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

Scott W. Linderman, David M. Blei, and Liam Paninski
Columbia University

December 1, 2016

Abstract

Recent advances in neural recording technologies have enabled simultaneous measurements of the majority of head ganglia neurons in immobilized *C. Elegans* [Kato et al., 2015]. 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 Introduction

The nematode *C. Elegans* is unique among the model organism of neuroscience. For decades, we have known the connectome of its 302 neurons, yet a comprehensive understanding of hows its neural activity reflects sensory processes and produces behavior still eludes us [White et al., 1986]. While this connectome, or wiring diagram, has provided invaluable information about the structure of these neural circuits, the critical missing piece has been large scale functional measurements of these circuits in action. With the advent of optical recording technologies, we now possess a mounting set of tools for measuring the activity of large fractions of these neurons simultaneously. These technologies provide exciting opportunities to probe the link between neural activity, sensory processing, and, ultimately, behavior.

Recently, Kato et al. [2015] have harnessed these methodological advances to study the coordinated activity of hundreds of neurons in head-fixed *C. Elegans*. Across multiple organisms, they have found that neural activity reliably traces out smooth trajectories in the subspace spanned by the first three principal components. Upon closer evaluation, they have found that different components of these trajectories correspond to different elements

of behavior, like forward and reverse crawling, dorsal and ventral turns, etc. These results raise a number of interesting questions: can we learn a (potentially nonlinear) dynamical system that approximates these low-dimensional dynamics; is there a more appropriate subspace, or collection of subspaces, for capturing these dynamics; can we gain statistical power by partial, noisy recordings across separate worms and trials; and, can we segment these trajectories in an unsupervised manner to glean further insight into the relationship between low-dimensional state and behavior?

We have developed a probabilistic framework for investigating these questions. The findings of Kato et al. [2015] — namely, that neural activity is low dimensional with behavior-specific dynamics — suggest that these multi-neuronal recordings may be aptly modeled in terms of a dynamic, low-dimensional latent states. Moreover, they suggest that neural activity may be parsed into a sequence of discrete actions, like turning or crawling forward. These two properties are naturally combined by *switching linear dynamical systems* (SLDS), a class of probabilistic time series characterized by co-evolving discrete and continuous latent states. We have extended this class of models with a hierarchical Bayesian framework that captures many of the nuances of these whole-brain recordings. In fitting these models to recordings from multiple *C. Elegans*, we characterize the globally nonlinear dynamics that govern neural activity, and we parse these recordings into an interpretable sequence of behavioral segments.

The remainder of this paper is structured as follows. In Section 2 we introduce the class of switching linear dynamical systems and provide some insight into the latent states and model parameters. Section 3 develops a hierarchical extension of the SLDS to share statistical strength by combining recordings from multiple organisms with partially overlapping sets of observed neurons. Then, Section 4 presents results of applying these models to a set of recordings from five separate *C. Elegans*. We conclude with directions for further research.

2 Switching Linear Dynamical Systems

The switching linear dynamical system (SLDS) is a powerful model for stochastic time series with nonlinear dynamics [Ackerson and Fu, 1970, Chang and Athans, 1978, Hamilton, 1990, Bar-Shalom and Li, 1993, Ghahramani and Hinton, 1996, Murphy, 1998, Fox et al., 2009], and as such, it is particularly well suited to the analysis of whole-brain activity. 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 instantaneous $\Delta F/F$ 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 $\boldsymbol{y}_{1:T}$.

The SLDS 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

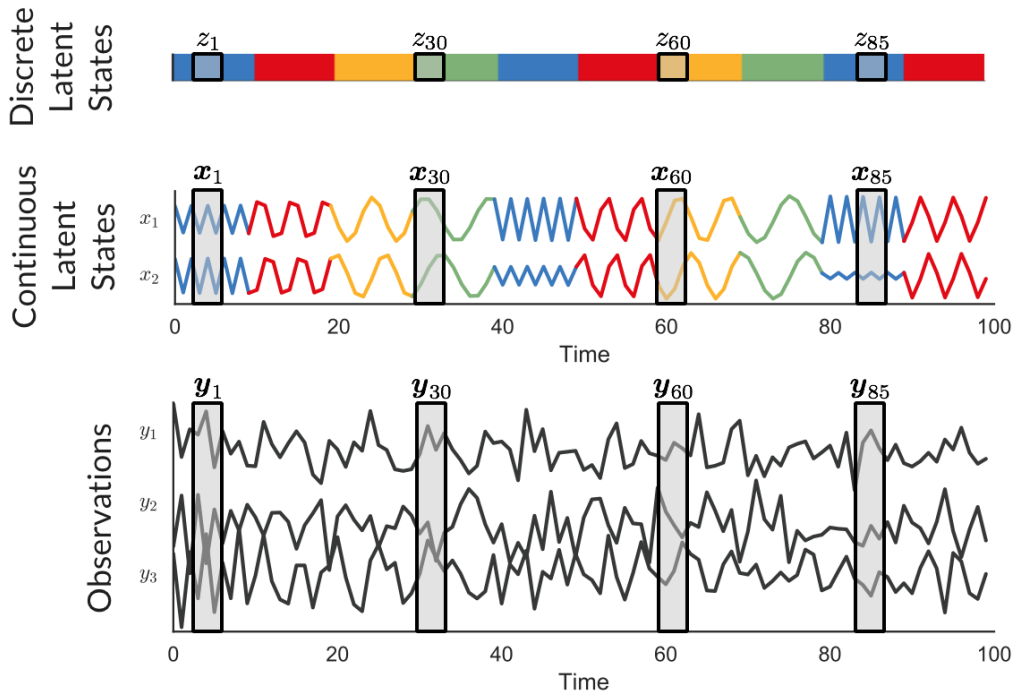

Figure 1: Simulated data from a switching linear dynamical system. The population stochastically switches between discrete states, $z_t \in \{1, \ldots, K\}$ (here, $K = 4$), each of which is color coded for visualization. These discrete states govern the linear dynamics of the continuous latent state, $\boldsymbol{x}_t \in \mathbb{R}^D$ (here, $D = 2$). For example, these states correspond to oscillatory dynamics with different frequencies. 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, $\boldsymbol{x}_t$, plus Gaussian noise. The correlations and dynamics in the observations are inherited from the dynamics of the latent states.

(iv) the observed neural activity is a linear function of the underlying states with additive Gaussian noise. Assumptions (i) and (iv) are justified by the low-dimensional trajectories that Kato et al. [2015] revealed with PCA (a linear dimensionality reduction method). Assumptions (ii) and (iii) follow from the trajectories naturally segment into discrete units, each with relatively simple dynamics.

These assumptions are formalized with the following generative model:

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

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,

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

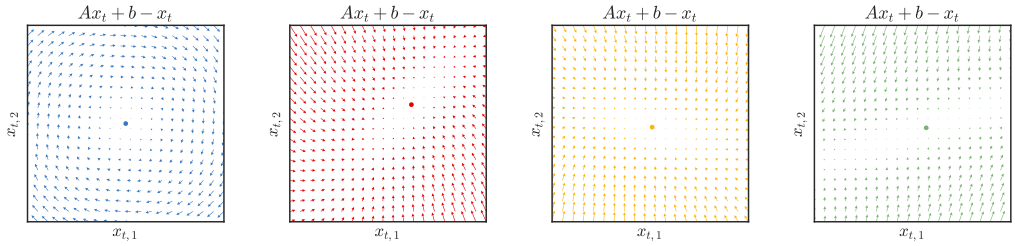

Figure 2: The dynamics corresponding to discrete latent state k may be visualized as a vector field where the arrows point to the expected next state, $\mathbb{E}[\boldsymbol{x}_{t+1}] = \boldsymbol{A}^{(k)}\boldsymbol{x}_t + \boldsymbol{b}^{(k)}$. Here, we show the dynamics for the first and fourth 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.

Next, the continuous latent state is imbued with linear Gaussian dynamics,

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

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

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

Thus, the parameters of the model are,

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

Figure 1 illustrates a sample from this generative model. The observations are a sequence of vectors, in this case they are $N = 3$ dimensional vectors, $\boldsymbol{y}_t$. These states are a noisy, linear projection of an underlying continuous latent state, $\boldsymbol{x}_t$, which is $D = 2$ dimensional in this case. Here, the continuous state switches between $K = 4$ discrete regimes (four different colors), each corresponding to different frequencies of oscillation. The discrete state, encoded by $z_t \in \{1, 2, 3, 4\}$ (blue, red, yellow, green), follows a simple Markov process with transition matrix $\boldsymbol{P} = \{\pi^{(k)}\}_{k=1}^{K}$ (not shown).

How should we interpret the these parameters? Linear dynamical systems can essentially capture simple dynamics, like rotations, exponential growth, and exponential decay. The stability of the system is determined by the eigenvalues of the dynamics matrix, $\boldsymbol{A}^{(k)}$ — if the eigenvalues all have modulus less than one, the system is asymptotically stable, and converges to a fixed point at $(\boldsymbol{I} - \boldsymbol{A}^{(k)})^{-1}\boldsymbol{b}^{(k)}$. If the largest eigenvalue is exactly one, the system oscillates in perpetuity. If any eigenvalue exceeds one in magnitude, the system diverges exponentially quickly. A variety of linear dynamical systems are illustrated in Figure 2. All of these systems are stable, and their fixed points are shown as dots.

While the capacity of linear dynamical systems is somewhat limited, the composition of many linear dynamical modes, as in an SLDS, yields a highly nonlinear system. This is particularly important for modeling neural data, which is believed to have nonlinear and

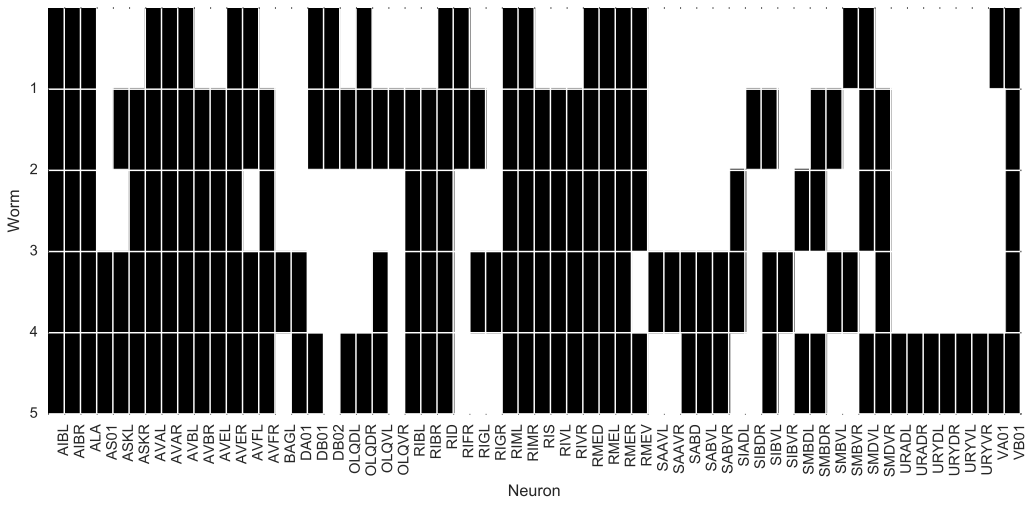

Figure 3: Identified neurons in the five worms from Kato et al. [2015]. Across all five worms, these 60 labeled neurons were identified in at least one worm. In addition to these labeled neurons, each worm contains roughly 70 unlabeled neurons that inform our estimate of that worm's latent states, but which we cannot share across worms.

often behavior-dependent dynamics.

Given observations, $\boldsymbol{y}_{1:T}$, our goal is to infer the discrete latent states, $bz_{1:T}$, and the continuous latent states, $\boldsymbol{x}_{1:T}$, as well as to learn the parameters, $\boldsymbol{\Theta}$. We take a Bayesian approach, specifying conjugate prior distributions for the parameters, and using Markov chain Monte Carlo (MCMC) to estimate the posterior distribution of the parameters and states given the observed data. Details of this MCMC algorithm are provided in Appendix A.

3 Hierarchical Switching Linear Dynamical Systems

The neural recordings of Kato et al. [2015] present additional opportunities to extend the SLDS. In these experiments, recordings have been made in five separate worms. Each recording contains roughly 100 of the worm's 302 neurons, but the subset of observed neurons varies from worm to worm. Moreover, while each of these 302 neurons has a unique name, only about 30 out of the 100 neurons could be reliably identified in each recording. For example, in the recording of worm A, we may observe 100 neurons of which 30 have been identified as, say, AAV, RIM, etc., while the remaining 70 are unidentified. Figure 3 shows the 60 labeled neurons which were identified in at least one of the worms. For example, ABL is identified in all five worms, whereas SIADL is only found in the second worm. In addition to these identified neurons, we also have roughly 70 unlabeled neurons for each worm. While these unlabeled neurons do not provide information that can be shared across worms, they do provide information about that worm's latent states. Our goal is to combine these recordings to extract a shared model of the neural dynamics of *C. Elegans*.

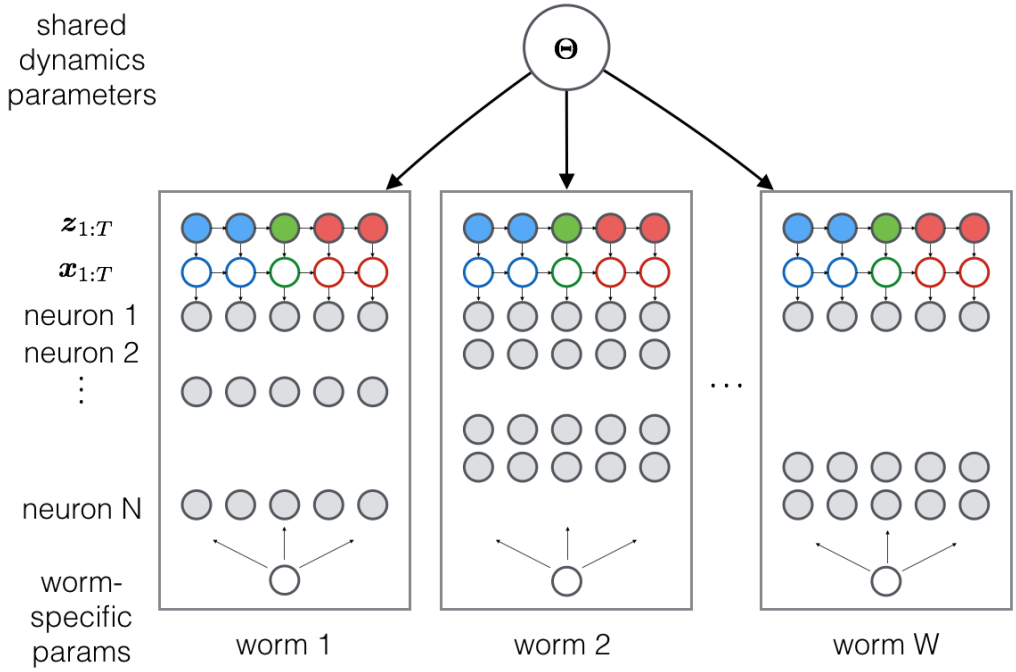

Figure 4: A schematic of the hierarchical SLDS for combining information across multiple worms in order to learn a shared dynamical system for *C. Elegans*. We have a set of shared set of model parameters, $\boldsymbol{\Theta}$, represented by the large node at the top of the graphical model. Each worm, w has its own set of discrete latent states (colored nodes) and continuous latent states (nodes with colored outlines) and observed neurons. For example, neuron 1 may appear in all neurons whereas neuron 2 may only be seen in worm 2. Thus, each worm provides information about the mapping from latent states to observations for only a subset of neurons. In terms of the model, this corresponds to a subset of rows in the matrix $\boldsymbol{C}$ and the vector $\boldsymbol{d}$. Finally, we allow each worm to have local parameters that capture, for example, the amount of fluorescent protein expressed in a particular neuron for a given worm.

We propose the following hierarchical SLDS (hSLDS) to combine data from multiple worms. Since these worms are genetically identical, we hypothesize that their dynamics should be the same. That is, they share the same set of K discrete states, as well as the parameters $\{\boldsymbol{A}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)}, \boldsymbol{\pi}^{(k)}\}_{k=1}^{K}$. We assume the following mapping from latent states to the neural activity of worm w, time t, and neuron n:

$$y_{w,t,n} = \gamma_{w,n}(\boldsymbol{c}_n^{\mathsf{T}}\boldsymbol{x}_{w,t} + d_n) + \epsilon_{w,t,n}, \tag{6}$$

$$\epsilon_{w,t,n} \sim \mathcal{N}(0, r_{w,n}^2). \tag{7}$$

Thus, the mapping from latent states to the activity of neuron n is partially shared across worms, in that all worms share the vector $\boldsymbol{c}_n$. These constitute the rows of an emission matrix, $\boldsymbol{C}$. Note that, for simplicity, we have required all discrete states, z_t to share the same emission matrix. However, we have introduced a set of worm-specific, scalar parameters, $\{\gamma_{w,n}$ and $r_{w,n}\}$. The gain, $\gamma_{w,n}$, captures differences in amplitude of the neural

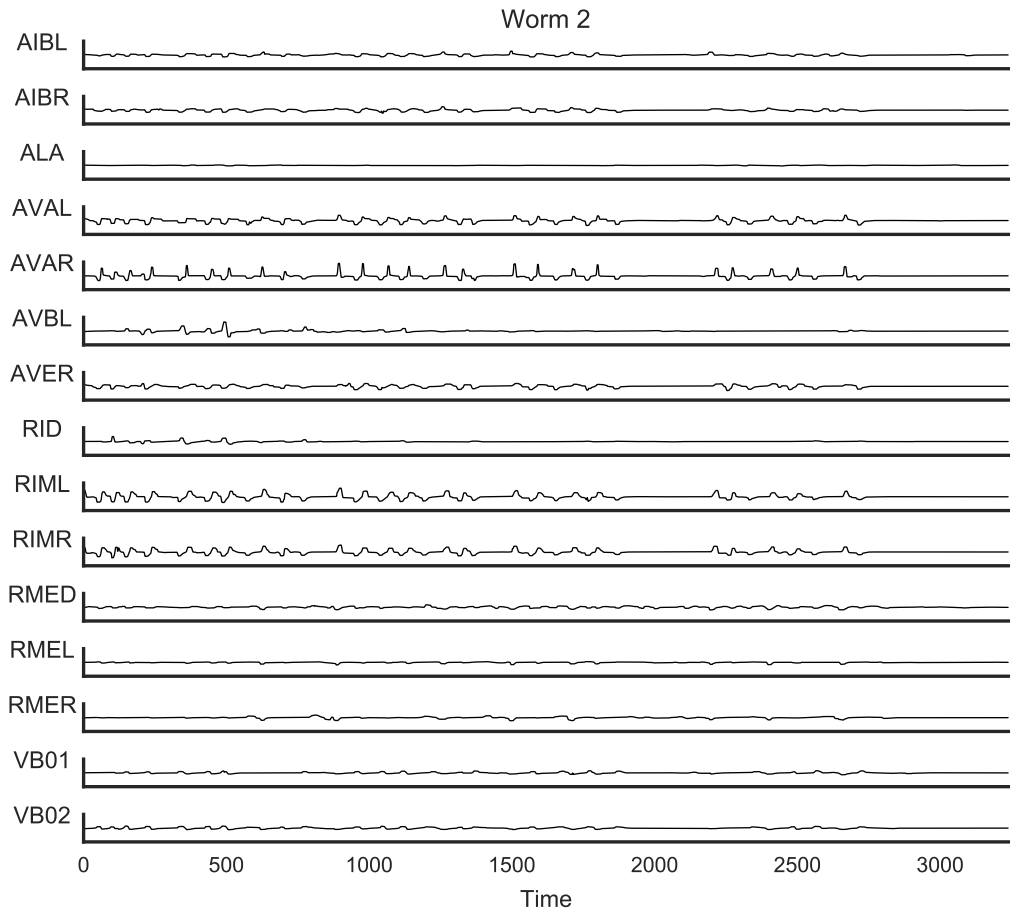

Figure 5: Fluorescence traces of 15 neurons in the second worm. These neurons were identified in all 5 worms. Axis limits are the same for all neurons.

signal that may arise from variation in fluorescent protein expression from worm to worm. The noise variance, $r^2_{w,n}$, enables neuron n to have different marginal variance from worm to worm, which may arise due to a number of factors, like protein expression levels and the clarity of a particular recording. Whereas the general SLDS allows for covariance in the noise from neuron to neuron, for simplicity we have assumed that this matrix is worm-specific and diagonal. Figure 4 provides a graphical representation of this model.

4 Results

We fit the hSLDS to recordings of five *C. Elegans* collected by Kato et al. [2015]. Each recording is 18 minutes long with a sampling frequency of 3Hz, yield 3240 time steps per worm. The recordings contain $N_w = 109$, 107, 131, 126, and 129 neurons, respectively. Of these, a subset have been manually identified. Figure 3 shows the neurons that were

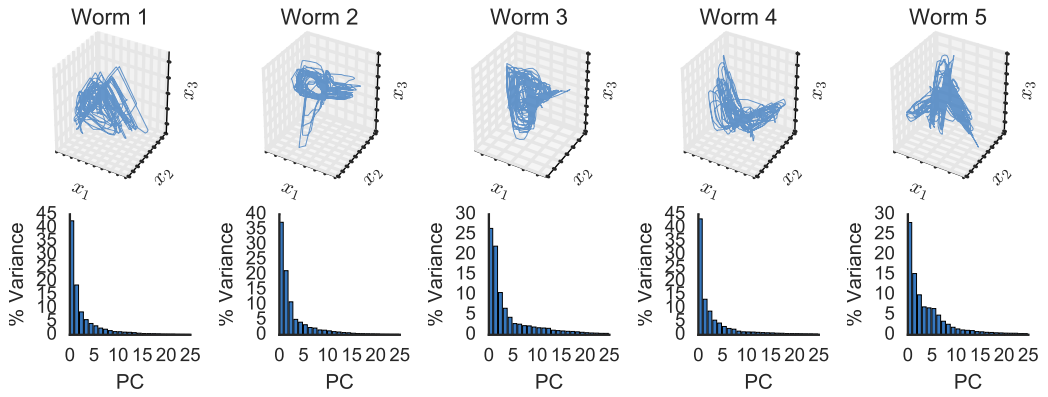

Figure 6: Principal components analysis of each individual worm. In all cases, the neural activity is relatively low dimensional, with large percentages of the variance captured by the first three principal components (PC's). Above, the projection of roughly 100-dimensional neural activity onto the subspace spanned by the first three PC's. The hierarchical SLDS will find a common subspace for all worms, and, furthermore, segment the low dimensional trajectories into a sequence of discrete pieces.

identified in each worm. We will capitalize on those neurons that are identified in multiple worms in order to learn a shared subspace of neural activity and a shared set of dynamical states for all *C. Elegans*. Figure 5 shows the neural activity of 15 identified neurons in the second worm. We immediately recognize repeating patterns of activity; these will correspond to characteristic "loops" in the low dimensional portrait of the population activity.

Neural Activity is Low Dimensional A fundamental assumption of the hSLDS is that neural activity is low dimensional. To test this assumption, we first run principal components analysis (PCA) on each worm's population activity (a $T \times N_w$ matrix) separately. We find that the top three principal components (PC's)capture a large fraction of the variance for all worms, as shown in the bottom panels of Figure 6. The top panels show the projection of neural activity onto the first three PC's. While the PCA projections show qualitatively similar patterns, like intersecting and orthogonal loops, the subspaces are not aligned across worms and the trajectories are not particularly smooth or interpretable. The hSLDS addresses these issues by learning a shared subspace across all worms, and segmenting the low dimensional trajectories into a sequence of states with simple, linear dynamics.

The hSLDS reveals a shared subspace and canonical dynamics across worms The hSLDS shares information across worms in order to learn a canonical subspace of neural activity and a common set of low-dimensional dynamics. Figure 7 shows the inferred continuous latent states, $\boldsymbol{x}_{1:T}$, of the hSLDS for each of the five worms. There are three points to note, First, in contrast to the PCA trajectories in Figure 6, the similarity in hSLDS trajectories is readily apparent. This is due to the fact that all worms share the same subspace, which is defined by the global emission matrix $\boldsymbol{C}$. Second, the hSLDS trajectories are considerably smoother than their PCA counterparts. This is a direct result of the set of

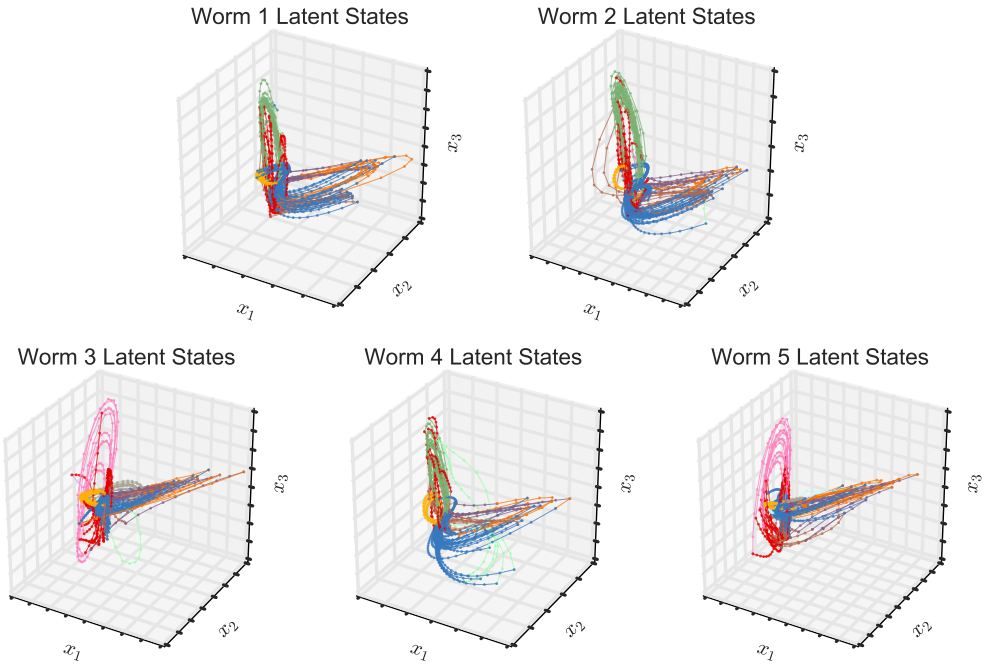

Figure 7: Latent states discovered by the hierarchical SLDS. Since the emission matrix, $\boldsymbol{C}$, is shared across worms, the coordinates of the latent states are shared as well. This reveals a common pattern of low dimensional activity characterized by interlocking and opposing loops. Colors denote the discrete latent state inferred for each part of the trajectory. For example, one loop begins with the latent state shooting out from the origin along the orange and purple states, and then returning via the blue state. Another begins with the green state, which swings upward and is then followed by a return via the red state. In worms 3 and 5, the upward swing is slightly tilted, as captured by the pink rather than the green state. While harder to see in this view, there is a third loop in yellow that lies in the (x_1, x_2) plane. See supplementary videos for further detail.

linear dynamics that comprise the hSLDS. Assuming that the latent state, $\boldsymbol{x}_{t+1}$, is linearly related to the preceding state, $\boldsymbol{x}_t$, induces an inductive bias toward smooth trajectories in latent space. This reduces the discontinuities and noise in the PCA trajectories and reveals the characteristic patterns of latent dynamics. Third, the hSLDS trajectories have been segmented into snippets, each corresponding to a different discrete latent state, z_t, with its own set of linear dynamics. These discrete states illustrated as different colors in the hSLDS trajectories. These provide a handle into understanding the latent dynamics of *C. Elegans* population activity, and a useful latent variable for relating activity to behavior.

Neural dynamics switch between a set of linear regimes As we have stated, the hSLDS segments the low dimensional trajectories into a sequence of discrete states, each governed by simple linear dynamics. These dynamics are shown in Figure 8. The states are ordered from most common (State 1) to least common (State 9). The model was fit with $K = 15$ states, but only 11 were used. The remaining two are omitted from this figure since they are used very rarely. The vector fields shown here point in the direction

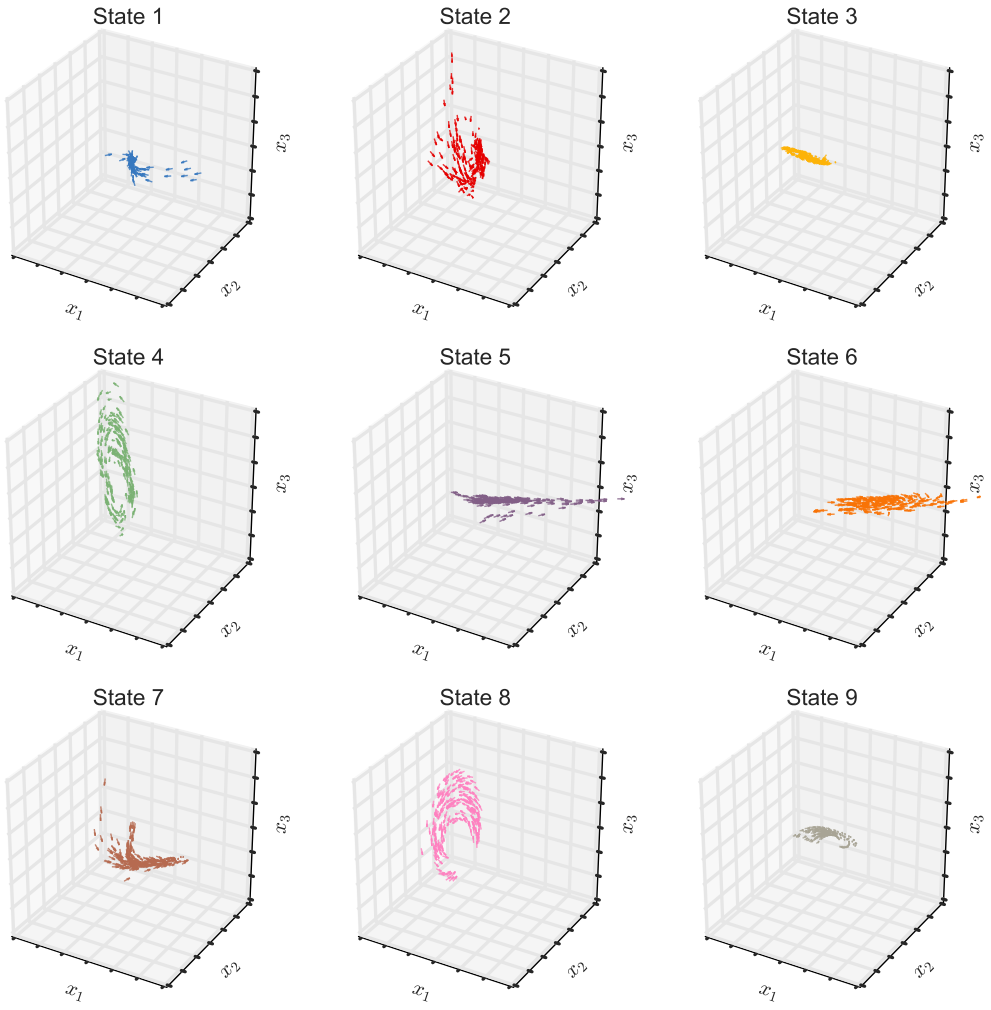

Figure 8: The discrete dynamical states that compose this switching linear dynamical system. The vector fields point in the direction of $\boldsymbol{A}^{(k)}\boldsymbol{x} + \boldsymbol{b}^{(k)}$ $(k = 1 \dots 9)$. While the dynamics are defined for all $\boldsymbol{x}$, they are only plotted at locations where the corresponding state is used. States are ordered from most common (1) to least common (9); see supplementary material for remainder of states. The most common states (1 and 2) simply decay to the origin. State 3 is a tight loop in the (x_1, x_2) plane. States 4 and 8 are larger loops in the (x_2, x_3) plane. States 5, 6, and 9 are sections of a third loop opposite to State 3. Finally, State 7 seems to bridge between loops.

of $\boldsymbol{A}^{(k)}\boldsymbol{x}_t + \boldsymbol{b}^{(k)}$ for $k = 1, \dots, 9$. While these dynamics are defined for all $\boldsymbol{x}_t \in \mathbb{R}^3$, for clarity, they are only plotted at the locations in latent space where the corresponding state is deployed.

The most common states (1 and 2) correspond to decays toward the origin. This is to be expected, since the neurons are quiescent most of the time (c.f. Figure 5). The remaining states capture different sections of the intersecting loops seen in Figure 7. State 3 is a short loop to the left in the (x_1, x_2) plane. States 4 and 8 correspond to the large upward swings that trace out loops in the (x_2, x_3) plane. States 5, 6, and 9 correspond to loops that swing

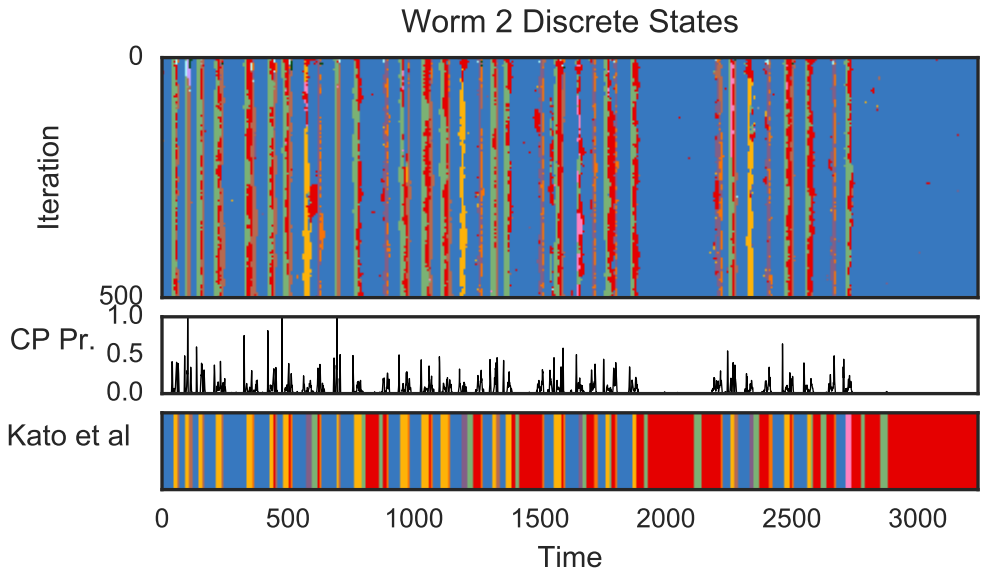

Figure 9: The inferred discrete states align with the manual segmentation of Kato et al. [2015]. *Top:* samples from the posterior distribution of discrete states, $z_{1:T}$. Each row is a sample, and each color denotes a distinct value from 1 to $K = 15$. *Middle:* from these samples we estimate the probability of a change-point (i.e. $z_t \neq z_{t-1}$). *Bottom:* The inferred change-points align with the manual segmentation of Kato et al. [2015]. Note, however, that the colors do not perfectly align since there is not a one-to-one correspondence between the inferred and manual states. For example, the blue and red states of Kato et al. [2015], which correspond to "REVSUS" and "SLOW," respectively, are combined into inferred State 1.

out to the right of the origin, again in the (x_1, x_2) plane. Finally, state 7 seems to connect the (x_2, x_3) loop to the (x_1, x_2) loop, bypassing the origin.

This discrete segmentation of the latent state trajectories serves multiple purposes. First, by stitching together discrete states with unique linear dynamics, the hSLDS models a globally *nonlinear* dynamical system. Second, these discrete latent states provide a valuable covariate for understanding switches in the neural activity. As we will see below, when we condition on the entry into these discrete states, characteristic patterns of population activity are revealed. Third, these states provide an interpretable summary of the complex, global brain activity. Next, we will show that these discrete states correspond to meaningful behaviors.

Discrete states correspond to distinct behaviors Kato et al. [2015] manually labeled segments of time corresponding to distinct behaviors, like forward or reverse crawling and dorsal or ventral turns. This process is both laborious and subjective. The hSLDS provides a principled and automated alternative to manual labeling, by automatically segmenting activity into snippets governed by linear dynamics. We find that the manual and hSLDS segmentations are closely aligned.

Figure 9 illustrates both the uncertainty of the discrete labeling as well as the alignment with the "ground truth" labels. The top panel shows samples from the posterior distribution

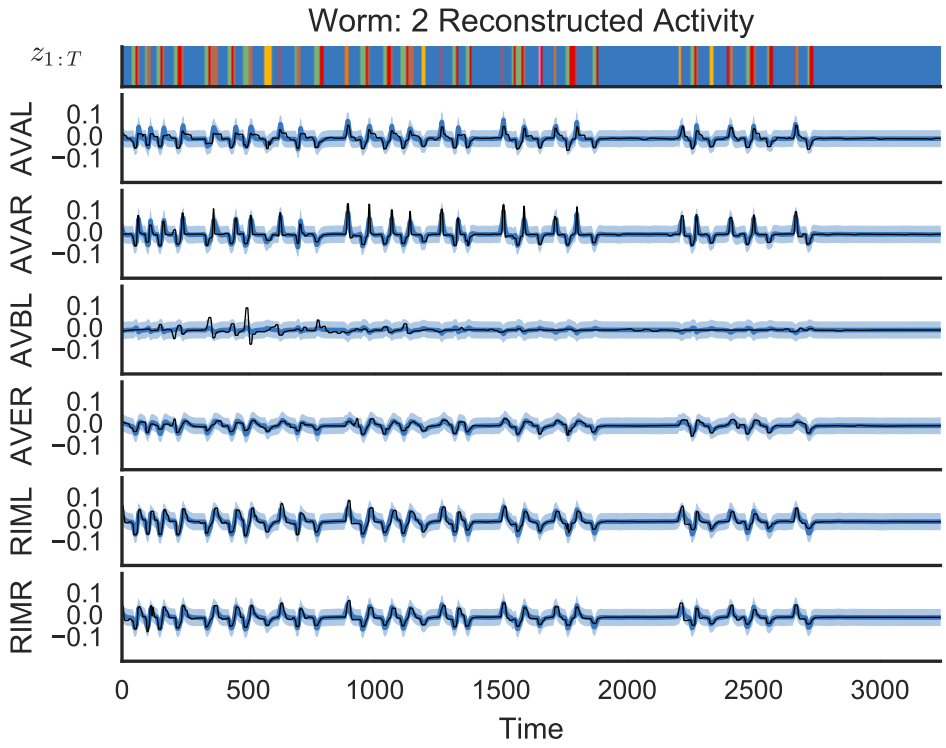

Figure 10: Observed (black) and smoothed (blue) activity of six neurons from worm 2. The smoothed activity is given by $\gamma_{w,n}(\boldsymbol{c}_n^\mathsf{T}\mathbb{E}[\boldsymbol{x}_{w,t}]+d_n)$. Shading denotes ± 1 standard deviation under the posterior given by $\sqrt{\gamma_{w,n}^2\boldsymbol{c}_n^\mathsf{T}\mathrm{Var}[\boldsymbol{x}_{w,t}]\boldsymbol{c}_n+\sigma_{w,n}^2}$. Inferred discrete states are shown above for reference.

over latent state sequences. Each row corresponds to a sample of $z_{1:T}$, and as before, the colors denote different discrete states. The variability from sample to sample highlights the uncertainty that is inherent in assigning a discrete label to a point in time — where one state ends and the next begins is not entirely clear from the noisy data.

From this sequence of discrete state samples we derive an estimate of the *change-point probability*, i.e. the probability that $z_t \neq z_{t-1}$. This is shown in the middle panel. In some instants it is eminently clear that the discrete state has switched; this is due to large changes in the dynamical trajectory that favor one state over another. More often, however, this probability is lower and distributed over a window of time. Thus, our Bayesian approach provides a principled means of assessing this uncertainty.

Finally, we see that these change-points align with the manual segmentation of Kato et al. [2015], which is shown in the bottom panel. Note that the manual and inferred state labels are not in one-to-one correspondence. This is to be expected given both the invariance to permutation (though both are sorted by usage), but more importantly because some manual states are grouped into one inferred state. For example, the red and blue manual states, which are labeled "REVSUS" and "SLOW", respectively, are combined in the blue inferred state. Likewise, the manual sequence of purple followed by green, which corresponds to a

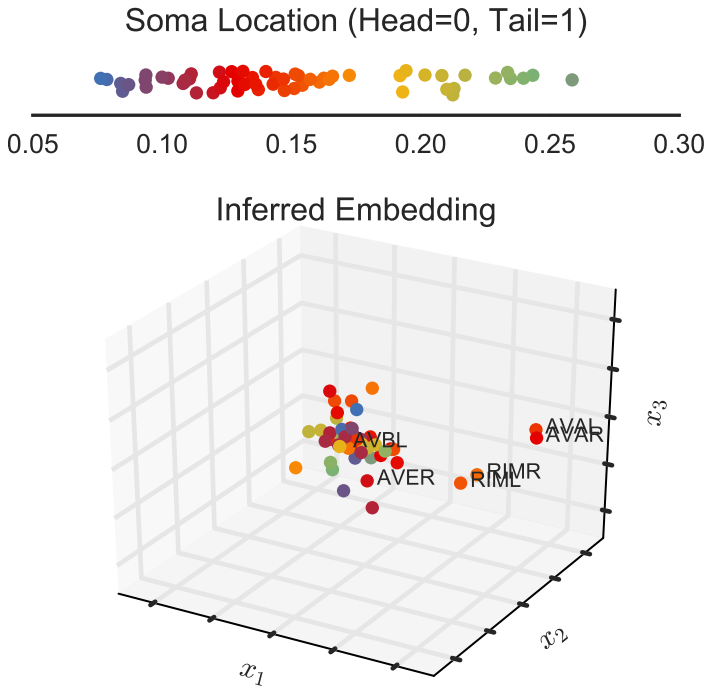

Figure 11: Three-dimensional embedding of neurons given by the rows of the emission matrix, $\boldsymbol{c}_n$. For reference, the true, one-dimensional location of the soma is shown above (jittered for clarity). While most neurons are clustered near the origin (reflecting the fact that most neurons have low amplitude activity), the more active neurons (shown above) stand out.

dorsal turn followed by a forward crawl, is combined in the yellow inferred state. Other states, like the yellow manual state, which corresponds to a ventral turn, are separated into two inferred states, here red and green. While the exact divisions between states are somewhat arbitrary, the change points in these trajectories are clearly revealed by the hSLDS.

The hSLDS captures salient patterns of neural activity As a generative model, the hSLDS provides a distribution over observed neural activity. This provides a valuable sanity check for our results — does the model capture the variability of the data? Figure 10 shows that indeed it does. We plot the observed traces of six neurons in black and the predicted traces under the posterior of the hSLDS in blue. With a Bayesian approach, we also obtain an estimate of the uncertainty of these predictions, shown here as shaded blue regions around the mean. We see that for most neurons, the predicted and observed trajectories are well aligned, indicating that the hSLDS with $D = 3$ dimensional latent states has sufficient capacity to capture the variability of the data; however, some spikes in the activity of neuron AVBL are smoothed over in this reconstruction.

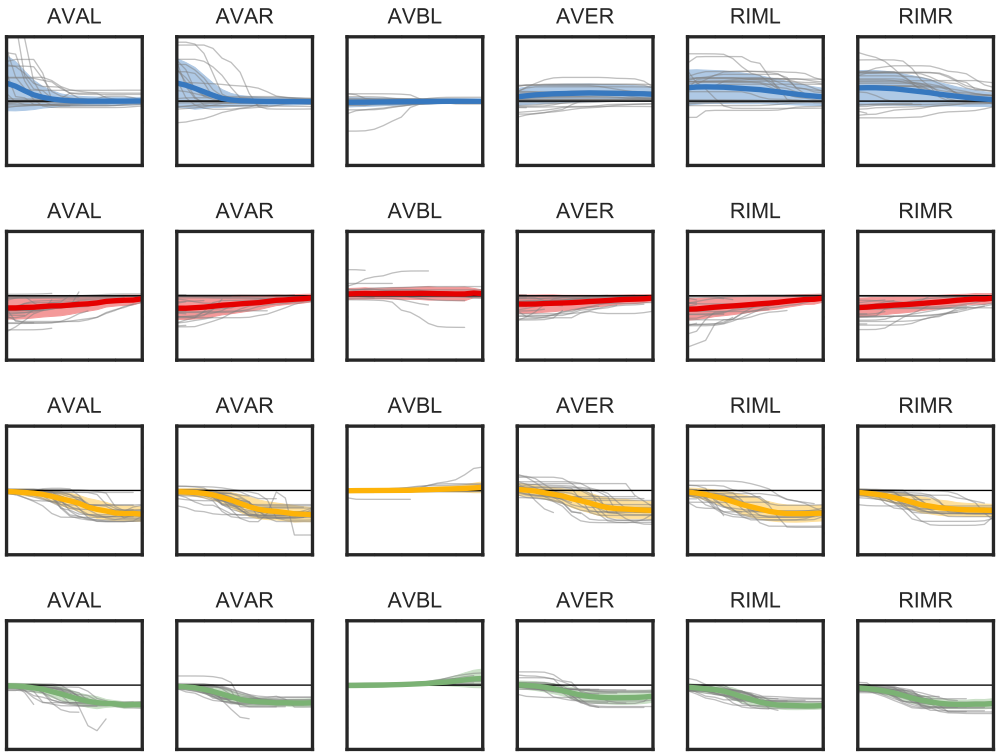

Figure 12: Neural activity conditioned upon entry into each discrete state. The x-axis is time since entering the discrete state (from 0 to 5 seconds), and the y-axis is the average neural response ± 2 standard deviations. Light gray traces show a sample of individual neural activity traces. Each row corresponds to the discrete state with the corresponding color. Here shown are States 1-4.

The hSLDS discovers a low-dimensional embedding of neurons In addition to yielding an interpretable, low-dimensional portrait of neural activity, the hSLDS also provides an embedding of neurons in latent space via the rows of the emission matrix, $\boldsymbol{C}$. Nearby neurons in latent space will have highly correlated activity, and neurons that are opposed will be anti-correlated. Figure 11 shows this embedding for the 60 neurons that are identified in at least one of the five worms. For reference, the true one-dimensional location of the neuron's soma along the head to tail axis is shown above. Most of the neurons are clustered around the origin, reflecting the fact that most neurons have little activity. However, two pairs of neurons stand out: (AVAL, AVAR) and (RIML, RIMR). These pairs of neurons have larger amplitude activity as shown above. These pairs both drive backward locomotion (reversal), and are correlated via a gap junction.

Discrete state onset reveal stereotyped population activity The discrete latent states partition neural activity into interpretable and behaviorally relevant segments. As such, they are natural covariates for neural activity. Figures 12 and 13 illustrate the distribution of neural responses upon entry into each of the top 8 discrete states. The x axis in these plots denotes time from 0 to 5 seconds after entry into the corresponding state. The

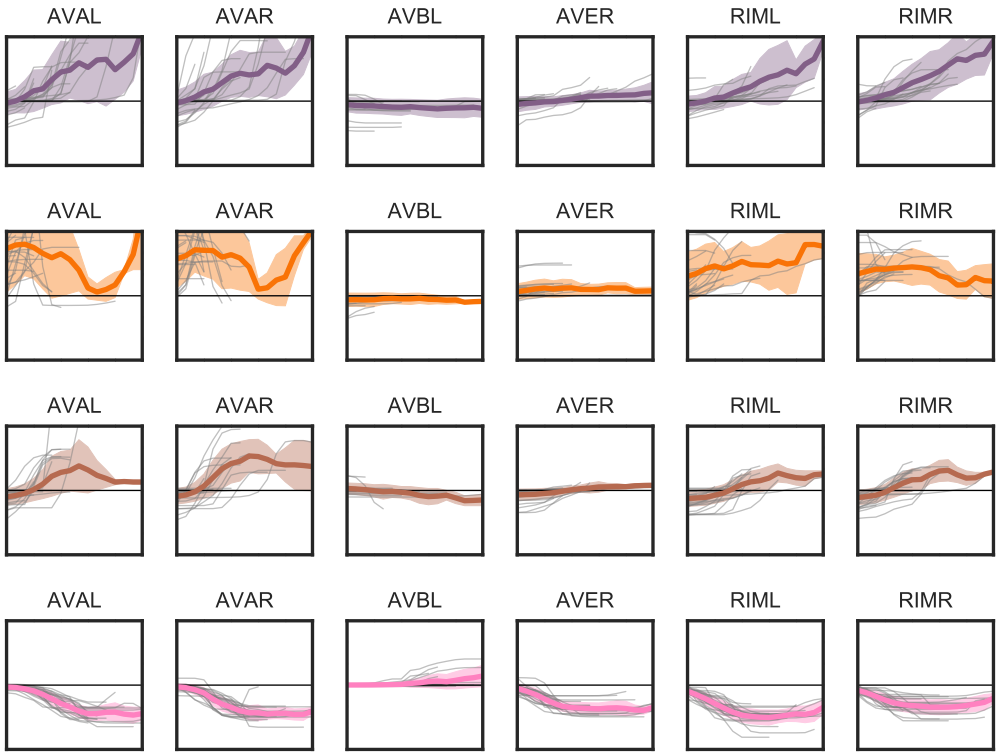

Figure 13: Same as figure above but for States 5-8.

colored line denotes the mean response of the corresponding neuron (same y limits in all panels), and the shaded regions denote ± 2 standard deviations. The light gray traces show a sample of randomly chosen neural responses. That AVA and RIM neurons are associated with reverse crawling suggests that states 2, 3, and 8 are related to forward locomotion (since the neural activity is decreasing) and states 5, 6, and 7 are related to the reverse locomotion (since the activity is increasing).

5 Discussion

The elucidation of global brain dynamics and their relation to behavior is perhaps the foremost goal of systems neuroscience. *C. Elegans*, with its well-studied connectome and corresponding array of optical imaging tools, provides many exciting opportunities to make progress toward this goal. To do so, however, requires statistical methods capable of revealing latent structure in complex neural data, and combining data across many worms. The hierarchical SLDS satisfies many of these demands, decomposing high-dimensional, nonlinear dynamics into a sequence of discrete latent states with low-dimensional, linear dynamics. Moreover, by sharing across worms while allowing for individual variability, the hSLDS reveals a common set of latent dynamics.

Many questions remain. First, as a generative model, the hSLDS falls somewhat short with its relatively limited model for discrete dynamics. It is clear from these portraits of neural activity that discrete states are deployed in very localized regions of latent space. Our recent work on recurrent switching linear dynamical systems (rSLDS) leverages this type of structure [Linderman et al., 2016]. In future work, we intend to combine the hSLDS and rSLDS and apply them to this data.

Second, our analyses of state-triggered neural activity have only scratched the surface. We intend to look more closely at the population responses in these distinct latent states and search for further insights into how this low-dimensional population activity drives behavior.

Third, we have not leveraged the availability of the *C. Elegans*connectome in this work. In theory, this connectome should provide strong information about the correlational structure of population activity. It remains an open question how best to combine this anatomical data with state space models like these.

Finally, while we have utilized the neurons that are found in multiple worms to learn a shared subspace of activity and a common set of discrete states, we could further improve upon this work by inferring neuron identities for the hundreds of unlabeled neurons. Their rough locations and the correlations across worms provides a substantial amount of information to tackle this problem.

References

Guy A Ackerson and King-Sun Fu. On state estimation in switching environments. *IEEE Transactions on Automatic Control*, 15(1):10–17, 1970.

Yaakov Bar-Shalom and Xiao-Rong Li. *Estimation and tracking*. Artech House, Boston, MA, 1993.

Chaw-Bing Chang and Michael Athans. State estimation for discrete systems with switching parameters. *IEEE Transactions on Aerospace and Electronic Systems*, (3):418–425, 1978.

Emily Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. Nonparametric Bayesian learning of switching linear dynamical systems. *Advances in Neural Information Processing Systems*, pages 457–464, 2009.

Zoubin Ghahramani and Geoffrey E Hinton. Switching state-space models. Technical report, University of Toronto, 1996.

Walter R Gilks. *Markov chain monte carlo*. Wiley Online Library, 2005.

James D Hamilton. Analysis of time series subject to changes in regime. *Journal of econometrics*, 45(1):39–70, 1990.

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.

Scott W Linderman, Andrew C Miller, Ryan P Adams, David M Blei, Liam Paninski, and Matthew J Johnson. Recurrent switching linear dynamical systems. *arXiv preprint arXiv:1610.08466*, 2016.

Kevin P Murphy. Switching Kalman filters. Technical report, Compaq Cambridge Research, 1998.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

John G White, Eileen Southgate, J Nichol Thomson, and Sydney Brenner. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1–340, 1986.

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,

$$p(\boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{\Theta} \mid \boldsymbol{y}_{1:T}) = \frac{p(\boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{\Theta})}{p(\boldsymbol{y}_{1:T})}. \tag{8}$$

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,

$$p(\boldsymbol{y}_{1:T}) = \int p(\boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{\Theta}) \, \mathrm{d}\boldsymbol{x}_{1:T} \, \mathrm{d}\boldsymbol{z}_{1:T} \, \mathrm{d}\boldsymbol{\Theta}. \tag{9}$$

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 [Gilks, 2005, Robert and Casella, 2013] offer one such approach.

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.

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 $p(\boldsymbol{z}_{1:T} \mid \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta})$ 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 $p(\boldsymbol{x}_{1:T} \mid \boldsymbol{z}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta})$ 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,

$$\begin{aligned} p(\boldsymbol{A}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)} \mid \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta}) \\ \propto p(\boldsymbol{A}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)}) \prod_{t=1}^T \left[\mathcal{N}(\boldsymbol{x}_t \mid \boldsymbol{A}^{(k)} \boldsymbol{x}_{t-1} + \boldsymbol{b}^{(k)}, \boldsymbol{Q}^{(k)} \right]^{\mathbb{I}[z_{t-1}=k]}. \end{aligned} \tag{10}$$

   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,

$$\begin{aligned} p(\boldsymbol{C}^{(k)}, \boldsymbol{d}^{(k)}, \boldsymbol{R}^{(k)} \mid \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta}) \\ \propto p(\boldsymbol{C}^{(k)}, \boldsymbol{d}^{(k)}, \boldsymbol{R}^{(k)}) \prod_{t=1}^T \left[\mathcal{N}(\boldsymbol{y}_t \mid \boldsymbol{C}^{(k)} \boldsymbol{x}_t + \boldsymbol{d}^{(k)}, \boldsymbol{R}^{(k)} \right]^{\mathbb{I}[z_t=k]}. \end{aligned} \tag{11}$$

   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,

$$p(\boldsymbol{\pi}^{(k)} \mid \boldsymbol{z}_{1:T}) \propto p(\boldsymbol{\pi}^{(k)}) \prod_{t=1}^T \left[\pi_{z_t}^{(k)} \right]^{\mathbb{I}[z_{t-1}=k]}. \tag{12}$$

   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,

$$p(\boldsymbol{\pi}^{(k)} \mid \boldsymbol{z}_{1:T}) = \mathrm{Dir}(\boldsymbol{\pi}^{(k)} \mid \widetilde{\boldsymbol{\alpha}}^{(k)}) \tag{13}$$

$$\widetilde{\alpha}_{k'}^{(k)} = \alpha + \sum_{t=1}^T \mathbb{I}[z_t = k'] \, \mathbb{I}[z_{t-1} = k]. \tag{14}$$

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.