AAS 26-947

STRUCTURE-PRESERVING DEEP KOOPMAN OPERATOR LEARNING OF PERIODIC ORBITS IN THE EARTH–MOON CR3BP

Simon Bauer* **and John R. Martin†**

Designing missions in the Earth–Moon system relies on repeatedly propagating trajectories in the Circular Restricted Three-Body Problem (CR3BP), a strongly nonlinear system whose motion is costly to integrate far into the future. Machine-learned models promise a fast and cheap alternative, as they reduce each propagation step to a matrix multiplication. One approach toward that goal is Koopman operator theory, trading the nonlinear dynamics for a linear operator that acts on observables of the state. Deep Koopman methods learn this operator from trajectory data. In practice, though, a learned operator that is accurate over a single step tends to drift out of phase or diverge when run forward over a long horizon. We incorporate known dynamics into the learning process, aiming to mitigate this issue. A classical change of coordinates, the Kustaanheimo–Stiefel (KS) transformation, exposes a simple oscillator hidden in the dynamics which we leverage in the learning process. The operator is preconditioned with known dynamical properties governing long-term stability rather than forced to discover them by training. Tested on Lyapunov orbits about the L_1 libration point, the resulting model stays accurate at least sixfold longer than an unconstrained counterpart, bounds the drift of the Jacobi constant, and learns an ordered internal representation. This work provides a principled step toward a future in which nonlinear astrodynamics systems can be modeled and controlled using familiar and time-proven linear methods.

INTRODUCTION

The Circular Restricted Three-Body Problem (CR3BP)[1] is the standard starting point for designing missions in the Earth–Moon system. It gives rise to libration points, periodic orbit families, and invariant manifolds that form the geometric framework of cislunar trajectory design.[2] It is also a strongly nonlinear system with a spectrally rich flow, combining the oscillation of periodic orbits, exponential stretching through stable and unstable manifolds, and mathematical singularities. This inherent complexity introduces challenges for purely numerical integration, such as the requirement of extremely small step sizes to accurately compute trajectories close to the primary bodies. However, mission design demands fast and repeatable propagation of these dynamics for onboard guidance, rapid trajectory screening, and uncertainty propagation, where thousands of trajectories must be integrated under tight computing budgets. Dealing with these challenges motivates the use of surrogate models that replace the nonlinear flow with a linear one. Such a linear model would be cheap to advance and, more valuably, would bring the dynamics within reach of the mature machinery of linear control, estimation, and stability analysis.

*Graduate Research Assistant, Department of Aerospace Engineering, University of Maryland, College Park, Martin Hall Bldg #088 College Park, MD 20742.

†Assistant Professor, Department of Aerospace Engineering, University of Maryland, College Park, Martin Hall Bldg #088 College Park, MD 20742.

Koopman operator theory offers a principled route to such a model, and its deep-learning variants make it possible to learn one from trajectory data alone. When a learned operator is applied recursively over hundreds of steps, small modeling errors compound. For a linear latent model this compounding has a unique spectral signature, and it produces two qualitatively distinct failure modes. The operator drives the latent state by K^k, so its eigenvalues govern the prediction. If any eigenvalue modulus differs from one, the latent amplitude, and through the decoder the predicted trajectory, grows or decays exponentially, spiraling off the orbit or collapsing toward a fixed point. If the eigenvalue arguments differ from the true orbital frequencies, the prediction remains on the orbit's geometric track but advances along it at the wrong rate, accumulating a phase error. Both failure modes are less visible in short-horizon training metrics and prohibitive for the downstream applications above. The CR3BP carries a great deal of known dynamical structure, so rather than hope a generic operator rediscovers it from data, we can supply it. The premise of this paper is that both failure modes can be addressed by construction, through the spectral structure of the operator, rather than by penalty terms or longer training horizons.

A CR3BP-specific challenge for data-driven methods is the $1/r^3$ gravity singularity near each primary. As a spacecraft approaches the Moon, numerical integrators require extremely small step sizes to maintain accuracy, producing unevenly spaced, stiff trajectory data. Machine learning models trained directly on such data must implicitly handle this singularity, which degrades generalization and forces unnecessarily complex architectures.

We address these challenges by combining Kustaanheimo–Stiefel (KS) regularization with a block-structured Koopman architecture (shown schematically in Figure 1). The regularization removes the Moon's singularity and reveals a harmonic oscillator as the leading-order dynamics of the transformed system. We embed this oscillator directly in the Koopman operator as a rotation block whose angle is set by the analytic two-body frequency and refined by a small learned, energy-dependent correction. We further constrain the learned residual block to be orthogonal, so that every eigenvalue of the full operator lies exactly on the unit circle for any value of the learned parameters. The aim is a reduced learning problem where the known physical structure is enforced by construction, and where the learned components can be evaluated in isolation from the known dynamics.

To quantify the benefit of our contribution, we compare our architecture against a baseline identical in encoder, decoder, loss, and training budget but free to learn a dense operator, both in KS and Cartesian coordinates. The two baselines separate the effect of the spectral constraint from that of the KS transformation. By decomposing the rollout error into its amplitude and phase channels, we show that the unit-circle constraint eliminates divergence and suppresses dephasing over multiple orbital periods while keeping the Jacobi constant's drift bounded. We then map how these gains scale with the training budget, sweeping the number of orbits, the samples per orbit, the latent dimension, and the training-rollout horizon. Latent-space and spectral analysis reveal how each mechanism influences performance, with the latent space arranging itself into an interpretable geometry under the KS coordinates and the operator's spectrum constrained to the unit circle.

BACKGROUND AND RELATED WORK

The Koopman operator

A dynamical system is normally described through its state, a point x that the dynamics carry forward one step to $F(x)$ under a generally nonlinear map F. Koopman operator theory takes

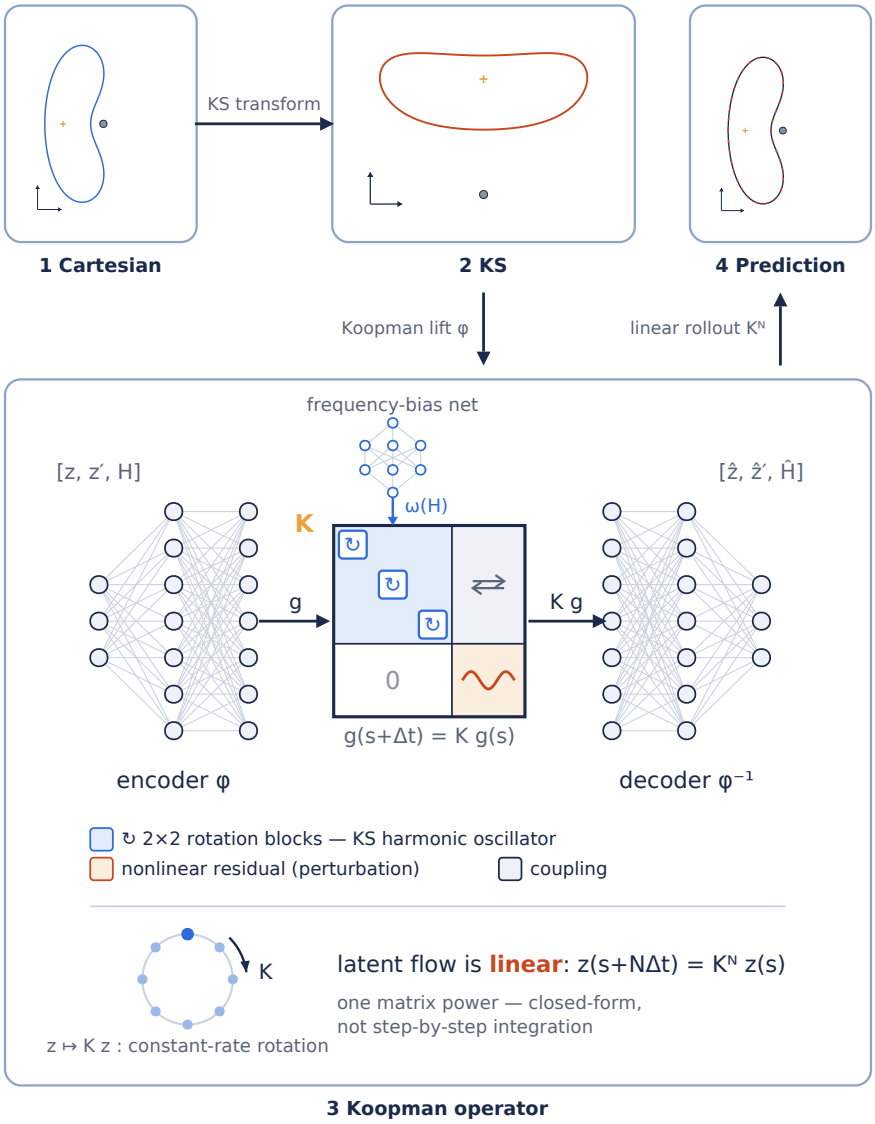

Figure 1. ***Overview of the structure-preserving Koopman pipeline, from Cartesian orbit through KS regularization and the block-structured operator to the final prediction***

another approach of tracking not the state itself but observables, scalar functions $g(x)$ that stand for measurements of the state, and asking how those measurements evolve.[3] The Koopman operator $\mathcal{K}$ is defined as the map that advances any observable by one step of the dynamics through composition with the flow, $(\mathcal{K}g)(x) = g(F(x))$, so the value a measurement takes one step ahead is $\mathcal{K}$ applied to the measurement now.

The property that makes this worthwhile is that $\mathcal{K}$ is a linear operator no matter how nonlinear F is, because composition acts linearly on functions. The cost is dimensionality, since $\mathcal{K}$ acts on an infinite-dimensional space of functions rather than on the finite-dimensional state. Koopman theory thus trades nonlinearity in a small space for linearity in a large one, and this linearity gives the dynamics a spectral decomposition. A Koopman eigenfunction φ obeys $\mathcal{K}\varphi = \lambda\varphi$ and therefore evolves by the constant complex factor λ at every step, its modulus setting growth or decay and

its argument setting an oscillation frequency.[4] Expanding an observable over such eigenfunctions rewrites the full nonlinear evolution as a superposition of these simple linear factors, and the resulting eigenvalues, eigenfunctions, and Koopman modes, the coefficients of that expansion, together characterize the geometry of the underlying flow, its equilibria, oscillations, and invariant sets.[5] No finite computation can carry the whole infinite-dimensional operator, so a practical model approximates it on a chosen dictionary of observables and identifies the resulting finite matrix K from trajectory data. The deep Koopman models discussed next learn that dictionary from data rather than fixing it in advance.

Deep Koopman models and data-driven CR3BP dynamics

Deep Koopman methods extend the classical theory by learning the dictionary of observables with a neural network rather than fixing it by hand.[6] Lusch, Kutz, and Brunton[7] showed that a deep autoencoder can discover a linearizing embedding directly from trajectory data, and that a structured operator handles systems whose spectrum mixes several kinds of behavior, the architecture that most directly motivates the present work. Otto and Rowley[8] jointly learn a Koopman-invariant subspace and a linear recurrent operator by minimizing multi-step prediction error. Later work makes long-horizon stability explicit, arguing that it is decided by the operator's spectrum and is better enforced through the operator's structure than left for the data to discover. Pan and Duraisamy[9] parameterize the operator so its eigenvalues are provably stable, and Azencot et al.[10] pair forward and backward operators under a consistency constraint to keep long-range predictions stable. Another related line of work constrains not the operator's spectrum but the predictor that drives the latent state. Ruiz-Morales et al.[11] prove that an idealized joint-embedding objective is minimized by encoders that span the Koopman-invariant subspace, and show empirically that constraining the linear predictor to a near-identity operator selects that solution, so the learned representation clusters by dynamical regime where a comparable reconstruction autoencoder entangles it.

Koopman methods have a growing presence in astrodynamics, along two broad lines. The first uses analytically constructed operators for dynamical analysis. Servadio, Arnas, and Linares[12] build a Koopman representation of the motion near the three-body libration points and recover Lyapunov and halo periodic orbits within the resulting linear framework, and a companion line applies the operator to uncertainty propagation and filtering.[13] The second learns the operator from data. Closest to our setting, Nehma, Tiwari, and Lingam[14] train a deep extended dynamic mode decomposition (EDMD) autoencoder that lifts the raw Cartesian rotating-frame state into a learned observable space and fits a single global, unconstrained Koopman matrix, while Hofmann et al.[15] compare analytical and learned Koopman operators for the perturbed Kepler and circular restricted three-body problems. Both learned approaches fit the operator without constraining its spectrum. A parallel effort preserves a different kind of structure. Izzo et al.[16] embed a learned perturbation inside a Hamiltonian neural ODE of three-body motion, keeping the symplectic structure of the vector field rather than the spectrum of a linear operator.

In Koopman learning broadly, long-horizon behavior is decided by the operator's spectrum, which is left free, pushed toward stability by a penalty, or shaped indirectly through the training objective. Within astrodynamics, Koopman and neural surrogates reproduce CR3BP motion but either construct the operator analytically, learn it without any spectral constraint, or preserve structure in the nonlinear vector field rather than in the linear operator that governs the rollout. Neither line of work sets the operator's spectrum to the values the physics prescribes, which is the gap this work fills. For a conservative system evolving on a periodic orbit, the Koopman spectrum restricted to the orbit is

purely unimodular,[4,5] so rather than encourage stability through a loss term we place every eigenvalue exactly on the unit circle by construction and leverage the dynamical representation uncovered by the KS transform.

Regularization and the KS transformation

The Kustaanheimo–Stiefel transformation is a classical tool for regularizing the equations of motion of a body moving under gravity. Introduced for the perturbed Kepler problem[17] and developed into a standard reference formulation,[18,19] it has long been applied to the restricted three-body problem to remove the singularity of a close approach to one of the primary bodies.[20,21] We follow the KS-regularized CR3BP formulation of Oguri,[22] developed for cislunar trajectory optimization, whose regularization about the Moon yields the singularity-free equations of motion we use below. Our contribution is to carry this structure into the deep Koopman setting, preconditioning the training data with the transform and building its leading-order dynamics directly into the learned operator.

We work in the standard rotating frame with the Moon at $[1-\mu, 0, 0]^T$ and mass ratio $\mu = 0.01215$. The Moon-centered position is $r_2 = [\xi,\ \eta,\ \zeta,\ 0]^T = [x{-}1{+}\mu,\ y,\ z,\ 0]^T$, with $r_{23} = \|r_2\| = \sqrt{\xi^2+\eta^2+\zeta^2}$ and $r_{13} = \sqrt{(\xi+1)^2+\eta^2+\zeta^2}$ the distances to the Moon and Earth. The untransformed equations of motion are

$$\begin{aligned}
\ddot{x} &= 2\dot{y} + x - \frac{1-\mu}{r_{13}^3}(x+\mu) - \frac{\mu}{r_{23}^3}(x-1+\mu) \\
\ddot{y} &= -2\dot{x} + y - \frac{1-\mu}{r_{13}^3}y - \frac{\mu}{r_{23}^3}y \\
\ddot{z} &= -\frac{1-\mu}{r_{13}^3}z - \frac{\mu}{r_{23}^3}z
\end{aligned} \tag{1}$$

whose μ/r_{23}^3 term diverges at close lunar approach.

The KS transformation writes the Moon-centered position as $r_2 = L(z)\, z = [\xi, \eta, \zeta, 0]^T$, $z \in \mathbb{R}^4$, with the KS matrix[17,18]

$$L(z) = \begin{bmatrix} z_1 & -z_2 & -z_3 & z_4 \\ z_2 & z_1 & -z_4 & -z_3 \\ z_3 & z_4 & z_1 & z_2 \\ z_4 & -z_3 & z_2 & -z_1 \end{bmatrix}, \qquad L(z)^T L(z) = \|z\|^2 I_4, \tag{2}$$

so $r_{23} = \|z\|^2$ and $L(z)^{-1} = L(z)^T/\|z\|^2$. With the Sundman time change $dt/ds = r_{23}$, which rescales physical time t by the lunar distance and defines the fictitious time s (primes denoting d/ds), the equations of motion become[22]

$$z'' = \tfrac{1}{2}\, h_2\, z + L^T M L\, z' + \tfrac{1}{2}\, \|z\|^2\, L^T f_2, \tag{3}$$

with the rotating-frame Coriolis matrix, the perturbing acceleration, and the two-body energy rela-

tive to the Moon

$$M = \begin{bmatrix} 0 & 2 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad f_2 = \begin{bmatrix} (\xi{+}1{-}\mu) - \frac{1-\mu}{r_{13}^3}(\xi{+}1) \\ \eta\big(1 - \frac{1-\mu}{r_{13}^3}\big) \\ -\frac{1-\mu}{r_{13}^3}\,\zeta \\ 0 \end{bmatrix}, \tag{4}$$

$$h_2 = H + \tfrac{1}{2}\big((\xi{+}1{-}\mu)^2 + \eta^2\big) + \frac{1-\mu}{r_{13}}.$$

Here H is the conserved Jacobi energy integral, related to the Jacobi constant by $C_J = -2H$, and the Moon-relative coordinates $[\xi, \eta, \zeta, 0]^T = L(z)\, z$ appearing in f_2 and h_2 are themselves functions of the KS state z. Only the singular Moon term μ/r_{23}^3 (with $r_{23} = \|z\|^2$) is regularized. The Earth-gravity and centrifugal contributions are smooth near the Moon. Augmenting the state with H gives $x = [z^T, z'^T, H]^T \in \mathbb{R}^9$, and physical time follows from $dt/ds = \|z\|^2$.

Dropping the third-body terms M and f_2 leaves the harmonic oscillator $z'' = \frac{1}{2}h_2\, z$, the classical KS reduction of Keplerian motion to a fixed-frequency oscillation,[18,19] with angular frequency $\sqrt{|h_2|/2}$. Since h_2 varies along an orbit and we want to construct a constant Koopman operator, we anchor it to the conserved H. Over one step Δs of fictitious time this fixes the rotation angle $\omega_0(H) = \sqrt{|H|/2}\,\Delta s$, constant over a rollout, on top of which we learn a small energy-dependent correction.

Cartesian data maps to KS by fixing the coordinate $z_4 = 0$. With the Moon-relative position ξ, η, ζ and $r_{23} = \|r_2\|$,

$$z_1 = \sqrt{\tfrac{r_{23}+\xi}{2}}, \quad z_2 = \frac{\eta}{2z_1}, \quad z_3 = \frac{\zeta}{2z_1}, \quad z_4 = 0, \qquad z' = \tfrac{1}{2}\,L(z)^T v_2, \tag{5}$$

with $v_2 = [\dot{x},\ \dot{y},\ \dot{z},\ 0]^T$. This branch degrades as $\xi \to -r_{23}$, where $z_1 \to 0$ and the divisions by z_1 diverge. For $\xi < 0$ we therefore swap the roles of z_3 and z_4 by taking the square root in $z_2 = \sqrt{(r_{23} - \xi)/2}$, which is large exactly when $\xi < 0$, and set

$$z_2 = \sqrt{\tfrac{r_{23}-\xi}{2}}, \quad z_1 = \frac{\eta}{2z_2}, \quad z_3 = 0, \quad z_4 = \frac{\zeta}{2z_2}, \tag{6}$$

so that z_3 is now the fixed coordinate and $z_4 = \zeta/(2z_2)$ carries the out-of-plane component, keeping the divisor away from zero.[18] The inverse maps $r_2 = L(z)\, z$ and $v_2 = 2\,L(z)\, z'/r_{23}$ return decoded predictions to Cartesian coordinates.

Periodic orbits, monodromy, and stability

A periodic orbit of period T is characterized by its monodromy matrix $\Phi(T)$, the state-transition matrix integrated over one period along the orbit.[2] Its eigenvalues, the Floquet multipliers, come in reciprocal pairs $(\lambda, 1/\lambda)$ owing to the Hamiltonian structure, with one trivial pair at unity along the flow direction. A multiplier pair on the unit circle spans a center (oscillatory) subspace. A real pair $(\mu_{\max}, 1/\mu_{\max})$ with $\mu_{\max} > 1$ spans a saddle with stable and unstable manifold directions along which perturbations contract or grow by a factor $\mu_{\max}$ per period. The associated continuous-time rate $\lambda_1 = \ln|\mu_{\max}|/T$ defines the Lyapunov time $\tau_\lambda = 1/\lambda_1$.[23] It is convenient to measure this growth in e-foldings, the number of times an error multiplies by $e \approx 2.718$. By definition one

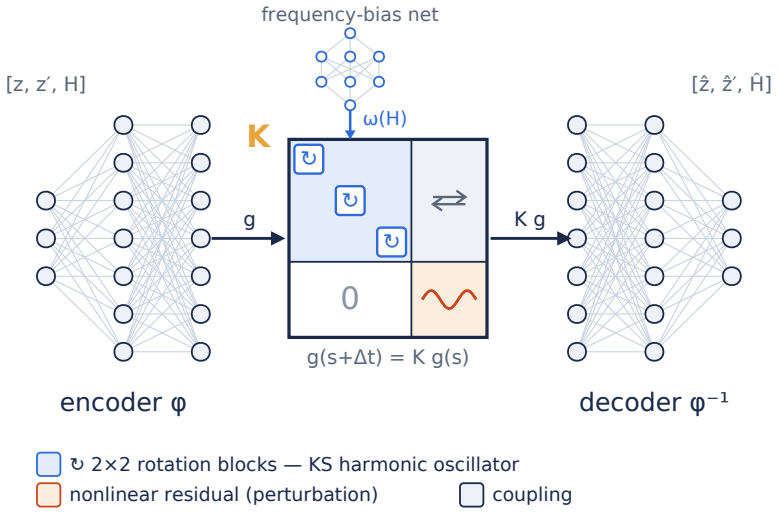

Figure 2. ***Structure-preserving deep Koopman architecture***

e-folding elapses per Lyapunov time τ_λ, so over a single period T a perturbation along the unstable direction grows by $|\mu_{\max}| = e^{\lambda_1 T}$, or equivalently by $\ln|\mu_{\max}|$ e-foldings per period.

The instability matters for two reasons. First, the L_1 Lyapunov family is strongly hyperbolic. Across the energy band investigated in this paper, the dominant multiplier ranges over $|\mu_{\max}| \sim 107$–1821, which is roughly 4.7 to 7.5 e-foldings of error growth per orbital period. Any error a model makes is therefore not merely carried forward but actively stretched by the true flow, so even a small per-step error can grow into a large deviation within a few periods, making multi-period validity a demanding test. Second, λ_1 sets the natural time unit for the validity metric deployed in this work. Normalizing a model's valid-prediction horizon by τ_λ measures it in e-foldings of the true flow rather than in raw time, which makes horizons comparable across orbits of very different instability.

STRUCTURE-PRESERVING DEEP KOOPMAN ARCHITECTURE

Our primary contribution is embedding pre-known dynamical structures into the learning process (shown schematically in Figure 2). We treat the Koopman operator as a two-part system, with a block free of learned parameters that captures the known oscillator dynamics exactly, and a learned block that accounts for the unmodeled perturbations. The autoencoder maps between the physical KS state and this structured latent space.

Encoder and decoder

The encoder $\psi_\theta : \mathbb{R}^9 \to \mathbb{R}^N$ receives the full KS state $x = [z^T,\ z'^T,\ H]^T$ and passes the eight KS coordinates (z, z') through unchanged, appending $N-8$ nonlinear features from a neural network. The decoder reverses this map, passing (z, z') straight back out while H is reconstructed by a separate network. The KS coordinates must remain undistorted because the rotation block acts on them directly, and any learned transformation of these coordinates would corrupt the exact oscillator structure.

Block-structured Koopman operator

At each step, the latent state is advanced by a block-structured Koopman matrix[5,7]

$$\begin{bmatrix} \Psi_{\text{lin}}^{k+1} \\ \Psi_{\text{nl}}^{k+1} \end{bmatrix} = \underbrace{\begin{bmatrix} A_{\text{lin}}(H) & A_{\text{coupling}} \\ 0 & A_{\text{nl}} \end{bmatrix}}_{K(H)} \begin{bmatrix} \Psi_{\text{lin}}^{k} \\ \Psi_{\text{nl}}^{k} \end{bmatrix} \tag{7}$$

The upper-left block $A_{\text{lin}}(H)$ has no trainable matrix entries. It is a pure rotation, the discrete propagator of the harmonic oscillator derived from Eq. (3), and its only learnable degrees of freedom are the few parameters of a small frequency correction network entering through the angle $\omega(H)$,

$$A_{\text{lin}}(H) = \begin{bmatrix} \cos\omega(H)\cdot I_4 & \sin\omega(H)\cdot I_4 \\ -\sin\omega(H)\cdot I_4 & \cos\omega(H)\cdot I_4 \end{bmatrix}, \qquad \omega(H) = \sqrt{\tfrac{1}{2}\big(|H| + b_\phi(H)^2\big)}\;\Delta s. \tag{8}$$

The angle combines the regularized oscillator's analytic leading-order rotation angle, which is $\omega_0(H) = \sqrt{|H|/2}\,\Delta s$ from Sec. , with a learned correction that depends on energy. Here $b_\phi(H)$ is a small multilayer perceptron of the conserved energy H whose output enters under the square root as an energy-like term $b_\phi(H)^2$ alongside $|H|$. Squaring it keeps the radicand non-negative, so the rotation rate is real and strictly positive for any learned parameters and no clamp is needed. The correction is therefore one-sided: the network can raise ω above its two-body value but not lower it, which is the direction the residual perturbation takes on these orbits. At initialization $b_\phi \equiv 0$, so $\omega(H) = \omega_0(H)$ exactly.

Each position–velocity pair (z_i, z_i') rotates by the angle $\omega(H)$, which is constant along a trajectory because H is conserved, and adapts automatically across trajectories at different energy levels. The learned submatrices handle the residual dynamics. The block $A_{\text{coupling}} \in \mathbb{R}^{8\times(N-8)}$ allows the nonlinear features to correct the KS coordinates, while $A_{\text{nl}} \in \mathbb{R}^{(N-8)\times(N-8)}$ governs the nonlinear features among themselves. The zero lower-left block is an architectural constraint that prevents the nonlinear features from corrupting the exact oscillator block. To enforce long-term stability consistent with the conservative nature of the CR3BP, A_{nl} is parameterized as $\exp(W - W^T)$ for a free matrix W. Since $W - W^T$ is skew-symmetric by construction, its matrix exponential is orthogonal, which places all eigenvalues of A_{nl} exactly on the unit circle for any choice of W.[24]

The spectrum of the full operator and the two failure modes

Because $K(H)$ in Eq. (7) is block upper-triangular, its spectrum is the union of the blocks' spectra,[25]

$$\operatorname{spec} K(H) \;=\; \operatorname{spec} A_{\text{lin}}(H) \cup \operatorname{spec} A_{\text{nl}} \;=\; \{e^{\pm i\omega(H)}\} \cup \{e^{i\theta_j}\}, \tag{9}$$

so every eigenvalue is unimodular by construction, for any learned W and A_{coupling}, and for any output of the frequency network b_ϕ, which sets the real angle $\omega(H)$ but cannot change a modulus.

This constraint targets the two ways a linear latent rollout fails. Each learned eigenvalue can be written as $\hat{\lambda}_j = \rho_j e^{i(\omega_j + \Delta\omega_j)}$, where ω_j is the true Koopman frequency of mode j and ρ_j and $\Delta\omega_j$ are its learned modulus and frequency errors. Propagated k steps forward, mode j then departs from the truth by the factor

$$\rho_j^k e^{ik\Delta\omega_j}, \qquad \left|\rho_j^k e^{ik\Delta\omega_j} - 1\right| \;\approx\; \sqrt{(k\ln\rho_j)^2 + (k\,\Delta\omega_j)^2} \quad \text{for small errors.} \tag{10}$$

The first term is the amplitude channel. Any $\rho_j \neq 1$ produces an exponential envelope $e^{k \ln \rho_j}$, diverging for $\rho_j > 1$ or collapsing onto a fixed point or limit set for $\rho_j < 1$. The second is the phase channel. A frequency error $\Delta\omega_j$ produces a phase drift growing linearly in k. The prediction stays on the orbit's track but arrives increasingly early or late. The structured parameterization sets $\rho_j = 1$ exactly, eliminating the amplitude channel for any learned parameters. It also fixes the dominant rotation at the analytic two-body frequency and learns only a low-dimensional, energy-dependent correction on top of it, driving the dominant-mode frequency error $\Delta\omega$ toward zero without ever leaving the unit circle, directly suppressing the phase channel. What remains learnable, the angles θ_j of A_{nl} and the coupling, is the part the data must determine. By contrast, a freely learned dense operator must discover $\rho_j = 1$ to numerical precision from finite, noisy gradients.

On an unstable periodic orbit these latent-error channels compound with the physical instability. Once the prediction leaves the orbit, the true flow stretches the deviation by $|\mu_{\max}|$ per period. Long-horizon validity therefore requires the model to keep its error both small and directed away from the unstable subspace.

Training loss and matched-baseline protocol

The model is trained with four loss terms[7]

$$\begin{aligned}\mathcal{L} = & \underbrace{\lambda_{\mathrm{rec}} \left\| \tilde{x}^k - g_\theta(\psi_\theta(\tilde{x}^k)) \right\|^2}_{\text{reconstruction}} + \underbrace{\lambda_{\mathrm{one}} \left\| \tilde{x}^{k+1} - g_\theta(K(H)\psi_\theta(\tilde{x}^k)) \right\|^2}_{\text{one-step prediction}} \\ & + \underbrace{\frac{\lambda_{\mathrm{multi}}}{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} \left\| \tilde{x}^{k+j} - g_\theta(K(H)^j \psi_\theta(\tilde{x}^k)) \right\|^2}_{\text{multi-step prediction}} + \underbrace{\frac{\lambda_{\mathrm{lin}}}{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} \left\| \psi_\theta(\tilde{x}^{k+j}) - K(H)^j \psi_\theta(\tilde{x}^k) \right\|^2}_{\text{latent linearity}}\end{aligned} \tag{11}$$

The first term enforces reconstruction accuracy, the second penalizes one-step prediction error in physical coordinates, and the third and fourth extend prediction accuracy over a rollout horizon $\mathcal{H}$, in physical space and in the latent space respectively. We use only these four terms, no energy or Jacobi penalty and no spectral regularizer. Whatever conservation behavior the model exhibits must therefore come from the operator's structure rather than the loss, which is the claim under test.

All comparisons in this paper are against free-operator baselines (K_{free}) that share the same encoder and decoder architecture, loss (Eq. (11)), latent dimension, training schedule, and data. The single difference is the operator. Each baseline replaces the structured $K(H)$ of Eq. (7) with a fully unconstrained dense matrix $K_{\mathrm{free}} \in \mathbb{R}^{N \times N}$, learned directly as a free parameter in the style of standard deep Koopman autoencoders.[7,14] No exponential map or other reparameterization is applied to K_{free}. Its spectrum is free to leave the unit circle, which is the freedom whose consequences we measure. Any performance gap between the structured model and the KS baseline is therefore attributable to the operator parameterization alone, not to coordinates, capacity, loss, or training budget.

EXPERIMENTAL SETUP

We evaluate on the planar L_1 Lyapunov family of the Earth–Moon CR3BP (Figure 3), drawn from a catalog of family members spanning $C_J \in [2.90, 3.15]$. The architecture and all claims are

formulated for the full three-dimensional KS state. The planar family is chosen as the controlled testbed because it is simultaneously easy to fit and strongly unstable, so long-horizon validity is still a demanding test even in 2D. Long-rollout experiments use an anchor orbit at $C_J \approx 3.00$ with period $T \approx 4.34$ (nondimensional) and dominant Floquet multiplier $|\mu_{\max}| \approx 287$, corresponding to a Lyapunov time $\tau_\lambda \approx 0.77$ and about 5.7 e-foldings of error growth per period. Figure 3 shows the family in the rotating frame, with the anchor orbit in black and the Moon and L_1 marked, next to the dominant multiplier of every catalog member against C_J, in which the band members are highlighted and the dashed line marks the anchor.

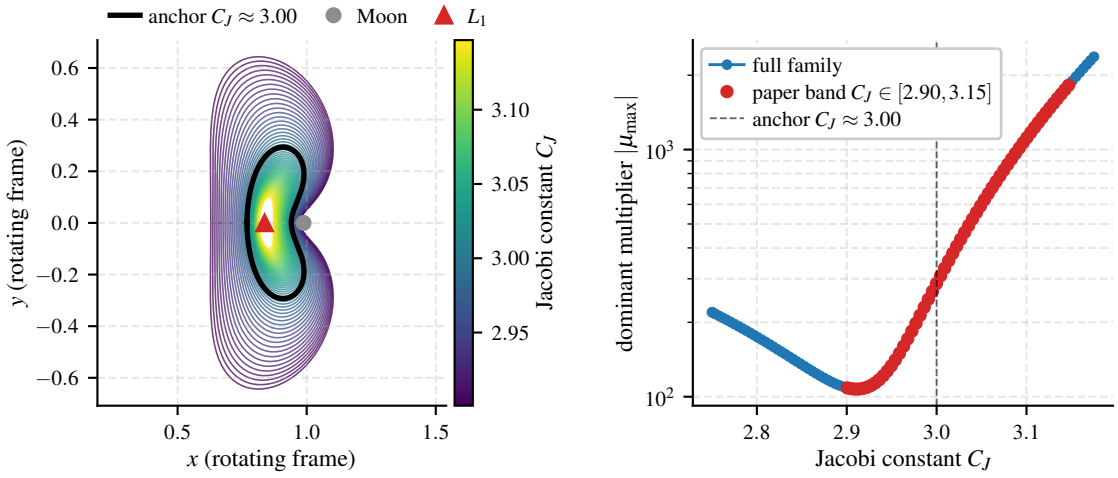

Figure 3. ***The L_1 Lyapunov testbed and its Floquet instability***

Training trajectories are generated by integrating the KS-regularized equations of motion in fictitious time at fixed step $\Delta s = 0.5$, sampling $n_{\text{starts}} = 41$ initial phases per orbit spaced uniformly in physical time, with no off-orbit perturbations, so the models see on-orbit data only. The two KS variants share the encoder/decoder architecture, latent dimension $N = 32$, loss weights, optimizer, schedule, and seeds detailed in the Appendix (Table 1), differing only in the operator. The structured model uses $K(H)$ of Eq. (7), and the KS baseline a freely learned dense K_{free}. A third, Cartesian baseline learns the same free operator in the raw rotating-frame state rather than in KS coordinates, isolating the contribution of the KS representation. In the figures the two baselines are labeled Free (KS) and Free (Cartesian). Each configuration is trained with 5 random seeds. Figures report mean $\pm$ one standard deviation across seeds.

Evaluation metrics

To present long-horizon fidelity clearly, we split it into four complementary metrics, each evaluated on held-out initial conditions and targeting one failure mode. Throughout, predicted trajectories are decoded to Cartesian coordinates and compared against numerically integrated truth extended over many periods. Three of these quantities, the prediction error, the off-orbit distance, and the Jacobi drift, oscillate within each period as the trajectory sweeps toward and away from the Moon. We therefore report them both per step and averaged over each orbital period, so the underlying growth trend reads off directly rather than being buried in the oscillation.

Multi-step prediction error We compute the root-mean-square error (RMSE) between the predicted and true Cartesian states as a function of rollout step, over 20 orbital periods. What matters more is how this error grows over the rollout, not its magnitude. A curve that grows throughout the rollout signals amplitude-channel failure, whereas one that saturates into a flat band signals a bounded error dominated by phase drift.

Phase–amplitude decomposition To separate the two channels we project the prediction onto the reference orbit. Let $\Gamma = \{x_{\mathrm{ref}}(\phi)\}_{\phi\in[0,2\pi)}$ be the reference periodic orbit parameterized by phase. At each rollout step we compute the nearest-point phase $\hat{\phi}(t) = \arg\min_\phi \|\hat{x}(t) - x_{\mathrm{ref}}(\phi)\|$ and the off-orbit distance $d(t) = \min_\phi \|\hat{x}(t) - x_{\mathrm{ref}}(\phi)\|$. The phase error $\Delta\phi(t) = \hat{\phi}(t) - \phi_{\mathrm{true}}(t)$ isolates dephasing. A secular, near-linear $\Delta\phi(t)$ with small $d(t)$ means the model tracks the geometric orbit at the wrong rate. The distance $d(t)$ isolates divergence. Growth in d means the prediction leaves the orbit altogether.

Jacobi constant drift We track the drift $|C_J(\hat{x}(t)) - C_J(x(0))|$ along the predicted trajectory, computed from the decoded state. A unit-circle spectrum keeps the latent dynamics bounded for all time, but the Jacobi constant is a nonlinear function of the decoded physical state. The structured operator therefore does not conserve C_J exactly. What it guarantees is that no latent mode can inject energy without bound, so physical C_J drift stays bounded rather than growing steadily. Since no variant is trained with a conservation penalty, this metric isolates how much of that bounded-drift behavior the operator structure provides on its own.

Validity horizon in Lyapunov units (VPT_λ) The valid prediction time (VPT), the first time the prediction error, normalized by the orbit's root-mean-square amplitude, exceeds a threshold $\varepsilon = 0.4$, is a standard measure of forecast horizon for chaotic systems.[26] On an unstable orbit, raw VPT conflates model quality with orbit instability, so we normalize by the Lyapunov time.

$$\mathrm{VPT}_\lambda = \frac{\mathrm{VPT}}{\tau_\lambda} = \mathrm{VPT} \cdot \frac{\ln|\mu_{\max}|}{T}, \tag{12}$$

i.e. the number of error e-foldings of the true flow the prediction survives. The normalization by τ_λ makes validity comparable across orbits of different instability, which matters for the family-level sweeps. We set $\varepsilon = 0.4$, a mid-range choice in the chaotic-forecasting literature, where thresholds in $[0.3, 0.5]$ are common.[26,27] A looser threshold grants every model more valid time, so the absolute VPT_λ scales with ε, but the comparison we draw does not. In an earlier evaluation, the structured variant's margin over both baselines held at every $\varepsilon \in \{0.2, 0.3, 0.4, 0.5\}$, with the two baselines indistinguishable from each other. Every rollout is scored over a fixed evaluation horizon of 40 orbital periods. A rollout whose error never crosses the threshold within that horizon has no measured VPT. Rather than drop it, which would bias the mean toward the rollouts that fail, we credit it the full horizon. A mean that contains such rollouts is therefore a lower bound, and we mark it with $\gtrsim$. Each reported value is a mean over five training seeds.

RESULTS

Long-rollout fidelity

Figure 4 shows 20-period rollouts from three starting phases, and Figure 5(a,b) plots the Cartesian position RMSE against the true orbit, per step and averaged per period. In the gallery, each rollout is shaded from light at its start to dark after twenty periods, so the passage of time is visible. The structured model stays on the true orbit as one clean loop from every start, while the KS baseline spreads into a widening band around the orbit and the Cartesian baseline spirals outward with growing amplitude. The error curves agree, and the per-period averages show this most clearly. Averaged per period, Figure 5(b) shows the structured model's error climbing for the first few periods and then saturating into a flat band near 10^{-2}, ending at 9.3×10^{-3}, while the KS baseline grows to 9.7×10^{-2} and the Cartesian baseline to 1.4×10^{-1}, both about an order of magnitude higher and still rising at the end of the rollout.

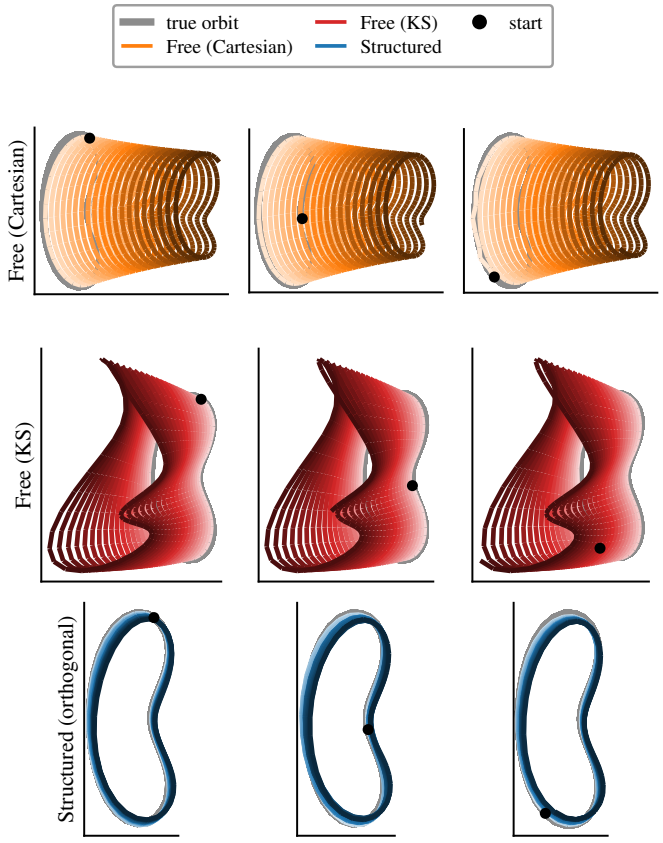

Figure 4. ***Twenty-period predicted rollouts versus the true orbit***

Figure 5(c) reports VPT_λ for the three variants as bars, mean over 5 seeds. The structured model reaches $\mathrm{VPT}_\lambda \gtrsim 223$, the KS baseline ≈ 34, and the Cartesian baseline ≈ 35, so the structured model runs at least sixfold longer than either baseline before its error crosses the threshold. The structured figure is a lower bound: scored over a 40-period horizon, 35 of its 45 rollouts never crossed the threshold at all and are credited the full horizon, whereas every baseline rollout broke within it. The structured model and the KS baseline are identical except for the latent operator and

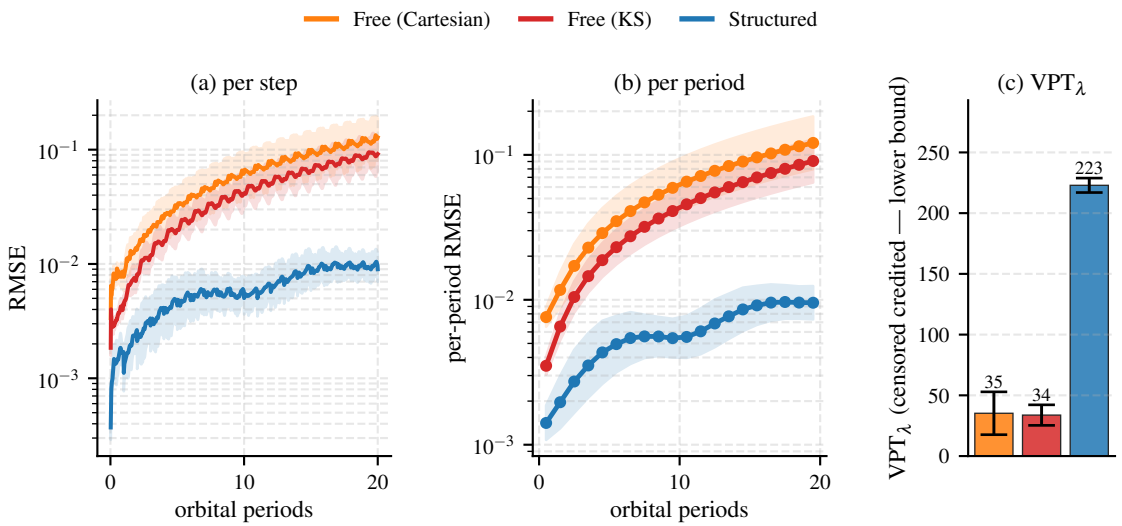

Figure 5. ***Prediction error and validity horizon***

share the same KS coordinates, so the gap between 223 and 34 comes from the operator alone. Learning a free dense operator in raw rotating-frame Cartesian coordinates gives 35, no better than the matched KS baseline.

Dephasing and divergence

Figure 6 plots the two error channels of Eq. (10) over 20 periods, phase error along the orbit and distance off the orbit. The structured model holds the phase error to about 0.06 rad, essentially no drift, while the KS baseline reaches about 39.7 rad and the Cartesian baseline about 68.8 rad. Both baselines drift about linearly once dephasing sets in, after about 5 periods for the Cartesian baseline and about 9 for the KS baseline. Off the orbit, the structured model stays about 10^{-2} away, while both baselines rise to about 0.2, roughly twentyfold larger. These are the two terms in Eq. (10). Phase drift comes from wrong eigenvalue angles, and off-orbit growth comes from eigenvalue moduli leaving one.

Jacobi-constant drift

Figure 7 plots $|C_J(t) - C_J(0)|$ over the rollout for the three variants, none of which uses a conservation penalty in training. The structured model stays in a bounded band with a peak of about 3.5×10^{-2}, while the KS baseline peaks at about 0.76 and the Cartesian baseline at about 0.43. Both baselines trend upward with the horizon and the structured model does not. The difference comes from the operator's spectrum: a unit-modulus operator cannot add energy without bound, so the decoded C_J stays in a band, while an operator with a modulus above one produces a growing envelope.

Data efficiency

Data-driven surrogates are often criticized for their large data requirements, a concern here since every training trajectory must be numerically integrated. Because our operator fixes part of the dynamics by construction, we expect it to need less data than a freely learned one. To test this

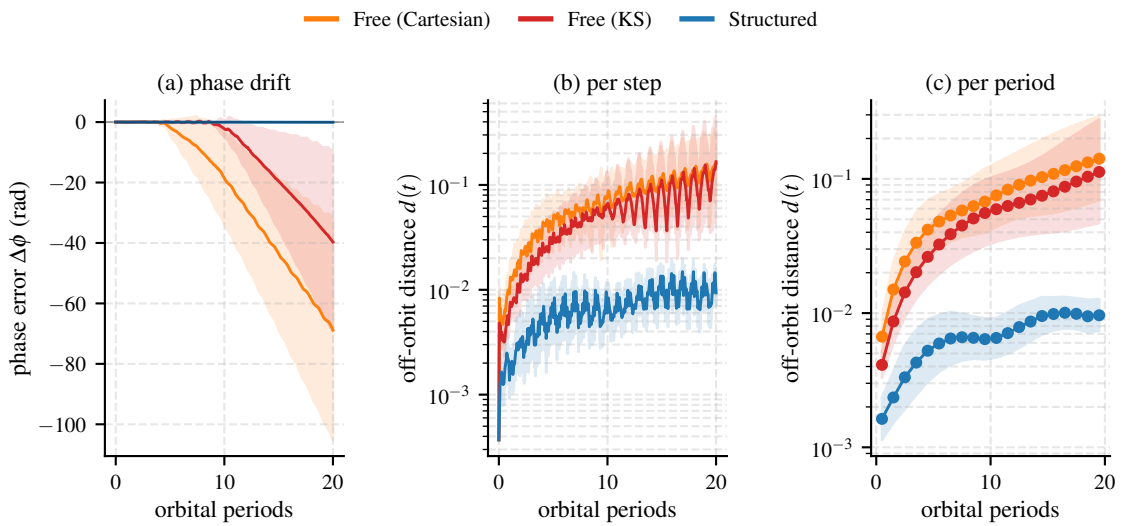

Figure 6. ***Phase drift and off-orbit divergence***

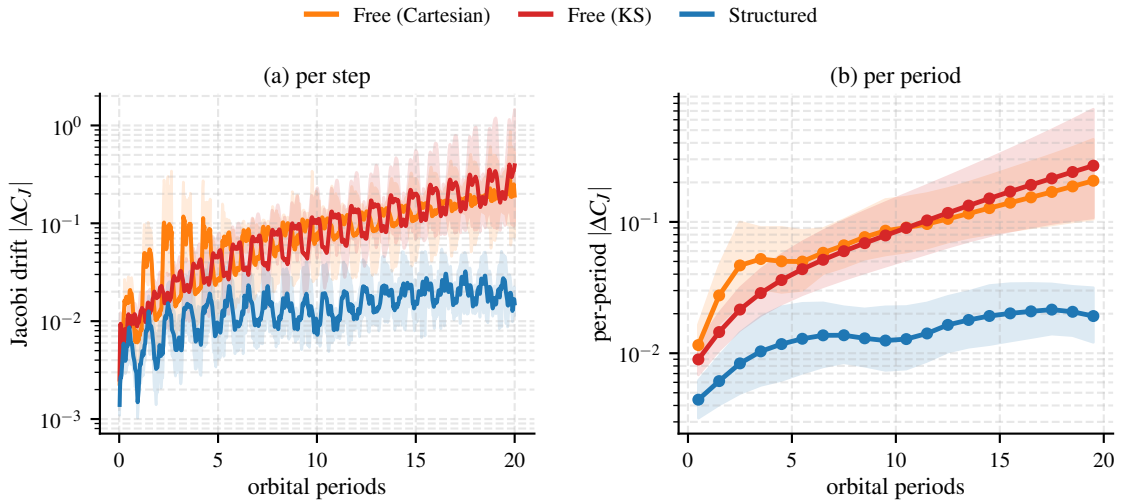

Figure 7. ***Jacobi-constant drift***

we sweep four axes, one at a time, and score every model with VPT_λ on a held-out set kept fixed for that axis, averaged over 5 seeds. The axes are the number of distinct training orbits $n_{\mathrm{orbits}} \in \{1, 2, 4, 8\}$, drawn uniformly from the C_J band, the number of phase starts per orbit $n_{\mathrm{starts}} \in \{3, 6, 8, 10, 18, 25, 41\}$, the latent dimension $N \in \{12, 14, 16, 20, 24, 28, 32, 48\}$, and the training-rollout horizon in orbital periods $\{0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5\}$.

Three of the four sweeps are scored on the anchor orbit and give the same result (Figure 8). Along the phase axis (a) the structured model reaches VPT_λ of a few hundred at every sampling density, even with only three phase-shifted starts per orbit, while both baselines lie near 30 to 75 throughout, so the structured model's VPT_λ is nearly independent of phase sampling density. One caveat applies to this axis, and to the family sweep below. Every point is trained for a fixed number of epochs

rather than a fixed number of optimizer steps, so a model trained on three phase starts passes over each sample far more often than one trained on forty-one. The flatness of the structured curve is therefore consistent both with genuine insensitivity to phase sampling and with the sparse settings compensating through extra passes over less data, and we read it as an upper bound on the structured model's low-data advantage rather than a clean data-efficiency measurement. Equalizing optimizer steps across the axis is left to future work. Along the latent-dimension axis (b) the structured model shows a threshold. At $N = 12$ only four latent dimensions remain for the nonlinear block, too few to realize its structure, and the model falls below the baselines, but from $N = 14$ upward it climbs steeply, reaching $\mathrm{VPT}_\lambda \approx 360$ at $N = 48$, while the baselines rise only gently to about 100. The training-horizon axis (c) repeats the pattern, since rollouts shorter than about half an orbital period leave all three variants near the floor, but once the horizon reaches half a period the structured model rises to the low hundreds and stays there, several times the baselines' plateau near 50.

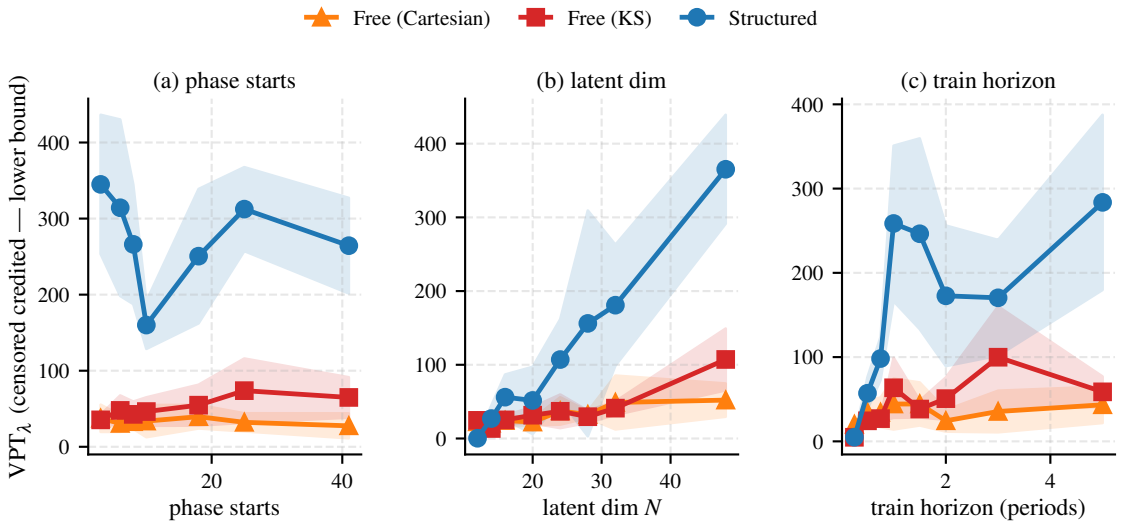

Figure 8. ***Data efficiency on the anchor orbit. (a) phase starts per orbit, (b) latent dimension, (c) training-rollout horizon***

The orbit family sweep (Figure 9) gives a less favorable result, because it is scored across the full C_J band rather than on the single anchor, so every model must generalize over a family rather than reproduce one orbit. Where the anchor sweeps reached VPT_λ of several hundred, all three variants here stay in the single digits, reaching only about 8 to 10 with eight training orbits: generalization across the family is substantially harder than single-orbit fidelity for every operator considered. The structured model retains a small advantage only at the scarcest end, holding $\mathrm{VPT}_\lambda \approx 6.3$ with a single training orbit against 2.2 to 2.8 for the two baselines, but that advantage narrows as orbits are added and the three variants converge by eight orbits. The unit-circle structure that substitutes for phase, latent, and horizon data on a fixed orbit improves VPT_λ only slightly once the model must span a range of energies, and no Koopman variant we tested, structured or free, reaches a useful horizon across the band. The rotation block already carries an energy-dependent frequency-bias network $b_\phi(H)$ meant to adapt the operator across the family. It sharpens the anchor-orbit frequency but does not measurably improve generalization over the band. We do not yet have a full account of why, but hypothesize that conditioning only the dominant rotation is insufficient and that closing the gap requires conditioning the full operator on the Jacobi constant across the family,

which we leave to future work.

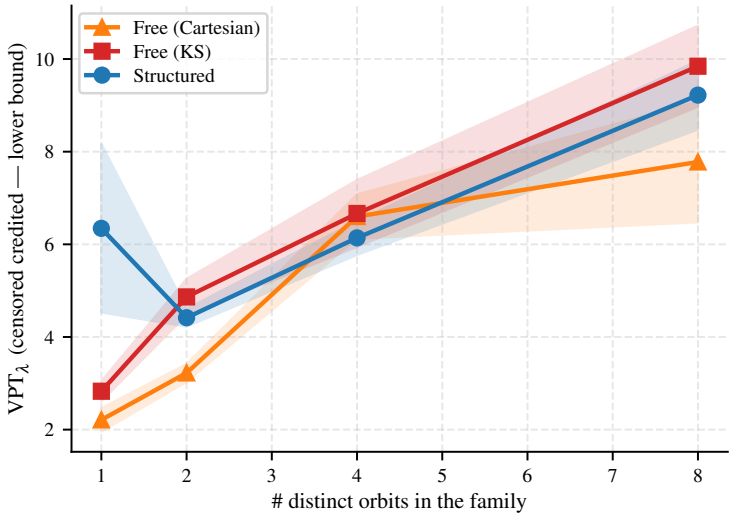

Figure 9. ***Data efficiency across the C_J band, versus the number of training orbits***

Latent geometry

The metrics so far measure how well each model predicts. The latent geometry shows what the encoder learns in order to do so, and whether that internal representation reflects the structure of the orbit. To visualize it we use UMAP, a nonlinear dimensionality-reduction method that projects the high-dimensional latent coordinates onto two dimensions while preserving their local neighborhood structure.[28] Figure 10 shows such a projection of the encoder latent coordinates for the three variants, colored by orbital phase. The encoder is identical across all three variants, so this figure isolates which latent geometry the encoder learns to fit the operator it is paired with. Both KS-coordinate variants, the structured model and the KS baseline, map the orbit onto smooth closed loops around which the phase advances monotonically, while the Cartesian baseline breaks into disconnected arcs with no consistent phase ordering. That ordered geometry comes from the KS coordinates, since both KS variants show it equally, not from the operator. The operator's effect is shown separately by the spectrum in Figure 11.

THE ROLE OF SPECTRAL STRUCTURE

The results separate the contributions of the three structural components. Figure 11 makes the operator's role concrete, plotting the eigenvalues of each trained operator in the complex plane against the unit circle drawn in grey: the structured spectrum lies on the unit circle, while both free operators damp most of their modes inside it and push the dominant one just outside. Both deviations degrade the rollout, since the damped modes shrink the predicted orbit while the expanding one compounds over a long rollout, and together they account for the amplitude error in Figure 5. The unit-circle constraint eliminates the amplitude channel. It is the reason the structured model neither diverges nor collapses over the 40-period evaluation horizon, and the reason its Jacobi-constant drift stays bounded without a penalty term. A norm-preserving latent operator cannot inject energy without bound, so the physical integral drifts within a bounded band, whereas an operator with a modulus

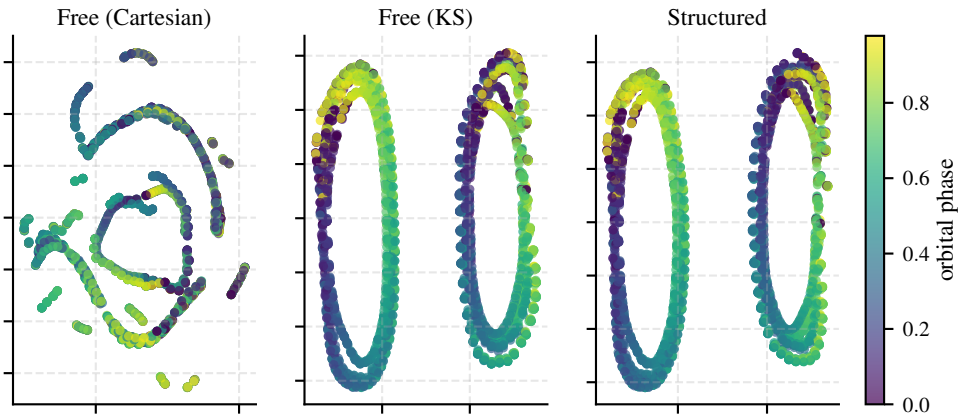

Figure 10. ***Latent geometry of the shared encoder***

above one lets it grow without bound. The physics-anchored frequency addresses the phase channel. Rather than asking gradient descent to discover each mode's frequency, it fixes the dominant rotation at the analytic two-body value and learns only a scalar, energy-dependent correction $b_\phi(H)$ on top, a low-dimensional, well-posed problem whose output, being a function of the conserved H, cannot perturb the operator's moduli. The residual frequency error this achieves is small. On the anchor orbit the structured model accumulates only 0.06 rad of phase error over twenty periods, a dominant-mode error of $\Delta\omega \approx 3 \times 10^{-3}$ rad per period, against several full revolutions of phase error for the two baselines. The KS coordinates make both constraints meaningful. Only in the regularized frame is the leading-order dynamics a constant-frequency oscillator to which a fixed rotation block can be matched.

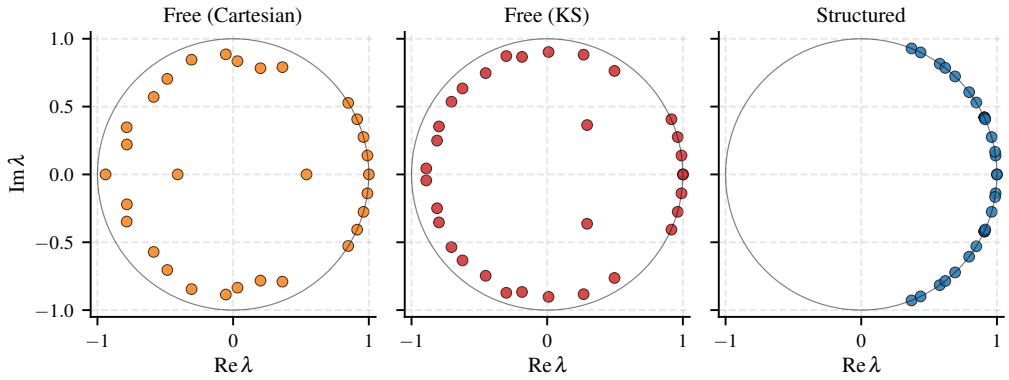

Figure 11. ***Trained-operator spectra***

For dynamics restricted to a periodic orbit, the Koopman spectrum is purely unimodular. The orthogonal constraint embeds that property as an inductive bias, which the results above support. The constraint also limits the model. The monodromy analysis shows the L_1 Lyapunov family carries a strong saddle, Floquet multipliers $(\mu_{\max}, 1/\mu_{\max})$ with $|\mu_{\max}| \sim 107$–1821 across the band. The dynamics near the orbit are governed by exactly the off-circle eigenvalues that the orthogonal parameterization forbids. An operator that must represent contraction and expansion in reciprocal

pairs while preserving the Hamiltonian structure points toward a symplectic parameterization as one candidate, for example $K = \exp(JS)$ with S symmetric, which admits hyperbolic pairs (e^{+s}, e^{-s}) without sacrificing volume preservation.

Strongly perturbed three-dimensional orbits such as near-rectilinear halo orbits pose a different difficulty. Their motion is far from a single harmonic. Close approaches concentrate the spectral content into many interacting harmonics with strong amplitude modulation, demanding more oscillator pairs, and more frequency relationships among them, than a single anchored rotation provides. We expect structure preservation to remain effective there, but the structure to preserve is richer. Harmonically related frequency combs, pre-known symplectic saddle pairs for the instability, and energy-conditioned operators across the family are all plausible additions; which of them closes the remaining gap is the subject of ongoing work.

CONCLUSION

We presented a structure-preserving deep Koopman framework for periodic orbits in the Earth–Moon CR3BP that enforces the physically correct operator spectrum by construction. KS regularization exposes a leading-order harmonic oscillator, embedded as a rotation whose frequency is anchored to the analytic two-body value of the Jacobi constant and refined by a small learned correction, and the learned residual block is constrained to be orthogonal, pinning every eigenvalue of the operator to the unit circle. In a controlled comparison on L_1 Lyapunov orbits against a baseline differing only in the operator parameterization, this spectral structure keeps predictions phase-accurate and bounded over twenty orbital periods, extends the validity horizon at least sixfold in Lyapunov units, and holds the Jacobi constant's drift to an order of magnitude smaller, all without a conservation penalty. Because we separate rollout error into an amplitude and a phase channel, each gain can be traced to a specific spectral property, the unit-circle modulus that prevents divergence and the anchored frequency that suppresses dephasing. Extending validity off the periodic orbit onto invariant manifolds and strongly perturbed orbits requires admitting hyperbolic spectrum in reciprocal pairs, motivating a symplectic generalization of this architecture as one candidate for future work.

ACKNOWLEDGMENT

Anthropic's Claude was used as a coding and writing assistant throughout this work, including the research code, the figures in this paper, and drafts of the text. The authors reviewed everything it produced and are responsible for the content of this paper.

REFERENCES

[1] V. G. Szebehely, *Theory of Orbits: The Restricted Problem of Three Bodies*. New York: Academic Press, 1967.

[2] W. S. Koon, M. W. Lo, J. E. Marsden, and S. D. Ross, *Dynamical Systems, the Three-Body Problem and Space Mission Design*. Marsden Books, 2011.

[3] B. O. Koopman, "Hamiltonian Systems and Transformation in Hilbert Space," *Proceedings of the National Academy of Sciences*, Vol. 17, No. 5, 1931, pp. 315–318, 10.1073/pnas.17.5.315.

[4] I. Mezić, "Spectral Properties of Dynamical Systems, Model Reduction and Decompositions," *Nonlinear Dynamics*, Vol. 41, No. 1–3, 2005, pp. 309–325, 10.1007/s11071-005-2824-x.

[5] M. Budišić, R. Mohr, and I. Mezić, "Applied Koopmanism," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 22, No. 4, 2012, p. 047510, 10.1063/1.4772195.

[6] S. L. Brunton, "Notes on Koopman Operator Theory," tech. rep., University of Washington, 2019. Lecture notes.

[7] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep Learning for Universal Linear Embeddings of Nonlinear Dynamics," *Nature Communications*, Vol. 9, No. 1, 2018, p. 4950, 10.1038/s41467-018-07210-0.

[8] S. E. Otto and C. W. Rowley, "Linearly Recurrent Autoencoder Networks for Learning Dynamics," *SIAM Journal on Applied Dynamical Systems*, Vol. 18, No. 1, 2019, pp. 558–593, 10.1137/18M1177846.

[9] S. Pan and K. Duraisamy, "Physics-Informed Probabilistic Learning of Linear Embeddings of Nonlinear Dynamics with Guaranteed Stability," *SIAM Journal on Applied Dynamical Systems*, Vol. 19, No. 1, 2020, pp. 480–509, 10.1137/19M1267246.

[10] O. Azencot, N. B. Erichson, V. Lin, and M. W. Mahoney, "Forecasting Sequential Data Using Consistent Koopman Autoencoders," *Proceedings of the 37th International Conference on Machine Learning (ICML)*, Vol. 119 of *PMLR*, 2020, pp. 475–485.

[11] P. Ruiz-Morales, D. Vanoost, D. Pissoort, and M. Verbeke, "Koopman Invariants as Drivers of Emergent Time-Series Clustering in Joint-Embedding Predictive Architectures," *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-26)*, Vol. 40, 2026, pp. 25176–25183, 10.1609/aaai.v40i30.39708.

[12] S. Servadio, D. Arnas, and R. Linares, "Dynamics Near the Three-Body Libration Points via Koopman Operator Theory," *Journal of Guidance, Control, and Dynamics*, Vol. 45, No. 10, 2022, pp. 1800–1814, 10.2514/1.G006519.

[13] S. Servadio, W. Parker, and R. Linares, "Uncertainty Propagation and Filtering via the Koopman Operator in Astrodynamics," *Journal of Spacecraft and Rockets*, Vol. 60, No. 5, 2023, pp. 1639–1655, 10.2514/1.A35688.

[14] G. Nehma, M. Tiwari, and M. Lingam, "Deep Learning Based Dynamics Identification and Linearization of Orbital Problems using Koopman Theory," *Aerospace Science and Technology*, Vol. 163, 2025, p. 110295. Preprint: arXiv:2403.08965, 10.1016/j.ast.2025.110295.

[15] C. Hofmann, G. Lavezzi, D. Wu, S. Servadio, and R. Linares, "Analytical and Data-Driven Koopman Operator for the Perturbed Kepler and Circular Restricted Three-Body Problems," *Acta Astronautica*, Vol. 234, 2025, pp. 314–328, 10.1016/j.actaastro.2025.04.034.

[16] D. Izzo, S. Origer, G. Acciarini, and F. Biscani, "High-Order Expansion of Neural Ordinary Differential Equation Flows," *Science Advances*, Vol. 11, No. 51, 2025, p. eady1348, 10.1126/sciadv.ady1348.

[17] P. Kustaanheimo and E. Stiefel, "Perturbation Theory of Kepler Motion Based on Spinor Regularization," *Journal für die reine und angewandte Mathematik*, Vol. 218, 1965, pp. 204–219, 10.1515/crll.1965.218.204.

[18] E. L. Stiefel and G. Scheifele, *Linear and Regular Celestial Mechanics: Perturbed Two-Body Motion, Numerical Methods, Canonical Theory*, Vol. 174 of *Grundlehren der mathematischen Wissenschaften*. Berlin, Heidelberg: Springer-Verlag, 1971.

[19] J. Waldvogel, "Fundamentals of Regularization in Celestial Mechanics and Linear Perturbation Theories," tech. rep., Seminar for Applied Mathematics, ETH Zürich, 2007. Lecture notes, Scottish Universities Summer School in Physics, Skye, Scotland.

[20] Y. N. Chelnokov, "Quaternion Regularization of the Equations of the Perturbed Spatial Restricted Three-Body Problem: I," *Mechanics of Solids*, Vol. 52, No. 6, 2017, pp. 613–639, 10.3103/S0025654417060036.

[21] M. Rossi and M. Guzzo, "A Kustaanheimo–Stiefel Regularization of the Elliptic Restricted Three-Body Problem and the Detection of Close Encounters with Fast Lyapunov Indicators," *Physica D: Nonlinear Phenomena*, Vol. 463, 2024, p. 134173, 10.1016/j.physd.2024.134173.

[22] K. Oguri, "Regularization of the Circular Restricted Three-Body Problem for Trajectory Optimization," *AAS/AIAA Astrodynamics Specialist Conference*, Broomfield, CO, Aug. 2024. Paper AAS 24-298.

[23] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach," *Physical Review Letters*, Vol. 120, No. 2, 2018, p. 024102, 10.1103/PhysRevLett.120.024102.

[24] M. Lezcano-Casado and D. Martínez-Rubio, "Cheap Orthogonal Constraints in Neural Networks: A Simple Parametrization of the Orthogonal and Unitary Group," *Proceedings of the 36th International Conference on Machine Learning (ICML)*, Vol. 97 of *PMLR*, 2019, pp. 3794–3803.

[25] J. W. Demmel, *Applied Numerical Linear Algebra*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1997, 10.1137/1.9781611971446.

[26] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, "Backpropagation Algorithms and Reservoir Computing in Recurrent Neural Networks for the Forecasting of Complex Spatiotemporal Dynamics," *Neural Networks*, Vol. 126, 2020, pp. 191–217, 10.1016/j.neunet.2020.02.016.

[27] J. A. Platt, S. G. Penny, T. A. Smith, T.-C. Chen, and H. D. I. Abarbanel, "A Systematic Exploration of Reservoir Computing for Forecasting Complex Spatiotemporal Dynamics," *Neural Networks*, Vol. 153, 2022, pp. 530–552, 10.1016/j.neunet.2022.06.025.

[28] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *arXiv preprint arXiv:1802.03426*, 2018.

HYPERPARAMETER AND TRAINING DETAILS

All three variants share every setting in Table 1 and differ only in the latent operator and, for the Cartesian baseline, in the input coordinates. The structured model uses $K(H)$ of Eq. (7), and the two baselines replace it with a dense $K_{\mathrm{free}} \in \mathbb{R}^{N\times N}$.

Table 1. Shared hyperparameters for all three variants

Setting	**Value**
Latent dimension N	32
Encoder hidden layers	$[64, 64, 64, 64]$
Decoder hidden layers	$[16, 16]$
Frequency-bias network $b_\phi(H)$	$[16, 16]$
Activation	SELU
Optimizer	AdamW
Learning rate (cosine to 10^{-5})	10^{-3}
Weight decay	2×10^{-4}
Gradient clip	0.5
Batch size	1024
Epochs	$20{,}000$
Loss weights $(\lambda_{\mathrm{rec}}, \lambda_{\mathrm{one}}, \lambda_{\mathrm{multi}}, \lambda_{\mathrm{lin}})$	$1.0,\ 1.0,\ 2.0,\ 1.0$
Rollout horizon $\mathcal{H}$	45
Fictitious-time step Δs	0.5
Phase starts per orbit n_{starts}	41
Train / validation split	$80/20$
Seeds per configuration	5