Network inference with latent variables

Danh-Tai Hoang[1], Junghyo Jo[2], and Vipul Periwal[1]

***Short Abstract* — We recently developed an entire data-driven approach for network inference based on free energy minimization. The method outperformed performance with published methods in predicting coupling strengths of stochastic processes, especially in the regimes of little observed samples and large variance of couplings. Using latent variables, we extend the method to recover the network when only a partial system is available. Our method can infer accurately the states of latent variables, the coupling strengths and the number of latent variables.**

***Keywords* — Network reconstruction, kinetic Ising model, latent variables, neural network.**

I. PURPOSE

PREDICTING network connections from observed data is a critical topic, not only in quantitative biology but also in other areas, more generally, data science [1]. Statistical methods have been developed based on naïve mean field [2], Thouless-Anderson-Palmer mean field [2], exact mean field [3], and maximum likelihood [4]. However, these methods work well only in the weak-coupling and large number of observed sample regimes. We recently developed an approach based on free energy minimization (FEM) and demonstrated that our method has a better performance than previous methods for strong-coupling regimes, especially in the limit of few observed samples [5].

Our aim is to extend the FEM method to infer network connections when available data does not contain every variable but some of them are unobserved.

II. METHOD

Our method combines FEM and maximum likelihood for latent variables and contains the following steps: (i) Assign the state of latent variables as random; (ii) Infer interaction between variables based on FEM [5]; (iii) Update the state of latent variables with a probability $P_2/(P_1 + P_2)$ where P_1 and P_2 represent the likelihood of systems before and after the updating; and (iv) Repeat steps (ii) and (iii) until the discrepancy of actual and its expectation value of variables becomes saturated. The number of latent variables can be estimated from the minima of the average of the discrepancy over the total number of variables.

III. RESULTS

We first tested our method to infer the coupling strengths in the kinetic Ising model in a system of 100 variables, using the states of only 60 variables. We could recover successfully the states of the 40 latent variables and the coupling strengths. We then applied our method to recover a neural network from neuronal activities in the salamander retina [6]. After inferring the neural dynamics couplings and the external local fields and the state of latent variables, we recovered the neuronal activities with an accuracy of 80% (Figure 1).

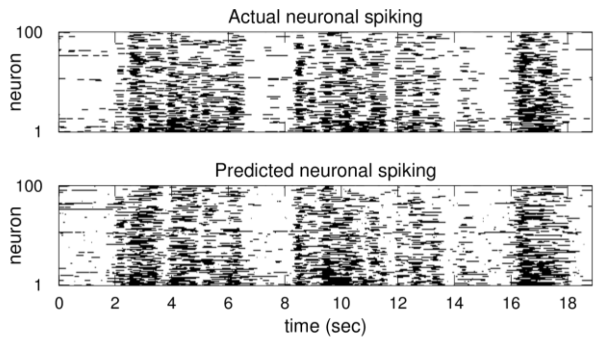

Figure 1. Inference of neuronal activity. Raster of 100 neuronal activities from experimental data (top) and our prediction (bottom).

IV. CONCLUSION

Extending our FEM method, we proposed an iterative algorithm with latent variables to infer the coupling strengths and the configuration of latent variables. Applying this to neuronal data, we can infer accurately activities. Besides better performance, our method is parameter-free and generalizes to many data types.

Acknowledgements: This work was supported by Intramural Research Program of the National Institutes of Health, NIDDK (D.-T.H., V.P.).

[1]Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. Email: danh-tai.hoang@nih.gov, vipulp@mail.nih.gov

[2]School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Korea. Email: jojunghyo@kias.re.kr

REFERENCES

[1] Schmidt M, Lipson H (2009) Distilling free-form natural laws from experimental data, Science, **324** (5923):81–5.

[2] Roudi Y and Hertz J (2011) Mean field theory for nonequilibrium network reconstruction, Phys. Rev. Lett. **106**, 048702.

[3] Mézard M and Sakellariou J (2011) Exact mean-field inference in asymmetric kinetic Ising systems, J. Stat. Mech., L07001.

[4] Zeng H-L, Alava M, Aurell E, Hertz J, and Roudi Y (2013) Maximum likelihood reconstruction for Ising models with asynchronous updates, Phys. Rev. Lett **110**, 210601.

[5] Hoang D-T, Song J, Jo J, and Periwal V (2018), "Stochastic causality inference with limited data from neuronal spiking to currency fluctuations", in preparation.

[6] Marre O, Tkacik G, Amodei D, Schneidman E, Bialek W, and Berry M-J (2017), Multi-electrode array recording from salamander retinal ganglion cells, IST Austria.