A Appendix / supplemental material

A.1 p-dLDS Algorithm

Algorithm 1 describes the proposed inference algorithm. In our experiments, we set $n = 1$ and $\eta = 10^{-4}$ and observe that the model converges. Below we use the notation hat notation for latent variable estimates or samples and the variable itself to represent the parameters of the variational distributions.

Algorithm 1 Variational EM for Probabilistic dLDS

Require: M observation dimension, N latent state dimension, K number of dynamic operators, S moving average window size, ξ SBL-DF trade-off parameter, n number of samples to estimate expectations, η sparsity threshold, θ model parameters.

// Initialize parameters
$\boldsymbol{c}_t \leftarrow \boldsymbol{0}$
$D_{i,j} \sim \mathcal{N}(0, \sigma^2)$
$f_{k,i,j} \sim \mathcal{N}(0, \sigma^2)$
$\widehat{\boldsymbol{x}}_t \leftarrow \boldsymbol{D}^+ \boldsymbol{y}_t$ ▷ Initialize latent state with PCA

while ELBO has not converged **do**
    // Update Latent State Posterior
    $\widehat{\boldsymbol{b}}_{1:T} \leftarrow \text{MovingAverage}_S(\widehat{\boldsymbol{x}}_{1:T})$
    $\widehat{\boldsymbol{c}}_t \sim q(\boldsymbol{c}_t)$
    $\widehat{\boldsymbol{F}}_t \leftarrow \sum_{k=1}^{K} \boldsymbol{f}_k \widehat{c}_{k,t}$
    $\boldsymbol{l}_{1:T}, \boldsymbol{\Sigma}_x \leftarrow \text{KalmanSmoother}(\boldsymbol{y}_{1:T}, \widehat{\boldsymbol{b}}_{1:T}, \widehat{\boldsymbol{F}}_{1:T}, \theta)$

    // Update Coefficient Posterior
    Initialize $q(\boldsymbol{c})$ and $q(\boldsymbol{\gamma})$ jointly with SBL-DF.
    Update $q(c_{t,k})$ with SGD over equation (10) for densities where $|c_{t,k}| > \eta$.
    Update $q(\gamma_t) \leftarrow \mathcal{IG}(\xi + \frac{n}{2}, \xi c_{t-1,k}^2 + \frac{\sum_{i=1}^{n}(\tilde{c}_{t,k,i} - c_{t,k})^2}{2})$

    // Update Parameters
    Update θ with SGD over equation (11).
end while

B Latent Variable Inference

B.1 Lemma 1 Derivation

Lemma 1. *Let the transition between any two state vectors $\boldsymbol{x}_t, \boldsymbol{x}_{t+1} \in \mathbb{R}^N$ be defined by the linear dynamics matrix $\boldsymbol{F}_t \in \mathbb{R}^{N \times N}$ and the dynamics offset $\boldsymbol{b}_t \in \mathbb{R}^N$. For any $\lambda > 0$, the objective,*

$$\arg\min_{\boldsymbol{F}_t, \boldsymbol{b}_t} \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t - \boldsymbol{F}_t \boldsymbol{x}_t - \boldsymbol{b}_t\|_2^2 + \lambda \|\boldsymbol{F}_t\|_2^2,$$

is minimized when $\boldsymbol{F}_t = \boldsymbol{0}$ and $\boldsymbol{b}_t = \boldsymbol{x}_{t+1} - \boldsymbol{x}_t$.

Proof. Let $\boldsymbol{r}_t = \boldsymbol{x}_{t+1} - \boldsymbol{x}_t$. We can rewrite the reconstruction objective in the following form,

$$\arg\min_{\boldsymbol{F}_t, \boldsymbol{b}_t} \|\boldsymbol{r}_t - \boldsymbol{F}_t \boldsymbol{x}_t - \boldsymbol{b}_t\|_2^2 + \lambda \|\boldsymbol{F}_t\|_2^2.$$

This objective is identical to the standard ridge regression with an unpenalized intercept term [13]. The solution is obtained by first centering the data, and then solving for the parameters using the solution for the standard Tikhonov regression. Below, we define the centered data as $\tilde{\boldsymbol{x}}_t$ and $\tilde{\boldsymbol{r}}_t$ for inputs and outputs respectively. Finally, we can use these values to obtain the following estimates of the parameters,

$$\widehat{\boldsymbol{b}}_t = \boldsymbol{\mu}_t \qquad \widehat{\boldsymbol{F}}_t = (\tilde{\boldsymbol{x}}_t^\top \tilde{\boldsymbol{x}}_t + \lambda \boldsymbol{I})^{-1} \tilde{\boldsymbol{x}}_t^\top \tilde{\boldsymbol{r}}_t$$

However, when there is only a single datapoint, we get that $\tilde{x}_t = 0$, which results in $\widehat{\boldsymbol{F}} = 0$. □

This result arises from having only a single observation for any dynamic transition, which leads to a singular design matrix. Although we can improve our estimate of $\boldsymbol{F}_t$ by collecting more samples along a given trajectory, this is impractical when dealing with naturalistic time-series. For instance, it may be infeasible to collect more data from the exact same initial condition in a naturalistic environment due to noise in the experimental setup. In chaotic systems, minor deviations can lead to drastically different outcomes over long time horizons. Even if it were possible to precisely control for the initial condition of the signal, the presence of dynamical noise can cause initially aligned time series to quickly drift out of alignment. Consequently, it is not uncommon to observe a single transition between any two time points, as it is not guaranteed that events across multiple trials will be well-aligned.

B.2 Lemma 2 Derivation

Lemma 2. *Let $l, b \in \mathbb{R}^N$ be independent random variables such that $l \sim p(l)$ and $b \sim p(b)$. Their sum $x = l + b$ is distributed according to $\mathcal{N}(\boldsymbol{\mu}_l + \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b)$ when 1) $p(b) = \mathcal{N}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$ and $p(l) = \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ and when 2) $p(b) = \delta(b - \boldsymbol{\mu}_b)$ and $p(l) = \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b)$.*

Proof. **Case 1.** Let $l \sim \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ and $b \sim \mathcal{N}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$. The sum of normal random variables follows a distribution that results from convolving their individual distributions,

$$\begin{aligned}
q(\boldsymbol{x}) &= q(\boldsymbol{l} + \boldsymbol{b}) \\
&= q(\boldsymbol{l}) * q(\boldsymbol{b}) \\
&= \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) * \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_b) \\
&= \mathcal{N}(\boldsymbol{\mu}_l + \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b)
\end{aligned}$$

This is a standard result from probability theory.

Case 2. Now let $l \sim \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b)$ and $b \sim \delta(b - \boldsymbol{\mu}_b)$. Similarly, the distribution of the sum of these variables is distributed according to their convolution,

$$\begin{aligned}
q(\boldsymbol{x}) &= q(\boldsymbol{l}) * q(\boldsymbol{b}) \\
&= \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b) * \delta(\boldsymbol{b} - \boldsymbol{\mu}_b) \\
&= \int_{-\infty}^{\infty} \mathcal{N}(\boldsymbol{x} - \boldsymbol{\tau}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b) \delta(\boldsymbol{\tau} - \boldsymbol{\mu}_b) d\boldsymbol{\tau} \\
&= \mathcal{N}(\boldsymbol{x} + \boldsymbol{\mu}_b; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b) \\
&= \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_l + \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_b),
\end{aligned}$$

where the fourth line is the result of the sifting property of delta distributions. Since the final distribution in Case 1 and Case 2 are identical, we complete the proof. □

B.3 Optimal $q(x)$ Update

The optimal coordinate ascent variational update is given by the following equation,

$$\begin{aligned}
\log q^*(\boldsymbol{x}) &\propto \mathbb{E}_{q(\boldsymbol{c}, \boldsymbol{\gamma})}[\log p(\boldsymbol{x}, \boldsymbol{c}, \boldsymbol{y}, \boldsymbol{\gamma} | \theta)] \\
&= \mathbb{E}_{q(\boldsymbol{c}, \boldsymbol{\gamma})}\Big[\log p(\boldsymbol{l}_1 | \theta) + \sum_{t=2}^{T} \log p(\boldsymbol{l}_t | \boldsymbol{l}_{t-1}, \boldsymbol{c}_t, \theta) + \sum_{t=1}^{T} \log p(\boldsymbol{y}_t | \boldsymbol{l}_t + \boldsymbol{b}_t, \theta)\Big] + C.
\end{aligned} \tag{12}$$

Conditioned on estimates of $\boldsymbol{b}_{1:T}$ and samples of $\boldsymbol{c}_{1:T}$, the factor graph of equation (12) corresponds exactly to a time-varing Linear Gaussian State Space Model. Thus we can leverage the efficient inference algorithms such as the Kalman filter and RTS smoother when computing the marginals of the variational distribution of $\boldsymbol{l}_{1:T}$.

C Generating Synthetic Examples

C.1 Noisy NASCAR

NASCAR data is generated by partitioning the two-dimensional state space into four regions according to the rules,

$$Z(\boldsymbol{x}) = \begin{cases} 1, & x_1 > 1 \\ 2, & x_1 < -1 \\ 3, & -1 \le x_1 \le 1, x_2 \ge 0 \\ 4, & -1 \le x_1 \le 1, x_2 < 0, \end{cases}$$

where $Z(\boldsymbol{x})$ is the ground truth switching state function that depends on the particular location $\boldsymbol{x}$. The ground truth dynamics matrices are defined as,

$$\boldsymbol{A}(\boldsymbol{x}) = \begin{cases} \begin{bmatrix} 0 & 0.1 \\ -0.1 & 0 \end{bmatrix}, & \text{when } Z(\boldsymbol{x}) = 1 \text{ or } 2 \\ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, & \text{when } Z(\boldsymbol{x}) = 3 \text{ or } 4, \end{cases}$$

and ground truth offsets are defined as,

$$\boldsymbol{b}(\boldsymbol{x}) = \begin{cases} \begin{bmatrix} 0 & 0.005 \end{bmatrix}^\top, & \text{when } Z(\boldsymbol{x}) = 1 \\ \begin{bmatrix} 0 & -0.005 \end{bmatrix}^\top, & \text{when } Z(\boldsymbol{x}) = 2 \\ \begin{bmatrix} 0.1 & 0 \end{bmatrix}^\top, & \text{when } Z(\boldsymbol{x}) = 3 \\ \begin{bmatrix} -0.1 & 0 \end{bmatrix}^\top, & \text{when } Z(\boldsymbol{x}) = 4. \end{cases}$$

Given the current location in state space $\boldsymbol{x}_t$, we can transition to the next point using the continuous time dynamics equation

$$\boldsymbol{x}_t = \text{expm}(\tau \boldsymbol{A}_{Z(\boldsymbol{x}_t)}) \boldsymbol{x}_{t-1} + \tau \boldsymbol{b}_{Z(\boldsymbol{x}_t)} + \boldsymbol{\nu}_t,$$

where each entry of the process noise is sampled from $\nu_{t,i} \sim \mathcal{N}(0, 10^{-4})$. To modulate the speed of the system, we uniformly sample a speed constant $\tau \in [0.1, 1]$, which is applied throughout each segment of the track. We use the continuous time formulation over the discrete-time formulation to ensure that changes to the speed do not distort the shape of the original system's state space. To generate noisy observations, we construct a linear emissions matrix with random variables such that each entry is given by $D_{i,j} \sim \mathcal{N}(0, 1)$.

C.2 Ramping Lorenz

In order to modulate the speed of the Lorenz system, we adjust the evaluation time points of an ODE integrator, specifically Runge-Kutta of the order 5(4) (RK54) as implemented in scipy's `solve_ivp` [9]. Ramping activity is generated randomly with the following procedure,

1. Uniformly sample an evaluation interval length $\tau \in [0.25, 1.5]$.

2. Construct a vector $\widetilde{T}$ that consists n evenly spaced numbers over the interval $[0, \tau]$. In our experiments, we set n to be 100.

3. Perform the transformation $\exp(\widetilde{T}) - 1$ to obtain a vector of ramped evaluation times.

4. Plug in the transformed evaluation times into the RK45 Solver to obtain latent trajectories.

Similar to the NASCAR experiment, we generate noisy observation from a randomly constructed linear emissions matrix such that each entry is given by $D_{i,j} \sim \mathcal{N}(0, 1)$.

C.3 Simulated Monkey Reaching Task

Our dataset is constructed from publicly available data and code from the center-out reach task in [22, 8]. We obtain latent factors from spiking networks that are trained to reproduce empirically measured EMG signals, given a 3-dimensional input that specifies the go input and the reach angle. In our experiments, these factors are considered ground truth. Our trained factor-based spiking network then generates spiking activity for 1200 neurons. Synaptic currents are used as inputs into the Weighted Sum of synaptic currents LFP proxy method (WSLFP) [29], as implemented in the `wslfp` Python package [17, 16]. As WSLFP is a function of the relative location of neurons and electrodes, we place neurons randomly within a 5 mm by 10 mm by 1 mm region and electrodes in a grid centered in this region. The result is a multi-channel LFP dataset with nonlinear dynamics and measurements characteristic of systems neuroscience.
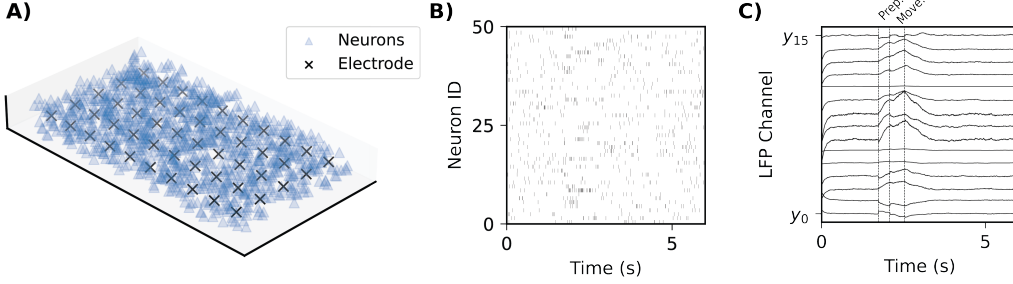

Figure 5: **Empirically-Derived Reach Experiment.** (**A**) 1,200 neurons are randomly placed into a 5 mm by 10 mm by 1 mm region. Electrodes are placed in a grid centered in this region (**B**) Spiking activity for a subset of neurons in an example trial produced from a factor-based spiking network. (**C**) First 15 channels in a simulated multi-channel LFP recording. Preparatory and Movement phases are marked by the dotted lines.

D Evaluation Metrics

D.1 Multi-step Inference

The multi-step inference performance is computed with the following R-squared metric,

$$R_k^2 = 1 - \frac{\sum_{t=0}^{T-k} \|\boldsymbol{y}_{t+k} - \widehat{\boldsymbol{y}}_{t+k}\|_2^2}{\sum_{t=0}^{T-k} \|\boldsymbol{y}_{t+k} - \bar{\boldsymbol{y}}\|_2^2}, \tag{13}$$

where k is the number of steps from the initial condition, $\bar{\boldsymbol{y}}$ is the mean estimator for each trajectory and $\widehat{\boldsymbol{y}}_{t+k}$ is the model prediction after applying the inferred dynamics for k steps. When testing, model parameters such as the dynamics and observation matrices are frozen, while specific latent variables are estimated based on the held-out data. In Table 1, we show results for $k = 100$.

D.2 Inferred Dynamics Error

We measure the accuracy of the latent dynamics with the mean squared error (MSE) of the inferred speed, defined as,

$$\text{MSE}_{\text{speed}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \|\dot{\boldsymbol{x}}_t - \boldsymbol{U}\widehat{\boldsymbol{x}}_t\|_2^2, \tag{14}$$

where the true speed $\dot{\boldsymbol{x}}_t = \boldsymbol{x}_{t+1} - \boldsymbol{x}_t$ is computed from the denoised ground truth latent state, and the predicted speed $\widehat{\dot{\boldsymbol{x}}}_t = \widehat{\boldsymbol{x}}_{t+1} - \widehat{\boldsymbol{x}}_t$ is computed using the model's 1-step prediction. Since latent trajectories are only identifiable up to a linear transformation, we align the inferred trajectories with the true trajectories using a least squares fit before computing this score. More specifically, we

17

find the optimal linear transformation $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ between the estimated and true states across all trajectories by solving,

$$\widehat{\boldsymbol{U}} = \arg\min_{\boldsymbol{U}} \frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{U}\widehat{\boldsymbol{x}}_t\|. \tag{15}$$

D.3 Inferred Latent State Space Error

Similarly, we measure the accuracy of the latent state space by computing the MSE after a linear alignment between trajectories from the inferred and true state space. We use this metric only for the reaching example, since the true observation function is a complex nonlinear function,

$$\text{MSE}_{\text{state}} = \frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{x_t} - \boldsymbol{U}\widehat{\boldsymbol{x}}_t\|_2^2. \tag{16}$$

The linear alignment $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ between the estimated and true states across all trajectories is computed by solving the least squares problem in equation (15).

D.4 Inferred switching rate error

Evaluating the accuracy of the switching behavior is a more difficult task. In fact, developing a procedure that matches predicted switch times with true switch times can lead to a complicated optimal transport procedure. To simplify the evaluation of switching times, we marginalize over time, and compare only the MSE of the switch rate defined as,

$$\text{MSE}_{\text{switch}} = \frac{1}{m} \sum_{i=1}^{m} \|r_i - \widehat{r}_i\|_2^2, \tag{17}$$

where m is the number of trials, r_i is the true switch rate for the ith trajectory, and $\widehat{r}_i = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\{z_t \neq z_{t-1}\}$ is the predicted switch rate. Intuitively, $\widehat{r}_i$ is the number of times that the state or dominant DO changes between consecutive time points normalized by the length of the interval T. In switching models, switch events are defined as a time point where the current inferred dynamical state differs from the state in the previous time step. Similarly in decomposed models, switch events are defined as time points where the active set of DOs change from the previous time step.

In the NASCAR example, r_i is defined with the number of transitions between ground truth segments. In the Lorenz example, r_i is defined by the number of times that the trajectory switches between the two lobes in addition to the number of ramping periods.

D.5 Reaching Classification Accuracy

We quantitatively evaluate the reaching experiment with a classification task. Here, we want to determine whether the learned systems can be used to distinguish between different reach directions. Recall that switched models infer a switching variable for each time point where $z_t \in \{1, \ldots, K\}$ while decomposed models infer a coefficient vector $\boldsymbol{c}_t \in \mathbb{R}^K$. Rather than viewing z_t as an index, we can equivalently view it as a one-hot encoded vector $z_t \in \{0, 1\}^K$ which describes whether a particular switching state is active at any given time. This matches the dimensionality of the variables in both switched and decomposed systems.

For simplicity, we focus on linear logistic regression classifiers in our experiment. If we let the inputs be z_t and c_t directly, then our classifiers quickly overfits since there are many more input features than trials. Specifically, the number of features scales linearly with the number of time points and systems $\mathcal{O}(TK)$. Instead, we marginalize over time and compute features from the estimated latent variables by averaging state activity over time. In switched models, this is the average one-hot encoding value over time. Similarly, this is the average coefficient value in decomposed models. However, for each dynamical state, we compute separate features for positive and negative coefficient values to prevent interference between them. In this setup, the input (feature) dimensionality scales according to $\mathcal{O}(K)$ while the output dimensionality of the linear classifiers are the reaching directions. For all

classifiers, we perform a grid search over the values $\{10^i\}_{i=-4}^{4}$ to identify an appropriate amount of L2 regularization. Top-k accuracies are a standard metric in machine learning [21, 3] and computed using the estimated class probabilities from the logistic regression classifier.

E Additional Results

E.1 Synthetic Dynamical Systems

Figure 6A demonstrates that our inference procedure converges to a local optimium while Figure 6B shows a full sweep of the multi-step inference metric. Tables 1 in the main paper reports the final value. For completeness, we include Tables 3 and 4 which reports the means across 5 seeds of each model, and includes the standard deviations in parenthesis.
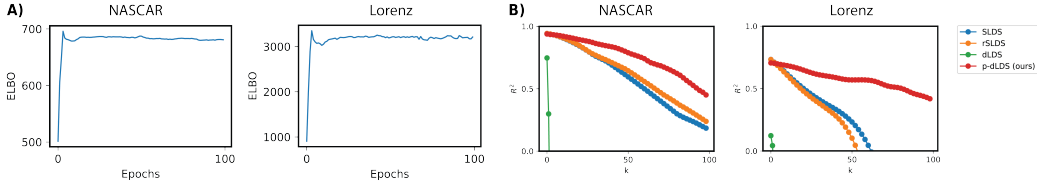

Figure 6: **(A)** ELBO converges in both synthetic dynamical systems. **(B)** Multi-step inference where k represents the number of steps. Tables 3 and 4 report the final values.

Table 3: Metrics for NASCAR. Bold means best performance. ($\uparrow$) indicates higher score is better while ($\downarrow$) indicates that lower is better. ✗ indicates that value diverged towards $-\infty$. All MSE values are $\times 10^{-3}$ while R^2 values are not scaled. We report means across 5 seeds and include standard deviation in parenthesis.

Model	Speed MSE ($\downarrow$)	Switch MSE ($\downarrow$)	100-step R^2 ($\uparrow$)
SLDS	0.0995 (0.021)	12.89 (1.30)	0.184 (0.024)
rSLDS	0.1065 (0.024)	13.17 (2.84)	0.238 (0.022)
dLDS	123.19 (23.13)	13.28 (5.31)	✗
p-dLDS (ours)	**0.033** (0.009)	**7.34** (3.40)	**0.450** (0.027)

Table 4: Metrics for Lorenz. Bold means best performance. ($\uparrow$) indicates higher score is better while ($\downarrow$) indicates that lower is better. ✗ indicates that value diverged towards $-\infty$. We report means across 5 seeds and include standard deviation in parenthesis.

Model	Speed MSE ($\downarrow$)	Switch MSE ($\downarrow$)	100-step R^2 ($\uparrow$)
SLDS	0.431 (0.233)	0.0204 (0.007)	-3.47 (1.052)
rSLDS	0.304 (0.040)	0.0208 (0.004)	-11.54 (1.353)
dLDS	1.123 (0.089)	0.1529 (0.070)	✗
p-dLDS (ours)	**0.141** (0.015)	**0.0137** (0.014)	**0.418** (0.079)

E.2 Reaching Task

For each model, we visualize the trial-averaged dynamic regime activity of each reach direction (Fig. 7). In SLDS, this is visualized by considering the discrete states as a one hot vector over time. When a dynamic regime is active, that state will have a value of 1 while the unactive states will have a value of 0. Thus the trial averaged value of each state must have a value in the interval $[0, 1]$. In dLDS, we plot the inferred coefficient value without any modification.

Although SLDS correctly identifies preparatory and movement phases using states 4 and 3 respectively, it fails to differentiate dynamics occurring outside of these expected phases, incorrectly grouping

19

unrelated regions together. Furthermore, the discrete formulation produces very similar activity patterns across all reach angles, obscuring any differences that are present. In dLDS, we observe that the features change smoothly and cyclically with the reach angle. However, the dynamic operator activity do not localize to the preparatory and movement phases due to a limited inference procedure.

Table 5: Inference performance for the reaching experiment (see Figure 3) on a held-out test set. Top-1 and Top-3 accuracies are obtained by predicting reach directions from latent variable features using linear classifiers. State and Dynamics MSE are computed with respect to true latent variables. We report standard deviations in parenthesis across 5 seeds.

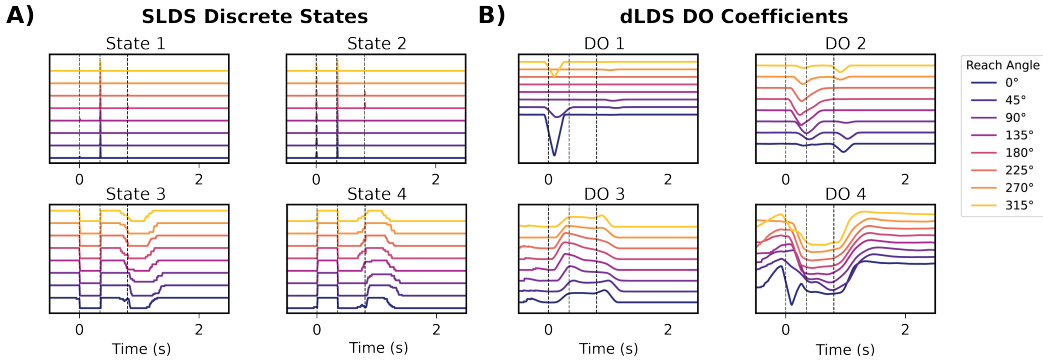| Model	Top-1 Acc.	Top-3 Acc.	State MSE ($\times 10^{-1}$)	Dynamics MSE ($\times 10^{-2}$)
SLDS	38.46 (2.84)	57.69 (7.53)	0.5289 (0.13)	0.3942 (0.23)
rSLDS	12.82 (3.05)	32.05 (8.31)	0.5503 (0.23)	292.41 (13.96)
dLDS	10.25 (5.97)	39.74 (10.29)	0.6742 (0.52)	35.680 (5.76)
pdLDS (ours)	**42.31** (3.50)	**70.51** (6.45)	**0.4061** (0.38)	**0.0567** (0.04)

Figure 7: Trial-averaged activity for **(A)** SLDS discrete states and **(B)** dLDS DO coefficients for each reach angle. The preparatory and movement phases occur between the dashed lines similar to Figure 3. Time 0 represents the onset of the stimulus.

F   Experimental Setup

F.1   Hyperparameter Settings

For switching models, we rely on the `ssm` package which allows for efficient Bayesian inference for a variety of state space models [25]. We set the variational posterior to `structured_meanfield`, and the fitting procedure to `laplace_em` as recommended by the developers. Additionally, we set the distributional form of the dynamics and emissions matrices to Gaussian.

The hyperparameters of dLDS primarily consists of the lagrange multipliers in the BPDN-DF objective including $\lambda_0, \lambda_1, \lambda_2$. We find the optimal value of these hyperparameters using a random search with a fixed budget of 1000 evaluations. For each hyperparameter, we uniformly sample over the log of the interval $[10^{-3}, 10^3]$ and evaluate it against the BPDN-DF objective. For the NASCAR experiment, we found that $\lambda_0 = 1.044, \lambda_1 = 0.254$, and $\lambda_2 = 0.023$ resulted in the best performance. For the Lorenz experiment, we found that $\lambda_0 = 0.628, \lambda_1 = 2.010$, and $\lambda_2 = 0.0124$ yielded the best performance.

For p-dLDS, the relevant hyperparameters consists of the SBL-DF dynamics tradeoff ξ, and the offset window size S. We use random search with a budget of 1000 samples to determine the values of S and ξ and fit a separate model for each set of hyperparameters. In the NASCAR experiment, we isolate the effect of the probabilistic inference procedure by setting $S = T$, removing the influence of the time-varying offset term. For ξ, we perform a random search by uniformly sample over the log of the interval $[10^{-3}, 10^3]$ and found that $\xi = 0.945$ was optimal. For the Lorenz experiment, we also optimize for the window size S by uniformly sample a discrete index on the interval $\{2, \ldots, T\}$.

20

For the Lorenz experiment, the optimized hyperparameters are $S = 85$ and $\xi = 8.928$. For the real dataset, the optimal offset is $S = 76$ which is smaller than the timescale of p-dLDS coefficient switching (around 150 time points), suggesting that the same DO dynamics may persist even as the fixed points of the system fluctuates throughout the experiment.

F.2 Hardware Specification

We perform hyperparameter sweep on our institution's HPC cluster using small-scale CPU resources which consists of Dual Intel Xeon Gold 6226 CPUs. Once hyperparameters have been optimized, it is possible to run each experiment within approximately 2 hours on the 2020 edition of the M1 Macbook Pro.

G   Description of Clinical Neurophysiology Data

Data was collected as part of a study investigating deep brain stimulation for treatment-resistant depression (TRD). The study is pre-registered in ClinicalTrials.gov (identifier NCT04106466). The study protocol was approved by the IRB (identifier IRB00066843). Informed consent was obtained from participants before participation in the trial. Patients receive no monetary compensation, but instead have their DBS electrodes and Summit RC+S IPG device provided free of charge. The analysis focused on LFP signals from a single participant with all personally identifiable information removed.

H   Limitations

While our proposed method demonstrates strong performance in our experiments, there are many limitations. For instance, our approach does not have a strong mechanism for generating future unseen coefficients. Our assumed coefficient transition model is primarily motivated by our desire to obtain smooth coefficients over time. However, we believe that they may be more complex transition models that can both capture persistent activity in challenging systems while also being an accurate forecaster, such as a deep learning based transition model. Another limitation of our approach is that our method assumes smoothness in the latent space. However, we do not explore the possibility of having sparse structure in the latent space which can be easily accomplished in BPDN-DF by adding an L1 penalty over x.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We propose a probabilistic treatment and an extended dynamics formulation in decomposed models (Section 3). We demonstrate that these changes reduce estimation errors and finds coherent structure where previous models fail in many challenging synthetic examples, and a noisy real-world example (Section 4).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: A limitations section is provided in Appendix H due to space constraints.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We provide proofs of our lemmas in B.1 and B.2.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide synthetic data details in Appendix C, metric definitions in Appendix D, and hyperparameter details in Appendix F.1. Moreover, we release our code with our submission.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide functions to generate the synthetic datasets from our experiments. Unfortunately, we are unable to release the neurophysiological dataset due to request from our collaborators.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We specify that all datasets are split 50:50 for train and test in the main paper. Moreover, we provide optimizer and hyperparameter settings in Appendix F.1.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We provide standard deviations in Appendix E.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

    Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

    Answer: [Yes]

    Justification: We provide details about hardware in Appendix F.2.

    Guidelines:

    - The answer NA means that the paper does not include experiments.
    - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
    - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
    - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

    Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

    Answer: [Yes]

    Justification: Most of our datasets are synthetically generated or derived from publicly available sources, which we refer to throughout the paper. For our real-world experiment, we include the IRB identifier.

    Guidelines:

    - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
    - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
    - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: Our proposed model aims to reveal coherent patterns in time-series data, serving as a scientific tool similar to PCA. As a probabilistic technique not designed for content generation or automated decision-making, we do not anticipate direct societal impacts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: Our method does not generate synthetic content that is at risk of abuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification:

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.

26

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We include details on where we obtained our data (either synthetically generated, or part of an existing dataset) in the main text and appendix. We also point to the commonly package used for SLDS and rSLDS in the appendix.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: We provide the IRB identifier in Appendix G which includes the informed consent form with experiment details.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: We include the IRB identifier number in Appendix G, but removed any information related to the institution conducting the experiments.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.