Extracting Governing Equations from Latent Dynamics via Multi-View Contrastive Learning

Paolo Muratore
EPFL
paolo.muratore@epfl.ch

Mackenzie Weygandt Mathis
EPFL
mackenzie.mathis@epfl.ch

Abstract

Identifying latent dynamical systems from noisy, high-dimensional measurements is a central problem at the intersection of representation learning, system identification, and scientific discovery. We present DYSCO (DYnamical Systems via Contrastive Optimisation), a multi-view temporal contrastive learning algorithm that jointly recovers latent trajectories and the governing dynamics from such observations, by leveraging multiple independent noisy views of the same underlying process to disentangle signal from noise. By parameterizing the dynamics in a structured functional basis, our framework further enables symbolic recovery of the governing equations within an affine gauge. We offer theoretical guarantees for strong identification up to an affine indeterminacy, extending prior identifiability results to the realistic setting of noisy nonlinear observations. Empirically, we demonstrate accurate recovery of both latent trajectories and flow fields across a diverse set of dynamical regimes (*e.g*, chaotic, oscillatory, and metastable) under both Gaussian and Poisson observation noise, the latter being particularly relevant for neural data, for which we offer a preliminary demonstration by recovering a known rotational structure in motor-cortical recordings.

1 Introduction

A central challenge in neuroscience and machine learning is recovering the computational laws governing a system from noisy, high-dimensional observations of its activity. This problem is classically articulated in Marr's tri-level framework (Marr, 1982), which distinguishes between: the computational level, *e.g.*, sorting numbers in ascending order, the algorithmic level, *e.g.*, a particular procedure, such as bubble-sort or merge-sort, and the implementation level, *i.e.*, the circuit of transistors switching on and off. While modern data-driven methods have made substantial progress at distilling the implementation level of raw neural activity into abstract latent representations (Schneider et al., 2023; Mathis and Mathis, 2026), the principled recovery of the dynamical laws governing these representations (*i.e.* the complete algorithm description of the system) remains an open challenge (Sussillo and Barak, 2013; Vyas et al., 2020; Versteeg et al., 2025). Bridging this gap is essential for interpretability, scientific discovery and the development of robust, generalizable models.

In many domains, ranging from neural population recordings to complex physical systems, the observable data are high-dimensional, noisy and generated through nonlinear measurement processes. The underlying system, however, is often governed by a low-dimensional latent dynamical process (Mastrogiuseppe and Ostojic, 2018). Recovering both the latent state and the governing dynamics from such observations constitutes a fundamental inverse problem at the intersection of representation learning, system identification, and scientific modeling. A key obstacle is that noisy, nonlinear observations sever the identifiability guarantees that underpin existing approaches: noise corrupts the signal on which both representation learning and symbolic methods rely, and without multiple independent views of the same latent state there is no principled mechanism to separate signal from nuisance. Recent advances in self-supervised and contrastive learning have demonstrated

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

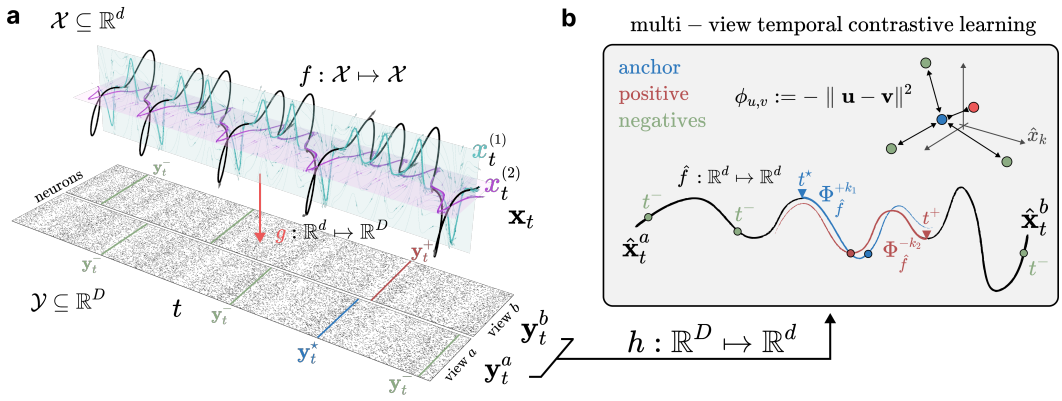


Figure 1: **Graphical overview.** (**a**) We consider a latent dynamical system (top) that evolves according to an unknown transition $f : \mathcal{X} \mapsto \mathcal{X}$. We observe the system via a non-linear mixing-channel $g : \mathbb{R}^d \mapsto \mathbb{R}^D$ that produces noisy high-dimensional observations $\boldsymbol{y}_t^a$. (**b**) We solve the identifiability problem by simultaneously learning an encoder $h : \mathbb{R}^D \mapsto \mathbb{R}^d$ and a symbolic vector field $\hat{f}$ by minimizing the multi-view temporal contrastive loss defined in (9).

that temporal structure provides a powerful inductive bias for extracting meaningful representations from sequential data. However, existing approaches primarily focus on learning useful representations, rather than explicitly identifying the underlying dynamical system that generates the data. In parallel, methods for symbolic regression and dynamical system discovery aim to recover explicit functional forms of the dynamics (Brunton et al., 2016), enabling interpretability and mechanistic insight. Yet, these approaches typically assume direct or partially observed access to the system state (but see (Champion et al., 2019) for a autoencoder approach to latent dynamical systems), limiting their applicability in realistic settings where observations are indirect and noisy. This disconnect highlights a key open problem: *can we simultaneously learn latent representations and recover the governing dynamical laws from high-dimensional, noisy observations, with theoretical guarantees of identifiability?*

In this work, we address this problem by introducing a multi-view temporal contrastive learning framework for the identification of latent dynamical systems. Our approach leverages multiple independent noisy views of the same underlying system to enforce consistency constraints that effectively disentangle signal from noise, while a structured parameterization of the dynamics enables symbolic recovery of the governing equations. Concretely, we consider a setting where a latent dynamical system evolves according to an unknown nonlinear map and is observed through multiple noisy, nonlinear measurement channels. We jointly learn (*i*) an encoder that inverts the observation process and recovers latent states, and (*ii*) a structured dynamics model expressed in a predefined functional basis. The learning signal is provided by a temporal contrastive objective that exploits both time evolution and cross-view consistency. As illustrated in our formulation (see Figure 1), this setup allows the model to align representations across views while respecting the underlying dynamics. Our main contributions are as follows:

- *(i) Multi-view contrastive system identification.* We propose DYSCO, a multi-view extension of dynamical contrastive learning that leverages temporal structure and multiple noisy views to jointly recover latent states and dynamics from high-dimensional observations.
- *(ii) Theoretical identifiability guarantees.* We show that, under asymptotic assumptions, multi-view contrastive learning identifies the latent dynamical system up to an affine transformation, extending prior results to the realistic setting of noisy, nonlinear observations.
- *(iii) Compatibility with symbolic regression.* We show how the affine gauge structure of the recovered representation is compatible with downstream symbolic-regression pipelines, and illustrate this on the Lorenz system.
- *(iv) Empirical validation across dynamical regimes.* We demonstrate accurate recovery of both latent trajectories and flow fields across a diverse set of systems, including chaotic, oscillatory, and metastable dynamics, even under substantial observational noise.

Taken together, our results suggest that contrastive learning, when appropriately structured, provides a principled route toward solving Marr's inverse problem: moving from raw observations at the implementation level to explicit, interpretable descriptions of the underlying computation.

Related Work Recovering latent dynamical systems from high-dimensional observations sits at the intersection of representation learning, system identification, and symbolic regression. The work most closely related to ours is (Laiz et al., 2025), which demonstrates that contrastive objectives can recover both latent representations and nonlinear dynamics up to an affine transformation. Our work builds on this foundation but departs from it in two important ways: first, we explicitly leverage multi-view observations to handle realistic noisy measurement processes, where (Laiz et al., 2025)'s identifiability relies on noiseless observations—a condition that multi-view consistency is specifically designed to relax by forcing the encoder to extract only the shared, noise-free component across views; second, we enforce a symbolic structure on the learned dynamics, moving beyond the switching-linear dynamical system formulation and enabling recovery of interpretable governing equations. We situate this contribution below within the broader literature.

Recent advances in self-supervised learning, and in particular contrastive learning, have demonstrated that temporal structure can provide a powerful signal for extracting latent representations. Methods such as Contrastive Predictive Coding (Oord et al., 2018) and its variants (Schneider et al., 2019; Henaff, 2020) leverage predictive objectives to learn embeddings that capture underlying dynamics. However, these approaches are typically optimized for representation quality rather than explicit recovery of the governing dynamical system. This gap has motivated a line of theoretical work connecting contrastive objectives to identifiable representation learning. In particular, nonlinear ICA frameworks (Hyvarinen and Morioka, 2016; Hyvarinen et al., 2019; Hälvä et al., 2021) show that temporal or auxiliary structure can render latent variables identifiable under suitable conditions (see (Hyvärinen et al., 2023) for a review), and recent results have further linked contrastive learning to inversion of the data-generating process (Zimmermann et al., 2021). Nevertheless, these works focus primarily on recovering latent states, leaving the identification of the underlying dynamics largely unaddressed. More recently, VAE-based techniques such as iLQR-VAE (Schimel et al., 2022) and Brenner et al. (2024) leveraged control-based variational inference and multimodal signals for dynamical reconstruction. DYSCO on the other hand is encoder-only and uses the correspondence between noisy measurements of an aligned latent state to recover the dynamics in a symbolic form.

In neuroscience, related efforts have focused on inferring latent population dynamics from noisy neural recordings. Sequential autoencoder models such as LFADS (Pandarinath et al., 2018b; Kim et al., 2021; Shah et al., 2025) and more recent contrastive approaches such as CEBRA (Schneider et al., 2023) learn low-dimensional representations that capture behaviorally relevant structure. While highly effective empirically, these methods typically either lack identifiability guarantees or do not recover explicit dynamical laws.

The symbolic identification of dynamical systems has been extensively studied in the context of sparse regression methods such as SINDy (Brunton et al., 2016; Klishin et al., 2025), which recovers governing equations by selecting a sparse combination of candidate basis functions. More recent approaches extend this idea using sequence (d'Ascoli et al., 2023) or diffusion-based models (Bastiani et al., 2025) for symbolic regression. These methods, however, typically assume direct (or partial) access to the system state. Closer in spirit, Champion et al. (2019) adapted SINDy within a variational autoencoder framework for discovery of latent dynamics. However this framework lacks any theoretical guarantee and the multi-term loss function requires careful balancing of the relevant terms. In contrast, our approach provides theoretical identifiability guarantees within an affine gauge using a single contrastive objective, with symbolic recovery as a downstream step.

2 Contrastive Learning Identifies Latent Dynamical Systems

Problem definition We study the following non-linear identification problem. We consider a latent (*i.e.*, non-observable) state variable $\boldsymbol{x}_t \in \mathcal{X} \subseteq \mathbb{R}^d$ whose stochastic discrete-time evolution can be described via a dynamical system of the form:

$$\boldsymbol{x}_{t+1} = F(\boldsymbol{x}_t) + C\boldsymbol{u}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $F : \mathcal{X} \mapsto \mathcal{X}$ is a bijective transition map, $\boldsymbol{u}_t \in \mathbb{R}^u$ is an external forcing input, $C \in \mathbb{R}^{d\times u}$ is the input coupling matrix and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^d$ is the latent system noise. We observe this system via the

non-linear observation channel:

$$\boldsymbol{y}_t^a = g(\boldsymbol{x}_t) + \boldsymbol{\xi}_t^a, \tag{2}$$

where $\boldsymbol{y}_t^a \in \mathbb{R}^D$, $D \geq d$ is the observed state, $g : \mathcal{X} \to \mathbb{R}^D$ is a non-linear injective mixing function from latents to observations and $\boldsymbol{\xi}_t^a$ denotes the observation noise assumed independent from the latent state, with a indexing independent realizations of the noise process. Our objective is the following: given a set of observed trajectories $\Upsilon \coloneqq \{\boldsymbol{y}_t^a\}_{t,a}^{T,V}$, with T the total trajectory length and $V \geq 2$ the total number of independent *views* of the system, reconstruct the latent states $\{\boldsymbol{x}_t\}_t^T$ and the dynamics F. In practice, we seek to learn a de-mixing function $h : \mathbb{R}^D \mapsto \mathcal{X}$ and a dynamics model $\widehat{F}$ such that the composition $r \coloneqq h \circ g$ is a trivial function (ideally, the identity) and $\widehat{F}$ recapitulates F. For the continuous-time implementation, we parameterize the transition F as a vector field f and consider the solution operator $\Phi_f^{t+k\Delta,t}(\boldsymbol{x}; \boldsymbol{u}_{[t,t+k\Delta]})$. In the autonomous deterministic case, $F = \Phi_f^{\Delta,0}$ is the sampled transition.

Model definition We follow a similar construction from (Laiz et al., 2025) and employ a contrastive learning framework for solving the identification problem. We introduce an encoder $h_\theta : \mathbb{R}^D \mapsto \mathbb{R}^d$, *i.e.*, a neural network parametrized by θ and a symbolic model for the dynamics $\hat{f}_\Xi : \mathbb{R}^d \mapsto \mathbb{R}^d$, represented as the span of a pre-defined set of library scalar functions $\Xi \coloneqq \{f_1, f_2, \ldots, f_n\} \in \mathbb{R}^n$ (Brunton et al., 2016). Typically in this work, we choose Ξ as the set of all monomials up to a given degree, *e.g.* for a two-dimensional system $\Xi^{(2)} = \left\{1, x, y, x^2, y^2, xy\right\}$ would represent the polynomial basis of degree two. The choice of the symbolic basis Ξ can be used to trade complexity (more terms, different function primitives f_i) for expressivity. The flow-field at a given point $\hat{f}_\Xi(\boldsymbol{x}_t)$ is then expressed as:

$$\hat{f}_\Xi(\boldsymbol{x}_t) = \boldsymbol{\Theta}\Xi(\boldsymbol{x}_t) = \sum_k^n \boldsymbol{\theta}_k f_k(\boldsymbol{x}_t), \tag{3}$$

where $\boldsymbol{\Theta} \in \mathbb{R}^{d \times n}$ is a learnable matrix of basis coefficients and $\Xi(\boldsymbol{x}_t) \in \mathbb{R}^n$ is the vector of scalar functions evaluated at $\boldsymbol{x}_t$. Given a vector field and initial state, we compute the solution using RK4. For compactness, we write the sampled solution operator as:

$$\Phi_f^k(\boldsymbol{x}_t; \boldsymbol{u}) \coloneqq \Phi_f^{t+k\Delta,t}(\boldsymbol{x}_t; \boldsymbol{u}_{[t,t+k\Delta]}), \qquad \Phi_f^0(\boldsymbol{x}_t) = \boldsymbol{x}_t. \tag{4}$$

For the discrete-time model, the corresponding solution rollout operator is the iterate F^k.

Finally, since the core objective of contrastive learning is to model the pairwise similarities of the data, we consider a similarity function $\phi : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$, which we will usually take to be the negative squared euclidean distance between two points. We can now define our model DYSCO as the following composition:

$$\psi_{ab}(\boldsymbol{y}_t^a, \boldsymbol{y}_\tau^b) \coloneqq \phi\left(\Phi_{\hat{f}}^{\pm k_1}\left(h(\boldsymbol{y}_t^a)\right), \Phi_{\hat{f}}^{\mp k_2}\left(h(\boldsymbol{y}_\tau^b)\right)\right) - \alpha\left(h(\boldsymbol{y}_\tau^b)\right), \tag{5}$$

where $k_{1,2} \in \{0, 1, \ldots, \kappa\}$ with κ a model hyperparameter defining the maximal time-integration horizon, $a, b \in V$ are view indices and $\alpha : \mathbb{R}^d \mapsto \mathbb{R}$ is a scalar potential, whose role is to supply the candidate-marginal correction in the optimal contrastive score $\log p(\boldsymbol{z}' \mid \boldsymbol{z}) - \log q(\boldsymbol{z}')$, up to an anchor-only term. Additionally, we have introduced the shorthand notation for backward rollouts:

$$\Phi_f^{-k} \coloneqq (\Phi_f^k)^{-1}, \tag{6}$$

which provide optional additional training signals (our main theorem only uses an active forward one-step constituent, but we found backward rollouts to stabilize training in practise).

Our model can then be viewed as implementing an amortized inference of the full multi-view model defined equivalently as:

$$\Psi_V\left(\boldsymbol{y}_t^{(V)}, \boldsymbol{y}_\tau^{(V)}\right) \coloneqq \phi\left(\Phi_{\hat{f}}^{\pm k_1}\left(H_V(\boldsymbol{y}_t^{(V)})\right), \Phi_{\hat{f}}^{\mp k_2}\left(H_V(\boldsymbol{y}_\tau^{(V)})\right)\right) - \alpha'\left(H_V(\boldsymbol{y}_\tau^{(V)})\right), \tag{7}$$

where $H_V : \mathbb{R}^{D \times V} \mapsto \mathbb{R}^d$ is a full-multi-view encoder, $\boldsymbol{y}_t^{(V)} \coloneqq \left\{\boldsymbol{y}_t^1, \ldots, y_t^V\right\}$ is the full set of observations at time t and $\alpha' : \mathbb{R}^d \mapsto \mathbb{R}$ is the corresponding scalar potential.

From here, our learning framework follows the standard contrastive learning setup with InfoNCE objective function (Oord et al., 2018), *i.e.* we learn to minimize the following expression for the log-likelihood:

$$\log p_\psi \left(\boldsymbol{y}_t^{a,\star} | \boldsymbol{y}_t^{b,+}, \{ \boldsymbol{y}_t^{c,-} \} \right) = \psi_{ab} \left(\boldsymbol{y}_t^{a,\star}, \boldsymbol{y}_t^{b,+} \right) - \log \sum_{\boldsymbol{y}_t^* \in \boldsymbol{y}_t^{b,+} \cup \{ \boldsymbol{y}_t^{c,-} \}} \exp \left(\psi_{a*} \left(\boldsymbol{y}_t^{a,\star}, \boldsymbol{y}_t^* \right) \right), \quad (8)$$

where we have used the following notation $\boldsymbol{y}_t^\star$, $\boldsymbol{y}_t^+$, $\{\boldsymbol{y}_t^-\}$ to denote, respectively, the reference (or anchor) and positive samples and the set of negative samples. The learning problem can then be fully specified as the minimization:

$$\min_\psi \mathcal{L}\left[\psi\right] = \min_\psi \mathbb{E}_{a,b,c} \mathbb{E}_{k_1,k_2} \mathbb{E}_{t,\tau} \left[-\log p_\psi \left(\boldsymbol{y}_{t \mp k_1}^a | \boldsymbol{y}_{t \pm k_2}^b, \{ \boldsymbol{y}_\tau^c \} \right) \right], \quad (9)$$

where all the expectations are computed by uniform sampling in the relevant ranges. Intuitively, the learning process is constructed in such a way as to exploit the temporal structure and available forward/backward solution operators of the latent dynamical system, while the enforced across-view consistency provides the encoder with the learning signal to effectively denoise the observations $\boldsymbol{y}_t^\alpha$ by forcing it to rely only on the shared component of the signal (the $g(\boldsymbol{x}_t)$ in (2)) and discard the nuisance (noise) component $\boldsymbol{\xi}_t^\alpha$. We offer a graphical overview of the problem in Figure 1.

We can now state our main theoretical result for the full-views model (7), formulated for the asymptotic limit $V \to \infty$ and the particular case of no external control $\boldsymbol{u}_t \equiv 0$. Our result extends Laiz et al. (2025) to the case of noisy observations. We then show that strong identification holds well in practice also for the DYSCO model (5) and the case of (known) external forcing.

Theorem 1 (Multi-view contrastive estimation of noisy dynamics)**.** *Consider the latent dynamical system $\boldsymbol{x}_{t+1} = F(\boldsymbol{x}_t) + \boldsymbol{\varepsilon}_t$ with $F : \mathbb{R}^d \mapsto \mathbb{R}^d$, a C^2 diffeomorphism and noise $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \Sigma)$, $\Sigma \succ 0$, independent of time. We observe the system via the non-linear observation channel* (2) *with $g : \mathbb{R}^d \mapsto \mathbb{R}^D$, $D \geq d$ a C^1 bi-Lipschitz embedding and sub-Gaussian noise $\boldsymbol{\xi}_t^a$ independent of $\boldsymbol{x}_t$ with $a = 1, \ldots, V$. Finally, consider the full multi-view model Ψ_V defined in* (7)*. Assume the following:*

- (A1) *The multi-horizon objective $\mathcal{L}[\Psi_V]$ is jointly realizable,* i.e. *there exists a single tuple $(H_V, \widehat{F}, \alpha)$ attaining the infimum of every active constituent $\mathcal{L}_V^{k_1,k_2}$ and $\lambda^{(1,0)} > 0$.*
- (A2) *In the infinite-view population limit, the population global minimizers have a stable C^1 limit for both representation and dynamics:*

$$H_V(\boldsymbol{y}_t^{(V)}) \to r(\boldsymbol{x}_t) \qquad \text{in probability}$$

  for $r \in C^1$ and $\widehat{F}_V \to \widehat{F} \in C^1$ on the learned support. The potential class represents the marginal correction. Density and observation regularity conditions are given in Appendix A.

Then, in the limit $V, T \to \infty$, any global minimizer identifies the latent state and the dynamics up to a common affine indeterminacy, i.e. *there exist $L \in \mathrm{GL}(d)$ and $\boldsymbol{b} \in \mathbb{R}^d$ such that for every point $\boldsymbol{x} \in \mathcal{U}$:*

$$r(\boldsymbol{x}) = L\boldsymbol{x} + \boldsymbol{b}, \quad (10)$$

moreover, the learned deterministic dynamics $\widehat{F}$ are the affine conjugate of the true dynamics:

$$\widehat{F}(\boldsymbol{z}) = LF\left(L^{-1}(\boldsymbol{z} - \boldsymbol{b})\right) + \boldsymbol{b}, \qquad \boldsymbol{z} \in r(\mathcal{U}). \quad (11)$$

Proof. See Appendix A for the full proof and subsequent remarks. □

3 Experimental Setup

We validate our theoretical results by verifying that our algorithm can recover known latent dynamical systems. We explored a wide variety of dynamical regimes and external forcing configurations, ranging from chaotic (*e.g.* Lorenz attractor) to metastable (*e.g.* heteroclinic) to periodic (*e.g.* Duffing

oscillator) with fast-and-slow dynamics (*e.g.* FitzHugh-Nagumo oscillator), proving that our system is robust and applicable to diverse, realistic scenarios relevant to machine learning and neuroscience.

Data Generating Process Given a dynamical system of the form (1) (or the equivalent continuous-time formulation specified via the corresponding ODE) we generated a collection of latent trajectories $\left\{\boldsymbol{x}_t^i\right\}_t^T$ of equal length $T = 2^{16}$ time steps, for different initial conditions $\boldsymbol{x}_0^i$. We forward-time integrated our system using the Euler-Maruyama scheme and reserved some latent trajectories $\boldsymbol{x}_t^i$ as a held-out validation set and report all our metrics on such examples. We followed the literature (Laiz et al., 2025; Zimmermann et al., 2022; Hyvarinen and Morioka, 2016) and parametrized the mixing function g as a 4-layer randomly initialized MLP with explicit control over the condition number for every layer matrix to ensure injectivity and final number of visible dimensions $D = 256$. The complete observation channel (2) is then obtained by passing the mixing-function's output through a noisy channel. In our experiments we modeled the noise either as a standard additive Gaussian process or as a non-homogeneous Poisson process, both independent across visible dimensions:

$$\boldsymbol{y}_t^a \sim \mathcal{N}\left(g(\boldsymbol{x}_t), \Sigma_\xi\right) \qquad \text{or} \qquad \boldsymbol{y}_t^a \sim \text{Poisson}\left(\left[g(\boldsymbol{x}_t)\right]_+ \Delta t\right), \tag{12}$$

where $[\,\cdot\,]_+$ is a rectification function to ensure positivity, which we simply take to be a linear rescaling of the output into a pre-defined range followed by a ReLU gate, while we fixed $\Sigma_\xi \equiv \sigma^2 \text{Id}$, $\sigma \in \mathbb{R}$. Different views a are obtained via independent realizations of the same process. We stress that the Poisson construction formally violates the assumptions of Theorem (1) since the observation-noise statistics depend on the latent state $\boldsymbol{x}_t$. However, we explored this particular form since it is an ubiquitous model of neural spiking dynamics (Pandarinath et al., 2018a) with $[g(\boldsymbol{x}_t)]_+$ representing the instantaneous firing rates, making it an important target for our method, and we empirically show that our model still works well in practice. In the cases where an external control signal $\boldsymbol{u}_t$ was present, we treated it as known input to our model and let it learn the appropriate input coupling matrix C.

Model architecture In our experiments the encoder network h is a 4-layer MLP with GELU activations and hidden layers of dimension 64, with appropriate input and output dimensions, i.e. D and d. In the experiments with Poisson observations, we included a custom input-normalization layer designed to mitigate the impact of a mean-dependent noise variance introduced by the Poisson sampling. We defined it as the sequential composition of the Freeman-Tukey transform (Freeman and Tukey, 1950), a short smoothing convolutional kernel and a final batch-normalization layer. For the dynamics model $\hat{f}$ we used a polynomial basis (the collection of all monomials of degree up to δ) of sufficient total degree δ to exactly capture the ground-truth dynamics (*e.g.* $\delta = 2$ for the case of the Lorenz system). We generate a global coefficient matrix $\boldsymbol{\Theta} = \text{reshape}(\text{MLP}(\boldsymbol{c}))$ from a learned global source $\boldsymbol{c} \in \mathbb{R}^{32}$, with a 2-layer MLP with hidden dimensions 64. Experiment-specific settings are given Appendix E. We computed the trajectory roll-outs (both forward and backward in time) for a maximum time horizon $\kappa = 8$ using the RK4 scheme implemented via the ODEINT function of the TORCHDIFFEQ package (Chen, 2018). We followed the recommendation in (Laiz et al., 2025) and omitted the scalar potential term since it was shown to have negligible impact on model accuracy while having a significant computational cost.

Evaluation metrics The evaluation metrics used to measure the model performance reflect the affine indeterminacy expected from Theorem (1). Given the ground-truth trajectory $\boldsymbol{x}_t$ and the model estimate $\hat{\boldsymbol{x}}_t$, we numerically estimate the optimal aligning affine transformation by solving the optimization problem:

$$L^\star, \boldsymbol{b}^\star \coloneqq \arg\min_{L,\boldsymbol{b}} \sum_t^T \parallel \boldsymbol{x}_t - (L\hat{\boldsymbol{x}}_t + \boldsymbol{b}) \parallel_2^2, \tag{13}$$

The aligned trajectory is $\hat{\boldsymbol{x}}_t^\star = L^\star \hat{\boldsymbol{x}}_t + \boldsymbol{b}^\star$, whose R^2 against $\boldsymbol{x}_t$ measures model performance on state recovery. Similarly, since tangent vectors transform with the linear part alone, for $\boldsymbol{z} = L\boldsymbol{x} + \boldsymbol{b}$:

$$\widehat{f}(\boldsymbol{z}) = Lf\big(L^{-1}(\boldsymbol{z} - \boldsymbol{b})\big), \qquad \widehat{C} = LC. \tag{14}$$

We therefore compute the $\text{dyn}R^2$ metric (Laiz et al., 2025) as the R^2 score between $f(\boldsymbol{x}_t)$ and $\hat{f}^\star(\hat{\boldsymbol{x}}_t) = L^\star \hat{f}(\hat{\boldsymbol{x}}_t)$.

Training details We trained our model with a combination of AdamW (Loshchilov and Hutter, 2017) and Muon (Jordan et al., 2024) optimizers: all hidden-layer matrix parameters were assigned

Table 1: **Identification of latent dynamical systems.** We present results for several (non-linear) dynamical systems, spanning several dynamical regimes from chaotic attractors to forced oscillators to point attractors and varying external forcing configuration. For a formal definition of each dynamical system we defer to the Appendix B. Each value is the mean across 3 runs and we report the standard deviation.

		Gaussian noise			Poisson noise	
Dynamical System	$\boldsymbol{u}_t$	σ	% R^2 (↑)	% dynR^2 (↑)	% R^2 (↑)	% dynR^2 (↑)
Duffing	✓	0.1	99.1 ± 0.9	97.8 ± 2.6	98.6 ± 0.6	94.4 ± 4.6
Lorenz	✗	1.0	93.5 ± 2.1	75.9 ± 6.6	94.4 ± 1.1	82.3 ± 3.8
FitzHugh–Nagumo	✓	0.2	99.7 ± 0.1	93.8 ± 0.8	99.7 ± 0.1	89.1 ± 2.9
Winner–Take–All	✓	0.2	95.1 ± 1.7	70.0 ± 5.4	97.4 ± 0.9	72.4 ± 3.1
Double–Well	✓	0.2	99.3 ± 0.5	93.8 ± 4.0	99.3 ± 0.4	92.7 ± 1.0
Stuart–Landau	✗	0.2	96.6 ± 0.8	78.4 ± 4.8	98.0 ± 1.4	83.1 ± 2.2
Heteroclinic	✓	0.2	99.6 ± 0.2	97.5 ± 0.8	99.6 ± 0.1	84.1 ± 5.5
Average			97.6 ± 2.6	87.4 ± 11.2	98.0 ± 2.0	85.4 ± 7.8

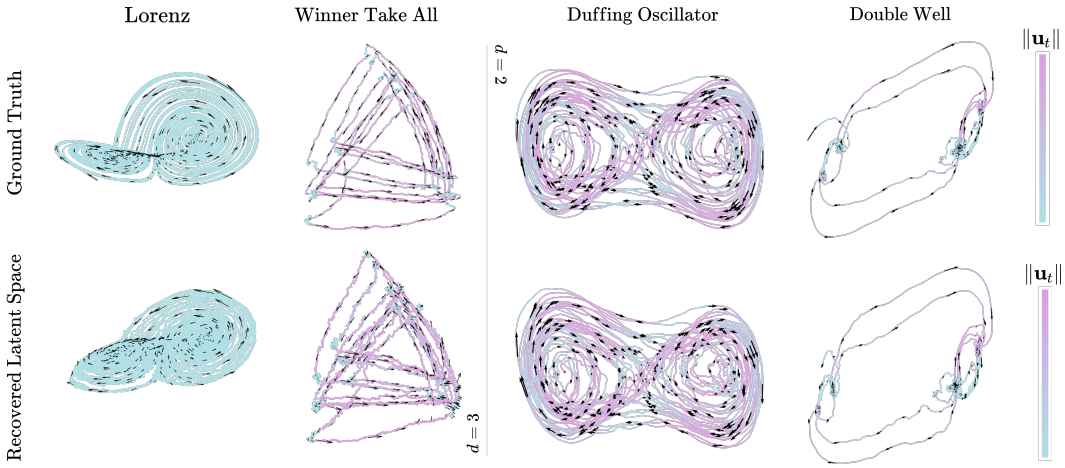


Figure 2: **Dynamical system recovery via multi-view contrastive learning**. Example phase space portraits of different dynamical systems (top-row) with corresponding affine-aligned recovered systems (bottom row) for either $d = 3$ (left) or $d = 2$ (right). Dynamical system trajectories are color-coded based on the external forcing magnitude $\| \boldsymbol{u}_t \|$. All reported systems are observed via a Poisson observation channel.

to the Muon optimizers, while the rest was optimized via AdamW. We used a learning rate of 3×10^{-4} and a weight decay of 10^{-4} (for both optimizers). We trained with a batch size of 4096 on a single NVIDIA RTX A4000 for 200 epochs. Depending on the system configuration, the training time was approximately two hours. Full configuration details are reported in Appendix E.

4 Results

Dynamical Systems In this section we report the identification results for several different dynamical systems. For a complete formal description of the dynamical systems considered in this study we refer to Appendix B. In Table 1 we have reported the results for both R^2 and dynR^2 metrics for the dynamical systems explored in this work, while Figure 2 reports some phase-portrait visual examples. Indeed, for a broad class of systems, spanning different dynamical regimes, external-forcing and noise configurations, our algorithm consistently achieves very high latent trajectory accuracies and good-to-very-good flow field accuracies. We note how even for the Winner-Take-All system

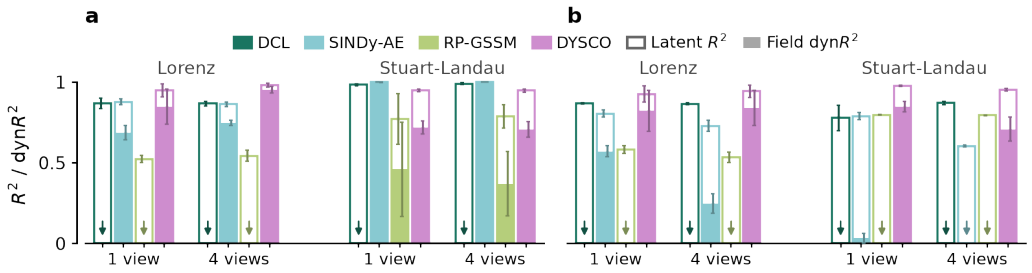


Figure 3: **Controlled baseline comparisons.** Latent R^2 (outlined bars) and field dynR^2 (filled bars) for Lorenz and Stuart–Landau, with one or four views. (**a**) Gaussian noise, $\sigma = 1.0$ for Lorenz and 0.2 for Stuart–Landau. (**b**) Poisson noise. Values are mean $\pm$ SEM over the three independent. For $\mathrm{dyn}R^2$ we reported $\max\left(0, \mathrm{dynR}^2\right)$, with arrows denoting clipped values.

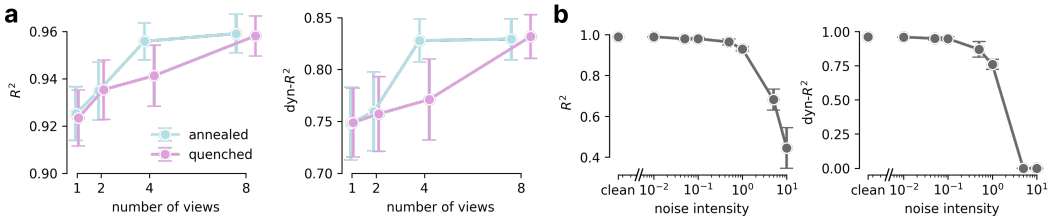


Figure 4: **Ablation results**. (**a**) Results for impact of available number of views on the final trajectory (left) and flow-field (right) model performances for the Lorenz dynamical system (observed via a Gaussian channel with standard deviation $\sigma = 1$) and for the two noise condition annealed (cyan) and quenched (plum). Markers are averages over 3 repetitions and error bars are the standard error of the mean. (**b**) Impact of the observation noise intensity σ on model performances for the Lorenz dynamical systems observed via a Gaussian channel of increasing variance σ^2 and for $|V| = 2$. The clean condition corresponds to the case $\sigma = 0$ (*i.e.* no observational noise).

where the $\mathrm{dyn}R^2$ metric shows average results, the phase portrait (Figure 2, second column) in fact shows excellent agreement with the ground truth. We take this fact to indicate how the metric is potentially affected by the many slow-velocity attractors, where the flow field strength gets overshadowed by the system and observational noise.

Controlled baseline comparisons We compare DYSCO with DCL (Laiz et al., 2025), SINDy-AE (Champion et al., 2019) and RP-GSSM (Hromadka et al., 2025) on Lorenz and Stuart–Landau (Figure 3). At each view count, methods receive the same raw observations, DYSCO additionally uses their alignment, while baselines treat views as additional trajectories. DCL uses switching-linear dynamics and assumes clean nonlinear observations, an assumption violated in our noisy-observation setting. DYSCO and SINDy-AE use polynomial libraries compatible with the benchmark fields. While SINDy-AE performs particularly well on Gaussian Stuart–Landau, DYSCO provides strong field recovery, with a leading gap particularly marked under Poisson noise, where all other baselines struggle. These results confirm that multi-view consistency is one critical mechanism enabling identification under realistic noisy observations.

Symbolic recovery While Theorem 1 guarantees identification of the ground-truth dynamical system up to an affine transformation, exact symbolic recovery is precluded by this gauge freedom. The coefficient representation $\boldsymbol{\Theta}$ of the latent flow field is gauge-dependent, since rotations or shifts in latent space mix the components of $\boldsymbol{\Theta}$ in a structured way. For the practically important case of a polynomial basis of degree δ, however, the basis is closed under affine transformations $\boldsymbol{x}'_t = \mathcal{A}\boldsymbol{x}_t \equiv L\boldsymbol{x}_t + \boldsymbol{b}$, and the induced action on the coefficients can be written explicitly:

$$\boldsymbol{\Theta}' = L\boldsymbol{\Theta}T^{(\delta)}_{\mathcal{A}^{-1}}, \tag{15}$$

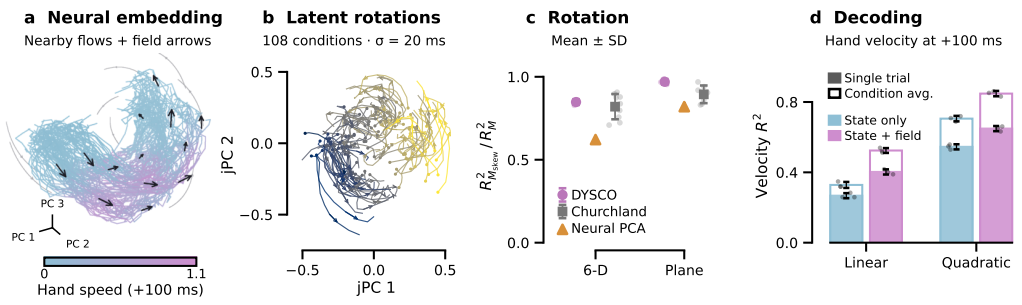


Figure 5: **DYSCO on motor-cortical recordings.** (**a**) Held-out condition means in 3D PCs, colored by hand speed at $+100$ ms. Dark arrows: projected field tangents (40 ms); grey curves: nearby 6D flows ($u = 0, \pm 80$ ms). (**b**) Training-fitted jPCA trajectories of DYSCO latents, colored by starting jPC1; we applied a light Gaussian smoothing with $\sigma = 20$ ms, arrows indicate temporal progression. (**c**) Skew/full R^2 ratios with neural-PCA and (Churchland et al., 2012) controls. (**d**) Linear and quadratic ridge decoding of hand velocity from state alone and state plus field. Filled bars: single trials; outlines: condition averages. DYSCO values: 3 seeds, mean $\pm$ SD.

where $T^{\delta}_{\mathcal{A}^{-1}}$ is a coefficient transport matrix defined via the relation:

$$\Xi_{\delta}\left(L^{-1}\left(\boldsymbol{x}_t - \boldsymbol{b}\right)\right) = T^{\delta}_{\mathcal{A}^{-1}} \Xi_{\delta}\left(\boldsymbol{x}_t\right). \tag{16}$$

This characterization defines the orbit of valid coefficient representations and opens the door to downstream symbolic regression: under the assumption that the underlying system satisfies *symbolic simplicity* (Brunton et al., 2016), we can attempt symbolic identification by seeking the sparsest representative within the orbit $\boldsymbol{\Theta}^{\star}$:

$$\boldsymbol{\Theta}^{\star} = \arg\min_{L,\boldsymbol{b}} \parallel L\boldsymbol{\Theta} T^{(\delta)}_{\mathcal{A}^{-1}} \parallel_0 . \tag{17}$$

We offer a proof of concept application of this technique in Appendix C.

Motor-cortical recordings To assess the applicability of DYSCO on real neural data, we considered the MC_Maze dataset (Pei et al., 2021), comprising 137 neurons recorded from macaque motor cortex during a reaching task (Churchland et al., 2012). We regarded repeated movement-aligned trials as approximate views of a shared condition-dependent trajectory and trained a six-dimensional DYSCO model with a quadratic field and the observed go cue as an external input (Figure 5a). The resulting latent space exhibited rotational organization (Figure 5a,b), with skew/full fit ratios in general agreement with the published neural values (Figure 5c). To inspect the learned equations, we chose a fixed basis whose first two coordinates span the training jPCA plane. The corresponding linear field components showed a clear rotational structure:

$$\begin{aligned} \dot{q}_1 &= -0.43 q_1 - 4.67 q_2 + r_1(\boldsymbol{q}, u), \\ \dot{q}_2 &= 5.59 q_1 + 0.46 q_2 + r_2(\boldsymbol{q}, u), \end{aligned} \tag{18}$$

where r_1, r_2 collect the remaining terms (constant, linear couplings to the other four coordinates, quadratic terms and observed-input response). Appendix F details the coordinate choice, full linear matrix and approximation errors. Finally, including the learned field improved single-trial hand-velocity decoding for both linear ($R^2 = 0.402 \pm 0.015$ versus 0.267 ± 0.014) and quadratic readouts (0.649 ± 0.015 versus 0.545 ± 0.015; Figure 5d).

Ablation results We investigated the properties of our learning framework via two ablation studies. We first gauged the impact of the number of total available views $|V|$ on the final model performances. While Theorem 1 guarantees identifiability asymptotically for $V \to \infty$, in practice we intuitively expected that a richer set of independent views would allow our model to achieve progressively better results. We focused on the Lorenz system and trained the same encoder and dynamics model on a fixed dataset of latent trajectories for a progressively increasing number of available views $a \in \{1, 2, 4, 8\}$. We explored both possible noise conditions: *annealed* and *quenched*, corresponding, respectively, to the case where the observation noise for a given latent state $\boldsymbol{x}_t$ is

resampled at every training iteration or not. The quenched configuration we argue is conceptually closer to the experimental reality where data acquisition is performed once, yielding frozen noise realizations. We performed our experiment for a Gaussian observation model with noise intensity $\sigma = 1$ and report our results in Figure 4a. Indeed we observed that model performance improved monotonically as a function of the number of views, with the annealed configuration offering minor performance advantages for small numbers of views, which is expected due to the richer resampling of the noise. Interestingly, even with a single available view, the model achieved good results: we speculate that for the particular case of a chaotic attractor where the system explores arbitrarily close trajectories in phase space, even a single view effectively allows for dense resampling of neighboring latent states.

We then explored how robust our system was for an increasingly noisy observation channel. We fixed $V = 2$ and trained our model on the Lorenz system with $\boldsymbol{\xi}_t^a \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$ Gaussian, for increasing noise intensity σ. Our results are showcased in Figure 4b, where the model consistently achieved good performances up to $\sigma^* \simeq 1$, which translates to a threshold signal-to-noise ratio of $\mathrm{SNR}^* \simeq 4\mathrm{dB}$. Interestingly, Figure 4b highlights a small region where trajectory recovery is still possible, but the dynamics is not. This observation is consistent with prior work (Laiz et al., 2025) which illustrated how, when noise dominates the dynamics, even simple contrastive objectives can recover meaningful latent structure while explicit dynamics modeling fails. This delineation helps clarify when standard contrastive learning suffices and when more structured approaches, such as ours, are required. Lastly, we explored the impact of the integration horizon κ on model performance in Appendix D, while an overcomplete-library diagnostic and forward-only objective ablation appear in Appendix G.

5 Discussion

In this work we have introduced a multi-view contrastive inference algorithm for the (symbolic) identification of latent dynamical systems under noisy observations. We showed how our model consistently achieves high quality recovery of both latent trajectories and dynamics under significant observational noise. From a theoretical perspective, our results reinforce and extend the emerging view that contrastive learning implicitly performs inversion of the data-generating process (Zimmermann et al., 2021). Compared to prior frameworks such as DCL (Laiz et al., 2025), our approach demonstrates that identifiability can be retained even under noisy nonlinear observation channels by leveraging the implicit denoising constraint imposed by multi-view consistency, thereby relaxing the idealized observation assumptions typically required in theory. Another important implication concerns the interplay between representation learning and scientific modeling. We opted for explicit symbolic modeling of the dynamics. Our results suggest that one can recover models that are both expressive and interpretable. In this sense, our framework provides a bridge between self-supervised learning and symbolic regression. Lastly, we argue that our setup is relevant for neuroscience, where measurements are inherently noisy but often trial-structured, making the assumption of repeated observations of a shared latent state natural. Achieving strong identifiability in experimentally realistic conditions thus offers a precious tool in the pursuit of understanding the algorithmic function of neural circuits via the lens of dynamical systems (Vyas et al., 2020). In Fig.5 we have provided initial evidence toward this goal, illustrating how DYSCO recovered a known rotational structure in motor-cortical recordings (Churchland et al., 2012). This, in turn, suggests that structured contrastive learning approaches might offer a viable path toward more interpretable, dynamics-level descriptions of neural computation.

Limitations A first limitation concerns the choice of latent dimensionality d and symbolic basis Ξ_f, which are assumed to be known. In practice, both quantities must be estimated from data, and poor choices may hinder accurate system identification (*e.g.*, the chosen basis is not expressive enough to capture the underlying dynamics, or it's overcomplete slowing convergence, see Appendix G). Nevertheless, this limitation can be mitigated in practice. Estimating the intrinsic dimensionality of data is a well-studied problem, with a range of established estimators available (Levina and Bickel, 2004; Facco et al., 2017; Muratore et al., 2022). Moreover, polynomial bases provide a flexible approximation class (*e.g.*, via Taylor expansions) while allowing explicit control over model complexity, making them a practical choice in many settings. Lastly, our current implementation disregards the contribution of the scalar potential α: this term can potentially become important for systems with non-uniform marginals, for example for systems with different fast-slow regimes. Future work should explore how to learn this term jointly and effectively.

References

Z. Bastiani, R. M. Kirby, J. Hochhalter, and S. Zhe. Diffusion-based symbolic regression. *arXiv preprint arXiv:2505.24776*, 2025.

M. Brenner, F. Hess, G. Koppe, and D. Durstewitz. Integrating multimodal data for joint generative modeling of complex dynamics. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 4482–4516. PMLR, 2024. URL `https://proceedings.mlr.press/v235/brenner24a.html`.

S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, Apr. 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1517384113.

K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, Nov. 2019. doi: 10.1073/pnas.1906995116.

R. T. Q. Chen. torchdiffeq, 2018. URL `https://github.com/rtqichen/torchdiffeq`.

M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy. Neural population dynamics during reaching. *Nature*, 487:51–56, 2012. doi: 10.1038/nature11129.

S. d'Ascoli, S. Becker, A. Mathis, P. Schwaller, and N. Kilbertus. ODEFormer: Symbolic Regression of Dynamical Systems with Transformers, Oct. 2023.

E. Facco, M. d'Errico, A. Rodriguez, and A. Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific reports*, 7(1):12140, 2017.

M. F. Freeman and J. W. Tukey. Transformations related to the angular and the square root. *The annals of mathematical statistics*, pages 607–611, 1950.

H. Hälvä, S. L. Corff, L. Lehéricy, J. So, Y. Zhu, E. Gassiat, and A. Hyvarinen. Disentangling Identifiable Features from Noisy Data with Structured Nonlinear ICA, Oct. 2021.

O. Henaff. Data-efficient image recognition with contrastive predictive coding. In *International conference on machine learning*, pages 4182–4192. PMLR, 2020.

S. Hromadka, K. Biegun, L. Fox, J. Heald, and M. Sahani. Maximum likelihood learning of latent dynamics without reconstruction. *arXiv preprint arXiv:2505.23569*, 2025. doi: 10.48550/arXiv.2505.23569.

A. Hyvarinen and H. Morioka. Unsupervised Feature Extraction by Time-Contrastive Learning and Nonlinear ICA. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

A. Hyvarinen, H. Sasaki, and R. Turner. Nonlinear ICA Using Auxiliary Variables and Generalized Contrastive Learning. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pages 859–868. PMLR, Apr. 2019.

A. Hyvärinen, I. Khemakhem, and H. Morioka. Nonlinear independent component analysis for principled disentanglement in unsupervised deep learning. *Patterns*, 4(10), Oct. 2023. ISSN 2666-3899. doi: 10.1016/j.patter.2023.100844.

K. Jordan, Y. Jin, V. Boza, Y. Jiacheng, F. Cesista, L. Newhouse, and J. Bernstein. Muon: An optimizer for hidden layers in neural networks, 2024. URL `https://kellerjordan.github.io/posts/muon/`.

T. D. Kim, T. Z. Luo, J. W. Pillow, and C. D. Brody. Inferring Latent Dynamics Underlying Neural Population Activity via Neural Differential Equations. In *Proceedings of the 38th International Conference on Machine Learning*, pages 5551–5561. PMLR, July 2021.

A. A. Klishin, J. Bakarji, J. N. Kutz, and K. Manohar. Statistical mechanics of dynamical system identification. *Physical Review Research*, 7(3):033181, Aug. 2025. doi: 10.1103/4d98-tdlp.

C. Kolb, L. Frost, B. Bischl, and D. Rügamer. Differentiable Sparsity via d-gating: Simple and Versatile Structured Penalization, Oct. 2025.

R. G. Laiz, T. Schmidt, and S. Schneider. Self-supervised contrastive learning performs non-linear system identification. In *ICLR*, page 38, 2025.

E. Levina and P. Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17, 2004.

I. Loshchilov and F. Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

C. Louizos, M. Welling, and D. P. Kingma. Learning Sparse Neural Networks through l0 Regularization, June 2018.

D. Marr. Vision: A computational investigation into the human representation and processing of visual information. *(No Title)*, 1982.

F. Mastrogiuseppe and S. Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623, 2018.

M. W. Mathis and A. Mathis. Joint modelling of brain and behaviour dynamics with artificial intelligence. *Nature Reviews Neuroscience*, 27:87–100, 2026. doi: 10.1038/s41583-025-00996-1. URL `https://doi.org/10.1038/s41583-025-00996-1`.

P. Muratore, S. Tafazoli, E. Piasini, A. Laio, and D. Zoccolan. Prune and distill: similar reformatting of image information along rat visual cortex and deep neural networks. *Advances in Neural Information Processing Systems*, 35:30206–30218, 2022.

A. v. d. Oord, Y. Li, and O. Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

C. Pandarinath, D. J. O'Shea, J. Collins, R. Jozefowicz, S. D. Stavisky, J. C. Kao, E. M. Trautmann, M. T. Kaufman, S. I. Ryu, L. R. Hochberg, J. M. Henderson, K. V. Shenoy, L. F. Abbott, and D. Sussillo. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, 15(10):805–815, Oct. 2018a. ISSN 1548-7105. doi: 10.1038/s41592-018-0109-9.

C. Pandarinath, D. J. O'Shea, J. Collins, R. Jozefowicz, S. D. Stavisky, J. C. Kao, E. M. Trautmann, M. T. Kaufman, S. I. Ryu, L. R. Hochberg, J. M. Henderson, K. V. Shenoy, L. F. Abbott, and D. Sussillo. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, 15(10):805–815, Oct. 2018b. ISSN 1548-7105. doi: 10.1038/s41592-018-0109-9.

F. Pei, J. Ye, D. Zoltowski, A. Wu, R. H. Chowdhury, H. Sohn, J. E. O'Doherty, K. V. Shenoy, M. T. Kaufman, M. Churchland, M. Jazayeri, L. E. Miller, J. Pillow, I. M. Park, E. L. Dyer, and C. Pandarinath. Neural latents benchmark '21: Evaluating latent variable models of neural population activity. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. URL `https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/979d472a84804b9f647bc185a877a8b5-Abstract-round2.html`.

M. Schimel, T.-C. Kao, K. T. Jensen, and G. Hennequin. iLQR-VAE: Control-based learning of input-driven dynamics with applications to neural data. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=wRODLDHaAiW`.

S. Schneider, A. Baevski, R. Collobert, and M. Auli. wav2vec: Unsupervised pre-training for speech recognition. *arXiv preprint arXiv:1904.05862*, 2019.

S. Schneider, J. H. Lee, and M. W. Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 617(7960):360–368, May 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06031-6.

N. P. Shah, B. A. Krasa, E. Kunz, N. Hahn, F. Kamdar, D. Avansino, L. R. Hochberg, J. M. Henderson, and D. Sussillo. Improved interpretability in LFADS models using a learned, context-dependent per-trial bias, Oct. 2025. ISSN 2692-8205.

D. Sussillo and O. Barak. Opening the Black Box: Low-Dimensional Dynamics in High-Dimensional Recurrent Neural Networks. *Neural Computation*, 25(3):626–649, Mar. 2013. ISSN 0899-7667. doi: 10.1162/NECO_a_00409.

C. Versteeg, J. D. McCart, M. Ostrow, D. M. Zoltowski, C. B. Washington, L. Driscoll, O. Codol, J. A. Michaels, S. W. Linderman, D. Sussillo, and C. Pandarinath. Computation-through-Dynamics Benchmark: Simulated datasets and quality metrics for dynamical models of neural activity, Feb. 2025.

S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy. Computation Through Neural Population Dynamics. *Annual Review of Neuroscience*, 43(1):249–275, July 2020. ISSN 0147-006X, 1545-4126. doi: 10.1146/annurev-neuro-092619-094115.

T. Wang and P. Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International conference on machine learning*, pages 9929–9939. PMLR, 2020.

H. Yang, W. Wen, and H. Li. DeepHoyer: Learning Sparser Neural Network with Differentiable Scale-Invariant Sparsity Measures, Jan. 2020.

R. S. Zimmermann, Y. Sharma, S. Schneider, M. Bethge, and W. Brendel. Contrastive Learning Inverts the Data Generating Process. In *Proceedings of the 38th International Conference on Machine Learning*, pages 12979–12990. PMLR, July 2021.

R. S. Zimmermann, Y. Sharma, S. Schneider, M. Bethge, and W. Brendel. Contrastive Learning Inverts the Data Generating Process, Apr. 2022.

Supplementary Material

A Proof of Theoretical Results

In this section we state the claim of the main theoretical result and provide a full proof. We then provide two remarks discussing the practical implementation choices for DYSCO.

Theorem 1 (Multi-view contrastive estimation of noisy dynamics)**.** *Consider the latent Markov process on* $\mathbb{R}^d$

$$\boldsymbol{x}_{t+1} = F(\boldsymbol{x}_t) + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma), \qquad \Sigma \succ 0,$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ *is a* C^2 *diffeomorphism. Let* $\mathcal{U} \subseteq \mathbb{R}^d$ *be a connected open region on which identification is claimed. Let* q *denote the marginal density of* $\boldsymbol{x}_t$*, and assume that the clean density ratio*

$$(\boldsymbol{x}, \boldsymbol{x}') \mapsto \frac{p(\boldsymbol{x}' \mid \boldsymbol{x})}{q(\boldsymbol{x}')}$$

is finite and continuous on $\mathcal{U} \times \mathcal{U}$*. Suppose that for each* t *we observe* V *independent noisy views*

$$\boldsymbol{y}_t^a = g(\boldsymbol{x}_t) + \boldsymbol{\xi}_t^a, \qquad a = 1, \dots, V,$$

where $g : \mathbb{R}^d \to \mathbb{R}^D$*,* $D \geq d$*, is a* C^1 *bi-Lipschitz embedding with lower bi-Lipschitz constant* $m_g > 0$*. Assume that the centered noises* $\boldsymbol{\xi}_t^a - \boldsymbol{\mu}_\xi$ *are independent, mean-zero, sub-Gaussian random vectors in* $\mathbb{R}^D$ *with sub-Gaussian parameter* σ_ξ^2*, uniformly in* t *and* a*, where* $\boldsymbol{\mu}_\xi = \mathbb{E}[\boldsymbol{\xi}_t^a]$*.*

Consider the all-views DYSCO population model with encoder $H_V : (\mathbb{R}^D)^V \to \mathbb{R}^d$*, dynamics* $\widehat{F}_V$*, and potential* α_V*, with rollout score*

$$\Psi_V^{k_1,k_2}(\boldsymbol{y}_t, \boldsymbol{y}'_\tau) = -\left\| \widehat{F}_V^{\pm k_1}(H_V(\boldsymbol{y}_t)) - \widehat{F}_V^{\mp k_2}(H_V(\boldsymbol{y}'_\tau)) \right\|^2 - \alpha_V(H_V(\boldsymbol{y}'_\tau)). \tag{19}$$

Let

$$\mathcal{L}_V[\Psi_V] = \sum_{k_1,k_2=0}^{\kappa} \lambda^{k_1,k_2} \mathcal{L}_V^{k_1,k_2}[\Psi_V]. \tag{20}$$

Assume the following.

(A1) *The multi-horizon objective is jointly realizable: there exists a single tuple* $(H_V, \widehat{F}_V, \alpha_V)$ *attaining the infimum of every active constituent* $\mathcal{L}_V^{k_1,k_2}$*. Moreover, the forward one-step constituent is active,* $\lambda^{1,0} > 0$*.*

(A2) *After fixing the affine gauge, consider any sequence of population global minimizers for which, along a subsequence, the learned all-views representation and dynamics have a stable* C^1 *limit:*

$$H_V(\boldsymbol{y}_t^{(V)}) \longrightarrow r(\boldsymbol{x}_t) \qquad \textit{in probability}$$

for $r \in C^1$*, and* $\widehat{F}_V \to \widehat{F}$ *in* C^1_{loc} *on the learned support. The potential class is rich enough to represent the marginal term* $\log q$ *on this support.*

Then in the limit $V, T \to \infty$ *every such limiting minimizer identifies the latent state and the deterministic dynamics up to a common affine indeterminacy. That is, on* $\mathcal{U}$ *there exist* $L \in \mathrm{GL}(d)$ *and* $\boldsymbol{b} \in \mathbb{R}^d$ *such that*

$$r(\boldsymbol{x}) = L\boldsymbol{x} + \boldsymbol{b}, \tag{21}$$

and

$$\widehat{F}(\boldsymbol{z}) = LF\left(L^{-1}(\boldsymbol{z} - \boldsymbol{b})\right) + \boldsymbol{b}, \qquad \boldsymbol{z} \in r(\mathcal{U}). \tag{22}$$

Proof. We prove the result in five steps.

Step 1: Independent views denoise the observation. For fixed t, define the empirical all-views average

$$\bar{\boldsymbol{y}}_t^{(V)} = \frac{1}{V}\sum_{v=1}^{V} \boldsymbol{y}_t^v = g(\boldsymbol{x}_t) + \frac{1}{V}\sum_{v=1}^{V} \boldsymbol{\xi}_t^v,$$

and set $\tilde{g}(\boldsymbol{x}) = g(\boldsymbol{x}) + \boldsymbol{\mu}_\xi$. Then

$$\bar{\boldsymbol{y}}_t^{(V)} - \tilde{g}(\boldsymbol{x}_t) = \frac{1}{V}\sum_{v=1}^{V} (\boldsymbol{\xi}_t^v - \boldsymbol{\mu}_\xi).$$

The right-hand side is sub-Gaussian with parameter σ_ξ^2/V. Hence, for a universal constant $C_0 > 0$, every $u > 0$, and every fixed t,

$$p\left(\left\|\bar{\boldsymbol{y}}_t^{(V)} - \tilde{g}(\boldsymbol{x}_t)\right\| > C_0\sigma_\xi\sqrt{\frac{D+u}{V}}\right) \le e^{-u}.$$

Equivalently, over any finite trajectory $t = 0, \dots, T$, a union bound gives, with probability at least $1-\delta$,

$$\max_{0\le t\le T} \|\bar{\boldsymbol{y}}_t^{(V)} - \tilde{g}(\boldsymbol{x}_t)\| \le C_0\sigma_\xi\sqrt{\frac{D+\log((T+1)/\delta)}{V}}.$$

In particular, for each fixed t,

$$\mathbb{E}\|\bar{\boldsymbol{y}}_t^{(V)} - \tilde{g}(\boldsymbol{x}_t)\| \le C\sigma_\xi\sqrt{\frac{D}{V}}.$$

Step 2: The all-views posterior concentrates on the clean latent state. Because g is bi-Lipschitz, $\tilde{g}$ has the same lower bi-Lipschitz constant. Define the oracle inverse estimator

$$\hat{\boldsymbol{x}}_t^{(V)} \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^d} \|\bar{\boldsymbol{y}}_t^{(V)} - \tilde{g}(\boldsymbol{x})\|.$$

The same projection argument as in the noiseless inverse problem gives

$$\|\hat{\boldsymbol{x}}_t^{(V)} - \boldsymbol{x}_t\| \le \frac{2}{m_g}\|\bar{\boldsymbol{y}}_t^{(V)} - \tilde{g}(\boldsymbol{x}_t)\|.$$

Therefore

$$\mathbb{E}\|\hat{\boldsymbol{x}}_t^{(V)} - \boldsymbol{x}_t\| \le C\frac{\sigma_\xi}{m_g}\sqrt{\frac{D}{V}} \longrightarrow 0.$$

where the L^1 convergence follows directly from the sub-Gaussian bound.

Let $\Pi_V(\cdot \mid \boldsymbol{y}_t^{(V)})$ be the posterior law of $\boldsymbol{x}_t$ given all views at time t. Since $\hat{\boldsymbol{x}}_t^{(V)}$ is measurable with respect to $\boldsymbol{y}_t^{(V)}$, the tower property yields

$$\mathbb{E}\left[\mathbb{E}[\|\hat{\boldsymbol{x}}_t^{(V)} - \boldsymbol{x}_t\| \mid \boldsymbol{y}_t^{(V)}]\right] = \mathbb{E}\|\hat{\boldsymbol{x}}_t^{(V)} - \boldsymbol{x}_t\| \to 0.$$

Thus the conditional expectation converges to zero in probability. Conditional Markov's inequality gives, for every $\varepsilon > 0$,

$$\Pi_V\left(\{\boldsymbol{x} : \|\boldsymbol{x} - \hat{\boldsymbol{x}}_t^{(V)}\| > \varepsilon\} \mid \boldsymbol{y}_t^{(V)}\right) \le \frac{\mathbb{E}[\|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_t^{(V)}\| \mid \boldsymbol{y}_t^{(V)}]}{\varepsilon} \to 0$$

in probability. Since also $\hat{\boldsymbol{x}}_t^{(V)} \to \boldsymbol{x}_t$ in probability, we conclude

$$\Pi_V(\cdot \mid \boldsymbol{y}_t^{(V)}) \Rightarrow \delta_{\boldsymbol{x}_t} \qquad \text{in probability.}$$

Step 3: The multi-horizon problem reduces to the active one-step problem. Let $\mathcal{A}$ be the set of active rollouts $\mathcal{A} = \{(k_1, k_2) : \lambda^{k_1,k_2} > 0\}$. For each active constituent define

$$m^{k_1,k_2} = \inf_{H_V, \widehat{F}_V, \alpha_V} \mathcal{L}_V^{k_1,k_2}(H_V, \widehat{F}_V, \alpha_V).$$

By joint realizability, a single tuple attains all these infima. Since every active term is bounded below by its own infimum,

$$\mathcal{L}_V(H_V, \widehat{F}_V, \alpha_V) \geq \sum_{(k_1,k_2)\in\mathcal{A}} \lambda^{k_1,k_2} m^{k_1,k_2},$$

and equality is attained by the jointly realizable tuple. Hence every global minimizer of the full objective minimizes each active constituent individually. In particular, since the forward one-step constituent is active, the identifiability analysis may be reduced to $\mathcal{L}_V^{1,0}$.

Step 4: The Bayes-optimal all-views score converges to the clean latent score. Write $\boldsymbol{y} = \boldsymbol{y}_t^{(V)}$ and $\boldsymbol{y}' = \boldsymbol{y}_{t+1}^{(V)}$. The Bayes-optimal one-step contrastive score is, up to an anchor-only term Wang and Isola (2020),

$$\psi_V^\star(\boldsymbol{y}, \boldsymbol{y}') = \log \frac{p_V(\boldsymbol{y}' \mid \boldsymbol{y})}{q_V(\boldsymbol{y}')} + c_V(\boldsymbol{y}),$$

where q_V is the marginal density of the negative all-views sample. Let

$$\Pi_V(d\boldsymbol{x} \mid \boldsymbol{y}) = p(\boldsymbol{x}_t \in d\boldsymbol{x} \mid \boldsymbol{y}_t^{(V)} = \boldsymbol{y})$$

and let $\Pi_V^+(d\boldsymbol{x}' \mid \boldsymbol{y}')$ be the posterior law of $\boldsymbol{x}_{t+1}$ under the marginal q. Bayes' rule gives

$$\frac{p_V(\boldsymbol{y}' \mid \boldsymbol{y})}{q_V(\boldsymbol{y}')} = \iint \frac{p(\boldsymbol{x}' \mid \boldsymbol{x})}{q(\boldsymbol{x}')} \Pi_V(d\boldsymbol{x} \mid \boldsymbol{y}) \Pi_V^+(d\boldsymbol{x}' \mid \boldsymbol{y}').$$

By Step 2, both posteriors concentrate:

$$\Pi_V(d\boldsymbol{x} \mid \boldsymbol{y}_t^{(V)}) \Rightarrow \delta_{\boldsymbol{x}_t}(d\boldsymbol{x}), \qquad \Pi_V^+(d\boldsymbol{x}' \mid \boldsymbol{y}_{t+1}^{(V)}) \Rightarrow \delta_{\boldsymbol{x}_{t+1}}(d\boldsymbol{x}').$$

Since the clean density ratio is continuous and finite on the relevant support,

$$\psi_V^\star(\boldsymbol{y}_t^{(V)}, \boldsymbol{y}_{t+1}^{(V)}) \longrightarrow \log p(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t) - \log q(\boldsymbol{x}_{t+1}) + c(\boldsymbol{x}_t).$$

Step 5: Score matching induces affine recovery. By (A1), every global minimizer of the full objective realizes the Bayes-optimal active one-step score. By (A2), along any stable limiting subsequence the model score therefore satisfies the clean score identity. At this point, we have reduced the problem to the clean observation case already tackled in Laiz et al. (2025). The remaining of the proof argument can thus proceed analogously as in Laiz et al. (2025). We substitute the expression 7 for our full-view model for ψ_V^* and the potential α' absorbs the marginal term $\log q(\boldsymbol{x}')$, while anchor-only terms are collected in $c'(\boldsymbol{x})$. Using the Gaussian transition density and absorbing constant factors into the positive definite matrix Λ, we obtain, for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{U}$,

$$-\left\| \widehat{F}(r(\boldsymbol{x})) - r(\boldsymbol{x}') \right\|_2^2 = -(F(\boldsymbol{x}) - \boldsymbol{x}')^\top \Lambda (F(\boldsymbol{x}) - \boldsymbol{x}') + c'(\boldsymbol{x}). \tag{23}$$

Taking the mixed derivative with respect to $\boldsymbol{x}$ and $\boldsymbol{x}'$ yields

$$J_r(\boldsymbol{x}')^\top J_{\widehat{F}}(r(\boldsymbol{x})) J_r(\boldsymbol{x}) = \Lambda J_F(\boldsymbol{x}),$$

up to the same harmless positive scalar absorbed into Λ. The right-hand side is invertible because $\Lambda \succ 0$ and F is a diffeomorphism. Hence $J_r(\boldsymbol{x}')$ is full rank. Moreover, for any two points $\boldsymbol{x}_1', \boldsymbol{x}_2'$, fixing $\boldsymbol{x}$ and comparing the two identities gives

$$J_r(\boldsymbol{x}_1')^\top J_{\widehat{F}}(r(\boldsymbol{x})) J_r(\boldsymbol{x}) = J_r(\boldsymbol{x}_2')^\top J_{\widehat{F}}(r(\boldsymbol{x})) J_r(\boldsymbol{x}).$$

The same identity also implies that the middle-right factor is invertible, so $J_r(\boldsymbol{x}_1') = J_r(\boldsymbol{x}_2')$. Thus J_r is constant on the connected region $\mathcal{U}$, and r is affine:

$$r(\boldsymbol{x}) = L\boldsymbol{x} + \boldsymbol{b}, \qquad L \in \mathrm{GL}(d).$$

It remains to identify the dynamics. Multiplying (23) by -1, both sides are strictly convex in $\boldsymbol{x}'$. The right-hand side has unique minimizer $\boldsymbol{x}' = F(\boldsymbol{x})$. Since r is affine and invertible, the left-hand side has unique minimizer satisfying $r(\boldsymbol{x}') = \widehat{F}(r(\boldsymbol{x}))$. The identity holds pointwise in $\boldsymbol{x}'$, so the minimizers coincide:

$$\widehat{F}(r(\boldsymbol{x})) = r(F(\boldsymbol{x})).$$

Writing $r(\boldsymbol{x}) = L\boldsymbol{x} + \boldsymbol{b}$, we obtain

$$\widehat{F}(\boldsymbol{z}) = LF\left(L^{-1}(\boldsymbol{z} - \boldsymbol{b})\right) + \boldsymbol{b}, \qquad \boldsymbol{z} \in r(\mathcal{U}),$$

which concludes the proof. □

Remark (Amortized Denoising)**.** *Theorem 1 offers theoretical guarantees in the ideal setting of the infinite views limit $V \to \infty$ and for an encoder with simultaneous access to the complete set of views. In practice, our model 5 departs from these idealized assumptions and tackles the denoising problem in an amortized setting with a relative small amount of views. We take the successful system identification achieved in our experiments as evidence that such technique works well in practice.*

Remark (Role of multi-step rollouts)**.** *The theorem allows the training objective to average over arbitrary rollout horizons (k_1, k_2). The proof uses only the one-step constituent $(k_1, k_2) = (1, 0)$, because under joint realizability any global minimizer of the full objective also minimizes each active constituent term individually. Thus the additional horizon terms do not weaken the identifiability conclusion but are not needed in the proof. However, in practice, for gradient-based optimizations, these additional terms can impose useful additional constraints that regularize the learning problem. For example, longer rollouts can help in the identification of the deterministic component in the presence of significant latent noise, at the cost of more expensive training iterations.*

B Dynamical Systems Definition

In this section we provide the continuous-time formulation for the dynamical systems evaluated in this work, alongside the model used for the external forcing $\boldsymbol{u}_t$ if present.

Duffing The Duffing oscillator is a non-linear second-order differential equation with external forcing that can equivalently be represented as a system of two first-order ordinary differential equations of the form:

$$\begin{aligned} \dot{x}_t &= v_t \\ \dot{v}_t &= -\delta v_t - \alpha x_t - \beta x_t^3 + \gamma \cos(\omega t) \end{aligned} \tag{24}$$

In our experiment we fixed $\alpha = -1$, $\beta = 1$, $\delta = 0.3$, $\gamma = 0.3$ and $\omega = 1.2$. Note that in this case $\boldsymbol{u}_t = \cos(\omega t)$ and $C = [0, \gamma]^\top$. We fixed $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon \mathrm{Id})$, $\sigma_\epsilon = 0.05$ and used as integration time-step $\Delta t = 0.05$ with integration horizon $\kappa = 8$.

Lorenz The Lorenz chaotic attractor is a system of three first-order ordinary differential equations and it is a classic system that exhibits chaotic behavior. It is defined as:

$$\begin{aligned} \dot{x}_t &= \sigma(y_t - x_t) \\ \dot{y}_t &= x_t(\rho - z_t) - y_t \\ \dot{z}_t &= x_t y_t - \beta z_t \end{aligned} \tag{25}$$

In our experiment we fixed $\rho = 28$, $\beta = 8/3$ and $\sigma = 10$. The latent noise was $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon \mathrm{Id})$, $\sigma_\epsilon = 0.5$ and used as integration time-step $\Delta t = 0.01$ with integration horizon $\kappa = 8$.

FitzHugh-Nagumo The FitzHugh-Nagumo dynamical system is a relaxation oscillator and represents a prototype of an excitable system (e.g. a neuron). It is a forced dynamical system and is defined as:

$$\begin{aligned} \dot{v}_t &= v_t - \frac{v_t^3}{3} - w_t + I_{\text{ext}} \\ \tau \dot{w}_t &= v_t + a - b w_t \end{aligned} \tag{26}$$

In our experiment we fixed $a = 0.7$, $b = 0.8$, $\tau = 12.5$ and we model the external input current I_{ext} as a square wave with random amplitudes $|I_{\text{ext}}| \sim U_{[0,1]}$, period $T \sim U_{[8,10]}$ and 50% duty-cycle. The latent noise was $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon \text{Id})$, $\sigma_\epsilon = 0.02$ and we used as integration time-step $\Delta t = 0.05$ with integration horizon $\kappa = 8$. Note that the Van-der-Pol oscillator is a special case of the FitzHugh–Nagumo model, with $a = b = 0$ and $C = [1, 0]^\top$.

Winner Take All The winner-take-all dynamical systems are a class of system defined by the property of state self-excitation paired with reciprocal inhibition. We consider the following particular form the system:

$$\dot{\boldsymbol{z}}_t = \boldsymbol{z}_t \left(\zeta - \boldsymbol{z}_t^2\right) - A\boldsymbol{z}_t + C\boldsymbol{u}_t, \tag{27}$$

where $\zeta = 1$ and A is the interaction matrix which we take to be a matrix of all ones with a zero diagonal. In our experiments the state $\boldsymbol{z}_t$ had dimensionality $\boldsymbol{z}_t \in \mathbb{R}^3$. The external forcing $\boldsymbol{u}_t \in \mathbb{R}^d$ was again a random square wave with amplitude $|I| \sim U_{[0.1,0.2]}$, period $T \sim U_{[10,20]}$, 50% duty-cycle and only one random component active per-cycle. We fixed $C = \text{Id}$. The latent noise was $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon \text{Id})$, $\sigma_\epsilon = 0.01$ and we used as integration time-step $\Delta t = 0.05$ with integration horizon $\kappa = 8$. This particular system is often used to model competitive dynamics (*e.g.* forced-choice decision-making in neuroscience).

Double-Well The double-well is a forced bistable system defined by the equation:

$$\begin{aligned} \dot{x}_t &= -\gamma \frac{\partial V}{\partial x} + \kappa \frac{\partial V}{\partial y} \\ \dot{y}_t &= -\gamma \frac{\partial V}{\partial x} - \kappa \frac{\partial V}{\partial y} + u_t \end{aligned} \tag{28}$$

$$V = \frac{x_t^4}{4} - \frac{x_t^2}{2} + \frac{y_t^2}{2}$$

where V is a potential function and we fixed $\gamma = 0.7$ and $\kappa = 1.2$. As forcing function u_t we used a square wave with amplitude $I \sim U_{[-2,+2]}$, period $T \sim U_{[8,20]}$ and 50% duty-cycle. For this system the control matrix is $C = [0, 1]^\top$. The latent noise was $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon \text{Id})$, $\sigma_\epsilon = 0.05$ and we used as integration time-step $\Delta t = 0.02$ with integration horizon $\kappa = 8$.

Stuart-Landau The Stuart-Landau system is a non-linear oscillator defined via the equation:

$$\begin{aligned} \dot{x}_t &= \mu x - \omega y - x\left(x^2 + y^2\right) \\ \dot{y}_t &= \omega x + \mu y - y\left(x^2 + y^2\right) \end{aligned} \tag{29}$$

In our experiment we fixed $\mu = \omega = 1$. The latent noise was $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon \text{Id})$, $\sigma_\epsilon = 0.01$ and we used as integration time-step $\Delta t = 0.02$ with integration horizon $\kappa = 8$.

Heteroclinic An heteroclinic orbit is defined as a path in phase space that connects two saddle points. By heteroclinic system we identify here a metastable system defined as:

$$\begin{aligned} \dot{x} &= x\left(1 - x - \alpha y - \beta z\right) \\ \dot{y} &= y\left(1 - y - \alpha z - \beta x\right) \\ \dot{z} &= z\left(1 - z - \alpha x - \beta y\right) \end{aligned} \tag{30}$$

where we fixed $\alpha = 2$ and $\beta = \frac{1}{2}$. The latent noise was $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon \text{Id})$, $\sigma_\epsilon = 0.01$ and we used as integration time-step $\Delta t = 0.02$ with integration horizon $\kappa = 8$.

C Symbolic Recovery from Affine Orbits

We offer here a proof-of-concept application for the symbolic recovery of a latent dynamical system via identification of the sparsest representative of the system's coefficient representation along the affine group orbit.

We illustrate this on the Lorenz system in a noiseless setting as a proof of concept; a systematic study of symbolic recovery under noise, including the development of robust gauge-aware sparse

regression, is an interesting direction for future work. In a noiseless configuration ($\sigma = 0$), our system achieves very accurate dynamics reconstruction ($\mathrm{dyn}R^2 = 99.11\%$) and we solved the optimization problem 17 via an L_1 proxy for sparsity with iterative pruning of small coefficients $|\theta_{kl}| \leq 0.1$. We solved 200 instances of the problem with different random initial conditions for both L and $\boldsymbol{b}$. Here we report the best matching solution found.

Ground Truth System	**Symbolic Recovery**
$\dot{x}_t = -10x_t + 10y_t$	$\dot{x}_t = -10.0x_t + 2.9y_t + 0.13x_t y_t$
$\dot{y}_t = 28.0x_t - 1.0x_t z_t - 1.0y_t$	$\dot{y}_t = -0.12x_t y_t - 2.1x_t z_t + 28.0x_t - 0.2y_t z_t + 5.8y_t$
$\dot{z}_t = 1.0x_t y_t - 2.7z_t$	$\dot{z}_t = -0.8x_t^2 + 0.9x_t y_t + 0.82x_t - 4.3z_t$

While the overall structure is not far from the ground-truth, the equations contain several small spurious terms with coefficients below 1, indicating that exact term recovery is not yet achieved and requires more advanced methods than simple L_1 thresholding within the gauge. Experimentally, we observed how minor imperfections in the flow field can hinder the symbolic regressor problem, which implicitly relies on exact term cancellations. Future work should experiment with more advanced frameworks, such as a paired teacher-student setup where a student network might learn to deviate from the exact affine group orbit in pursuit of significant sparsity gains. Additionally, more faithful approximations to the sparsity optimization problem 17, *e.g.* better L_0 approximations Louizos et al. (2018); Yang et al. (2020); Kolb et al. (2025), might also be used to improve upon these initial attempts.

D Ablation Study on Integration Horizon

One of the flexible components of the DYSCO algorithm 5 is the maximal integration time-horizon κ. This parameter controls how far along the flow field the latent states are evolved before being compared via the similarity score. We argued in our second remark in Appendix section A how longer integration horizons might be beneficial in practice, despite Theorem 1 claiming identification even for a single step of forward time-integration. Here we empirically investigate the impact of this hyperparameter on the model performances.

We consider the Lorenz system under Gaussian observation noise $\xi_t^a \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$ with intensity $\sigma = 0.3$. We trained our standard architecture (4-layer MLP encoder, polynomial functional basis, see Methods for details) for 200 epochs and measured both R^2 and $\mathrm{dyn}R^2$ metrics on the held-out validation set every 10 epochs. We explored the following time integration horizons $\kappa = 2, 4, 6, 8, 10$ and report our results in Figure 6. We note that indeed longer horizons result in improved quality for both latent trajectories and flow fields, with a saturating (or peaking) effect after $\kappa = 8$. In practice, however, longer integration horizons come with a direct computational cost: on our hardware (NVIDIA RTX A4000) a full training run with $\kappa = 2$ lasted $\sim$ 50min, while an identical configuration for $\kappa = 10$ took $\sim$ 2h 20min; with the majority of the overhead coming from the ODEINT rollout. Furthermore, we noted how longer rollouts ($\kappa = 10$) tended to suffer from training instability, probably due to the inherent difficulty of bootstrapping flow-fields from random initializations that are stable for longer horizons. For these reasons, we fixed $\kappa = 8$ for our main experiments.

E Controlled Baseline Protocol and Training Budgets

The controlled comparisons use fixed noisy Lorenz and Stuart–Landau observations with one, two and four views; Figure 3 displays one and four views. DYSCO uses view correspondences, while the baselines receive the same observations as additional trajectories. Gaussian observation noise has $\sigma = 1.0$ for Lorenz and $\sigma = 0.2$ for Stuart–Landau. The Poisson condition uses the same raw observations across methods at each view count.

Baseline Methods. DCL uses a differentiable switching-linear model with 200 modes and $20,000$ negative samples. SINDy-AE uses a degree-three polynomial library, encoder widths $64/32$, and its reconstruction/dynamics objectives. RP-GSSM uses a Gaussian latent state-space model with a

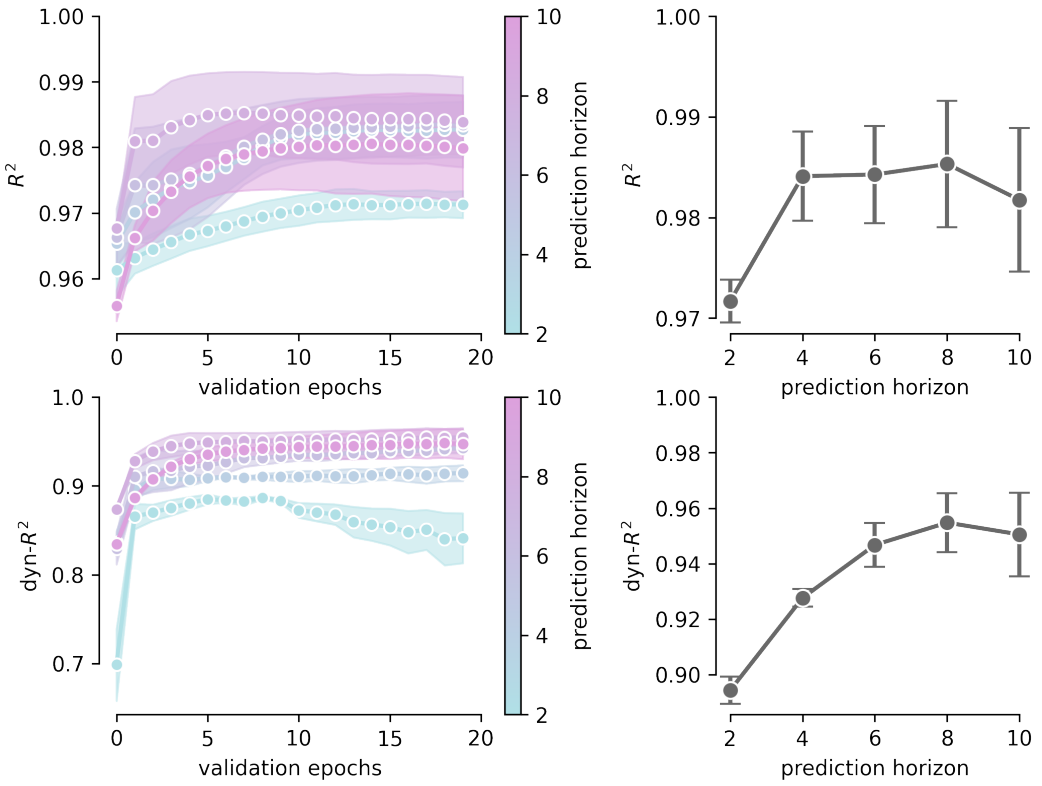


Figure 6: **Ablation on time horizon** κ. Ablation result for varying time integration horizons $\kappa = 2, 4, 6, 8, 10$ on the Lorenz system with Gaussian noise of intensity $\sigma = 0.3$. We trained for 200 epochs and evaluated the model on the validation set every 10 epochs (left column). We report the maximum achieved score on the right column for both the R^2 metric (top row) and $\mathrm{dyn}R^2$ metric (bottom row). Each marker is the average over 3 runs $\pm$ standard error of the mean.

nonlinear recognition network, sequence length 100, and stabilized transition parameters. Architectures, derivative preprocessing, batch units and optimization differ across methods. The comparison controls raw observational information, with additional alignment metadata supplied to DYSCO.

Architectures and configured schedules. The coefficient network generates one global coefficient matrix from a learned source vector. In these controlled comparisons, Lorenz DYSCO uses source dimension 32, hidden widths $128/128$, a quadratic library and learning rate 10^{-4}; Stuart–Landau uses source dimension 32, widths $64/64$, a cubic library and learning rate 3×10^{-4}. Lorenz configurations use 200 or 500 epochs, and Stuart–Landau configurations use 200 epochs. The neural model's active polynomial term uses source dimension 8 and one width-32 hidden layer. Table 2 summarizes the configured schedules.

Method	Configured schedule	Batch size	Learning rate
DYSCO	200 or 500 epochs	4096	10^{-4} or 3×10^{-4}
DCL	30,000 iterations	2048	3×10^{-4}
SINDy-AE	10,001 epochs + 1,001 refinement	8000	10^{-3}
RP-GSSM	10,000 iterations	32 sequences	10^{-3}

Table 2: Configured baseline training budgets from 149 run configurations. Batch units and per-step objectives vary across methods; RP-GSSM uses sequences of length 100.

F Motor-Cortical Data and Analysis Protocol

Data and fitting. We use MC_Maze motor-cortical spike counts from the Neural Latents Benchmark (Pei et al., 2021), in 20-ms bins over $[-250, 450)$ ms relative to movement onset. Conditions are defined by trial type and version, retaining groups with sufficient repetitions for four-view training. The split contains 1,721 training and 574 disjoint validation trials across 108 matched conditions, with 137 observed neurons. We learn six latent coordinates and a degree-two polynomial field with a linear observed-input term. The go cue is rendered as a Gaussian pulse with width parameter 40 ms. The input transform combines Freeman–Tukey variance stabilization, a learned five-bin convolution and batch normalization. We trained for 500 epochs and used the saved validation-loss-selected checkpoints for analysis. Behavioral variables are used for evaluation.

Rotational analysis. We adapt the jPCA analysis of Churchland et al. (2012) over $[-50, 150)$ ms. After subtracting the cross-condition mean separately at each time, we fit the six-PC transform and skew-symmetric linear generator on training condition means. The dominant plane corresponds to the largest imaginary eigenfrequency. We orient its axes using training data to maximize initial spread along jPC1 and make net rotation counterclockwise. Plotted curves in Figure 5 use validation repetitions, colored by their displayed starting jPC1 coordinate. Gaussian smoothing ($\sigma = 20$ ms, truncated at three standard deviations) is applied to the full projected time series before cropping, solely for display. Fitted axes and quantitative results use unsmoothed arrays.

Table 3 reports the numerical value of the neural analysis for the skew/full ratios and the measured rotational frequencies. DYSCO uncertainties summarize three independent initializations, while neural PCA is a deterministic same-data control.

Quantity	DYSCO	Neural PCA	Churchland et al.
Six-dimensional skew/full ratio	0.848 ± 0.017	0.622	$\approx 0.820 \pm 0.076$ ($n = 9$)
Dominant-plane skew/full ratio	0.970 ± 0.017	0.820	$\approx 0.894 \pm 0.053$ ($n = 7$)
Angular frequency (rad/s)	6.057 ± 0.993	7.419	A: 8.8 / 9.8; B: 12.2 / 13.5

Table 3: **Rotational analysis of MC_Maze.** DYSCO: mean $\pm$ sample SD over three seeds; neural PCA: a deterministic same-data control. Values for the neural data are extracted from Churchland et al. (2012), Fig. 6b: mean $\pm$ sample SD across nine six-dimensional and seven best-plane supplied points, summarizing variation among the supplied points. Angular frequencies are reported for monkeys A/B, fast/slow, while MC_Maze here is monkey J.

Velocity probes and absolute field accuracy. We employed ridge probes to predict the signed two-dimensional hand velocity at $+100$ ms. We selected regularization strength by sample-wise five-fold cross-validation on the training set, using nine logarithmically spaced values from 10^{-4} to 1 and unscaled features. The final evaluation uses disjoint validation trials from the same conditions. Condition-average probes are fitted separately to training condition means and evaluated on validation condition means. Table 4 reports the results for linear and quadratic features of the state and learned field. The quadratic expansion of state and field includes field squares and state–field interactions, increasing the feature count from 27 to 90. Since the field is a function of the state and observed input, these results characterize the usefulness of its features under each readout class. The learned field's separate full-window held-out derivative R^2 is 0.138 ± 0.005, with cosine agreement 0.355 ± 0.002.

Projected field and nearby flows. For the geometric illustration in Figure 5a, we use seed 43, selected by training cross-validation of the state-plus-field velocity probe. The PCA projection is fitted to training condition means and applied to all 108 validation condition means. Dark arrows show 40 ms times the field evaluated on individual validation trials with their observed go-cue inputs, averaged within condition and projected into the same PCs. To illustrate the field around the observed states, we generate nearby starting points around twelve evenly spaced condition indices at -150, -10, $+130$, $+270$ and $+410$ ms. Offsets combine -0.5, 0 and $+0.5$ training-state standard deviations along PC1–3, preserving the other three coordinates at each anchor. Eight starting points are selected using projected separation and open display regions. We integrate the full six-dimensional

Readout features	Features	Single-trial R^2	Condition-mean R^2
State	6	0.267 ± 0.014	0.328 ± 0.017
State and field	12	0.402 ± 0.015	0.524 ± 0.014
Quadratic state	27	0.545 ± 0.015	0.705 ± 0.016
Quadratic state and linear field	33	0.545 ± 0.015	0.705 ± 0.014
Quadratic state and field	90	0.649 ± 0.015	0.848 ± 0.015

Table 4: **Velocity decoding from state and field features.** Mean $\pm$ sample SD over seeds 42–44, with the same trials and $+100$-ms velocity targets in each comparison. Quadratic features include all linear, square and pairwise-product terms. The 33-feature readout appends the six field coordinates to the quadratic state library; the 90-feature readout expands all twelve state and field coordinates jointly. Ridge intercepts are fitted separately.

field from these points for 80 ms in both directions, with the go-cue input fixed at $u = 0$, using fixed-step RK4 with a 2.5-ms step. The grey curves and their arrows show these nearby flows, with all six coordinates evolving during integration.

Symbolic field in a rotational gauge. To better illustrate the potential application of DYSCO in analyzing the symbolic form of the fitted dynamics, we inspected the same frozen seed-43 model to see whether a clean rotational structure was evident in the dynamics itself. Given the gauge-dependence of the symbolic form (Appendix C), we seeked a particular gauge where the rotational form might become immediately evident: a natural choice is the rotation identified by the jPCA construction itself. In practise, we fixed an orthogonal basis $U \in \mathbb{R}^{6\times 6}$, whose first two columns span the training-fitted jPCA plane, with the orientation described in the main text, while for the remaining columns we picked the principal directions of training-state variation in the orthogonal complement. With $\boldsymbol{m}$ the mean of all training states, we defined:

$$\boldsymbol{q} = U^\top(\boldsymbol{z} - \boldsymbol{m}), \qquad \widehat{f}_q(\boldsymbol{q}, u) = U^\top \widehat{f}(\boldsymbol{m} + U\boldsymbol{q}, u). \tag{31}$$

The fixed origin and basis preserve the autonomous part of the field under a time-independent affine change of coordinates. The quadratic model can therefore be written as:

$$\widehat{f}_q(\boldsymbol{q}, u) = \boldsymbol{c} + A\boldsymbol{q} + \sum_{1\le i\le j\le 6} \boldsymbol{b}_{ij} q_i q_j + \boldsymbol{d}\, u, \tag{32}$$

with A identifying the linear part of the field in this particular gauge (see Figure 7a), while the remaining terms capture the nonlinear and input-dependent contributions. The linear part of the field is also reported in the main text in Eq. (18), where we rounded the coefficients to two significant figures for display.

This linear terms shows a clear rotational structure in its dominant 2×2 block (q_1 and q_2), with the antisymmetric part of this block being characterized by a rate $(A_{21} - A_{12})/2 = 5.13$ rad/s. This block describes the local linear contribution at the training mean, where the field has nonzero drift $\boldsymbol{c}$, which differs from the jPCA frequency in Table 3.

The remainders in Eq. (18) contain c_i, linear couplings to $q_3, \ldots, q_6$, all quadratic terms and the observed-input response $d_i u$. We then quantified the relative importance of these terms by examining their contribution to the overall field. Completly omitting them gives relative RMS errors of 73.6% and 25.3% in the first two field components at held-out states with $u = 0$. Omitting every dependence on $q_3, \ldots, q_6$ gives 26.7% relative RMS error in the two-component field, indicating a measurable but not dominant contribution of the remaining four coordinates to the rotational dynamics.

Progressive term removal. To quantify the cost of simplifying the equations, we rank the 168 scalar coefficients of the state library (one constant, six linear and 21 quadratic terms in each of six equations) by their RMS contribution on training states. For coefficient θ_{ij} multiplying monomial ϕ_j, the relative contribution is:

$$s_{ij} = \frac{|\theta_{ij}|\sqrt{\langle \phi_j(\boldsymbol{q})^2 \rangle_{\text{train}}}}{\max_k |\theta_{ik}|\sqrt{\langle \phi_k(\boldsymbol{q})^2 \rangle_{\text{train}}}}. \tag{33}$$

We remove coefficients in ascending order of s_{ij}, one at a time, keeping all six input coefficients fixed and performing no refitting. For each of the 169 retained-count settings, Figure 7b reports the relative RMS difference from the full learned field at validation states with $u = 0$. We evaluate both the complete field and its first two components (which by construction define the rotating plane):

$$E_S(K) = \frac{\left\| P_S\left(\widehat{f}_q^{(K)} - \widehat{f}_q\right)\right\|_{F,\mathrm{val}}}{\left\| P_S \widehat{f}_q \right\|_{F,\mathrm{val}}}, \qquad S \in \{\{1,\dots,6\},\{1,2\}\}, \tag{34}$$

where K is the number of retained state-library coefficients and P_S selects the indicated components. For example, retaining coefficients above 5% of the largest training RMS contribution in each equation leaves 112 coefficients and gives 10.8% plane-field error; a 10% threshold leaves 80 coefficients and gives 15.8% error. These errors measure the approximation of the fitted polynomial field, with all retained contributions summed before evaluation. The resulting curves indicate that the learned field can be substantially simplified at a modest cost in accuracy, with the first two rotating components being somewhat more robust to simplification than the full six-dimensional field.

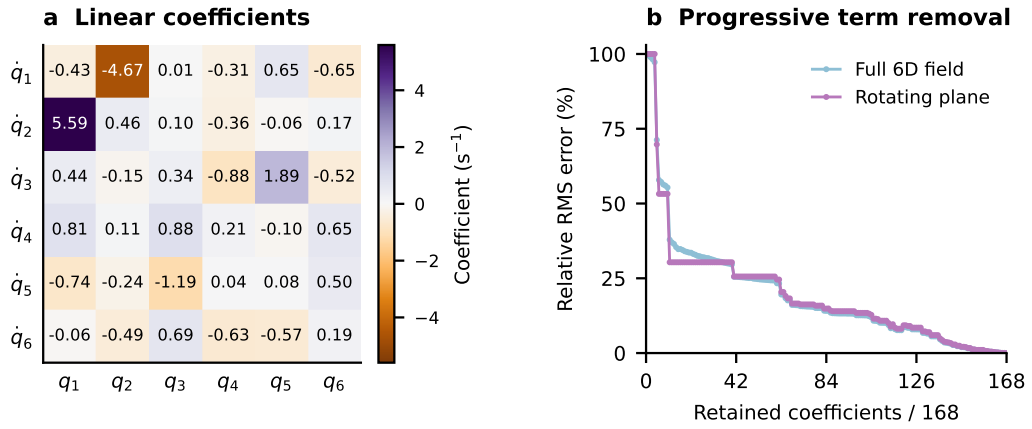


Figure 7: **Symbolic structure of the learned neural field.** Seed 43 in the fixed coordinates of Eq. (31). (**a**) Full linear coefficient matrix A, with entries rounded for display. (**b**) Relative RMS approximation error at held-out states with $u = 0$ as individual coefficients are removed in ascending order of their relative training RMS contribution. Each curve contains all 169 settings, from 168 retained state-library coefficients to zero. Cyan evaluates all six field components; orchid evaluates the two rotating components. Input coefficients remain fixed.

G Forward-Only Objective and Overcomplete-Library Diagnostics

Forward-only objective. We validated the theoretical prediction from Theorem 1 that a forward-only loss is sufficient for identifying the underlying dynamics. We considered the four-view Gaussian Lorenz system and compared forward-only training (at horizon $k = 8$) against the full DYSCO objective. We report results in Table 5 for three seed-matched 200-epoch runs. Indeed, we find that forward-only training is sufficient for recovering the system; althought the additional terms of the full DYSCO objective improve field recovery accuracy in this experiment, indicating how their inclusion provides additional valuable training signal.

Objective	n	Latent R^2 (%)	Field dynR^2 (%)
Full objective	3	98.0 ± 1.2	95.1 ± 2.1
Forward only ($k = 8$)	3	96.0 ± 0.8	87.5 ± 2.1

Table 5: Four-view Gaussian Lorenz objective ablation. Mean $\pm$ SEM over seeds 42–44; the forward-only variant retains only the forward/encoded contrastive term. The extra objective terms improve field fit in this setting; the theorem uses an active forward one-step constituent.

Overcomplete-library diagnostic. As a preliminary assessment of DYSCO resilience against symbolic library misspecification, we measured the effects of an overcomplete-library by comparing a quadratic and cubic field library on the Lorenz system at a matched seed. A cubic library doubles the coefficient count for this system from 30 to 60, resulting in a more complex model to optimize. We train the two DYSCO configurations for 200 epochs and report the results in Table 6. We observe that the cubic library yields a lower field precision in this single-run comparison, although the quality degradation is gradual and not catastrophic.

Library	Coefficients	Latent R^2 (%)	Field dynR^2 (%)
Quadratic	30	99.15	97.22
Cubic	60	97.21	85.94

Table 6: Overcomplete-library diagnostic on four-view Gaussian Lorenz: single matched seed 42, trained for 200 epochs. Entries are the scores of the individual runs.

H Autonomous Lorenz Climate

To assess the long-term stability of the learned DYSCO dynamics, we evaluate our stored DYSCO checkpoints on the Lorenz system for additional 50-time-unit autonomous rollouts from 24 held-out initial states per model, using the already-computed affine alignment and a 10-time-unit burn-in for attractor statistics. Figure 8a,b shows representative long-term trajectories for the Lorenz system with Gaussian and Poisson noise, respectively. Figure 8c,d shows the corresponding power spectra and survival rates of the trajectories (i.e. time for which the trajectories remains stable). Table 7 reports values for all seven available four-view checkpoints evaluations.

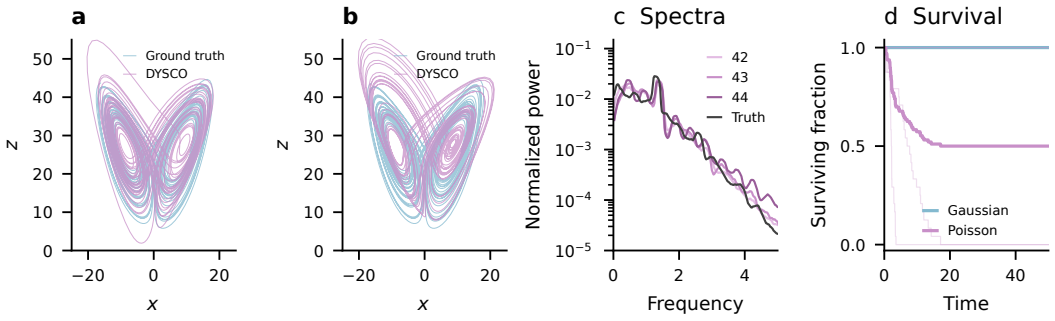


Figure 8: **Autonomous Lorenz dynamics after four-view training.** (**a,b**) Ground-truth and affine-aligned learned trajectories for Gaussian seed 42 and selected successful Poisson seed 45, respectively; first cached initial condition, after a 10-time-unit burn-in. Chaotic trajectories can diverge in phase. (**c**) Gaussian power spectra for seeds 42–44 (orchid shades), averaged over surviving trajectories and coordinates with Hann-window/nine-bin smoothing. (**d**) Survival among 24 initial conditions over 50 time units: thick solid lines average three Gaussian seeds (42–44) and four Poisson seeds (42–45); thin transparent solid lines show individual seeds.

Noise	Seed	dynR^2	Survival	Overlap	Spectrum H	Switch rate
Gauss	42	0.972	100%	0.502	0.153	0.759
Gauss	43	0.971	100%	0.562	0.192	0.864
Gauss	44	0.909	100%	0.333	0.191	0.934
Poisson	42	0.939	0%	0.000	—	—
Poisson	43	0.518	100%	0.038	0.784	1.928
Poisson	44	0.934	0%	0.238	—	—
Poisson	45	0.965	100%	0.372	0.181	0.827

Table 7: Four-view Lorenz autonomous rollouts over 50 time units. All available four-view climate evaluations are shown; seed 45 is an additional Poisson run. Statistics discard the first 10 time units; overlap is $1 - \mathrm{TV}$ (higher is better), H is spectral Hellinger distance (lower is better). Dashes indicate unavailable spectral/switch statistics after early escape. Gaussian ground-truth switching rate is 0.579 per time unit. High local field accuracy does not guarantee autonomous stability.

For the case of Gaussian noise, we measured local dynR^2 0.951 ± 0.036 with survival rate hitting 100% and attractor occupancy overlap of 0.466 ± 0.119 and spectral Hellinger distance of 0.179 ± 0.022 (mean $\pm$ SD). The lobe-switching rate is 0.852 ± 0.088 per time unit, compared with 0.579 for the matched ground-truth cohort. Together this metrics indicate a good qualitative long-term behavior of the learned dynamics, although some appreciable statistical differences remain.

For the case of Poisson noise we observed how some seeds yield unstable long-term dynamics, resulting in less than 100% survival. The surviving seeds yield local dynR^2 0.950 ± 0.024, attractor occupancy overlap of 0.205 ± 0.167, spectral Hellinger distance of 0.483 ± 0.429 and lobe-switching rate of 1.378 ± 0.556 per time unit, indicating a more significant deviation from the ground-truth dynamics. This preliminary analysis suggests that the Poisson noise case is more challenging for long-term stability, and that further work is needed to improve the robustness of the learned dynamics under this noise model.

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: The abstract and the introduction present our new algorithm for latent dynamical system identification, claiming both theoretical guarantees and experimental validation. The main theoretical result is presented in section 2, while sections 3 and 4 showcase our experimental results.

2. **Limitations**

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

   Answer: [Yes]

   Justification: We provide a dedicated subsection on the main limitations of our algorithm in our Discussion section 5.

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 a full proof for our Theorem 1 in Appendix A.

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: For all our experiments we provide a dedicated section describing the data generating process, model architecture and optimizer used (section 3). In Appendix B we provide a full description of the dynamical systems considered in our experiments with corresponding parameters. Furthermore, we will fully open-source the code upon acceptance.

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 the source code used in our experiments, together with configuration files and training scripts for ease-of-use.

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) necessary to understand the results?

   Answer: [Yes]

   Justification: In our Section 3 we describe our full data generating process, model architecture, dataset size, training details (optimizers, used hardware, approximate training time). Additionally Appendix B collects all the parameters for each dynamical systems considered.

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: For our two core metrics (R^2 and $\mathrm{dyn}R^2$) we provide average with standard deviations computed over 3 runs in both our main results (Table 1 and **??**) and our ablations (Figure 4, Figure 6).

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: In our **Training details** subsection of Section 3 we provide details about the hardware used to run all our experiments and the approximate training time required for a typical experiment.

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: We conform with the NeurIPS Code of Ethics.

10. **Broader impacts**

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

    Answer: [N/A]

    Justification: Our main contribution is the introduction of a novel algorithm for latent dynamical system identification. Given the foundational nature of this research, we do not foresee potential negative societal impacts, while the potential benefits (neuroscience applications) are currently far downstream.

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., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper poses no such risks.

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: In our experiments we use several existing open python libraries (e.g. TORCHDIFFEQ), which we properly cite.

13. **New assets**

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

    Answer: [N/A]

    Justification: The paper does not release new assets.

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: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

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: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method developed in this research does not involve LLMs as any important components.