Provably Stable Neural Dynamics via Koopman Operator Certificates

Aryan Dadwal [1]

Abstract

Learning neural forward models of dynamical systems that remain stable over long rollout horizons is a fundamental challenge in scientific computing and physics-informed machine learning. We introduce Koopman-Stable Neural Dynamics, a deep Koopman architecture that learns a finite-dimensional, stable latent approximation of Koopman-style dynamics. The latent transition operator is parameterized as the matrix exponential of a continuous-time generator $G = -S + A$, where $S \succ 0$ is symmetric positive definite and A is skew-symmetric. This yields a native stability certificate: $V(z) = \|z\|^2$ is a strict latent Lyapunov function. We prove that this latent certificate transfers to practical Input-to-State Stability in the original state space under bi-Lipschitz conditions on the encoder–decoder pair, and establish a limitation theorem characterizing when strictly stable reduced representations cannot exactly capture a given dynamical system. Experiments on four benchmarks—the Duffing oscillator, an unstable saturating node, the Lorenz attractor, and the 1D viscous Burgers' PDE—demonstrate that the certified model maintains bounded predictions where unconstrained baselines diverge catastrophically.

1. Introduction

Learning forward models of dynamical systems from data is central to many areas of physics, from fluid dynamics and plasma simulation to climate modeling and model-based control (Nagabandi et al., 2018; Pathak et al., 2018; Hafner et al., 2019). Deep Koopman networks (Lusch et al., 2018; Li et al., 2017) and neural ODEs (Chen et al., 2018) are powerful paradigms for learning such dynamics, but they lack formal guarantees that long-horizon rollouts remain bounded. In practice, even a spectral radius marginally exceeding unity causes exponential divergence over hundreds of time steps, rendering models useless for downstream tasks in experimental physics or engineering control.

Koopman operator theory (Koopman, 1931) provides a principled framework: any nonlinear system admits a (possibly infinite-dimensional) linear representation in a lifted function space. Data-driven approximations such as EDMD (Williams et al., 2015) and deep Koopman variants learn finite-dimensional projections of this operator. Yet these methods treat stability as an afterthought—either hoping the learned operator is stable, or adding soft spectral penalties that provide no hard guarantees and remain brittle across initializations.

In the broader Koopman theory literature, learning finite-dimensional linear representations relies on finding appropriate observables and understanding the limits of exact embeddings. A faithful representation perfectly separates states, while a reduced representation discards information to maintain a finite-dimensional or stable bottleneck. We view our learned latent coordinates as an approximate reduced Koopman representation whose primary role is to preserve long-horizon boundedness, rather than providing an exact finite-dimensional linearization of arbitrary dynamics.

Physics-informed approaches that embed structural priors—Hamiltonian networks (Greydanus et al., 2019), Lagrangian networks (Cranmer et al., 2020), or Lyapunov-penalized models (Kolter & Manek, 2019)—have advanced this agenda, but they either restrict the model class too severely or rely on soft constraints that can be violated at test time. A gap remains for architectures that guarantee stability by construction while retaining sufficient expressivity for complex physical systems.

Contributions. We make the following contributions:

1. **Native stability certificate.** We parameterize the Koopman operator as $K = \exp(\Delta(-S + A))$ with $S \succeq \varepsilon I$ and $A^\top = -A$, guaranteeing $\rho(K) < 1$

[1]Indian Institute of Technology, Jodhpur. Correspondence to: Aryan Dadwal <b24bs1070@iitj.ac.in>.

Accepted at ICML 2026 Workshop on Decision-Making from Offline Datasets to Online Adaptation: Black-Box Optimization to Reinforcement Learning., Seoul, South Korea. 2026. Copyright 2026 by the author(s).

by construction. The identity-metric quadratic form $V(z) = \|z\|^2$ serves as a parameter-free Lyapunov function with contraction rate $\alpha = e^{-2\varepsilon\Delta}$.

2. **Conditional stability transfer.** We prove that latent ISS transfers to practical ISS in the original state space under bi-Lipschitz conditions on the encoder–decoder, with an explicit error bound decomposing into a Koopman residual and a reconstruction error.

3. **Limitation theorem.** We prove that strictly stable finite-dimensional reduced representations cannot exactly capture any system whose attractor is not a single equilibrium—including limit cycles and strange attractors.

4. **Empirical validation.** We demonstrate on four benchmarks, including the 1D viscous Burgers' PDE (d_x=64), that the certified model maintains bounded trajectories where unconstrained baselines and neural ODEs diverge.

2. Related Work

Koopman operator methods. Since the operator-theoretic revival for dynamical systems highlighted the role of spectral properties and eigenfunctions (Mezić, 2005), data-driven approximations have become prominent. Classical methods like Dynamic Mode Decomposition (DMD) (Schmid, 2010) and EDMD (Williams et al., 2015) use fixed polynomial or time-delayed observables, while Generalized Laplace Analysis computes spectra directly from trajectories (Mohr & Mezić, 2014). Deep Koopman networks (Lusch et al., 2018; Takeishi et al., 2017; Li et al., 2017; Morton et al., 2018) learn the dictionary end-to-end but do not enforce stability of the learned operator. Several works constrain K to be orthogonal or contractive via projected gradient descent (Mamakoukas et al., 2023), but these lack the continuous-time generator structure that yields a clean Lyapunov certificate.

Stable neural dynamical models. Stability-constrained neural ODEs have been explored via Lyapunov-based loss penalties (Manek & Kolter, 2019; Kolter & Manek, 2019) and input-convex architectures (Chen et al., 2021). Neural ODEs (Chen et al., 2018) learn continuous-time dynamics but provide no stability guarantees. These approaches either (a) rely on soft penalties that can be violated at test time, or (b) severely restrict the model class. Our generator parameterization $G = -S + A$ occupies a deliberate middle ground: it guarantees stability under the identity-metric Lyapunov function $V(z) = z^\top z$—a restricted but practically useful subclass of all Hurwitz-stable matrices—while providing a hard, parameter-free certificate that requires no learned Lyapunov function.

Physics-informed neural networks. Structure-preserving approaches such as Hamiltonian (Greydanus et al., 2019) and Lagrangian (Cranmer et al., 2020) networks embed physical conservation laws directly into the architecture. Our work is complementary: rather than encoding energy conservation, we enforce dissipativity, which is the relevant structural prior for systems with friction, damping, or viscosity—common in fluid dynamics, plasma systems, and condensed matter.

3. Method

3.1. Problem Setup

Consider a discrete-time nonlinear dynamical system $x_{t+1} = f(x_t, u_t)$ where $x_t \in \mathbb{R}^{d_x}$ is the state, $u_t \in \mathbb{R}^{d_u}$ is the control input, and f is an unknown, potentially unstable transition map. Our goal is to learn a model that (i) accurately predicts multi-step rollouts $\hat{x}_{1:T}$ given an initial condition x_0 and input sequence $u_{0:T-1}$, and (ii) provides a formal guarantee that the predicted trajectory remains bounded for arbitrary horizon T.

3.2. Architecture

We lift the state into a latent space $\mathbb{R}^{d_z}$ via a learned encoder $\phi_\theta : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$ and decode via $\psi_\theta : \mathbb{R}^{d_z} \to \mathbb{R}^{d_x}$. The latent dynamics are governed by:

$$z_{t+1} = Kz_t + Bu_t, \qquad \hat{x}_t = \psi_\theta(z_t), \tag{1}$$

where $K \in \mathbb{R}^{d_z \times d_z}$ is the discrete-time Koopman operator and $B \in \mathbb{R}^{d_z \times d_u}$ is a learned input matrix.

Native stability via generator parameterization. We parameterize K as the exact discretization of a continuous-time generator:

$$K = \exp(\Delta\, G), \qquad G = -S + A, \tag{2}$$

where $S = S_{\text{raw}}^\top S_{\text{raw}} + \varepsilon I \succeq \varepsilon I$ is symmetric positive definite, $A = A_{\text{raw}} - A_{\text{raw}}^\top$ is skew-symmetric, $S_{\text{raw}}, A_{\text{raw}} \in \mathbb{R}^{d_z \times d_z}$ are unconstrained learnable parameters, and $\Delta > 0$ is the discretization step. Since all eigenvalues of G have real part $\leq -\varepsilon < 0$, the matrix exponential K has spectral radius strictly less than one. No projection, clipping, or penalty is required.

Equilibrium alignment. To ensure that the origin in the state space maps to the origin in the latent space, we construct ϕ_θ and ψ_θ as bias-free MLPs with ReLU activations. Since $\sigma(0) = 0$ for ReLU, this structurally guarantees $\phi_\theta(0) = 0$ and $\psi_\theta(0) = 0$, eliminating steady-state bias without soft penalties.

Lipschitz control. We apply spectral normalization (Miyato et al., 2018) to every linear layer in ϕ_θ and ψ_θ, bounding the per-layer operator norm to 1. This bounds the upper

Lipschitz constant of the encoder and decoder. To encourage the lower Lipschitz bound (injectivity), we include an isometric regularization term in the training loss.

3.3. Training Objective

The composite loss $\mathcal{L}$ balances four objectives:

$$\mathcal{L} = \alpha\,\mathcal{L}_{\text{fwd}} + \beta\,\mathcal{L}_{\text{rec}} + \gamma\,\mathcal{L}_{\text{lin}} + \delta\,\mathcal{L}_{\text{iso}}, \tag{3}$$

where $\mathcal{L}_{\text{fwd}} = \|x_{1:T} - \hat{x}_{1:T}\|^2$ is the forward prediction loss, $\mathcal{L}_{\text{rec}} = \|x_{0:T} - \psi_\theta(\phi_\theta(x_{0:T}))\|^2$ is the autoencoder reconstruction loss, $\mathcal{L}_{\text{lin}} = \|\hat{z}_{1:T} - \phi_\theta(x_{1:T})\|^2$ is the Koopman invariance loss that minimizes the Koopman residual, and $\mathcal{L}_{\text{iso}} = \mathbb{E}[(\|z_i - z_j\| - \|x_i - x_j\|)^2]$ is the isometric penalty that promotes bi-Lipschitz injectivity of the encoder.

4. Theoretical Results

We state our assumptions, followed by the main theorems. Full proofs are deferred to Section A.

Assumption 4.1 (Lifting and decoder)**.** Let $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ be a compact, forward-invariant set under f. The encoder ϕ_θ and decoder ψ_θ are locally bi-Lipschitz on $\mathcal{X}$: there exist constants $l_\phi, L_\phi, l_\psi, L_\psi > 0$ such that for all $x, x' \in \mathcal{X}$, $l_\phi\|x - x'\| \le \|\phi_\theta(x) - \phi_\theta(x')\| \le L_\phi\|x - x'\|$, and analogously for ψ_θ. The equilibrium $x^* = 0$ satisfies $\phi_\theta(0) = 0$ and $\psi_\theta(0) = 0$.

Assumption 4.2 (Bounded residuals)**.** The Koopman approximation residual $r_t = \phi_\theta(f(x_t, u_t)) - K\phi_\theta(x_t) - Bu_t$ is uniformly bounded: $\|r_t\| \le \bar{r}$ for all $x_t \in \mathcal{X}$. The reconstruction error satisfies $\sup_{x\in\mathcal{X}} \|x - \psi_\theta(\phi_\theta(x))\| \le \bar{e}$.

Theorem 4.3 (Lifted exponential stability)**.** *Under the parameterization (2), the autonomous system $z_{t+1} = Kz_t$ is globally exponentially stable. The function $V(z) = \|z\|^2$ is a Lyapunov function satisfying $V(z_{t+1}) \le \alpha\, V(z_t)$ with contraction rate $\alpha = e^{-2\varepsilon\Delta} < 1$.*

Theorem 4.4 (Input-to-State Stability in the latent space)**.** *The forced system $z_{t+1} = Kz_t + Bu_t$ is ISS. For $\lambda = e^{-\varepsilon\Delta} < 1$:*

$$\|z_t\| \le \lambda^t\|z_0\| + \frac{\|B\|_2}{1-\lambda} \sup_{0\le\tau<t} \|u_\tau\|. \tag{4}$$

Theorem 4.5 (Practical ISS transfer to state space)**.** *Under Assumptions 4.1 and 4.2, the state-space prediction error is uniformly bounded:*

$$\|x_t - \hat{x}_t\| \le \frac{L_\psi \bar{r}}{1-\lambda}(1-\lambda^t) + \bar{e} \le \frac{L_\psi \bar{r}}{1-\lambda} + \bar{e}. \tag{5}$$

The strict stability ($\lambda < 1$) prevents compounding of errors, and the bound decomposes into the approximate Koopman residual ($\bar{r}$) and the autoencoder reconstruction quality ($\bar{e}$).

Remark 4.6 (Scope of the state-space guarantee). The latent certificate (Theorems 4.3 and 4.4) is *unconditional*: it holds for any parameter values by construction. The state-space transfer (Theorem 4.5) is *conditional* on the bi-Lipschitz quality of the encoder–decoder. Our architecture enforces the upper Lipschitz bound via spectral normalization but only encourages the lower bound via $\mathcal{L}_{\text{iso}}$. Enforcing global injectivity would require invertible architectures at significant computational cost.

Theorem 4.7 (Limitation of strictly stable reduced representations)**.** *Let f be a dynamical system possessing a compact invariant set $\mathcal{A}$ that is not a single equilibrium point (e.g., a limit cycle, a strange attractor, or multiple isolated equilibria). Then no exact finite-dimensional Koopman embedding with a continuous single-valued decoder ψ and strictly stable K ($\rho(K) < 1$) exists for f on $\mathcal{A}$.*

This result directly explains the elevated MSE on the Lorenz system (Section 5): the Lorenz attractor is a compact invariant set that is not a single equilibrium, so any strictly stable embedding must incur a nonzero approximation residual $\bar{r} > 0$. The model sacrifices local accuracy to maintain global boundedness—a principled trade-off inherent to reduced representations, which is clearly preferable to catastrophic divergence.

5. Experiments

We evaluate the proposed architecture on four benchmark systems of increasing difficulty, ranging from low-dimensional ODEs to a spatially-discretized PDE. All results are reported as mean $\pm$ standard deviation over 3 random seeds. Full experimental details are in Section B.

5.1. Benchmark Systems

Forced damped Duffing oscillator. A 2D nonlinear oscillator with cubic restoring force: $\dot{x}_1 = x_2$, $\dot{x}_2 = -\delta x_2 - \alpha x_1 - \beta x_1^3 + u$, with $\delta = 0.5$, $\alpha = -1$, $\beta = 1$. This system has a single stable attractor and serves as the primary prediction accuracy benchmark.

Unstable saturating node. A 2D system with a linearly unstable origin stabilized by saturating nonlinearities: $\dot{x}_1 = 0.5x_1 + 0.1x_2 - \tanh(x_1) + u$, $\dot{x}_2 = 0.1x_1 + 0.5x_2 - \tanh(x_2)$. This is a stress test for the stability certificate: unconstrained models must learn to stabilize inherently unstable open-loop dynamics.

Lorenz system. The standard Lorenz attractor ($\sigma = 10, \rho = 28, \beta = 8/3$) with additive forcing. This chaotic system directly probes Theorem 4.7: no finite-dimensional stable linear embedding can exactly represent the strange attractor.

1D viscous Burgers' equation. A spatially-discretized PDE with $d_x = 64$ grid points, periodic boundary conditions,

and viscosity $\nu = 0.02$. This benchmark tests scalability to high-dimensional state spaces where classical EDMD is inapplicable.

5.2. Baselines

We compare five model variants: (1) **Full Certified**: our architecture with $K = \exp(\Delta(-S + A))$, spectral normalization, and bias-free layers; (2) **Soft Penalty**: unconstrained K with a soft spectral penalty $\lambda_{\text{pen}} \max(0, \rho(K) - 0.99)$; (3) **Unconstrained**: unconstrained K with no stability enforcement; (4) **Neural ODE**: a discrete-time residual network $x_{t+1} = x_t + \Delta f_\theta(x_t, u_t)$ trained with MSE; (5) **EDMD**: classical Extended Dynamic Mode Decomposition with degree-2 polynomial dictionary (omitted for Burgers due to dictionary explosion at $d_x = 64$).

5.3. Main Results

Table 1 presents the main results. The key findings are:

The stability certificate prevents catastrophic divergence. On the Lorenz system, the Unconstrained ($\rho(K) = 1.091$), Soft Penalty ($\rho(K) = 1.074$), and Neural ODE baselines all achieve 100% violation rate—every test trajectory diverges to infinity. The Full Certified model maintains bounded predictions (max norm 2.29) with 0% violations, demonstrating the graceful degradation predicted by Theorem 4.7.

The certificate scales to PDE-level dimensionality. On the 64-dimensional Burgers PDE, the certified model achieves the best short-horizon MSE (0.04) with a maximum state norm of only 0.60. The Unconstrained baseline matches short-horizon MSE but drifts to a max norm of 65.7 on longer rollouts. Classical EDMD is inapplicable at this dimensionality due to combinatorial dictionary explosion.

The accuracy–stability trade-off is favorable. On the Duffing system, the certified model achieves comparable short-horizon MSE (0.40 vs. 0.51 for Unconstrained) while maintaining a $1000\times$ lower maximum state norm (1.11 vs. 1097). The Neural ODE achieves the lowest MSE (0.22) on this simple system but offers no stability guarantee.

5.4. Ablation Study

Table 2 validates each component of the architecture:

Generator parameterization is essential. Removing the $G = -S + A$ structure (Unconstrained K) causes the spectral radius to exceed unity (1.017) and the maximum state norm to explode to 79.7, with enormous variance across seeds. The Soft Penalty baseline achieves a low spectral radius (0.990) but provides no hard guarantee—it can be violated under distribution shift.

Spectral normalization controls representation geometry. Without spectral normalization, the Koopman residual magnitude increases nearly $5\times$ (from 0.59 to 2.83) and the maximum state norm doubles, indicating that the encoder inflates the latent representation, amplifying reconstruction errors as predicted by the L_ψ factor in Theorem 4.5.

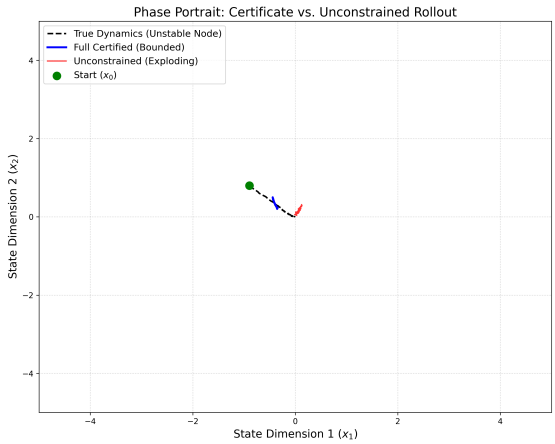

Figure 1. Phase portrait on the Unstable Node system. The Full Certified model (blue) remains bounded near the true trajectory (black dashed), while the Unconstrained model (red) diverges. The stability certificate structurally prevents unbounded rollouts.

Bias-free layers prevent steady-state drift. The "No Eq. Alignment" configuration achieves competitive rollout MSE but exhibits a steady-state error of 0.677 when rolled out from the origin with zero input. All other configurations with bias-free layers achieve exactly zero steady-state error.

6. Limitations

We identify three limitations. **(1) Conditional bi-Lipschitz guarantee.** While spectral normalization bounds the upper Lipschitz constant and the isometric loss encourages the lower bound, a formal guarantee of global injectivity would require invertible architectures (e.g., iResNets), which we leave to future work. **(2) Identity-metric restriction.** The $P = I$ Lyapunov certificate excludes non-normal stable dynamics with large transient growth. Extending to learnable $P \succ 0$ would broaden the model class at the cost of requiring a learned certificate. **(3) Scaling.** The dense matrix exponential is $O(d_z^3)$. Our Burgers PDE experiment ($d_z = 64$) shows this is tractable for moderate dimensions, but scaling to $d_z > 1000$ would require structured approximations (e.g., diagonal or block-diagonal generators).

7. Conclusion

We presented Koopman-Stable Neural Dynamics, a neural forward model with a native stability certificate derived from Koopman operator theory. By parameterizing the latent transition operator via $K = \exp(\Delta(-S + A))$ and

Table 1. Benchmark results across four systems. The Full Certified model maintains 0% violation rate on all systems. Unconstrained baselines, soft penalties, and neural ODEs diverge on the Lorenz system (100% violation). On the 64-dimensional Burgers PDE, the certified model achieves competitive short-horizon MSE while maintaining bounded rollouts. Values: mean ± std over 3 seeds.

System	Model	d_x	MSE (t=50) ↓	Max Norm ↓	$\rho(K)$	Viol. Rate ↓
Duffing	Full Certified	2	0.40 ± 0.08	$\mathbf{1.11 \pm 0.11}$	0.999	0.00
	Soft Penalty	2	0.54 ± 0.09	1.64 ± 0.32	0.979	0.00
	Unconstrained	2	0.51 ± 0.10	1097 ± 491	1.014	0.06
	Neural ODE	2	0.22 ± 0.01	1.49 ± 0.07	—	0.00
	EDMD (deg=2)	2	$\mathbf{0.28 \pm 0.05}$	2.72 ± 0.23	0.991	0.00
Unstable Node	Full Certified	2	0.45 ± 0.14	$\mathbf{1.28 \pm 0.07}$	1.000	**0.00**
	Soft Penalty	2	0.76 ± 0.19	1.53 ± 0.71	0.987	0.00
	Unconstrained	2	0.62 ± 0.15	20509 ± 25392	1.007	0.54
	Neural ODE	2	0.09 ± 0.01	2.55 ± 0.54	—	0.00
	EDMD (deg=2)	2	$\mathbf{0.07 \pm 0.02}$	1645 ± 1878	1.015	0.19
Lorenz	Full Certified	3	286 ± 1	$\mathbf{2.29 \pm 0.27}$	0.999	**0.00**
	Soft Penalty	3	202 ± 11	$> 10^{17}$	1.074	1.00
	Unconstrained	3	216 ± 12	∞	1.091	1.00
	Neural ODE	3	$\mathbf{168 \pm 5}$	$> 10^{12}$	—	1.00
	EDMD (deg=2)	3	298 ± 5	98.8 ± 3.2	1.005	0.00
Burgers PDE	Full Certified	64	$\mathbf{0.04 \pm 0.00}$	$\mathbf{0.60 \pm 0.06}$	1.000	**0.00**
	Soft Penalty	64	0.04 ± 0.00	0.40 ± 0.09	0.909	0.00
	Unconstrained	64	0.04 ± 0.00	65.7 ± 63.2	1.033	0.00
	Neural ODE	64	0.16 ± 0.00	4.82 ± 0.13	—	0.00

Table 2. Ablation study on the Unstable Node system, validating the necessity of each architectural component. Only the biased model exhibits nonzero steady-state error. Values: mean ± std over 3 seeds.

Configuration	$\rho(K)$	Max Norm ↓	Resid. ↓	SS Err ↓
Full Model	1.000	$\mathbf{1.38 \pm 0.13}$	0.59	**0.000**
No Spec. Norm	0.999	2.78 ± 0.22	2.83	0.000
Unconstr. K	1.017	79.7 ± 94.1	18.5	0.000
Soft Pen. K	0.990	1.93 ± 1.10	0.67	0.000
No Eq. Align.	1.000	1.89 ± 0.23	0.44	0.677

enforcing structural constraints—bias-free layers, spectral normalization, isometric regularization—we obtain a model that is provably stable by construction. Our theoretical analysis provides an explicit error bound that decomposes rollout error into a Koopman residual and a reconstruction error, and a limitation theorem that precisely characterizes when and why the certificate incurs an approximation cost. Experiments on four benchmark systems validate that the certificate prevents catastrophic divergence while maintaining competitive prediction accuracy, with particular relevance for long-horizon simulation of physical systems where safety and reliability are non-negotiable.

Impact Statement

This paper presents work whose goal is to advance reliable and physically grounded AI systems for scientific simulation. Provably stable forward models could improve safety in downstream applications such as robotic control and engineering design, where unbounded model predictions can lead to physical harm. We see no specific negative societal consequences that must be highlighted beyond those generally associated with advances in machine learning for physical systems.

References

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Chen, Y.-C., Shi, Y., and Zhang, B. Learning stable deep dynamics models. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., and Ho, S. Lagrangian neural networks. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.

Greydanus, S., Dzamba, M., and Yosinski, J. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Hafner, D., Lillicrap, T., Ba, J., and Norouzi, M. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2019.

Kolter, J. Z. and Manek, G. Learning stable deep dynamics models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Koopman, B. O. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

Li, Q., Dietrich, F., Bollt, E. M., and Kevrekidis, I. G. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. *Chaos*, 27(10):103111, 2017.

Lusch, B., Kutz, J. N., and Brunton, S. L. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, 2018.

Mamakoukas, G., Castano, M. L., Tan, X., and Murphey, T. D. Local Koopman operators for data-driven control of robotic systems. In *Robotics: Science and Systems*, 2023.

Manek, G. and Kolter, J. Z. Learning stable deep dynamics models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Mezić, I. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41 (1-3):309–325, 2005.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018.

Mohr, R. and Mezić, I. Construction of eigenfunctions for scalar-type operators via laplace averages with connections to the koopman operator. *arXiv preprint arXiv:1403.6559*, 2014.

Morton, J., Jameson, A., Kochenderfer, M. J., and Witherden, F. Deep dynamical modeling and control of unsteady fluid flows. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Nagabandi, A., Kahn, G., Fearing, R. S., and Levine, S. Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. In *IEEE International Conference on Robotics and Automation*, 2018.

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E. Model-free prediction of large spatiotemporally chaotic systems from data. *Physical Review Letters*, 120(2):024102, 2018.

Schmid, P. J. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656: 5–28, 2010.

Takeishi, N., Kawahara, Y., and Yairi, T. Learning Koopman invariant subspaces for dynamic mode decomposition. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Williams, M. O., Kevrekidis, I. G., and Rowley, C. W. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

A. Full Proofs

A.1. Proof of Theorem 4.3

Consider the continuous-time system $\dot{z} = Gz = (-S + A)z$. We compute the time derivative of the Lyapunov candidate $V(z) = \|z\|^2 = z^\top z$:

$$\frac{d}{dt}V(z) = z^\top(G + G^\top)z = z^\top(-S + A - S + A^\top)z = z^\top(-2S)z = -2z^\top Sz. \tag{6}$$

Since A is skew-symmetric ($A + A^\top = 0$) and $S \succeq \varepsilon I$, we have $z^\top Sz \geq \varepsilon\|z\|^2$, giving $\frac{d}{dt}V(z) \leq -2\varepsilon V(z)$. By Grönwall's inequality, $V(z(t)) \leq e^{-2\varepsilon t}V(z(0))$. Since $K = \exp(\Delta G)$ is the exact flow map at time Δ, we obtain $V(Kz) = V(z(\Delta)) \leq e^{-2\varepsilon\Delta}V(z)$. Setting $\alpha = e^{-2\varepsilon\Delta} < 1$ completes the proof. □

A.2. Proof of Theorem 4.4

From Theorem 4.3, $\|K\|_2 \leq e^{-\varepsilon\Delta} = \lambda < 1$. For $z_{t+1} = Kz_t + Bu_t$, by the triangle inequality: $\|z_{t+1}\| \leq \lambda\|z_t\| + \|B\|_2\|u_t\|$. Unrolling from $\tau = 0$ to t:

$$\|z_t\| \leq \lambda^t\|z_0\| + \|B\|_2 \sum_{\tau=0}^{t-1} \lambda^{t-1-\tau}\|u_\tau\| \leq \lambda^t\|z_0\| + \frac{\|B\|_2}{1-\lambda}\sup_\tau \|u_\tau\|. \tag{7}$$

This satisfies ISS with $\beta(s,t) = \lambda^t s$ ($\mathcal{KL}$) and $\gamma(s) = \frac{\|B\|_2}{1-\lambda}s$ ($\mathcal{K}$). □

A.3. Proof of Theorem 4.5

Define the latent error $e_t = \phi_\theta(x_t) - \hat{z}_t$ where $\hat{z}_t$ is the predicted latent state. Both satisfy $e_{t+1} = Ke_t + r_t$ with $e_0 = 0$. Unrolling gives $\|e_t\| \leq \bar{r}\sum_{\tau=0}^{t-1}\lambda^\tau \leq \frac{\bar{r}}{1-\lambda}$. The state-space error is:

$$\|x_t - \hat{x}_t\| \leq \|\psi_\theta(\phi_\theta(x_t)) - \psi_\theta(\hat{z}_t)\| + \|x_t - \psi_\theta(\phi_\theta(x_t))\| \leq L_\psi\|e_t\| + \bar{e} \leq \frac{L_\psi\bar{r}}{1-\lambda} + \bar{e}.$$ □

A.4. Proof of Theorem 4.7

Suppose for contradiction that an exact embedding exists with $\bar{r} = 0$, continuous ψ, and $\rho(K) < 1$. Since $\rho(K) < 1$, every trajectory satisfies $z_t = K^t z_0 \to 0$ as $t \to \infty$. By continuity of ψ, every state trajectory satisfies $x_t = \psi(z_t) \to \psi(0)$. Therefore, the ω-limit set of every orbit is $\{\psi(0)\}$, a single point. But the compact invariant set $\mathcal{A}$ (which may be a limit cycle, strange attractor, or collection of multiple equilibria) contains trajectories whose ω-limit sets are not single points. This is a contradiction. □

B. Experimental Details

Architecture. For ODE benchmarks ($d_x \in \{2, 3\}$), the encoder ϕ_θ and decoder ψ_θ are 2-hidden-layer MLPs with hidden dimension 64, ReLU activations, no biases, and spectral normalization on all linear layers. The latent dimension is $d_z = 32$. For the Burgers PDE ($d_x = 64$), we use hidden dimension 128 and $d_z = 64$. All systems use $\varepsilon = 10^{-4}$.

Training. For ODE benchmarks, we generate 128 training trajectories of length 50 and 32 test trajectories of length 500 (200 for ablation) using RK4 integration with $dt = 0.05$. For the Burgers PDE, we use 64 training trajectories of length 30 and 16 test trajectories of length 200 with $dt = 0.001$. Initial conditions are sampled from $\mathcal{N}(0, I)$ (scaled by 0.5 for Burgers). We train for 80–100 epochs using Adam (lr $= 2 \times 10^{-3}$), cosine annealing, and gradient clipping at norm 1.0. Loss weights: $\alpha = \beta = \gamma = 1.0$, $\delta = 0.1$. The Neural ODE baseline uses the same hidden dimensions and is trained with plain MSE loss.

Compute. All experiments were run on a single CPU (Intel, 8 GB RAM). Total wall-clock time for the full suite (4 benchmarks $\times$ 5 models $\times$ 3 seeds + ablation) was approximately 15 minutes.