Scalable Inference of Neural Dynamical Systems

Alyson K. Fletcher, U.C. Santa Cruz.

Abstract

Fluorescent calcium imaging provides a potentially powerful tool for inferring connectivity in large neural circuits. However, a key challenge in using calcium imaging for connectivity detection is that current systems often have a temporal response and frame rates that can be orders of magnitude slower than the underlying neural spiking process. Bayesian inference methods based on expectation-maximization (EM) have been proposed to overcome these limitations, but are often computationally demanding since the E-step in the EM procedure typically involves state estimation for a high-dimensional nonlinear dynamical system. In this work, we propose a computationally scalable method based on a hybrid of loopy belief propagation and approximate message passing (AMP). The key insight is that a neural system as viewed through calcium imaging can be factorized into simple scalar dynamical systems for each neuron with linear interconnections between the neurons. Using the structure, the updates in the proposed hybrid AMP methodology can be computed by a set of one-dimensional state estimation procedures and linear transforms with the connectivity matrix. The method extends earlier works by incorporating more general nonlinear dynamics and responses to stimuli.

1. Introduction

Determining connectivity in populations of neurons is fundamental to understanding neural computation and function. In recent years, calcium imaging has emerged as a promising technique for measuring synaptic activity and mapping neural microcircuits [1–4]. Fluorescent calcium-sensitive dyes and genetically-encoded calcium indicators can be loaded into neurons, which can then be imaged for spiking activity either in vivo or in vitro. Current methods enable imaging populations of hundreds to thousands of neurons with very high spatial resolution. Using two-photon microscopy, imaging can also be localized to specific depths and cortical layers [5]. Calcium imaging also has the potential to be combined with optogenetic stimulation techniques such as in [6].

However, inferring neural connectivity from calcium imaging remains a mathematically and computationally challenging problem. Unlike anatomical methods, calcium imaging does not directly measure connections. Instead, connections must be inferred indirectly from statistical relationships between spike activities of different neurons. In addition, the measurements of the spikes from calcium imaging are indirect and noisy. Most importantly, the imaging introduces significant temporal blurring of the spike times: the typical time constants for the decay of the fluorescent calcium concentration, \([\text{Ca}^{2+}]\), can be on the order of a second – orders of magnitude slower than the spike rates and inter-neuron dynamics. Moreover, the calcium imaging frame rate remains relatively slow – often less than 100 Hz. Hence, determining connectivity typically requires super-resolution of spike times within the frame period.

To overcome these challenges, the recent work [7] proposed a Bayesian inference method to estimate functional connectivity from calcium imaging in a systematic manner. Unlike “model-free” approaches such as in [8], the method in [7] assumed a detailed functional model of the neural dynamics with unknown parameters including a connectivity weight matrix \(W\). The model parameters including the connectivity matrix can then be estimated via a standard EM procedure [9]. While the method is general, one of the challenges in implementing the algorithm is the computational complexity. Specifically, the E-Step in the EM procedure requires a state estimation of a high-dimensional nonlinear dynamical system. Approximations such blockwise Gibbs sampling may be slow.

In [10], an approximate method for the E-step was proposed using a combination of loopy belief propagation and approximate message passing. The key insight is that a network of \(N\) neurons decomposes into \(N\) low-dimensional nonlinear dynamical systems with linear interconnections. The Calcium imaging fluorescence can also be modeled as \(N\) additional low-dimensional dynamical systems. This decomposition of a high-dimensional system into low-dimensional nonlinear dynamical systems with linear interconnections becomes
amenable to graphical model methods.

In this short note, we review this method and suggest that the technique can be extended to more general classes of nonlinear processes. Responses to stimuli can also be modeled.

2. Model for the Neuron and Calcium Dynamics

We consider a recurrent network of $N$ neurons indexed by $i = 1, \ldots, N$. All dynamics are approximated in discrete time with some time step $\Delta$ with time indexed by $k = 0, \ldots, T - 1$. We let $s^k_i$ be the number of spikes (action potentials) in time bin $k$ from neuron $i$. We suppose that the system is exposed to a stimuli represented by a vector $u^k = (u^k_1, \ldots, u^k_M)$. The components of the vector could represent stimuli including audio or visual data. The goal is to develop a functional model relating the stimuli and spike process that elucidates the connectivity between the neurons.

To this end, we assume that the spiking process of each neuron is described by some nonlinear difference equations of the form,

$$
\begin{align*}
\tau_v v^{k+1}_i &= f_v(v^k_i, q^k_i, a^k_{vi}, \lambda_{vi}) \\
\lambda^k_i &= g_v(v^k_i, q^k_i, a^k_{vi}, \lambda_{vi}),
\end{align*}
$$

where $f_v(\cdot)$ and $g_v(\cdot)$ are known functions, $v^k_i$ is an internal state of the neuron, $a^k_{vi}$ is noise process and $\lambda_{vi}$ are unknown parameters of the system. The terms $q^k_i$ incorporate the effects of the stimuli and spikes from other neurons, which we assume has a linear relation of the form

$$
q^k_i = \sum_{l=0}^{L-1} \sum_{j=1}^N W_{ij} s^k_j - \ell + \sum_{m=1}^M F_{im} u^k_m,
$$

where $W_{ij}$ represents the influence of a spike in neuron $j$ to neuron $i$ at time delay $\ell$ and $F_{im}$ represents the influence of stimuli $u^k_m$ on neuron $i$. The matrices $W$ and $F$ thus describe the connectivity and direct stimulus response of each neuron. The parameters $\lambda_{vi}$ can include, for example, the threshold levels and integration time constants.

The model (1) and (2) is extremely general. Together they can incorporate, for example, both the linear nonlinear Poisson (LNP) integration as studied in [7] as well as the integrate and fire model in [10]. Our sole critical assumption is that the state dimension $v^k_i$ is small – usually one or two dimensional. Thus, the overall system is described as a set of low-dimensional nonlinear dynamical systems (1) with linear interconnections (2).

For the observations, we assume a similar decom-

position of the form

$$
\begin{align*}
c^{k+1}_i &= f_c(c^k_i, s^k_i, d^k_{ci}, \lambda_{ci}) \\
y^k_i &= g_c(v^k_i, q^k_i, d^k_{ci}, \lambda_{ci}),
\end{align*}
$$

where $c^k_i$ is an internal state in the observation process and $y^k_i$ is the observation. For calcium imaging, $y^k_i$ would be the measured calcium fluorescence and $c^k_i$ the fluorescent calcium concentration which generally decays as a low-pass filter of the spikes. Details of the models can be found in [11]. Again, we have a nonlinear low-dimensional system for each neuron $i$.

3. Parameter Estimation via Message Passing

3.1. EM Estimation

Let $\theta$ be set of all the unknown parameters,

$$
\theta = \{W, F, \lambda_{ci}, \lambda_{vi}, i = 1, \ldots, N\},
$$

which includes the connectivity matrix $W$, stimuli response matrix $F$ and the parameters in neuron dynamics (1) and calcium dynamics (3). Our goal is to estimate $\theta$ from the observations $y$. For calcium imaging, estimating $\theta$ from $y$ enables estimating the connectivity and stimulus response from the calcium imaging.

We consider a regularized maximum likelihood (ML) estimate

$$
\hat{\theta} = \arg \max_{\theta} L(y|\theta) + \phi(\theta),
$$

where $L(y|\theta)$ is the negative log likelihood of $y$ given the parameters $\theta$ and $\phi(\theta)$ is some regularization function. As we will see below, the regularization function $\phi(\theta)$ can be used to impose constraints or priors on the parameters.

Exact computation of $\hat{\theta}$ in (5) is generally intractable, since the observations $y$ depend on the unknown parameters $\theta$ through a large set of hidden variables and nonlinear difference equations. Similar to [7], we thus use a standard EM procedure [9]. To apply the EM procedure to the calcium imaging problem, let $x$ be the set of hidden variables,

$$
x = \{v, z, q, s\},
$$

where $v$ are the internal states of the neurons in (1) $z$ are set of the observation states in (3), $s$ the spike outputs and $q$ the linearly combined spike inputs (2).

The EM procedure alternately estimates distributions on the hidden variables $x$ given the current parameter estimate for $\theta$ (the E-step); and then updates the
estimates for parameter vector $\theta$ given the current distribution on the hidden variables $x$ (the M-step).

- **E-Step**: Given parameter estimates $\hat{\theta}^\ell$, estimate
  \[
P(x|y, \hat{\theta}^\ell),
  \tag{7}
\]
  which is the posterior distribution of the hidden variables $x$ given the observations $y$ and current parameter estimate $\hat{\theta}^\ell$.

- **M-Step** Update the parameter estimate via the minimization,
  \[
  \hat{\theta}^{\ell+1} = \arg \min_{\theta} \mathbb{E} [L(x, y|\theta)\hat{\theta}^\ell] + \phi(\theta),
  \tag{8}
\]
  where $L(x, y|\theta)$ is the joint negative log likelihood,
  \[
  L(x, y|\theta) = -\log p(x, y|\theta).
  \tag{9}
\]
  In (8) the expectation is with respect to the distribution found in (7) and $\phi(\theta)$ is the parameter regularization function.

The next two sections will describe how we approximately perform each of these steps.

### 3.2. E-Step estimation via Approximate Message Passing

The computationally challenging step is the E-step. Estimating the posterior (7) is, in essence, a nonlinear state estimation problem whose complexity grows exponentially in the state dimension. Following [10], we use an approximate messaging passing method that exploits the separable structure of the system: For the remainder of this section, we will assume the parameters $\theta$ in (4) are fixed to the current parameter estimate $\hat{\theta}^\ell$. Then, under the assumptions of Section 2, the joint probability distribution function of the variables can be written in a factorized form,

\[
P(x, y) = P(q, v, s, z, y) = \frac{1}{Z} \prod_{k=0}^{T-1} \psi_{\text{neuron}}^k(q_k, v_k, s_k) \psi_{\text{obs}}^k(s_k, z_k, y_i),
\tag{10}
\]

where $Z$ is a normalization constant; $\psi_{\text{neuron}}^k(q_k, v_k, s_k)$ is the potential function for the neuron dynamics model (1) relating the summed spike inputs $q_k$ to the membrane voltages $v_k$ and spike outputs $s_k$; $\psi_{\text{obs}}^k(s_k, z_k, y_i)$ is a potential function for the observation process (3) relating the spike outputs $s_k$ to the observation state $z_k$ and observed values $y_i$; and the term $\psi_{\text{obs}}^k(q_k := W:s_k + Fu_k)$ indicates that the distribution is to be restricted to the set satisfying the linear constraints (2).

The factorization (10) thus divides the density into three terms: $N$ terms for the neuron dynamics; $N$ terms for the observation dynamics and $T$ terms for linear constraints between the neurons. That is, the system is divided into $2N$ low-dimensional nonlinear systems coupled only by linear constraints. This decomposition enables efficient approximate inference of the density (7) via belief propagation in a manner using a graphical models based method from [12]. That method combines standard loopy belief propagation [13] and approximate message passing methods [14]. Using the decomposition structure, the method reduces the high-dimensional nonlinear state estimation problem to a sequence of low-dimensional estimation problems on the individual dynamical systems (1) and (3) as well as scalar AWGN estimation problems on the outputs and inputs of the linear constraints. The method [12] and related methods [15] have also been used in other neural estimation problems [16].

### 3.3. Approximate M-step Optimization

The M-step (8) is computationally relatively simple. The key parameters $W$ and $F$ in (2) have a linear relationship between the components of the variables in the vector $x$ in (6). Thus, using a Gaussian approximation on the hidden states $x$, the log likelihood term in (8) becomes quadratic in $W$ and $F$. Hence, using any convex regularization function $\phi(W)$ we can then easily minimize (8).

The regularization function can incorporate many biologically relevant constraints. For example, using an $\ell_1$-type constraint, as is standard in the LASSO problem [17], can impose sparsity in the connectivity. The response matrix $W$ over time may also have a low-rank structure due to the similar shape of the time responses from different neurons.

### Conclusions

We have proposed a scalable method for inferring connectivity in neural systems from calcium imaging. The method is based on factorizing the systems into scalar dynamical systems with linear connections. Once in this form, state estimation – the key computationally challenging component of the EM estimation – is tractable via approximating message passing methods. The method can incorporate complex nonlinear dynamics in the neurons and calcium imaging process as well as structural constraints in the connectivity and response matrices.
Acknowledgements

This work was supported under ONR grant N00014-15-1-2677 and NSF award 1254204. The author would like to thank Lav Varshney, Bruno Olshausen, Mitya Chklovskii for their valuable comments.

References