Discrete and continuous latent states of neural activity in *Caenorhabditis Elegans*

Scott W. Linderman and Liam Paninski
Columbia University

July 27, 2016

Abstract

Recent advances in neural recording technologies have enabled simultaneous measurements of the majority of head ganglia neurons in immobilized C. elegans [1]. Moreover, since some neurons are known to reliably indicate the onset or offset of particular behaviors, like ventral and dorsal turns, behavioral state can be decoded from the simultaneous population recordings. These datasets provide unique visibility into the relationship between neural activity and behavior. While it seems clear that activity is inherently lower dimensional than the number of neurons due to strong correlations between cells, the nature of the latent brain state remains unclear. For example, is brain state better thought of as discrete or continuous, or perhaps a combination of the two? Does it obey linear or nonlinear dynamics? We propose a generative approach to probing these questions. We model the neural activity as a *switching linear dynamical system* (SLDS), with both discrete and continuous latent states, and conditionally linear dynamics. We then analyze the posterior distribution over states implied by the neural recordings and find that the discrete states correspond to stereotypical motor sequences. In contrast to previous work, these states are exposed in an entirely unsupervised manner.

1 Model

Assume the instantaneous neural activity at time t for a population of N neurons is represented as a vector, $\boldsymbol{y}_t \in \mathbb{R}^N$. In calcium imaging settings, the entries in this vector may be instaneous $\Delta F/F$ measurements, or another signal that captures neural activity. In this experiment, we use the smoothed time derivative of $\Delta F/F$. Over the course of an experiment, we measure a sequence of vectors, which we combine into a matrix denoted by $\boldsymbol{y}_{1:T}$.

Our model is based on the following assumptions: (i) the instantaneous neural activity, $\boldsymbol{y}_t$, reflects an underlying, low-dimensional latent state; (ii) this state has a discrete component, $z_t \in \{1, \ldots, K\}$, and a continuous component, $\boldsymbol{x}_t \in \mathbb{R}^D$; (iii) the continuous latent state has linear dynamics governed by the corresponding discrete latent state; and (iv) the observed neural activity is a linear function of the underlying states with additive Gaussian noise.

Discuss the motivation for these modeling assumptions.

These assumptions are combined in a switching linear dynamical system, which we

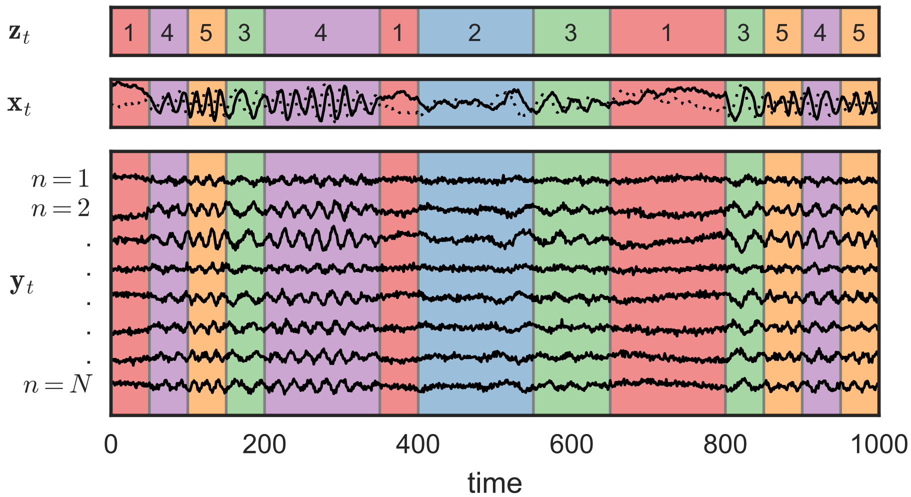

Figure 1: Simulated data from a switching linear dynamical system. The population stochastically switches between discrete states, $z_t \in \{1, \ldots, K\}$ (here, $K = 5$), each of which is color coded for visualization. These discrete states govern the linear dynamics of the continuous latent state, $\boldsymbol{x}_t \in \mathbb{R}^D$ (here, $D = 2$). For example, state $z_t = 1$ corresponds to a simple random walk, whereas state $z_t = 5$ has fast, oscillatory dynamics. Finally, the observed signals, $\boldsymbol{y}_t \in \mathbb{R}^N$ (here, $N = 8$), are obtained via a linear transformation of the underlying, continuous state, $\boldsymbol{x}_t$. The correlations and dynamics in the observations are inherited from the dynamics of the latent states.

formalize with the following generative model:

$$p(\boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T} \,|\, \boldsymbol{\Theta}) = p(\boldsymbol{\Theta}) \prod_{t=1}^{T} p(z_t \,|\, z_{t-1}, \boldsymbol{\Theta})\, p(\boldsymbol{x}_t \,|\, z_{t-1}, \boldsymbol{x}_{t-1}, \boldsymbol{\Theta})\, p(\boldsymbol{y}_t \,|\, z_t, \boldsymbol{x}_t, \boldsymbol{\Theta}). \quad (1)$$

Our beliefs about the dynamics of these latent states are encoded in the form of the conditional distributions for z_t and $\boldsymbol{x}_t$. First, we assume the discrete states follow a Markov process,

$$p(z_t \,|\, z_{t-1}, \boldsymbol{\Theta}) \sim \mathrm{Discrete}(\boldsymbol{\pi}^{(z_{t-1})}). \quad (2)$$

Next, we imbue the continuous latent states with linear Gaussian dynamics,

$$p(\boldsymbol{x}_t \,|\, \boldsymbol{x}_{t-1}, z_{t-1}, \boldsymbol{\Theta}) \sim \mathcal{N}(\boldsymbol{A}^{(z_{t-1})}\boldsymbol{x}_{t-1} + \boldsymbol{b}^{(z_{t-1})}, \boldsymbol{Q}^{(z_{t-1})}). \quad (3)$$

Finally, we impose the assumption of linear observations via the conditional distribution,

$$p(\boldsymbol{y}_t \,|\, \boldsymbol{x}_t, z_t, \boldsymbol{\Theta}) \sim \mathcal{N}(\boldsymbol{C}^{(z_t)}\boldsymbol{x}_t + \boldsymbol{d}^{(z_t)}, \boldsymbol{R}^{(z_t)}). \quad (4)$$

Thus, the parameters of the model are,

$$\boldsymbol{\Theta} = \left\{ \boldsymbol{A}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)}, \boldsymbol{C}^{(k)}, \boldsymbol{d}^{(k)}, \boldsymbol{R}^{(k)}, \boldsymbol{\pi}^{(k)} \right\}_{k=1}^{K}. \quad (5)$$

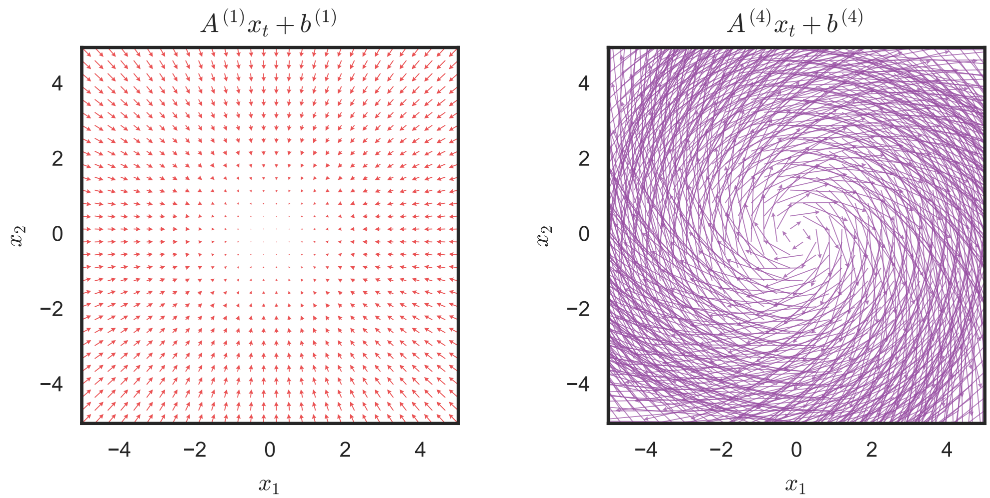

Figure 2: The dynamics corresponding to discrete latent state k may be visualized as a vector field where the arrows point to the expected next state, $\mathbb{E}[\boldsymbol{x}_{t+1}] = \boldsymbol{A}^{(k)}\boldsymbol{x}_t + \boldsymbol{b}^{(k)}$. Here, we show the dynamics for the first and fourth discrete states in the synthetic example from Figure 1. The first discrete state is a random walk with a slight decay toward the origin; the fourth discrete state corresponds to fast, oscillatory dynamics.

Discuss interpretation of the parameters. Set the stage for visualizing $\boldsymbol{A}^{(k)}$ in later sections.

2 Results

References

[1] Saul Kato, Harris S Kaplan, Tina Schrödel, Susanne Skora, Theodore H Lindsay, Eviatar Yemini, Shawn Lockery, and Manuel Zimmer. Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. *Cell*, 163(3):656–669, 2015.

A Bayesian Inference for Switching Linear Dynamical Systems

Our goal is to estimate the posterior probability of a sequence of latent states and a set of parameters given the observed data. From Bayes' rule, we have,

$$p(\boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{\Theta} \,|\, \boldsymbol{y}_{1:T}) = \frac{p(\boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{\Theta})}{p(\boldsymbol{y}_{1:T})}. \tag{6}$$

The numerator is the joint probability given by Eq. (1), and the denominator, $p(\boldsymbol{y}_{1:T})$, which is also known as the *marginal likelihood*, is given by an integral over possible latent states and parameters,

$$p(\boldsymbol{y}_{1:T}) = \int p(\boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{\Theta}) \, \mathrm{d}\boldsymbol{x}_{1:T} \, \mathrm{d}\boldsymbol{z}_{1:T} \, \mathrm{d}\boldsymbol{\Theta}. \tag{7}$$

Unfortunately, this integral is not efficiently computable for complex models like the SLDS, forcing us to seek approximate inference methods instead. Markov chain Monte Carlo (MCMC) methods offer one such approach. cite

To construct our MCMC algorithm, we iteratively sample one set of latent states or parameters from its conditional distribution, holding the rest fixed, in a technique known as Gibbs sampling . There are five main sets of parameters to sample, detailed below. cite

1. *Gibbs sampling the discrete latent states, $\boldsymbol{z}_{1:T}$:*

   Given the continuous latent states, $\boldsymbol{x}_{1:T}$, and the parameters, $\boldsymbol{\Theta}$, the conditional distribution over discrete latent states is the same as in a standard hidden Markov model. A joint sample from $p(\boldsymbol{z}_{1:T} \,|\, \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta})$ can be generated using the forward filtering backward sampling (FFBS) algorithm.

2. *Gibbs sampling the continuous latent states, $\boldsymbol{x}_{1:T}$:*

   Given the discrete latent states, $\boldsymbol{z}_{1:T}$, the observations, $\boldsymbol{y}_{1:T}$, and the parameters, $\boldsymbol{\Theta}$, the conditional distribution of the continuous latent states is linear and Gaussian. As with the discrete latent states, a joint sample of $p(\boldsymbol{x}_{1:T} \,|\, \boldsymbol{z}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta})$ can be generated using an FFBS algorithm.

3. *Gibbs sampling the dynamics parameters, $\{\boldsymbol{A}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)}\}_{k=1}^{K}$:*

   For fixed latent state sequences, the dynamics model reduces to a simple multivariate regression problem. We have,

$$\begin{aligned} &p(\boldsymbol{A}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)} \,|\, \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta}) \\ &\qquad \propto p(\boldsymbol{A}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)}) \prod_{t=1}^{T} \left[\mathcal{N}(\boldsymbol{x}_t \,|\, \boldsymbol{A}^{(k)} \boldsymbol{x}_{t-1} + \boldsymbol{b}^{(k)}, \, \boldsymbol{Q}^{(k)} \right]^{\mathbb{I}[z_{t-1}=k]}. \end{aligned} \tag{8}$$

   If the prior distribution is the form of a matrix normal inverse Wishart (MNIW) prior, then this conditional distribution will be as well.

4. *Gibbs sampling the observation parameters, $\{\boldsymbol{C}^{(k)}, \boldsymbol{d}^{(k)}, \boldsymbol{R}^{(k)}\}_{k=1}^{K}$:*

   As with the dynamics parameters, for fixed latent state sequences, the observation model is also a multivariate regression problem. We have,

$$\begin{aligned} &p(\boldsymbol{C}^{(k)}, \boldsymbol{d}^{(k)}, \boldsymbol{R}^{(k)} \,|\, \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta}) \\ &\qquad \propto p(\boldsymbol{C}^{(k)}, \boldsymbol{d}^{(k)}, \boldsymbol{R}^{(k)}) \prod_{t=1}^{T} \left[\mathcal{N}(\boldsymbol{y}_t \,|\, \boldsymbol{C}^{(k)} \boldsymbol{x}_{t} + \boldsymbol{d}^{(k)}, \, \boldsymbol{R}^{(k)} \right]^{\mathbb{I}[z_{t}=k]}. \end{aligned} \tag{9}$$

This, too, is conjugate when the prior distribution assumes the form of a matrix normal inverse Wishart (MNIW) distribution.

5. *Gibbs sampling the Markov parameters,* $\{\boldsymbol{\pi}^{(k)}\}_{k=1}^{K}$*:*

   Finally, we must sample the Markov transition matrix. We separate this into its K rows, each of which specifies a probability distribution, $p(z_t \mid z_{t-1} = k) = \boldsymbol{\pi}^{(k)}$. For a fixed discrete latent state sequence, the conditional distribution of $\boldsymbol{\pi}^{(k)}$ is,

$$p(\boldsymbol{\pi}^{(k)} \mid \boldsymbol{z}_{1:T}) \propto p(\boldsymbol{\pi}^{(k)}) \prod_{t=1}^{T} \left[\pi_{z_t}^{(k)}\right]^{\mathbb{I}[z_{t-1}=k]}. \tag{10}$$

   If the prior distribution is $p(\boldsymbol{\pi}^{(k)}) = \mathrm{Dir}(\boldsymbol{\pi}^{(k)} \mid \alpha)$, then this conditional distribution is a Dirichlet as well,

$$p(\boldsymbol{\pi}^{(k)} \mid \boldsymbol{z}_{1:T}) = \mathrm{Dir}(\boldsymbol{\pi}^{(k)} \mid \widetilde{\boldsymbol{\alpha}}^{(k)}) \tag{11}$$

$$\widetilde{\alpha}_{k'}^{(k)} = \alpha + \sum_{t=1}^{T} \mathbb{I}[z_t = k']\, \mathbb{I}[z_{t-1} = k]. \tag{12}$$

Each one of these five steps leaves the desired posterior distribution as the unique stationary distribution of the Markov chain. Thus, by iterating these steps, the sampled states and parameters will eventually be distributed according to their posterior probability given the observed data. Critically, the rate at which the Markov chain converges to its stationary distribution is determined in part by the correlation between the sampled latent states at one iteration and those at the next. If the chain only makes minor updates to the latent state sequence, it will likely take a long time to converge to the desired posterior distribution. By performing joint, "block" updates of $\boldsymbol{x}_{1:T}$ and $\boldsymbol{z}_{1:T}$ in steps 1 and 2, we find that the latent state sequences are able to be explored more efficiently.