performed in this paper.

More generally, in order to set the parameter  $M$  of the number of samples used to model the posterior distribution of hidden neuronal activities, we tried different  $M$  such as  $M = 10, 50, 100$  and  $150$ . The value of  $M = 50$  was the highest value that we could try on the above mentioned computer configuration for the GLM neuronal networks. The value  $M = 150$  was the highest that we could try for the synfire model. Generally, we found that  $M = 10$  sufficed to obtain a solution, however, the noise in that solutions was high and correlated but disconnected neurons frequently were identified as connected (in the case of the synfire model). Using a larger

hidden activity sample  $M$ , thus, appeared to be necessary.  $M = 50$  appeared to be sufficient to achieve satisfactory reconstructions for the weakly correlated cortical neuronal models, but appeared to be lacking in the case of the strongly correlated synfire model.  $M$  above 100 appeared to be satisfactory in the case of the weakly correlated cortical neuronal network as well as the synfire model.

**Table 1** The parameters of the linear neuronal activity model used in the numerical simulations in this paper.  $\mathcal{U}(a, b)$  is a uniform probability distribution on  $[a, b]$ .

Number of input neurons, $N$	100
Input weights distribution, $W$	$\mathcal{U}(0, W_{max})$ , $W_{max} = 2$
Connection probability, $s$	10%
Observations sparseness, $p$	20%
Sample size, $T$	500 samples

**Table 2** The parameters of the toy GLM synfire model used in the numerical simulations in this paper.

Number of neurons	10
Base log-firing rate ( $\exp(b_i)$ )	15 Hz
Average firing rate	33 Hz
Simulation time step	1 msec
Max observation duration	50 sec
Neurons per imaging block	20%
Excitatory conn. strength ( $w_{ij}$ )	13
Excitatory PSP rise time	1 msec
Excitatory PSP decay time	1 msec
Refractory time	1 msec

### 3 Results

#### 3.1 The correctness of the shotgun neuronal connectivity estimation

One of the biggest challenges of the functional connectivity estimation in neuroscience remains the presence of unobserved or hidden inputs in neuronal population activity data. The shotgun connectivity estimation is a promising approach for alleviating this problem, consisting in imaging a large neuronal population using small groups of random neurons and reconstructing the complete connectivity matrix from such partial observations.

In Materials and Methods, we study analytically the problem of the shotgun connectivity

estimation in a linear model of neuronal population activity. We show that the shotgun estimation can be always posed as the problem of estimating the input connectivity of one “output” neuron at a time, given a partially observed population of neuronal inputs. This is because in typical network models of neuronal activity the activity of neurons is conditionally independent given the activity of their presynaptic neuronal populations and the respective connection weights.

In the considered linear neuronal activity model, we explicitly calculate the observations likelihood  $P(X|\mathcal{W})$  and demonstrate that the MLE in this case converges to the correct complete connectivity matrix as long as all possible input-input and input-output pairs of neurons are observed together in at least a fraction of the observations. We also explicitly derive the hidden inputs bias and the variance of the shotgun estimator in this model.

It is possible to further extend these results to more general settings. For that, we assume first that the activity of a neuronal population,  $\mathcal{X}$ , can be described by a general parametric model probability density  $P(\mathcal{X}|\mathcal{W})$ , given a parameter  $\mathcal{W}$ . We say that an estimator  $\hat{\mathcal{W}}$  is consistent as long as  $\hat{\mathcal{W}} \rightarrow \mathcal{W}$  as the sample size tends to infinity. Given a set of partial neuronal activity observations  $X \subset \mathcal{X}$ , we recall from the asymptotic estimation theory that the MLE is guaranteed to provide a consistent estimator as long as the observations distribution  $P(X|\mathcal{W})$

**Table 3** The parameters of the realistic GLM cortical neuronal network used in the numerical simulations in this paper.  $\text{Exp}(\lambda)$  is the exponential probability distribution with mean  $\lambda$ , and  $N_p(\mu, \sigma)$  is the truncated-normal probability distribution with mean  $\mu$ , standard deviation  $\sigma$ , and lower bound  $p$ .

Number of neurons	50
Base log-firing rate ( $\exp(b_i)$ )	10 Hz
Average firing rate	10 Hz
Excitatory neurons	80%
Inhibitory neurons	20%
Connectivity sparseness	10%
Simulation time step	1 msec
Max observation duration	3000 sec
Neurons per imaging block	20%
Excitatory PSP peak	$\text{Exp}(0.5)$ mV
Inhibitory PSP peak	$\text{Exp}(-2.3)$ mV
Excitatory PSP rise time	1 msec
Inhibitory PSP rise time	1 msec
Excitatory PSP decay time	$N_5(10, 2.5)$ msec
Inhibitory PSP decay time	$N_{10}(20, 5)$ msec
Refractory time	$N_1(2, 0.5)$ msec

possesses no observation-indistinguishable parameter sets. That is,  $P(X|\mathcal{W})$  is such that no two distinct parameters  $\mathcal{W} \neq \mathcal{W}'$  result in identical distributions  $P(X|\mathcal{W}) \equiv P(X|\mathcal{W}')$ . The above condition is commonly referred to in statistics as the identifiability property.

The above observation can be made most readily by considering the average observed-data log-likelihood maximized by the MLE. More specifically, let us consider a collection of neuronal population activity observations  $\{X^{(k)}, k = 1 \dots n\}$ . This can be understood as a collection of repeated experiments or a collection formed from different segments of the same experiment, corresponding to its different time-intervals.

In the MLE, we maximize the log-likelihood of the observed data, whereas by the observed data here we understand the collection of all neuronal activity observations  $X^{(k)}$ . The average observed-data log-likelihood in the limit of large  $n$  converges in probability to

$$l(\hat{\mathcal{W}}|\{X^{(k)}\}) = \frac{1}{n} \sum_{k=1}^n \log P(X^{(k)}|\hat{\mathcal{W}}) \quad (42)$$

$$\approx E_{P(X|\mathcal{W})}[\log P(X|\hat{\mathcal{W}})],$$

where the last line identifies the expected log-likelihood function  $l(\hat{\mathcal{W}}) = E[\log P(X|\hat{\mathcal{W}})]$ , and the expectation is with respect to the true distribution of  $X$ ,  $P(X|\mathcal{W})$ . Well known Gibbs inequality then tells us that Eq. (42) achieves its global maximum when and only when  $P(X|\hat{\mathcal{W}}) \equiv P(X|\mathcal{W})$  for all  $X$ . The above standard argument leads to nontrivial insights about the shotgun neuronal connectivity estimation below.

From the perspective of the identifiability of the connectivity estimation problems in neuronal activity models with the lack of complete observations, we first point out that the problem of hidden inputs in that setting arises because the identifiability condition becomes broken. That is, different models of hidden neuronal connectivity can produce the same distribution of the observed data. In the classical example of the hidden inputs problem, for example, a direct connection is inferred between two unconnected observed neurons because a correlated input fed into these neurons by a third neuron also can reproduce the data collected on the observed neurons. The violation of identifiability leads to multiple network models being able to reproduce the same empirical distribution  $P(X|\mathcal{W})$  and, therefore, achieve the global maximum of the log-likelihood (42). However, the key observation at this stage is that the

true connectivity  $\mathcal{W}$  still remains a ML solution. Thus, it is not that the hidden inputs problem somehow skews the neuronal connectivity estimation. It is just that such estimation is no longer unique.

To examine this state of things in the shotgun estimation, we recognize initially that the expected log-likelihood (42) in this case should be re-written as,

$$l(\hat{\mathcal{W}}) = E_{\mathcal{S}} \left\{ E_{P(X_{\mathcal{S}}|\mathcal{W})} [\log P(X_{\mathcal{S}}|\hat{\mathcal{W}})] \right\}. \quad (43)$$

Here,  $\mathcal{S}$  denotes a particular subpopulation of observed neurons and the external average is over such subpopulations  $\mathcal{S}$  imaged during a shotgun sampling experiment.  $X_{\mathcal{S}}$  refers to the part of the neuronal population's activity observed in  $\mathcal{S}$ , and  $P(X_{\mathcal{S}}|\mathcal{W})$  is the probability distribution of such activities.

Once again, we recall the observation above that the true connectivity matrix  $\hat{\mathcal{W}} = \mathcal{W}$  is necessarily a global maximizer of Eq. (43), by Gibbs inequality. More significantly,  $\hat{\mathcal{W}} = \mathcal{W}$  maximizes Eq. (43) by simultaneously maximizing all individual terms  $E_{P(X_{\mathcal{S}}|\mathcal{W})} [\log P(X_{\mathcal{S}}|\hat{\mathcal{W}})]$ , since evidently  $P(X_{\mathcal{S}}|\hat{\mathcal{W}}) = P(X_{\mathcal{S}}|\mathcal{W})$  for all  $\mathcal{S}$  whenever  $\hat{\mathcal{W}} = \mathcal{W}$ . We now ask whether there can exist another model  $\mathcal{W}' \neq \mathcal{W}$  such that achieves the same global maximum. We note that, since  $\hat{\mathcal{W}} = \mathcal{W}$  achieves the global maximum of Eq. (43) by simultaneously globally maximizing all  $E_{P(X_{\mathcal{S}}|\mathcal{W})} [\log P(X_{\mathcal{S}}|\hat{\mathcal{W}})]$ , any other such maximizer of Eq. (43) must also have this property. In turn, this implies that any such  $P(X_{\mathcal{S}}|\mathcal{W}')$  must match the marginal distributions  $P(X_{\mathcal{S}}|\mathcal{W})$  on all inspected  $\mathcal{S}$ .

We thus arrive at the following conclusion:

**Definition 1** Assume that a statistical model of neuronal population activity  $P(\mathcal{X}|\mathcal{W})$  and a set of neuronal subpopulations  $\mathbf{S} = \{\mathcal{S}\}$  are such that for any two different model parameters  $\mathcal{W}$  and  $\mathcal{W}'$  there exist at least one  $\mathcal{S}$  in  $\mathbf{S}$  such that the distributions  $P(X_{\mathcal{S}}|\mathcal{W})$  and  $P(X_{\mathcal{S}}|\mathcal{W}')$  are not identically equal, where  $X_{\mathcal{S}}$  is the restriction of  $\mathcal{X}$  to  $\mathcal{S}$ . Then, we say that the model  $P(\mathcal{X}|\mathcal{W})$  is uniquely identified by the set of distributions  $\mathbf{P}(\mathbf{S}) = \{P(X_{\mathcal{S}}|\mathcal{W}), \mathcal{S} \in \mathbf{S}\}$ .

**Definition 2** Assume that for any neuronal subpopulation  $\mathcal{S}$  in a set of neuronal subpopulations  $\mathbf{S}$  there exist a subpopulation  $\mathcal{S}'$  in a different set  $\mathbf{S}'$  such that  $\mathcal{S} \subset \mathcal{S}'$ . Then we say that  $\mathbf{S}'$  completely covers  $\mathbf{S}$ .

**Theorem 1** *Assume a statistical model of neuronal population activity  $P(\mathcal{X}|\mathcal{W})$  is uniquely identified by a set of distributions  $\{P(X_{\mathcal{S}}|\mathcal{W}), \mathcal{S} \in \mathbf{S}\}$ . Then, for any set of partial observations of the neuronal population activity in this model on a set of neuronal subpopulations  $\mathbf{S}'$  that completely covers  $\mathbf{S}$ , the ML estimator defined as the argmin of (43) is consistent.*

In other words, if the shotgun sampling  $\mathbf{S}'$  covers all neuronal subpopulations  $\mathcal{S}$  from the identifying set of partial neuronal activity distributions  $\mathbf{P}(\mathbf{S})$ , then the shotgun ML estimator is guaranteed to converge to the correct  $\mathcal{W}$ . The condition on  $\mathbf{S}'$  to completely cover  $\mathbf{S}$  ensures that any  $P(X_{\mathcal{S}}|\mathcal{W})$  can be recovered from some  $P(X_{\mathcal{S}'}|\mathcal{W})$  such that  $\mathcal{S} \subset \mathcal{S}'$ , by marginalization.

Theorem 1 replaces the question of the consistency of the MLE in the shotgun sampling by the question of finding the identifying sets  $\mathbf{P}(\mathbf{S})$  for a given neuronal activity model. It shows that, whenever an identifying set  $\mathbf{P}(\mathbf{S})$  is covered by a sampling scheme, either using non-deterministic or deterministic protocol, the respective inference problem possesses the identifiability property, whereas in general such identifiability property for an inference problem with only partial observations data is unknown and, in fact, cannot be guaranteed even when the original full problem is identifiable. It is also worth noting that in Theorem 1 it is not important how exactly the mapping between the identifying sets  $\mathbf{P}(\mathbf{S})$  and  $\mathcal{W}$  is constructed, or by using which statistics of the distributions in  $\mathbf{P}(\mathbf{S})$  the parameter  $\mathcal{W}$  can be extracted and how. As long as such a mapping exists, the full parameter  $\mathcal{W}$  is recoverable by Theorem 1 by using the MLE.

It remains to establish the identifying sets of given neuronal activity models. In certain cases, such as in the linear neuronal activity model, the identifying set can be rather immediately established (Section 2.2.5), providing the sufficient conditions for the shotgun sampling to be successful. However, in general, different neuronal activity models require separate investigation into their identifying sets. From general arguments, we can conjecture that in a large class of network models of neuronal activity the set of all pair-wise input-output neuronal activity distributions such as  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  would suffice to form such a set, see below. However, in general the question of finding the

identifying sets  $\mathbf{P}(\mathbf{S})$  of a neuronal activity model is nontrivial, and even the existence of nontrivial identifying sets  $\mathbf{P}(\mathbf{S})$  cannot be always guaranteed.

Specifically, consider an arbitrary probability density  $P(\mathcal{X})$  defined on a discrete grid of  $d$  points in  $n$  dimensions, with  $d^n$  points in total. Such  $P(\mathcal{X})$  has  $d^n - 1$  free parameters. All marginal distributions of up to  $n-1$  dimensions together provide  $n(d-1) + n(n-1)/2(d^2-1) + n(n-1)(n-2)/3!(d^3-1) + \dots < (1+d)^n - d^n < nd^{n-1}$  linear conditions on  $P(\mathcal{X})$ . Therefore, even all marginal distribution  $P(X_{\mathcal{S}})$  together are not sufficient to determine  $P(\mathcal{X})$  uniquely whenever  $d > n$ .

As a specific counter-example, let us consider a toy model  $P(\mathcal{X}_1, \mathcal{X}_2|\mathcal{W})$  with 2 neurons in 4 possible states ( $n = 2$  and  $d = 4$ ). Specifying such a general model requires providing  $4^2 - 1 = 15$  different probability values  $P(\mathcal{X}_1, \mathcal{X}_2)$ . However, the two marginal distributions  $P(\mathcal{X}_1)$  and  $P(\mathcal{X}_2)$  provide only 8 linear constraints on  $P(\mathcal{X}_1, \mathcal{X}_2)$ :  $P(\mathcal{X}_1) = \sum_{\mathcal{X}_2} P(\mathcal{X}_1, \mathcal{X}_2)$  and  $P(\mathcal{X}_2) = \sum_{\mathcal{X}_1} P(\mathcal{X}_1, \mathcal{X}_2)$ . Clearly, an infinite number of  $P(\mathcal{X}_1, \mathcal{X}_2)$  can be suggested perfectly matching these constraints for almost any  $P(\mathcal{X}_1)$  and  $P(\mathcal{X}_2)$ . In this case, any partial observation of  $\mathcal{X}$  is insufficient—a complete observation is required.

We ran into the problems in the above example because the number of the degrees of freedom of the model probability distribution  $P(\mathcal{X})$  was simply too high. However, in network models of neuronal population activity  $P(\mathcal{X}|\mathcal{W})$  is generally a much more constrained distribution, which is typically completely specified by a single connectivity matrix  $\mathcal{W}$  of  $N^2$  elements, where  $N$  is the number of neurons. Counting the degrees of freedom in this case implies that there can be at most  $N^2$  independent distributions in the set of all marginal distributions  $P(X_{\mathcal{S}}|\mathcal{W})$ . Choosing any subset of  $N^2$  independent such distributions, thus, will uniquely fix all other distributions and, therefore, the complete model  $\mathcal{W}$ .

It is natural to conjecture (but we cannot prove this now rigorously) that such a subset can be chosen in the form of the  $N^2$  input-output distributions  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$ . Other choices may also exist. For example, if the stationary distribution of the model (37),  $\pi(\mathcal{X}_t)$ , is such that a unique mapping is formed by  $\pi(\mathcal{X}_t) \rightarrow \mathcal{W}$ , then a set of  $N^2$  independent same-time

distributions such as  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  may do just as well for specifying  $\mathcal{W}^1$ .

In summary, we suggest, based on the counting of degrees of freedom, that one can expect the set of the observations of all distinct input-output neuronal pairs  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  to be sufficient for recovering the complete neuronal connectivity matrix in general network models of neuronal population activity. At the same time, the above argument cannot constitute a rigorous proof. Therefore, we must present it here as a conjecture open for further investigation.

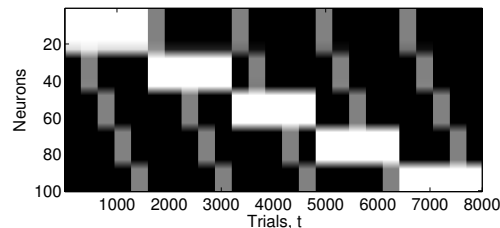
### 3.2 Alternative organizations of the shotgun neuronal population activity sampling

The results obtained in Section 3.1 have certain implications for the organization of the shotgun connectivity estimation experiments. We see that, to be able to reconstruct a complete neuronal connectivity, it is only important that the identifying set  $\mathbf{P}(\mathbf{S})$  is covered by the observations, and the manner in which such coverage is provided is not important. Per the conjecture of Section 3.1, any imaging organization of a neuronal network that furnishes the observations of all input-output neuronal pairs in that network would suffice to uniquely identify the complete neuronal connectivity matrix.

A particularly advantageous and conceptually simple alternative organization of neuronal activity sampling from this perspective is the block-wise round-robin sampling illustrated in Figure 2. In this approach, a neuronal population is imaged as a series of contiguous blocks, one input and one output block imaged simultaneously for a set number of observations  $T_b$ . The blocks are moved through the population so that all possible combinations of the input and the output blocks are inspected.

The block-wise round-robin strategy can be straightforwardly implemented using the exist-

<sup>1</sup> In this respect, note that the pairwise same-time distributions  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  in general may not be sufficient to identify  $\mathcal{W}$  since these are symmetric and, in the worst case scenario, only provide  $N(N+1)/2$  independent constraints on  $\mathcal{W}$ . In that sense, the time-shifted distributions  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  are guaranteed to provide  $N^2$  different constraints. One can also look at the distributions of triples  $P(\mathcal{X}_{it}, \mathcal{X}_{jt}, \mathcal{X}_{kt})$  from  $\pi(\mathcal{X}_t)$ . Alternatively, if the distributions  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  contain different statistics (such as mean and variance) that essentially independently relate to  $\mathcal{W}$ , then measuring  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  may also prove sufficient for identifying  $\mathcal{W}$ .



**Fig. 2** Block-wise round-robin neuronal population activity sampling strategy. In this strategy, the neuronal population is imaged as a sequence of contiguous blocks of input and output neurons. During one section of the experiment, the neurons in one input and one output block are observed simultaneously for a set of observations  $T_b$ . All possible combinations of the input and the output blocks are observed over the entire experiment. The figure illustrates the block-wise round-robin sampling strategy applied to a hypothetical population of 100 neurons with 20 neurons observed per each input and output block. White color indicates the neurons in the output blocks and gray color indicates the neurons in the input blocks. In each observation, the activity of all marked neurons is observed simultaneously.

ing fluorescent microscopy tools by scanning two field-of-views of a confocal or two-photon microscope over the neuronal population. Another advantage of this strategy is a simpler numerical connectivity estimation problem.

Block-wise round-robin sampling strategy is sufficient for collecting all input-output as well as same-time pairs of neurons and, in that sense, is sufficient for the recovery of the complete neuronal connectivity in the sense of the conjecture about the identifying sets of network models of neuronal activity in Section 3.1.

### 3.3 The impact of the missing data on the shotgun connectivity estimation

In Section 2.2.4 we obtained some key properties of the shotgun connectivity estimator in the linear neuronal activity model. It is interesting to return to these results now, from the positions of the discussion of the last sections.

In particular, let us consider a segment of a block-wise round-robin neuronal activity sampling experiment with a fixed input and output neuronal blocks. We can roughly relate the activity of the observed neurons in this scenario as,

$$Z_t = f(\mathcal{W}_{ZX} X_t + \zeta_t + \epsilon_t), \quad (44)$$

where  $Z_t$  is the column-vector of the activities of the neurons in the output block,  $X_t$  is the



column-vector of the activities of the neurons in the input block, and we introduced a new random variable  $\zeta = \mathcal{W}_{ZY}Y$ , which represents the combined input of the hidden neurons into the neuronal outputs.

Posed from this perspective, the shotgun connectivity estimation appears now as the problem of estimating a block of the connectivity matrix  $\mathcal{W}_{ZX}$  given the observations of the activities of all relevant input and output neurons, whereas the presence of the unobserved neurons enters solely in the form of an additional noise  $\zeta$ . The noise  $\zeta$  is both structured and correlated with  $X$ . The knowledge of that structure is required to successfully remove the associated bias from the estimates of  $\mathcal{W}_{ZX}$ . However,  $\zeta$  also introduces additional statistical uncertainty into the estimates of  $\mathcal{W}_{ZX}$ . We can estimate the variance of  $\zeta$  as  $\text{var}(\zeta) = \text{var}(\mathcal{W}_{ZY}Y) \approx |\mathcal{W}_{ZY}|^2 \text{var}(Y)$ , leading to a rough estimate of the variance of the estimator  $\hat{\mathcal{W}}_{ZX}$  as,

$$\begin{aligned} \text{var}(\hat{\mathcal{W}}_{ZX}) &\propto (T_b \text{var}(X))^{-1} \text{var}(\epsilon + \zeta) \\ &\approx (T_b \text{var}(X))^{-1} (\text{var}(\epsilon) + |\mathcal{W}_{ZY}|^2 \text{var}(Y)), \end{aligned} \quad (45)$$

where  $|\mathcal{W}_{ZY}|^2$  is understood as the average 2-norm of the rows of  $\mathcal{W}_{ZY}$ . In the limit where the size of the hidden neuronal population is large, we can write that approximately,

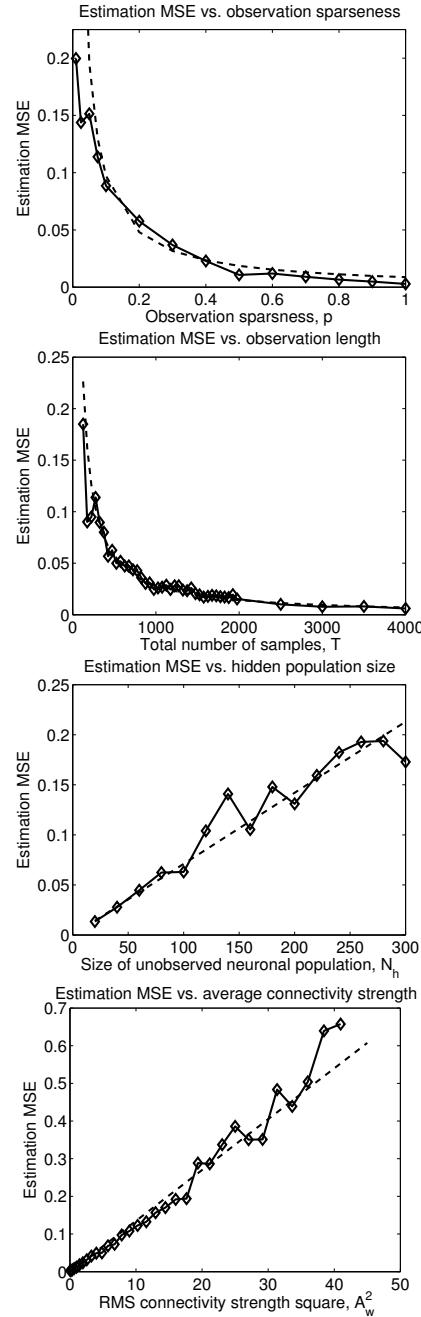
$$\text{var}(\hat{\mathcal{W}}_{ZX}) \approx s N_h A_w^2 / T_b, \quad (46)$$

where  $N_h$  is the number of neurons in the unobserved neuronal population,  $A_w$  is the root mean square average of the (nonzero) neuronal connectivity weights,  $s$  is the sparsity of the connectivity matrix, and  $T_b$  is the number of observations observing given input-output neuronal pairs.

Eq. (46) provides a useful approximation for the error of the shotgun connectivity estimator. We see that the primary source of that error is the uncontrolled fluctuations of the hidden inputs in the observed neurons. The error is affected by both the size of the hidden population and the strength of its coupling, and is reduced in the proportion to the total number of the observations of different input-output neuronal pairs.

### 3.4 Numerical simulations

In this section we demonstrate the recovery of the complete neuronal connectivity matrix from



**Fig. 3** The properties of the shotgun connectivity estimator in relation to the missing data. From top to bottom, the posterior error of the shotgun estimator is shown in relation to the observations sparseness, the total number of observations, the size of the hidden population, and the rms average connectivity strength. The results of numerical simulations are shown with solid lines and the theoretical predictions are shown with dashed lines. The model parameters are described in Table 1.

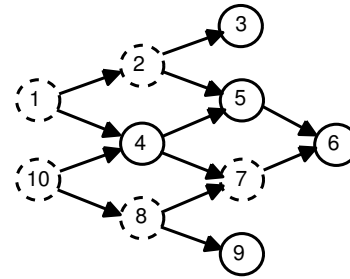
partial neuronal activity observations using simulated neuronal population activity data.

We first examine the problem of such a connectivity estimation in a small model synfire network of  $N = 10$  neurons, Figure 4. Synfire model represents one of the worst cases of the hidden inputs problem in the functional connectivity inference, as the correlations can propagate down the synfire chain over large distances and emulate false connections among distant neurons.

We first illustrate the connectivity estimation in this model in the population of 5 neurons indicated in Figure 4 with solid circles. This corresponds to the typical neuroimaging situation, in which a fixed neuronal population is continuously observed whereas the rest of the neurons or their inputs are hidden. As a measure of connectivity, we present in Figure 5 the calculation of the time-shifted correlogram (of lag 1 bin) for the selected neuronal population, widely used in the literature as a functional connectivity measure, and a generalized linear model (GLM) connectivity matrix estimation. We observe that either the correlogram and the GLM connectivity matrix produce spurious connections in this neuronal population, namely, such seen between the neurons 4 and 9, 3 and 6, and 9 and 6.

Next, in Figure 6 we perform the estimation of this neuronal connectivity matrix using the block-wise round-robin sampling strategy and the EM estimation algorithm described in Section 2.3.2. Despite the fact that at most 4 neurons are observed in this situation at a time, the complete neuronal connectivity matrix is recovered rather well. The strength of the false-positive connections between the neurons 4-9, 3-6 and 9-6 is also reduced by a factor of 2 to 3 compared to the reconstruction in Figure 5. At the same time, the obtained connectivity matrix is substantially more noisy and requires more observations to suppress that noise. Typically, we observe that the accuracy of the shotgun connectivity estimation with the observation time  $T$  is comparable to that using the full observations and the observations duration  $p^2T$ , essentially in agreement with Eq. (46), the bottom panel in Figure 6.

In Figure 7, we apply the shotgun approach to the estimation of the complete connectivity matrix in a model of realistic weakly coupled cortical neuronal network with  $N = 50$  neurons. The shotgun estimation again allows



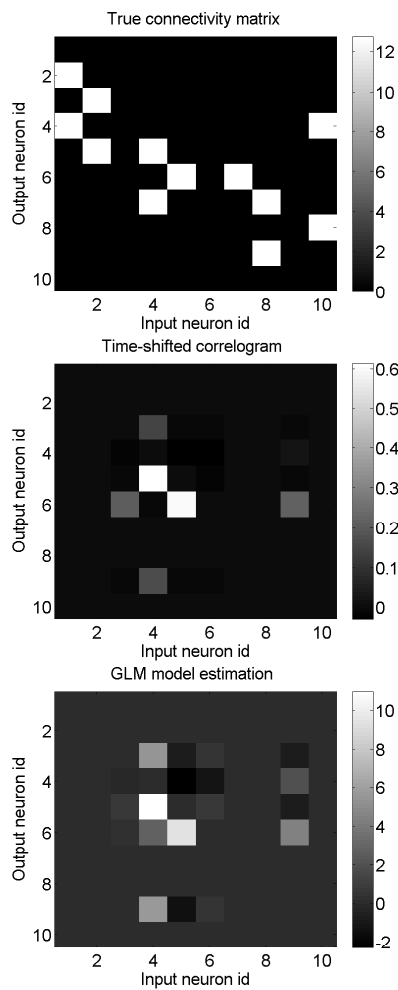
**Fig. 4** A model fragment of synfire neuronal circuit illustrating the resolution of the hidden inputs situation in the shotgun connectivity estimation.

recovering the complete neuronal connectivity matrix from partial neuronal population activity observations. Once again, the data size necessary to achieve a given accuracy is significantly greater than that required when using the complete neuronal activity. Specifically, if a good connectivity matrix reconstruction could be obtained in this simulation with all neurons observed from a total of  $T = 60$  seconds of neuronal activity data, in  $p = 20\%$  block-wise round-robin sampling case the imaging time necessary for a similar reconstruction accuracy approached  $T = 3000$  seconds, the lower panel in Figure 7.

## 4 Discussion and Conclusions

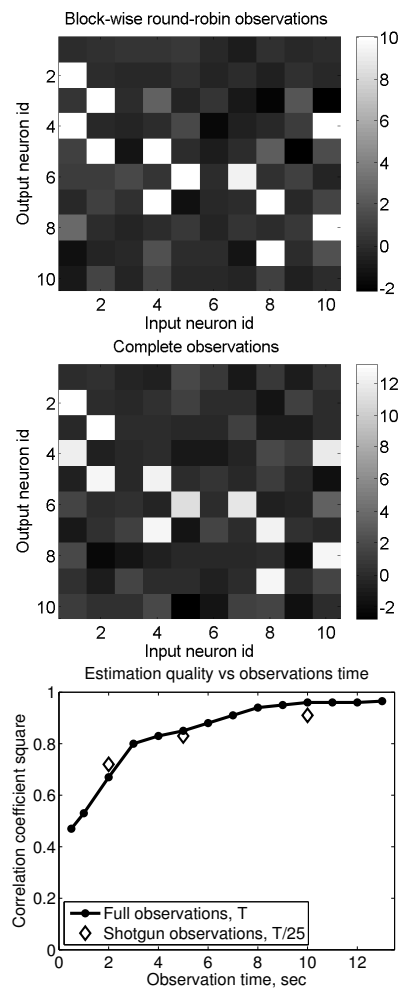
The shotgun sampling solution of the common inputs problem is a promising new approach for the functional estimation of the neuronal connectivity in large neuronal networks in the brain without requiring the simultaneous imaging of entire neuronal populations. By statistically estimating the neuronal connectivity from a collection of partial observations of different neuronal sub-populations, the shotgun sampling offers a possibility of recovering the connectivity matrix in realistically large neuronal circuits using limited imaging resources.

In this paper, we investigate analytically and in simulations the properties of such shotgun neuronal connectivity estimation. In Theorem 1 of Section 3.1 we establish the sufficient and necessary conditions for the complete connectivity matrix of a neuronal population to be recoverable from shotgun-type incomplete neuronal activity observations. In Section 2.2.2, 2.2.3 and 2.2.4, we discuss the shotgun connectivity estimation in a linear neuronal activity model



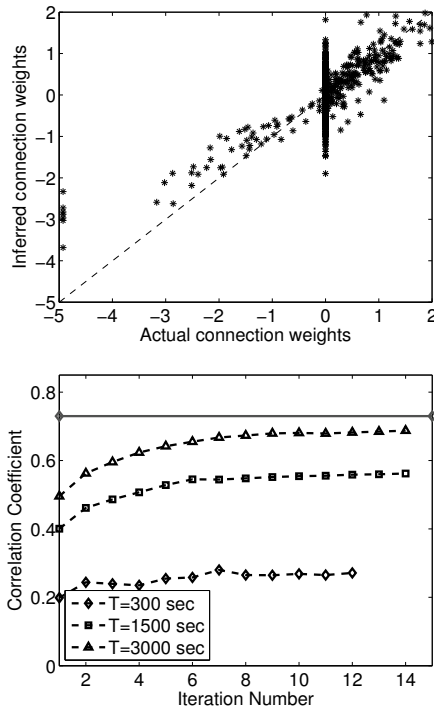
**Fig. 5** The hidden inputs problem in the model synfire circuit in Figure 4. The top panel shows the true connectivity matrix for the circuit in Figure 4. The middle panel shows the time-shifted correlogram calculated for the population if neurons indicated in Figure 4 with solid circles, from  $T = 10$  seconds of observation. The bottom panel shows the GLM connectivity matrix estimation for the same population from  $T = 10$  seconds of neuronal activity data. The simulation parameters are as described in Table 2.

more explicitly, and analytically derive its shotgun observations likelihood function, associated maximum-likelihood estimator, and the key properties of that estimator. In Section 2.2.5, 2.2.6 and 2.2.7, we explicitly establish the sufficient conditions for the correctness of the shotgun connectivity estimation in linear and exponential GLM neuronal activity models as well as in general spiking models of neuronal population activity, where the neuronal firing rates are described by a general nonlinear function of linearly summed inputs. The latter, in particular, covers a great variety of general network models of neuronal populations activity.



**Fig. 6** Estimation of the complete neuronal connectivity in the model synfire circuit in Figure 4 using the block-wise round-robin activity sampling. The top panel shows the result of the neuronal connectivity reconstruction using this approach and  $T = 50$  seconds of neuronal activity data. The middle panel shows, for comparison, a similar reconstruction but using the complete observations and  $T = 10$  seconds of neuronal activity data. The bottom panel compares the quality of the estimated connectivity matrices for the block-wise round-robin approach (diamonds) and the complete data (solid line). The shotgun approach points are placed at the "equivalent" time calculated as  $T' = p^2 T$ . The simulation parameters are given in Table 2.

In the linear and the exponential-GLMs of neuronal activity in the limit of large number of neurons, we find that the shotgun estimation is guaranteed to produce the complete connectivity matrix whenever all possible input-output pairs of neurons as well as all the pairs of same-time neuronal activities had been observed in at least some parts of a shotgun sampling experiment. In these models, constraining all elements of the same-time and time-shifted co-



**Fig. 7** The shotgun connectivity estimation in a model of a realistic weakly coupled cortical neuronal network. The upper panel shows the reconstruction of the complete connectivity matrix using  $p = 0.2$  block-wise round-robin neuronal activity observations in such network and  $T = 3000$  seconds of neuronal activity data. The multiplicative bias seen in the reconstructed weights is the finite time-discretization bias discussed in (Mishchenko et al., 2011) (not corrected here). The reconstructed connection weights reproduce the true connectivity, however, are rather noisy. The bottom panel shows the correlation coefficient of the reconstructed connection weights vs the true connectivity as a function of the EM algorithm’s iteration number, for  $T = 300, 1500$  and  $3000$  seconds of neuronal activity data. Solid gray line represents the baseline reconstruction produced using the complete observations and  $T = 60$  seconds of neuronal activity data. The simulation parameters are given in Table 3.

variance matrices,  $\Sigma = E[\mathcal{X}_t \mathcal{X}_t^T]$  and  $\Sigma_1 = E[\mathcal{X}_{t+1} \mathcal{X}_t^T]$  respectively, suffices to uniquely identify the full connectivity matrix as well as offers a simple way for deducing the connectivity matrix with modest computational effort, as discussed in Section 2.2.5 and 2.2.6. For general spiking GLM of neuronal population activity, we find that the set of observations of all neuronal activities in the form  $(\mathcal{X}_{it}, \mathcal{X}_{jt})$ ,  $(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$  and  $(\mathcal{X}_{it}, \mathcal{X}_{jt}, \mathcal{X}_{kt})$  can be guaranteed to allow the recovery of the full connectivity matrix, as discussed in Section 2.2.7.

We find that, whenever the full connectivity matrix is uniquely identified by a set of obser-

vations, the observations likelihood function is as well guaranteed to have the correct connectivity matrix as a unique global maximum. At the same time, observations likelihood cannot be guaranteed to have a unique local optimum. Already in the case of the linear neuronal activity model, where the MLE solution in the full-observations case is globally and locally unique, we find the possibility of two local optima—the true solution  $\hat{\mathcal{W}}_{true} = \mathcal{W}$  and the “mirror” solution  $\hat{\mathcal{W}}_{mirror} \approx -\mathcal{W}$ , Section 2.2.2.

The mirror optimum-like solutions clearly will need to be avoided in practice. Simple heuristics may be able to achieve this objective in many cases, such as inspecting the likelihood of the “mirror” solution for any found locally optimal  $\hat{\mathcal{W}}$ . However, this strategy may not be plausible when the fraction of observed neurons  $p$  tends to zero, as we can show that the difference between the likelihoods of the true and the mirror solutions in this case quickly tends to zero. Alternatively, the resolution of the mirror-optima can be based on the use of a-priori information. Such a-priori information may include the identity of one or several excitatory or inhibitory neurons in the population or the relative abundances of the excitatory and inhibitory neuronal populations, which are rather easy to assess. By requiring given neurons in the reconstruction to be excitatory or by requiring a particular split between the abundances of the excitatory and inhibitory neuronal populations, the sign of the solution can be fixed trivially.

We find that the lack of the observations of the entire neuronal population in the shotgun estimation leads to an increase in the statistical error of the connectivity estimator, warranting a respective increase in the imaging time necessary to suppress that error. We find that the required shotgun sampling data size scales proportionally to the number of neurons in the unobserved neuronal populations and the average square neuronal connectivity strength. The data size also increases as the inverse square of the fraction of neurons in the shotgun observations. This scaling is inopportune for the reconstructions of neuronal connectivity where the fraction of observed neurons will remain small.

In Section 2.3.2, we propose an exact numerical approach for solving the shotgun connectivity estimation problem in general settings, and demonstrate its applications to the estimation of the complete neuronal connectivity ma-

trix in different model neuronal circuits, including linear neuronal models, a spiking synfire network, and a small realistic weakly coupled cortical neuronal network. In all cases we find that the shotgun sampling succeeds in recovering the complete neuronal connectivity matrix from partial activity observations.

In Section 3.1, we prove important Theorem 1 that sets the basis for the subsequent investigation of the correctness of the shotgun connectivity estimation in different neuronal population activity models. This theorem also established important results about the design of possible shotgun-type neuronal connectivity imaging experiments. Theorem 1 establishes that any sparse neuronal activity imaging that covers a particular set of marginal "identifying" neuronal activity distributions is sufficient for the recovery of the complete connectivity matrix. In the linear model and the exponential-GLM of neuronal activity such sets of marginal neuronal activity distributions can be explicitly shown to consist of all time-shifted and same-time neuronal activity pairs, such as  $P(\mathcal{X}_{it}, \mathcal{X}_{jt})$  and  $P(\mathcal{X}_{i,t+1}, \mathcal{X}_{jt})$ . It is not important how such set of neuronal activity distributions was covered, what is important for the identifiability of full connectivity matrix is that such set is completely obtained. These results open new possibilities for investigating alternative shotgun-type neuronal population activity sampling organizations optimized for different experiment parameters.

In Section 3.2, we propose a particularly advantageous such alternative organization consisting of block-wise sequential scanning of a neuronal population's activity. In this approach, the neurons are observed in two contiguous input and output blocks, sequentially, with all possible combinations of input and output blocks imaged during the experiment. This sampling organization guarantees the coverage of all same-time and time-shifted neuronal activity pairs and furthermore has the advantage of allowing straightforward implementations using existing fluorescent microscopy tools as well as simpler associated numerical connectivity estimation problem. At the same time, truly random sampling of the neuronal activity in a large neuronal population proposed in (Turaga et al., 2013; Keshri et al., 2013) is significantly more challenging, both from the experimental implementation stand-point and the solution of associated estimation problem. Thus proposed

"block-wise round-robin" sampling strategy is simple conceptually and can be straightforwardly implemented experimentally, which hopefully will allow it to be realized in experimental designs by researchers in the near future.

Another particularly important issue raised by this study in Sections 3.3 and 3.4 is that of the scalability of the numerical shotgun connectivity estimation. We demonstrated here the shotgun connectivity estimation in simulated neuronal networks of up to  $N = 50$  neurons. In practice, the reconstructions of real neuronal circuits will require solving this estimation problem for thousands and even millions of neurons. The SMC EM procedure described here is efficient, requiring  $O(N^2 M^2 T)$  time for the E-step and  $O(N^2 MT)$  time for the M-step, as well as parallelizable, making the solution of the above problems possible using the high-performance computing infrastructures currently existing in the world. Faster sampling schemes for hidden neuronal population's activity may be proposed, for example, by using the fast Metropolis-Hastings algorithm described in (Mishchenko and Paninski, 2011), reducing the cost of the E-step to  $O(N^2 MT)$ .

At the same time, in this work the requirement to store up to  $M$  examples of the entire hidden neuronal population's activity was the most significant burden on the numerical computation. For example, for  $N = 10^4$  neurons,  $M = 100$  EM samples, and  $T = 10^4$  seconds of neuronal activity data recorded at 100 Hz, the required sample of the hidden neuronal activities constitutes and requires storing in the computer memory of staggering  $10^{12}$  neuronal activity states. While this problem can be solved by partitioning over the nodes of a supercomputing infrastructure, research of alternative approaches for reducing the memory footprint of the shotgun connectivity estimation appears to be of relevance.

Among possible solutions is the use of alternative sampling schemes such as the block-wise round-robin strategy above, where the connectivity matrix estimation can be solved as a sequence of smaller problems, in which only the part of the experiment corresponding to a given input and output blocks is considered. While not affecting  $N$  and  $M$ , this reduces  $T$  to at most such containing the given input and output neuronal blocks. Another alternative is to make use of the sufficient statistics of the connectivity estimation problem in a given neu-

ronal activity model. This would allow using the smaller sufficient statistics in the place of large hidden neuronal population activity samples. Finally, one other approach can consist in using approximate models for the combined inputs from hidden neuronal populations. For example, one can model the inputs into observed neurons from hidden population using a multivariate Gaussian distribution constructed using the current estimate of the connectivity matrix in the hidden and the observed neuronal population sectors.

Several issues remain open in this work and require future investigation. Among these is the issue of the identifying sets of marginal neuronal activity distributions for given neuronal population activity models. We proposed here, based on the degrees-of-freedom counting, that the sets of all pair-wise input-output neuronal activity distributions would suffice to uniquely constrain any network model of neuronal population activity parametrized by a single connectivity matrix  $\mathcal{W}$  of  $N^2$  weights, where  $N$  is the number of neurons in the model. How can this conjecture be established rigorously? What are the identifying sets of the generalized linear model of neuronal activity defined by Eqs. (40-41) for general  $f(\cdot)$ ? What are the sufficient statistics of different network models of neuronal activity, that can be used in the connectivity matrix estimation? Do time-shifted correlation matrices provide the sufficient statistics for network models of neuronal activity?

Another set of questions is related to the design of the shotgun-like neuronal activity sampling strategies. Does the block-wise round-robin strategy offer the same speed of accumulating the information about different input-output neuronal pairs as the random sampling in the original shotgun proposal? What are other strategies that allow uniformly collecting pairwise input-output neuronal activity measurements? What are their important properties? Are all identifying sets equivalent from the point of view of accruing information about neuronal connectivity or some sets are more advantageous than the others? We hope that this work will stimulate further discussion regarding these questions.

**Acknowledgements** The author acknowledges the support via the TUBITAK ARDEB 1001 research grant number 113E611 (Ankara, Turkey), Toros University BAP grant number TUBAP135001 (Mersin, Turkey), and Bilim Akademisi—The Science Academy’s

young investigator award grant under the BAGEP program (Istanbul, Turkey). The author wants to thank Liam Paninski for the key discussion leading to the development of this work, and Daniel Soudry for valuable comments on the manuscript’s draft. The author is also thankful to the anonymous reviewers, whose comments allowed the author to significantly improve the manuscript.

### A Sequential Monte Carlo Expectation Maximization algorithm for numerical solution of the shotgun connectivity estimation problem

The EM algorithm (Dempster et al., 1977) is the standard method of statistical inference in the presence of missing data. Briefly, the EM algorithm produces at least a locally maximum likelihood estimate of the parameters of a model  $P(X, Y|\theta)$  given a set of observations  $X$  with the data  $Y$  missing,  $\hat{\theta} = \arg \max_{\theta} \sum_Y P(\theta, Y|X)$ . The EM algorithm produces a sequence of parameter estimates  $\hat{\theta}^k$  by iteratively maximizing the functions  $Q(\theta|\hat{\theta}^k)$ ,

$$Q(\theta|\hat{\theta}^k) = E_{P(Y|X, \hat{\theta}^k)}[\log P(X, Y, \theta)], \quad (47)$$

where  $Q(\theta|\hat{\theta}^k)$  at each step is calculated by constructing  $M$  samples of the unavailable data  $Y$  from  $P(Y|X, \hat{\theta}^k)$  and using the following average,

$$Q(\theta|\hat{\theta}) = \frac{1}{M} \sum_{l=1}^M \log P(X, Y^l, \theta). \quad (48)$$

In the case of the shotgun sampling, the sampling step of the EM algorithm can be implemented using the forward-backward algorithm (Rabiner, 1989) and the sequential Monte-Carlo method also known as the Particle Filtering (Godsill et al., 2001). In this case, the distribution of the hidden neuronal activities at every observation is modeled by a sample of  $M$  hidden neurons’ activity configurations,  $Y_t^l \sim P(Y_t|X, \mathcal{W})$ , each referred to as a “particle”.

In order to produce this sample, it is advantageous to reformulate the sampling problem  $Y_t \sim P(Y_t|X, \mathcal{W})$  in a more convenient way as applying to drawing a sample of the complete neuronal activity configurations  $\mathcal{X}_t$  in such a way that the activity of the parts of the neuronal population observed at time  $t$  match the available neuronal activity data  $X_t$ . In this sense, we view the activity of the entire neuronal population  $\mathcal{X}_t$  as “hidden” and the mapping of  $\mathcal{X}_t$  onto the subset of observed neurons,  $\mathcal{X}_t \mapsto X_t$ , as the observation. In this form, the problem becomes that of sampling the sequence of the hidden states  $\mathcal{X}_t$  from a Hidden Markov Model with the observations  $X_t$ . This problem now can be efficiently solved using the standard forward-backward algorithm.

Forward-backward algorithm consists of two passes. In the first forward pass, a sequence of samples of hidden neural activity states is produced according to  $P(\mathcal{X}_t|X_{1:t}, \mathcal{W})$ , where  $X_{1:t}$  refers to the collection of all observed neuronal activities up to and including the time  $t$ . Each sample in this sequence contains  $M$  examples of the complete neuronal population activity,  $\mathcal{X}_t \sim P(\mathcal{X}_t|X_{1:t}, \mathcal{W})$ , while the entire sequence

contains  $T$  such samples  $t = 1 \dots T$ , where  $T$  is the number of the observations,  $\{\mathcal{X}_t^k, k = 1 \dots M, t = 1 \dots T\}$ .

Forward pass samples can be constructed iteratively by drawing the first sample  $\mathcal{X}_0$  from the prior distribution  $P(\mathcal{X}_0)$ , and then constructing each next sample according to,

$$\mathcal{X}_t \sim P(\mathcal{X}_t | X_{1:t}) = \mathcal{Z}^{-1} \sum_{\mathcal{X}_{t-1}} P(\mathcal{X}_t | \mathcal{X}_t) P(\mathcal{X}_t | \mathcal{X}_{t-1}) P(\mathcal{X}_{t-1} | X_{1:t-1}). \quad (49)$$

Here  $\mathcal{Z}$  is a normalization constant to be calculated below and we stopped writing parameter  $\mathcal{W}$  in the probability densities for brevity.

According to Eq. (49), the forward pass step at each  $t$  can be realized by taking the previous sample's particles  $\mathcal{X}_{t-1}^k \sim P(\mathcal{X}_{t-1} | X_{1:t-1})$  and "moving" them according to the transition probabilities

$$P(\mathcal{X}_{t-1}^k \rightarrow \mathcal{X}_t^k) = \mathcal{Z}^{-1} P(\mathcal{X}_t | \mathcal{X}_t^k) P(\mathcal{X}_t^k | \mathcal{X}_{t-1}^k). \quad (50)$$

Eq. (50) can be simplified by noting that  $P(\mathcal{X}_t | \mathcal{X}_t)$  has the effect of only restricting the moves  $\mathcal{X}_{t-1}^k \rightarrow \mathcal{X}_t^k$  to such that make the activity patterns of the neurons observed in  $\mathcal{X}_t^k$  match the available observation  $X_t$ ,

$$P(\mathcal{X}_t | \mathcal{X}_t^k) \propto \begin{cases} 1 & \text{if } X_t^k = X_t \\ 0 & \text{otherwise} \end{cases} \quad (51)$$

By using this and taking advantage of the factorization of the probabilities  $P(\mathcal{X}_t | \mathcal{X}_{t-1})$  over individual neurons  $i$ ,  $P(\mathcal{X}_t | \mathcal{X}_{t-1}) = \prod_i P(\mathcal{X}_{i,t} | \mathcal{X}_{i,t-1})$ , we can obtain the normalization constant  $\mathcal{Z}$  explicitly as,

$$\mathcal{Z} = E_{\mathcal{X}_{t-1}} [P(\mathcal{X}_t | \mathcal{X}_{t-1})] = \frac{1}{M} \sum_k P(\mathcal{X}_t | \mathcal{X}_{t-1}^k). \quad (52)$$

With this simplification, we arrive at the final forward step algorithm as follows:

#### Forward Step

- (i) Select one  $\mathcal{X}_{t-1}^k$  from the previous  $t-1$  sample  $\mathcal{X}_{t-1}^k \sim P(\mathcal{X}_{t-1} | X_{1:t-1})$  with the probability

$$p(k) = 1/M \cdot P(\mathcal{X}_t | \mathcal{X}_{t-1}^k) / \mathcal{Z} = P(\mathcal{X}_t | \mathcal{X}_{t-1}^k) / \sum_k P(\mathcal{X}_t | \mathcal{X}_{t-1}^k),$$

where

$$P(\mathcal{X}_t | \mathcal{X}_{t-1}) = \sum_{\mathcal{X}_t} P(\mathcal{X}_t | \mathcal{X}_t) P(\mathcal{X}_t | \mathcal{X}_{t-1}).$$

- (ii) Set in  $\mathcal{X}_t^k$  the activity of the neurons  $i$  observed in observation  $t$  as  $\mathcal{X}_{i,t}^k = X_{i,t}$ ;
- (iii) Set in  $\mathcal{X}_t^k$  the activity of the neurons  $i'$  not observed in observation  $t$  as  $\mathcal{X}_{i',t}^k \sim P(\mathcal{X}_{i',t} | \mathcal{X}_{i',t-1}^k)$ .

In the backward pass, the samples  $(\mathcal{X}_{t-1}, \mathcal{X}_t) \sim P(\mathcal{X}_{t-1}, \mathcal{X}_t | X)$  need to be constructed for each  $t$  conditional on the all observations  $X = \{X_t, t = 1 \dots T\}$ . These samples can be constructed using the following relationship that we adopt here from (Paninski et al., 2010),

$$P(\mathcal{X}_t, \mathcal{X}_{t+1} | X) = P(\mathcal{X}_t | X_{1:t}) \frac{P(\mathcal{X}_{t+1} | \mathcal{X}_t)}{P(\mathcal{X}_{t+1} | X_{1:t})} P(\mathcal{X}_{t+1} | X), \quad (53)$$

where  $P(\mathcal{X}_{t+1} | X_{1:t}) = \sum_{\mathcal{X}_t} P(\mathcal{X}_{t+1} | \mathcal{X}_t) P(\mathcal{X}_t | X_{1:t}) = E_{\mathcal{X}_t} [P(\mathcal{X}_{t+1} | \mathcal{X}_t)]$ , the average being over the forward pass sample  $\mathcal{X}_t^k \sim P(\mathcal{X}_t | X_{1:t})$ .

According to Eq. (53), the backward step can be constructed by first combining into pairs the forward pass samples  $t$ ,  $\mathcal{X}_t^k \sim P(\mathcal{X}_t | X_{1:t})$ , and the backward pass samples  $t+1$ ,  $\mathcal{X}_{t+1}^l \sim P(\mathcal{X}_{t+1} | X)$ , and then weighing these with the weights  $w_t^{kl} = P(\mathcal{X}_{t+1}^l | \mathcal{X}_t^k) / \sum_k P(\mathcal{X}_{t+1}^l | \mathcal{X}_t^k)$ . Evidently, thus formed pairs  $(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l)$  are distributed according to  $(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l) \sim P(\mathcal{X}_t | X_{1:t}) P(\mathcal{X}_{t+1} | X)$ , and the expectation value of any functional  $F(\mathcal{X}_t, \mathcal{X}_{t+1})$  over  $P(\mathcal{X}_t, \mathcal{X}_{t+1} | X)$  can be calculated by using such pairs as  $E[F] = 1/M \sum_{kl} F(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l) w_t^{kl}$ . In addition,  $P(\mathcal{X}_t | X) = \sum_{\mathcal{X}_{t+1}} P(\mathcal{X}_t, \mathcal{X}_{t+1} | X)$  and the next backward pass sample for observation  $t$ ,  $\mathcal{X}_t^k \sim P(\mathcal{X}_t | X)$ , can be constructed by drawing with replacement  $\mathcal{X}_t^k$  from  $(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l)$  with probabilities  $p(k) \propto \sum_l w_t^{kl}$ .

Thus, we arrive at the final backward step algorithm as follows:

#### Backward Step

- (i) Form  $M^2$  pairs  $(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l)$  for each available forward pass sample  $\mathcal{X}_t^k \sim P(\mathcal{X}_t | X_{1:t})$  and backward pass sample  $\mathcal{X}_{t+1}^l \sim P(\mathcal{X}_{t+1} | X)$ ;
- (ii) Calculate the weights  $w_t^{kl} = P(\mathcal{X}_{t+1}^l | \mathcal{X}_t^k) / \sum_k P(\mathcal{X}_{t+1}^l | \mathcal{X}_t^k)$ ;
- (iii) As the next backward pass sample  $\mathcal{X}_t^k \sim P(\mathcal{X}_t | X)$  select with replacement  $\mathcal{X}_t^k$  from the pairs  $(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l)$  with the probabilities  $p(k) = 1/M \sum_l w_t^{kl}$ ;
- (iv) The expectations values of a functional  $F(\mathcal{X}_t, \mathcal{X}_{t+1})$ ,  $E_{P(\mathcal{X}_t, \mathcal{X}_{t+1} | X)} [F(\mathcal{X}_t, \mathcal{X}_{t+1})]$ , are given by  $E[F] = 1/M \sum_{kl} F(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l) w_t^{kl}$ .

In the optimization step of the EM algorithm we maximize with respect to  $\mathcal{W}$  the following function,

$$Q(\mathcal{W} | \hat{\mathcal{W}}) = E_{P(Y|X, \hat{\mathcal{W}})} [\log P(X, Y, \mathcal{W})] = \log P(\mathcal{W}) + E_{P(\mathcal{X}_0 | X, \hat{\mathcal{W}})} [\log P(\mathcal{X}_0)] + \sum_t E_{P(\mathcal{X}_{t-1}, \mathcal{X}_t | X, \hat{\mathcal{W}})} [\log P(\mathcal{X}_t | \mathcal{X}_{t-1}, \mathcal{W})]. \quad (54)$$

In order to calculate  $Q(\mathcal{W} | \hat{\mathcal{W}})$  it is sufficient to know the samples  $(\mathcal{X}_t^k, \mathcal{X}_{t+1}^l) \sim P(\mathcal{X}_t | X_{1:t}) P(\mathcal{X}_{t+1} | X)$  and the weights  $w_t^{kl}$ . Moreover,  $Q(\mathcal{W} | \hat{\mathcal{W}})$  can be split into a sum over the rows of the matrix  $\mathcal{W}$ ,  $W_i$ , as  $Q(\mathcal{W} | \hat{\mathcal{W}}) = \sum_i Q(W_i | \hat{\mathcal{W}})$ , with  $Q(W_i | \hat{\mathcal{W}})$  given by

$$Q(W_i | \hat{\mathcal{W}}) = \log P(W_i) + E_{P(\mathcal{X}_0 | X, \hat{\mathcal{W}})} [\log P(\mathcal{X}_{i0})] + \sum_t E_{P(\mathcal{X}_{t-1}, \mathcal{X}_t | X, \hat{\mathcal{W}})} [\log P(\mathcal{X}_{it} | \mathcal{X}_{i,t-1}, W_i)]. \quad (55)$$

Thus, the optimization of Eq. (54) can be solved for each row  $i$  independently, reducing the complexity of the problem from quadratic in the number of neurons  $N$  to linear. Finally, inhomogeneous Poisson point-process models of neuronal activity with log-concave rate functions result in the problems  $Q(W_i | \hat{\mathcal{W}})$  that are convex, which allows their efficient numerical optimization for very large  $N$  using the standard gradient descent methods (Paninski, 2004).

## B The calculation of posterior log-likelihood for the linear neuronal activity model with input neuronal populations.

In this appendix we calculate the marginal likelihood

$$P(Z_t, X_t | \hat{W}) \propto \int dY_t \exp\left(- (Z_t - W X_t)^2 / 2 - \mathcal{X}_t^T C^{-1} X_t / 2\right), \quad (56)$$

of the model (3), where the input neuronal activities are distributed according to a correlated Gaussian distribution with covariance matrix  $C$ ,

$$P(\mathcal{X}_t) \propto \exp(-\mathcal{X}_t^T C^{-1} X_t / 2)$$

and the integration is performed over the part of  $\mathcal{X}_t$ ,  $Y_t$ , that is not observed during observation  $t$ . The part of  $\mathcal{X}_t$  observed during observation  $t$ , respectively, is held fixed at  $X_t$ .  $W$  is a single row-vector from the full connectivity matrix  $\mathcal{W}$  corresponding to the input connection weights of one ‘‘output’’ neuron.

The calculation of Eq. (56) can be simplified if we represent the integral in an invariant form by introducing  $\delta$ -functions, which will restrict the integration over  $\mathcal{X}_t$  to the hyperplane of the observed neuronal activities  $X_t$ , namely,

$$\int dY_t \exp\left(- (Z_t - W X_t)^2 / 2 - \mathcal{X}_t^T C^{-1} X_t / 2\right) = \int d\mathcal{X}_t \prod_{i=1}^{i=m} \delta(\mathcal{X}_{it} - X_{it}) \times \exp\left(- (Z_t - W X_t)^2 / 2 - \mathcal{X}_t^T C^{-1} X_t / 2\right). \quad (57)$$

Here  $m$  is the number of the observed neuronal inputs and w.l.o.g. we assumed that the observed inputs  $X_t$  comprise the first  $m$  elements of  $\mathcal{X}_t$ . We now replace the  $\delta$ -functions in Eq. (57) with their Fourier representation,  $\delta(x) = \frac{1}{2\pi} \int dk e^{-ikx}$ , yielding

$$\int d\mathcal{X}_t dK \prod_{i=m+1}^{i=N} \delta(K_i) \exp\left(- (Z_t - W X_t)^2 / 2 - iK^T (\mathcal{X}_t - \bar{X}_t) - \mathcal{X}_t^T C^{-1} X_t / 2\right), \quad (58)$$

where  $\bar{X}_t$  is a full-size column-vector of neuronal inputs, with the first  $m$  elements equal to  $X_t$  and the rest of the elements zero (these do not affect the integral since  $K_i = 0$  for  $i > m$ ). In Eq. (58), the integral over  $\mathcal{X}_t$  now can be taken explicitly as a Gaussian, resulting in

$$\int dK \prod_{i=m+1}^{i=N} \delta(K_i) \exp(i\bar{X}_t^T K) \sqrt{\det \Gamma} \times \exp\left(- Z_t^2 / 2 + (Z_t W - iK^T) \Gamma (Z_t W^T - iK) / 2\right), \quad (59)$$

where the matrix  $\Gamma$  is identified from the part of the argument of the exponential in Eq. (58) quadratic

in  $\mathcal{X}_t$ ,  $\Gamma^{-1} = (C^{-1} + W^T W)$ . We expand the second term under the exponential in Eq. (59) as

$$\int dK \prod_{i=m+1}^{i=N} \delta(K_i) \sqrt{\det \Gamma} \times \exp\left(- Z_t^2 / 2 + Z_t^2 W \Gamma W^T / 2 + i\bar{X}_t^T K - iZ_t W \Gamma K - K^T \Gamma K / 2\right). \quad (60)$$

The  $\delta$ -functions in Eq. (60) can be used subsequently to restrict the integration over  $K$  to only such values where  $K_i = 0$  for all  $m < i \leq N$ . Thus, we rewrite this integration as

$$\int dK_X \sqrt{\det \Gamma} \times \exp\left(- Z_t^2 / 2 + Z_t^2 W \Gamma W^T / 2 + i(\bar{X}_t^T - iZ_t W \Gamma)_X K_X - K_X^T \Gamma_X K_X / 2\right), \quad (61)$$

where the subscript  $X$  means restriction to the first  $m$  elements, as contained in the observed set of neuronal inputs  $X_t$ . Thus obtained integration over  $K_X$  is again Gaussian, and so we can perform it explicitly producing

$$\int dY_t \exp(- (Z_t - W X_t)^2 / 2 - \mathcal{X}_t^T C^{-1} X_t / 2) \propto \sqrt{\frac{\det \Gamma}{\det \Gamma_X}} \exp\left(- Z_t^2 / 2 + Z_t^2 W \Gamma W^T / 2 - (\bar{X}_t^T - Z_t W \Gamma)_X \Gamma_X^{-1} (\bar{X}_t - Z_t \Gamma W^T)_X / 2\right). \quad (62)$$

A simple check is in order now. By assuming  $C = I$  (uncorrelated inputs), we obtain by repeatedly using Woodbury lemma,

$$\begin{aligned} \Gamma &= (I + W^T W)^{-1} = I - W^T W / (1 + W^2) \\ \Gamma_X &= I - W_X^T W_X / (1 + W^2) \\ \Gamma_X^{-1} &= I + W_X^T W_X / (1 + W_X^2) \\ (W \Gamma) &= W / (1 + W^2) \\ (W \Gamma)_X &= W_X / (1 + W^2) \end{aligned} \quad (63)$$

where  $W_X$  and  $W_Y$  are the restrictions of  $W$  to the subsets of the neuronal inputs  $X_t$  and  $Y_t$ , respectively, and  $W^2 = W W^T$ . For  $\det \Gamma$  and  $\det \Gamma_X$ , we then obtain  $\det \Gamma = (1 + W^2)^{-1}$  and  $\det \Gamma_X = (1 + W_Y^2) / (1 + W^2)$ , therefore,

$$\det \Gamma / \det \Gamma_X = 1 / (1 + W_Y^2).$$

Similarly, we calculate

$$\begin{aligned} -Z_t^2 + Z_t^2 W \Gamma W^T &= -Z_t^2 / (1 + W^2) \\ X_t^T \Gamma_X^{-1} X_t &= X_t^2 + (W_X X_t)^2 / (1 + W_Y^2) \\ Z_t (W \Gamma)_X \Gamma_X^{-1} X_t &= Z_t (W_X X_t) / (1 + W_Y^2) \\ Z_t^2 (W \Gamma)_X \Gamma_X^{-1} (\Gamma W^T)_X &= -Z_t^2 W_X^2 / ((1 + W^2)(1 + W_Y^2)) \end{aligned} \quad (64)$$

Combining all of Eqs.(64), we obtain for Eq. (61) and the case  $C = I$ ,

$$\int dY_t \exp(- (Z_t - W X_t)^2 / 2 - \mathcal{X}_t^T C^{-1} X_t / 2) = (1 + W_Y^2)^{-1/2} \exp\left(- \frac{1}{2} \left[ \frac{Z_t^2}{1 + W_Y^2} + \frac{2Z_t W_X X_t}{1 + W_Y^2} - \frac{(W_X X_t)^2}{1 + W_Y^2} \right] - \frac{X_t^2}{2}\right), \quad (65)$$

which is the same as Eq. (??) in the main text.



In the case of general  $C$ , we calculate similarly by repeatedly using Woodbury lemma

$$\begin{aligned} \Gamma &= C - C \frac{W^T W}{1 + W C W^T} C \\ \Gamma_X &= C_{XX} - C_{X*} \frac{W^T W}{1 + W C W^T} C_{*X} \\ \Gamma_X^{-1} &= C_{XX}^{-1} + C_{XX}^{-1} \frac{C_{X*} W^T W C_{*X}}{1 + B^2} C_{XX}^{-1} \\ (W\Gamma) &= \frac{WC}{(1 + W C W^T)} \\ (W\Gamma)_X &= \frac{W C_{*X}}{(1 + W C W^T)} \end{aligned} \quad (66)$$

where

$$B^2 = W C W^T - W C_{*X} C_{XX}^{-1} C_{X*} W^T,$$

and  $C_{XX}$  is the square block of the full covariance matrix  $C$  corresponding to the observed inputs  $X_t$ , while  $C_{X*}$  and  $C_{*X}$  are the rectangular blocks of the full covariance matrix containing all the rows or the columns corresponding to the observed inputs  $X_t$ . Then, for the determinant factor we obtain,

$$\det \Gamma = \det C / (1 + W C W^T)$$

and

$$\det \Gamma_X = \det C_{XX} (1 + B^2) / (1 + W C W^T).$$

Consequently,

$$\sqrt{\frac{\det \Gamma}{\det \Gamma_X}} = \sqrt{\frac{\det C}{\det C_{XX}}} \cdot (1 + B^2)^{-1/2}. \quad (67)$$

By considering

$$C = \begin{bmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{bmatrix},$$

we can also reorder the quantity  $B^2$  as

$$B^2 = W_Y (C_{YY} - C_{YX} C_{XX}^{-1} C_{XY}) W_Y^T.$$

We see that  $B^2$  plays the role here of  $W_Y^2$  in Eq. (65). We further proceed to simplify all the terms under the exponent in Eq. (62) using Eqs.(66). This calculation is highly tedious, however, its result is obtained in the form,

$$\begin{aligned} \int dY_t \exp(-Z_t - W \mathcal{X}_t)^2 / 2 - \mathcal{X}_t^T C^{-1} \mathcal{X}_t / 2) \propto \\ \sqrt{\frac{\det C}{\det C_{XX}}} (1 + B^2)^{-1/2} \\ \times \exp\left(-\frac{1}{1+B^2} \frac{Z_t^2}{2} - \frac{1}{2} X_t^T C_{XX}^{-1} X_t \right. \\ \left. + \frac{Z_t W C_{*X} C_{XX}^{-1} X_t}{1+B^2} - \frac{1}{2} \frac{(W C_{*X} C_{XX}^{-1} X_t)^2}{1+B^2}\right). \end{aligned} \quad (68)$$

As a consistency check, we can verify that Eq. (68) reduces to Eq. (65) when  $C = I$ . Finally, we rewrite Eq. (68) more concisely as

$$\begin{aligned} \int dY_t \exp(-Z_t - W \mathcal{X}_t)^2 / 2 - \mathcal{X}_t^T C^{-1} \mathcal{X}_t / 2) \propto \\ \sqrt{\frac{\det C}{\det C_{XX}}} (1 + B^2)^{-1/2} \\ \times \exp\left(-\frac{1}{2} \frac{(Z_t - W C_{*X} C_{XX}^{-1} X_t)^2}{1+B^2} - \frac{1}{2} X_t^T C_{XX}^{-1} X_t\right) \end{aligned} \quad (69)$$

## References

- [Abeles, 1991]Abeles, M. (1991). *Corticonics*. Cambridge University Press.
- [Braitenberg and Schuz, 1998]Braitenberg, V. and Schuz, A. (1998). *Cortex: statistics and geometry of neuronal connectivity*. Springer, Berlin.
- [Brillinger, 1988]Brillinger, D. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59:189–200.
- [Brillinger, 1992]Brillinger, D. (1992). Nerve cell spike train data analysis: a progression of technique. *Journal of the American Statistical Association*, 87:260–271.
- [Chornoboy et al., 1988]Chornoboy, E., Schramm, L., and Karr, A. (1988). Maximum likelihood identification of neural point process systems. *Biological Cybernetics*, 59:265–275.
- [Cossart et al., 2003]Cossart, R., Aronov, D., and Yuste, R. (2003). Attractor dynamics of network up states in the neocortex. *Nature*, 423:283–288.
- [Dempster et al., 1977]Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal Royal Stat. Soc., Series B*, 39:1–38.
- [Djurisic et al., 2004]Djurisic, M., Antic, S., Chen, W. R., and Zecevic, D. (2004). Voltage imaging from dendrites of mitral cells: EPSP attenuation and spike trigger zones. *J. Neurosci.*, 24(30):6703–6714.
- [Godsill et al., 2001]Godsill, S., Doucet, A., and West, M. (2001). Maximum a Posteriori Sequence Estimation Using Monte Carlo Particle Filters. *Annals of the Institute of Statistical Mathematics*, 53(1):82–96.
- [Gomez-Urquijo et al., 2000]Gomez-Urquijo, S. M., Reblet, C., Bueno-Lopez, J. L., and Gutierrez-Ibarluzea, I. (2000). Gabaergic neurons in the rabbit visual cortex: percentage, distribution and cortical projections. *Brain Res*, 862:171–9.
- [Grewe et al., 2010]Grewe, B., Langer, D., Kasper, H., Kampa, B., and Helmchen, F. (2010). High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature Methods*, pages 399–405.
- [Iyer et al., 2006]Iyer, V., Hoogland, T. M., and Saggau, P. (2006). Fast functional imaging of single neurons using random-access multiphoton (RAMP) microscopy. *J Neurophysiol*, 95(1):535–545.
- [Keshri et al., 2013]Keshri, S., Pnevmatikakis, E., Pakman, A., Shababo, B., and Paninski, L. (2013). A shotgun sampling solution for the common input problem in neuronal connectivity inference. *arXiv:1309.3724*.
- [Koch, 1999]Koch, C. (1999). *Biophysics of Computation*. Oxford University Press.
- [Kulkarni and Paninski, 2007]Kulkarni, J. and Paninski, L. (2007). Common-input models for multiple neural spike-train data. *Network: Computation in Neural Systems*, 18:375–407.
- [Lefort et al., 2009]Lefort, S., Tomm, C., Floyd Sarria, J.-C., and Petersen, C. C. H. (2009). The excitatory neuronal network of the c2 barrel col-

- umn in mouse primary somatosensory cortex. *Neuron*, 61:301–16.
- [Mishchenko et al., 2011]Mishchenko, Y., Vogelstein, J., and Paninski, L. (2011). A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. *Annals of Applied Statistics*, 5:1229–61.
- [Nguyen et al., 2001]Nguyen, Q. T., Callamaras, N., Hsieh, C., and Parker, I. (2001). Construction of a two-photon microscope for video-rate  $\text{Ca}^{2+}$  imaging. *Cell Calcium*, 30(6):383–393.
- [Nykamp, 2007]Nykamp, D. (2007). A mathematical framework for inferring connectivity in probabilistic neuronal networks. *Mathematical Biosciences*, 205:204–251.
- [Nykamp, 2005]Nykamp, D. Q. (2005). Revealing pairwise coupling in linear-nonlinear networks. *SIAM J Applied Mathematics*, 65(6):2005–2032.
- [Ohki et al., 2005]Ohki, K., Chung, S., Ch'ng, Y., Kara, P., and Reid, C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, 433:597–603.
- [Paninski, 2004]Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15:243–262.
- [Paninski et al., 2010]Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29:107–126.
- [Paninski et al., 2004]Paninski, L., Fellows, M., Shoham, S., Hatsopoulos, N., and Donoghue, J. (2004). Superlinear population encoding of dynamic hand trajectory in primary motor cortex. *J. Neurosci.*, 24:8551–8561.
- [Pillow and Latham, 2007]Pillow, J. and Latham, P. (2007). Neural characterization in partially observed populations of spiking neurons. *NIPS*.
- [Pillow et al., 2008]Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A., Chichilnisky, E., and Simoncelli, E. (2008). Spatiotemporal correlations and visual signaling in a complete neuronal population. *Nature*, 454:995–999.
- [Plesser and Gerstner, 2000]Plesser, H. and Gerstner, W. (2000). Noise in integrate-and-fire neurons: From stochastic input to escape rates. *Neural Computation*, 12:367–384.
- [Rabiner, 1989]Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 72(2):257–286.
- [Reddy et al., 2008a]Reddy, G., Kelleher, K., Fink, R., and Saggau, P. (2008a). Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nature neuroscience*, 11:713–720.
- [Reddy et al., 2008b]Reddy, G., Kelleher, K., Fink, R., and Saggau, P. (2008b). Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nature Neuroscience*, 11(6):713–720.
- [Rigat et al., 2006]Rigat, F., de Gunst, M., and van Pelt, J. (2006). Bayesian modelling and analysis of spatio-temporal neuronal networks. *Bayesian Analysis*, 1:733–764.
- [Salome et al., 2006]Salome, R., Kremer, Y., Dieudonne, S., Leger, J.-F., Krichevsky, O., Wyart, C., Chatenay, D., and Bourdieu, L. (2006). Ultrafast random-access scanning in two-photon microscopy using acousto-optic deflectors. *Journal of Neuroscience Methods*, 154(1-2):161–174.
- [Sayer et al., 1990]Sayer, R. J., Friedlander, M. J., and Redman, S. J. (1990). The time course and amplitude of epsps evoked at synapses between pairs of ca3/cal neurons in the hippocampal slice. *J. Neurosci.*, 10:826–36.
- [Stevenson et al., 2008a]Stevenson, I., Rebesco, J., Hatsopoulos, N., Haga, Z., Miller, L., and Kording, K. (2008a). Inferring network structure from spikes. *Statistical Analysis of Neural Data meeting*.
- [Stevenson et al., 2009]Stevenson, I. H., Rebesco, J. M., Hatsopoulos, N. G., Haga, Z., Miller, L. E., and Kording, K. P. (2009). Bayesian inference of functional connectivity and network structure from spikes. *IEEE Trans. Neural Systems and Rehab.*, 17:203–13.
- [Stevenson et al., 2008b]Stevenson, I. H., Rebesco, J. M., Miller, L. E., and Kording, K. P. (2008b). Inferring functional connections between neurons. *Curr. Opin. Neurobiol.*, 18:582–8.
- [Stosiek et al., 2003]Stosiek, C., Garaschuk, O., Holthoff, K., and Konnerth, A. (2003). In vivo two-photon calcium imaging of neuronal networks. *Proceedings of The National Academy Of Sciences Of The United States Of America*, 100(12):7319–7324.
- [Truccolo et al., 2005]Truccolo, W., Eden, U., Fellows, M., Donoghue, J., and Brown, E. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *Journal of Neurophysiology*, 93:1074–1089.
- [Tsien, 1989]Tsien, R. Y. (1989). Fluorescent probes of cell signaling. *Ann. Rev. Neurosci.*, 12:227–253.
- [Turaga et al., 2013]Turaga, S., Buesing, L., Packer, A., Dagleish, H., Pettit, N., Hausser, M., and Macke, J. (2013). Inferring neural population dynamics from multiple partial recordings of the same neural circuit. *NIPS*.
- [Vidne et al., 2012]Vidne, M., Ahmadian, Y., Shlens, J., Pillow, J., Kulkarni, J., Litke, A., Chichilnisky, E., Simoncelli, E., and Paninski, L. (2012). The impact of common noise on the activity of a large network of retinal ganglion cells. *Journal of Computational Neuroscience*, 33:97–121.
- [Vidne et al., 2009]Vidne, M., Kulkarni, J., Ahmadian, Y., Pillow, J., Shlens, J., Chichilnisky, E., Simoncelli, E., and Paninski, L. (2009). Inferring functional connectivity in an ensemble of retinal ganglion cells sharing a common input. *COSYNE*.
- [Wallace et al., 2008]Wallace, D., zum Alten Borgloh, S., Astori, S., Yang, Y., Bausen, M., Kugler, S., Palmer, A., Tsien, R., Sprengel, R., Kerr, J., Denk, W., and Hasan, M. (2008). Single-spike detection in vitro and in vivo with a genetic  $\text{Ca}^{2+}$  sensor. *Nature methods*, 5(9):797–804.

- [Yuste et al., 2006]Yuste, R., Konnerth, A., Masters, B., et al. (2006). *Imaging in Neuroscience and Development, A Laboratory Manual*.
- [Mishchenko and Paninski, 2011]Mishchenko, Y., Paninski, L. (2011). Efficient methods for sampling spike trains in networks of coupled neurons. *The Annals of Mathematical Statistics*, 5(3):1893-1919.
- [Cotton et al., 2013]Cotton, R. J., Froudarakis, E., Storer, P., Saggau, P., and Tolias, A. S. (2013). Three-dimensional mapping of microcircuit correlation structure. *Frontiers in Neural Circuits*, 7:151.
- [Furedi and Komlos, 1981]Furedi, Z. and Komlos, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica*, 1: 233.
- [Theis et al., 2015]Theis, L., Berens, P., Froudarakis, E., Reimer, J., Roman-Roson, M., Baden, T., Euler T., Tolias AS., and Bethge, M. (2015). Supervised learning sets benchmark for robust spike detection from calcium imaging signals. *bioRxiv*, 010777.
- [Yatsenko et al., 2015]Yatsenko, D., Josi, K., Ecker, A. S., Froudarakis, E., Cotton, R. J., and Tolias, A. S. (2015). Improved Estimation and Interpretation of Correlations in Neural Circuits. *PLoS Computational Biology*, 11:e1004083.
- [Vogelstein et al., 2010]Vogelstein J. T., Packer A. M., Machado T. A., Sippy T., Babadi B., Yuste R., Paninski L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104:3691.
- [Vogelstein et al., 2009]Vogelstein, J., Watson, B., Packer, A., Yuste, R., Jedynak, B., and Paninski, L. (2009). Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical Journal*, 97:636.
- [Soudry et al., 2015]Soudry, D., Keshri, S., Stinson, P., Oh M-H., Iyengar, G., Paninski, L. (2015). Efficient "Shotgun" inference of neural connectivity from highly sub-sampled activity data. *PLOS Computational Biology*, 11:e1004464.