Estimation of cerebral network structure

Jonathan Schiefer and Stefan Rotter

Bernstein Center Freiburg & Faculty of Biology, University of Freiburg

Our goal is to design and test new methods to reconstruct causal interactions between local populations of neurons based on population signals measured at multiple sites. This involves a number of challenges for which the usage of high performance computing is essential. These challenges occur in both the data analysis itself, as well as in large-scale network simulations to validate the data analysis.

1 Introduction

Inference of brain connectivity is of great importance both in basic research and for clinical applications. To this end, we have devised a new method for inferring effective connectivity from a complete set of covariances of neuronal activity measured at multiple sites. In other words, we are looking for causal relations reflected by measured brain signals. Our method exploits the covariance structure of the activity very efficiently. This makes our method robust with regard to temporal resolution and renders it applicable to various data types occurring in clinical practice, including electrocortigograms (ECoG) and functional magnetic resonance imaging (fMRI).

Our inference method uses statistical data analysis and machine learning. Since in the analysis of real data the ground truth is unknown, we employ biophysically informed large-scale numerical simulations to validate the results of data analysis. For both fields of application high performance computing is essential.

2 Data analysis

Inspired by [4], which focused on estimating effective connectivity among single neurons using spike train covariances, we extended the method to continuous population signals recorded from the brain.

As model we use a linear system with a vector of continuous signals $Y(t)$ given by

$$\begin{aligned} Y(t) &= X(t) + \int G(\tau) Y(t-\tau) d\tau \\ &= X(t) + (G * Y)(t) \end{aligned} \tag{1}$$

where $X(t)$ is the external input, and $G(\tau)$ is a matrix of kernels describing the pairwise interaction between nodes. Fourier transformation and rearranging equation (1) yields

$$\hat{Y}(\omega) = (\mathbb{1} - \hat{G}(\omega))^{-1} \hat{X}(\omega).$$

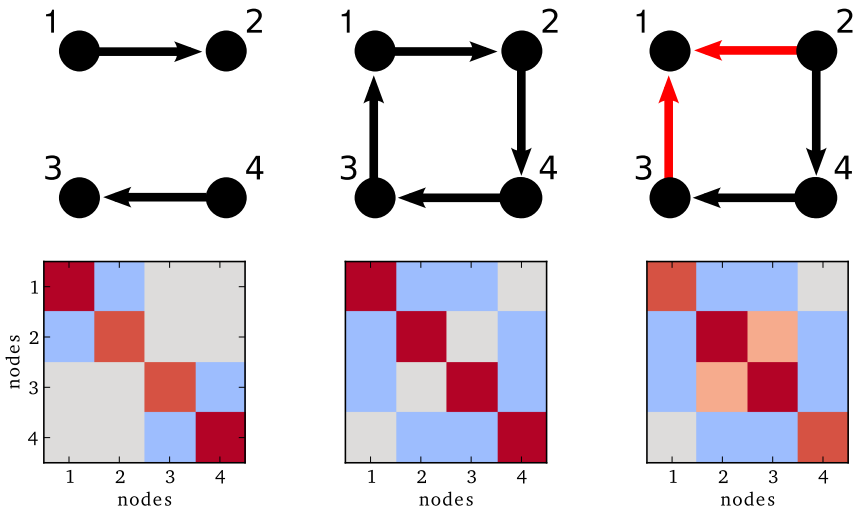

Figure 1: Small networks with 4 nodes each and different connectivity between them (top). They exhibit different activity dynamics, which is also reflected in the associated covariance matrix (bottom). In general, many different networks can lead to the same covariance matrix, and inferring connectivity from covariance is not possible; the inverse problem is ill-defined. However, assuming that the network is sparse (only a small fraction of links are actually present), the method of L^1-minimization (using gradient descent) helps to disambiguate the solution. The second order statistics of activity is particularly sensitive to collider structures (red arrows) in a network, and our method is able to exploit exactly these entries in the inverse covariance matrix ("precision matrix") to infer both the existence and the direction of connections [3] (image source: Schiefer et al, arXiv:1708.02423 [q-bio.NC], 2017).

The covariance matrix C for Y is then given by

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

where the matrix

$$\hat{D}(\omega) = \mathbb{E}[\hat{X}(\omega)\hat{X}^*(\omega)],$$

is diagonal, provided the components of external input are pairwise uncorrelated.

We can now reconstruct $\hat{G}(\omega)$ from equation (2)

$$\tilde{C}^{-1} = B^*B = B^*U^*UB \tag{3}$$

with $B = \sqrt{\hat{D}^{-1}}(\mathbb{1} - \hat{G})$ and a unitary matrix $U \in \mathrm{Mat}_n(\mathbb{C})$. We are looking for the sparsest matrix G with this property. This means that starting with an initial matrix B_0, we have to minimize the (modified) L^1-Norm of

$$\|UB_0\|_1 = \sum_{i \neq j} |(UB_0)_{ij}|$$

with respect to U. For the optimization procedure, we use a gradient descent algorithm described in [3]. The number of times this optimization has to be executed depends on the required frequency resolution. Additionally, the estimation of the connectivity can be improved using bootstrapping. This introduces the additional cost that the number of boots necessary to achieve significant results is quite high.

The optimization via gradient descent is a procedure where no parallelization is possible. In total, with a frequency resolution of 256 DFT points and 10 000 boots, the optimization has to be performed $129 \times 10\,000 = 1\,290\,000$ times. This task is split up by calculating the bootstrap covariance matrices and then performing the optimization for each boot separately, for all frequency bands. This results in single core jobs where each job takes approximately 20 minutes.

3 Simulation

For validation of the estimation of effective connectivity, we use biophysically informed numerical simulations of very large networks. These simulations are supposed to generate the same data types as the ones we use for the network inference. For simulation ECoG data we set up a simulation scheme using NEST (Neural simulation tool [2]).

As shown in [1], downscaling the size of networks has limitations regarding effective connectivity and covariances. This means that reducing the number of neurons in a network can either keep the effective connectivity of the network or the covariance structure of the resulting activity unchanged, but not both at the same time. In our project we are studying the relation of these measures, so we have a strong interest in avoiding any downscaling as far as possible. This means the networks we simulate in terms of size should be as biologically realistic as possible.

Since the commonly used ECoG electrode grids have 64 macro-contacts each, we also simulated a network of 64 mesoscopic sub-networks, consisting of 1 000 LIF-neurons each. Each of the 64 sub-networks represents the neural population recorded by one ECoG electrode. Within each sub-network the neurons are randomly connected with a connection probability of 10%. These subnetworks are connected by excitatory neurons, which form additional long-range connections to neurons in other populations. The extracellular fields that are induced by synaptic activity in every individual neuron are then accumulated over each population and recorded. These 64 signals are then used for estimating the meso-scale connectivity among the 64 populations. In our approach, the relevant aspects for the estimation of effective connectivity are the first and second order signal statistics. Therefore, our simulation should be consistent with measured data regarding these properties. To achieve this, we developed a new model of ongoing activity projecting to the recorded populations from other parts of the brain.

For the simulation we use the parallelization methods provided by NEST including MPI and multi-threading. These make the simulation of 64 000 neurons plus external input and measuring devices feasible.

4 Conclusion

As described above the availability of high performance computing (here, the HPC cluster NEMO with high speed interconnect) is essential for our project In both data analysis and simulation, HPC is a key factor to achieve results in a biologically realistic setting. In data analysis, NEMO enables us to apply bootstrap methods which would not be feasible without HPC. In the case of numerical simulations we are able to treat bigger networks, which reduce the unwanted scaling effects that are detrimental to our method. Even with up-to-date high-end HPC facilities it is currently not possible to simulate networks of the same sizes as they occur in human brains, but the new possibilities offered by NEMO represent an important step in the right direction.

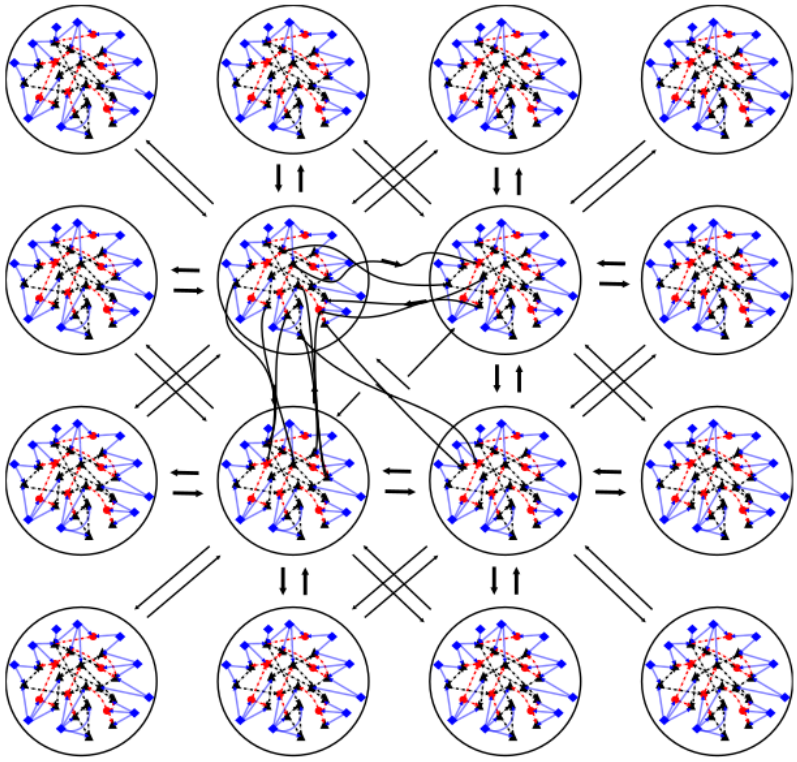

Figure 2: Scheme of the simulated network. Each circle represents the neural population underneath one ECoG electrode. Within each population neurons are randomly connected with 10% probability, but excitatory neurons form additional long-range connections to neurons in other populations according to a pre-defined connectivity scheme on a mesoscopic scale. The goal now is to reconstruct this mesoscopic long-range connectivity from the recorded accumulated extracellular fields (induced by synaptic activity) of all neurons in each population [5].

Acknowledgements

Funded by the German Research Foundation (DFG, grant EXC 1086). Supported by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG, grant INST 39/963-1 FUGG). We thank V. Pernice, A. Niederbühl and L.M. Schäfer for providing their software for data analysis and numerical simulations to us.

References

[1] Sacha Jennifer Van Albada, Moritz Helias, and Markus Diesmann. Scalability of Asynchronous Networks Is Limited by One-to-One Mapping between Effective Connectivity and Correlations. *PLoS Computational Biology*, 2015.

[2] Hannah Bos, Abigail Morrison, Alexander Peyser, Jan Hahne, Moritz Helias, Susanne Kunkel, Tammo Ippen, Jochen Martin Eppler, Maximilian Schmidt, Alex Seeholzer, Mikael Djurfeldt, Sandra Diaz, Janne Morén, Rajalekshmi Deepu, Teo Stocco, Moritz Deger, Frank Michler, and Hans Ekkehard Plesser. Nest 2.10.0, December 2015.

[3] Alexander Niederbühl. Inferring network structure from covarying neural activity. Diploma Thesis in Biology, 2014.

[4] Volker Pernice and Stefan Rotter. Reconstruction of sparse connectivity in neural networks from spike train covariances. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03008, 2013.

[5] Lisa Schäfer. Modelling neuronal population signals in networks of spiking networks. Master Thesis in Biology, 2015.