IDENTIFYING NEURAL DYNAMICS USING INTERVENTIONAL STATE SPACE MODELS

FLATIRON INSTITUTE

UNIVERSITY OF MICHIGAN

Amin Nejatbakhsh[1], Yixn Wang[2]

[1] Center for Computation Neuroscience, Flatiron Institute; [2] University of Michigan, Michigan

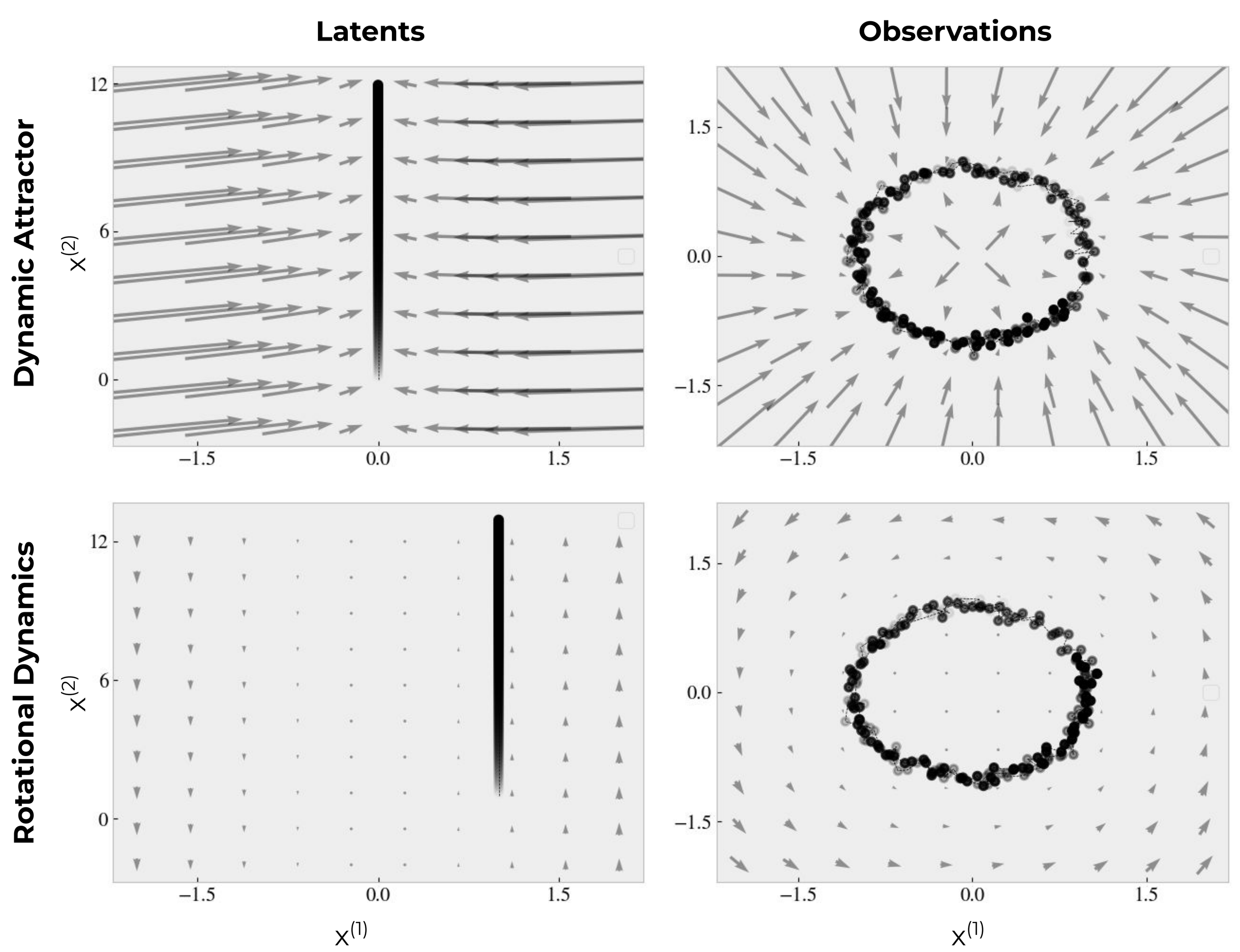
ICML 2025
Poster #W-417

Introduction: Identifying Dynamics

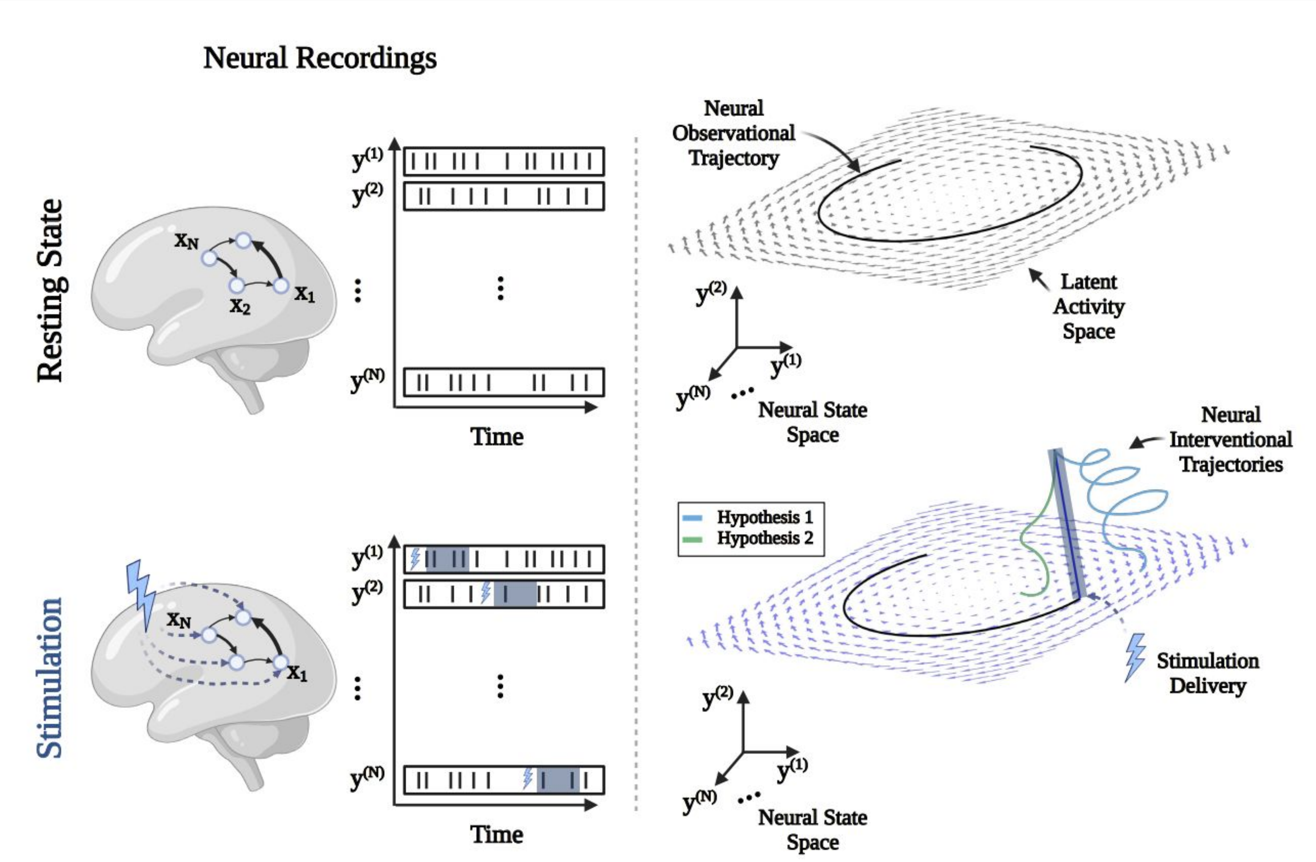
Questions

- Which dynamical system model generated my data?
- Is motor cortex using continuous attractors or not?
- What are the latent variables underlying the observations?

The two models below are indistinguishable from the observational data alone



Approach: Interventional Models



Can we use interventional data to identify the dynamics?

Intuition

- Interventions kick the state of the system outside of its attractor manifold, thereby allowing for the exploration of the state space and collecting more information about the dynamics
- However, interventional data alone is not sufficient for identification, we also need **interventional models** that properly leverage the interventional data

	Dynamics	Observations
SSM	$x_{t+1} = Ax_t + Bu_t + \epsilon_t,$	$y_t \sim P(y_t \mid f_\theta(x_t)).$
iSSM	$x_{t+1} = 1\{Bu_t = 0\} \otimes Ax_t + Bu_t + \epsilon_t,$	$y_t \sim P(y_t \mid f_\theta(x_t)).$

Theoretical Results: iSSM is Identifiable

Interventional State Space Models

Latents $\boldsymbol{x}_{t+1} = \mathbf{1}\{\boldsymbol{B}\boldsymbol{u}_t = 0\} \otimes \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{\epsilon}_t$ **Observations** $\boldsymbol{y}_t \sim P(\boldsymbol{y}_t \mid f_{\boldsymbol{\theta}}(\boldsymbol{x}_t))$.

- $\boldsymbol{u}_t \in \mathbb{R}^M$ interventional input to individual channels at time t.
- $\boldsymbol{y}_t \in \mathbb{R}^N$ neural responses at time t, e.g. N-vector that concatenates the spike counts or calcium activities of all neurons.
- $\boldsymbol{x}_t \in \mathbb{R}^D$ D-dimensional time-dependent latent variable.
- $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q})$ and $\otimes$ denotes element-wise multiplication; $f_{\boldsymbol{\theta}}(.)$ generic nonlinear function mapping latents to observations.
- $\boldsymbol{A} \in \mathbb{R}^{D\times D}$ captures latent dynamics; $\boldsymbol{B} \in \mathbb{R}^{D\times M}$ captures the effect of neural perturbations on latent dynamics.
- If the intervention $\boldsymbol{u}_t$ is zero, the model follows observational dynamics.
- In the presence of an intervention, the model decouples the intervened node from its parents.

Identification Assumptions

- **Assumption 1 (Observation model).** The function $P(\boldsymbol{y}_t \mid \boldsymbol{z}_t)$, where $\boldsymbol{z}_t = f_{\boldsymbol{\theta}}(\boldsymbol{x}_t)$, is **bounded complete** in $\boldsymbol{y}_t$.
  - [e.g. including exponential families, location-scale families, and nonparametric regression models]
- **Assumption 2 (Mixing function).** The mixing function $f_{\boldsymbol{\theta}}(.)$ is piecewise linear, continuous, and injective.
  - [e.g. including (deep) ReLU networks]
- **Assumption 3 (Faithfulness).** There does not exist a non-zero vector $\boldsymbol{V}$ such that $Cov(\boldsymbol{V}^\top \boldsymbol{x}_{t+1}, \boldsymbol{V}^\top \boldsymbol{x}_t) = 0, \forall t$
  - [Loosely, each latent dimension has at least one (non-trivial) causal parent from the previous timestep]

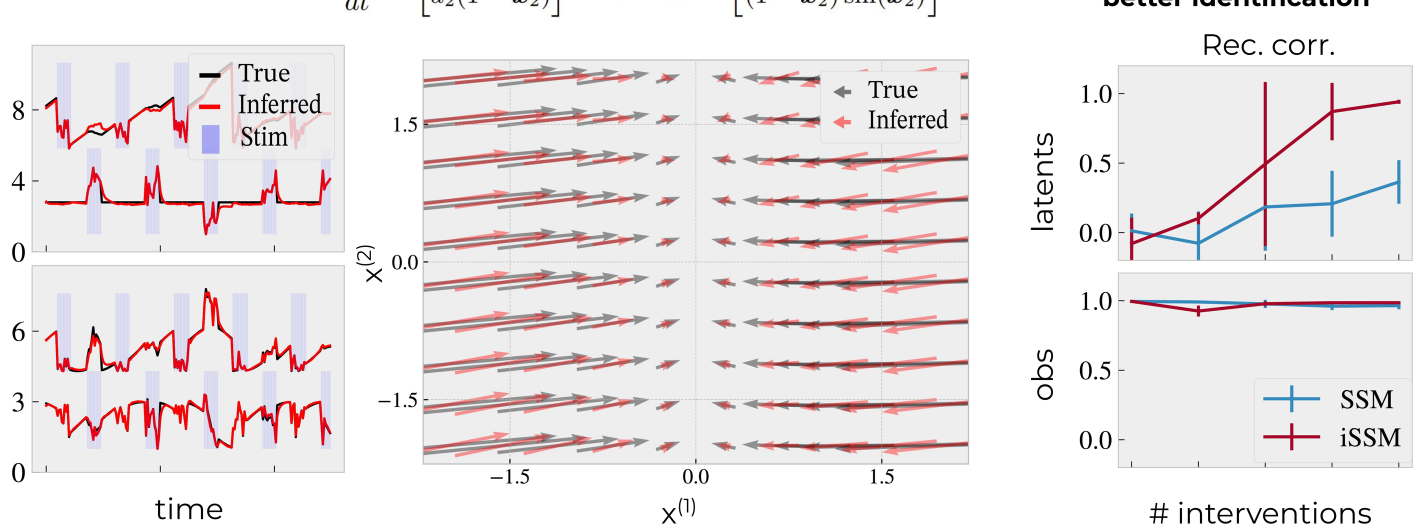
Identifiability Guarantees

- **Theorem (Block identifiability of iSSM and generalization to unseen interventions).**
  - Under Assumptions 1-3, the latent dynamics $\boldsymbol{A}$ and the mixing function of $f_{\boldsymbol{\theta}}(.)$ can be **block-identified** up to permutation, and shifting and scaling.
  - Given a single intervention trial, one can **separate out** the intervened latents from the un-intervened ones.
  - Can extrapolate to **novel, unseen interventions** as long as they only touch upon already separated latents.
- **Corollary (Identifiability of iSSM under sufficiently diverse interventions).**
  - If the interventions satisfy the **unordered pairs condition** (Hyttinen et al., 2013),
  - then the iSSM is **identifiable** up to permutation, along with coordinate-wise scaling and shifting.
  - The distribution under **any novel interventions** is also **identifiable**.

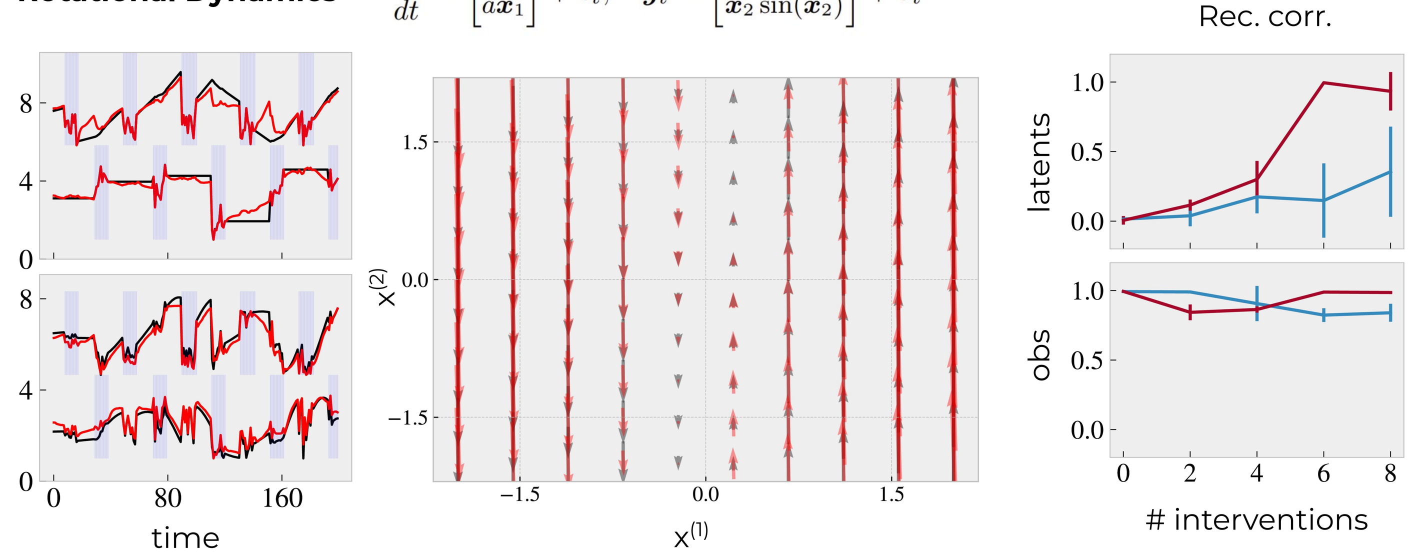
Results: Models of Motor Cortex

Dynamic Attractor $\frac{dx}{dt} = \begin{bmatrix} a_1 x_1 \\ a_2(1-x_2) \end{bmatrix} + \epsilon_t, \quad y_t = \begin{bmatrix} (1-x_1)\cos(x_2) \\ (1-x_2)\sin(x_2) \end{bmatrix} + \delta_t$

More interventions lead to better identification



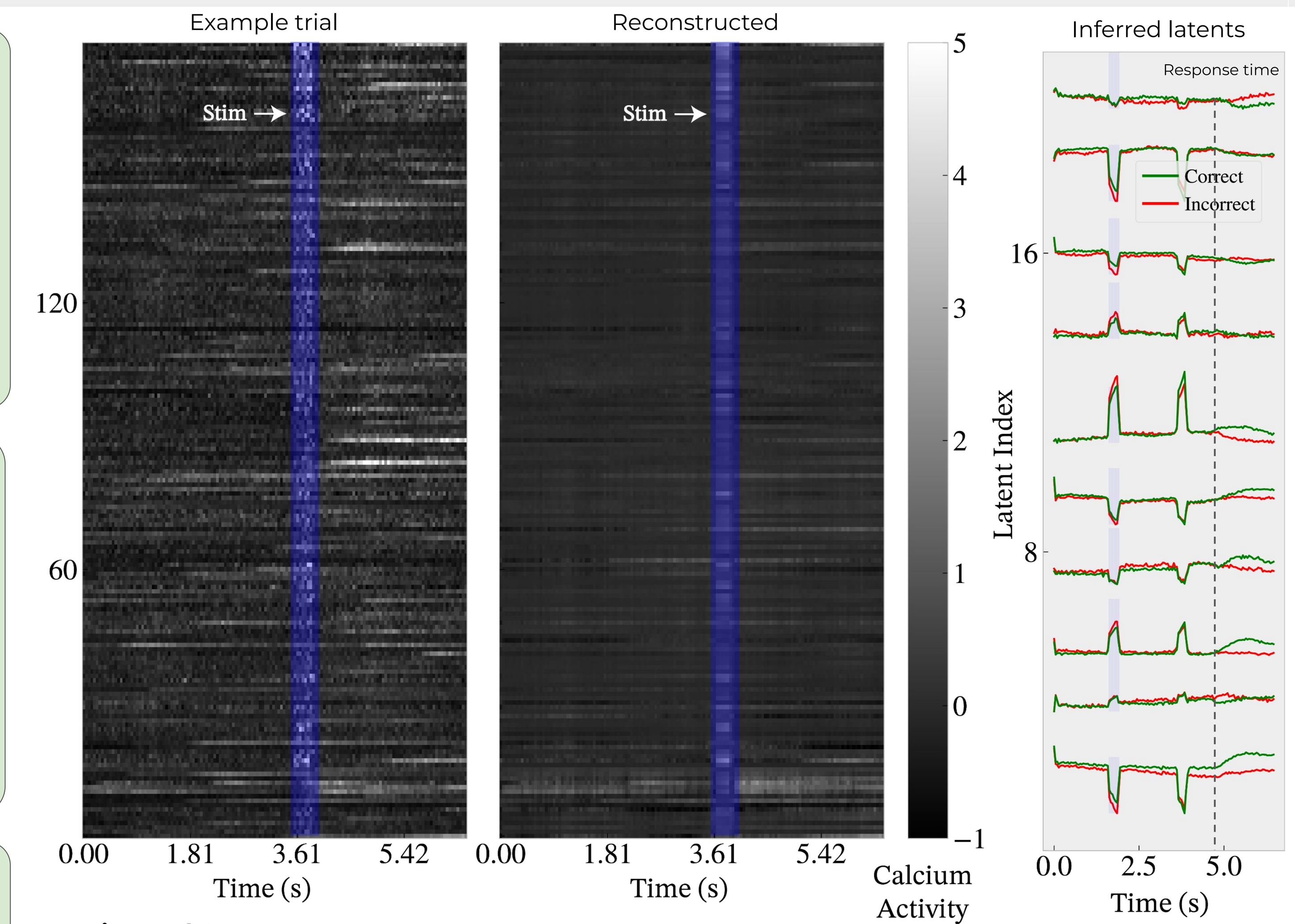
Rotational Dynamics $\frac{dx}{dt} = \begin{bmatrix} 0 \\ a x_1 \end{bmatrix} + \epsilon_t, \quad y_t = \begin{bmatrix} x_1 \cos(x_2) \\ x_2 \sin(x_2) \end{bmatrix} + \delta_t$



* For higher dimensional results check the paper

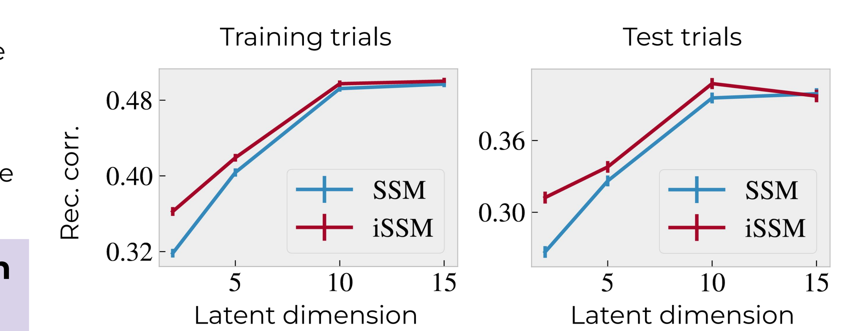
iSSM identifies the latents and the underlying flow field

Results: Optogenetics in Mouse ALM



Experimental Setup

- ALM neurons in mice recorded in response to short-term memory task
- Targeted optogenetic stimulation was delivered during the delay period
- Photostimulation was delivered right at the start of the delay period, or after 1 or 2 sec



iSSM latents distinguish between correct and incorrect trials

Results: Models of Working Memory

Experimental Setup

- Low-rank (LR) vs. functionally feed-forward (FF) are proposed as models of persistent activity in working memory
- Data was generated from LR and FF models in response to stimulation
- iSSM and SSM were fit using linear dynamics and linear observation models

Noisy linear dynamics

$$\frac{d\boldsymbol{x}}{dt} = \mathbf{A}\boldsymbol{x} + d\boldsymbol{w}_t$$

LR connectivity

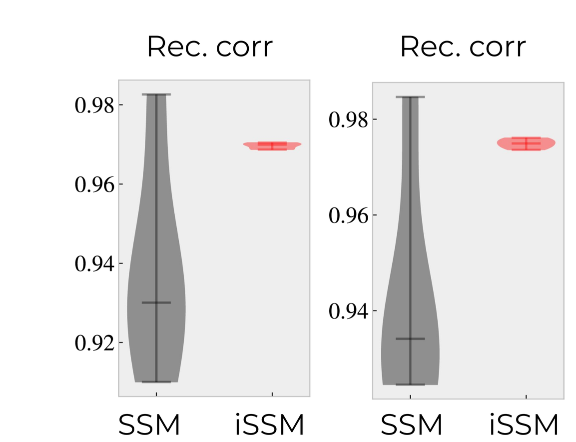
$$\mathbf{T} = \mathbf{I}_D^{(1)} + 0.5\left[\mathbf{0}_{D\times(D-1)} \quad \mathbf{1}_{D\times 1}\right]$$
$$\mathbf{A} = -\mathbf{I}_D + \mathbf{O}\mathbf{T}\mathbf{O}^\top$$

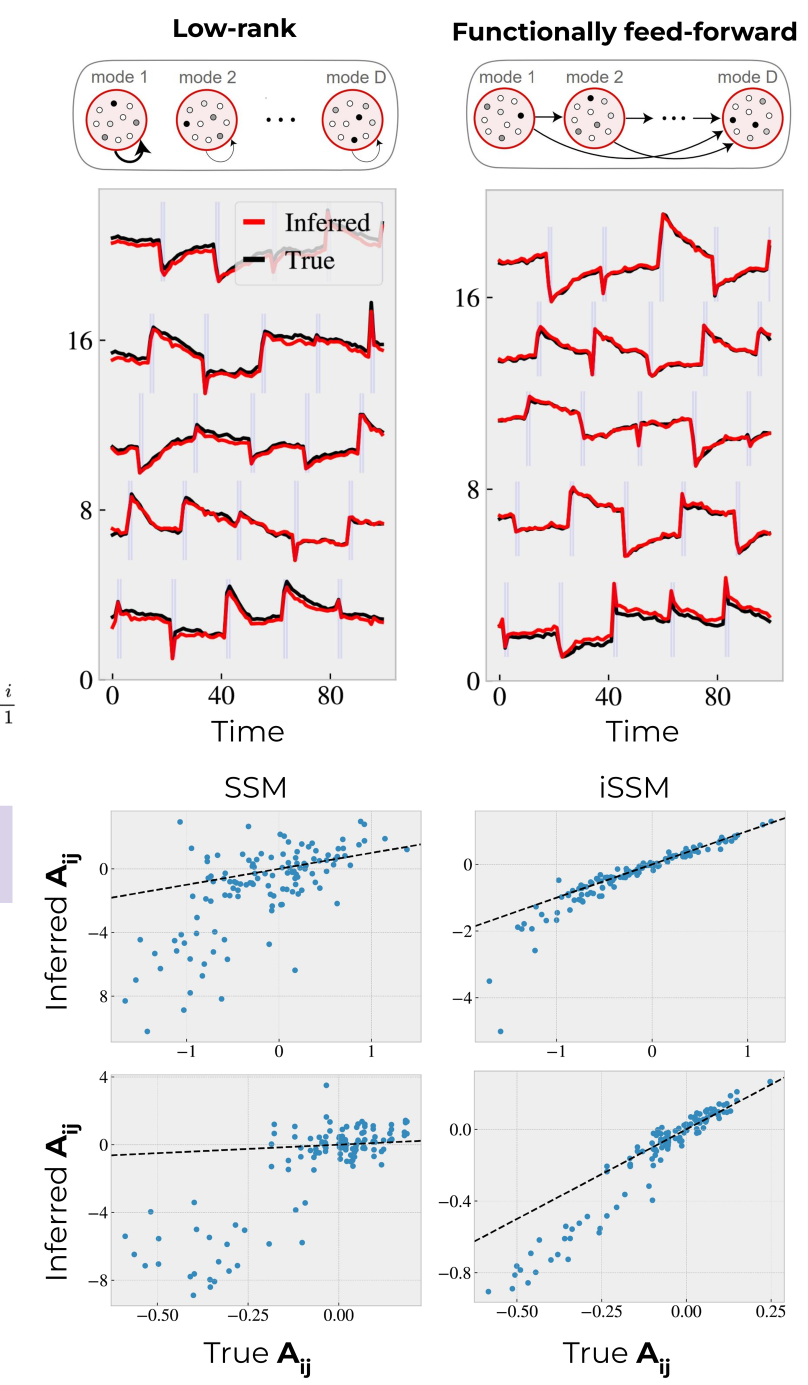
$\mathbf{O}$ is a randomly generated orthogonal matrix

FF connectivity

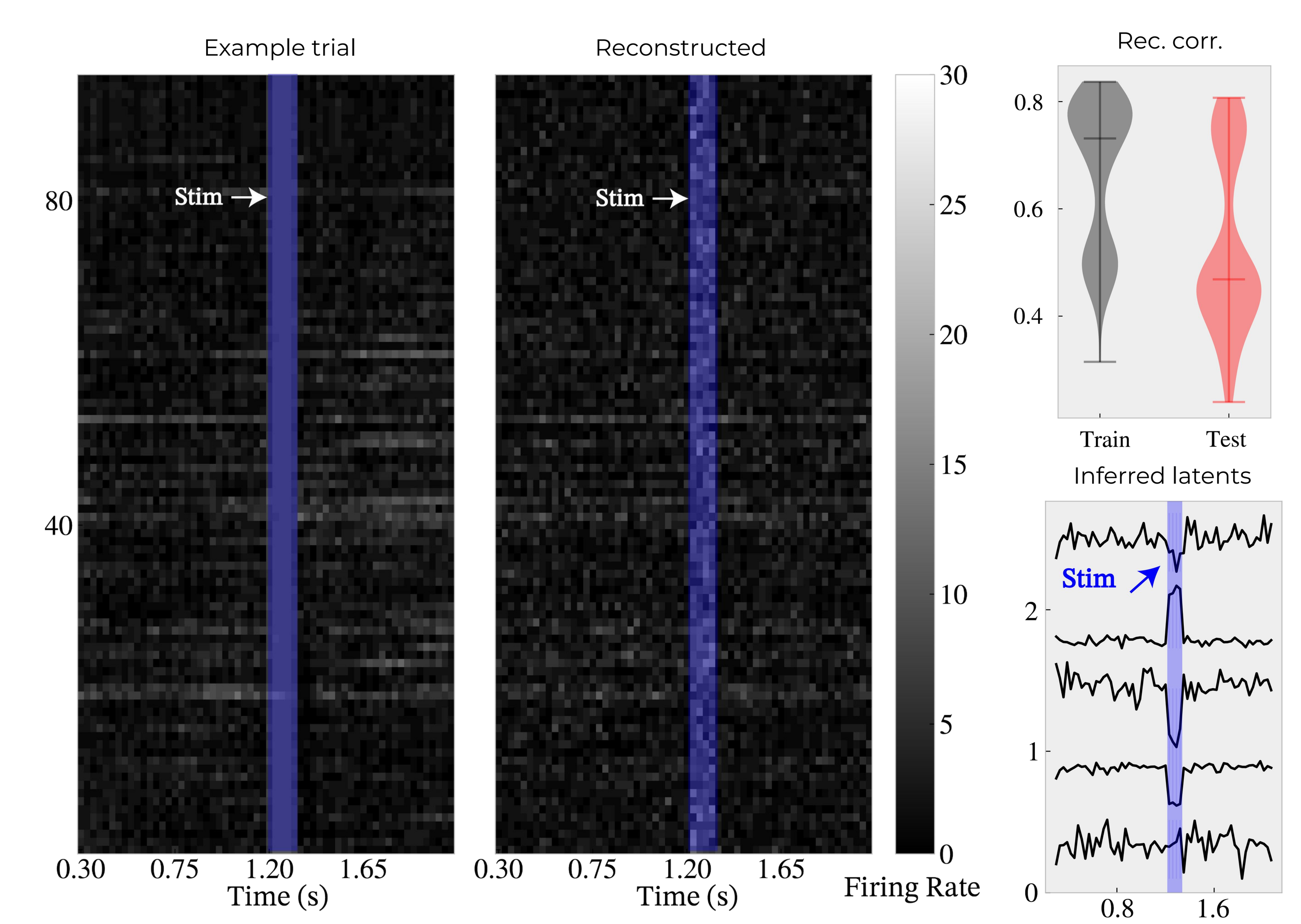
$$\mathbf{\Lambda} = \mathrm{diag}(\exp(\ell_1), \exp(\ell_2), \ldots, \exp(\ell_D)), \quad \text{where } \ell_i = \frac{D-i}{D-1}$$
$$\mathbf{A} = -\mathbf{I}_D + \mathbf{O}\mathbf{\Lambda}\mathbf{O}^\top$$

iSSM identifies the latents and the connectivity matrix



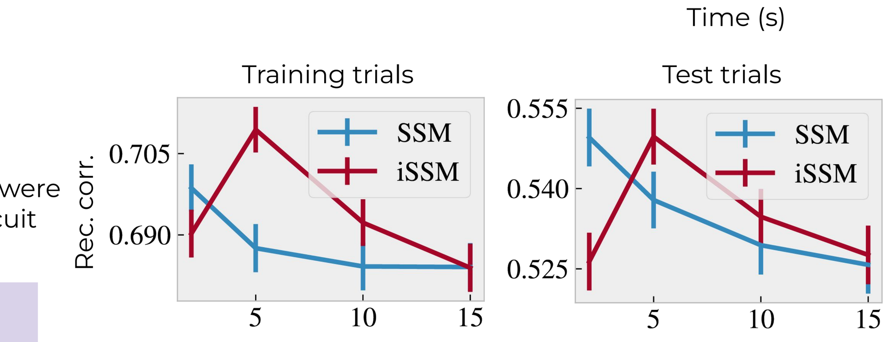


Results: Micro-Stimulation in Primate dlPFC



Experimental Setup

- dlPFC neurons were recorded in macaque monkeys during quiet wakefulness (resting).
- The electrode
- array included 96 electrodes that were also used for delivering micro-circuit electrical stimulations



iSSM better generalizes to test interventions

Summary & References

Summary

1. Here we proposed **iSSM**, a framework for joint modeling of observational and interventional data.
2. We provided theoretical results showing that the iSSM model, when fitted on interventional data, leads to the **identifiability of latents as well as dynamics and emissions**.
3. We showed in the **models of motor cortex** and **working memory** with **linear dynamics** and **linear or nonlinear emissions** iSSM leads to model identifiability.
4. We showed an application of iSSM to **calcium recordings from the mouse ALM** region with **targeted photostimulation** delivered by channels that targeted groups of neurons.
5. We showed an application of iSSM to **electrophysiological recordings** from the **macaque monkey prefrontal cortex with micro-stimulation** delivered by the same recording electrodes.

Future Directions

(1) Interventional models with nonlinear dynamics (2) Modeling interventions applied to neurons as opposed to latents (3) Better inference algorithms

GitHub

ICML International Conference On Machine Learning

References

1. Galgali, A., et al. Residual dynamics resolves recurrent contributions to neural computation. Nature Neuroscience, 26(2):326–338, 2023.
2. Qian, W., et al. Partial observation can induce mechanistic mismatches in data-constrained models of neural dynamics. bioRxiv, pp. 2024–05, 2024.
3. Daie, K., et al. Targeted photostimulation uncovers circuit motifs supporting short-term memory. Nature neuroscience, 24(2):259–265, 2021.
4. Nejatbakhsh, A., et al. Predicting the effect of micro-stimulation on macaque prefrontal activity based on spontaneous circuit dynamics. Physical Review Research, 5(4):043211, 2023.