SPEC: EXACT STRUCTURAL CONSTRAINTS ARE CHEAP, SPECIFYING THEM IS NOT

Anonymous authors
Paper under double-blind review

ABSTRACT

Exactness is the defining property of structure-preserving dynamics models, yet the specification it rests on, which law to impose, under which metric, and in which coordinates, is never itself tested, because a prediction loss cannot tell a correct specification from a wrong one. To address this gap, we present SPEC, a staged framework that makes the specification computable from trajectory data and applies an exact constraint only where the data support one. Its foundation is a per-step identity, $(Ux)^\top M(Ux) - x^\top Mx = w^\top \Sigma_M(A)\, w$, which reduces the structural question to the sign of one symmetric matrix and yields, under the Euclidean block metric, a library of five exact constructions together with the boundaries beyond which none exists. Three conditional stages then decide what the identity leaves open, separating a statistical decision taken on finite trajectories from a guarantee that holds at every parameter value. Experiments on analytically labelled systems isolate what each stage contributes. Discover recovers coordinates without a reconstruction weight, at median invariant R^2 of 0.73 to 0.86 against 0.10 to 0.22 for a comparison objective. Refusing unsupported blocks costs 4% more out-of-family error than an oracle router, and enforcing every block costs 39%. Over the fitting horizon, exact enforcement matches a penalty of equal capacity in modal-energy error while holding the selected latent law to 10^{-15} where the penalty reaches only 10^{-3}. Accuracy is therefore limited by neither stage, but by what sits behind them: the metric the construction is instantiated in, for which we give the exact form in the damped case, and the decision rule, which measures agreement inside the selection window rather than validity beyond it.

1 INTRODUCTION

A learned dynamics model is usually applied autoregressively, so an error in one step becomes the input to the next, and a rollout can diverge within tens of steps even when its single-step accuracy is high (Sanchez-Gonzalez et al., 2020; Brandstetter et al., 2022). Structure-preserving architectures address this by encoding a physical law into the parameterisation, so the guarantee follows from the architecture rather than from the fit and holds at every value of the learned weights. For example, Hamiltonian and symplectic models preserve conservative structure, contractive formulations enforce decay, and equivariant ones transform consistently under a specified group action, an approach with roots in geometric integration (Hairer et al., 2006; Greydanus et al., 2019; Zhong et al., 2020b; Lohmiller & Slotine, 1998; Finzi et al., 2020; Lee et al., 2021). This has improved long-horizon prediction wherever the governing law and its representation are known in advance.

When the governing structure must instead be inferred from trajectories, three decisions precede any construction. One must determine which constraint family the data support, under which metric the generator satisfies the corresponding condition, and in which coordinates the constraint is expressible. We refer to this triple as the *constraint specification*. An error in any one of them leaves the guarantee formally intact while voiding its meaning: a conservative model applied to an irreversible process conserves exactly and predicts badly (Lee et al., 2021; Zhong et al., 2020a), a law invisible in one coordinate system becomes explicit in another (Cranmer et al., 2020), and a mismatched metric associates the generator with a constraint it does not support (Lohmiller & Slotine, 1998; Manchester & Slotine, 2017). Existing work relaxes parts of this problem, removing the requirement for canonical coordinates or softening symmetry assumptions (Chen et al., 2021; Finzi et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

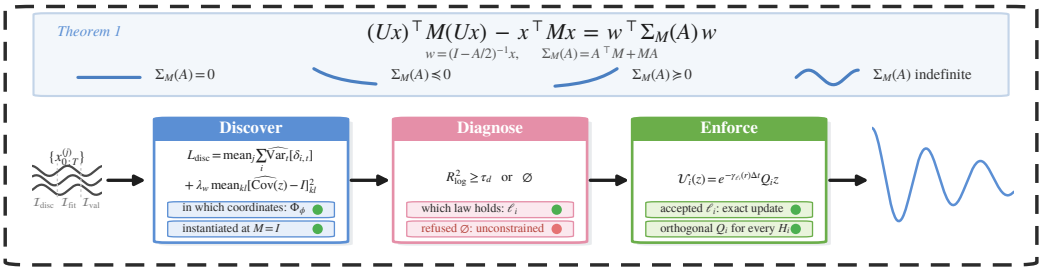

Figure 1: **Overview of SPEC.** The identity (top) fixes what a generator can guarantee, and the three stages (bottom) decide the rest from data.

2021; Wang et al., 2022; Yang et al., 2023), but addresses the decisions one at a time rather than all three from the same data, and none of it tests the specification once made, since a prediction loss is not sensitive to a guarantee that holds whatever the fit.

We first resolve which metric a given generator is aligned to. For a Cayley update $U = \text{Cay}(A)$, an auxiliary vector $w = (I - A/2)^{-1}x$, and $\Sigma_M(A) = A^\top M + MA$, we derive the per-step identity

$$(Ux)^\top M(Ux) - x^\top Mx = w^\top \Sigma_M(A)\, w, \tag{1}$$

which holds for every admissible A, every symmetric M, and every state x. The question of what a generator can guarantee therefore reduces to the sign of one symmetric matrix. Under a Euclidean block metric this yields five exact constructions, obtained by pairing an orthogonal Cayley component that preserves the block norm with a scalar radial update that prescribes its evolution, together with the boundaries beyond which no such construction exists. What the identity cannot supply is which of the five the data support, and in which coordinates.

To settle those two questions we introduce **SPEC** (**S**elective **P**rescription of **E**xact **C**onstraints), which takes them in turn through three conditional stages (Fig. 1). *Discover* searches for coordinates in which candidate radial laws are expressible, using an objective invariant to blockwise rescaling so that no reconstruction weight is needed to fix the scale of the representation. *Diagnose* fits each automatically searched candidate (conservation, exponential and algebraic decay) on one interval, scores it on a disjoint one, and either selects a supported law or refuses the block, while the sphere and region laws are used only when their radii are supplied. *Enforce* applies the matched construction to what survives, so that an accepted guarantee holds at every value of the trainable parameters while a refused block is left free.

We evaluate each stage against what it claims on systems with analytically known structure. Without any reconstruction term, Discover recovers coordinates in which the invariant is expressible at median R^2 of 0.73 to 0.86 across three analytic systems, against 0.10 to 0.22 for a comparison objective trained under the same budget. On mixed-regime systems, refusing unsupported blocks costs 4% more out-of-family error than an oracle router while enforcing every block costs 39%. Over the fitting horizon, exact enforcement matches a penalty of equal capacity in prediction error while holding the selected latent law to 10^{-15} where the penalty reaches only 10^{-3}, although the penalty predicts better at longer horizons. Two things bound what any of this can achieve. The metric is one, since the budget operator of a damped mode's shifted generator is indefinite and its Euclidean radius cannot follow a single exponential. The decision rule is the other, since a held-out score measures agreement inside the selection window rather than beyond it.

Contributions. **(i)** An exact per-step identity that reduces the structural question to the sign of one symmetric matrix, yielding five exact constructions under the Euclidean block metric together with the boundaries beyond which none exists. **(ii)** SPEC, which decides the specification from trajectory data or refuses, separating a statistical decision taken on finite trajectories from a construction that holds at every parameter value. **(iii)** A separation of what exactness delivers from what it does not: it holds the selected latent law exactly at no cost in modal-energy error over the fitting horizon, while long-horizon accuracy depends on rate estimation and on the metric, for which we give the exact form in the damped case. **(iv)** An information limit on the decision itself. Any score computed inside the selection window measures agreement there rather than validity beyond it, which makes the two failure modes move oppositely under the threshold, and we show that no threshold and no scale-free structure test removes the resulting trade-off.

2 Related work

Structure-preserving learned dynamics. Geometric integration established conditions under which numerical updates preserve invariants and symplectic structure (Cooper, 1987; Hairer et al., 2006). These principles enter learned dynamics through Hamiltonian and Lagrangian models (Greydanus et al., 2019; Cranmer et al., 2020), symplectic networks (Jin et al., 2020), and GENERIC architectures for reversible and irreversible dynamics (Grmela & Öttinger, 1997; Zhang et al., 2022). Related approaches use contraction, Lyapunov, or equivariance conditions for stability or symmetry guarantees (Lohmiller & Slotine, 1998; Manchester & Slotine, 2017; Revay et al., 2024; Finzi et al., 2020; Cheng et al., 2024). In each, the property is selected in advance rather than chosen by comparing candidates on trajectory data.

Data-driven structure discovery. A complementary line infers dynamical structure from observations. Sparse identification and symbolic regression recover governing equations or interpretable relations (Brunton et al., 2016; Udrescu & Tegmark, 2020). Koopman and autoencoder methods learn coordinates in which nonlinear dynamics admit simpler representations (Champion et al., 2019; Azencot et al., 2020), and others infer conserved quantities, symmetries, or approximately equivariant structure (van der Ouderaa et al., 2024; Finzi et al., 2021; Wang et al., 2022; Yang et al., 2023). These identify relations consistent with the data, but a discovered relation does not by itself yield a model in which it holds exactly for every parameter value.

Exact constrained parameterisations. Cayley transforms give differentiable parameterisations of orthogonal matrices (Helfrich et al., 2018), and exponential maps support optimisation over orthogonal, unitary, and more general Lie groups (Lezcano-Casado & Martínez-Rubio, 2019). Contraction-based models and stabilised neural differential equations construct dynamics satisfying a specified stability or manifold condition (Revay et al., 2024; Cheng et al., 2024; White et al., 2023). These maintain a selected constraint through construction, but not which constraint the trajectories support, or in which coordinates. Extended related work is in Appendix B.

3 Method

3.1 Problem formulation

Let $\mathcal{D} = \{x^{(j)}_{0:T_j}\}_{j=1}^{N_{\mathrm{traj}}}$ denote trajectories sampled at interval Δt. An encoder Φ_ϕ maps each state to B latent blocks,

$$z_t = \Phi_\phi(x_t) = (z_{1,t}, \ldots, z_{B,t}), \qquad z_{i,t} \in \mathbb{R}^{d_b}, \tag{2}$$

and a decoder D_ψ maps the updated latent state to observation space. Each block i is associated with the radial observable $r_{i,t} := \|z_{i,t}\|_2$, so the coordinate map determines the representation in which the Euclidean block metric is evaluated. Full notation is given in Table 9 of Appendix A. Let $\mathcal{K}_{\mathrm{con}}$ denote the library of exactly constructible radial laws and let $\mathcal{K}_i \subseteq \mathcal{K}_{\mathrm{con}}$ be the candidate set evaluated for block i. From $\mathcal{D}$, the specification procedure returns either a fitted law label $\ell_i \in \mathcal{K}_i$ with parameters $\widehat{\eta}_{i,\ell_i}$, or the refusal label $\varnothing$. For every accepted or externally specified decision, the goal is to construct a latent update $\mathcal{U}_i$ satisfying

$$\|\mathcal{U}_i(z_i)\|_2 = f_{\ell_i}\left(\|z_i\|_2; \widehat{\eta}_{i,\ell_i}\right) \quad \text{for every admissible value of its free parameters.} \tag{3}$$

No structural statement is imposed when $\ell_i = \varnothing$. Constraint specification therefore comprises coordinate discovery, law selection or refusal, and exact conditional construction.

3.2 Exact characterisation and constructions

The relation between a generator and the quantity its Cayley update changes is characterised below.

Theorem 1 (Exact budget congruence)**.** *Let $M = M^\top$, let $S_A := I - A/2$ be invertible, and define*

$$U := \mathrm{Cay}(A) = S_A^{-1}(I + A/2), \qquad \Sigma_M(A) := A^\top M + MA.$$

Then

$$U^\top M U - M = S_A^{-\top} \Sigma_M(A) S_A^{-1}. \tag{4}$$

Consequently $U^\top M U - M$ and $\Sigma_M(A)$ have the same inertia: one vanishes, is negative semidefinite, or is positive semidefinite exactly when the other is.

Equation 4 is a congruence by an invertible matrix, so the inertia claim follows from Sylvester's law. Substituting $w = S_A^{-1}x$ yields the per-step form

$$(Ux)^\top M(Ux) - x^\top Mx = w^\top \Sigma_M(A)w, \tag{5}$$

so zero, negative-semidefinite, positive-semidefinite, and indefinite $\Sigma_M(A)$ correspond to preservation, nonincrease, nondecrease, and state-dependent change of $x^\top Mx$. The complete argument and the well-posedness conditions are in Appendices C.1–C.3.

The learned coordinates are equipped with the Euclidean block metric $M = I$. For each block, an unconstrained matrix H_i determines the skew-symmetric generator $K_i := H_i - H_i^\top$, whose Cayley transform $Q_i := \mathrm{Cay}(K_i)$ satisfies $Q_i^\top Q_i = I$ for every H_i. The following proposition separates the radial and orthogonal components of the block update.

Proposition 1 (Exact radial closure)**.** *Let $Q^\top Q = I$ and let $\alpha : [0,\infty) \to [0,\infty)$. The update*

$$\mathcal{U}_\alpha(z) = \alpha(r)Qz, \qquad r = \|z\|_2, \tag{6}$$

satisfies $r^+ = \alpha(r)r$ exactly. Thus, any radial map f satisfying $f(0) = 0$ is realised by $\alpha(r) = f(r)/r$ for $r > 0$, with the origin defined separately.

For $\alpha_i(r) = e^{-\gamma_i(r)\Delta t}$, the block update and its induced radial recurrence are

$$\mathcal{U}_{i,\gamma}(z) = e^{-\gamma_i(r)\Delta t}\, Q_i z, \qquad r^+ = e^{-\gamma_i(r)\Delta t} r. \tag{7}$$

The five constructions instantiate Eq. 7 with different rate functions, each integrating in closed form:

$$\begin{array}{ll}
\mathrm{con} : \gamma = 0 & r_k = r_0 \\
\mathrm{exp} : \gamma = \bar{\gamma} & r_k = e^{-\bar{\gamma}k\Delta t} r_0 \\
\mathrm{alg} : \gamma(r) = \frac{1}{2p\Delta t}\log(1 + 2pc\Delta t\, r^{2p}) & r_k = (r_0^{-2p} + 2pck\Delta t)^{-1/(2p)} \\
\mathrm{sph} : \gamma(r) = \gamma_0(r^2 - r_*^2) & r_* \text{ invariant, attracting for } 0 < \mu < 1 \\
\mathrm{reg} : \gamma(r) = \gamma_0[r^2 - R^2]_+ & r \le R \;\Rightarrow\; r^+ \le R
\end{array} \tag{8}$$

Every line is algebraic, holding for each H_i because Eq. 6 makes the radius independent of the skew parameters. What the trajectories supply is which line, and with which η. Proofs, and the boundaries of the family, are in Appendices C.6–C.10.

3.3 The SPEC framework

As illustrated in Figure 1, the three stages operate on disjoint intervals $\mathcal{I}_{\mathrm{disc}}$, $\mathcal{I}_{\mathrm{fit}}$, and $\mathcal{I}_{\mathrm{val}}$ for representation learning, parameter fitting, and held-out law selection.

Discover. For adjacent observations in $\mathcal{I}_{\mathrm{disc}}$, define the block log rate

$$\delta_{i,t}^{(j)} := \frac{\log r_{i,t+1}^{(j)} - \log r_{i,t}^{(j)}}{\Delta t}. \tag{9}$$

The encoder and decoder are estimated by minimising

$$L_{\mathrm{discover}} = \underbrace{\operatorname*{mean}_j \sum_{i=1}^{B} \widehat{\mathrm{Var}}_t\big[\delta_{i,t}^{(j)}\big]}_{L_{\mathrm{rate}}} + \lambda_{\mathrm{rec}} \operatorname*{mean}_{(j,t),\,c}\big[D_\psi(\Phi_\phi(x_t^{(j)})) - x_t^{(j)}\big]_c^2 + \lambda_w \operatorname*{mean}_{k,l}\big[\widehat{\mathrm{Cov}}(z) - I\big]_{kl}^2, \tag{10}$$

where $\widehat{\mathrm{Var}}_t$ is the temporal variance within one trajectory and $\widehat{\mathrm{Cov}}$ is the sample covariance of z pooled over $\mathcal{I}_{\mathrm{disc}}$. Each term is a mean over the indices it runs over, so λ_w weights the mean squared entry of $\widehat{\mathrm{Cov}}(z) - I$ rather than its Frobenius norm. The rate term L_{rate} is invariant under positive blockwise rescaling, which is why it needs no reconstruction term to fix the scale of z. It is, however, minimised by any constant encoder, for which $\delta \equiv 0$, and driven towards zero by a large constant offset in z. The whitening term excludes the first and is therefore applied at every setting, including the $\lambda_{\mathrm{rec}} = 0$ used for the headline results, with $\lambda_w = 0.1$ throughout, but it constrains only the covariance of z and does not exclude the second.

Algorithm 1 SPEC. Blue holds at every parameter value, green is estimated from finite trajectories.

Input: Trajectories $\mathcal{D}$, block count B, threshold τ_d, optional targets $\{r_{*,i}, R_i\}$

1: Split each trajectory into disjoint $\mathcal{I}_{\text{disc}}, \mathcal{I}_{\text{fit}}, \mathcal{I}_{\text{val}}$
2: *Discover*: estimate Φ_ϕ on $\mathcal{I}_{\text{disc}}$, then freeze it and set $r_{i,t} \leftarrow \|[\Phi_\phi(x_t)]_i\|_2$ ▷ Eq. 10
3: **for** $i = 1, \dots, B$ **do** ▷ *Diagnose*
4: Fit $\widehat{\eta}_{i,k}$ on $\mathcal{I}_{\text{fit}}$ and score $R^2_{\log}(i,k)$ on $\mathcal{I}_{\text{val}}$, for each $k \in \mathcal{K}_i$
5: $\ell_i \leftarrow k_i^*$ if $R^2_{\log}(i,k_i^*) \geq \tau_d$ and $\widehat{\eta}_{i,k_i^*} \in \mathcal{E}_{k_i^*}$, else $\varnothing$ ▷ Eq. 11
6: **end for**
7: **for** $i = 1, \dots, B$ **do** ▷ *Enforce*
8: $\mathcal{U}_i(z) \leftarrow e^{-\gamma_{\ell_i}(\|z\|_2)\Delta t}\,\mathrm{Cay}(H_i - H_i^\top)\,z$ if $\ell_i \neq \varnothing$, else $\mathcal{U}_i \leftarrow F_{\xi_i}$ ▷ exact for every H_i
9: **end for**
10: Train D_ψ, $\{H_i\}$ and $\{F_{\xi_i}\}$ on the task loss, with Φ_ϕ frozen
11: **return** Φ_ϕ, D_ψ, $\{\mathcal{U}_i\}_{i=1}^B$

Diagnose. Candidate parameters $\eta_{i,k}$ are fitted on $\mathcal{I}_{\text{fit}}$, propagated recursively on $\mathcal{I}_{\text{val}}$, and scored by $R^2_{\log}(i,k)$. The resulting label is

$$k_i^* = \arg\max_{k \in \mathcal{K}_i} R^2_{\log}(i,k), \qquad \ell_i = \begin{cases} k_i^*, & R^2_{\log}(i,k_i^*) \geq \tau_d \text{ and } \widehat{\eta}_{i,k_i^*} \in \mathcal{E}_{k_i^*}, \\ \varnothing, & \text{otherwise}, \end{cases} \tag{11}$$

where $\mathcal{E}_k$ is the parameter range the construction of law k admits, so a candidate that scores above the threshold but cannot be realised exactly is refused rather than clipped. Conservation, exponential decay, and algebraic decay are searched automatically. The sphere candidate is evaluated only when r_* is supplied, and a forward-invariant region requires an externally specified boundary R.

Enforce. An accepted or externally specified block is assigned the matched construction from Eq. 7, with the parameters $\widehat{\eta}_{i,\ell_i}$ that were scored. A refused block remains unconstrained under F_{ξ_i}.

Algorithm 1 gives the procedure. Line 2 is the only step that fits the coordinate map, and it freezes Φ_ϕ before anything is scored, so no later estimate can change the coordinates in which a law was selected. The estimated steps can be wrong and the construction cannot: line 6 holds for every H_i whatever line 5 decided, and line 7 optimises $\{H_i\}$ without touching an accepted guarantee. Sphere and region candidates, the scoring definition, and the batched form are in Appendix D.

4 Experiments

Our experiments answer four questions. **Q1:** Does the per-step budget identity hold to floating-point precision across dimensions and sign regimes? **Q2:** Does the Discover objective, whose rate term is invariant under blockwise rescaling, recover coordinates in which candidate laws are expressible without a reconstruction weight? **Q3:** What does selective refusal cost against enforcing everything or nothing, and where does it fail? **Q4:** What does exact enforcement buy once capacity is held fixed, and what limits it?

Setup. Q1 checks Eq. 5 under $M = I$ across dimensions, sign regimes and floating-point formats. Q2 compares L_{rate} with L_{div}, which scores a log-linear fit within a block partition against random re-partitions, on three analytic systems and on a damped grid with a linear encoder, modal reference coordinates and a phase-shuffled control. Q3 and Q4 run on analytical modal coordinates: Q3 routes mixed-regime systems at $\tau_d = 0.30$ and probes the decision boundary with out-of-family blocks, and Q4 imposes the same fitted rates through one aggregate law, a per-block penalty and the exact construction. Full configurations are given in Appendix F.1.

4.1 Q1: Numerical precision of the exact construction

The identity holds for every admissible A and every M by direct expansion. With the Cayley update $U = \mathrm{Cay}(A) = (I - A/2)^{-1}(I + A/2)$ and the auxiliary vector $w = (I - A/2)^{-1}x$, substituting

$Ux = (I + A/2)w$ and $x = (I - A/2)w$ gives

$$(Ux)^\top M(Ux) - x^\top Mx = \big[(I + A/2)w\big]^\top M\big[(I + A/2)w\big] - \big[(I - A/2)w\big]^\top M\big[(I - A/2)w\big]$$
$$= 2w^\top\Big(\tfrac{A^\top}{2}M + M\tfrac{A}{2}\Big)w = w^\top \Sigma_M(A)\, w. \tag{12}$$

Results. Residuals remain at floating-point precision across all sign regimes and dimensions. In float64 the maximum across all 280 cases is 1.07×10^{-13}, and the indefinite case matches the sign-definite regimes in absolute residual. Its larger relative residual is a property of the denominator rather than of the computation: $w^\top \Sigma_M(A) w$ passes through zero when $\Sigma_M(A)$ is indefinite, so the worst ratio over 70 draws is taken near a cancellation. Float64 medians fall near 2×10^{-15} and float32 medians near 5×10^{-7} outside the zero regime, within one order of the respective machine epsilon, confirming that residuals arise entirely from floating-point rounding. At $d_b = 2$ the deployed Enforce step is an analytic blockwise rotation costing $\mathcal{O}(Bd_b^2)$ operations per step, with operator storage of Bd_b^2 elements. Table 1 and Figure 2 report the full residual breakdown.

Table 1: **Budget-identity residuals in finite precision.** Each cell aggregates 70 draws (7 dimensions $\times$ 10 seeds) under $M = I$. Relative residual is the maximum over the 70 draws of $|\text{res.}|/|(Ux)^\top M(Ux) - x^\top Mx|$. The zero regime has $\Sigma_M(A) = 0$ identically, so the relative column is undefined and shown as a dash.

	float64			float32		
Sign regime	Median	Max	Relative	Median	Max	Relative
Zero	1.78×10^{-15}	8.53×10^{-14}	—	0	7.63×10^{-6}	—
Nonpositive	1.83×10^{-15}	1.07×10^{-13}	2.5×10^{-14}	4.92×10^{-7}	8.11×10^{-6}	2.1×10^{-6}
Nonnegative	2.33×10^{-15}	7.46×10^{-14}	1.3×10^{-14}	5.07×10^{-7}	1.62×10^{-5}	2.9×10^{-6}
Indefinite	2.16×10^{-15}	7.13×10^{-14}	1.3×10^{-12}	5.81×10^{-7}	1.19×10^{-5}	2.5×10^{-4}

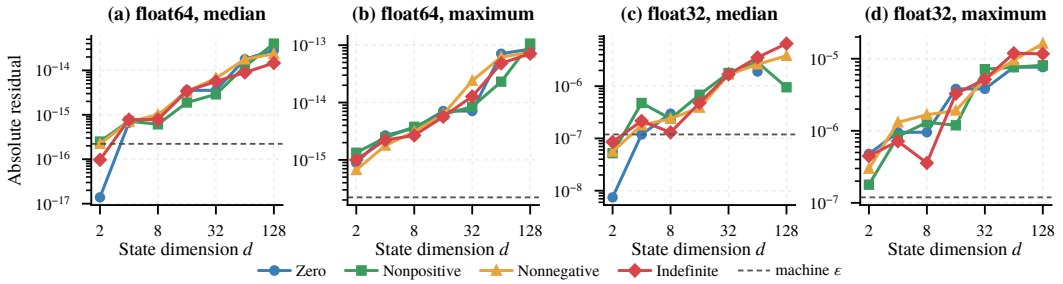

Figure 2: **Budget-identity residuals by dimension, sign regime, and floating-point format.** Panels **(a)–(d)** show float64 median, float64 maximum, float32 median, and float32 maximum across seven dimensions for four sign regimes. Dashed lines mark the machine epsilon of each format. In panel (c), 41 of the 70 float32 draws in the zero regime have residual exactly zero, so the medians at $d = 16$, 32 and 128 are zero and do not appear on the logarithmic axis.

4.2 Q2: Does coordinate discovery make candidate laws expressible?

Results. At $\lambda_{\text{rec}} = 0$ the log-rate objective reaches median invariant R^2 of .728, .767, and .858 on the three systems, against .095, .216, and .219 for the comparison objective, with disjoint seed distributions in every case (Fig. 3d, Table 2). Across the full sweep the two differ in kind rather than degree: the log-rate median varies by at most .16 over three orders of magnitude in λ_{rec}, while the comparison median moves by .81 on the pendulum alone (Fig. 3a–c). A tuned weight lets L_{div} reach .902 there, above the best log-rate median of .833, so what L_{rate} provides is robustness to a hyperparameter rather than a uniformly higher score. The comparison objective is sensitive to the numerical environment as well, since its 1200-step optimisation amplifies the BLAS reduction order, so the reported values reproduce only at the recorded thread count and R^2 moves by up to .18 otherwise, whereas the log-rate objective is bitwise identical across thread counts. The rate term is invariant under positive blockwise rescaling, which is why it needs no reconstruction weight to fix the scale of z, and it is also why any constant encoder minimises it. The learned encoders place a median of 91% of each block radius into a constant offset, which the whitening term of Eq. 10

does not constrain. The invariant is carried by a per-trajectory radial level on top of that offset, which predicts it on held-out trajectories at R^2 of 0.88 to 0.99, so the reported R^2 is a linear readout of a nearly constant radius whose small variation across trajectories tracks the conserved quantity (Appendix F.3). Centring the latent removes that level and lowers R^2 on two of the three systems.

Table 2: **Coordinate discovery on three analytic systems.** Median invariant R^2 over ten encoder seeds at $\lambda_{\text{rec}} = 0$ and at the best sweep weight, with $\lambda^\star$ in parentheses.

	$\lambda_{\text{rec}} = 0$		Best (at $\lambda^\star$)	
System	L_{rate}	L_{div}	L_{rate}	L_{div}
Pendulum	**.728**	.095	.833 (50)	**.902** (50)
Lotka-Volterra	**.767**	.216	**.767** (0)	.386 (10)
Kepler	**.858**	.219	**.884** (10)	.237 (1)

The full sweep appears in Appendix F.2.

Table 3: **Damped-system grid: evaluations reaching $R^2_{\log} > 0.8$ (%).** Reported by system dimension N and pooled over damping scales, horizons, and learning rates.

	System dimension N				
Coordinates	5	10	20	40	Pooled
Modal (reference)	99.0	91.2	90.2	64.5	86.2
L_{rate} (learned)	6.5	2.9	0.2	0.1	2.4
L_{div} (learned)	16.0	11.3	2.0	0.6	7.5
Phase-shuffled	0.0	0.0	0.0	0.0	0.0

A breakdown by damping scale appears in Appendix F.3.

The damped-system grid reverses the ordering. Modal coordinates make candidate laws expressible in 86.2% of evaluations at $R^2_{\log} > 0.8$, falling to 64.5% at $N = 40$ and 66.5% at the strongest damping (Table 3). Against that reference the comparison objective reaches 7.5% and the log-rate objective 2.4%, and their median $R^2_{\log}$ peaks at $2T_{\text{char}}$ and falls below that of the phase-shuffled control from $5T_{\text{char}}$ onward (Appendix F.3). Since neither objective uses reconstruction, the reversal does not come from weight pressure, and it is consistent with the weaker encoder class and shorter budget used on this grid. The learned coordinates fall more than an order of magnitude short of the reference, so at this scale coordinate discovery rather than enforcement is what limits the pipeline.

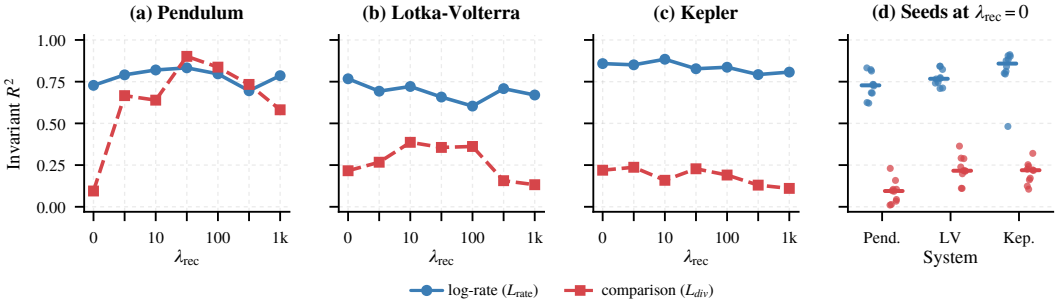

Figure 3: **Coordinate discovery across systems and seeds. (a)–(c)** Invariant R^2 against the reconstruction weight, medians over ten encoder seeds. The log-rate curve is nearly flat, while the comparison objective is strongly weight dependent and overtakes it only on the pendulum. **(d)** Seed-level R^2 at $\lambda_{\text{rec}} = 0$, one marker per seed with lines at group medians. The distributions are disjoint in all three systems. The damped-system grid is in Appendix F.3.

4.3 Q3: What refusal costs, and where the rule fails

Results. As shown by the routing errors in Figure 4a and Table 4, selective refusal recovers most of the oracle's advantage at a small cost. Relative to an oracle with access to the true labels, SPEC incurs 4% more out-of-family error, whereas enforcing every block incurs 39%. Here a refused block holds its last observed radius, and all errors are in log radius. Holding every block therefore matches the oracle out-of-family error, since no out-of-family block is altered, but leaves each in-family block at a constant radius while its true radius decays. Every accepted block is enforced with the law that was scored, and blocks whose fitted parameters fall outside the admissible range of their construction are refused rather than clipped, yielding an in-family error of .0125. The residual 4% is attributable to blocks accepted in error, which raises the question

	Hold	All	SPEC	Oracle
In-family ↓	5.07	.0125	.0125	.0125
Out-of-family ↓	.585	.814	.610	.585
Blocks enforced	none	all	accepted	in-family
True labels	no	no	no	yes

Table 4: **Log-radius rollout error by routing rule at $\tau_d = 0.30$.** Medians over 60 cells. A refused block holds its last observed radius.

of when such errors occur. On blocks constructed as two exponentials with rate ratio κ, enforcement becomes harmful at $\kappa = 5$, where 89% of them are still accepted, and refusal begins only at $\kappa = 8$ (Fig. 4b). Beyond that point the score correlates negatively with benefit, with a Spearman coefficient between -0.50 and -0.56, so the accepted blocks are those that enforcement damages most. Blocks whose law is correct but whose evidence is noisy exhibit the opposite failure, being refused where enforcement would have been beneficial. The two respond to the threshold in opposite directions: regret is minimised at $\tau = 0.8$ for the first and at $\tau = -0.5$ for the second (Fig. 4c). Figure 5 shows why. On wrong law forms the regret divides between blocks accepted in error and blocks refused in error, so raising the threshold trades one against the other, whereas under observation noise the score never accepts in error and every unit of regret comes from refusal, which at $\sigma = 0.3$ leaves 0.9% of correct blocks accepted while enforcement would still have delivered a benefit above 5. Substituting a residual-structure test separates the two mechanisms, whose lag-one autocorrelations are 0.999 and 0, without separating the decisions, since the decision depends on an amplitude that a scale-free statistic discards (Appendix F.5). The autocorrelation is 0.9992 to four digits at every rate ratio from $\kappa = 1.1$ to $\kappa = 20$, while the residual amplitude that tracks the cost of enforcing grows from 2×10^{-5} to 6×10^{-3} over the same range. The score therefore measures agreement inside the selection window rather than validity beyond it.

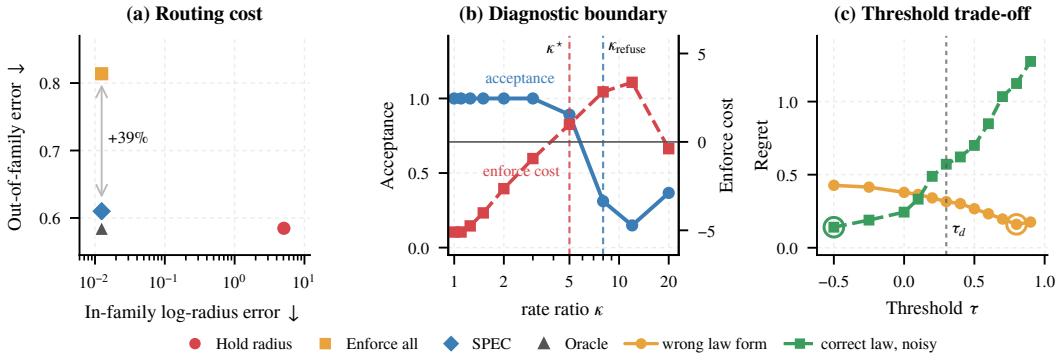

Figure 4: **What refusal costs, and where the rule fails. (a)** Log-radius rollout error of each routing rule, where a refused block holds its last observed radius. **(b)** On two-rate blocks, enforcement becomes harmful at $\kappa^\star = 5$ while refusal begins only at $\kappa_{\text{refuse}} = 8$. **(c)** Regret against threshold for the two failure modes, with the optima circled at opposite ends.

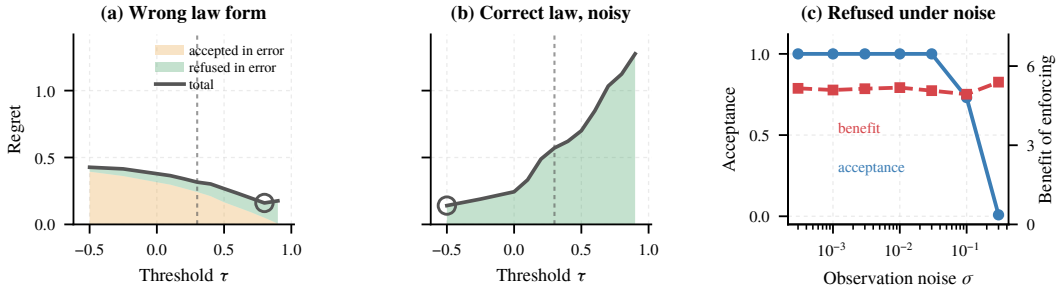

Figure 5: **How the threshold trades the two failures. (a)–(b)** Regret against threshold, split into blocks accepted in error and blocks refused in error, with the optimum circled and the deployed $\tau_d = 0.30$ dashed. On wrong law forms both contribute, whereas under noise the score never accepts in error, so all regret comes from refusal. **(c)** Acceptance of correct exponential blocks under observation noise, against the benefit enforcement would have delivered.

4.4 Q4: What exact enforcement buys, and what limits it

Results. In a representative cell (Fig. 6) the ten modal decay rates span a factor 1.48, a difference that is negligible within the fitting window but compounds beyond it. One aggregate law therefore overdamps the slow modes and underdamps the fast ones, and by $100\,T_{\text{fit}}$ its ratio of predicted to true energy departs from one by more than twelve orders of magnitude. Over the grid, one law per mode reduces truth error at $100\,T_{\text{fit}}$ by a factor of $29.4\times$, the median of the per-cell ratios (Table 5). This gain comes from having one law per block rather than from exactness, since the penalty, which also

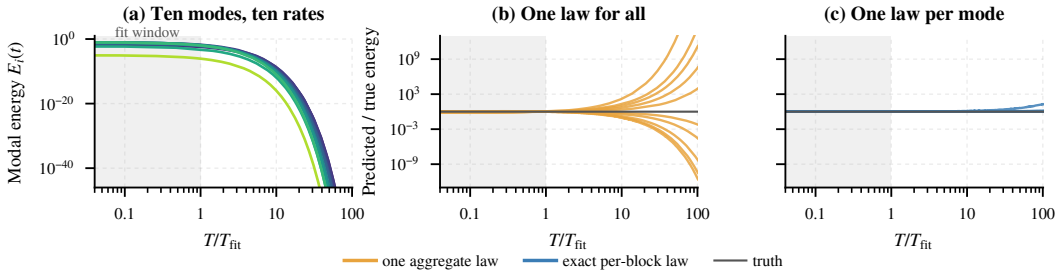

Figure 6: **Why one law cannot cover ten modes.** A representative cell with $N = 10$ and $\rho = 1$, shaded over the fitting window. **(a)** The ten modal energies, coloured by decay rate. **(b)** Ratio of predicted to true energy under a single aggregate law, which overdamps the slow modes and underdamps the fast ones. **(c)** The same ratio under one exact law per mode, on the same axis.

Table 5: **Truth error and deviation from the imposed law across horizons.** Medians over 80 configuration-seed cells, with the penalty shown at $\beta = 1$. Truth error is the log modal-energy error of the extrapolated block radius against the analytical modal solution, averaged over modes, trajectories and the window of length T after T_{fit}.

	Truth error ↓			Deviation from imposed law ↓		
T/T_{fit}	Aggregate	Penalty	Exact	Aggregate	Penalty	Exact
1	.123	.0207	.0225	9.4e−15	1.5e−3	9.4e−15
2	.241	.0214	.0259	1.9e−14	2.9e−3	1.8e−14
5	.584	.0250	.0411	4.6e−14	7.3e−3	4.5e−14
10	1.16	.0364	.0585	9.2e−14	1.5e−2	9.1e−14
20	2.33	.0558	.0900	1.9e−13	2.9e−2	1.8e−13
50	5.82	.105	.198	4.6e−13	7.3e−2	4.5e−13
100	11.6	.206	.377	9.2e−13	1.5e−1	9.1e−13

The aggregate law is also an exact construction, which is why its deviation matches the exact per-block column. The gain by damping scale is in Appendix F.6.

fits one rate per block, comes within 8% of the exact truth error at the fitting horizon. What exactness adds is the guarantee: at the fitting horizon the rollout radius deviates from the law imposed on it by 10^{-15} under the exact construction against 10^{-3} under the penalty, and the two stay eleven orders of magnitude apart at every horizon. The penalty's truth error is nonetheless lower at every horizon, with the gap widening beyond T_{fit}. Its radius also follows a single exponential, but at a rate that differs from the fitted one by 1.4×10^{-4} in relative terms, so the gap reflects rate estimation. All three errors stay above 1.5×10^{-2} because of the metric. The budget operator of the shifted generator $A_\omega + \gamma I$ is indefinite, so the Euclidean radius of a damped mode oscillates around an exponential with log half-amplitude $\frac{1}{2}\operatorname{artanh}(\gamma/\omega)$, which no single-exponential law can follow, and Proposition 7 gives the metric under which the radius decays exactly. Pooled by damping, the gain over the aggregate law at $100\,T_{\text{fit}}$ falls from $263.9\times$ at $\rho = 0.5$ to $4.5\times$ at $\rho = 5.0$ (Fig. 9c). The limit lies in the construction rather than the data: on analytical modal coordinates, an unconstrained linear block map, not restricted to rotations, reaches a log-radius error of 10^{-12} against 10^{-2} for the enforced law (Appendix F.5). Exact enforcement therefore matches the penalty in modal-energy error over the fitting horizon while holding the selected latent law exactly, and long-horizon accuracy is set by rate estimation and by the metric.

5 Conclusion

We address a question that structure-preserving dynamics models take as settled: which law should be imposed, under which metric, and in which coordinates? We introduced SPEC, which makes this specification computable from trajectory data, combining a per-step identity that reduces what a generator can guarantee to the sign of one symmetric matrix with three conditional stages that choose the coordinates, select a law or refuse, and apply the matched construction. Evaluating each stage against what it claims shows that holding a selected law exactly is cheap, and that what limits the result is the specification behind it, namely the metric in which the construction is instantiated and a decision rule that measures agreement inside the selection window rather than validity beyond it. As learned models are rolled out further beyond the data they were fitted on, specifying structure correctly will matter more than enforcing it.

THE USE OF LARGE LANGUAGE MODELS (LLMS)

In this work, we used generative AI tools to edit the manuscript for readability. None of the tasks requiring disclosure are applicable to this work. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Omri Azencot, N. Benjamin Erichson, Vanessa Lin, and Michael Mahoney. Forecasting sequential data using consistent koopman autoencoders. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 475–485. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/azencot20a.html.

A. Bhatt, D. Floyd, and B. E. Moore. Second order conformal symplectic schemes for damped Hamiltonian systems. *Journal of Scientific Computing*, 66(3):1234–1259, 2016. doi: 10.1007/s10915-015-0062-z.

Johannes Brandstetter, Daniel Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations*, 2022.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016. doi: 10.1073/pnas.1517384113.

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

B. Chang, M. Chen, E. Haber, and E. H. Chi. AntisymmetricRNN: a dynamical system view on recurrent neural networks. In *International Conference on Learning Representations*, 2019.

Yuhan Chen, Takashi Matsubara, and Takaharu Yaguchi. Neural symplectic form: Learning hamiltonian equations on general coordinate systems. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 16659–16670. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/8b519f198dd26772e3e82874826b04aa-Paper.pdf.

Jing Cheng, Ruigang Wang, and Ian R. Manchester. Learning stable and passive neural differential equations. *arXiv preprint arXiv:2404.12554*, 2024. doi: 10.48550/arXiv.2404.12554.

G. J. Cooper. Stability of Runge–Kutta methods for trajectory problems. *IMA Journal of Numerical Analysis*, 7(1):1–13, 1987. doi: 10.1093/imanum/7.1.1.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. *arXiv preprint arXiv:2003.04630*, 2020. URL https://arxiv.org/abs/2003.04630.

N. B. Erichson, O. Azencot, A. Queiruga, L. Hodgkinson, and M. W. Mahoney. Lipschitz recurrent neural networks. In *International Conference on Learning Representations*, 2021.

Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3165–3176. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/finzi20a.html.

Marc Finzi, Gregory Benton, and Andrew G Wilson. Residual pathway priors for soft equivariance constraints. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 30037–30049. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/fc394e9935fbd62c8aedc372464e1965-Paper.pdf.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/26cd8ecadce0d4efd6cc8a8725cbd1f8-Paper.pdf.

M. Grmela and H. C. Öttinger. Dynamics and thermodynamics of complex fluids. I. Development of a general formalism. *Physical Review E*, 56(6):6620–6632, 1997.

A. Gu, K. Goel, and C. Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Albert Gu, Tri Dao, Stefano Ermon, Atri Rudra, and Christopher Ré. HiPPO: Recurrent memory with optimal polynomial projections. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 1474–1487. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/102f0bb6efb3a6128a3c750dd16729be-Paper.pdf.

E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Series in Computational Mathematics*. Springer, 2nd edition, 2006. doi: 10.1007/3-540-30666-8.

Kyle Helfrich, Devin Willmott, and Qiang Ye. Orthogonal recurrent neural networks with scaled Cayley transform. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1969–1978. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/helfrich18a.html.

N. J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, 2008.

A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna. Lie-group methods. *Acta Numerica*, 9: 215–365, 2000. doi: 10.1017/S0962492900002154.

Jet Propulsion Laboratory. JPL Horizons On-Line Ephemeris System. NASA/JPL, 2023. URL https://ssd.jpl.nasa.gov/horizons/.

P. Jin, Z. Zhang, A. Zhu, Y. Tang, and G. E. Karniadakis. SympNets: Intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems. *Neural Networks*, 132:166–179, 2020.

F. M. Lasagni. Canonical Runge–Kutta methods. *Zeitschrift für angewandte Mathematik und Physik*, 39(6):952–953, 1988. doi: 10.1007/BF00945133.

Kookjin Lee, Nathaniel Trask, and Panos Stinis. Machine learning structure preserving brackets for forecasting irreversible processes. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 5696–5707. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/2d1bcedd27b586d2a9562a0f8e076b41-Paper.pdf.

Mario Lezcano-Casado and David Martínez-Rubio. Cheap orthogonal constraints in neural networks: A simple parametrization of the orthogonal and unitary group. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3794–3803. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/lezcano-casado19a.html.

W. Lohmiller and J.-J. E. Slotine. Contraction analysis: The discrete-time case. MIT-NSL Report 971101, MIT Nonlinear Systems Laboratory, nov 1997.

Winfried Lohmiller and Jean-Jacques E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. doi: 10.1016/S0005-1098(98)00019-3.

D. S. Mackey, N. Mackey, and F. Tisseur. Structured tools for structured matrices. *Electronic Journal of Linear Algebra*, 10:106–145, 2003. doi: 10.13001/1081-3810.1101.

Ian R. Manchester and Jean-Jacques E. Slotine. Control contraction metrics: Convex and intrinsic criteria for nonlinear feedback design. *IEEE Transactions on Automatic Control*, 62(6):3046–3053, 2017. doi: 10.1109/TAC.2017.2668380.

R. McLachlan and M. Perlmutter. Conformal Hamiltonian systems. *Journal of Geometry and Physics*, 39(4):276–300, 2001. doi: 10.1016/S0393-0440(01)00020-1.

M. Revay, R. Wang, and I. R. Manchester. Recurrent equilibrium networks: flexible dynamic models with guaranteed stability and robustness. *IEEE Transactions on Automatic Control*, 69(5):2855–2870, 2024. doi: 10.1109/TAC.2023.3294101.

T. K. Rusch and S. Mishra. Coupled oscillatory recurrent neural network (coRNN): An accurate and (gradient) stable architecture for learning long time dependencies. In *International Conference on Learning Representations*, 2021.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8459–8468. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/sanchez-gonzalez20a.html.

J. M. Sanz-Serna. Runge–Kutta schemes for Hamiltonian systems. *BIT Numerical Mathematics*, 28 (4):877–883, 1988. doi: 10.1007/BF01954907.

R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The Annals of Mathematical Statistics*, 35(2):876–879, 1964. doi: 10.1214/aoms/1177703591.

Yu. B. Suris. The canonicity of mappings generated by Runge–Kutta type methods when integrating the systems $\ddot{x} = -\partial u/\partial x$. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 29(2): 202–211, 1989.

Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020. doi: 10.1126/sciadv.aay2631.

Tycho F. A. van der Ouderaa, Mark van der Wilk, and Pim de Haan. Noether's razor: Learning conserved quantities. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 135943–135965. Curran Associates, Inc., 2024. doi: 10.52202/079017-4318. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/f5332c8273d02729730a9c24dec2135e-Paper-Conference.pdf.

Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 23078–23091. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/wang22aa.html.

Alistair White, Niki Kilbertus, Maximilian Gelbrecht, and Niklas Boers. Stabilized neural differential equations for learning dynamics with explicit constraints. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 12929–12950. Curran Associates, Inc., 2023. doi: 10.52202/075280-0568. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/2a4179ef39846557e99f6bfac580ea2e-Paper-Conference.pdf.

Jianke Yang, Robin Walters, Nima Dehmamy, and Rose Yu. Generative adversarial symmetry discovery. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 39488–39508. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/yang23n.html.

Z. Zhang, Y. Shin, and G. E. Karniadakis. GFINNs: GENERIC formalism informed neural networks for deterministic and stochastic dynamical systems. *Philosophical Transactions of the Royal Society A*, 380(2229):20210207, 2022. doi: 10.1098/rsta.2021.0207.

Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. Dissipative SymODEN: Encoding Hamiltonian dynamics with dissipation and control into deep learning. *arXiv preprint arXiv:2002.08860*, 2020a. URL https://arxiv.org/abs/2002.08860.

Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. Symplectic ODE-net: Learning Hamiltonian dynamics with control. In *International Conference on Learning Representations*, 2020b. URL https://openreview.net/forum?id=ryxmb1rKDS.

Appendix

Table of Contents

A Datasets, implementation details, theorems, and notation **15**

B Extended related work **20**

B.1 Structure preserving learned dynamics . . . 20
B.2 Governing equation and law discovery . . . 20
B.3 Coordinate and representation discovery . . . 21
B.4 Invariant and symmetry discovery . . . 21
B.5 Certified stability and invariant sets . . . 21
B.6 Exact constrained parameterisations . . . 21
B.7 Position of SPEC . . . 22

C Proofs for the SPEC construction **23**

C.1 Proof of the exact budget congruence . . . 23
C.2 Metric-preserving Cayley parameterisation . . . 23
C.3 Dissipative well-posedness and monotonicity . . . 24
C.4 Proof of exact radial closure . . . 24
C.5 Angular invariance of the radial guarantee . . . 25
C.6 Proofs of the five radial guarantees . . . 25
C.7 Coordinate covariance and scale invariance . . . 27
C.8 The Euclidean block metric as a design choice . . . 28
C.9 Blockwise composition and the exactness boundary . . . 28
C.10 Construction boundaries and coverage . . . 29

D SPEC algorithm **31**

D.1 Candidate radial maps . . . 31
D.2 Candidate fitting and validation . . . 31
D.3 Complete procedure . . . 32

E Blockwise computational cost **33**

F Extended experimental results **34**

F.1 Experimental configuration . . . 34
F.2 Q2: Full reconstruction-weight sweep . . . 34
F.3 Q2: Damped-system grid and what R^2 measures . . . 34
F.4 Q2: Real-data case study (NL-A3 orbit) . . . 35
F.5 Q3: Diagnostic classification and decision rules . . . 36
F.6 Q4: Aggregated error across horizons . . . 37

A DATASETS, IMPLEMENTATION DETAILS, THEOREMS, AND NOTATION

Table 6: **Source of datasets used in this research.**

Dataset	Details
Public datasets	
JPL Horizons (NL-A3)	Heliocentric ephemeris of the Earth-Moon barycentre (2020). Twenty non-overlapping windows of 400 steps drawn from one continuous orbit, so the segments are dependent. Used for the real-data case study (Appendix F.4) (Jet Propulsion Laboratory, 2023).
Synthetic datasets	
Simple pendulum	Mechanics. $H{=}p^2/2{-}\cos q$, $\Delta t{=}10^{-3}$, 48 trajectories from 4 energy levels $\times$ 12 initial angles.
Lotka-Volterra	Population dynamics. $\alpha{=}1.0$, $\beta{=}0.1$, $\gamma{=}1.5$, $\delta{=}0.075$, 48 trajectories, invariant $C{=}\delta V{-}\gamma \ln V{+}\beta W{-}\alpha \ln W$.
Kepler orbit (2D)	Mechanics. 6 eccentricities $\times$ 8 orbital phases, angular momentum $L{=}xp_y{-}yp_x$, leapfrog integrator, $\Delta t{=}0.01$.
Spring-mass chain (fixed M)	Mechanics. $N{=}10$, $M{=}I$, K tridiagonal, $a{=}0.02$, $b{=}0.002$, $\Delta t{=}0.01$, law-fitting window of 2000 steps, $n_{\mathrm{traj}}{=}8$.
Spring-mass chain (random M, K)	Mechanics. $N{\in}\{5, 10, 20\}$, $M, K{\sim}U(0.7, 1.4)$, otherwise as above.
Spring-mass chain (heavy damping)	Mechanics. $N{=}10$, $a{=}0.10$, $b{=}0.010$, $\Delta t{=}0.01$, $n_{\mathrm{traj}}{=}8$.
Spring-mass chain (N-ρ grid)	Mechanics. $N{\in}\{5, 10, 20, 40\}$, $\rho{\in}\{0.5, 1.0, 2.5, 5.0\}$, 5 seeds, laws evaluated in closed form. Five of the 1500 modes, all at $N{=}40$ and $\rho{=}5$, are overdamped and use the corresponding closed form. Used for the enforcement experiment. The Discover grid uses the same N with $\rho{\in}\{0.5, 1.0, 5.0\}$.
Mixed-regime hybrid	Mechanics. Rayleigh modes ($a{=}0.02$, $b{=}0.002$) in family and coupled pendula (coupling 0.3) out of family, combined into $B{\in}\{10, 20\}$ blocks at in-family fractions $\{0.25, 0.50, 0.75\}$, $\Delta t{=}0.01$, $n_{\mathrm{seeds}}{=}10$.
Van der Pol	Mechanics. $\mu{=}1.0$, $\Delta t{=}0.005$, $T{=}20000$, $n_{\mathrm{traj}}{=}50$, $x, y{\sim}U(-3, 3)$.
Diagnostic instances	Mixed. 280 supported and 180 unsupported blocks, $T{=}4000$, $n_{\mathrm{seeds}}{=}10$. Supported: Rayleigh chain, attracting sphere, random-M chain. Unsupported: pendulum, Van der Pol, coupled nonlinear, heavy-damped with noise.
Graded out-of-family blocks	Mixed. Three generators with a continuous difficulty knob: two-rate decay at rate ratio κ, drifting rate at drift ϵ, and exponential decay under observation noise σ. Used to locate the diagnostic boundary (Appendix F.5).
Simple pendulum (HNN baseline)	Mechanics. $H{=}p^2/2{-}\cos q$, $\Delta t{=}0.05$, $n_{\mathrm{traj}}{=}50$, seed$=0$. Used for the Hamiltonian baseline.

Table 9: **Glossary and notation.**

Symbol	Description
System and dynamics	
n	State dimension
Δt	Time step, listed per system in Table 7
$x \in \mathbb{R}^n$	State vector
$x_{t+1} = Ux_t$	One-step update
$U \in \mathbb{R}^{n\times n}$	Update matrix
$A \in \mathbb{R}^{n\times n}$	Dynamics generator
θ	Learnable parameters
$K \in \mathbb{R}^{n\times n}$	Skew generator, $K^\top = -K$
H_i	Unconstrained per-block parameter, with $K_i = H_i - H_i^\top$
$K_i,\ Q_i$	Per-block skew generator and the orthogonal operator $Q_i = \mathrm{Cay}(K_i)$
N	Number of modes in the spring-mass chain, $N \in \{5, 10, 20, 40\}$ on the enforcement grid

Table 9: **Glossary and notation** (continued).

Symbol	Description
$\omega_i,\ \omega_{d,i}$	Undamped and damped modal frequency, $\omega_d = \sqrt{\omega^2 - \gamma^2}$
$a,\ b$	Rayleigh damping coefficients, with $\gamma_i = (a + b\omega_i^2)/(2\omega_i)$
ρ	Damping scale multiplying (a, b), $\rho \in \{0.5, 1.0, 2.5, 5.0\}$ on the enforcement grid
$\zeta = \gamma/\omega$	Damping ratio, with M_γ positive definite for $\zeta < 1$
Budget identity	
$M \in \mathbb{R}^{n\times n}$	Alignment metric, required to be positive definite for dissipative well-posedness. The learned coordinates use $M = I$
$w = (I - A/2)^{-1}x$	Auxiliary vector
$\Sigma_M(A) = A^\top M + MA$	Budget operator, whose sign determines the per-step change of $x^\top Mx$
$\mathfrak{g}_M = \{A : \Sigma_M(A) = 0\}$	Conservation algebra
$G_M = \{U : U^\top MU = M\}$	Conservation group
$\mathrm{Cay}(A) = (I - A/2)^{-1}(I + A/2)$	Cayley map, defined wherever $I - A/2$ is invertible
M_γ	Metric under which a damped mode contracts exactly, off-diagonal entry γ/ω, eigenvalues $1 \pm \zeta$
Guarantee library	
$\mathcal{K}_{\mathrm{con}}$	Library of exactly constructible radial laws
$\mathcal{K}_i \subseteq \mathcal{K}_{\mathrm{con}}$	Candidate set evaluated for block i
$\mathcal{K}_{\mathrm{auto}} = \{\mathrm{con}, \mathrm{exp}, \mathrm{alg}\}$	Candidates searched without external input
$\mathrm{con},\ \mathrm{exp},\ \mathrm{alg}$	Labels of conservation, exponential decay and algebraic decay
$\mathrm{sph},\ \mathrm{reg}$	Labels of the attracting sphere and the forward-invariant region, enabled only when r_* or R is supplied
$\varnothing$	Refusal label, carrying no structural claim
ℓ_i	Label assigned to block i, either a member of $\mathcal{K}_i$ or $\varnothing$
$f_\ell(r; \eta)$	Radial map of law ℓ with parameters η
$\mathcal{E}_\ell$	Parameter range law ℓ admits. A candidate outside it is refused rather than projected
γ	Constant prescribed decay rate, giving conservation at $\gamma = 0$ and exponential decay at $\gamma > 0$
$\gamma_i,\ \hat{\gamma}_i$	Per-mode decay rate and fitted per-block decay rate
$\gamma(r)$	State-dependent rate function
γ_0	Rate coefficient of the attracting sphere and the forward-invariant region
$\sigma(\vartheta)$	Sigmoid applied to the sphere parameter, keeping μ inside $(0, 1)$
$r_*,\ R$	Target radius of the sphere and boundary radius of the region, both supplied externally
$\mu = \gamma_0 r_*^2 \Delta t$	Stability parameter of the sphere map, with local attraction for $0 < \mu < 1$
p	Degree of the algebraic decay law, $p > 0$
f	Continuous strictly increasing law with $f(0) = 0$, covered by the extension in Appendix C.6
F_{ξ_i}	Unconstrained update applied to a refused block
Block structure and encoder	
B	Number of latent blocks, 4 in the Discover runs
d_b	Width of each latent block, 2
$\Phi_\phi,\ D_\psi$	Encoder and decoder
$z_{i,t}$	Latent block i at time t
$r_{i,t} = \|z_{i,t}\|_2$	Block radius, the observable on which every guarantee is stated
$r_i(t) = \sqrt{q_i(t)^2 + (p_i(t)/\omega_i)^2}$	Block energy-norm radius in modal phase-space coordinates
$\delta_{i,t}^{(j)}$	Block log rate, $(\log r_{i,t+1}^{(j)} - \log r_{i,t}^{(j)})/\Delta t$
$L_{\mathrm{rate}}(\Phi)$	Log-rate term of the Discover objective, the temporal variance of $\delta_{i,t}$ within a trajectory, summed over blocks and averaged over trajectories, invariant under positive blockwise rescaling

Table 9: **Glossary and notation** (continued).

Symbol	Description
$L_{div}(\Phi)$	Comparison objective scoring a log-linear fit within a block partition against random re-partitions
L_{rec}	Reconstruction loss
λ_{rec}	Reconstruction weight, set to 0 for all headline results
L_{white}	Whitening term, the mean squared entry of $\widehat{\mathrm{Cov}}(z) - I$, applied with weight 0.1 at every setting including $\lambda_{\mathrm{rec}} = 0$
$\mathcal{I}_{\mathrm{disc}}$	Interval on which the encoder is estimated
$\mathcal{I}_{\mathrm{fit}}$	Interval on which candidate parameters are fitted
$\mathcal{I}_{\mathrm{val}}$	Disjoint interval on which candidates are scored and selected
Diagnostic	
$R^2_{\log}(i,k)$	Log-radius score of candidate k on block i, computed on $\mathcal{I}_{\mathrm{val}}$ and invariant under positive block rescalings
$\tau,\ \tau_d$	Acceptance threshold and its deployed value 0.30
$\widehat{\eta}_{i,k}$	Parameters of candidate k fitted on $\mathcal{I}_{\mathrm{fit}}$
k_i^*	Highest-scoring candidate for block i
κ	Rate ratio of the two-exponential out-of-family blocks
σ	Observation noise of the noisy exponential blocks
ρ_1	Lag-one autocorrelation of the log-radius residual, used to analyse the limit of the decision rule
Experiments	
$Q(t) = \sum_i r_i(t)^2$	Aggregate block energy, the observable used by the single-law baseline
$\bar{\gamma}(t)$	Effective aggregate decay rate
H	Hamiltonian, $p^2/2 - \cos q$ for the pendulum
$T_{\mathrm{fit}} = \lvert\mathcal{I}_{\mathrm{fit}}\rvert\Delta t$	Duration of the law-fitting interval
T_{eval}	Evaluation horizon, reported as a multiple of T_{fit}
$T_{\mathrm{char}} = 1/(\zeta\omega)$	Characteristic decay time
$\varepsilon_{\mathrm{truth}}$	Log modal-energy error of the extrapolated block radius against the analytical modal solution
$\varepsilon_{\mathrm{fit}}$	Deviation of the fitted law from the truth, with log half-amplitude $\frac{1}{2}\,\mathrm{artanh}(\gamma/\omega)$
$\varepsilon_{\mathrm{law}}$	Deviation of the rollout from the law that was imposed, held at floating-point precision by the exact construction

Table 7: **Architecture, optimisation, and evaluation settings.** All values are fixed before evaluation.

Setting	Value	Scope
Architecture		
Encoder	3 hidden layers, 64 units, Tanh, linear output	Analytic systems
Encoder	Linear map, no hidden layer	Damped-system grid
Decoder	Same topology as the corresponding encoder	Representation learning
Block count B	4 (Discover), $\{10, 20\}$ (routing), N (enforcement)	Per-experiment
Block width d_b	2	All reported runs
Optimisation		
Optimiser	Adam, no weight decay, full batch	Representation learning
Learning rate	3×10^{-3} (pendulum), $\{3 \times 10^{-3}, 10^{-2}\}$ (Discover grid)	Per-experiment
Discover steps	1200 (pendulum), 2400 (Discover grid)	Per-experiment
Whitening weight	0.1, applied at every setting including $\lambda_{\mathrm{rec}} = 0$	Representation learning
Reconstruction weight	$\lambda_{\mathrm{rec}} = 0$ for the headline results, swept over $\{0, 1, 10, 50, 100, 500, 1000\}$	Coordinate comparison
Time and evaluation		
Δt	0.01 (Diagnose and mechanics), 10^{-3} (pendulum), 0.05 (HNN baseline), $T_{\mathrm{char}}/200$ (modal rollouts)	Per-system
T_{fit}	One characteristic time	Discover and Enforce grids
Threshold τ_d	0.30 for the diagnostic evaluation. The routing benchmark also reports $\tau_d = 0$	Selection and routing
Numerical floors	Squared block norm clamped below at 10^{-30}, equivalently a radius floor of 10^{-15}. A floor of 10^{-300} on the log radius and on the $R^2_{\log}$ denominator	All radial computations
Seeds and hardware		
Random seeds	10 (Discover), 5 calibration and 5 test (Diagnose), 5 (Enforce)	Per-experiment
Thread counts	Q1 four; the analytic sweep four on the pendulum and one on Lotka-Volterra and Kepler; the damped grid two; Q4 one. Q3 is closed form and thread invariant	Per-experiment
Hardware	CPU. Software versions are recorded in the run manifests	Validated experiments

Table 8: **Theoretical results, grouped by topic.** The final column gives the subsection of Appendix C in which each statement is proved.

No.	Type	Statement	Proof
Budget identity			
1	Lemma	Budget identity: $(Ux)^\top M(Ux) - x^\top Mx = w^\top \Sigma_M(A)w$, exact for every admissible A and M	§C.1
2	Proposition	$\mathfrak{g}_M$ is a linear subspace of $\mathbb{R}^{n\times n}$	§C.2
3	Proposition	For every $A \in \mathfrak{g}_M$, the Cayley map satisfies $U^\top MU = M$	§C.2
4	Proposition	Arbitrary intersections $\bigcap_a \mathfrak{g}_{M_a}$ are linear subspaces, so a generator can support several conservation laws at once	§C.2
Well-posedness and obstructions			
5	Theorem	Dissipative well-posedness: $\Sigma_M(A) \preceq 0$ and $M \succ 0$ imply that $I - A/2$ is invertible	§C.3
6	Proposition	Pole placement: Cay is undefined exactly when $2 \in \mathrm{spec}(A)$, and a shifted map removes the singularity without changing group membership or coverage	§C.10
7	Proposition	No nonconstant polynomial map takes values in a bounded set, so a compact matrix group admits no polynomial parameterisation and a rational or transcendental map is required	§C.10
8	Proposition	No linear group action keeps a ball forward invariant without preserving the norm, so the regional construction requires a state-dependent rate	§C.10
Radial dynamics			
9	Proposition	Exact radial closure: the scalar recursion $r_{t+1} = e^{-\gamma(r_t)\Delta t} r_t$ holds independently of dimension and skew matrix K	§C.4
10	Proposition	Attracting sphere: $0 < \mu < 1$ is sufficient for local exponential attraction, and the parameterisation $\gamma_0 = \sigma(\vartheta)/(r_*^2\Delta t)$ makes $\mu = 1$ unreachable	§C.6
11	Proposition	Angular invariance: the radius sequence is independent of the skew parameters, so $\partial\mathcal{L}/\partial H = 0$ for any objective depending on the rollout only through the radii	§C.5
Coordinate invariance			
12	Theorem	The budget $w^\top \Sigma_M(A)w$ is a scalar invariant under any linear coordinate change	§C.7
13	Proposition	Objectives built on $d\|z\|_2^2/dt$ are scale dependent, with an explicit counterexample via $h(z_1, z_2) = (z_1, 2z_2)$	§C.7
14	Theorem	L_{rate} satisfies $L_{\mathrm{rate}}(h \circ \Phi) = L_{\mathrm{rate}}(\Phi)$ for all blockwise positive rescalings h	§C.7
15	Theorem	$R^2_{\log}$ is invariant under $r_i \to c_i r_i$ for any $c_i > 0$	§C.7
Metric alignment			
16	Proposition	Damped modal metric: the budget operator of the shifted generator $A_\omega + \gamma I$ is indefinite, so the Euclidean radius of the flow cannot follow a single exponential. Under M_γ with off-diagonal entry γ/ω the same construction gives $\Psi^\top M_\gamma \Psi = e^{-2\gamma\Delta t} M_\gamma$, and M_γ is positive definite exactly when $\gamma < \omega$	§C.8
Composition and boundary cases			
17	Proposition	Accepted block statements hold simultaneously under the product update, and a refused block carries no structural claim	§C.9
18	Proposition	The budget identity holds for arbitrary non-symmetric M, because the antisymmetric part contributes nothing to $w^\top \Sigma_M(A)w$	§C.1
19	Proposition	The Cayley image is exactly $\{U \in SO(n) : -1 \notin \mathrm{spec}(U)\}$, whose complement in $SO(n)$ is null. Coverage statements do not extend to $O(n)$	§C.10
20	Proposition	For any continuous strictly increasing f with $f(0) = 0$, the rate $\hat{\gamma}(r) = -(1/\Delta t)\ln(f(r)/r)$ is well defined, and a fixed point with $\|f'(r_*)\| < 1$ is locally exponentially attracting	§C.6

B EXTENDED RELATED WORK

This section expands the comparison in Section 2. Prior work relevant to SPEC spans two broad objectives: discovering structure from observations and enforcing structure through model construction. Within these objectives, we distinguish methods for structure-preserving dynamics (§B.1), governing equation and law discovery (§B.2), coordinate and representation discovery (§B.3), invariant and symmetry discovery (§B.4), certified stability and invariant sets (§B.5), and exact constrained parameterisation (§B.6). Table 10 summarises their relationship to the components of SPEC.

Table 10: **Relationship between SPEC and the principal lines of prior work.**

Line of work	Primary objective	Connection to SPEC	Key distinction
Structure-preserving dynamics	Encode a known physical or geometric property	Provides the architectural principle used in the Enforce stage	The constraint family and alignment metric are fixed before training, whereas SPEC selects both from trajectory evidence
Equation and law discovery	Infer governing relations from observations	Motivates testing which candidate law is supported by held-out trajectory evidence	Outputs a fitted relation without exact enforcement, whereas SPEC additionally enforces the selected law exactly for every parameter value
Coordinate and representation discovery	Find representations that expose simple dynamics or conserved quantities	Provides the closest connection to the Discover stage	Coordinates are chosen for prediction quality or linearisation, whereas SPEC searches for coordinates in which blockwise structural laws are expressible and validates them on a held-out criterion
Certified stability and invariance	Construct dynamics satisfying a specified certificate	Provides guarantee classes related to Diagnose and Enforce	The certificate type is specified externally, whereas SPEC diagnoses the appropriate guarantee from trajectory evidence before constructing the exact parameterisation
Constrained matrix parameterisation	Maintain a selected matrix constraint during optimisation	Provides the algebraic mechanism underlying the Cayley construction	Maintains a pre-selected constraint, whereas SPEC uses the per-step budget identity to determine which guarantees are achievable under a given generator-metric pair

B.1 STRUCTURE PRESERVING LEARNED DYNAMICS

Geometric numerical integration studies how a numerical update can preserve invariants and geometric structure. Quadratic invariant preservation for Runge Kutta methods was characterised by Cooper (1987). Related work established conditions for symplectic Runge Kutta schemes (Lasagni, 1988; Sanz-Serna, 1988; Suris, 1989) and conformal integrators (McLachlan & Perlmutter, 2001; Bhatt et al., 2016). These results are treated systematically by Hairer et al. (2006).

Structure preserving learning incorporates related principles into trainable dynamical models. Hamiltonian neural networks represent conservative dynamics through a learned Hamiltonian (Greydanus et al., 2019). Lagrangian neural networks and SympNets extend this approach to generalized mechanical coordinates and symplectic updates (Cranmer et al., 2020; Jin et al., 2020). GENERIC based architectures represent reversible and irreversible components through specified operators and degeneracy conditions (Grmela & Öttinger, 1997; Zhang et al., 2022), while dissipative extensions incorporate selected forms of energy loss (Zhong et al., 2020a).

These models cover several important forms of physical structure. Their guarantees follow from the architecture after the relevant conservation, dissipation, or geometric relation has been selected. The selection of that relation from multiple candidate behaviors is a separate problem.

B.2 GOVERNING EQUATION AND LAW DISCOVERY

A separate body of work infers dynamical laws directly from observations. Sparse identification of nonlinear dynamics uses sparse regression over a candidate library to recover governing equations (Brunton et al., 2016). Symbolic regression searches for interpretable expressions that explain observed relations (Udrescu & Tegmark, 2020). Related methods combine neural representations with sparse equation discovery to identify governing equations in learned coordinates (Champion et al., 2019).

These methods focus on recovering an equation or functional relation that explains the observations. Their output is generally selected according to predictive fit, sparsity, symbolic complexity, or a combination of these criteria. Such selection does not by itself produce an architecture for which the recovered relation holds exactly for every parameter value. Moreover, a relation that fits a bounded observation window may not retain the same validity over a held out horizon.

The Discover and Diagnose stages of SPEC are closest to this line of work. Their objective is not to recover an unrestricted governing equation. Instead, they test whether a member of a candidate structural family extrapolates and determine which exact guarantee, if any, is supported.

B.3 Coordinate and representation discovery

The form of a dynamical law depends on the coordinates in which it is represented. Koopman methods seek observables under which nonlinear dynamics admit a linear representation. Neural Koopman models learn such observables jointly with the latent dynamics (Azencot et al., 2020). Autoencoder based sparse discovery similarly learns coordinates together with a parsimonious governing equation (Champion et al., 2019).

Coordinate dependence also appears in structure preserving models. Hamiltonian formulations are naturally expressed in canonical variables, which may be unavailable in observed data. Lagrangian models operate in generalized coordinates (Cranmer et al., 2020), while neural symplectic form methods learn geometric structure when canonical position and momentum variables are unknown (Chen et al., 2021).

These methods seek coordinates that simplify prediction, linearise the dynamics, or expose a selected mechanical structure. The Discover stage of SPEC instead optimises a Fisher metric objective for coordinates in which candidate blockwise laws become expressible. The resulting coordinates are subsequently evaluated by the out of window diagnostic rather than accepted from reconstruction quality alone.

B.4 Invariant and symmetry discovery

Several methods infer conserved quantities or symmetries from data rather than assuming them in advance. Invariant discovery methods search for functions that remain constant along observed trajectories. Recent approaches combine conservation law discovery with model selection to recover candidate conserved quantities and their associated symmetries (van der Ouderaa et al., 2024).

Symmetry discovery methods infer transformations under which the observed dynamics or data distribution remain consistent (Yang et al., 2023). Other work permits controlled departures from a specified symmetry through residual pathways or approximate equivariance (Finzi et al., 2021; Wang et al., 2022).

These methods determine whether a particular invariant or symmetry is compatible with observations. SPEC considers a different but related decision. It assigns each retained coordinate block to one of several evolution laws, including conservation, decay, attraction, and forward invariance, and refuses enforcement when no candidate law extrapolates.

B.5 Certified stability and invariant sets

Contraction analysis certifies convergence by requiring neighboring trajectories to approach one another under a differential metric (Lohmiller & Slotine, 1998). The discrete time formulation provides related conditions for iterative systems (Lohmiller & Slotine, 1997), while control contraction metrics extend the framework to nonlinear control synthesis (Manchester & Slotine, 2017).

These ideas have informed learnable dynamical models. Recurrent equilibrium networks use contraction based conditions to construct stable recurrent dynamics (Revay et al., 2024). Stable neural differential equations parameterise vector fields with Lyapunov stability, exponential stability, or passivity guarantees (Cheng et al., 2024). Stabilized neural differential equations enforce specified manifold constraints through corrective dynamics (White et al., 2023). Antisymmetric and Lipschitz recurrent models constrain the recurrent operator to control stability and long range propagation (Chang et al., 2019; Erichson et al., 2021; Rusch & Mishra, 2021).

These approaches provide certificates for a selected stability property or invariant set. SPEC differs in the decision preceding certification. The guarantee type is diagnosed from held out trajectory behavior before the corresponding exact parameterisation is constructed.

Structured state space models form a neighboring but distinct line. HiPPO and S4 use structured state transitions for efficient long sequence modelling (Gu et al., 2020; 2022), while Mamba introduces input dependent selective state updates (Gu & Dao, 2024). Although these models use structured dynamics, they should not be identified with contraction certificates or exact physical constraint selection.

B.6 Exact constrained parameterisations

Exact matrix constraints can be maintained through differentiable parameterisations. Scaled Cayley transforms parameterise orthogonal recurrent matrices (Helfrich et al., 2018). Exponential maps convert elements of a Lie algebra into orthogonal, unitary, and more general connected compact Lie groups (Lezcano-Casado &

Martínez-Rubio, 2019). Their foundations are developed in structured matrix theory, matrix function theory, and Lie group integration (Mackey et al., 2003; Higham, 2008; Iserles et al., 2000). Sinkhorn normalisation provides a related construction for the doubly stochastic set (Sinkhorn, 1964).

These methods establish how a selected matrix constraint can be satisfied throughout optimisation. For the conservative case, the SPEC construction recovers the role of the Cayley map in producing a metric preserving update. The per step identity extends this construction by relating the generator, metric, and blockwise change through the sign structure of $\Sigma_M(A)$. This relation supports the five guarantee types considered by SPEC.

B.7 Position of SPEC

The preceding methods address different parts of structural learning. Equation discovery identifies candidate laws. Koopman and representation learning search for coordinates that simplify the observed dynamics. Invariant and symmetry discovery identify particular forms of structure. Certified models and constrained parameterisations enforce a selected property through model construction.

SPEC connects these objectives within a single conditional procedure. Discover searches for coordinates in which such laws become expressible. Diagnose assigns a supported guarantee type or refuses enforcement. Enforce constructs the corresponding Cayley parameterisation. The resulting model therefore combines trajectory based specification with a guarantee that holds exactly for every parameter value.

C PROOFS FOR THE SPEC CONSTRUCTION

C.1 PROOF OF THE EXACT BUDGET CONGRUENCE

Proof. Let

$$S_A := I - A/2, \qquad T_A := I + A/2.$$

Both are polynomials in A and therefore commute. Since S_A is invertible,

$$U = S_A^{-1} T_A = T_A S_A^{-1}.$$

Consequently,

$$\begin{aligned} U^\top M U - M &= S_A^{-\top} T_A^\top M T_A S_A^{-1} - S_A^{-\top} S_A^\top M S_A S_A^{-1} \\ &= S_A^{-\top} \left(T_A^\top M T_A - S_A^\top M S_A \right) S_A^{-1}. \end{aligned} \tag{13}$$

Direct expansion of the middle term yields

$$T_A^\top M T_A - S_A^\top M S_A = A^\top M + M A = \Sigma_M(A). \tag{14}$$

Substitution of Eq. 14 into Eq. 13 proves Eq. 4.

Because S_A^{-1} is invertible, $S_A^{-\top} \Sigma_M(A) S_A^{-1}$ is congruent to $\Sigma_M(A)$. Sylvester's law of inertia therefore implies that the two matrices have identical numbers of positive, negative, and zero eigenvalues. This proves the inertia and sign equivalences in Theorem 1. □

Per-step consequence. For $w = S_A^{-1} x$, multiplication of Eq. 4 on the left and right by $x^\top$ and x yields

$$(Ux)^\top M (Ux) - x^\top M x = w^\top \Sigma_M(A) w.$$

If $\Sigma_M(A)$ is indefinite, inertia preservation implies the existence of states producing a positive change and states producing a negative change.

Remark on nonsymmetric matrices. The algebraic expansion remains valid for an arbitrary square matrix M. However, $x^\top M x$ and the scalar budget depend only on

$$\mathrm{sym}(M) := \frac{M + M^\top}{2}.$$

Moreover, semidefinite order is defined for symmetric matrices. Accordingly, M is assumed symmetric throughout the main text, with $M \succ 0$ whenever $x^\top M x$ is interpreted as a squared norm.

C.2 METRIC-PRESERVING CAYLEY PARAMETERISATION

Proposition 2 (Metric-preserving Cayley map)**.** *The set*

$$\mathfrak{g}_M := \{A : A^\top M + M A = 0\}$$

is a linear subspace. For every $A \in \mathfrak{g}_M$ such that $I - A/2$ is invertible,

$$\mathrm{Cay}(A)^\top M \, \mathrm{Cay}(A) = M. \tag{15}$$

If $M = M^\top \succ 0$, every matrix $A = M^{-1} K$ with $K^\top = -K$ belongs to $\mathfrak{g}_M$.

Proof. The map

$$A \mapsto A^\top M + M A$$

is linear. Its kernel $\mathfrak{g}_M$ is therefore a linear subspace.

Define

$$B_- := I - A/2, \qquad B_+ := I + A/2.$$

For $A \in \mathfrak{g}_M$,

$$B_+^\top M B_+ - B_-^\top M B_- = A^\top M + M A = 0. \tag{16}$$

Multiplication of Eq. 16 from the left by $B_-^{-\top}$ and from the right by B_-^{-1} yields

$$(B_+ B_-^{-1})^\top M (B_+ B_-^{-1}) = M.$$

Because B_+ and B_- are polynomials in A, they commute. Therefore

$$B_+ B_-^{-1} = B_-^{-1} B_+ = \mathrm{Cay}(A),$$

which proves Eq. 15.

Finally, if $A = M^{-1}K$ and $K^\top = -K$, then

$$A^\top M + MA = K^\top + K = 0.$$

□

Proposition 3 (Simultaneous conservation)**.** *Let $\{M_a\}_{a\in\mathcal{A}}$ be an arbitrary family of symmetric matrices, indexed by a set $\mathcal{A}$ that need not be finite. Then*

$$\bigcap_{a\in\mathcal{A}} \mathfrak{g}_{M_a}$$

is a linear subspace of $\mathbb{R}^{n\times n}$. Every A in this intersection satisfies $\mathrm{Cay}(A)^\top M_a \,\mathrm{Cay}(A) = M_a$ simultaneously for all $a \in \mathcal{A}$, whenever $I - A/2$ is invertible.

Proof. Fix $a \in \mathcal{A}$. The map

$$\Sigma_{M_a} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}, \qquad A \mapsto A^\top M_a + M_a A,$$

is linear, so $\mathfrak{g}_{M_a} = \ker \Sigma_{M_a}$ is a linear subspace. The intersection of any family of linear subspaces of a vector space is again a linear subspace, because the zero matrix lies in every member, and if A and A' lie in every member and $\lambda \in \mathbb{R}$, then $A + \lambda A'$ lies in every member by closure of each. The second statement follows by applying Proposition 2 separately for each a, which is permitted because A lies in $\mathfrak{g}_{M_a}$ for every a and the invertibility hypothesis does not depend on a. □

A generator can therefore support several quadratic conservation laws at once, and the admissible generators still form a linear space.

C.3 Dissipative well-posedness and monotonicity

Proposition 4 (Dissipative well-posedness)**.** *Let $M = M^\top \succ 0$ and*

$$\Sigma_M(A) = A^\top M + MA \preceq 0.$$

Then $I - A/2$ is invertible and

$$\mathrm{Cay}(A)^\top M \,\mathrm{Cay}(A) \preceq M. \tag{17}$$

Proof. Suppose that $I - A/2$ were singular. Then there would exist $u \neq 0$ satisfying $Au = 2u$. Evaluation of the budget operator at u yields

$$u^\top \Sigma_M(A) u = (Au)^\top M u + u^\top M (Au) = 4u^\top M u > 0,$$

which contradicts $\Sigma_M(A) \preceq 0$. Hence $I - A/2$ is invertible.

For any x, the per-step consequence of Theorem 1 yields

$$x^\top \Big(\mathrm{Cay}(A)^\top M \,\mathrm{Cay}(A) - M \Big)\, x = w^\top \Sigma_M(A) w \leq 0.$$

The matrix inside the quadratic form is symmetric. Since the inequality holds for every x, Eq. 17 follows. □

C.4 Proof of exact radial closure

Proof. Orthogonality gives

$$\|Qz\|_2 = \|z\|_2.$$

For the update in Eq. 6,

$$\|\mathcal{U}_\alpha(z)\|_2 = \alpha(r)\|Qz\|_2 = \alpha(r) r.$$

For a radial map f satisfying $f(0) = 0$, the choice

$$\alpha(r) = \frac{f(r)}{r}, \qquad r > 0,$$

therefore gives

$$\|\mathcal{U}_\alpha(z)\|_2 = f(r).$$

At $r = 0$, define $\mathcal{U}_\alpha(0) = 0$. This proves Proposition 1. □

The exponential rate parametrisation corresponds to

$$\alpha(r) = e^{-\gamma(r)\Delta t},$$

for which

$$\|\mathcal{U}_\gamma(z)\|_2 = e^{-\gamma(r)\Delta t} r. \tag{18}$$

C.5 Angular invariance of the radial guarantee

The radial statements of Section 3.2 are asserted for every value of the trainable angular parameters. The following proposition is the formal content of that assertion.

Proposition 5 (Angular invariance)**.** *Fix a block width d_b and a measurable $\alpha : [0,\infty) \to [0,\infty)$. For $H \in \mathbb{R}^{d_b \times d_b}$ set $K = H - H^\top$ and $Q = \mathrm{Cay}(K)$, and define the update*

$$\mathcal{U}_H(z) = \alpha(\|z\|_2)\, Qz.$$

Let $z_0 \in \mathbb{R}^{d_b}$ be an initial state, let $z_{t+1} = \mathcal{U}_H(z_t)$, and let $r_t = \|z_t\|_2$. Then for every $T \geq 0$ the radius sequence $(r_0, \ldots, r_T)$ is independent of H. Consequently, if a loss $\mathcal{L}$ depends on the rollout only through that sequence, so that $\mathcal{L} = \Lambda(r_0, \ldots, r_T)$ for some $\Lambda : [0,\infty)^{T+1} \to \mathbb{R}$, then

$$\frac{\partial \mathcal{L}}{\partial H} = 0 \qquad \textit{identically in } H.$$

Proof. Because K is skew-symmetric and $I - K/2$ is invertible for every K of that form, $Q = \mathrm{Cay}(K)$ satisfies $Q^\top Q = I$ by Proposition 2 with $M = I$. Hence $\|Qz\|_2 = \|z\|_2$ for every z. Applying this to one step,

$$r_{t+1} = \|\alpha(r_t) Q z_t\|_2 = \alpha(r_t)\|Qz_t\|_2 = \alpha(r_t)\, r_t,$$

where the scalar $\alpha(r_t)$ is nonnegative and therefore passes through the norm. The right-hand side involves H nowhere. By induction on t, starting from $r_0 = \|z_0\|_2$, which does not involve H, the whole sequence $(r_0, \ldots, r_T)$ is a function of r_0 and α alone.

Therefore $H \mapsto \mathcal{L} = \Lambda(r_0, \ldots, r_T)$ is a constant function on $\mathbb{R}^{d_b \times d_b}$, and a constant function has vanishing derivative at every point. □

Gradient updates on H therefore cannot break an accepted radial guarantee, which is what makes the guarantee hold for every value of the trainable parameters rather than only at convergence. The same computation shows that the radial law carries no information about H, so the angular component is estimated from the prediction loss through the decoder, as in Algorithm 2.

C.6 Proofs of the five radial guarantees

All five statements follow from Proposition 1. The individual arguments below make each guarantee boundary explicit.

Conservation. For $\gamma(r) = 0$,

$$r^+ = e^0 r = r.$$

Thus, the block radius is conserved for every Q and every state.

Exponential decay. For a constant $\bar{\gamma} \geq 0$,

$$r_{t+1} = e^{-\bar{\gamma}\Delta t} r_t.$$

Iteration yields

$$r_t = e^{-\bar{\gamma} t \Delta t} r_0$$

exactly. Conservation is the boundary case $\bar{\gamma} = 0$.

Algebraic decay. Set

$$\gamma(r) = \frac{1}{2p\Delta t}\log\bigl(1 + 2pc\Delta t\, r^{2p}\bigr), \qquad c, p > 0.$$

Then

$$\begin{aligned} r^+ &= r\exp\left[-\frac{1}{2p}\log\bigl(1 + 2pc\Delta t\, r^{2p}\bigr)\right] \\ &= r\left(1 + 2pc\Delta t\, r^{2p}\right)^{-1/(2p)}. \end{aligned} \tag{19}$$

Raising both sides of Eq. 19 to the power $-2p$ gives

$$(r^+)^{-2p} = r^{-2p} + 2pc\Delta t.$$

For $r_0 > 0$, induction therefore yields

$$r_t = \left(r_0^{-2p} + 2pc\, t\Delta t\right)^{-1/(2p)}.$$

This is an exact discrete power law. The origin is treated separately. If $r_0 = 0$, then $\gamma(0) = \frac{1}{2p\Delta t}\log(1+0) = 0$, so $\alpha = 1$ and $r_1 = r_0 = 0$. By induction $r_t = 0$ for all $t \geq 0$. The simpler rate $\gamma(r) = cr^{2p}$ does not yield this identity and represents only a small-step approximation.

Attracting sphere. Let $r_* > 0$ and define

$$f(r) = r\exp\bigl[-\gamma_0(r^2 - r_*^2)\Delta t\bigr].$$

Since $f(r_*) = r_*$, the sphere of radius r_* is invariant. Differentiation yields

$$f'(r) = e^{-\gamma_0(r^2 - r_*^2)\Delta t}\left(1 - 2\gamma_0 r^2\Delta t\right),$$

and therefore

$$f'(r_*) = 1 - 2\mu, \qquad \mu := \gamma_0 r_*^2\Delta t.$$

The fixed radius is locally exponentially attracting whenever

$$|1 - 2\mu| < 1,$$

which is equivalent to

$$0 < \mu < 1.$$

The argument is the one given for general monotone laws at the end of this section, with contraction factor any ρ in $(|1 - 2\mu|, 1)$. The condition is sufficient and not necessary. At $\mu = 1$ one has $f'(r_*) = -1$, the linearisation is inconclusive, and stability would depend on higher-order terms. The parameterisation

$$\gamma_0 = \frac{\sigma(\vartheta)}{r_*^2\Delta t}$$

gives $\mu = \sigma(\vartheta) \in (0, 1)$ for every finite ϑ, so the inconclusive case is unreachable by construction and does not have to be resolved. This result establishes local attraction. It does not establish global convergence from the zero state or from arbitrary initial radii.

Forward-invariant region. Let

$$\gamma(r) = \gamma_0[r^2 - R^2]_+, \qquad \gamma_0 > 0.$$

If $r \leq R$, then

$$[r^2 - R^2]_+ = 0.$$

Proposition 1 therefore implies

$$r^+ = r \leq R.$$

Hence the closed ball

$$\{z : \|z\|_2 \leq R\}$$

is forward invariant. The statement does not require a trajectory outside the ball to enter it within one step.

Extension to general monotone laws. The five constructions are instances of a single scheme. Let $f : [0, \infty) \to [0, \infty)$ be continuous, strictly increasing, and satisfy $f(0) = 0$. Strict monotonicity together with $f(0) = 0$ gives $f(r) > 0$ for every $r > 0$, so the rate

$$\hat{\gamma}(r) := -\frac{1}{\Delta t} \ln\left(\frac{f(r)}{r}\right), \qquad r > 0,$$

is well defined and continuous on $(0, \infty)$, and Proposition 1 gives $r^+ = e^{-\hat{\gamma}(r)\Delta t} r = f(r)$ exactly. The origin is handled separately by $\mathcal{U}_\alpha(0) = 0$.

Suppose in addition that f is continuously differentiable near a fixed point $r_* > 0$ with $f(r_*) = r_*$ and $|f'(r_*)| < 1$. Choose ρ with $|f'(r_*)| < \rho < 1$. By continuity of f' there is $\delta > 0$ such that $|f'(r)| \le \rho$ for all r with $|r - r_*| \le \delta$. For such r the mean value theorem gives

$$|f(r) - r_*| = |f(r) - f(r_*)| \le \rho\, |r - r_*| \le \delta,$$

so the interval $[r_* - \delta, r_* + \delta]$ is forward invariant and the iterates satisfy $|r_k - r_*| \le \rho^k |r_0 - r_*|$. The fixed point is therefore locally exponentially attracting. Monotonicity is used only to guarantee positivity of f on $(0, \infty)$, and the stability argument holds for any continuously differentiable f with the stated fixed-point condition.

C.7 Coordinate covariance and scale invariance

Proposition 6 (Coordinate covariance and scale invariance)**.** *The budget scalar is invariant under an invertible linear change of coordinates. The rate term L_{rate} of the Discover objective is invariant under positive blockwise rescaling. The diagnostic score is invariant when the observed and predicted radii, including any supplied target radius, are rescaled together, provided the validation targets are non-degenerate (positive denominator in Eq. 28). These statements do not imply invariance under arbitrary nonlinear coordinate transformations.*

Proof.

Budget covariance. Let P be invertible and define

$$A' := PAP^{-1}, \qquad M' := P^{-\top} M P^{-1}, \qquad w' := Pw.$$

Direct substitution yields

$$\Sigma_{M'}(A') = P^{-\top} \Sigma_M(A) P^{-1}.$$

Consequently,

$$(w')^\top \Sigma_{M'}(A') w' = w^\top \Sigma_M(A) w.$$

Discover log-rate objective. Under

$$z'_{i,t} = a_i z_{i,t}, \qquad a_i > 0,$$

the transformed log rate is

$$\begin{aligned} \delta'_{i,t} &= \frac{\log(a_i r_{i,t+1}) - \log(a_i r_{i,t})}{\Delta t} \\ &= \delta_{i,t}. \end{aligned} \tag{20}$$

Therefore, the temporal variance of each blockwise log rate is unchanged. The remaining two terms of Eq. 10 are not invariant. The reconstruction term measures error in observation space, and the whitening term compares $\widehat{\mathrm{Cov}}(z)$ against the identity, so both fix the scale of z that L_{rate} leaves free, although neither constrains the mean of z.

Diagnostic score. Suppose that the observed and predicted radii, together with any supplied r_* or R, are multiplied by $a_i > 0$. Their logarithms and the validation mean are shifted by $\log a_i$. The residuals and centered observations in Eq. 28 are unchanged. Therefore, $R^2_{\log}$ is invariant under this joint rescaling. □

For comparison, objectives based directly on $d\|z\|_2^2/dt$ are scale dependent. For $z = (z_1, z_2)^\top$ and $h(z) = (z_1, 2z_2)^\top$,

$$\frac{d}{dt}\|z\|_2^2 = 2z_1\dot{z}_1 + 2z_2\dot{z}_2,$$

whereas

$$\frac{d}{dt}\|h(z)\|_2^2 = 2z_1\dot{z}_1 + 8z_2\dot{z}_2.$$

The two quantities differ whenever $z_2\dot{z}_2 \neq 0$.

C.8 THE EUCLIDEAN BLOCK METRIC AS A DESIGN CHOICE

Section 3.2 instantiates the identity at $M = I$, which gives the orthogonal Cayley component and the scalar radial update. The identity itself holds for every symmetric M. This section records what the Euclidean choice costs on a damped modal block and exhibits the metric under which the same construction is exact.

Proposition 7 (Damped modal metric)**.** *Fix $\omega > 0$ and $0 \le \gamma < \omega$, and take the block coordinates $z = (\eta, \dot\eta/\omega)$ in which a damped mode obeys $\dot z = A_\omega z$ with*

$$A_\omega = \begin{pmatrix} 0 & \omega \\ -\omega & -2\gamma \end{pmatrix}.$$

Write $\Psi = \exp(\Delta t\, A_\omega)$ and

$$M_\gamma := \begin{pmatrix} 1 & \gamma/\omega \\ \gamma/\omega & 1 \end{pmatrix}.$$

Then the following hold.

(i) *The Euclidean budget operator $\Sigma_I(A_\omega) = \mathrm{diag}(0, -4\gamma)$ is negative semidefinite and singular, while that of the shifted generator $N = A_\omega + \gamma I$ is $\Sigma_I(N) = 2\,\mathrm{diag}(\gamma, -\gamma)$, which is indefinite for every $\gamma > 0$. Consequently, for $\gamma > 0$ the flow $\exp(tA_\omega)$ is not a scalar multiple of an orthogonal matrix at every $t \ge 0$.*

(ii) *$\Sigma_{M_\gamma}(A_\omega) = -2\gamma M_\gamma$, and therefore*

$$\Psi^\top M_\gamma \Psi = e^{-2\gamma\Delta t} M_\gamma.$$

(iii) *M_γ has eigenvalues $1 \pm \gamma/\omega$ and is positive definite exactly when $\gamma < \omega$. On that range the radius $r_{M_\gamma}(z) = \sqrt{z^\top M_\gamma z}$ satisfies $r_{M_\gamma}(\Psi z) = e^{-\gamma\Delta t} r_{M_\gamma}(z)$ for every z, with no state dependence.*

Proof. For *(i)*, both operators follow by inspection. Since $\exp(tA_\omega) = e^{-\gamma t}\exp(tN)$ and $\mathrm{tr}\, N = 0$, so that $\det\exp(tN) = 1$, the matrix $\exp(tA_\omega)$ is a scalar multiple of an orthogonal matrix exactly when $\exp(tN)$ is orthogonal. If this held for every $t \ge 0$, differentiating $\exp(tN)^\top \exp(tN) = I$ at $t = 0$ would give $N^\top + N = 0$, contradicting $\Sigma_I(N) \ne 0$ for $\gamma > 0$. It can hold at individual times: because $N^2 = -\omega_d^2 I$ with $\omega_d = \sqrt{\omega^2 - \gamma^2}$, one has $\exp(tN) = \cos(\omega_d t) I + \sin(\omega_d t) N/\omega_d$, which equals $(-1)^k I$ at $t = k\pi/\omega_d$.

For *(ii)*, write $\zeta = \gamma/\omega$ and compute

$$A_\omega^\top M_\gamma + M_\gamma A_\omega = \begin{pmatrix} -2\gamma & -2\gamma\zeta \\ -2\gamma\zeta & -2\gamma \end{pmatrix} = -2\gamma M_\gamma.$$

Set $P(t) := \Psi(t)^\top M_\gamma \Psi(t)$ with $\Psi(t) = \exp(tA_\omega)$. Differentiating,

$$\dot P = \Psi^\top \big(A_\omega^\top M_\gamma + M_\gamma A_\omega\big)\Psi = -2\gamma P,$$

a scalar linear equation with $P(0) = M_\gamma$, whose solution is $P(t) = e^{-2\gamma t} M_\gamma$.

For *(iii)*, the eigenvalues of M_γ are $1 \pm \zeta$, so positive definiteness is equivalent to $\zeta < 1$. The radial statement follows from *(ii)* by taking square roots. □

Part *(i)* is the source of the residual reported in Section 4.4. Because $\Sigma_I(A_\omega)$ is singular, the Euclidean radius decays only when the state has a component along the damped direction, so its instantaneous rate depends on the state. After the average decay $e^{-\gamma t}$ is removed, the remainder is governed by $\Sigma_I(N)$, whose indefiniteness makes the Euclidean radius oscillate around an exponential rather than follow one. Writing $\zeta = \gamma/\omega$ for the off-diagonal entry of M_γ, the exact solution gives $r(t)^2 e^{2\gamma t} = C^2(1 + \zeta\cos\chi)$ for a phase χ advancing at $2\omega_d$, so the log radius oscillates with half-amplitude $\frac{1}{2}\,\mathrm{artanh}(\zeta)$, which is $\gamma/(2\omega)$ to first order.

Part *(iii)* bounds the construction. At $\gamma \ge \omega$ the mode is overdamped, M_γ is no longer positive definite, and r_{M_γ} is not a norm. An overdamped mode has two distinct real decay rates, while the update $\alpha(r)Qz$ applies one scalar factor to both directions of the block, so no metric in this family reproduces it.

C.9 BLOCKWISE COMPOSITION AND THE EXACTNESS BOUNDARY

Proposition 8 (Blockwise composition)**.** *Suppose each accepted block update $\mathcal{U}_i$ satisfies its assigned radial statement*

$$r_i^+ = f_i(r_i)$$

for every value of its free parameters. Then the product update satisfies all assigned block statements simultaneously. Blocks labelled ∅ carry no structural claim.

Proof. The latent state belongs to the direct product

$$\mathcal{Z} = \mathbb{R}^{d_b} \times \cdots \times \mathbb{R}^{d_b}.$$

For every accepted block i, Proposition 1 implies

$$\|z_{i,t+1}\|_2 = f_i(\|z_{i,t}\|_2)$$

independently of every other block. Therefore, all accepted block statements hold simultaneously. For $\ell_i = \varnothing$, the update F_{ξ_i} is unrestricted, and no radial statement is asserted. □

The decoder does not alter the latent recurrence when a rollout is performed entirely in latent space and decoded only for prediction or evaluation. For an observed state, the corresponding encoded quantity is

$$C_i(x) = \|[\Phi_\phi(x)]_i\|_2.$$

If a decoded prediction is re-encoded at every step, the composition $\Phi_\phi \circ D_\psi$ introduces an additional error term. Exactness under repeated re-encoding would require

$$\Phi_\phi(D_\psi(z)) = z$$

throughout the rollout. This condition is not assumed by the present construction.

C.10 Construction boundaries and coverage

This section establishes the limits of the construction: why a rational map is required, which part of the group the Cayley map misses, which singularities are removable, and why forward invariance of a bounded region cannot be obtained from a linear group action.

Proposition 9 (No polynomial parameterisation)**.** *Let $G \subseteq \mathbb{R}^{n\times n}$ be bounded and let $P : \mathbb{R}^N \to \mathbb{R}^{n\times n}$ be a polynomial map with $P(\mathbb{R}^N) \subseteq G$. Then P is constant. In particular, no nonconstant polynomial map takes values in a compact matrix group such as $O(n)$.*

Proof. Suppose P is not constant. Then some entry $P_{jk} : \mathbb{R}^N \to \mathbb{R}$ is a nonconstant polynomial, so there exist $a, b \in \mathbb{R}^N$ with $P_{jk}(a) \neq P_{jk}(b)$. Define the univariate polynomial

$$q(s) := P_{jk}\big(a + s(b-a)\big), \qquad s \in \mathbb{R}.$$

Then $q(0) = P_{jk}(a) \neq P_{jk}(b) = q(1)$, so q is nonconstant and has degree at least one. A univariate real polynomial of degree at least one satisfies $|q(s)| \to \infty$ as $|s| \to \infty$, so P_{jk} is unbounded on $\mathbb{R}^N$. This contradicts $P(\mathbb{R}^N) \subseteq G$ with G bounded. □

A parameterisation of a compact matrix group by an unconstrained Euclidean parameter must therefore be nonpolynomial. The Cayley map is rational and the matrix exponential is transcendental (Helfrich et al., 2018; Lezcano-Casado & Martínez-Rubio, 2019).

Proposition 10 (Coverage of the Cayley map)**.** *Let K range over the skew-symmetric matrices in $\mathbb{R}^{n\times n}$ and let $U = \mathrm{Cay}(K)$. Then*

$$\mathrm{Cay}\big(\{K : K^\top = -K\}\big) = \{U \in SO(n) : -1 \notin \mathrm{spec}(U)\},$$

and the complement of this set within $SO(n)$ has measure zero with respect to the Haar measure. In particular no U with -1 in its spectrum has a preimage, and $-I$ has none when n is even.

Proof.

The image lies in $SO(n)$. Orthogonality is Proposition 2 with $M = I$. For the determinant, skew-symmetry gives $\det(I - K/2) = \det\big((I - K/2)^\top\big) = \det(I + K/2)$, hence

$$\det U = \frac{\det(I + K/2)}{\det(I - K/2)} = 1.$$

No image point has -1 in its spectrum. From $U = (I - K/2)^{-1}(I + K/2)$,

$$I + U = (I - K/2)^{-1}\big[(I - K/2) + (I + K/2)\big] = 2(I - K/2)^{-1},$$

which is invertible, so $\det(I + U) \neq 0$ and -1 is not an eigenvalue of U. Taking $U = -I$ gives $I + U = 0$, which is not invertible, so $-I$ is not in the image.

Every such U is attained. Let $U \in SO(n)$ with $-1 \notin \operatorname{spec}(U)$, so that $I + U$ is invertible. Set $K := 2(U - I)(U + I)^{-1}$. Since U is orthogonal, $U^\top = U^{-1}$, and

$$K^\top = 2(U + I)^{-\top}(U - I)^\top = 2(U^{-1} + I)^{-1}(U^{-1} - I).$$

Writing $U^{-1} + I = U^{-1}(I + U)$ and $U^{-1} - I = U^{-1}(I - U)$ gives $K^\top = 2(I + U)^{-1}(I - U) = -K$, so K is skew-symmetric. Rearranging the definition of K gives $(I - K/2)U = I + K/2$, which is $\operatorname{Cay}(K) = U$.

The excluded set is null. The map $U \mapsto \det(I + U)$ is the restriction to $SO(n)$ of a polynomial in the matrix entries, hence real analytic on the connected analytic manifold $SO(n)$. It is not identically zero, because $\det(I + I) = 2^n \neq 0$. The zero set of a real-analytic function that does not vanish identically on a connected analytic manifold has measure zero with respect to any smooth measure, in particular the Haar measure. □

The restriction to $SO(n)$ cannot be relaxed. When n is odd, every $U \in O(n)$ with $\det U = -1$ has -1 as an eigenvalue, because its remaining eigenvalues are either $+1$ or come in conjugate pairs of unit modulus whose product is $+1$. That component lies entirely outside the image and is not a null subset of $O(n)$.

Proposition 11 (Pole placement)**.** $\operatorname{Cay}(A)$ *is undefined exactly when* $2 \in \operatorname{spec}(A)$. *For* $s \neq 0$ *define the shifted map*

$$\operatorname{Cay}_s(A) := \operatorname{Cay}(A/s) = (I - A/2s)^{-1}(I + A/2s).$$

Then $\operatorname{Cay}_s(A)$ *is defined for all but finitely many* s*, and if* $A \in \mathfrak{g}_M$ *then* $\operatorname{Cay}_s(A) \in G_M$ *for every such* s*. The image set is unchanged, that is* $\operatorname{Cay}_s(\mathfrak{g}_M) = \operatorname{Cay}(\mathfrak{g}_M)$.

Proof. The matrix $I - A/2$ is singular if and only if there is $u \neq 0$ with $(I - A/2)u = 0$, that is $Au = 2u$, which is the statement $2 \in \operatorname{spec}(A)$. Applying this to A/s shows that $\operatorname{Cay}_s(A)$ is undefined exactly when $2s \in \operatorname{spec}(A)$. Since $\operatorname{spec}(A)$ is finite, at most n values of s are excluded.

If $A \in \mathfrak{g}_M$ then $A/s \in \mathfrak{g}_M$, because $\mathfrak{g}_M$ is a linear subspace by Proposition 3 applied to a single metric, and linear subspaces are closed under scalar multiplication. Proposition 2 then gives $\operatorname{Cay}(A/s) \in G_M$. The same closure property shows that $A \mapsto A/s$ is a bijection of $\mathfrak{g}_M$ onto itself, so the two image sets coincide. □

The shift removes the singularity at an individual generator at the cost of reparameterising which A produces which U, which is the mechanism behind the scaled and pivoted variants used in practice (Helfrich et al., 2018). It does not enlarge coverage, since the set excluded by Proposition 10 is a property of the image.

Proposition 12 (No linear group action is forward invariant on a ball)**.** *Let* $R > 0$ *and let* $\mathcal{G} \subseteq GL(n, \mathbb{R})$ *be a group under matrix multiplication such that*

$$\|x\|_2 \leq R \implies \|Ux\|_2 \leq R \qquad \text{for every } U \in \mathcal{G}.$$

Then every $U \in \mathcal{G}$ *is orthogonal, so* $\|Ux\|_2 = \|x\|_2$ *for every* x*.*

Proof. Write $\|\cdot\|$ for the operator norm induced by the Euclidean norm. The hypothesis states that U maps the closed ball of radius R into itself, which by homogeneity is equivalent to $\|U\| \leq 1$. Because $\mathcal{G}$ is a group, $U^{-1} \in \mathcal{G}$, so the same hypothesis gives $\|U^{-1}\| \leq 1$. For any x,

$$\|Ux\|_2 \leq \|x\|_2 \qquad \text{and} \qquad \|x\|_2 = \|U^{-1}Ux\|_2 \leq \|Ux\|_2,$$

so the two are equal. A linear map preserving the Euclidean norm satisfies $U^\top U = I$. □

A construction built from a matrix group acting linearly on the latent state can therefore realise $r^+ = r$ and nothing else. It cannot contract a state that starts inside the ball, and it cannot bring a state from outside the ball into it. The forward-invariant construction of Section 3.2 is accordingly not a group action. Its update $z \mapsto e^{-\gamma_0[r^2 - R^2]_+ \Delta t} Qz$ is nonlinear in z through the radial factor.

D SPEC ALGORITHM

Algorithm 2 specifies the complete three-stage procedure. Notation follows Table 9. Each trajectory is partitioned chronologically into disjoint representation, fitting, and validation intervals before optimisation. The encoder is estimated on the representation interval and frozen before candidate laws are fitted, so the coordinate map is not selected using the horizon on which extrapolation is measured.

D.1 CANDIDATE RADIAL MAPS

For each latent block, candidate radii are obtained by iterating a scalar map. The maps corresponding to conservation, exponential decay, algebraic decay, attraction to a specified sphere, and forward invariance of a specified region are

$$f_{\mathrm{con}}(r) = r, \tag{21}$$

$$f_{\exp}(r;\bar{\gamma}) = e^{-\bar{\gamma}\Delta t} r, \qquad \bar{\gamma} \geq 0, \tag{22}$$

$$f_{\mathrm{alg}}(r;c,p) = r\left(1 + 2pc\Delta t\, r^{2p}\right)^{-1/(2p)}, \qquad c > 0, \tag{23}$$

$$f_{\mathrm{sph}}(r;\gamma_0,r_*) = r\exp\left[-\gamma_0(r^2 - r_*^2)\Delta t\right], \tag{24}$$

$$f_{\mathrm{reg}}(r;\gamma_0,R) = r\exp\left[-\gamma_0[r^2 - R^2]_+\Delta t\right]. \tag{25}$$

Here p is fitted as a continuous parameter. Write $\mathcal{E}_k$ for the set of parameters that law k admits, namely $\bar{\gamma} \geq 0$ for exp, a negative log-log slope for alg, which is the form in which the algebraic candidate is fitted, and $\mu = \gamma_0 r_*^2 \Delta t \in (0,1)$ for sph. A candidate whose fitted parameters fall outside $\mathcal{E}_k$ cannot be realised exactly, so Eq. 11 refuses it rather than projecting it onto the admissible set. Projection would enforce a law other than the one that was scored.

The automatic candidate set implemented in Diagnose is

$$\mathcal{K}_{\mathrm{auto}} = \{\mathrm{con}, \exp, \mathrm{alg}\}. \tag{26}$$

The sphere candidate sph is added only when a target radius r_* is supplied. In that case only γ_0 is fitted. Joint estimation of r_* and γ_0 is not included. The regional construction reg is not selected by trajectory fit. Its label is assigned only when a radius R is specified externally.

D.2 CANDIDATE FITTING AND VALIDATION

Let $\widehat{r}_{i,t}^{(j,k)}(\eta)$ denote the recursive radius prediction for trajectory j, block i, candidate k, and parameters η. Each fitting rollout is initialized from the first observed radius in $\mathcal{I}_{\mathrm{fit}}$. The candidate parameters are estimated only on that interval,

$$\widehat{\eta}_{i,k} = \arg\min_{\eta} \sum_{(j,t)\in\mathcal{I}_{\mathrm{fit}}} \left(\log r_{i,t}^{(j)} - \log \widehat{r}_{i,t}^{(j,k)}(\eta)\right)^2. \tag{27}$$

The validation rollout is initialized from the first observed radius in $\mathcal{I}_{\mathrm{val}}$ and propagated recursively without refitting. Define

$$y_{i,t}^{(j)} := \log r_{i,t}^{(j)}, \qquad \widehat{y}_{i,t}^{(j,k)} := \log \widehat{r}_{i,t}^{(j,k)},$$

and let $\overline{y}_i$ denote the validation mean. The held-out score is

$$R^2_{\log}(i,k) = 1 - \frac{\displaystyle\sum_{(j,t)\in\mathcal{I}_{\mathrm{val}}} \left(y_{i,t}^{(j)} - \widehat{y}_{i,t}^{(j,k)}\right)^2}{\displaystyle\sum_{(j,t)\in\mathcal{I}_{\mathrm{val}}} \left(y_{i,t}^{(j)} - \overline{y}_i\right)^2}. \tag{28}$$

The denominator of Eq. 28 carries an additive term $|\mathcal{I}_{\mathrm{val}}|\,\epsilon^2$ with $\epsilon = 10^{-6}$, applied on every evaluation, so that a block whose log radius is constant to within ϵ is scored against that floor rather than against round-off in its own variance. Radii are computed as $\sqrt{\max(\|z\|_2^2, 10^{-30})}$, and a floor of 10^{-300} is applied to the log radius and to the denominator of Eq. 28. These floors are numerical conventions that do not alter the exact radial constructions, whose statements concern the unclamped latent recurrence.

Algorithm 2 SPEC: Discover, Diagnose, and Enforce

Input: Trajectories $\mathcal{D}$, block count B, block width d_b, threshold τ_d, and optional targets $\{r_{*,i}, R_i\}$
Output: Encoder Φ_ϕ, decoder D_ψ, and block updates $\{\mathcal{U}_i\}_{i=1}^B$

Partition each trajectory chronologically into disjoint $\mathcal{I}_{\text{disc}}$, $\mathcal{I}_{\text{fit}}$, and $\mathcal{I}_{\text{val}}$

Stage 1: Discover
Estimate Φ_ϕ and D_ψ on $\mathcal{I}_{\text{disc}}$ by minimizing L_{discover} in Eq. 10
Freeze Φ_ϕ and compute $z_{i,t} = [\Phi_\phi(x_t)]_i$ and $r_{i,t} = \|z_{i,t}\|_2$

Stage 2: Diagnose
for $i = 1, \ldots, B$ **do**
  if a forward-invariant radius R_i is specified **then**
    $\ell_i \leftarrow \text{reg}$ ▷ R supplied; ball interior cannot identify exterior behavior
    continue
  end if
  $\mathcal{K}_i \leftarrow \mathcal{K}_{\text{auto}}$
  if an attracting radius $r_{*,i}$ is supplied **then**
    $\mathcal{K}_i \leftarrow \mathcal{K}_i \cup \{\text{sph}\}$ ▷ sphere evaluated only when r_* is known
  end if
  for each $k \in \mathcal{K}_i$ **do**
    Fit $\widehat{\eta}_{i,k}$ on $\mathcal{I}_{\text{fit}}$ using Eq. 27
    Initialize from the first validation radius and propagate f_k recursively
    Compute $R^2_{\log}(i,k)$ on $\mathcal{I}_{\text{val}}$
  end for
  $k_i^* \leftarrow \arg\max_{k \in \mathcal{K}_i} R^2_{\log}(i,k)$
  if $R^2_{\log}(i,k_i^*) \geq \tau_d$ **and** $\widehat{\eta}_{i,k_i^*} \in \mathcal{E}_{k_i^*}$ **then**
    $\ell_i \leftarrow k_i^*$
  else
    $\ell_i \leftarrow \varnothing$ ▷ score below τ_d, or a law that cannot be constructed exactly
  end if
end for

Stage 3: Enforce
for $i = 1, \ldots, B$ **do**
  if $\ell_i = \varnothing$ **then**
    $\mathcal{U}_i \leftarrow F_{\xi_i}$ ▷ unconstrained block update
  else
    $K_i \leftarrow H_i - H_i^\top$
    $Q_i \leftarrow \text{Cay}(K_i)$
    Take γ_{ℓ_i} from Eqs. 21–25 with the parameters $\widehat{\eta}_{i,\ell_i}$ that were scored
    $\mathcal{U}_i(z) \leftarrow e^{-\gamma_{\ell_i}(\|z\|_2)\Delta t} Q_i z$
  end if
end for
Freeze the diagnosed or supplied radial parameters
Optimize D_ψ, the skew parameters $\{H_i\}$, and the refusal updates $\{F_{\xi_i}\}$ using the task loss, with Φ_ϕ fixed
return Φ_ϕ, D_ψ, and $\{\mathcal{U}_i\}_{i=1}^B$

D.3 Complete procedure

Algorithm 2 states the complete SPEC procedure, in which Discover trains the encoder, Diagnose assigns a label to each block, and Enforce constructs the matched update. The encoder is frozen after Discover because changing the coordinate map after Diagnose would change the observable on which the law was selected, while the decoder, skew parameters, and refusal updates remain trainable for the prediction task. The exact radial statement is unaffected by the skew matrices because $\text{Cay}(H_i - H_i^\top)$ remains orthogonal for every H_i.

E BLOCKWISE COMPUTATIONAL COST

Let the latent dimension be

$$d_z = Bd_b,$$

where B is the number of blocks and d_b is the block width. Construction or refactorisation of a dense Cayley operator requires a factorisation of dimension d_z and costs

$$O(d_z^3)$$

per parameter update.

The blockwise construction instead performs B independent factorisations of dimension d_b. Its construction cost is therefore

$$O(Bd_b^3) = O(d_z d_b^2). \tag{29}$$

The corresponding operator storage is

$$O(Bd_b^2) = O(d_z d_b).$$

Once the block operators or their factorisations are cached, application to one latent state costs

$$O(Bd_b^2) = O(d_z d_b)$$

per rollout step. An implementation that refactorises the operators at every time step instead incurs the cubic construction cost at each step.

The deployed Enforce step uses $d_b = 2$, for which $\mathrm{Cay}(K_i)$ has a closed form and no factorisation is needed, so its cost is $O(Bd_b^2)$ with no caching required.

This notation separates the block count B from the block width d_b and from the Rayleigh damping coefficient b used in the synthetic systems.

F EXTENDED EXPERIMENTAL RESULTS

F.1 EXPERIMENTAL CONFIGURATION

Table 11 lists the configuration of each experiment.

Table 11: **Experimental configuration.** Q3 and Q4 are run on analytical modal coordinates, which isolates the decision and the enforcement from coordinate discovery. Datasets and every fixed hyperparameter are listed in Appendix A.

Experiment	Configuration
Q1: budget identity	Both sides of Eq. 5 under $M = I$, for $d \in \{2, 4, 8, 16, 32, 64, 128\}$, all four sign regimes of $\Sigma_M(A)$, ten seeds per regime, in float32 and float64.
Q2: analytic systems	L_{rate} against L_{div}, which scores an in-window log-linear fit within a partition of the blocks against the same quantity under random re-partitions, redrawn at every step. The two objectives are trained with the same whitening term of Eq. 10 at $\lambda_w = 0.1$ and the same reconstruction weight, and differ only in this first term. Pendulum ($H = p^2/2 - \cos q$), Lotka-Volterra ($C = \delta V - \gamma \ln V + \beta W - \alpha \ln W$) and Kepler orbit ($L = xp_y - yp_x$), each with 48 trajectories and a three-layer MLP encoder with $B = 4$ blocks of width $d_b = 2$, trained from ten seeds at every $\lambda_{\text{rec}} \in \{0, 1, 10, 50, 100, 500, 1000\}$.
Q2: damped grid	$N \in \{5, 10, 20, 40\}$ and $\rho \in \{0.5, 1.0, 5.0\}$ with a linear encoder, analytical modal coordinates as a reference, and phase-shuffled trajectories as a negative control.
Q3: routing	Mixed-regime systems at $\tau_d = 0.30$, over $B \in \{10, 20\}$ and in-family fractions $\{0.25, 0.50, 0.75\}$.
Q3: decision boundary	Out-of-family blocks generated at a controlled distance from the library, as sums of two exponentials with rate ratio κ or as single exponentials under observation noise σ.
Q4: enforcement	The same fitted rates imposed on the same blocks through one aggregate law over every block, a per-block penalty that holds each law approximately, and the exact per-block construction that holds it by parameterisation. $N \in \{5, 10, 20, 40\}$ and $\rho \in \{0.5, 1.0, 2.5, 5.0\}$ with five seeds, evaluated from T_{fit} to $100\,T_{\text{fit}}$.

F.2 Q2: FULL RECONSTRUCTION-WEIGHT SWEEP

Table 12 gives the complete predefined sweep for all three analytic systems.

Table 12: **Invariant R^2 across reconstruction weights for three systems.** Median over ten encoder seeds, with the best median of each column in bold. The log-rate median varies by at most 0.16 in any system, while the comparison median varies by up to 0.81 on the pendulum.

	Pendulum		Lotka-Volterra		Kepler	
λ_{rec}	L_{rate}	L_{div}	L_{rate}	L_{div}	L_{rate}	L_{div}
0	.728	.095	**.767**	.216	.858	.219
1	.791	.666	.693	.267	.851	**.237**
10	.820	.639	.721	**.386**	**.884**	.159
50	**.833**	**.902**	.658	.356	.827	.228
100	.797	.837	.604	.362	.837	.190
500	.695	.734	.708	.157	.793	.130
1000	.786	.581	.670	.132	.807	.110

F.3 Q2: DAMPED-SYSTEM GRID AND WHAT R^2 MEASURES

Table 13 breaks the damped-system grid down by damping scale, complementing the by-dimension view in the main text, and Table 14 gives the median $R^2_{\log}$ by horizon. The modal reference stays above 0.97 at every horizon, while both learned objectives peak at $2T_{\text{char}}$, at 0.07 for L_{rate} and 0.31 for L_{div}, and fall below the phase-shuffled control from $5T_{\text{char}}$ onward. Figure 7 shows the quantity behind the invariant R^2 directly.

Where the coordinates are right, the block radius is a function of the invariant alone, so the points at each level collapse onto a single value.

Constant offset. On the selected block of the reported encoders, a constant offset accounts for 91% to 96% of the radius, and the remaining variation is a per-trajectory level that predicts the invariant on held-out trajectories at R^2 of 0.97, 0.99 and 0.88 for the pendulum, Lotka-Volterra and Kepler systems. On a rerun with eight held-out trajectories, centring the latent inside the loss removes that level and lowers median R^2 from 0.76 to 0.36 on the pendulum and from 0.85 to 0.03 on Kepler, while Lotka-Volterra re-encodes the invariant in the fluctuation amplitude and rises from 0.74 to 0.86.

Table 13: **Damped-system grid: evaluations with $R^2_{\log} > 0.8$ (%) by damping scale ρ.** Pooled over system dimension, horizon, and learning rate.

Coordinates	$\rho = 0.5$	$\rho = 1.0$	$\rho = 5.0$
Modal (reference)	99.4	92.8	66.5
L_{rate} (learned)	6.8	0.4	0.1
L_{div} (learned)	15.5	6.0	0.9
Phase-shuffled	0.0	0.0	0.0

Table 14: **Damped-system grid: median $R^2_{\log}$ by horizon.** Pooled over system dimension, damping scale, seed, block and learning rate, with 960 evaluations per cell.

	Horizon (T_{char})				
Coordinates	0.5	1	2	5	10
Modal (reference)	.9755	.9940	.9986	.9998	.9999
L_{rate} (learned)	−.385	−.121	.067	−.412	−3.86
L_{div} (learned)	−.085	.117	.311	−.414	−4.51
Phase-shuffled	−.447	−.382	−.337	−.310	−.318

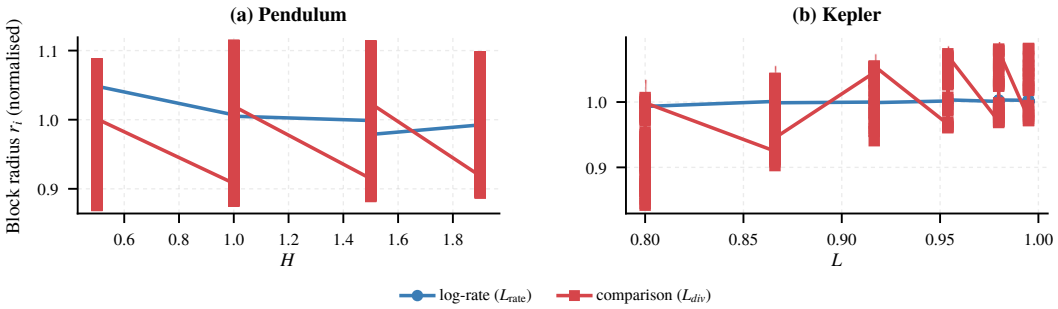

Figure 7: **Block radius against the conserved quantity, seed 0 at $\lambda_{\text{rec}} = 0$.** Each point is one time step of one trajectory, with the median at each level joined by a line. Under the log-rate objective the radius tracks the invariant with little spread at each level, whereas under the comparison objective the spread dominates, which is what the invariant R^2 measures. Values are for a single seed and differ from the medians in Table 2.

F.4 Q2: Real-data case study (NL-A3 orbit)

The three analytic systems use synthetic trajectories with ground-truth invariants. Table 15 reports results on the NL-A3 orbit system, which uses JPL Horizons ephemeris data for the Earth-Moon barycentre (heliocentric, 2020). The 20 trajectory segments are non-overlapping 400-step windows of one continuous orbit (Fig. 8a), so the segments are dependent rather than independent and seed-to-seed spread understates true uncertainty. The orbital energy varies only slightly across segments (relative standard deviation 3.4×10^{-3}), making invariant recovery structurally harder than on the pendulum or Kepler benchmarks. The block radius stays flatter under the log-rate objective (Fig. 8b), although the comparison objective leads at all seven reconstruction weights (Fig. 8c), reversing the ordering on the analytic systems.

Table 15: **NL-A3 orbit invariant R^2 at $\lambda_{\text{rec}} = 0$.** Median and seed range over ten encoder seeds. The two distributions overlap, and the comparison objective leads at all seven tested weights.

Objective	Median	Min	Max
L_{rate}	.514	.060	.728
L_{div}	.694	.071	.801

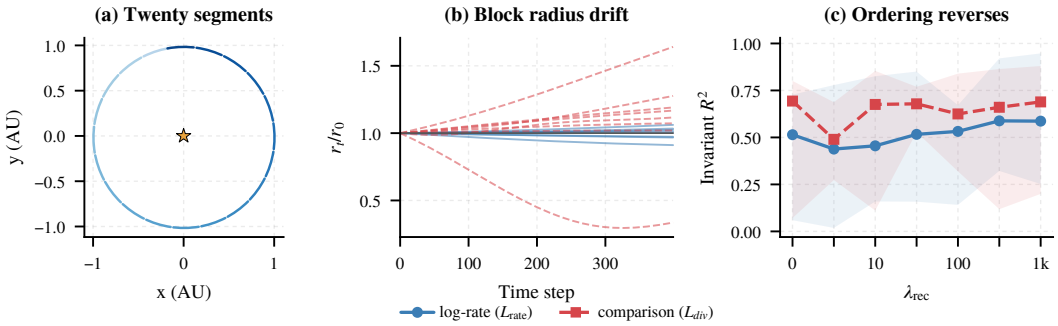

Figure 8: **NL-A3 orbit case study. (a)** The 20 segments, non-overlapping 400-step windows spanning one orbit of the Earth-Moon barycentre around the Sun, coloured by segment index. **(b)** Normalised block radius over a segment for two representative encoder seeds under each objective. **(c)** Invariant R^2 across reconstruction weights, with medians over ten seeds and shaded seed ranges.

F.5 Q3: Diagnostic classification and decision rules

The main text reports the cost of refusal and the two regimes in which the decision rule fails. This section reports the classification performance underlying those decisions, a residual-structure test used in place of the score, and the baseline against which a refused block is compared.

Classification. Three libraries are compared on disjoint test seeds: exponential decay alone, the automatic set of conservation, exponential decay and algebraic decay, and that set extended with an attracting-sphere law when r_* is supplied. The test split contains 120 exponential-decay, 20 attracting-sphere and 90 unsupported blocks, and every number below applies the constructibility gate of Eq. 11. At $\tau_d = 0.30$ the three-family library attains ROC-AUC 0.945 and PR-AUC 0.975, accepts all 120 exponential-decay blocks and accepts 3 of the 90 unsupported ones. The same three blocks pass in every library, so the false-acceptance rate is constant at .033. Adding the conditional sphere raises ROC-AUC to 0.984, an improvement not resolved at five seeds, so the three-family library is the deployable default since it requires no supplied r_*. Of the 12 sphere blocks that library accepts, four are named correctly and eight are assigned to algebraic decay, so the stage is reliable at refusal rather than at naming. Table 16 gives seed-level bootstrap intervals, which aggregate per-seed estimates and therefore differ from the pooled figures above.

Table 16: **Conditional-sphere uncertainty at $\tau_d = 0.30$.** Percentile bootstrap intervals for the median over five test seeds, with the constructibility gate applied. With four sphere blocks per test seed the interval for sphere acceptance is correspondingly wide.

Metric	Seed median	95% interval
ROC-AUC	.978	[.964, 1.000]
PR-AUC	.988	[.985, 1.000]
Overall sensitivity	.964	[.857, 1.000]
Specificity	.944	[.944, 1.000]
Exponential-decay acceptance	1.000	[1.000, 1.000]
Sphere acceptance	.750	[.000, 1.000]
Family accuracy given acceptance	.926	[.889, 1.000]
Refusal rate	.413	[.391, .457]

Residual-structure test. The statistic is the lag-one autocorrelation ρ_1 of the log-radius residual on $\mathcal{I}_{\text{val}}$, computed per block after refitting the selected family on that interval. It separates the two failure mechanisms completely, with median 0.9992 on wrong-form blocks and -0.0065 on noisy correct blocks, and an AUC

of 1.0. Used as a refusal rule in place of $R^2_{\log}$, it does not serve the decision. Every wrong-form block has $\rho_1 = 0.9992$ to four digits for κ from 1.1 to 20, so the rule refuses all of them at any threshold, including those with $\kappa \leq 3$, where enforcement lowers the error of every block. Its regret on wrong-form blocks stays between 1.92 and 2.21 across thresholds, against 0.160 for $R^2_{\log}$ at its best threshold. The rule also refuses every drifting-rate block, and, because the ripple of a damped mode gives in-family residuals a lag-one autocorrelation of $\cos(2\omega_d \Delta t)$, between 13% and 100% of correct in-family blocks depending on the threshold. Table 17 shows the cause: ρ_1 is constant across κ, while the residual amplitude, which the correlation normalises away, tracks the cost of enforcing. Combining the statistic with $R^2_{\log}$ does not help, since a rule requiring both inherits the refusals of each.

Table 17: **Residual structure against enforcement cost on two-rate blocks.** Medians per rate ratio κ, with enforcement cost at the fitting horizon, negative when enforcing lowers error.

κ	ρ_1	Residual std	Enforcement cost	$R^2_{\log}$ acceptance
1.1	0.9992	2.0×10^{-5}	-5.10	1.00
2	0.9992	1.0×10^{-3}	-2.64	1.00
3	0.9992	2.3×10^{-3}	-0.94	1.00
5	0.9992	3.9×10^{-3}	$+1.03$	0.89
8	0.9992	5.2×10^{-3}	-0.27	0.31
12	0.9992	5.5×10^{-3}	-0.30	0.15
20	0.9992	5.9×10^{-3}	-0.30	0.37

Refused-block baseline. In the routing benchmark a refused block holds its last observed radius. On the same analytical modal coordinates, an unconstrained least-squares linear block map, fitted on the fitting windows and rolled out per trajectory, recovers the true one-step map to round-off. At the horizon of Table 4 its in-family log-radius error is 5.8×10^{-13}, against 1.25×10^{-2} for the enforced law and 5.07 for the held radius, and at ten times that horizon the three are 5.8×10^{-12}, 1.0×10^{-1} and 50.7.

F.6 Q4: Aggregated error across horizons

Figure 9 aggregates the three enforcement settings over the grid, complementing the single-cell view in the main text.

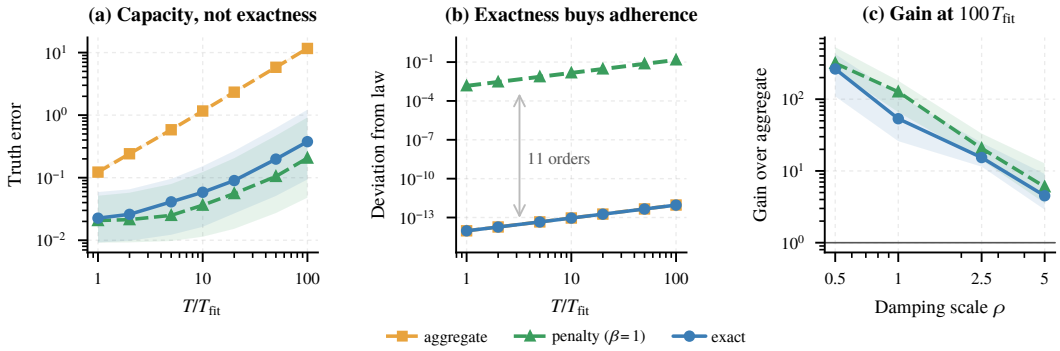

Figure 9: **Aggregated enforcement error across horizons.** **(a)** Truth error against horizon, with interquartile bands over 80 configuration-seed cells. The penalty matches the exact construction at T_{fit}, so the gain over the aggregate law comes from capacity. **(b)** Deviation from the imposed law. The aggregate and exact curves coincide, since both are exact constructions, while the penalty sits eleven orders of magnitude above at every horizon. **(c)** Gain over the aggregate law at $100\,T_{\text{fit}}$ by damping scale.