**\section{Abstract}**

Consistent improvements in neural recording technology have made the role of theory in neuroscience as crucial as ever. Theoretical models must explain more and more properties of neuronal activity as they are rapidly discovered. To account for additional scientific findings, models often increase in complexity making theoretical work more challenging. Specifically, a critical step in theoretical neuroscience is determining how parameterizations of neural circuit models govern the emergent properties of interest (those motivated by scientific findings.) The gold standard is to derive these relationships analytically using mathematical expertise. However, models of sufficient complexity to describe neural processing may not be amenable to such analytic approaches. We propose a machine learning approach for characterizing parameter-behavior relationships as an alternative to such analytic or brute-force simulation approaches. Degenerate solution networks (DSNs) identify the full distribution of model parameters that result in an emergent property of interest. We demonstrate how DSNs facilitate exploratory analyses and scientific hypothesis testing beyond modern practice in models of the stomatogastric ganglion (STG), primary visual cortex (V1), superior colliculus (SC), and in recurrent neural networks (RNNs).

**\section{Introduction}**

Theoretical neuroscientists design and test models of neural computation, assessing a model’s quality by its recapitulation of experimentally recorded neural activity. Developing a theory for a particular neural computation involves a.) mathematically defining the emergent properties of the neural activity that signify the computation, b.) designing a parameterized model of the brain area(s) executing the computation, and then c.) characterizing the model parameters that produce these emergent properties. In idealized practice of theory, scientists perform step c.) through analytic derivation. These derivations often rely on tools developed in fields such as physics [cite], computer science[cite] and engineering [cite]. It is often necessary for theoretical neuroscientists to extend these formalisms [cite], or build an entirely new class of methodology [cite].

Such gold standard theoretical practice is not always possible for addressing neuroscientific questions. This is because the relation of the model parameterization to its behavior is often analytically intractable. In such cases, theorists search for structure in simulated model activity with respect to the parameters. This approach becomes computationally intractable as the number of parameters increases, and glaringly lacks a probabilistic treatment of the model.

The primary focus of research in statistical neuroscience has been to develop inference methods for generative models of neural data that scale to the growing size of such data sets. However, these generative models, for which we have designed inference procedures, are rarely the models employed in theoretical neuroscience. Historically, this model disconnect can be due to the intractability of inference for such theoretical models. For example, the nonlinear dynamical nature of many neural circuit models [cite] makes calculation of the likelihood function (let alone its gradient) intractable. Encouragingly, over the past couple of years, we have seen the introduction of likelihood-free variational inference approaches (cite Dustin, Alireza’s spiking network inference) and the use of generative adversarial networks (cite Yashar) for learning model parameterizations that produce observed data points in this intractable likelihood function setting. Even so, there is still a considerable amount of work to be done tailoring these effective machine learning methods to the needs of theorists.

Importantly, we point out that this disconnect between the generative models used in statistical and theoretical neuroscience research is not simply tractability vs intractability of inference, but arises primarily from the data-agnostic nature of much work in theoretical neuroscience. Certainly, theorists are cognizant of the neural data that goes into scientific findings, and that motivates the definition of their research questions. These findings allow the definition of emergent properties of neural activity, which are a satisfactory substitute for data points collected in an experiment when working with such complex, abstract theoretical models. Statistical machine learning is largely focused on algorithms for inference on a generative model given a collection of its observations, not an emergent property: a more abstract specification of the statistics of its observations. Here, we introduce a novel approach to learning in likelihood-free models, degenerate solution networks (DSNs), which learn the full (maximum entropy) distribution of model parameters that result in an emergent property of interest.

(add more punchline stuff to this para)

In this study, we demonstrate how DSNs facilitate exploratory analyses, scientific hypothesis testing, and experimental design beyond modern practices in models of the stomatogastric ganglion (STG), primary visual cortex (V1), superior colliculus (SC), and recurrent neural networks (RNNs).

**\section{Results}**

**\subsection{Inverting neural models given their emergent properties.}**

More text about why this is useful and important.

The big figure at the beginning \\

\begin{figure}

\begin{center}

**\includegraphics**[scale=0.2]{figs/fig1/fig1.pdf}

\caption{blah blah blah}

\end{center}

\end{figure}

**\subsection{Stomatogastric ganglion}**

\begin{figure}

\begin{center}

**\includegraphics**[scale=0.4]{figs/fig2/fig2.pdf}

\caption{STG figure. Also what a DSN does...}

\end{center}

\end{figure}

**\subsection{Primary visual cortex}**

\begin{figure}

\begin{center}

**\includegraphics**[scale=0.35]{figs/fig3/fig3.pdf}

\caption{V1 stuff}

\end{center}

\end{figure}

**\subsection{Superior colliculus}**

\begin{figure}

\begin{center}

**\includegraphics**[scale=0.4]{figs/fig4/fig4.pdf}

\caption{SC stuff}

\end{center}

\end{figure}

**\subsection{Computation in low-rank RNNs}**

\begin{figure}

\begin{center}

**\includegraphics**[scale=0.4]{figs/fig5/fig5.pdf}

\caption{Low-rank RNN stuff}

\end{center}

\end{figure}

**\section{Conclusion}**