Theories of neural computation are enhanced by the modern inference engine.

Sean R. Bittner, Agostina Palmigiano, Alex T. Piet, Chunyu A. Duan, Francesca Mastrogiuseppe, Srdjan Ostojic, Carlos D. Brody, Kenneth D. Miller, and John P. Cunningham.

1 Abstract

To understand brain function, theoretical neuroscientists design mathematical models of neural activity, which capture the signature features of computation. Historically, the gold standard of theoretical neuroscience has been to analytically derive the relationship between model parameters and emergent properties of computation. However, this becomes infeasible as biologically realistic constraints are included in the neural system model. Therefore, research on such realistic models focuses on the examination of simulated activity. To attain a more formalized and interpretable understanding of realistic neuroscientific models, we propose the widespread adoption of the modern inference engine - technological advancements from probabilistic machine learning - to learn distributions of parameters that map to the emergent properties of computation. We demonstrate that the modern inference engine can usher in a new era of theoretical neuroscience by producing novel insights into network syncing in the stomatogastric ganglion, neuron-type input-responsivity in primary visual cortex, rapid task switching in superior colliculus, and approximate Bayesian inference in recurrent neural networks.

(150 word limit)

2 Introduction

Mathematical modeling has become a key part of modern neuroscience [1]. A theory of neural computation describes a neural system with a set of equations (i.e. a model) motivated by the laws of nature and neurophysiological observations. The key challenge for theorists is the description of how the model parameters govern the computational function of the neural system, which is characterized by mathematically defined features or "emergent properties" of computation. In idealized practice, theorists analytically derive how model parameters govern these emergent properties. Historical examples of this gold standard of theory include derivations of memory capacity in associative neural networks [2], chaos and autocorrelation timescales in random neural networks [3], and the paradoxical effect in E/I networks [4]. Unfortunately, as biological realism is introduced into neural circuit models, theory through analytic derivation becomes infeasible.

In fact, neural circuit models are designed in the context of misaligned incentives. Models are kept simplistic with unrealistic assumptions (e.g. symmetry, gaussianity) to facilitate analytic derivations. On the other hand, complexity is introduced for biological relevance. In models, simplicity sacrifices biological realism for derivations, and complexity sacrifices derivations for biological realism. When biological realism is the focus of a study, standard practice is to examine simulated activity from the model [5] (cite a bunch here). Visualization or regression is used to understand the model, however theorists strive for a more formalized understanding of these complex models.

Classically, statistical inference is a formalized way of describing the probabilistic relationship between observed data and model parameters. However, statistical inference is impracticable in neural system models, because the likelihood functions are generally intractable. Research in neural data analysis, which has enhanced our knowledge of a[kass], b[brown], c[paninski], d[jpcunni], e[pillow], focuses on the development of statistically inferable models for neural data sets, where such likelihood functions are tractable. Likelihood function intractability thus creates a gap between the models analyzed by theoreticians (motivated by laws of nature and physiology) and the probabilistic models developed by neural data scientists (constrained by tractability of inference) (Figure 1?). Theoretical neuroscientists are careful about model creation, where neural data analysts are practical. Neural data analysts are careful about inference of model parameters, where theoretical neuroscientists are practical. This motivates the question: can we start doing *careful* inference in *careful* models?

Advancements in probabilistic machine learning have led to transformative changes in industrial applications like image processing (sparse cite), speech recognition (sparse cite), text classification (sparse cite), and more. We call the generalizable components of this groundbreaking technology (deep learning, stochastic gradient descent, GPU parallelization, etc.) the “modern inference engine.” (Point to work from Cunningham/Paninski using modern inference engine for neuroscientific phenomenological models (PfLDS, BehaveNET)?) In this study, we use the modern inference engine to bypass the perceived intractability of statistical inference in realistic models of neural systems. (Introduce SDNs?) We demonstrate the widespread applicability of this approach by producing novel insights into network syncing in the stomatogastric ganglion (STG), neuron-type input-responsivity in primary visual cortex (V1), rapid task switching in superior colliculus (SC), and approximate Bayesian inference in recurrent neural networks (RNNs).

3 Results

3.1 Degenerate solution networks

- To translate progress in neural data analysis to theoretical neuro, need to key steps.
  - 1. Need to learn parameter distributions of biologically realistic (not just phenom.) models.
  - 2. Must be able to condition on emergent properties of interest, not simply computationally convenient sufficient statistics of data sets.
- Bayesian data scientists will say experimental data is all that matters.
- *Transition*: This is untrue when working in a creative, exploratory modeling setting.

Edgy contrarian point about theorists and data

- Common misconception: theoreticians rarely attempt to directly reproduce experimental data.
- Instead, they work with (abstracted?) mathematical definitions of emergent properties.

DSNs

- We introduce DSNs, which bridge methodology in these subfields of comp neuro.
- Combine ideas from MEFNs (cite Gabe) and LFVI (cite Dustin) to learn a deep probability distribution of theoretical model parameterizations z that produce the emergent properties of interest $T(x)$ (see Appendix).
- Explain deep probability distributions.
- DSNs are deep probability distributions of theoretical model parameters, which are optimized to be maximally random (maximum entropy) while producing the specified value of emergent

properties:

$$\begin{aligned} q_\theta^*(z) &= \operatorname*{argmax}_{q_\theta \in Q} H(q_\theta(z)) \\ \text{s.t. } & E_{z\sim q_\theta}\left[E_{x\sim p(x|z)}\left[T(x)\right]\right] = \mu \end{aligned} \tag{1}$$

Worked example: STG

- For example, consider the STG.
- Explain this STG circuit, emergent property of interest.
- For our choices of STG as model and network syncing as emergent property, we use a DSN to learn a distribution on STG conductance parameters that produces network syncing.
- Emphasize utility of DSN using Hessian.
- An equivalent conceptualization is that DSNs do Bayesian inference (see Appendix).
- Punchline about DSNs and transition to V1.

3.2 Exploratory analysis of a theoretical model

Will focus on this once result finalized. Have a lot of text to pull from.

3.3 Identifying sufficient mechanisms of task learning

Will focus on this once V1 and LRRNN finalized. Have a lot of text to pull from.

3.4 Conditioning on computation with interpretable models of RNNs

Will focus on this once result finalized. Have a lot of text to pull from.

4 Discussion

- Summarize the key methodlogical demonstrations from the results section.
- Talk big picture: If we know we can't analytically derive these things, we need an alternative characterization. Simulate and examine isn't cutting it. We need to be leveraging the modern

inference engine to gain this understanding. Bayesian probability is the framework we should use for this formalism.

- Expand on idea of posterior predictive checks / hypothesis testing / exploratory analyses of models themselves. Give the whole, we don't even understand the models we're developing pitch.
- Elaborate on idea of conditioning on flexibly defined statistics i.e. emergent properties. Emphasize how this is practical. Link to sufficient statistics, esp. commonly used in phenom models like spike counts etc.
- Summarize the respective strengths SNPE and DSN.
- Link conditioning on task execution with work done today with RNNs. Basically, we're training overparameterized models with regression, and get a distribution (we have no prob treatment of). Emphasize utility of low-dim interpretable parameterizations.
- A paragraph on bridging large scale recordings with theory.

References

[1] Larry F Abbott. Theoretical neuroscience rising. *Neuron*, 60(3):489–495, 2008.

[2] John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10):3088–3092, 1984.

[3] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural networks. *Physical review letters*, 61(3):259, 1988.

[4] Misha V Tsodyks, William E Skaggs, Terrence J Sejnowski, and Bruce L McNaughton. Paradoxical effects of external modulation of inhibitory interneurons. *Journal of neuroscience*, 17(11):4382–4388, 1997.

[5] Gabrielle J Gutierrez, Timothy OLeary, and Eve Marder. Multiple mechanisms switch an electrically coupled, synaptically inhibited neuron between competing rhythmic oscillators. *Neuron*, 77(5):845–858, 2013.

[6] Brendan K Murphy and Kenneth D Miller. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron*, 61(4):635–648, 2009.

[7] Hirofumi Ozeki, Ian M Finn, Evan S Schaffer, Kenneth D Miller, and David Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4):578–592, 2009.

[8] Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417, 2015.

[9] Henry Markram, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and Caizhi Wu. Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience*, 5(10):793, 2004.

[10] Bernardo Rudy, Gordon Fishell, SooHyun Lee, and Jens Hjerling-Leffler. Three groups of interneurons account for nearly 100% of neocortical gabaergic neurons. *Developmental neurobiology*, 71(1):45–61, 2011.

[11] Ashok Litwin-Kumar, Robert Rosenbaum, and Brent Doiron. Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *Journal of neurophysiology*, 115(3):1399–1409, 2016.

[12] Carsten K Pfeffer, Mingshan Xue, Miao He, Z Josh Huang, and Massimo Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience*, 16(8):1068, 2013.

[13] Mario Dipoppa, Adam Ranson, Michael Krumin, Marius Pachitariu, Matteo Carandini, and Kenneth D Harris. Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *Neuron*, 98(3):602–615, 2018.

[14] Chunyu A Duan, Marino Pagan, Alex T Piet, Charles D Kopec, Athena Akrami, Alexander J Riordan, Jeffrey C Erlich, and Carlos D Brody. Collicular circuits for flexible sensorimotor routing. *bioRxiv*, page 245613, 2018.