ON PARAMETERS OF NONLINEAR SCALAR DYNAMICS FROM VIDEO: INVARIANTS, CALIBRATION, AND IDENTIFIABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Physical parameter estimation from video aims to recover the parameters of a known family of governing dynamical equations from pixel observations. Existing identifiability theory for this setting has focused on linear time-invariant (LTI) second-order systems, leaving open what can be identified for nonlinear scalar dynamics. We develop an identifiability theory for nonlinear scalar second-order ODEs, organized by how their velocity dependence interacts with changes of the learned state coordinate. Under a shared non-collapsed state map and explicit same-state velocity-coverage conditions, we show that parameter identifiability depends on the ODE family: some parameters are uniquely identifiable, while in other families only invariant parameter combinations are identifiable or external physical calibration is required. For laws that are at most linear in velocity, compatibility forces affine coordinate alignment, yielding explicit parameter relations, invariants, and calibration conditions. This affine conclusion extends to broader finite velocity-feature families when coordinate curvature can be separated from the declared velocity dependence. For families admitting a squared-velocity term, nonlinear coordinate ambiguity can remain; a law-derived normalization instead enables affine comparison between canonical laws. Experiments on synthetic systems and real pendulum and free-fall videos support the predicted parameter relations, coverage effects, and calibration requirements.

1 INTRODUCTION

Physical parameter estimation from video aims to recover the parameters of a known family of governing dynamical equations from pixel observations. Recent video models increasingly seek to predict and interact with the physical world (Bruce et al., 2024; Assran et al., 2025), but predictive success alone does not establish which governing physical parameters are identifiable from visual observations. A more direct line of work therefore asks the inverse question: given a prescribed dynamical model, can its governing parameters be recovered from video? Physics-as-inverse-graphics methods estimate latent physical states and parameters using known differential equations (Jaques et al., 2020), while more recent encoder-based approaches estimate parameters of known continuous dynamics directly from video (Garcia et al., 2025). IRIS further systematizes this setting by pairing real videos with governing equations and independently measured physical parameters, enabling evaluation of parameter recovery and identifiability (Khanbayov et al., 2026). These developments establish physical parameter recovery from video as a concrete scientific-inference problem. Identifiability guarantees have begun to emerge in restricted settings, most notably for linear time-invariant dynamics, while corresponding results for nonlinear scalar systems remain substantially less understood. We discuss broader connections to video prediction, quantitative sensing, and physics-oriented evaluation in Appendix B.

For scalar linear time-invariant (LTI) second-order ODEs, Wang et al. (2026) establish affine coordinate alignment and exact coefficient identification under shared-state-map and trajectory-coverage assumptions. LTI is a favorable special case: the remaining family-preserving affine transformations leave its coefficients unchanged. But this no longer holds in general for nonlinear dynamics. For example, in the cubic Duffing law $z'' + \delta z' + \alpha z + \beta z^3 = 0$, rescaling the state by $\hat{z} = \lambda z$ leaves δ and α invariant under this transformation but maps β to β/λ^2. Thus even affine coordinate ambiguity

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.

can induce parameter ambiguity, while some velocity-dependent families can admit broader nonlinear coordinate ambiguity. This motivates our central question: *for nonlinear scalar dynamics, which parameter information remains identifiable from video, and when is external physical calibration sufficient to resolve the remaining ambiguity?*

Answering this question requires more than fitting a richer ODE. The learned state coordinate and the dynamical parameters must be analyzed jointly, because multiple coordinate–parameter descriptions may remain compatible with the same observed dynamics. Two issues then become central. First, passive videos constrain this ambiguity only through the states and velocities realized by the observed trajectories, making identifiability dependent on motion coverage as well as on the equation form. Second, external physical measurements may resolve residual parameter ambiguity, but their sufficiency must be characterized rather than replaced by an arbitrary normalization of the learned representation. We therefore develop a unified identifiability analysis for nonlinear scalar second-order ODEs, organized by how the velocity structure of the governing law interacts with coordinate changes. The physical and fitted dynamics belong to the same family fixed before training, and a shared encoder provides the scalar coordinate used for ODE fitting. Under a common non-collapsed C^2 state map and explicit same-state physical-velocity coverage, we characterize compatible parameter sets, identifiable invariants, and sufficient conditions for physical calibration.

Contributions. ① We extend identifiability theory from LTI dynamics to nonlinear scalar second-order ODEs with nonlinear state dependence and structured velocity dependence. We characterize when learned coordinates must be affine and how the remaining coordinate freedom transforms physical parameters; for polynomial-velocity families in which a squared-velocity term can absorb coordinate curvature, we instead obtain affine comparison after a law-derived canonical normalization. Under explicit representation conditions, we further establish realizability of the semilinear compatible parameter sets, yielding necessary and sufficient identification criteria. ② We turn the theory into a constructive tool for family-specific parameter analysis. Coefficient-weight rules expose invariant parameter combinations, rank conditions give sufficient local criteria for physical calibration, and an operational audit connects the required family, shared-map, coverage, canonical-coordinate, parameter-comparison, and calibration conditions to empirical diagnostics. ③ We validate the theory on synthetic systems and real pendulum and free-fall videos, supporting the predicted parameter relations, coverage effects, and calibration behavior.

2 Problem Setup and Identification Targets

Observations. We observe M passive video clips as timestamped frames $\mathcal{D} = \{(t_{m,k}, x_k^{(m)})\}$ at known recorded times. For the structural analysis, each clip samples an ideal frame path $x^{(m)}(t)$ and has an unobserved physical trajectory $z^{(m)} \in C^2(I^{(m)}; \mathbb{R})$. Clips may have different initial conditions; no controlled input is applied or observed. Full sampling notation is given in Appendix A.

Declared dynamics. Before fitting, we fix a parameter space Θ and a declared family of laws $F_\xi : D_\xi \times \mathbb{R} \to \mathbb{R}$, $\xi \in \Theta$, where each D_ξ is an open state interval. The physical trajectories share one unknown parameter $\theta \in \Theta$: each satisfies $z'' = F_\theta(z, z')$ on its interval, so all realized states lie in D_θ. The primary class studied in Section 3 is the semilinear family

$$z'' + a_\xi(z)z' + b_\xi(z) = 0, \qquad \xi \in \Theta, \tag{1}$$

with continuous a_ξ, b_ξ on D_ξ; equivalently, $F_\xi(u, v) = -a_\xi(u)v - b_\xi(u)$. Section 4 treats velocity-dependent extensions; the notation F_ξ does not extend the guarantees to arbitrary nonlinear families.

Direct encoder and shared state map. A single encoder E_ϕ, shared across all clips, maps each frame to a scalar latent, giving sampled outputs $\hat{z}_k^{(m)} = E_\phi(x_k^{(m)})$ and ideal paths $\hat{z}^{(m)}(t) = E_\phi(x^{(m)}(t))$. The learned latent state $\hat{z}$ is not assumed to equal the physical state z; Definition 2.1 below is the hypothesis relating the two.

Write $R_z = \bigcup_{m=1}^{M} z^{(m)}(I^{(m)})$ for the realized physical-state range; the domain condition above gives $R_z \subset D_\theta$. Fix a nonempty open interval $U \subset R_z$. All coverage assumptions and identification claims are stated on U: the data constrain the declared laws only at states the videos visit.

Definition 2.1 (Latent-state consistency)**.** The encoder trajectories are latent-state-consistent on U if there exists one map $f \in C^2(U)$, common to all clips, such that

$$\hat{z}^{(m)}(t) = f\Big(z^{(m)}(t)\Big) \quad \text{for every } (m,t) \text{ with } z^{(m)}(t) \in U. \tag{2}$$

The map is non-collapsed on U when $f' \not\equiv 0$ there; f' may still vanish at individual states.

Definition 2.1 adapts the state-consistency hypothesis of Wang et al. (2026) to multiple clips and the window U. Controlled capture, such as a static camera, fixed exposure and white balance, near-constant lighting, and a uniform background, supports this hypothesis in practice; it remains a modeling assumption rather than a certified property.

Identification target. A fitted parameter $\eta \in \Theta$ specifies a law in the same declared family, imposed directly on $\hat{z}$, with $f(U) \subset D_\eta$. At interior points with $z^{(m)}(t) \in U$, its continuous-time residual is $\mathcal{R}^{(m)}_{\eta,\phi}(t) := \hat{z}^{(m)\prime\prime}(t) - F_\eta(\hat{z}^{(m)}(t), \hat{z}^{(m)\prime}(t))$. A triple (E_ϕ, f, η) is an *admissible report on* U when it satisfies Definition 2.1, the common map is non-collapsed, all domain conditions hold, and the residual vanishes at the covering points: the interior trajectory points that realize the velocity coverage stated in each result. Our target is any parameter functional $\mathcal{J}$ satisfying $\mathcal{J}(\eta) = \mathcal{J}(\theta)$ for every admissible report; taking $\mathcal{J}$ to be the identity gives full parameter identification. When dynamics leave residual freedom, we ask which external physical calibrations identify further quantities.

Fitting objective. In practice, (ϕ, η) is estimated from the sampled frames: within each clip, time derivatives of $\hat{z}$ are formed by three-point centered differences, giving a sampled residual $\widehat{r}_{m,k}(\phi,\eta)$ at interior indices (Appendix C.1). The fitting objective is

$$\mathcal{L}_{\text{total}}(\phi,\eta) = \frac{1}{N_{\text{int}}} \sum_{m=1}^{M} \sum_{k=1}^{T_m-1} \widehat{r}_{m,k}(\phi,\eta)^2 + \lambda_{\text{var}} \mathcal{L}_{\text{var}}, \tag{3}$$

where $N_{\text{int}} = \sum_{m=1}^{M}(T_m - 1)$, $\lambda_{\text{var}} > 0$, and the per-clip variance-floor penalty $\mathcal{L}_{\text{var}}$ discourages a collapsed, near-constant encoder output. The identifiability results in this paper concern the ideal continuous-time setting; in practice we minimize this objective or the variants stated with each experiment, and Appendix C.2 quantifies the discretization and sampling gap between the two.

3 The Semilinear Core: Orbits, Invariants, and Anchors

We study parameter identification in the semilinear family (1), under the setup of Section 2. Trajectory coverage first restricts the learned coordinate to an affine map; the declared family then determines invariant quantities and how physical anchors can further constrain compatible parameters.

3.1 Coordinate Alignment and Compatible Parameters

Velocity coverage. The videos have *three-slope level coverage* on a nonempty open interval U if at least three pairwise-distinct physical velocities are realized at interior trajectory points at every $u \in U$. The points may come from one clip or several clips; the requirement concerns velocities at the same physical state, not the number of videos or repeated observations alone.

At the covering points, where $\hat{z}^{(m)} = f(z^{(m)})$ by Definition 2.1, the physical and fitted laws are

$$z^{(m)\prime\prime}(t) + a_\theta(z^{(m)}(t))z^{(m)\prime}(t) + b_\theta(z^{(m)}(t)) = 0, \tag{4}$$

$$\hat{z}^{(m)\prime\prime}(t) + a_\eta(\hat{z}^{(m)}(t))\hat{z}^{(m)\prime}(t) + b_\eta(\hat{z}^{(m)}(t)) = 0. \tag{5}$$

Theorem 3.1 (Three-slope affine collapse)**.** *Let $U \subset R_z$ be a nonempty open interval and suppose the videos have three-slope level coverage on U. Assume latent-state consistency through one $f \in C^2(U)$ with $f(U) \subset D_\eta$ and $f' \not\equiv 0$ on U. Let the coefficient functions in (4)–(5) be continuous on their respective domains, and assume the physical and encoder-induced paths satisfy those two laws at all covering points. Then $f(u) = \lambda u + \tau$ on U for constants $\lambda \neq 0$ and $\tau \in \mathbb{R}$; that is, $\hat{z}^{(m)}(t) = \lambda\, z^{(m)}(t) + \tau$ whenever $z^{(m)}(t) \in U$. Moreover, for every $u \in U$,*

$$a_\eta(\lambda u + \tau) = a_\theta(u), \qquad b_\eta(\lambda u + \tau) = \lambda b_\theta(u). \tag{6}$$

Fix a covering point with state $z^{(m)}(t) = u$ and velocity $z^{(m)\prime}(t) = v$. Substituting $\hat{z}^{(m)} = f(z^{(m)})$ into (5) and eliminating $z^{(m)\prime\prime}$ with (4) gives

$$0 = f''(u)v^2 + f'(u)\big(a_\eta(f(u)) - a_\theta(u)\big)v + b_\eta(f(u)) - f'(u)b_\theta(u). \tag{7}$$

Three distinct realized velocities are roots of this quadratic polynomial, so its coefficients vanish. In particular, $f'' = 0$ throughout U; non-collapse then gives a nonzero affine slope. The remaining identities are (6), which we call the *orbit equations*. The full proof is in Appendix H. Three slopes are sharp for a family-uniform affine-collapse guarantee over unrestricted continuous semilinear coefficients (Proposition E.1); narrower families may permit conclusions from fewer slopes.

From coordinates to parameters. Affine alignment does not alone determine η: the transformed law must also match a member of the declared coefficient family on the transported interval. Call $(\lambda, \tau) \in \mathbb{R}^* \times \mathbb{R}$ an *affine witness* for $\eta \in \Theta$ when $\lambda U + \tau \subset D_\eta$ and (6) holds on U. Define

$$\mathcal{O}_U(\theta) := \{\eta \in \Theta : \eta \text{ admits an affine witness on } U\}.$$

This is the *family-compatible outer set* on U; it need not be a group orbit or an equivalence class for general interval-local coefficients.

Theorem 3.2 (Affine compatibility and singleton identification)**.** *Fix a video collection whose physical trajectories are generated by θ in the semilinear family of Section 2, with three-slope level coverage on a nonempty open interval $U \subset R_z$. Every admissible report (E_ϕ, f, η) on U satisfies $\eta \in \mathcal{O}_U(\theta)$. Thus $\mathcal{O}_U(\theta) = \{\theta\}$ implies $\eta = \theta$ for every such report.*

Any functional constant on $\mathcal{O}_U(\theta)$ is therefore identifiable. Under explicit representation assumptions, the next result shows that this outer set is also realizable.

Proposition 3.3 (Realizability of affine-compatible parameters)**.** *Under the semilinear setup and coverage assumptions of Theorem 3.2, let $\mathscr{E}$ be an encoder class. Assume:*

(i) there exists $E_0 \in \mathscr{E}$ such that $E_0(x^{(m)}(t)) = z^{(m)}(t)$ whenever $z^{(m)}(t) \in U$;
(ii) $E \in \mathscr{E}$ implies $\lambda E + \tau \in \mathscr{E}$ for every $\lambda \neq 0$ and $\tau \in \mathbb{R}$.

For every $\eta \in \mathcal{O}_U(\theta)$ and each of its affine witnesses (λ, τ), the encoder $E_\eta = \lambda E_0 + \tau$ realizes an admissible report with $f(u) = \lambda u + \tau$. Consequently,

$$\{\eta \in \Theta : \exists\, (E, f, \eta) \text{ admissible on } U,\ E \in \mathscr{E}\} = \mathcal{O}_U(\theta). \tag{8}$$

Corollary 3.4 (Exact identification criterion)**.** *Under Proposition 3.3, a parameter functional $\mathcal{J}$ is identified across admissible reports in $\mathscr{E}$ if and only if it is constant on $\mathcal{O}_U(\theta)$. In particular, the full parameter is identified if and only if $\mathcal{O}_U(\theta) = \{\theta\}$.*

In other words, Theorem 3.2 shows that every admissible report must lie in $\mathcal{O}_U(\theta)$, but this alone does not imply that every point in the set is achievable by an encoder. Proposition 3.3 supplies the converse under assumptions (i)–(ii): if the physical coordinate can be represented and the encoder class is closed under affine output transformations, then every $\eta \in \mathcal{O}_U(\theta)$ is realized by an admissible encoder. Hence, under these representation assumptions, $\mathcal{O}_U(\theta)$ is exactly the parameter ambiguity left by the videos, and a quantity is identifiable precisely when it is constant on this set. Without these assumptions, only the containment result of Theorem 3.2 is guaranteed.

3.2 Parameter Invariants

Once translation is excluded, the degree of each monomial in a polynomial family sets its response to scaling; this is a calculation within that subclass, not an assumption on all semilinear laws.

Translation lock. A polynomial family is *translation locked* on a parameter stratum when every compatible affine witness has $\tau = 0$. A structural top gap suffices: the maximal declared degree $n \geq 1$ in the damping or restoring support has a nonzero coefficient at θ while degree $n - 1$ is structurally absent (Lemma H.1); Appendix J.2 illustrates the unlocked case on the Helmholtz family.

Once $\tau = 0$, the parameter transformation is diagonal in the polynomial coefficients.

Proposition 3.5 (Weight rule for a translation-locked polynomial family)**.** *Let U be a nonempty open interval and consider a translation-locked polynomial family with fixed structural supports I_a, I_b.*

Table 1: Family-specific coordinate ambiguity, identifiable parameter information, and sufficient physical calibration under the assumptions of Theorem 3.1. Details are given in Appendix J.

Declared family	Coordinate gauge	Invariant information	Sufficient parameter calibration
LTI $z'' + \delta z' + \alpha z = 0$	$\lambda u + \tau$, $\alpha\tau = 0$	δ, α	None
Angular pendulum $z'' + \delta z' + \omega^2 \sin z = 0$ $(\omega > 0)$	$\sigma u + 2\pi k$	δ, ω	None
Normalized Van der Pol $z'' + \mu(z^2 - 1)z' + z = 0$ $(\mu \neq 0)$	σu	μ	None
Cubic Duffing $z'' + \delta z' + \alpha z + \beta z^3 = 0$ $(\beta \neq 0)$	λu	δ, α, $\operatorname{sign}(\beta)$	One unsigned amplitude anchor for β
Quintic Duffing $z'' + \delta z' + \alpha z + \beta z^3 + \gamma z^5 = 0$ $(\beta\gamma \neq 0)$	λu	δ, α, $\operatorname{sign}(\beta)$, γ/β^2	One unsigned amplitude anchor for β, γ

Write the physical law as $a_\theta(u) = \sum_{i \in I_a} A_i(\theta) u^i$ *and* $b_\theta(u) = \sum_{j \in I_b} B_j(\theta) u^j$. *Assign weights* $w(A_i) = i$ *and* $w(B_j) = j - 1$. *For any compatible report* $\eta \in \mathcal{O}_U(\theta)$ *with scale witness* $u \mapsto \lambda u$,

$$A_i(\eta) = \lambda^{-w(A_i)} A_i(\theta), \qquad B_j(\eta) = \lambda^{-w(B_j)} B_j(\theta). \tag{9}$$

Every weight-zero coefficient, and every well-defined product or ratio whose net weight is zero, is therefore invariant across compatible reports.

Example: Duffing invariants, ambiguity, and calibration. Returning to cubic Duffing, $z'' + \delta z' + \alpha z + \beta z^3 = 0$ with $\beta \neq 0$, translation is locked and every compatible report satisfies

$$(\delta_\eta, \alpha_\eta, \beta_\eta) = (\delta_\theta, \alpha_\theta, \beta_\theta/\lambda^2).$$

Hence δ, α, and $\operatorname{sign}(\beta)$ are invariant, whereas the magnitude of β depends on the latent scale. Importantly, this is a realizable ambiguity rather than only an outer possibility: by Proposition 3.3, every allowed scale is realized by an encoder on the same videos, so $|\lambda| \neq 1$ produces genuinely different cubic coefficients even under exact continuous compatibility. For quintic Duffing with $\beta\gamma \neq 0$, $\gamma_\eta = \gamma_\theta/\lambda^4$, making γ/β^2 invariant. A single known nonzero unsigned physical displacement between states in U fixes $|\lambda|$ and therefore calibrates these scale-dependent coefficients, while coordinate reflection may remain. Thus the physical parameters can be calibrated without selecting a unique latent coordinate. Table 1 summarizes the corresponding conclusions for the other families.

3.3 Physical Calibration and Family-Specific Recovery

An external anchor supplies physical information not fixed by the ODE fit. For a translation-locked family, a signed pairing of $u_\star \in U \setminus \{0\}$ with $\hat{z}_\star = f(u_\star)$ fixes $\lambda = \hat{z}_\star/u_\star$ and hence all coefficients through (9). An unsigned amplitude pairing fixes only $|\lambda|$; reflection remains if both sign branches are admitted, and only odd-weight coefficients change under it. Variance normalization is not a physical anchor, and anchors outside U require a separately justified extension of the witness.

For general families, an anchor differential with rank equal to the dimension of the regular lifted witness–parameter set gives sufficient local calibration (Proposition G.1); disconnected branches require separate checks. Table 1 distinguishes parameter recovery from full coordinate calibration. All conclusions are relative to U and the stated parameterization. For the polynomial and real-analytic examples, the coefficient identities extend across the connected common domain.

3.4 Approximate Residuals

Theorem H.3 bounds the affine mismatch by residual size, level-matching error, and physical-velocity design conditioning. These are deterministic transfers, not pixel-to-parameter rates. Appendix H.6 gives the bounds and shows why encoder-coordinate conditioning is no substitute for physical-unit conditioning; Appendix C covers discretization and approximate state consistency.

4 BEYOND SEMILINEARITY: EXTENDING PARAMETER IDENTIFICATION

The semilinear core allows nonlinear state dependence but only the velocity features 1 and v. To extend its coordinate-to-parameter argument, consider the chain rule

$$\hat{z}'' = f'(z)z'' + \underbrace{f''(z)(z')^2}_{\text{coordinate curvature}} .$$

If the declared velocity features exclude v^2, sufficient coverage again forces raw affine alignment. If the family admits v^2, we instead compare laws after a law-derived coordinate change, under the polynomial-velocity hypotheses of Section 4.2. Both routes constrain the original ODE parameters; unique recovery remains family-dependent.

4.1 WHEN RAW AFFINE ALIGNMENT STILL HOLDS

Consider a fixed velocity-feature space V and laws

$$z'' = F_\xi(z, z'), \qquad F_\xi(u, \cdot) \in \mathsf{V} \quad (u \in D_\xi). \tag{10}$$

The space V is finite-dimensional, consists of continuous functions, contains 1, and is *dilation stable*: $\varphi(\alpha\,\cdot) \in \mathsf{V}$ for every $\varphi \in \mathsf{V}$ and $\alpha \in \mathbb{R}$, including zero. This closure keeps a rescaled velocity law within the same feature space.

Theorem 4.1 (Curvature-column exclusion)**.** *Retain the interval, latent-state-consistency, non-collapse, and domain assumptions of Theorem 3.1. Replace its two semilinear laws by (10), satisfied exactly at the covering points, and its three-slope coverage by the rank condition below. Suppose V has the properties above and $v^2 \notin \mathsf{V}$. For any basis $\varphi_1, \dots, \varphi_K$, require at every $u \in U$ realized physical velocities $v_1(u), \dots, v_{N_u}(u)$ at interior points such that*

$$M_u := \big[\, \varphi_1(v_i) \;\cdots\; \varphi_K(v_i)\; v_i^2 \,\big]_{i=1}^{N_u}, \qquad \operatorname{rank} M_u = K + 1. \tag{11}$$

Then $f(u) = \lambda u + \tau$ on U, with $\lambda \neq 0$ and $\tau \in \mathbb{R}$, and

$$F_\eta(\lambda u + \tau, \lambda v) = \lambda F_\theta(u, v) \quad (u \in U,\ v \in \mathbb{R}). \tag{12}$$

Full rank separates $f''(u)v^2$ from the declared features; the proof is in Appendix I. For $\mathsf{V} = \operatorname{span}\{1, v\}$ with basis $(1, v)$, the design is $N_u \times 3$; three distinct velocities select a nonsingular Vandermonde submatrix and recover the semilinear result.

Odd quadratic drag. For $F_\xi(u, v) = c_{0,\xi}(u) + c_{1,\xi}(u)v + c_{d,\xi}(u)v|v|$, any four distinct velocities including a positive and a negative value give full rank. Affine compatibility then implies

$$c_{d,\eta}(\lambda u + \tau) = \frac{c_{d,\theta}(u)}{|\lambda|}.$$

One-sided velocities cannot separate $v|v|$ from v^2; the rank proof and an exact one-sided construction, with its family restriction, are given in Appendix I.

Whenever the theorem applies, (12), with $\lambda U + \tau \subset D_\eta$, defines a family-compatible outer set and yields the same route to invariants and calibration as Section 3. Orbit terminology applies to restricted laws with transported domains; the parameter projection at fixed U need not be an equivalence class for general interval-local families.

4.2 WHEN AFFINE COMPARISON REQUIRES NORMALIZATION

A squared-velocity channel can absorb coordinate curvature, so raw affine alignment need not hold; the remedy is a corrective coordinate computed from each fitted law.

A coordinate determined by the law. Consider the fixed polynomial-velocity family

$$z'' = F_\xi(z, z') := \sum_{r=0}^{p} c_{r,\xi}(z)(z')^r, \qquad \xi \in \Theta. \tag{13}$$

The finite degree p and the coefficient forms are declared before fitting; $c_{r,\xi}$ are continuous on D_ξ, missing channels are zero-padded, polynomial dependence is required in velocity. The key point is that each law determines its own corrective coordinate: for each ξ, choose $u_{\xi,\star} \in D_\xi$ and let ψ_ξ solve

$$\psi_\xi'' + c_{2,\xi}\psi_\xi' = 0, \qquad \psi_\xi' > 0, \tag{14}$$

with $\psi_\xi(u_{\xi,\star}) = 0$ and $\psi_\xi'(u_{\xi,\star}) = 1$. Then ψ_ξ maps D_ξ diffeomorphically onto $J_\xi := \psi_\xi(D_\xi)$, and in the coordinate $q = \psi_\xi(z)$ the squared-velocity channel vanishes. We call the transformed equation the *canonical law* of ξ; its coefficients $\bar{c}_{r,\xi}$ are explicit (Appendix I). Normalization is a calculation from the fitted law, not physical calibration. The model case is a pendulum with angle q observed through the projection $y = \sin q$ on $|q| < \pi/2$ (Figure 9): the projected law has $c_2(y) = -y/(1-y^2)$, and the normalizer is exactly $\psi(y) = \arcsin y$ up to affine, so the canonical coordinate is the angle itself (Appendix F). The narrow angular-pendulum family itself excludes the projected law, so this example does not establish ambiguity within that family.

Parameter identification after normalization. Comparing each law in its own canonical coordinate restores the affine picture of Section 3. Call the canonical laws of θ and η *affine-compatible* on $\psi_\theta(U)$ if one affine map $A(q) = \lambda q + \tau$, $\lambda \neq 0$, carries $\psi_\theta(U)$ into J_η and matches their coefficients; the exact identities are (44) in Appendix I. Projecting to the declared parameters gives the canonical family-compatible outer set

$$\mathcal{O}_U^{\mathrm{can}}(\theta) := \{\eta \in \Theta : \text{the canonical laws are affine-compatible on } \psi_\theta(U)\}. \tag{15}$$

This is a set of *original parameter values* η: normalization changes how laws are compared, not the inference target. The transported-domain and equivalence-class caveats of Section 4.1 apply here.

Theorem 4.2 (Canonical compatibility containment for passive video)**.** *Let $U \subset R_z$ be a nonempty open interval, with the coefficient continuity and domains of (13). Set $d_v = \max\{2, p\}$. Assume at least $d_v + 1$ pairwise-distinct physical velocities are realized at interior trajectory points at every $u \in U$. Then every admissible report of Section 2 satisfies $\eta \in \mathcal{O}_U^{\mathrm{can}}(\theta)$ and*

$$\psi_\eta\Big(\hat{z}^{(m)}(t)\Big) = A\Big(\psi_\theta(z^{(m)}(t))\Big) = \lambda\psi_\theta(z^{(m)}(t)) + \tau \tag{16}$$

on every covered frame in U, for one affine A common to all clips.

The proof combines passive coefficient matching and exact raw–canonical compatibility (Proposition I.1 and Theorem I.2, Appendix I). Functionals constant on this outer set are identifiable; a singleton guarantees $\eta = \theta$. Constant viscous damping is already invariant before normalization (Corollary I.3). Physical anchors constrain $f = \psi_\eta^{-1} \circ A \circ \psi_\theta$, so a raw amplitude anchor need not be a canonical scale constraint. Local anchor counting proceeds as in Section 3.3 (Proposition G.1).

Table 7 summarizes the two theoretical routes and their conditions, and Figure 2 summarizes the workflow from joint training to parameter recovery. These results cover the stated feature-space and polynomial-velocity settings rather than arbitrary scalar ODEs; Appendix O provides an automated tool for analyzing new declared families under the same framework.

5 Experiments

We test our theoretical predictions using synthetic and real videos. A shared encoder maps each frame to a scalar latent z; the encoder and dynamical coefficients are learned jointly without state-coordinate targets. Section 5.1 tests coordinate ambiguity and invariant parameter information; Section 5.2 tests the roles of velocity coverage and canonical normalization; Section 5.3 examines physical parameter recovery through anchor calibration on three real cases.

Evaluation metrics. We use three metrics: (1) coordinate NRMSE e_{map} measures held-out coordinate recovery after a non-test-fitted alignment; (2) parameter relative error e_θ measures recovery at the corresponding parameter-orbit point or in calibrated physical units; and (3) dynamics relative RMSE e_{dyn} measures agreement of the transformed law with reference acceleration or force. Definitions and reporting conventions are in Appendix L, with interpretation limits in Appendix D.

5.1 Coordinate ambiguity and parameter information

To test the parameter information predicted in Section 3, we simulate the ODE forms in Table 1 and the quadratic/Helmholtz family (Appendix J.2), rendering pendulum or spring-oscillator videos

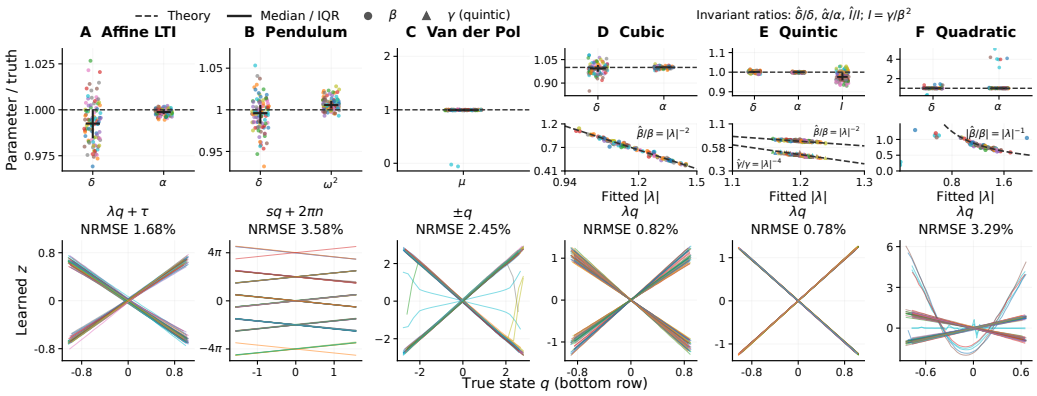


Figure 1: **Coordinate maps explain parameter orbits.** Top: invariant ratios and coefficient orbits; bottom: learned coordinates and allowed affine maps. Colors match across rows; D–F reuse training-fitted scales for testing.

Table 2: **Accurate coordinates do not ensure accurate drag coefficients.** Ten IC collections $\times$ five optimizer seeds per condition. Rank is the minimum for $[1, v, v|v|, v^2]$ over shared positions. K counts speed magnitudes; e_θ tests the scale-corrected mean drag coefficient.

Velocity design	Rank	Coordinate e_{map} (%)	Parameter e_θ (%)
One-way, $K = 6$	$3/4$	1.18 [1.08, 1.40]	44.97 [15.07, 59.87]
Two-way, $K = 1$	$2/4$	1.08 [0.91, 1.34]	6.95 [4.85, 8.65]
Two-way, $K = 2$	$4/4$	1.07 [0.93, 1.22]	1.34 [0.85, 2.60]

(Figure 3). Each family has fixed physical coefficients (Table 10) and rendering, with ten initial-condition (IC) collections crossed with ten optimizer seeds, giving 100 fits. IC collections vary the initial physical states; optimizer seeds control model initialization and training sampling. Allowed coordinate maps fitted on training references are frozen for held-out coordinate and parameter comparisons. Appendix M.1 gives the observations, clip splits, and training details.

Figure 1 supports the predicted link between coordinate ambiguity and parameter recovery. Affine LTI, pendulum, and normalized Van der Pol retain their respective affine, sign/periodic, and sign gauges while recovering nearly invariant coefficients. Cubic and quintic coefficients follow $\hat{\beta}/\beta \simeq |\lambda|^{-2}$ and $\hat{\gamma}/\gamma \simeq |\lambda|^{-4}$, while δ, α and the quintic ratio γ/β^2 remain nearly invariant (Eq. (9)); quadratic follows $\hat{\beta}/\beta \simeq \lambda^{-1}$ on its observed origin branch. Using the same training-fitted λ for coordinate and parameter-orbit comparisons gives median coordinate NRMSE of 0.78–3.58% and parameter relative error of 0.31–2.56% (Table 12). To examine ODE-family misspecification, we also fit affine-LTI models to the same observations and ICs with matched training settings: the declared nonlinear family gives lower test dynamics error e_{dyn} in 92–100 of 100 paired fits per family (Appendix M.1.6).

5.2 From velocity coverage to canonical coordinates

Using synthetic videos, we test Section 4's two routes: velocity coverage for raw affine alignment, and canonical normalization for laws with a squared-velocity channel.

Velocity coverage and parameter recovery. We test Section 4.1 on spring-cart videos governed by $q'' = -kq - \delta q' - \kappa q'|q'|$, with $(k, \delta, \kappa) = (0.8, 0.18, 0.22)$, and jointly learn an encoder and $z'' = c_0(z) + c_1(z)z' + c_d(z)z'|z'|$. Full rank of $[1, v, v|v|, v^2]$ predicts affine coordinates and the drag relation $c_d = -\kappa/|\lambda|$. All conditions use 24 training clips and 5,000 updates; observations, protocol details, and additional results are provided in Appendix M.2.

In Table 2, coordinate NRMSE stays near 1% while parameter errors differ substantially. One-way motion gives $v|v| = v^2$, so six speeds still cannot separate drag from coordinate curvature. Two directions and two magnitudes give full rank and reduce median drag error to 1.34%. Accurate coordinate fitting alone therefore does not ensure accurate parameter recovery; Figure 8 reports all six coverage conditions.

Canonical normalization. We test whether the fitted-law normalization in Section 4.2 removes coordinate distortion under the projected observation $y = \sin q$ (Figure 9). One encoder and law are

Table 3: **Law-derived normalization.** Median [Q25,Q75] over three fits.

Raw coordinate error e_{map} (%)	Canonical coordinate error e_{map} (%)	Restoring-force error e_{dyn} (%)
5.633 [5.629, 5.635]	0.305 [0.303, 0.305]	0.193 [0.183, 0.199]

Table 4: **Real-video coordinate and physical parameter recovery.** Estimates are mean $\pm$ SD; errors are median [Q25,Q75]. L is in m and g in $\mathrm{m/s^2}$.

Experiment	Parameter	Reference	Estimate	e_θ (%)	e_{map} (%)
Pendulum (nonlinear)	L	0.50	0.5150 ± 0.0012	2.92 [2.82, 3.17]	5.04 [4.38, 5.83]
Pendulum (LTI)	L_{eff}	0.50	0.5646 ± 0.0005	12.92 [12.83, 12.97]	7.13 [5.46, 7.45]
Side fall	g	9.81	9.370 ± 0.771	5.10 [3.21, 7.42]	3.75 [3.22, 5.14]
Overhead fall	g	9.81	10.146 ± 0.029	3.35 [3.26, 3.39]	5.57 [5.57, 5.60]

jointly trained on three equal-duration videos, released near $45°, 60°, 70°$. Raw and Canonical use the same held-out videos and matched affine readouts fitted on training references (Appendix M.3). From the result, normalization reduces coordinate NRMSE from 5.633% to 0.305% while recovering the restoring force to 0.193% relative RMSE (Table 3). Figure 10 shows the removal of systematic distortion. Appendix M.3.5 reports detailed results, followed by additional velocity-coverage and repeated-observation controls in Appendix M.3.6.

5.3 Physical parameter recovery from real videos

We use three real-video cases—real pendulum, side-view fall, and overhead fall—to test invariant coefficients, affine parameter orbits, and canonical normalization, respectively, and then obtain physical parameter estimates. The real pendulum recordings come from IRIS (Khanbayov et al., 2026); the side-view and overhead falls are self-captured. Each case contains multiple videos from repeated recordings under multiple initial conditions. All three cases use encoder-only training with an ODE-residual loss and no frame-reconstruction target. Observations, data details, training procedures, full evaluation, and additional results and discussion are provided in Appendix N.

Table 4 summarizes physical parameter recovery after anchor calibration. *(1) Pendulum.* With known gravity, the nonlinear pendulum coefficient yields a length of 0.5150 m against the IRIS reference 0.50 m, reducing median parameter error from the matched LTI model's 12.92% to 2.92%. This comparison shows the estimation cost of replacing the nonlinear restoring law by its small-angle approximation. *(2) Side fall.* A ball-diameter anchor converts the scale-corrected acceleration to $\hat{g} = 9.370\,\mathrm{m/s^2}$, with median error 5.10%. *(3) Overhead fall.* The learned curvature coefficient is $\hat{\rho} = 2.072 \pm 0.024$, close to the projective value $\rho = 2$. Canonical normalization reduces median coordinate NRMSE from 15.85% to 5.57%; a fixed post-fit release-height calibration then gives $\hat{g} = 10.146\,\mathrm{m/s^2}$, with median error 3.35%. These results illustrate how invariant recovery, affine parameter orbits, and canonical normalization combine with external calibration in real-video settings.

6 Discussion and limitations

We developed an identifiability analysis beyond LTI for the nonlinear scalar families studied here. It characterizes compatible parameter values, identifies invariant combinations, and gives sufficient conditions for recovering further parameter information through physical calibration. Synthetic and real-video experiments support the predicted parameter relations, coverage effects, and calibration behavior. The central implication is that parameter identification need not determine a unique state coordinate: residual coordinate freedom can leave the quantities of interest unchanged. At the studied quadratic-velocity boundary, canonical normalization enables law comparison, while physical calibration supplies separate information about the original parameters.

Limitations. The proposed analysis covers prescribed scalar autonomous ODE families and the stated velocity structures on regular coordinate intervals. It does not yet extend to coupled multidimensional systems, where finite compact smooth trajectories leave off-trajectory coordinate freedom that the scalar argument does not resolve. Appendix K explains this boundary, why the scalar setting remains informative, and the gap between structural identification and empirical parameter recovery.

Reproducibility Statement

The appendix provides complete proofs of the retained results, family-specific calculations, calibration conditions, and the sampled-data error interface. The experimental sections give their methods and reporting conventions, and the supplementary material includes all code for the synthetic experiments used to evaluate the theoretical predictions.

LLM Usage

Generative AI assistants were used as assistive tools for language editing, notation and cross-reference checks, literature-search support, proof-audit suggestions, mathematical derivation and revision, experimental planning, interpretation checks, and verification scripts. All AI-assisted suggestions, mathematical arguments, citations, experiments, and scripts were independently reviewed by the authors. The authors take full responsibility for the final content of the submission.

References

Kartik Ahuja, Jason Hartford, and Yoshua Bengio. Properties from mechanisms: An equivariance perspective on identifiable representation learning. In *International Conference on Learning Representations*, 2022.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Mojtaba Komeili, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, Sergio Arnaud, Abha Gejji, Ada Martin, Francois Robert Hogan, Daniel Dugas, Piotr Bojanowski, Vasil Khalidov, Patrick Labatut, Francisco Massa, Marc Szafraniec, Kapil Krishnakumar, Yong Li, Xiaodong Ma, Sarath Chandar, Franziska Meier, Yann LeCun, Michael Rabbat, and Nicolas Ballas. V-JEPA 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.

Yongsheng Bai, Halil Sezen, Alper Yilmaz, and Rongjun Qin. Bridge vibration measurements using different camera placements and techniques of computer vision and deep learning. *Advances in bridge engineering*, 4(1):25, 2023.

Markus W. Baumgartner, Anson Lei, Joe Watson, and Ingmar Posner. Disentangling dynamical systems: Causal representation learning meets local sparse attention. In *Proceedings of the Fifth Conference on Causal Learning and Reasoning*, volume 323 of *Proceedings of Machine Learning Research*, pp. 119–165. PMLR, 2026.

Johannes G. Borgqvist, Alexander P. Browning, Fredrik Ohlsson, and Ruth E. Baker. Framing local structural identifiability and observability in terms of parameter-state symmetries. *arXiv preprint arXiv:2603.11387*, 2026.

Jake Bruce, Michael D Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, et al. Genie: Generative interactive environments. In *Forty-first international conference on machine learning*, 2024.

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.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Sandra Díaz-Seoane, Xabier Rey Barreiro, and Alejandro F Villaverde. Strike-goldd 4.0: user-friendly, efficient analysis of structural identifiability and observability. *Bioinformatics*, 39(1): btac748, 2023.

R. Dong, C. Goodbrake, H. Harrington, and G. Pogudin. Differential elimination for dynamical models via projections with applications to structural identifiability. *SIAM Journal on Applied Algebra and Geometry*, 7(1):194–235, 2023. doi: 10.1137/22M1469067.

Wei Guo Foo, Julien Heyd, and Joël Merker. Normal forms of second-order ordinary differential equations $y_{xx} = J(x, y, y_x)$ under fibre-preserving maps. *Complex Analysis and its Synergies*, 9: 10, 2023. doi: 10.1007/s40627-023-00121-x.

Alejandro Castañeda Garcia, Jan Warchocki, Jan van Gemert, Daan Brinks, and Nergis Tomen. Learning physics from video: Unsupervised physical parameter estimation for continuous dynamical systems. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 27924–27933, 2025.

Rodrigo González Laiz, Tobias Schmidt, and Steffen Schneider. Self-supervised contrastive learning performs non-linear system identification. In *International Conference on Learning Representations*, 2025.

Florian Hofherr, Lukas Koestler, Florian Bernard, and Daniel Cremers. Neural implicit representations for physical parameter inference from a single video. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2093–2103, 2023.

Hoon Hong, Alexey Ovchinnikov, Gleb Pogudin, and Chee Yap. Sian: software for structural identifiability analysis of ode models. *Bioinformatics*, 35(16):2873–2874, 2019. doi: 10.1093/bioinformatics/bty1069.

Bronisław Jakubczyk. Existence and uniqueness of realizations of nonlinear systems. *SIAM Journal on Control and Optimization*, 18(4):455–471, 1980. doi: 10.1137/0318034.

Miguel Jaques, Michael Burke, and Timothy Hospedales. Physics-as-inverse-graphics: Unsupervised physical parameter estimation from video. In *International Conference on Learning Representations*, 2020.

Miguel Jaques, Martin Asenov, Michael Burke, and Timothy Hospedales. Vision-based system identification and 3d keypoint discovery using dynamics constraints. In *Proceedings of the 4th Annual Learning for Dynamics and Control Conference*, volume 168 of *Proceedings of Machine Learning Research*, pp. 316–329, 2022.

Krishna Murthy Jatavallabhula, Miles Macklin, Florian Golemo, Vikram Voleti, Linda Petrini, Martin Weiss, Breandan Considine, Jérôme Parent-Lévesque, Kevin Xie, Kenny Erleben, et al. gradSim: Differentiable simulation for system identification and visuomotor control. In *International Conference on Learning Representations*, 2021.

Sonia Joseph, Quentin Garrido, Randall Balestriero, Matthew Kowal, Thomas Fel, Shahab Bakhtiari, Blake Richards, and Mike Rabbat. Interpreting physics in video world models. *arXiv preprint arXiv:2602.07050*, 2026.

Rama Krishna Kandukuri, Jan Achterhold, Michael Moeller, and Joerg Stueckler. Physical representation learning and parameter identification from video using differentiable physics. *International Journal of Computer Vision*, 130(1):3–16, 2022.

Bingyi Kang, Yang Yue, Rui Lu, Zhijie Lin, Yang Zhao, Kaixin Wang, Gao Huang, and Jiashi Feng. How far is video generation from world model: A physical law perspective. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 28991–29017, 2025.

Robert M Kanko, Elise K Laende, Gerda Strutzenberger, Marcus Brown, W Scott Selbie, Vincent DePaul, Stephen H Scott, and Kevin J Deluzio. Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system. *Journal of biomechanics*, 122:110414, 2021.

Rasul Khanbayov, Mohamed Rayan Barhdadi, Erchin Serpedin, and Hasan Kurban. Iris: A real-world benchmark for inverse recovery and identification of physical dynamic systems from monocular video. In *European Conference on Computer Vision*, pp. 366–383. Springer, 2026.

Ilya Kossovskiy and Dmitri Zaitsev. Normal form for second order differential equations. *Journal of Dynamical and Control Systems*, 24(4):541–562, 2018. doi: 10.1007/s10883-017-9380-9.

Boris Kruglikov. Point classification of second-order ODEs: Tresse classification revisited and beyond. In Boris Kruglikov, Valentin Lychagin, and Eldar Straume (eds.), *Differential Equations: Geometry, Symmetries and Integrability*, volume 5 of *Abel Symposia*, pp. 199–221. Springer, Berlin, Heidelberg, 2009. doi: 10.1007/978-3-642-00873-3_10.

Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9:4950, 2018. doi: 10.1038/s41467-018-07210-0.

Paolo Muratore and Mackenzie Weygandt Mathis. Extracting governing equations from latent dynamics via multi-view contrastive learning. *arXiv preprint arXiv:2606.13260*, 2026.

Omar Mustafa. Position-dependent mass lagrangians: Nonlocal transformations, Euler–Lagrange invariance and exact solvability. *Journal of Physics A: Mathematical and Theoretical*, 48(22): 225206, 2015. doi: 10.1088/1751-8113/48/22/225206.

Peter J. Olver. *Equivalence, Invariants and Symmetry*. Cambridge University Press, Cambridge, 1995. ISBN 978-0-521-47811-3. doi: 10.1017/CBO9780511609565.

A. Paliathanasis, S. Moyo, and P. G. L. Leach. A geometric interpretation for the algebraic properties of second-order ordinary differential equations. *Mathematical Methods in the Applied Sciences*, 48:6912–6917, 2025. doi: 10.1002/mma.10726.

A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 08 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp358.

Yulia Rubanova, Ricky T. Q. Chen, and David K. Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Héctor J. Sussmann. Existence and uniqueness of minimal realizations of nonlinear systems. *Mathematical Systems Theory*, 10(1):263–284, 1976. doi: 10.1007/BF01683278.

Antonios Tragoudaras, Chenyu Zhang, Daniil Cherniavskii, Antonios Vozikis, Thijmen Nijdam, Derck WE Prinzhorn, Mark Bodracska, Nicu Sebe, Andrii Zadaianchuk, and Efstratios Gavves. Evaluating newtonian mechanics in video generative models with real physical systems. *arXiv preprint arXiv:2504.02918*, 2025.

Scott D Uhlrich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay S Chaudhari, Jennifer L Hicks, and Scott L Delp. Opencap: Human movement dynamics from smartphone videos. *PLoS computational biology*, 19(10):e1011462, 2023.

Alejandro F. Villaverde and Gemma Massonis. On testing structural identifiability by a simple scaling method: Relying on scaling symmetries can be misleading. *PLOS Computational Biology*, 17(10): e1009032, 2021. doi: 10.1371/journal.pcbi.1009032.

Yuanyuan Wang, Wenjie Wang, Kun Zhang, and Mingming Gong. Physics from video: Identifiability of time-invariant second-order odes under minimal trajectory conditions. In *International Conference on Machine Learning*, 2026.

Yan Xu and James MW Brownjohn. Review of machine-vision based methodologies for displacement measurement in civil structures. *Journal of Civil Structural Health Monitoring*, 8(1):91–110, 2018.

Dingling Yao, Caroline Muller, and Francesco Locatello. Marrying causal representation learning with dynamical systems for science. In *Advances in Neural Information Processing Systems*, volume 37, pp. 71705–71736, 2024. doi: 10.52202/079017-2290.

A Notation and Domain Conventions

A.1 Notation

Table 5 collects the main symbols. Primes on continuous-time paths denote time derivatives; primes on a state map or normalizer denote derivatives with respect to its scalar argument. The operators $D_{1,m}, D_{2,m}$ denote numerical differences instead.

Table 5: Notation index.

Symbol	Meaning
$\mathcal{D}$	Observed collection of timestamped video frames.
$m, M, I^{(m)}, T_m$	Clip index, number of clips, underlying time interval, and final frame index; clip m has $T_m + 1$ frames.
$t_{m,k}, \Delta t_m$	Recorded timestamp and known uniform within-clip sampling interval. All dynamical derivatives use this time coordinate.
$x^{(m)}(t), x_k^{(m)}$	Ideal frame path and its recorded sample at $t_{m,k}$.
$z^{(m)}(t)$	Unobserved data-generating scalar physical coordinate.
E_ϕ	Shared per-frame scalar encoder. Its output enters the ODE residual directly.
$\hat{z}^{(m)}(t), \hat{z}_k^{(m)}$	Ideal encoder trajectory and its sampled values.
$\Theta;\ \xi, \theta, \eta$	Declared parameter space; a generic member, the physical parameter, and a fitted encoder-coordinate parameter.
F_ξ, D_ξ	Declared right-hand-side law and its open state domain; $z'' = F_\xi(z, z')$. The form of F_ξ is specialized by each result.
a_ξ, b_ξ	Semilinear coefficients in $z'' + a_\xi(z)z' + b_\xi(z) = 0$.
$c, \hat{c}$	Physical and reported equilibrium coordinates in the affine-LTI family; distinct from the residual-coefficient vector $c(u)$.
f	Analysis-only shared state map, $f \in C^2(U)$; $\hat{z} = f(z)$ on the stated interval.
R_z	Union of realized physical state values, without velocity information; $R_z \subset D_\theta$.
U	Nonempty open physical-state interval on which coverage and identification claims are asserted.
$K = [u_-, u_+] \Subset U$	Nondegenerate compact interval contained in U, used for quantitative stability; K also denotes a basis size or a speed-magnitude count where stated locally.
$\mathcal{R}_{\eta,\phi}^{(m)}(t)$	Continuous-time fitted-law residual.
$\mathcal{J}(\xi)$	Parameter functional whose agreement across admissible reports is the identification target.
$D_{1,m}, D_{2,m}$	Three-point centered first- and second-difference operators within clip m.
$\widehat{r}_{m,k}, N_{\text{int}}$	Sampled residual and total number $\sum_m (T_m - 1)$ of interior sample indices.
$\mathcal{L}_{\text{ODE}}, \mathcal{L}_{\text{var}}, \mathcal{L}_{\text{total}}$	Residual loss, variance-floor penalty, and total sampled fitting objective.
$p, d_v, c_{r,\xi}$	Declared velocity degree, augmented degree $d_v = \max\{2, p\}$, and continuous coefficient functions in the polynomial-velocity specialization.
V	Fixed finite-dimensional velocity-feature space, with constants and dilation stability where required.
$\boldsymbol{\phi}_{\text{sl}}(v)$	Semilinear augmented feature vector $(v^2, v, 1)^\top$.
$\Phi_u, \sigma_{\min}(\Phi_u), \sigma_\star$	Physical-velocity design, its smallest singular value, and a positive lower bound. For semilinear matching, rows are $\boldsymbol{\phi}_{\text{sl}}(v_i)^\top$. Raw conditioning is scale-dependent.
M_u	Augmented feature design in the curvature-exclusion theorem, including the extra v^2 column.
ψ_ξ, J_ξ	Boundary-only normalizer satisfying $\psi_\xi'' + c_{2,\xi}\psi_\xi' = 0$ and its image $J_\xi = \psi_\xi(D_\xi)$.
$\bar{c}_{r,\xi}$	Coefficients of the canonical law with quadratic channel set to zero.
$A(q) = \lambda q + \tau$	Affine witness between canonical laws; the raw witness is $\psi_\eta^{-1} \circ A \circ \psi_\theta$.
g	Canonicalized state map $g = \psi_\eta \circ f \circ \psi_\theta^{-1}$; distinct from the gravity constant g in the real-video experiments.
$\mathcal{O}_U(\theta)$	Family-compatible affine parameter outer set on U, obtained by projecting out the affine witness.

Continued on next page

Table 5 (continued)

Symbol	Meaning
$\mathcal{O}_U^{\mathrm{can}}(\theta)$	Canonical compatible parameter outer set, restricted back to the declared Θ.
$\mathcal{S}_U(\theta)$	Lifted affine witness–parameter set retaining (λ, τ, η) rather than only η.
$\mathcal{S}_U^{\mathrm{can}}(\theta)$	Canonical lifted set; raw physical anchors are evaluated through the law-dependent inverse normalizer.
$d_{\mathcal{S}}$	Local manifold dimension used by the anchor-rank criterion.
$I_a, I_b;\ A_i, B_j$	Structural polynomial supports and coefficients of u^i in damping and u^j in restoring terms.
$w(A_i) = i,\ w(B_j) = j - 1$	Scaling weights in the translation-locked semilinear polynomial family.
$\varepsilon, h, V, \delta_v$	Residual tolerance, level mismatch, physical-velocity bound, and pairwise velocity separation; h also denotes a height where stated locally.
$\mathcal{T}_{\mathrm{sel}}, \varepsilon_{\mathrm{sc}}$	Selected clip–time pairs and uniform value/first-/second-derivative error for approximate continuous state consistency.
$c(u), L_c$	Semilinear residual-coefficient vector and its Lipschitz constant.
c_0, Λ	Bounds $0 < c_0 \le \lvert f' \rvert \le \Lambda$ used only in quantitative results; c_0 here is not $c_{0,\xi}$.
ξ_j, σ_0	Additional sample perturbation and its bound in Appendix C.2; ξ_j is distinct from a law index ξ.
$\delta_0, \delta_1, \delta_2$	Local value, first-difference, and second-difference errors relative to a smooth reference path.

A.2 Domain and Regularity Conventions

Each declared law is evaluated only on its stated domain. A shared map used on $U \subset R_z$ must satisfy $f(U) \subset D_\eta$. An affine witness $u \mapsto \lambda u + \tau$ must map U into D_η; a canonical affine witness must map $\psi_\theta(U)$ into J_η. These inclusions apply to all displayed coefficient identities and inverse-normalizer expressions. Sampled residuals likewise require their central encoded states to lie in the fitted coefficient domain.

A law identity proved on U is only a restricted-law statement unless a polynomial or real-analytic identity theorem is explicitly invoked on the connected common domain. No such extension is assumed for general continuous or merely smooth coefficients. Quantitative results state their compact intervals, bounds, and additional regularity locally. In particular, the C^4 condition used for centered-difference error bounds is not a standing hypothesis of the exact identification theory. Here $K \Subset U$ means that K is compact and contained in the open interval U.

Latent-state consistency in Definition 2.1 is imposed along the continuous trajectories whenever the physical state lies in U, whereas exact fitted-law compatibility is required only at the covering points specified by each result. The local condition is local in the physical coordinate, not a restriction to parameters near θ.

For general interval-local coefficients, the fixed-U parameter compatibility projection need not be symmetric and is therefore called an outer set. Literal law orbits carry their domains along with the coordinate change. Where a global law action is defined, intersecting an ambient law orbit with the declared family provides an orbit-based interpretation. These distinctions do not change the report-containment conclusions.

For semilinear laws, the right-hand-side convention is $F_\xi(u, v) = -a_\xi(u)v - b_\xi(u)$. In polynomial-velocity notation this means $c_{0,\xi} = -b_\xi$, $c_{1,\xi} = -a_\xi$, and all higher channels zero. Missing polynomial channels are zero-padded when matching two laws.

B Extended related work

We organize the literature around the distinction between predicting visual motion, estimating dynamical parameters, and establishing which parameter information is identifiable. These objectives connect the application background in the Introduction to the representation, dynamical, and calibration assumptions used in our analysis.

Video prediction, physical evaluation, and measurement. Genie learns action-controllable environments from videos without action labels (Bruce et al., 2024), while V-JEPA 2 learns predictive video representations and uses an action-conditioned model for robotic planning (Assran et al., 2025). Physics-oriented evaluations examine a different aspect of these models: whether their generated motion respects physical principles (Kang et al., 2025; Tragoudaras et al., 2025). In particular, Morpheus uses real physical settings and conservation-law-based metrics to evaluate Newtonian consistency. Work on interpreting video encoders also studies whether physical variables can be decoded from learned representations (Joseph et al., 2026). Prediction, physical consistency, and variable decodability are relevant to physical understanding, but none alone implies unique recovery of a prescribed equation's parameters. Our task concerns parameter information determined by observations and a declared dynamical family, rather than the physical plausibility of generated videos.

Vision-based sensing provides a practical motivation for quantitative inference from video. Xu & Brownjohn (2018) organize structural-displacement measurement around camera calibration, target tracking, and conversion to physical displacement; Bai et al. (2023) investigate camera placements and visual methods for bridge-vibration measurement. In biomechanics, Kanko et al. (2021) evaluate video-based spatiotemporal gait measurements, and OpenCap combines smartphone videos, learned motion estimation, and musculoskeletal simulation to estimate human movement kinematics and dynamics (Uhlrich et al., 2023). These works illustrate the value of geometric and physical calibration, rather than implying that all physical quantities follow from pixels alone. Our scalar results do not cover their full structural or multibody models; they address which parameter quantities dynamics constraints can determine and what external measurements add when the state coordinate is learned.

Physical parameter estimation from video. Physics-as-inverse-graphics estimates latent states and physical parameters using known differential equations and a generative observation model (Jaques et al., 2020). Differentiable-physics approaches couple visual inference to simulation (Kandukuri et al., 2022); gradSim additionally combines differentiable multiphysics simulation and rendering for pixel-based system identification and control (Jatavallabhula et al., 2021). Other methods use dynamics constraints for three-dimensional keypoint discovery (Jaques et al., 2022) or combine neural implicit appearance models with parameterized ODEs for single-video inference (Hofherr et al., 2023). More directly related to our fitting interface, Garcia et al. (2025) estimate parameters of known continuous governing equations without frame prediction and introduce Delfys75 for real-video evaluation. IRIS expands this evaluation setting with controlled real videos, governing equations, independently measured parameters, and protocols addressing accuracy, identifiability, extrapolation, robustness, and equation selection (Khanbayov et al., 2026). These methods already address nonlinear physical models. Our contribution is not to introduce nonlinear video-based estimation, but to characterize its parameter identification guarantees in the stated shared-encoder setting. We fix the family in advance, so governing-equation selection is outside our task.

Learning coordinates and governing equations. Sparse identification of nonlinear dynamics selects active terms from a candidate equation library (Brunton et al., 2016). SINDy autoencoders jointly learn coordinates and sparse equations (Champion et al., 2019), while learned Koopman embeddings seek coordinates with a simpler dynamical evolution (Lusch et al., 2018). Neural ODE and latent ODE models provide continuous-time learning architectures (Chen et al., 2018; Rubanova et al., 2019). These approaches motivate treating representation and dynamics jointly, but optimizing predictive fit or sparsity does not by itself specify which physical coefficients must agree across compatible coordinates. We instead fix the equation family and analyze its parameter transformations. A sparse or normalized representation is not automatically a physically calibrated one.

Identifiable dynamical representations and the LTI case. Nonlinear realization theory studies state descriptions up to isomorphism under its minimality and regularity assumptions (Sussmann, 1976; Jakubczyk, 1980). In representation learning, Ahuja et al. (2022) characterize ambiguity through transformations compatible with known mechanisms or mechanism classes. Yao et al. (2024) study time-invariant, trajectory-specific physical parameters and connect their identification to causal representation learning; Baumgartner et al. (2026) use local causal structure in parameter disentanglement. DynCL obtains latent-system identification through contrastive learning under stochastic-transition assumptions (González Laiz et al., 2025). DYSCO studies affine recovery from noisy multi-view observations under asymptotic assumptions and also examines sparse symbolic recovery along affine coefficient orbits (Muratore & Mathis, 2026). These are substantive nonlinear identification results, not merely predictive models. Our complementary question concerns deterministic compatibility of a

shared scalar frame encoder with a fixed second-order ODE family, using realized same-state velocity coverage.

The closest direct video-based starting point is Wang et al. (2026), which establishes affine coordinate alignment and coefficient identification for scalar LTI second-order ODEs. We extend the structural analysis to nonlinear state and structured velocity dependence, where affine alignment need not fix every coefficient and canonical-law comparison may be needed. We distinguish family-compatible outer sets from the parameter reports realizable by an encoder class. For semilinear laws, true-coordinate representability and closure under invertible affine output transformations make these sets exactly realizable (Proposition 3.3), yielding necessary and sufficient identification criteria. This existence result is distinct from statistical consistency or convergence of a learning objective. The LTI-specific damping-regime and stochastic finite-sample results remain separate contributions of Wang et al. (2026).

Structural identifiability, invariants, and calibration. Structural identifiability asks whether a specified state–output model uniquely determines parameters from ideal observations, including when some states are unobserved. SIAN and STRIKE-GOLDD provide computational tests for their respective model classes (Hong et al., 2019; Díaz-Seoane et al., 2023); differential elimination offers another algebraic route (Dong et al., 2023). Borgqvist et al. (2026) connect locally identifiable combinations to invariants of output-preserving parameter–state symmetries. Villaverde & Massonis (2021) show that examining scaling symmetries alone can miss non-identifiability. Our coefficient-weight rule is therefore used after affine reduction and a family-specific translation-lock argument, not as a test for arbitrary nonlinear systems. Our added structure is the explicit relation between passive-trajectory coverage, admissible encoder coordinates, and compatible parameter sets. External anchors constrain a lifted set retaining both the coordinate witness and the parameters (Appendix G). The restricted-Jacobian criterion is a sufficient local condition; it neither settles all singular cases nor removes disconnected branches automatically. The calculus and audits support family-specific analysis, rather than constituting a general automatic structural-identifiability solver.

Structural uniqueness and accuracy under finite noisy data are different questions. Profile-likelihood analysis explicitly studies structural and practical non-identifiability and parameter uncertainty (Raue et al., 2009). Our semilinear residual-stability results are conditional deterministic bounds, not confidence intervals or pixel-to-parameter statistical rates. The reporting checks in Appendix D distinguish fitted-law agreement, state-consistency evidence, velocity coverage, admissible parameter comparisons, and physical calibration. They connect the theory to empirical evaluation without treating finite audits as proofs of continuous-time premises.

ODE equivalence and coordinate normalization. Equivalence and invariants of differential equations are classical topics (Olver, 1995; Kruglikov, 2009), with modern normal-form results for second-order equations under point and fibre-preserving maps (Kossovskiy & Zaitsev, 2018; Foo et al., 2023). Geometric elimination of quadratic derivative terms and transformations of position-dependent-mass Lagrangians provide context for our normalization (Paliathanasis et al., 2025; Mustafa, 2015). We use the classical one-dimensional elimination of the squared-velocity channel; the normalizing integral itself is not new. Our analysis restricts transformations to the scalar state while preserving recorded time, connects law compatibility to passive coefficient-matching conditions, and restricts canonical affine compatibility back to the original parameter family. The resulting fixed-interval parameter sets are not automatically equivalence classes. Canonical normalization is determined by the law and does not itself supply physical units. It can be applied after fitting for law comparison, or incorporated directly into an integral fitting objective.

C Sampled Objectives and Error Accounting

Figure 2 connects the shared-encoder objective in Eq. (3) to parameter recovery. Training uses frames and timestamps. The fitted law determines any required canonical normalizer; family-compatible parameter sets then determine which targets are already identified and which need external calibration. Synthetic reference states are used separately to evaluate these predictions. Appendix D states what these comparisons establish. Below we specify the sampled objective and variance floor, then relate sampled to continuous residuals. Domain conventions are in Appendix A.2.

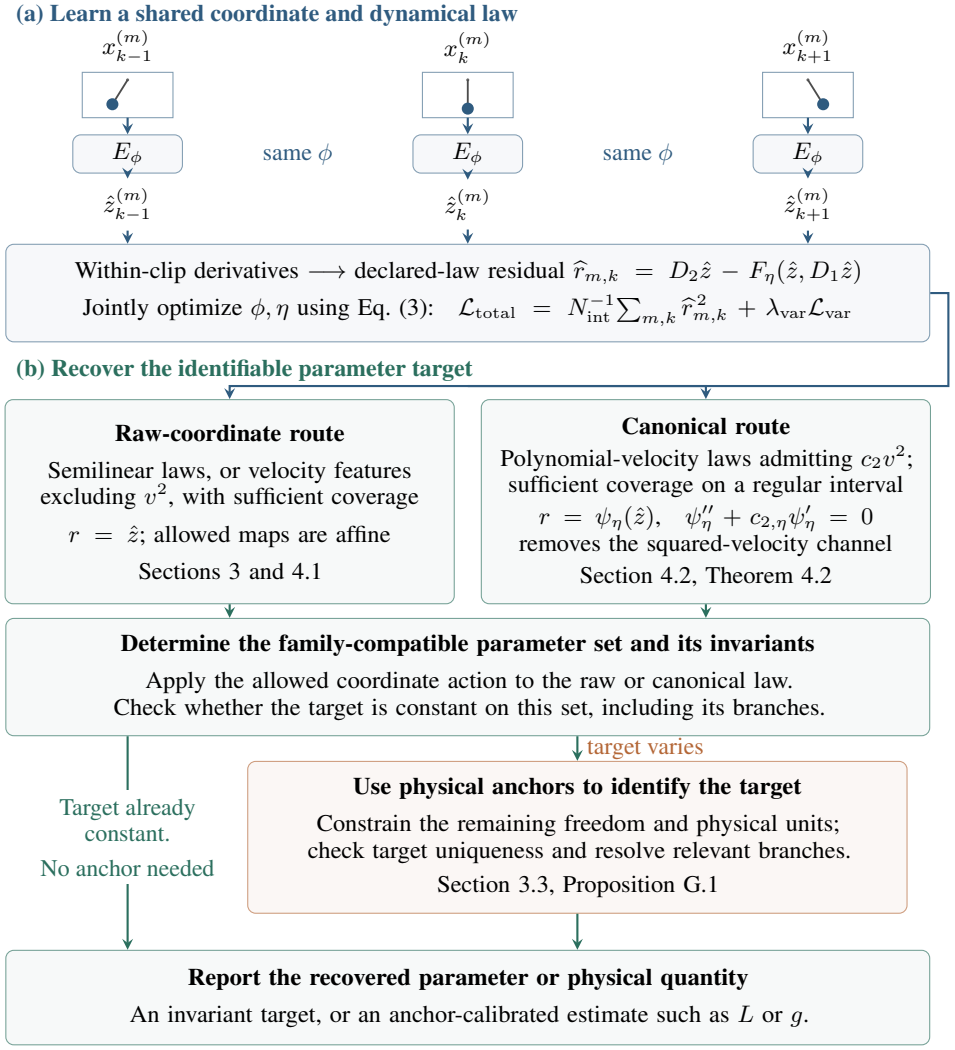


Figure 2: **From joint video fitting to parameter recovery.** The declared family and coverage conditions determine the raw or canonical route. In the canonical route, $\psi_\eta(u_\star) = 0$ and $\psi_\eta'(u_\star) = 1$ fix a normalizer, and $r'' = \psi_\eta' F_\eta + \psi_\eta''(\hat{z}')^2$ is the transformed law. Normalization is computed from the fitted law; external anchors are added only when the desired target is not yet identified. Parameter sets are understood with the containment and realizability qualifications in Appendix D. Synthetic coordinate and parameter errors evaluate these predictions using reference states. The diagram uses z for physical state and $\hat{z}$ for learned state; experimental plots use q and z, respectively.

C.1 Numerical Derivatives and Anti-Collapse Regularization

Clip $m = 1, \ldots, M$ supplies frames $x_k^{(m)} = x^{(m)}(t_{m,k})$, $k = 0, \ldots, T_m$, on a uniform time grid $t_{m,k} = t_{m,0} + k\Delta t_m$ with known $\Delta t_m > 0$. Thus T_m is the final frame index and each clip has

$T_m + 1$ frames. For the sampled objective, assume $T_m \geq 2$ so every included clip has an interior three-frame stencil. The sampled encoder outputs are $\hat{z}_k^{(m)} = E_\phi(x_k^{(m)}) = \hat{z}^{(m)}(t_{m,k})$.

Following the encoder-only interface of Wang et al. (2026), define for any sequence $q^{(m)} = (q_k^{(m)})_{k=0}^{T_m}$

$$\begin{aligned}(D_{1,m}q^{(m)})_k &= \frac{q_{k+1}^{(m)} - q_{k-1}^{(m)}}{2\Delta t_m}, \\ (D_{2,m}q^{(m)})_k &= \frac{q_{k+1}^{(m)} - 2q_k^{(m)} + q_{k-1}^{(m)}}{\Delta t_m^2}, \qquad k = 1, \ldots, T_m - 1.\end{aligned} \tag{17}$$

No endpoint derivative or cross-clip stencil is used.

At interior indices, the sampled fitted-law residual is

$$\widehat{r}_{m,k}(\phi, \eta) = (D_{2,m}\hat{z}^{(m)})_k - F_\eta\Big(\hat{z}_k^{(m)}, (D_{1,m}\hat{z}^{(m)})_k\Big). \tag{18}$$

The objective in Eq. (3) can be written as

$$\mathcal{L}_{\text{total}}(\phi, \eta) = \underbrace{\frac{1}{N_{\text{int}}} \sum_{m=1}^{M} \sum_{k=1}^{T_m - 1} \widehat{r}_{m,k}(\phi, \eta)^2}_{\mathcal{L}_{\text{ODE}}} + \lambda_{\text{var}}\mathcal{L}_{\text{var}}, \qquad N_{\text{int}} = \sum_{m=1}^{M}(T_m - 1), \tag{19}$$

where $\lambda_{\text{var}} > 0$ and $\mathcal{L}_{\text{var}}$ is a variance-floor penalty that discourages constant encodings.

For a polynomial-velocity family, the residual in (18) is

$$\widehat{r}_{m,k}(\phi, \eta) = (D_{2,m}\hat{z}^{(m)})_k - \sum_{r=0}^{p} c_{r,\eta}(\hat{z}_k^{(m)})\big((D_{1,m}\hat{z}^{(m)})_k\big)^r. \tag{20}$$

In the semilinear case it is $(D_{2,m}\hat{z}^{(m)})_k + a_\eta(\hat{z}_k^{(m)})(D_{1,m}\hat{z}^{(m)})_k + b_\eta(\hat{z}_k^{(m)})$.

To specify the penalty in (19), define

$$\widehat{\mu}_m = \frac{1}{T_m + 1}\sum_{k=0}^{T_m} \hat{z}_k^{(m)}, \qquad \widehat{\text{Var}}_m = \frac{1}{T_m + 1}\sum_{k=0}^{T_m}(\hat{z}_k^{(m)} - \widehat{\mu}_m)^2, \tag{21}$$

and set

$$\mathcal{L}_{\text{var}} = \frac{1}{M}\sum_{m=1}^{M}\left[\max\left\{0, s_{\text{floor}} - \sqrt{\widehat{\text{Var}}_m + \epsilon_{\text{stab}}}\right\}\right]^2, \tag{22}$$

where $\lambda_{\text{var}} > 0$, $\epsilon_{\text{stab}} > 0$ is numerical, and $s_{\text{floor}} > \sqrt{\epsilon_{\text{stab}}}$. A constant encoding c has zero ODE residual whenever $F_\eta(c, 0) = 0$, equivalently $b_\eta(c) = 0$ in the semilinear family. The penalty discourages such encodings but does not prove non-collapse on a particular interval, latent-state consistency, or identifiability. Its scale is a fitting convention, not a physical anchor.

C.2 From Sampled Residuals to Continuous Compatibility

The explicit constants below concern the semilinear residual. A general smooth F_η admits an analogous transfer with suitable local Lipschitz bounds. Additional regularity is stated here, rather than assumed for all exact results.

Approximate continuous state consistency. For each clip, let $e^{(m)}(t) = \hat{z}^{(m)}(t) - f(z^{(m)}(t))$, and assume this error is C^2 on a neighborhood of every selected time. Let $\mathcal{T}_{\text{sel}}$ collect the selected interior clip–time pairs, and define

$$\varepsilon_{\text{sc}} = \sup_{(m,t)\in\mathcal{T}_{\text{sel}}} \max_{r=0,1,2}\left|\frac{d^r}{dt^r}e^{(m)}(t)\right|. \tag{23}$$

Suppose $|z'| \le V$, $|f'| \le \Lambda$, and both $f(z)$ and $f(z) + e^{(m)}$ lie in a compact interval $J \subset D_\eta$ at the selected points. Assume $\|a_\eta\|_{\infty,J} \le A$ and that a_η, b_η are respectively L_a, L_b-Lipschitz on J. Write $R_\eta[y] = y'' + a_\eta(y)y' + b_\eta(y)$. Then

$$\left| R_\eta[\hat{z}^{(m)}](t) - R_\eta[f \circ z^{(m)}](t) \right| \le \varepsilon_{\mathrm{sc}}\big(1 + A + L_a \Lambda V + L_b\big). \tag{24}$$

The bound follows by expanding the difference and using $|y'| \le \Lambda V$. Thus a residual tolerance for the smooth encoder path transfers to the state-consistent reference path by adding the right-hand side of (24).

Remark C.1 (Centered-stencil error accounting). Fix one clip and an interior sample time t_k, suppressing m. Let $\bar{z}(t) = \hat{z}(t)$ be the smooth reference encoder path and $e(t) = \bar{z}(t) - f(z(t))$. To allow an additional, disjoint sample-level perturbation, write the sequence supplied to the stencil as

$$\hat{z}_j = \bar{z}(t_j) + \xi_j, \qquad |\xi_j| \le \sigma_0.$$

The sampling model of Section 2 is the case $\xi_j = 0$. If acquisition or encoder errors are already included in e, they must not also be counted as a separate ξ_j perturbation. Suppose $\bar{z} \in C^4([t_{k-1}, t_{k+1}])$, and let $M_r = \|\bar{z}^{(r)}\|_\infty$ on this interval. With D_1, D_2 denoting the corresponding within-clip operators,

$$\begin{aligned} \delta_0 &:= |\hat{z}_k - \bar{z}(t_k)| \le \sigma_0, \\ \delta_1 &:= |(D_1\hat{z})_k - \bar{z}'(t_k)| \le \frac{M_3}{6}\Delta t^2 + \frac{\sigma_0}{\Delta t}, \\ \delta_2 &:= |(D_2\hat{z})_k - \bar{z}''(t_k)| \le \frac{M_4}{12}\Delta t^2 + \frac{4\sigma_0}{\Delta t^2}. \end{aligned} \tag{25}$$

Hence truncation bias is second order, while sample perturbations are amplified at orders Δt^{-1} and Δt^{-2}.

Define, at t_k,

$$\begin{aligned} R_{\mathrm{cont}} &= \bar{z}'' + a_\eta(\bar{z})\bar{z}' + b_\eta(\bar{z}), \\ R_{\mathrm{disc},k} &= (D_2\hat{z})_k + a_\eta(\hat{z}_k)(D_1\hat{z})_k + b_\eta(\hat{z}_k). \end{aligned}$$

If $\hat{z}_k, \bar{z}(t_k) \in J \subset D_\eta$, $\|a_\eta\|_{\infty,J} \le A$, a_η, b_η are respectively L_a, L_b-Lipschitz on J, and $|\bar{z}'(t_k)| \le \bar{V}$, then

$$|R_{\mathrm{disc},k} - R_{\mathrm{cont}}| \le \delta_2 + A\delta_1 + (L_a\bar{V} + L_b)\delta_0. \tag{26}$$

This follows by adding and subtracting $a_\eta(\hat{z}_k)\bar{z}'(t_k)$. Thus $|R_{\mathrm{disc},k}| \le \varepsilon_{\mathrm{emp}}$ yields a continuous residual bound after adding the right-hand side of (26).

Combining the error layers. Under their respective hypotheses, (26) first transfers a sampled residual bound to the smooth encoder path, and (24) then transfers it to $f \circ z$. The near-level stability results separately account for state mismatch h and amplify the resulting errors through the physical-velocity design conditioning. These are deterministic transfers, not a conversion of encoder-coordinate singular values into physical-unit conditioning. The stochastic analysis of the LTI-specific centered-difference estimator in Wang et al. (2026, Theorem 4.8) is separate from these nonlinear orbit and deterministic error calculations.

D INTERPRETING THE EXPERIMENTAL EVIDENCE

The experiments evaluate three linked questions: whether a shared coordinate map transfers to held-out clips, whether the coefficients follow the family-specific parameter relations, and whether external anchors recover physical quantities. The metrics in Appendix L measure accuracy for these comparisons. Table 6 connects the comparisons to the assumptions of the theory, so that empirical agreement can be interpreted at the appropriate level.

Table 6: **Empirical checks and the conclusions they support.** The checks probe the stated premises; finite observations do not certify their continuous-time versions.

Question	Evidence to report	Scope of the conclusion
Declared family	Coefficient restrictions and domains; test-law error and matched-family comparisons.	A small fitting residual alone does not establish that the declared family contains the physical law.
Shared state map	One frozen map across clips; held-out coordinate error and, when available, derivative comparisons.	Coordinate agreement supports transfer of the map. A shared encoder alone does not establish latent-state consistency.
Velocity coverage	Physical coordinate, level tolerance, distinct speeds and signs, feature rank and conditioning.	The coefficient-matching result requires its stated coverage. Failure of a sufficient rank condition does not itself prove nonidentifiability.
Canonical normalization	Fitted c_2, normalizer basepoint, regular interval, and transformed-law comparison.	The comparison is restricted to the valid chart. Normalization removes the squared-velocity channel but does not determine physical units.
Parameter recovery	Allowed coordinate action and branches, parameter-orbit error, and the specified invariants.	Agreement of selected coefficients or invariants need not establish full law equivalence or identify every parameter.
Physical calibration	Anchor values, units, domain, remaining branches, and a uniqueness argument.	Calibrated estimates use additional physical information. A regular full-rank anchor map gives local uniqueness, with branches checked separately.

Coordinate comparison, normalization, and calibration. An invariant comparison tests a specified parameter functional; an aligned comparison uses the coordinate and coefficient transformations declared by the family. A calibrated comparison adds external physical information. In Figure 2, normalization is computed from the learned law before these comparisons: it selects a canonical coordinate but leaves its affine freedom. For the projected pendulum, the endpoints of $|y| < 1$ are chart singularities, and comparisons remain within the regular interval. A geometry check for a fixed observable supports that observable's model; it does not also validate a learned encoder. These distinctions apply to both synthetic and real videos.

Structural conclusions and finite-data accuracy. The general containment results provide compatible parameter outer sets; constancy on such a set is sufficient for identification. Variation on an outer set establishes realizable ambiguity only with an additional construction. For the semilinear case, Proposition 3.3 and Corollary 3.4 supply this connection under representability and affine-closure assumptions. The same converse is not asserted for canonical sets solely from raw–canonical law equivalence. Empirical agreement does not establish optimizer convergence to all compatible reports.

Error and uncertainty. The stability bounds require physical-velocity conditioning, which encoder-coordinate singular values cannot replace. Appendix C separates sampling and state-consistency errors. Proposition G.1 treats anchors as exact and guarantees local uniqueness at regular points; it does not quantify anchor measurement error. Repeated fits with shared observations or anchors describe protocol variability, not confidence intervals over independent physical measurements. The experimental appendices report parameter accuracy and initialization sensitivity separately.

E SHARPNESS OF THREE-SLOPE COVERAGE

Three-slope level coverage is the geometric hypothesis used by the semilinear affine reduction. This appendix shows the requirement is sharp: two slopes admit non-affine aliases.

Proposition E.1 (Two-slope sharpness for the continuous semilinear class)**.** *Let $U \subset R_z$ be a nonempty open interval, let $a_\theta, b_\theta \in C(U)$, and suppose the physical covered points obey*

$$z'' + a_\theta(z)z' + b_\theta(z) = 0.$$

Assume that at every $u \in U$ the set of covered velocities is exactly $\{v_1(u), v_2(u)\}$, where $v_1, v_2 \in C(U)$ and $v_1(u) \neq v_2(u)$. For any C^2 diffeomorphism $f : U \to f(U)$, define continuous target coefficients by

$$a_\eta(f(u)) := a_\theta(u) - \frac{f''(u)}{f'(u)}\big(v_1(u) + v_2(u)\big), \tag{27}$$

$$b_\eta(f(u)) := f'(u)b_\theta(u) + f''(u)v_1(u)v_2(u). \tag{28}$$

Then the transformed coordinate $\hat{z} = f(z)$ satisfies $\hat{z}'' + a_\eta(\hat{z})\hat{z}' + b_\eta(\hat{z}) = 0$ at every covered point. Consequently, two same-level velocities cannot force affine collapse uniformly over unrestricted continuous semilinear coefficient functions.

Proof. Because f is a diffeomorphism, (27)– (28) define single-valued continuous functions on the open interval $f(U)$. At a covered point with state u and velocity v, the target-law residual is

$$\begin{aligned}
&\hat{z}'' + a_\eta(\hat{z})\hat{z}' + b_\eta(\hat{z}) \\
&\quad = f''(u)v^2 + f'(u)\big(a_\eta(f(u)) - a_\theta(u)\big)v + b_\eta(f(u)) - f'(u)b_\theta(u) \\
&\quad = f''(u)\big(v - v_1(u)\big)\big(v - v_2(u)\big).
\end{aligned}$$

It vanishes on both realized branches. Since f may be chosen nonlinear, the data do not imply affine collapse in this unrestricted class. □

For a conservative branch pair $v_2 = -v_1$, the construction simplifies to

$$a_\eta \circ f = a_\theta, \qquad b_\eta \circ f = f'b_\theta - f''v_1^2.$$

For example, one orbit of $z'' + z = 0$ with amplitude A has $v_1(u)^2 = A^2 - u^2$ on any $U \Subset (-A, A)$, so every such nonlinear f is compatible with the continuous semilinear law $a_\eta \circ f = 0$ and $b_\eta \circ f = f'u - f''(A^2 - u^2)$. This law-space obstruction proves that two slopes do not guarantee affine collapse. No claim is made that a particular encoder architecture realizes every such nonlinear witness.

F Canonical coordinates and the projected-pendulum example

Section 4.2 removes the squared-velocity channel using a coordinate computed from each law. The explicit integral (41), its positive derivative, and uniqueness up to an invertible affine map are proved together in Appendix I. The transformed coefficients (43) need not be polynomial in the canonical state. This normalization permits comparison of laws without providing a physical calibration of the state.

Regular-chart intuition. For the angular pendulum, $y = \sin q$ on $|q| < \pi/2$ creates a squared-velocity term through the chain rule. The inverse chart $q = \arcsin y$ removes it, while its derivative diverges at $y = \pm 1$. Those endpoints are outside the regular chart. The full laws (49)–(50) and their derivation appear in Corollary I.4 and its proof. This illustrates a change of law representation; the narrow angular-pendulum family does not admit the projected law as another report.

Parameter-recovery routes. Table 7 summarizes the assumptions and conclusions of Sections 3–4. The semilinear case is contained in the curvature-excluding feature route; the canonical route has the separate polynomial-velocity scope stated in Section 4.2.

Table 7: Parameter-recovery routes within a fixed declared family, conditional on a shared non-collapsed state map, exact compatibility, and the stated domain and coverage assumptions. The semilinear row is a special case of the finite-feature route; the rows are not an exhaustive partition of scalar ODEs.

Family and required design	Coordinate relation	Parameter conclusion
Semilinear; three distinct physical velocities at each covered state (Theorem 3.1)	Raw coordinates are affine-related	Affine compatible outer set; invariant combinations, singleton identification, and sufficient local physical anchors
Finite dilation-stable library with constants, excluding v^2; full augmented rank (Theorem 4.1)	Raw coordinates are affine-related	Affine compatibility analysis, restricted to the declared family
Polynomial in velocity, admitting $c_2(z)v^2$; $d_v + 1$ distinct physical velocities per state (Theorem 4.2)	Raw relation may be nonlinear; canonical relation is affine	Canonical compatible outer set in the original parameter space; invariants and local physical-anchor analysis
Required coverage or shared-map premise fails or remains unverified	The affected guarantee is unavailable	Acquire additional evidence, supply a separate family-specific argument, or withhold the affected claim

G PHYSICAL ANCHORS AND LOCAL CALIBRATION

External anchors constrain coordinate witnesses as well as reported parameters. A single local criterion applies to the raw and canonical compatibility sets.

Proposition G.1 (Local anchor-rank criterion)**.** *Let $\mathcal{S}$ be a $d_{\mathcal{S}}$-dimensional C^1 manifold near s_0, and let $\mathcal{A} : \mathcal{S} \to \mathbb{R}^k$ be a C^1 vector of external scalar-anchor predictions. If*

$$\operatorname{rank} D\mathcal{A}_{s_0} = d_{\mathcal{S}},$$

then s_0 is locally the unique point satisfying $\mathcal{A}(s) = \mathcal{A}(s_0)$. If $k < d_{\mathcal{S}}$ and $\mathcal{A}$ has locally constant maximal rank k, its local level set has dimension $d_{\mathcal{S}} - k > 0$. Thus fewer than $d_{\mathcal{S}}$ scalar anchors are insufficient at such regular points.

The proof is in Appendix H.4. Full differential rank is sufficient for local uniqueness, not necessary: singular but locally injective anchor maps are possible. Disconnected branches and global uniqueness require separate checks.

Raw affine witnesses. The family-compatible outer set projects out the affine witness. To retain the coordinate information on which physical anchors act, define

$$\mathcal{S}_U(\theta) := \{(\lambda, \tau, \eta) \in \mathbb{R}^* \times \mathbb{R} \times \Theta : (\lambda, \tau) \text{ is an affine witness for } \eta \text{ on } U\}. \tag{29}$$

Its parameter projection is $\mathcal{O}_U(\theta)$, but the lifted set also retains coordinate symmetries that leave the mechanism parameters unchanged. At a regular point where $\mathcal{S}_U(\theta)$ is locally a C^1 manifold, Proposition G.1 applies to any C^1 anchor predictions restricted to that manifold. A paired physical state $u_\star \in U$ predicts the encoded value $\lambda u_\star + \tau$. Anchors outside U require a separately justified extension of the witness.

Canonical witnesses. For the polynomial-velocity family, the affine witness relates law-dependent normalized coordinates. Define

$$\mathcal{S}_U^{\mathrm{can}}(\theta) := \left\{ (\lambda, \tau, \eta) \in \mathbb{R}^* \times \mathbb{R} \times \Theta : \begin{array}{l} A(q) = \lambda q + \tau, \quad A(\psi_\theta(U)) \subset J_\eta, \\ A \text{ satisfies (44)} \end{array} \right\}. \tag{30}$$

Its parameter projection is $\mathcal{O}_U^{\mathrm{can}}(\theta)$. The associated raw-coordinate witness is

$$f_{\lambda,\tau,\eta;\theta}(u) = \psi_\eta^{-1}\big(\lambda \psi_\theta(u) + \tau\big). \tag{31}$$

Apply Proposition G.1 wherever this lifted set is locally a C^1 manifold and the anchor predictions are C^1 on it. Raw position and amplitude anchors must be evaluated through (31); because ψ_η

depends on the reported law, an unsigned raw amplitude anchor does not automatically fix $|\lambda|$. In the semilinear slice, where $c_r = 0$ for every $r \geq 2$, choosing identity normalizers identifies the raw and canonical lifted sets.

Both applications concern the full lifted compatibility set. Full rank therefore gives local uniqueness in every compatible report subset, whereas the lower bound on the number of anchors need not apply to a smaller subset realized by a particular encoder architecture. Parameter recovery can require less information than fixing the entire coordinate witness. For interval-local laws, the fixed-U parameter projection is an outer set, not automatically an orbit or an equivalence class.

H PROOFS FOR THE SCALAR GAUGE-ORBIT THEORY

We use the admissible-report convention of Section 2 and Definition 2.1: the common map holds along the continuous paths on U, is non-collapsed there, and satisfies $f(U) \subset D_\eta$. The fitted semilinear law holds at the covering points. Affine witnesses and the outer set $\mathcal{O}_U(\theta)$ are as defined in Section 3, with domains transported as in Appendix A.2.

H.1 PROOF OF THEOREM 3.1

Proof. Fix $u \in U$. At a covered point with physical state $z = u$ and physical velocity $z' = v$, latent-state consistency holds on a time neighborhood of the point because U is open. Differentiating it twice gives

$$\hat{z}' = f'(u)v, \qquad \hat{z}'' = f''(u)v^2 + f'(u)z''.$$

Substitution of the physical and encoder-side semilinear laws therefore gives the quadratic residual $R_u(v)$ in (7). By three-slope level coverage, there are three pairwise distinct realized physical velocities v_1, v_2, v_3 at level u, and compatibility gives $R_u(v_i) = 0$ for all three. A polynomial of degree at most two with three distinct roots is identically zero. Hence its three coefficients vanish:

$$f''(u) = 0, \qquad f'(u)\big(a_\eta(f(u)) - a_\theta(u)\big) = 0, \qquad b_\eta(f(u)) - f'(u)b_\theta(u) = 0.$$

Since u was arbitrary, $f'' = 0$ on the interval U, so $f(u) = \lambda u + \tau$ there. The non-collapse premise gives $\lambda \neq 0$; therefore division of the middle identity by $f' = \lambda$ and substitution into the last identity yield

$$a_\eta(\lambda u + \tau) = a_\theta(u), \qquad b_\eta(\lambda u + \tau) = \lambda b_\theta(u),$$

which are exactly the orbit equations (6). □

H.2 PROOF OF THEOREM 3.2

Proof. Let (E_ϕ, f, η) be any admissible state-consistent encoder report on U. Theorem 3.1 applies, so $f(u) = \lambda u + \tau$ on U with $\lambda \neq 0$, and the orbit equations (6) hold. Thus $\eta \in \mathcal{O}_U(\theta)$. If $\mathcal{O}_U(\theta) = \{\theta\}$, membership immediately gives $\eta = \theta$. □

H.3 PROOF OF AFFINE-COMPATIBLE ENCODER REALIZABILITY

Proof of Proposition 3.3. The inclusion $\subseteq$ in (8) follows from Theorem 3.2. For the reverse inclusion, fix $\eta \in \mathcal{O}_U(\theta)$ and an affine witness (λ, τ). By definition, $\lambda \neq 0$, $\lambda U + \tau \subset D_\eta$, and, for every $u \in U$,

$$a_\eta(\lambda u + \tau) = a_\theta(u), \qquad b_\eta(\lambda u + \tau) = \lambda b_\theta(u). \tag{32}$$

Let $E_\eta = \lambda E_0 + \tau$, which belongs to $\mathscr{E}$ by assumption. For every clip, on the times with $z^{(m)}(t) \in U$,

$$\hat{z}^{(m)}(t) := E_\eta(x^{(m)}(t)) = \lambda z^{(m)}(t) + \tau.$$

The affine map extends smoothly to the state domain, is common to all clips, and is non-collapsed. Its image on U lies in D_η. Since U is open and each physical path is continuous, the displayed equality holds on a time neighborhood of every interior point under consideration. The encoder-induced path is therefore C^2 there, with $\hat{z}^{(m)\prime} = \lambda z^{(m)\prime}$ and $\hat{z}^{(m)\prime\prime} = \lambda z^{(m)\prime\prime}$. Using (32),

$$\begin{aligned}\hat{z}^{(m)\prime\prime} &+ a_\eta(\hat{z}^{(m)})\hat{z}^{(m)\prime} + b_\eta(\hat{z}^{(m)}) \\ &= \lambda\big[z^{(m)\prime\prime} + a_\theta(z^{(m)})z^{(m)\prime} + b_\theta(z^{(m)})\big] = 0.\end{aligned}$$

Thus the reported law holds at every such interior point, including the covering points. All admissibility conditions are satisfied, proving the reverse inclusion. The construction uses the same videos and recorded time, without a time rescaling. Three-slope coverage is needed for the forward containment, not for this reverse construction. □

Proof of Corollary 3.4. The identity witness gives $\theta \in \mathcal{O}_U(\theta)$. If $\mathcal{J}$ is constant on this set, (8) implies $\mathcal{J}(\eta) = \mathcal{J}(\theta)$ for every admissible report in $\mathscr{E}$. If it is not constant, some $\eta \in \mathcal{O}_U(\theta)$ satisfies $\mathcal{J}(\eta) \neq \mathcal{J}(\theta)$; Proposition 3.3 realizes a report with that value. Taking $\mathcal{J}$ to be the identity proves the full-parameter claim. □

Representation and domain qualifications. Exact physical-coordinate representation is assumed, not inferred from shared weights, universal approximation, or finite training error. An unconstrained final affine layer can supply output-affine closure, but does not establish the existence of E_0. Fixed output bounds, fixed-scale normalizations or imposed physical anchors can restrict the report class; the proposition must then be applied only after checking its hypotheses. No compatibility is asserted outside U. Even when realizability holds, transported-domain restrictions may prevent an interval-local compatible set from being an equivalence class. The result concerns the semilinear report convention; it does not automatically extend to nonlinear raw-to-canonical transformations.

H.4 Proofs for invariants and anchors

Proof of Proposition G.1. Restrict the C^1 map $\mathcal{A}$ to the $d_{\mathcal{S}}$-dimensional manifold $\mathcal{S}$ near s_0. If its differential has rank $d_{\mathcal{S}}$ at s_0, that maximal rank persists locally. The constant-rank theorem then makes the local level set $\mathcal{A}(s) = \mathcal{A}(s_0)$ zero-dimensional, so s_0 is locally the unique representative satisfying the anchors. More generally, if the restricted rank is constant and equal to r near s_0, the local level set has dimension $d_{\mathcal{S}} - r$. In particular, $k < d_{\mathcal{S}}$ independent scalar anchors leave dimension $d_{\mathcal{S}} - k > 0$. This regular-level argument does not rule out isolated fibers at singular points, and full differential rank is a sufficient local criterion, not a necessary condition for an arbitrary nonlinear anchor to isolate a point. □

H.5 Translation lock and the weight rule

Lemma H.1 (Top-gap test for translation lock)**.** *Let U be a nonempty open interval and let the declared polynomial family have fixed damping and restoring supports I_a and I_b. Fix $\theta \in \Theta$. Every affine witness for every $\eta \in \mathcal{O}_U(\theta)$ has $\tau = 0$ if either of the following holds:*

1. *$n := \max I_b \geq 1$, $B_n(\theta) \neq 0$, and $n - 1 \notin I_b$; or*
2. *$m := \max I_a \geq 1$, $A_m(\theta) \neq 0$, and $m - 1 \notin I_a$.*

Thus either structural top gap is sufficient to make the surviving affine witness a pure scaling $u \mapsto \lambda u$.

Proof of Lemma H.1. Both laws expand over the structural support. Write

$$b_\eta(w) = \sum_{k \in I_b} B_k(\eta) w^k,$$

and impose $b_\eta(\lambda u + \tau) = \lambda b_\theta(u)$ on U; both sides are polynomials, so the identity holds for all $u \in \mathbb{R}$. Since $n = \max I_b$, no term of $b_\eta(\lambda u + \tau)$ has degree exceeding n, and only the index $k = n$ reaches degree n: matching gives $B_n(\eta)\lambda^n = \lambda B_n(\theta)$, hence $B_n(\eta) = \lambda^{1-n} B_n(\theta) \neq 0$, independently of τ. At degree $n-1$, the index $k = n$ contributes $\binom{n}{1} B_n(\eta)\lambda^{n-1}\tau = n\, B_n(\eta)\lambda^{n-1}\tau$ through the binomial expansion; the index $k = n - 1$ is absent because $n - 1 \notin I_b$; and every index $k < n - 1$ contributes only to degrees at most $k < n - 1$. On the right the coefficient is $\lambda B_{n-1}(\theta) = 0$. Hence $n\, B_n(\eta)\lambda^{n-1}\tau = 0$, and since $B_n(\eta) \neq 0$, $\lambda \neq 0$, and $n \geq 1$, we conclude $\tau = 0$. With $\tau = 0$ the witness reduces to $u \mapsto \lambda u$.

The damping-side top-gap lock is identical. If $a_\eta(\lambda u + \tau) = a_\theta(u)$ and the structural damping support I_a has top degree $m \geq 1$ with $A_m(\theta) \neq 0$ and $m - 1 \notin I_a$, then matching the top coefficient gives $A_m(\eta)\lambda^m = A_m(\theta) \neq 0$; matching degree $m - 1$ gives $m A_m(\eta)\lambda^{m-1}\tau = 0$. Thus $\tau = 0$ by the same argument. □

Proof of Proposition 3.5. By translation lock every witness has $\tau = 0$, so for any $\eta \in \mathcal{O}_U(\theta)$ with witness scale λ the orbit equations (6) read $a_\eta(\lambda u) = a_\theta(u)$ and $b_\eta(\lambda u) = \lambda b_\theta(u)$. Since U is open, these polynomial identities can be matched coefficientwise. Substituting the monomial expansions and matching the coefficient of u^i (respectively u^j),

$$A_i(\eta)\lambda^i = A_i(\theta), \qquad B_j(\eta)\lambda^j = \lambda B_j(\theta),$$

which is (9) with $w(A_i) = i$ and $w(B_j) = j - 1$. A rational monomial $\prod_k C_k^{p_k}$ transforms by $\lambda^{-\sum_k p_k w(C_k)}$, so it is invariant whenever its total weight vanishes and its denominator is nonzero. Finally, at a paired state anchor $u_\star \in U \setminus \{0\}$ with $\hat{z}_\star = f(u_\star)$, translation lock gives $\hat{z}_\star = \lambda u_\star$. A signed anchor therefore fixes $\lambda = \hat{z}_\star / u_\star$ and hence every coefficient through (9); an unsigned anchor fixes only $|\lambda|$ and may leave the reflection action on odd weights. □

H.6 Proofs for the stability results

Quantitative slope coverage. For $\delta_v > 0$ and $V < \infty$, the collection has (δ_v, V)-*slope coverage on* U if for every $u \in U$ there are three realized velocities v_1, v_2, v_3 at level u with $|v_i - v_j| \geq \delta_v$ for $i \neq j$ and $|v_i| \leq V$.

Lemma H.2 (Quantitative exact-level affine collapse)**.** *Let $K = [u_-, u_+] \Subset U$ be a nondegenerate compact interval, assume latent-state consistency with $f \in C^2$ near K, (δ_v, V)-slope coverage on K, $0 < c_0 \leq |f'|$, exact physical trajectories, and absolute encoder-side residual at most ε at every covering point, i.e. $|R_u(v_i)| \leq \varepsilon$ for each covering velocity v_i at level u. Then, for every $u \in K$,*

$$|f''(u)| \leq \frac{3\varepsilon}{\delta_v^2}, \quad |a_\eta(f(u)) - a_\theta(u)| \leq \frac{6V\varepsilon}{c_0\delta_v^2}, \quad |b_\eta(f(u)) - f'(u)b_\theta(u)| \leq \frac{3V^2\varepsilon}{\delta_v^2}.$$

For every $u_0 \in K$, set $\ell_{u_0}(u) = f(u_0) + f'(u_0)(u - u_0)$. Then

$$\sup_{u \in K} |f(u) - \ell_{u_0}(u)| \leq \frac{3\varepsilon}{2\delta_v^2} \operatorname{diam}(K)^2.$$

Proof of Lemma H.2. For each $u \in K$, the exact physical law and the chain rule give the quadratic residual $R_u(v)$ given by the right-hand side of (7). At the three covering velocities, $|R_u(v_i)| \leq \varepsilon$. Lagrange interpolation writes $R_u(v) = \sum_{i=1}^3 R_u(v_i) L_i(v)$, where

$$L_i(v) = \frac{(v - v_j)(v - v_k)}{(v_i - v_j)(v_i - v_k)}, \qquad \{i, j, k\} = \{1, 2, 3\}.$$

The denominator has absolute value at least δ_v^2, so the quadratic, linear, and constant coefficients of L_i have absolute values at most $1/\delta_v^2$, $2V/\delta_v^2$, and V^2/δ_v^2, respectively. Summing the three residual contributions bounds $|f''|$, $|f'(a_\eta \circ f - a_\theta)|$, and $|b_\eta \circ f - f' b_\theta|$ by $3\varepsilon/\delta_v^2$, $6V\varepsilon/\delta_v^2$, and $3V^2\varepsilon/\delta_v^2$. Divide the middle bound by $|f'| \geq c_0$. Finally, Taylor's theorem gives $|f(u) - \ell_{u_0}(u)| \leq \frac{1}{2} \sup_K |f''| \, |u - u_0|^2$, proving the stated uniform bound for every $u_0 \in K$. □

Theorem H.3 (Near-level, design-conditioned affine stability)**.** *Let $K = [u_-, u_+] \Subset U$ be a nondegenerate compact interval, where $K \Subset U$ means that K is compactly contained in U. Assume there is a map $f \in C^2$ on an open neighborhood of K such that $\hat{z}^{(m)} = f \circ z^{(m)}$ on a time neighborhood of every selected realized point used below, and assume that the physical trajectories satisfy (4) exactly at those points. For $v \in \mathbb{R}$, write $\boldsymbol{\phi}_{\mathrm{sl}}(v) = (v^2, v, 1)^\top$. The chain-rule residual at level u and velocity v has coefficient form $\boldsymbol{\phi}_{\mathrm{sl}}(v)^\top c(u)$, where*

$$c(u) := \begin{pmatrix} f''(u) \\ f'(u)\big(a_\eta(f(u)) - a_\theta(u)\big) \\ b_\eta(f(u)) - f'(u)b_\theta(u) \end{pmatrix}. \tag{33}$$

Assume $0 < c_0 \leq |f'|$ on K, and suppose c is L_c-Lipschitz on K. This is a direct regularity assumption on the composite residual coefficients; it follows, for example, from $f \in C^{2,1}$ near K and local Lipschitz regularity of $a_\theta, b_\theta, a_\eta, b_\eta$ on the relevant ranges. Fix $u_0 \in K$ and define $\ell_{u_0}(u) := f(u_0) + f'(u_0)(u - u_0)$. For every target level $u \in K$, suppose there are three realized points with physical levels $u_i \in K$ and velocities v_i such that

$$|u_i - u| \leq h, \qquad |v_i| \leq V, \qquad \Phi_u := \begin{pmatrix} \boldsymbol{\phi}_{\mathrm{sl}}(v_1)^\top \\ \boldsymbol{\phi}_{\mathrm{sl}}(v_2)^\top \\ \boldsymbol{\phi}_{\mathrm{sl}}(v_3)^\top \end{pmatrix}, \qquad \sigma_{\min}(\Phi_u) \geq \sigma_\star > 0.$$

If the absolute encoder-side residual at each of these three realized points is at most ε, then

$$\|c(u)\|_2 \le B, \qquad B := \frac{\sqrt{3}}{\sigma_\star}\left(\varepsilon + L_c h\sqrt{V^4+V^2+1}\right). \tag{34}$$

Consequently the same right-hand side bounds $|f''(u)|$ and the restoring compatibility error, and after division by c_0 it bounds the damping compatibility error. Moreover,

$$\sup_{u\in K} |f(u) - \ell_{u_0}(u)| \le \tfrac{1}{2} B \operatorname{diam}(K)^2. \tag{35}$$

Gauge dependence of design conditioning. The matrix Φ_u above uses physical velocities. Under an exact affine map, $\hat{v}_i = \lambda v_i$ and the analogous encoder-velocity matrix satisfies

$$\widehat{\Phi}_u = \Phi_u \operatorname{diag}(\lambda^2, \lambda, 1). \tag{36}$$

Thus rank is gauge invariant but the raw smallest singular value is not. Column normalization yields a scale-free rank/shape diagnostic, whereas the raw bound (34) remains a physical-unit statement and requires a calibrated or bounded scale.

Proof of Theorem H.3. At a realized point with physical level u_i and velocity v_i, the exact physical equation and the chain rule give the encoder-side residual

$$R_{u_i}(v_i) = \boldsymbol{\phi}_{\mathrm{sl}}(v_i)^\top c(u_i).$$

For a fixed target u, add and subtract $c(u)$ to obtain

$$\begin{aligned} |\boldsymbol{\phi}_{\mathrm{sl}}(v_i)^\top c(u)| &\le |R_{u_i}(v_i)| + \|\boldsymbol{\phi}_{\mathrm{sl}}(v_i)\|_2 \, \|c(u_i) - c(u)\|_2 \\ &\le \varepsilon + \sqrt{V^4+V^2+1}\, L_c h. \end{aligned}$$

Stacking the three inequalities and using the Euclidean norm gives

$$\|\Phi_u c(u)\|_2 \le \sqrt{3}\left(\varepsilon + L_c h\sqrt{V^4+V^2+1}\right).$$

Because $\sigma_{\min}(\Phi_u) \ge \sigma_\star$, inversion on the three-dimensional coefficient space yields (34). The first and third components of $c(u)$ are bounded directly; dividing the second component by $|f'(u)| \ge c_0$ bounds the damping compatibility error. Taking the supremum of the first-component bound over K and applying Taylor's theorem around any $u_0 \in K$ proves the final near-affine statement. □

Column-normalized diagnostic. If each nonzero column is normalized to unit Euclidean norm, then

$$\overline{\widehat{\Phi}}_u = \overline{\Phi}_u \operatorname{diag}(1, \operatorname{sign}\lambda, 1),$$

so the column-normalized singular values are invariant. This normalized matrix is a scale-free rank/shape diagnostic; it cannot replace the raw physical-unit conditioning in Theorem H.3, because the residual coefficients and their error bounds also change units under the gauge.

Approximate state consistency. When the encoder path only approximates $f \circ z$, use $f \circ z$ as the state-consistent reference path. Under the regularity, domain, and error hypotheses of Appendix C, (24) gives the additional residual budget needed for this replacement. The bounds above then apply to the reference path with that inflated budget, provided their remaining hypotheses hold.

I Proofs for feature separation and the canonical connection gauge

This appendix supplies the intermediate coefficient-matching and law-equivalence statements and proves the results used in Section 4 and Appendix F. All coefficients use the right-hand-side convention: in particular, a polynomial-velocity law is written as $z'' = \sum_r c_r(z)(z')^r$. All identities are restricted to the open intervals on which the displayed compositions are defined.

I.1 FEATURE SEPARATION

Proof of Theorem 4.1. Fix $u \in U$. At any realized point with $z = u$ and $z' = v$, latent-state consistency holds on a time neighborhood of that point because U is open and the trajectory is continuous. It may therefore be differentiated twice at the point:

$$\hat{z}' = f'(u)v, \qquad \hat{z}'' = f''(u)v^2 + f'(u)z''.$$

Substituting the physical and encoder-induced feature laws gives

$$R_u(v) := f''(u)v^2 + f'(u)F_\theta(u, v) - F_\eta(f(u), f'(u)v) = 0 \tag{37}$$

at every realized velocity at level u.

The function $v \mapsto f'(u)F_\theta(u, v)$ lies in V because $F_\theta(u, \cdot) \in \mathsf{V}$. The function $v \mapsto F_\eta(f(u), f'(u)v)$ also lies in V: this is dilation stability with dilation factor $f'(u)$, including the case $f'(u) = 0$. Consequently $R_u \in \mathsf{V} + \operatorname{span}\{v^2\}$. Since $v^2 \notin \mathsf{V}$, this sum is direct. Use the basis $\varphi_1, \ldots, \varphi_K$ specified in the theorem. In the ordered basis $(\varphi_1, \ldots, \varphi_K, v^2)$ of the direct sum, the evaluation matrix is exactly M_u in (11), which has column rank $K + 1$. Since (37) vanishes at all its rows, the coefficient vector of R_u is zero, so $R_u \equiv 0$ as a function of v. In particular, the coefficient of the last basis element v^2 is $f''(u) = 0$.

The state u was arbitrary, hence $f'' = 0$ on the interval U and $f(u) = \lambda u + \tau$ there. Since $f' \not\equiv 0$, its constant derivative λ is nonzero. Returning to $R_u \equiv 0$ and substituting this affine form gives $F_\eta(\lambda u + \tau, \lambda v) = \lambda F_\theta(u, v)$ for all $u \in U$ and all $v \in \mathbb{R}$, which is (12). □

Odd drag: rank and affine scaling. Consider $\mathsf{V} = \operatorname{span}\{1, v, v|v|\}$ and four distinct velocities including both signs. If $P(v) = a + bv + cv|v| + dv^2$ vanishes at all four, its restrictions to the two half-lines are quadratics. If three sample velocities lie in the same closed half-line, that quadratic vanishes identically; the remaining, strictly opposite-sign velocity then forces the other quadratic to vanish as well. This also covers a sample at zero. Otherwise the samples are $p, q, -r, -s$ with $p, q, r, s > 0$, and

$$P(v) = k_+(v - p)(v - q) \quad (v \geq 0), \qquad P(v) = k_-(v + r)(v + s) \quad (v \leq 0).$$

Equality of their constant and linear coefficients gives $k_+pq = k_-rs$ and $-k_+(p + q) = k_-(r + s)$, which force $k_+ = k_- = 0$. Thus the four-column design has full rank, including when zero is sampled. Since $(\lambda v)|\lambda v| = \lambda|\lambda|v|v|$, (12) yields

$$c_{0,\eta}(\lambda u + \tau) = \lambda c_{0,\theta}(u), \qquad c_{1,\eta}(\lambda u + \tau) = c_{1,\theta}(u), \qquad c_{d,\eta}(\lambda u + \tau) = \frac{c_{d,\theta}(u)}{|\lambda|}.$$

One-sided velocities and the declared family. For $F(u, v) = c_0(u) + c_1(u)v + c_d(u)v|v|$ with continuous coefficients on U, let $f : U \to f(U)$ be any C^2 diffeomorphism. Define coefficients on $f(U)$ by

$$\widetilde{c}_0 \circ f = f'c_0, \qquad \widetilde{c}_1 \circ f = c_1, \qquad \widetilde{c}_d \circ f = \frac{f'' + f'c_d}{f'|f'|}.$$

For the corresponding $\widetilde{F}$, direct substitution gives

$$f''(u)v^2 + f'(u)F(u, v) - \widetilde{F}(f(u), f'(u)v) = f''(u)\big(v^2 - v|v|\big).$$

Hence the transformed law is exact for all physical $v \geq 0$, including orientation-reversing f, but generally fails for $v < 0$ wherever $f'' \neq 0$. For a physical $v \leq 0$ branch instead, replace the numerator by $f'c_d - f''$. The unrestricted class of continuous coefficient functions on the transported interval is closed under these constructions. Membership in a fixed parametric family, coefficient ties, and any sign restrictions must be checked separately; one-sided rank failure alone does not prove non-identifiability in every such family.

In particular, for the coverage experiment of Appendix M.2, the algebraic control $f(u) = u + 5u^3$ certifies this unrestricted one-sided construction, not an exact alternative within its state-polynomial 5/3/5 fitted family. Indeed its physical $c_0(u) = -0.8u$ requires $\widetilde{c}_0(u + 5u^3) = -0.8u - 12u^3$. If $\widetilde{c}_0$ were a polynomial, degree comparison would make it linear, while its linear term forces $\widetilde{c}_0(y) = -0.8y$, whose cubic coefficient after composition is -4, not -12. Thus this twin is outside that polynomial family.

I.2 Polynomial coefficient matching

For the fixed family (13), set $d_v = \max\{2, p\}$ and zero-pad missing channels, including $c_2 \equiv 0$ when $p < 2$. The following statement does not require non-collapse.

Proposition I.1 (Passive coefficient matching)**.** *Let $U \subset R_z$ be a nonempty open interval. Assume latent-state consistency through $f \in C^2(U)$ with $f(U) \subset D_\eta$, and at least $d_v + 1$ pairwise-distinct physical velocities realized at interior trajectory points at every $u \in U$. If the physical and encoder-induced paths satisfy the declared laws at those points, then, for every $u \in U$,*

$$c_{r,\eta}(f(u))f'(u)^r = f'(u)c_{r,\theta}(u), \qquad 0 \le r \le p,\ r \ne 2, \tag{38}$$

$$c_{2,\eta}(f(u))f'(u)^2 = f''(u) + f'(u)c_{2,\theta}(u). \tag{39}$$

When both declared laws exclude the quadratic channel, (39) reduces to $f'' = 0$. With non-collapse this recovers the primary invertible affine workflow. When the channel is present, the same identity is the one-dimensional connection transformation law.

Proof of Proposition I.1. Fix $u \in U$. As in the preceding proof, differentiating the local factorization $\hat{z} = f(z)$ twice and substituting the two polynomial-velocity laws shows that the residual at physical velocity v is exactly

$$R_u(v) = f''(u)v^2 + f'(u)\sum_{r=0}^{p} c_{r,\theta}(u)v^r - \sum_{r=0}^{p} c_{r,\eta}(f(u))(f'(u)v)^r. \tag{40}$$

It belongs to $\operatorname{span}\{1, v, \dots, v^{d_v}\}$ with $d_v = \max\{2, p\}$ and vanishes at every realized velocity used by the proposition hypothesis. Its monomial feature matrix contains a square Vandermonde submatrix and therefore has full column rank, so the coefficient vector of R_u is zero and hence $R_u \equiv 0$ as a polynomial.

For $r \ne 2$, the coefficient of v^r is

$$f'(u)c_{r,\theta}(u) - c_{r,\eta}(f(u))f'(u)^r,$$

so its vanishing gives (38). The coefficient of v^2 is

$$f''(u) + f'(u)c_{2,\theta}(u) - c_{2,\eta}(f(u))f'(u)^2,$$

and its vanishing gives (39). Whenever both quadratic coefficients of the declared laws are identically zero, including every $p \le 1$ family, the last identity is $f''(u) = 0$. Since u was arbitrary, all identities hold on U. □

I.3 Construction and uniqueness of the normalizer

For each law choose $u_{\xi,\star} \in D_\xi$ and define

$$\psi_\xi(u) := \int_{u_{\xi,\star}}^{u} \exp\left(-\int_{u_{\xi,\star}}^{s} c_{2,\xi}(r)\,dr\right) ds. \tag{41}$$

We verify the claims in Section 4.2, including the basepoint values $\psi_\xi(u_{\xi,\star}) = 0$ and $\psi'_\xi(u_{\xi,\star}) = 1$. Suppress the law index, let c_2 be continuous on the open interval D, and write

$$h(u) := \exp\left(-\int_{u_\star}^{u} c_2(r)\,dr\right).$$

The fundamental theorem of calculus gives $h \in C^1(D)$, $h(u) > 0$, $h(u_\star) = 1$, and $h'(u) = -c_2(u)h(u)$. Defining $\psi(u) = \int_{u_\star}^{u} h(s)ds$ therefore gives $\psi \in C^2(D)$,

$$\psi' = h > 0, \qquad \psi'' = -c_2\psi',$$

which proves (14). Strict positivity of ψ' makes ψ strictly increasing. The inverse-function theorem gives a local C^2 inverse at every point, strict monotonicity makes these local inverses agree globally, and the image $J = \psi(D)$ is open. Hence $\psi : D \to J$ is a C^2 diffeomorphism.

For a trajectory satisfying (13), set $q = \psi(z)$. Then

$$\begin{aligned} q'' &= \psi''(z)(z')^2 + \psi'(z)z'' \\ &= \big(\psi''(z) + c_2(z)\psi'(z)\big)(z')^2 + \sum_{r\neq 2} \psi'(z)c_r(z)(z')^r \\ &= \sum_{r\neq 2} \psi'(z)^{1-r} c_r(z)(q')^r, \end{aligned}$$

where $q' = \psi'(z)z'$ was used in the last line, which identifies the transformed coefficients. Restoring the law index, the canonical equation is

$$q'' = \sum_{\substack{0\le r\le p \\ r\neq 2}} \bar{c}_{r,\xi}(q)(q')^r, \qquad q \in J_\xi. \tag{42}$$

Its coefficients are, explicitly,

$$\bar{c}_{r,\xi}(\psi_\xi(u)) := \psi_\xi'(u)^{1-r} c_{r,\xi}(u) \quad (0 \le r \le p,\ r \neq 2), \qquad \bar{c}_{2,\xi} := 0; \tag{43}$$

$\bar{c}_{r,\xi}(q)$ multiplies $(q')^r$ and need not be polynomial in q.

Finally, let ϕ be any other C^2 solution of $\phi'' + c_2\phi' = 0$ with ϕ' nowhere zero. Because both derivatives solve the first-order equation $y' = -c_2 y$,

$$\left(\frac{\phi'}{\psi'}\right)' = \frac{\phi''\psi' - \phi'\psi''}{(\psi')^2} = 0.$$

Thus $\phi' = \lambda\psi'$ for a constant $\lambda \neq 0$, and integration gives $\phi = \lambda\psi + \tau$. This proves uniqueness up to affine post-composition, including normalizers defined using different basepoints.

I.4 Canonical affine compatibility

Two canonical laws are affine-compatible on $\psi_\theta(U)$ if some $A(q) = \lambda q + \tau$, $\lambda \neq 0$, maps $\psi_\theta(U)$ into J_η and satisfies, for every $0 \le r \le p$, $r \neq 2$,

$$\bar{c}_{r,\eta}(A(q))\lambda^r = \lambda \bar{c}_{r,\theta}(q), \qquad q \in \psi_\theta(U). \tag{44}$$

Theorem I.2 (Canonical connection boundary)**.** *Let the θ- and η-laws be two scalar polynomial-velocity laws of the same declared degree p on open intervals, with missing channels zero-padded, continuous coefficient functions, and the normalizers of (14) under the stated basepoint convention. Let $U \subset D_\theta$ be a nonempty open interval. The following are equivalent.*

1. *There is a non-collapsed $f \in C^2(U)$ with $f(U) \subset D_\eta$ satisfying all raw compatibility identities (38)–(39).*
2. *There is an affine map $A(q) = \lambda q + \tau$, $\lambda \neq 0$, with $A(\psi_\theta(U)) \subset J_\eta$ satisfying the canonical coefficient identities (44) on $\psi_\theta(U)$.*

The witnesses correspond bijectively through

$$A = \psi_\eta \circ f \circ \psi_\theta^{-1}, \qquad f = \psi_\eta^{-1} \circ A \circ \psi_\theta. \tag{45}$$

The choice of basepoint, origin, scale, or orientation for either normalizer changes A only by invertible affine pre- or post-composition.

Proof of Theorem I.2. Assume statement 1 and define $g = \psi_\eta \circ f \circ \psi_\theta^{-1}$ on $\psi_\theta(U)$. Put $x = \psi_\theta(u)$. A first differentiation gives

$$g'(x) = \frac{\psi_\eta'(f(u))f'(u)}{\psi_\theta'(u)}. \tag{46}$$

Since $d/dx = (1/\psi_\theta'(u))d/du$, differentiating once more yields

$$\begin{aligned} g''(x) &= \frac{1}{\psi_\theta'(u)} \frac{d}{du}\left(\frac{\psi_\eta'(f(u))f'(u)}{\psi_\theta'(u)}\right) \\ &= \frac{\psi_\eta'(f(u))}{\psi_\theta'(u)^2}\left[f''(u) + f'(u)c_{2,\theta}(u) - c_{2,\eta}(f(u))f'(u)^2\right]. \end{aligned} \tag{47}$$

To obtain the last line, we used $\psi_\theta'' = -c_{2,\theta}\psi_\theta'$ and $\psi_\eta'' = -c_{2,\eta}\psi_\eta'$. The bracket vanishes by (39), so $g'' = 0$. Therefore $g(x) = \lambda x + \tau$ on the interval $\psi_\theta(U)$. The factors ψ_θ' and ψ_η' in (46) are strictly positive. Since $f' \not\equiv 0$, $g' \not\equiv 0$; because $g' = \lambda$ is constant, $\lambda \neq 0$. Equation (46) then also gives

$$f'(u) = \lambda \frac{\psi_\theta'(u)}{\psi_\eta'(f(u))} \neq 0 \quad \text{for every } u \in U,$$

which proves the automatic upgrade from the non-collapse premise to a local diffeomorphism.

It remains to transform the nonquadratic coefficients. For $r \neq 2$, using (43), (46), and (38), we obtain

$$\begin{aligned}
\bar{c}_{r,\eta}(g(x))g'(x)^r &= \psi_\eta'(f(u))^{1-r} c_{r,\eta}(f(u)) \left(\frac{\psi_\eta'(f(u))f'(u)}{\psi_\theta'(u)} \right)^r \\
&= \frac{\psi_\eta'(f(u))}{\psi_\theta'(u)^r} c_{r,\eta}(f(u)) f'(u)^r \\
&= \frac{\psi_\eta'(f(u))f'(u)}{\psi_\theta'(u)^r} c_{r,\theta}(u) \\
&= g'(x)\psi_\theta'(u)^{1-r} c_{r,\theta}(u) = g'(x)\bar{c}_{r,\theta}(x).
\end{aligned}$$

Substituting $g = A$ and $g' = \lambda$ gives (44), proving statement 2.

Conversely, assume statement 2 and define $f = \psi_\eta^{-1} \circ A \circ \psi_\theta$. Each factor is a C^2 diffeomorphism on the relevant interval and A' is the nonzero constant λ, so f is a C^2 diffeomorphism from U onto its image. Reversing the coefficient calculation in the preceding paragraph gives (38). Moreover $\psi_\eta \circ f \circ \psi_\theta^{-1} = A$ has second derivative zero. Applying the identity (47) and using positivity of its prefactor gives (39). Thus statement 1 holds.

The two constructions in (45) are inverse because ψ_θ and ψ_η are bijective. If the normalizers are replaced by $\widetilde{\psi}_\theta = B_\theta \circ \psi_\theta$ and $\widetilde{\psi}_\eta = B_\eta \circ \psi_\eta$, where B_θ, B_η are invertible affine maps by the uniqueness result above, then the canonical witness becomes $B_\eta \circ A \circ B_\theta^{-1}$, again invertible affine. Existence of a canonical affine witness, and hence the canonical family-compatible outer set, is unchanged. □

Corollary I.3 (Constant linear damping is state-gauge invariant)**.** *Under Theorem I.2, interpret any absent linear channel as $c_{1,\xi} \equiv 0$. Then*

$$c_{1,\eta}(f(u)) = c_{1,\theta}(u), \qquad u \in U. \tag{48}$$

In particular, if both laws use $c_{1,\xi} \equiv -\delta_\xi$, then $\delta_\eta = \delta_\theta$, even when the raw state map is nonlinear and c_2 is nonzero.

Proof of Corollary I.3. If $p = 0$, both zero-padded linear coefficients vanish, so (48) and the constant-damping conclusion are immediate. If $p \geq 1$, taking $r = 1$ in (38) gives $c_{1,\eta}(f(u))f'(u) = f'(u)c_{1,\theta}(u)$. Theorem I.2 upgrades the non-collapse premise to $f'(u) \neq 0$ at every $u \in U$, so division by $f'(u)$ proves (48). Substituting $c_{1,\xi} \equiv -\delta_\xi$ gives the constant-damping claim. □

Proof of Theorem 4.2. Let (E_ϕ, f, η) be an admissible report as in Section 2, with the physical-velocity coverage in Theorem 4.2. The common map satisfies Definition 2.1 along the continuous trajectories on U, so the chain rule applies at the covering points. Proposition I.1 gives all raw compatibility identities. Theorem I.2 then gives an affine A satisfying (44); by the definition (15), $\eta \in \mathcal{O}_U^{\mathrm{can}}(\theta)$.

At a covered frame, state consistency gives $E_\phi(x^{(m)}(t)) = f(z^{(m)}(t))$. Post-composing by ψ_η and using $\psi_\eta \circ f = A \circ \psi_\theta$ from (45) yields

$$\psi_\eta\Big(\hat{z}^{(m)}(t)\Big) = \psi_\eta(f(z^{(m)}(t))) = A(\psi_\theta(z^{(m)}(t))),$$

which is (16). The same f, and therefore the same A, is common to all clips by latent-state consistency. □

I.5 PROJECTED-PENDULUM EXAMPLE

Consider the angular law

$$q'' + \delta q' + \omega^2 \sin q = 0. \tag{49}$$

Corollary I.4 (Projected pendulum: apparent inertia and exact normalizer)**.** *Let q satisfy (49) and set $y = \sin q$ on a branch contained in $(-\pi/2, \pi/2)$. Then y satisfies*

$$y'' + \delta y' + \omega^2 y\sqrt{1-y^2} + \frac{y}{1-y^2}(y')^2 = 0. \tag{50}$$

In the right-hand-side convention, $c_2(y) = -y/(1-y^2)$ and the normalizer is $\psi(y) = \arcsin y$ up to an affine transformation. The endpoints $y = \pm 1$ lie outside the regular chart.

Proof of Corollary I.4. On a branch contained in $(-\pi/2, \pi/2)$, $y = \sin q$ has $\cos q = \sqrt{1-y^2} > 0$. Differentiation gives

$$y' = \cos q\, q', \qquad y'' = \cos q\, q'' - \sin q\, (q')^2.$$

Substitute $q'' = -\delta q' - \omega^2 \sin q$, $q' = y'/\sqrt{1-y^2}$, and $\sin q = y$:

$$y'' = -\delta y' - \omega^2 y\sqrt{1-y^2} - \frac{y}{1-y^2}(y')^2.$$

Moving all terms to the left proves (50). In the right-hand-side convention, the quadratic coefficient is $c_2(y) = -y/(1-y^2)$. Therefore

$$\psi'(y) = \exp\left(-\int c_2(y)dy\right) = \frac{1}{\sqrt{1-y^2}}$$

up to a nonzero multiplicative constant, and integration gives $\psi(y) = \arcsin y$ up to an affine post-composition. At $y = \pm 1$ this derivative diverges and the forward map has derivative $d(\sin q)/dq = 0$, so neither the coefficient-domain nor local diffeomorphism hypotheses hold there. □

J PROOFS FOR THE FAMILY-SPECIFIC CONCLUSIONS

J.1 PROOFS OF TABLE 1

Proof. Under the coverage and family assumptions of Table 1, Theorem 3.1 supplies an affine witness $f(u) = \lambda u + \tau$ on U, with $\lambda \neq 0$, satisfying (6). Tildes denote reported parameters, with the same family conventions as the physical parameters.

LTI. The orbit equations give $\tilde{\delta} = \delta$ and $\tilde{\alpha}(\lambda u + \tau) = \lambda\alpha u$ on U. Matching coefficients yields $\tilde{\alpha} = \alpha$ and $\alpha\tau = 0$. Conversely, every $\lambda \neq 0$ and τ satisfying that restriction is a witness, including arbitrary translations when $\alpha = 0$.

Damped pendulum. With $\omega, \tilde{\omega} > 0$, the orbit equations give

$$\tilde{\delta} = \delta, \qquad \tilde{\omega}^2 \sin(\lambda u + \tau) = \lambda\omega^2 \sin u \qquad \forall u \in U. \tag{51}$$

Let $\kappa = \lambda\omega^2/\tilde{\omega}^2$. The identity $\sin(\lambda u + \tau) = \kappa \sin u$ is first obtained only on the covered nonempty interval U; by analyticity of sine, it extends to all of $\mathbb{R}$. Differentiating twice gives $-\lambda^2 \sin(\lambda u + \tau) = -\kappa \sin u$. Combining with the original identity yields $(\lambda^2 - 1)\kappa \sin u = 0$ for all u, so $\lambda = \pm 1$. If $\lambda = 1$, then $\sin(u + \tau) = \kappa \sin u$ forces $\sin\tau = 0$ and $\cos\tau = \kappa > 0$, hence $\tau = 2\pi k$ and $\kappa = 1$. If $\lambda = -1$, the same expansion gives $\tau = 2\pi k$ and $\kappa = -1$. In both cases $\tilde{\omega} = \omega$, and direct substitution verifies every witness $u \mapsto \sigma u + 2\pi k$ with $\sigma \in \{\pm 1\}$ and $k \in \mathbb{Z}$.

Normalized Van der Pol. Because both physical latent and encoder-induced laws are restricted to the normalized Van der Pol family, the orbit equations are

$$\tilde{\mu}((\lambda u + \tau)^2 - 1) = \mu(u^2 - 1), \qquad \lambda u + \tau = \lambda u \qquad \forall u \in U. \tag{52}$$

These polynomial identities are first obtained on U and then extend to all $u \in \mathbb{R}$. The second equation gives $\tau = 0$. Matching coefficients in $\tilde{\mu}(\lambda^2 u^2 - 1) = \mu(u^2 - 1)$ gives $\tilde{\mu} = \mu$ and then $\lambda^2 = 1$; direct substitution verifies both signs.

Cubic Duffing. The orbit equations give

$$\tilde{\delta} = \delta, \qquad \tilde{\alpha}(\lambda u + \tau) + \tilde{\beta}(\lambda u + \tau)^3 = \lambda(\alpha u + \beta u^3) \qquad \forall u \in U.$$

The polynomial identity extends from U to all $u \in \mathbb{R}$. Its u^3 and u^2 coefficients give $\tilde{\beta}\lambda^3 = \lambda\beta$ and $3\tilde{\beta}\lambda^2\tau = 0$. Since $\lambda\beta \neq 0$, we obtain $\tau = 0$, $\tilde{\beta} = \beta/\lambda^2$, and then $\tilde{\alpha} = \alpha$. Direct substitution verifies every scale $\lambda \neq 0$.

Quintic Duffing. The orbit equations give

$$\tilde{\delta} = \delta, \qquad \tilde{\alpha}(\lambda u + \tau) + \tilde{\beta}(\lambda u + \tau)^3 + \tilde{\gamma}(\lambda u + \tau)^5 = \lambda(\alpha u + \beta u^3 + \gamma u^5) \qquad \forall u \in U. \quad (53)$$

The polynomial identity is first obtained on the covered nonempty interval U and therefore holds for all $u \in \mathbb{R}$. Its u^5 and u^4 coefficients give $\tilde{\gamma}\lambda^5 = \lambda\gamma$ and $5\tilde{\gamma}\lambda^4\tau = 0$. Since $\lambda\gamma \neq 0$, we have $\tilde{\gamma} \neq 0$ and hence $\tau = 0$. Matching the u^3 and u coefficients then yields $\tilde{\beta} = \beta/\lambda^2$ and $\tilde{\alpha} = \alpha$; the u^5 identity also gives $\tilde{\gamma} = \gamma/\lambda^4$. Direct substitution proves the converse, so this is the exact quintic-Duffing orbit.

The Duffing amplitude-anchor statements follow from the same scale action $u \mapsto \lambda u$. In the quintic case, $\tilde{\beta} = \beta/\lambda^2$ and $\tilde{\gamma} = \gamma/\lambda^4$, while δ, α, $\operatorname{sign}(\beta)$, and γ/β^2 are invariant. A signed calibration of a nonzero physical latent value fixes λ. An unsigned amplitude calibration fixes only $|\lambda|$, but this is already sufficient because both nonzero-weight coefficients transform through even powers of λ. The same argument for the cubic family uses only $\tilde{\beta} = \beta/\lambda^2$. Each anchor must pair a nonzero physical state in U with its encoder value; an anchor outside U needs a separately justified extension of the affine witness. □

J.2 Failure without a top gap: Helmholtz

A missing degree away from the degree immediately below the top need not lock translations. For $z'' + \delta z' + \alpha z + \beta z^2 = 0$ on $\mathbb{R}$, with $\alpha\beta \neq 0$, both coefficient signs allowed, and no restriction on nonzero scales, the damping identity gives $\tilde{\delta} = \delta$. Coefficient matching in $\tilde{\alpha}(\lambda u+\tau)+\tilde{\beta}(\lambda u+\tau)^2 = \lambda(\alpha u+\beta u^2)$ gives $\tilde{\beta} = \beta/\lambda$, $\tilde{\alpha}+2\beta\tau/\lambda = \alpha$, and $\tau(\tilde{\alpha}+\beta\tau/\lambda) = 0$. Solving these identities yields

$$(\tau, \tilde{\alpha}, \tilde{\beta}) = (0, \alpha, \beta/\lambda) \quad \text{or} \quad (\tau, \tilde{\alpha}, \tilde{\beta}) = \left(\frac{\alpha\lambda}{\beta}, -\alpha, \frac{\beta}{\lambda}\right).$$

Thus $f(u) = u + 1$ exchanges $z'' + z + z^2 = 0$ and $y'' - y + y^2 = 0$. Before calibration, the translation branch flips $\operatorname{sign}(\alpha)$, and δ and α^2 form a complete set of parameter-orbit invariants for this unrestricted nonzero stratum: the two branches connect either sign of α, and the nonzero scale connects any two nonzero values of β with matching δ, α^2. If $0 \in U$, an origin anchor $f(0) = 0$ removes the translated branch and makes α itself invariant, but leaves the scale λ free, with $\tilde{\beta} = \beta/\lambda$. After that origin calibration, one additional signed state anchor at $u_0 \in U \setminus \{0\}$ fixes λ and hence β. An unsigned anchor fixes only $|\lambda|$ and leaves the admissible reflection branch unresolved. Anchors outside U require a separately justified extension of the affine witness to their locations.

K Detailed Limitations

This appendix integrates the scope of scalar coverage with the limitations of parameter recovery. We first explain the obstruction to directly extending our coordinate argument to multiple dimensions, then discuss what the scalar setting does cover, and finally distinguish these structural results from finite-video estimation.

K.1 Multidimensional scope and the coverage obstruction

Our results concern a scalar configuration variable z; the associated first-order phase-space variable (z, z') is already two-dimensional. The restriction is therefore to one physical coordinate, not to one phase-space component or one unknown parameter. The theory does not directly cover coupled configuration vectors $\boldsymbol{q} \in \mathbb{R}^d$ with $d > 1$.

The difficulty is not only additional algebra. In the ideal continuous-time setting, a nonconstant scalar trajectory can traverse a nonempty open coordinate interval. By contrast, the compact images of finitely many smooth trajectories in a higher-dimensional configuration space have empty interior. They can leave an open region unvisited, where a coordinate transformation can be changed without affecting its values or derivatives on the observed curves. The following result makes this limitation precise.

Proposition K.1 (Finite passive-curve obstruction)**.** *Let $d > 1$, let $U \subset \mathbb{R}^d$ be a nonempty open set, and let $C \subset U$ be the compact image of finitely many physical latent trajectories with empty interior. Then there exists a non-affine C^∞ diffeomorphism $h_\varepsilon : U \to h_\varepsilon(U)$ such that $h_\varepsilon(\boldsymbol{q}) = \boldsymbol{q}$, $Dh_\varepsilon(\boldsymbol{q}) = I$, and $D^2 h_\varepsilon(\boldsymbol{q}) = 0$ for all $\boldsymbol{q} \in C$. Thus realized positions, velocities, accelerations, and pointwise residuals along the clips are unchanged, while the coordinate map is non-affine away from the data. This is not a family-preserving parameter transformation; it only explains why coordinate-level multi-dimensional recovery needs additional structure.*

Proof. Because C is compact with empty interior in the open set U, choose a point $p \in U \setminus C$ and a radius $r > 0$ such that the closed ball $\overline{B}(p, r)$ is contained in $U \setminus C$. Let $\rho \in C_c^\infty(B(p, r))$ be a nonzero bump function and let e_1 be the first coordinate vector. Define

$$h_\varepsilon(q) = q + \varepsilon \rho(q) e_1. \tag{54}$$

Extend ρ by zero to $\mathbb{R}^d$ and choose $0 < |\varepsilon| \operatorname{Lip}(\rho) < 1$. Then, for all $q, q' \in \mathbb{R}^d$,

$$\|h_\varepsilon(q) - h_\varepsilon(q')\| \geq \big(1 - |\varepsilon| \operatorname{Lip}(\rho)\big) \|q - q'\|,$$

so h_ε is globally injective. Its Jacobian is everywhere invertible, and the inverse-function theorem therefore makes it a smooth diffeomorphism from U onto its image. The map is non-affine because ρ is nonzero and compactly supported. Since the support of ρ is disjoint from C, ρ, $D\rho$, and $D^2\rho$ vanish on C. Hence $h_\varepsilon(q) = q$, $Dh_\varepsilon(q) = I$, and $D^2 h_\varepsilon(q) = 0$ for all $q \in C$.

For a realized physical latent trajectory $q(t) \in C$, the transformed trajectory $h_\varepsilon(q(t))$ has the same value, first derivative, and second derivative as $q(t)$ at every sampled time along the clip. Therefore any pointwise residual evaluated only on the realized clips is unchanged, even though the coordinate map is non-affine away from the data. □

The proposition concerns recovery of a coordinate map on an open region, not parameter identification in every fixed multidimensional ODE family. The constructed transformation is not asserted to preserve such a family or to change its parameters. It therefore does not exclude parameter recovery under additional restrictions. Rather, it shows why matching passive trajectories and their derivatives alone cannot determine an unrestricted coordinate extension. Restricted coordinate classes or richer observations may change this boundary, but require a separate analysis.

K.2 Why scalar dynamics remain informative

The scalar setting is not synonymous with linear dynamics or a single unknown coefficient. The semilinear family permits nonlinear restoring forces and state-dependent damping, including the pendulum, cubic and quintic Duffing, and normalized Van der Pol examples. The extensions treat specified nonlinear velocity features and polynomial-velocity laws, including quadratic terms induced by nonlinear observation coordinates. The restriction limits the number of coupled coordinates, rather than reducing the analysis to the LTI case.

Dimension also changes how passive coverage can constrain the coordinate map. On a scalar interval U, sufficiently many velocities at each state turn the chain-rule residual into coefficient identities throughout U. For semilinear laws, these force $f'' = 0$; at the studied quadratic-velocity boundary, the corresponding identities force an affine relation after law-derived normalization. This supports parameter analysis across the covered interval rather than only along isolated observed points. Merely visiting every state in U is not sufficient: the required velocity variation must also be present. The two-slope construction in Appendix E shows why repeated visits along the same restricted velocity branches need not supply it.

Within this scope, the parameter conclusions are already distinct: angular pendulum parameters can be identified despite residual coordinate symmetries, while Duffing coefficients can retain scale

dependence that physical calibration resolves. These are different outcomes of the theory, not failures to obtain the same conclusion for every family. The scope is nevertheless restricted: the family is prescribed and autonomous, the velocity dependence must satisfy the stated structural conditions, and canonical comparison is local to regular coordinate intervals. It is not a classification of arbitrary nonlinear ODEs or a global multi-chart result.

K.3 From structural guarantees to empirical recovery

Within the covered model classes, a separate gap remains between structural identification and estimation from sampled videos. The exact results assume a shared non-collapsed C^2 map and the required physical-velocity coverage. Finite observations provide only sampled evidence about these continuous premises. A low residual alone establishes neither the state map nor the correctness of the declared family. The audits in Appendix D describe which evidence should be reported without treating it as a proof of those premises.

The semilinear stability results quantify deterministic effects of residuals, level mismatch, and design conditioning; Appendix C separates numerical differentiation and state-consistency errors. They do not provide a statistical pixel-to-parameter rate. Likewise, the encoder construction in Proposition 3.3 establishes realizability under its representation assumptions, not convergence of training to that construction. Structural uniqueness, existence of a compatible encoder, and numerical accuracy are different claims.

Physical calibration has a similar distinction. The rank criterion in Appendix G supplies sufficient local uniqueness on a regular lifted set, with disconnected branches requiring separate checks. It does not by itself quantify the effect of measurement error in the anchors. Extending the recovery guarantees to jointly account for image-derived errors, motion-coverage conditioning, and physical-anchor uncertainty would connect the structural analysis more directly to empirical parameter accuracy.

L Evaluation metrics

The three primary metrics assess complementary aspects of recovery across the synthetic and real-video experiments: coordinate agreement, parameter agreement under the permitted coordinate transformation or calibration, and agreement of the resulting dynamical function. The formulas below define dimensionless ratios; smaller values indicate closer agreement. In Figure 2, coordinate and parameter comparisons follow the chosen raw or canonical route; physical-unit comparisons additionally use the stated anchors. The metrics assess empirical accuracy within those conventions. Their relation to the conditional identifiability results is discussed in Appendix D.

Coordinate NRMSE e_{map} (%). For each fit, let c index test clips and t the valid frames within a clip:

$$e_{\mathrm{map}} = \mathrm{median}_c \, \frac{\mathrm{RMS}_t(\widetilde{x}_{ct} - x_{ct})}{\mathrm{std}_t(x_{ct})}. \tag{55}$$

Here x is the physical state in synthetic experiments, or tracked angle, pixel position, or inverse radius in real videos. The prediction $\widetilde{x}$ uses a coordinate map fitted on the stated non-test references and frozen for evaluation. Unrestricted affine comparisons fit the encoding as a function of the reference and invert the fitted map; family-constrained comparisons retain their allowed sign, period, scale, or branch restrictions. Raw and canonical representations use their respective fitted maps. NRMSE is computed within each clip, then summarized by the clip median; the standard deviation uses divisor N_c. This is a coordinate-recovery error: it tests whether a frozen readout of the encoder recovers the reference physical state on new clips, rather than evaluating an integrated trajectory rollout.

Parameter relative error e_θ (%). For coefficients expressed in the same comparison coordinate, use

$$e_{\theta,j} = \frac{|\hat{\theta}_j - \theta_j^{\mathrm{ref}}|}{|\theta_j^{\mathrm{ref}}|}, \qquad e_\theta = \max_{j \in J} e_{\theta,j}. \tag{56}$$

The frozen coordinate map determines the reference point on the parameter orbit. For example, under $z = \lambda q$, a learned cubic coefficient is compared with $\beta^{\mathrm{ref}} = \beta/\lambda^2$. After external calibration, the

estimate is instead compared in physical units. The six-family experiment takes the maximum over its nonzero dynamical coefficients J; coverage uses the scale-corrected mean drag coefficient, and real videos use the calibrated scalar L or g. Normalization reports the individual errors $e_{\theta,j}$. Thus this metric tests parameter agreement at the coordinate map fixed from reference pairs.

Dynamics relative RMSE e_{dyn} (%). For a transformed law or channel $\widetilde{F}$ and its physical reference F, define

$$e_{\mathrm{dyn}} = \frac{\mathrm{RMS}_{\mathcal{S}}(\widetilde{F} - F)}{\mathrm{RMS}_{\mathcal{S}}(F)}. \tag{57}$$

The evaluation set $\mathcal{S}$ contains all held-out frames for the matched-LTI and overhead experiments, or the stated physical-state grid for the normalized pendulum's restoring-force and damping channels. Samples receive equal weights within this set. This metric tests whether the transformed law reproduces the reference acceleration or force over the evaluation domain, rather than comparing individual coefficients or integrated trajectory rollouts. The RMS denominator also accommodates nonzero constant acceleration such as gravity.

Aggregation across fits. All three errors are multiplied by 100 and reported as percentages. They are computed per fit before reporting median [Q25, Q75] across the stated cohort. Quartiles use linear interpolation between ordered observations. Physical estimates use mean and sample SD (divisor $n - 1$). These summaries describe variability over the stated IC/video groups and optimizer seeds. Each experimental protocol specifies its coordinate constraints, reference quantities, and evaluation domain.

M Synthetic experiments: protocols and supporting results

The appendix follows the main experiments. Appendix M.1 supports Section 5.1, including the matched affine-LTI comparison. Appendices M.2 and M.3 separately support the velocity-coverage and projection experiments in Section 5.2. Each experiment follows the same sequence: observations and data, theoretical predictions, model and training, evaluation, and results and discussion. Additional analyses follow the main results where applicable.

M.1 Six-family experiments

M.1.1 Observations and data

The six-family experiment tests the coordinate and parameter relations of Section 3. The five families in Table 1 are supplemented by the quadratic/Helmholtz family, whose two affine branches are described in Appendix J.2. The LTI experiment uses an affine extension with a learned equilibrium $\hat{c}$, which absorbs translations while leaving δ and α invariant. We simulate each prescribed law at fixed physical coefficients and render the resulting state trajectories. The angular pendulum is shown as a moving pendulum; the other families use spring-oscillator renderings (Figure 3). The simulated states provide evaluation references and are not coordinate targets for encoder training.

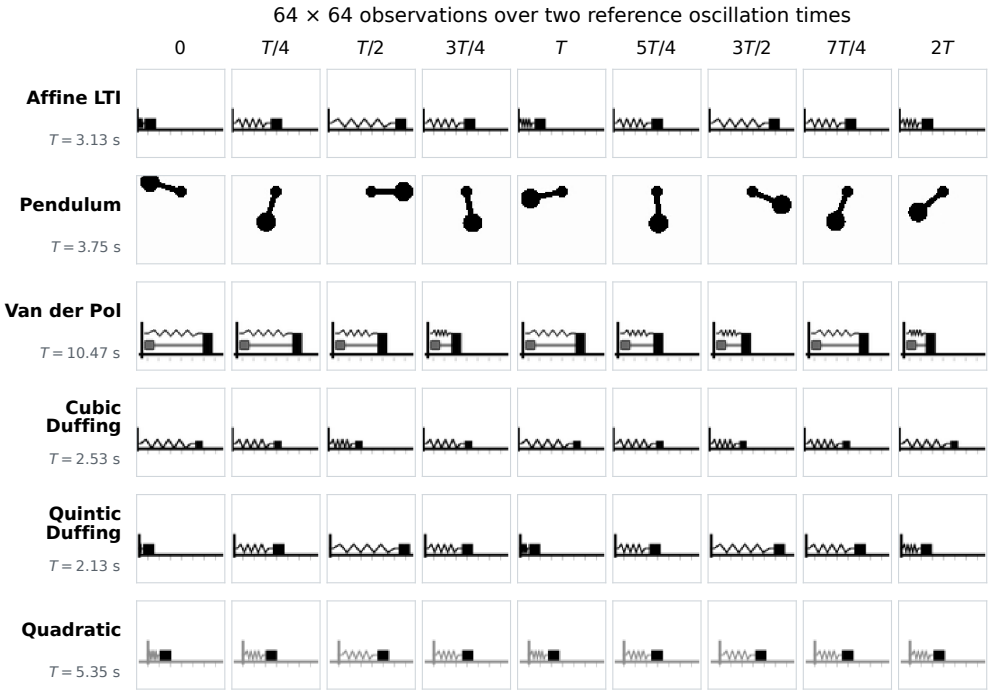


Figure 3: **Temporal evolution over two reference oscillation times.** Nine observations at $0, T/4, \ldots, 2T$ show amplitude and timing changes. Each row starts at a turning point; T is the first interval between same-type turning points, not an assumed constant period. Damped systems ring down, whereas Van der Pol approaches a limit cycle. All frames are 64×64.

Table 8: **Synthetic observations and data splits.** Clip counts are per IC collection. Each family uses ten collections $\times$ ten optimizer seeds. Training clips supply images and timestamps for fitting the encoder and law. Test clips supply the reported held-out errors (coordinate NRMSE e_{map} (%)). The permitted coordinate comparison is fitted using training reference pairs and frozen on test clips.

Family / experiment	Pixels / Hz / duration	Train / test
Affine LTI	64^2 / 60 / 6 s	16 / 8
Pendulum	64^2 / 60 / 8 s	10 / 4
Van der Pol	64^2 / 60 / 12 s	6 / 4
Cubic Duffing	64^2 / 60 / 8 s	10 / 4
Quintic Duffing	64^2 / 60 / 3 s	16 / 8
Quadratic / Helmholtz	64^2 / 60 / 8 s	3 / 4

For each ODE family, physical coefficients (Table 10) and the rendering map are fixed. Table 9 specifies the nominal initial position–velocity pairs (q_0, v_0), with $v_0 = q'(0)$; the learned initial value is $z_0 = E_\phi(x_0)$ and is not prescribed separately. Starting from these pairs, ten data seeds generate ten trajectory collections; the different initial conditions produce different observed video sequences. Each collection contains the training and test clips listed in Table 8, with split assignments inherited from the prescribed trajectory categories. For the pendulum, initial angles remain fixed and initial angular velocities receive independent multiplicative perturbations in $[0.95, 1.05]$. For the other five families, initial positions and velocities are perturbed independently by the same factors. Each fixed video collection is then fitted with ten optimizer seeds, yielding $10 \times 10 = 100$ fits per family. Data seeds control initial-condition variation, while optimizer seeds control model initialization and training sampling.

Table 9: **Nominal initial states before IC perturbation.** Pairs are (q_0, v_0) in the physical coordinate. Each $\pm$ includes both signs; in the pendulum row, angle and velocity signs vary independently. LTI and quintic use the same IC design.

Family	Training pairs	Test pairs
LTI / quintic	$(\pm a, 0)$, $a \in \{0.25, 0.55, 0.85, 1.15\}$; $(0, \pm b)$, $b \in \{0.8, 1.8, 2.8, 3.4\}$	$(\pm a, 0)$, $a \in \{0.65, 1.00\}$; $(0, \pm b)$, $b \in \{0.40, 1.00\}$
Pendulum	$(\pm\pi/2, \pi h)$, $h \in \{-1/2, -1/4, 0, 1/4, 1/2\}$	$(\pm\pi/2, \pm\pi/8)$
Van der Pol	$(0, \pm 0.5)$, $(0, \pm 2.5)$, $(\pm 3, 0)$	$(0, \pm 1.25)$, $(\pm 2.75, 0)$
Cubic	$(0, \pm\sqrt{4p^2 + 2p^4})$, $p \in \{0.40, 0.55, 0.70, 0.85, 1.00\}$	$(0, \pm\sqrt{4p^2 + 2p^4})$, $p \in \{0.625, 0.925\}$
Quadratic	$(0, -1.2)$, $(0, 0.8)$, $(0, 1.2)$	$(-0.4, 0.65)$, $(0.4, -0.65)$, $(-0.15, 1.0)$, $(0.15, -1.0)$

M.1.2 Theoretical predictions

Under the shared-state and velocity-coverage assumptions of Theorem 3.1, a compatible coordinate map is affine. Requiring the transformed law to remain in the declared family further restricts this map and determines the corresponding parameter orbit (Table 10). Thus invariant coefficients should be preserved across fits, while coefficients that depend on coordinate scale should follow the same scale inferred from the learned coordinate. The quadratic family also requires checking both permitted affine branches. These experiments test coordinate recovery and parameter recovery against the same family-permitted relation.

For experimental plots, write q for physical state and $z = E_\phi(x)$ for learned state (the theory's z and $\hat{z}$, respectively), with hats on learned coefficients. The comparison map is determined by the declared dynamical family.

Table 10: Declared laws, fixed physical coefficients and permitted coordinate actions. Nonzero λ is estimated from training coordinate pairs. The quintic invariant uses $\beta \neq 0$; the plotted physical coefficients are positive.

Learned law / physical coefficients	Allowed map / parameter consequence
$z'' + \hat{\delta} z' + \hat{\alpha}(z - \hat{c}) = 0$ $(\delta, \alpha, c) = (0.2, 4, 0)$	$z \simeq \lambda q + \tau$; $\hat{c} \simeq \lambda c + \tau$; δ, α invariant.
$z'' + \hat{\delta} z' + \hat{\omega}^2 \sin z = 0$ $(\delta, \omega^2) = (0.15, 4)$	$z \simeq sq + 2\pi n$; $s = \pm 1$, $n \in \mathbb{Z}$; δ, ω^2 invariant.
$z'' - \hat{\mu}(1 - z^2) z' + z = 0$ $\mu = 1.5$	$z \simeq \pm q$; μ invariant.
$z'' + \hat{\delta} z' + \hat{\alpha} z + \hat{\beta} z^3 = 0$ $(\delta, \alpha, \beta) = (0.2, 4, 4)$	$z \simeq \lambda q$; $\hat{\beta} \simeq \beta/\lambda^2$; $\delta, \alpha, \operatorname{sign} \beta$ invariant.
$z'' + \hat{\delta} z' + \hat{\alpha} z + \hat{\beta} z^3 + \hat{\gamma} z^5 = 0$ $(\delta, \alpha, \beta, \gamma) = (0.2, 4, 3, 2)$	$z \simeq \lambda q$; $\hat{\gamma} \simeq \gamma/\lambda^4$; also γ/β^2 invariant.
$z'' + \hat{\delta} z' + \hat{\alpha} z + \hat{\beta} z^2 = 0$ $(\delta, \alpha, \beta) = (0.18, 1.6, 0.8)$	Origin: $z \simeq \lambda q$, $\hat{\alpha} \simeq \alpha$, $\hat{\beta} \simeq \beta/\lambda$. Shifted: $z \simeq \lambda(q + \alpha/\beta)$, $\hat{\alpha} \simeq -\alpha$.

M.1.3 Model and training

LTI, cubic, quintic and quadratic use per-frame CNNs with three convolutional blocks (base width 32), GroupNorm, GELU, a 128-unit MLP and a scalar output. Inputs are 64×64; cubic frames use Lanczos resizing. The encoder and dynamical coefficients are optimized jointly. The $\mathcal{L}_{\text{var}}$ penalty introduced in Section 2 uses the standard-deviation threshold $s_{\text{floor}} = 0.02$ (Eq. 22). Table 11 lists the optimization settings for all six families. The pendulum and Van der Pol scalar output biases are initialized in $[-3\pi, 3\pi]$ and $[-0.5, 0.5]$, respectively.

Table 11: **Training settings for the six-family experiment.** All use Adam and the final checkpoint after the stated number of updates (optimizer steps). Learning rates are encoder / dynamical law. Each family has 100 fits.

Family	IC collections	Seeds per set	Updates	$\mathrm{lr}_{\mathrm{enc}}/\mathrm{lr}_{\mathrm{ode}}$
Affine LTI	10	10	2,000	$0.0005/0.003$
Pendulum	10	10	2,500	$0.001/0.005$
Van der Pol	10	10	5,000	$0.001/0.003$
Cubic Duffing	10	10	8,000	$0.0005/0.1$
Quintic Duffing	10	10	8,000	$0.0005/0.003$
Quadratic / Helmholtz	10	10	6,000	$0.0005/0.003$

M.1.4 Evaluation

After training, freeze the encoder and fitted law. Estimate the family-permitted map $\hat{f}$ from training reference pairs (q, z): LTI uses an affine least-squares map, the pendulum selects a sign and 2π shift, and Van der Pol selects a sign. Cubic and quintic use $\lambda = \sum qz / \sum q^2$. Quadratic fits both permitted branches on training data and freezes the branch and scale with the smaller coordinate error (Table 10). No test states are used to select or refit this map.

On test clips, set $x = q$ and $\widetilde{x} = \hat{f}^{-1}(z)$ in Eq. (55): compute coordinate NRMSE within each clip and take the clip median to obtain one e_{map} per fit. For parameter evaluation, transform the physical law with the same frozen map and branch to obtain θ^{ref}. Equation (56) then gives e_θ as the largest relative error over the listed nonzero dynamical coefficients, retaining coefficient signs and excluding the LTI equilibrium. Thus the coordinate and parameter checks test the same predicted relation, rather than choosing a scale from coefficient ratios. The 100 per-fit errors are summarized using the conventions in Appendix L; the acceleration comparison below uses e_{dyn} from Eq. (57).

M.1.5 Results and discussion

Figure 1 shows the learned coefficients and coordinate maps. Table 12 summarizes the 100 fits per family; Figures 4 and 5 separate these fits by IC collection, showing the ten optimizer seeds within each collection.

Table 12: **Parameter relative error and coordinate recovery.** Percentages are median [Q25, Q75] over all 100 fits per family.

Family (100 fits)	Parameter e_θ (%)	Coordinate e_{map} (%)
Affine LTI	0.85 [0.52, 1.65]	1.68 [1.53, 1.90]
Pendulum	1.44 [0.84, 2.17]	3.58 [2.27, 5.99]
Van der Pol	0.31 [0.13, 0.61]	2.45 [1.69, 3.71]
Cubic	2.56 [1.60, 4.60]	0.82 [0.68, 1.08]
Quintic	1.68 [1.02, 2.29]	0.78 [0.62, 1.06]
Quadratic	2.47 [1.31, 4.37]	3.29 [1.96, 6.08]

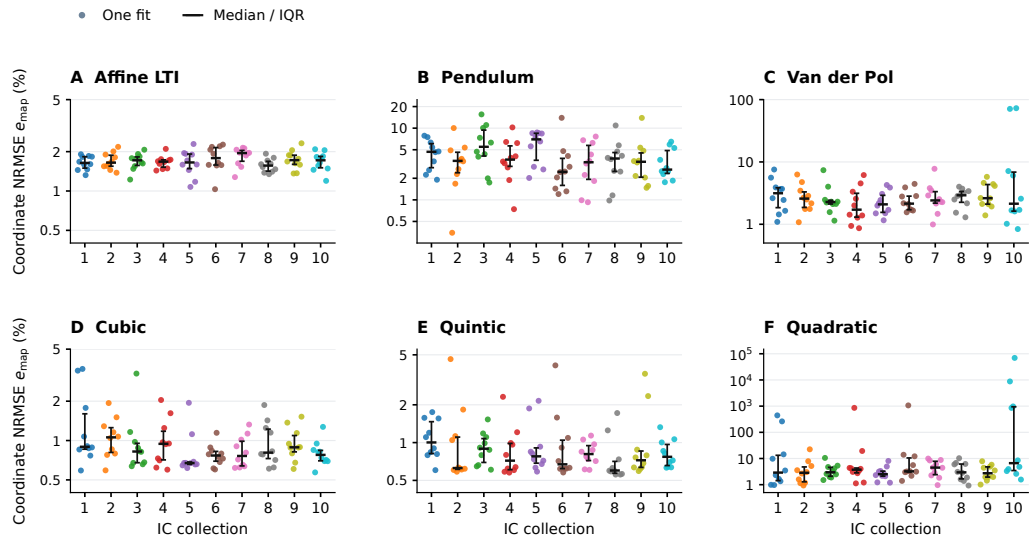


Figure 4: **Test coordinate NRMSE e_{map} by IC collection.** Each family contains ten IC collections, each with ten optimizer seeds. Dots show e_{map}; black bars give each collection's median and interquartile range. Colors match Figure 1; IC indices are local to each family. The logarithmic vertical limits vary by family.

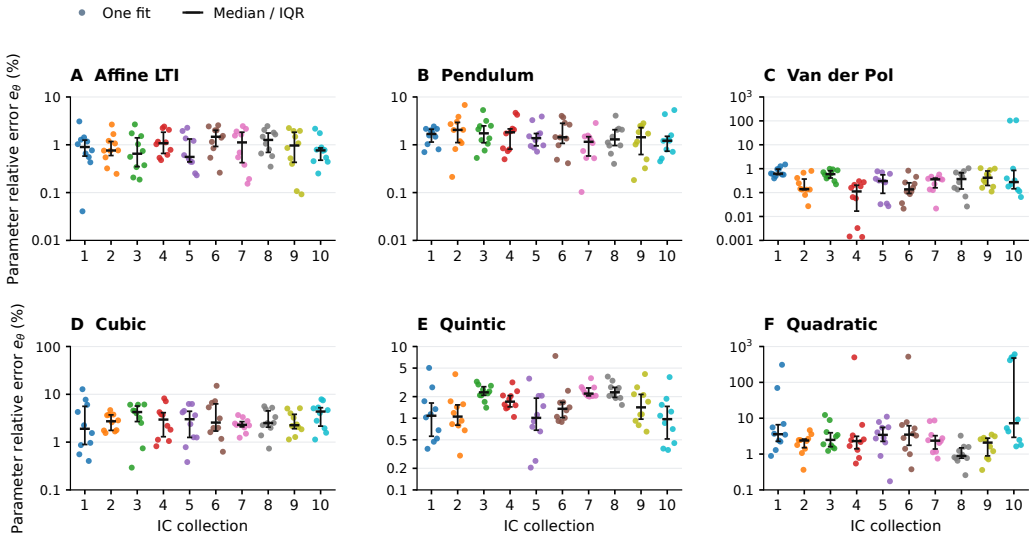


Figure 5: **Parameter relative error e_θ by IC collection.** The same fits and IC groups as in Figure 4, now showing e_θ. Within each group, horizontal offsets follow optimizer-seed order and are identical in both figures. Bars show median and interquartile range.

Median parameter relative errors range from 0.31% to 2.56%. Cubic and quintic combine sub-1% median coordinate NRMSE with the predicted β/γ scaling in Figure 1. Pendulum and Van der Pol have median coordinate errors of $3.58\%/2.45\%$ and parameter errors of $1.44\%/0.31\%$. The IC breakdown shows both variation across trajectory collections and the spread across optimizer seeds within each collection. These comparisons assess finite-data agreement with the family-specific relations. Proposition 3.3 constructs exact alternative encoders under representability and affine-closure assumptions; the empirical errors do not certify these assumptions or every continuous-time premise (Appendix D).

M.1.6 ADDITIONAL ANALYSIS: MATCHED AFFINE-LTI COMPARISON

This experiment probes the scope of affine-LTI modeling and the loss in dynamical accuracy caused by ODE-family misspecification. We keep the observations and ICs unchanged, train the encoder jointly with an affine-LTI law, and then evaluate the fitted coefficients on trajectories from the held-out test videos. The affine-LTI dataset provides a matched-family control, while the five nonlinear datasets test the cost of restricting the learned dynamics to this family. For each of the six families, we pair the 100 fitted models with 100 affine-LTI controls, $z'' + \hat{\delta} z' + \hat{\alpha}(z - \hat{c}) = 0$. Each pair shares the video data, encoder architecture and initial weights, training-window sampling, optimizer settings, and update budget (Table 11); the LTI equilibrium $\hat{c}$ is learned jointly. Pairing covers ten IC collections and ten optimizer seeds per family.

For each model, fit $z \simeq \lambda q + \tau$ on training frames and freeze it on test trajectories. Evaluate e_{dyn} using the stored physical states and derivatives, with

$$\widetilde{F} = \lambda^{-1} \hat{F}(\lambda q + \tau, \lambda q'), \qquad F = q''.$$

Here $\hat{F}$ uses the fitted coefficients without refitting on test data. Equation (57) measures how accurately the learned dynamical law reproduces physical acceleration along the held-out trajectories, including the effect of family misspecification; it does not integrate a rollout or compare coefficients from different families. The LTI oracle fits acceleration directly to $(q, q', 1)$ on training trajectories with the matching coefficient-sign constraints and uses the same test comparison. Wins count paired fits with lower e_{dyn} for the declared family model.

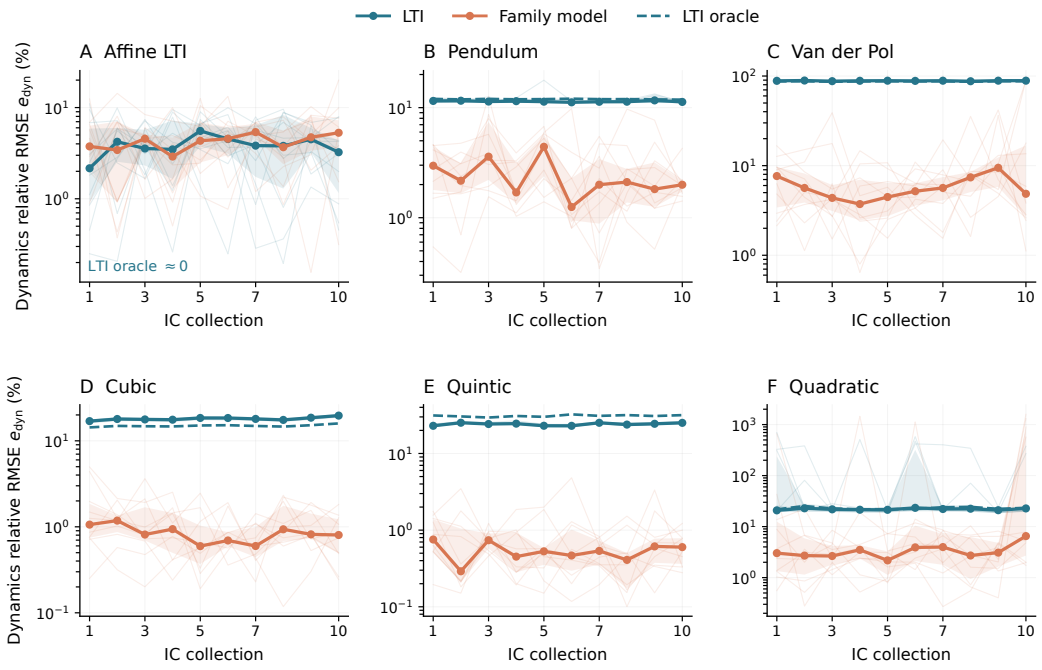


Figure 6: **Matched affine-LTI comparison across IC collections.** Each family has ten IC collections and ten paired optimizer seeds per collection. Faint lines show individual seeds; solid curves and bands give the within-IC median and interquartile range of test dynamics relative RMSE e_{dyn} (%), defined in Eq. (57) and evaluated on physical accelerations. Dashed curves show the training-fitted physical-state LTI oracle. Both methods in A use the affine-LTI family.

Table 13: **Test dynamics relative RMSE e_{dyn} (%).** Physical-acceleration comparison using Eq. (57); median [Q25, Q75] over 100 pairs per family. Wins count pairs with a smaller e_{dyn} for the declared family model.

Dataset	Affine LTI (%)	Family model (%)	Family wins
Affine LTI	3.86 [1.96, 6.58]	4.56 [2.16, 6.84]	46/100
Pendulum	11.34 [11.20, 11.70]	2.17 [1.43, 3.89]	99/100
Van der Pol	88.48 [88.01, 88.79]	5.80 [3.07, 9.21]	98/100
Cubic	17.92 [17.46, 18.53]	0.81 [0.57, 1.39]	100/100
Quintic	24.29 [23.46, 25.02]	0.54 [0.33, 0.89]	100/100
Quadratic	22.03 [20.42, 23.85]	3.15 [1.70, 6.68]	92/100

Across the five nonlinear datasets, the declared family has lower median dynamics error e_{dyn} in all ten IC collections and wins 92–100 of the 100 individual pairs. On the affine-LTI dataset, both columns instantiate the same model family and have comparable error distributions; their small difference is due to finite-precision divergence from a different parameter traversal order in global gradient clipping. A controlled replay with a common accumulation order reproduces matching training trajectories.

M.2 Velocity coverage and parameter recovery

M.2.1 Observations and data

The physical system is $q'' = -kq - \delta q' - \kappa q'|q'|$, with spring coefficient $k = 0.8$, linear damping $\delta = 0.18$, and quadratic drag $\kappa = 0.22$. Each IC collection contains 24 training and eight test clips of 1 s, rendered at 96×96 pixels and 60 Hz (61 frames).

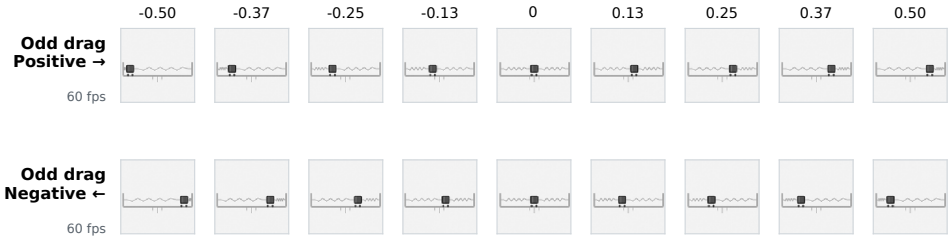


Figure 7: Opposite odd-drag crossings. Both rows show 96×96 inputs at 60 Hz, aligned at the same position at $t = 0$. The central frames have the same geometry despite opposite velocities; labels give time in seconds.

We vary the number K of speed magnitudes and whether trajectories cross $q = 0$ in one or both directions. As in Appendix M.1, an *IC collection* groups videos generated from one prescribed set of physical trajectory conditions. Here the conditions are the crossing velocities at $q = 0$, which determine the simulated trajectories. Ten data seeds independently perturb the base magnitudes $\{0.8, 1.35, 1.95\}$ by factors in $[0.97, 1.03]$. One-way $K = 1$ uses the middle magnitude; $K = 3$ uses all three; $K = 6$ uses six values spanning the same envelope. Two-way $K = 1$ uses both signs of the middle magnitude; $K = 2$ uses both signs of the smallest and largest magnitudes; $K = 3$ uses all three. Appearance repetitions give 24 training clips in every condition. The test pools use crossing speeds near $\{1.1, 1.7\}$ for one-way designs and $\{-1.1, 1.1\}$ for two-way designs. Evaluation below uses the four positive-direction test clips near 1.1 shared by all six conditions within each IC collection.

M.2.2 Theoretical predictions

The experiment tests the curvature-column exclusion in Theorem 4.1. A nonlinear coordinate map contributes $f''(q)v^2$ to the acceleration. Although $v|v|$ differs from v^2 over two-way motion, these columns coincide for positive velocities (and differ only by sign for negative velocities). One-way

motion therefore cannot separate drag from coordinate curvature, regardless of the number of speeds. With two directions and two distinct speed magnitudes, the augmented feature matrix $[1, v, v|v|, v^2]$ has rank four, giving the theorem's raw-affine conclusion under its remaining assumptions. The corresponding drag relation is $c_d = -\kappa/|\lambda|$ (Eq. (58)). The test is whether this change in coverage improves parameter recovery even when coordinate errors are small.

M.2.3 Model and training

Each of the ten IC collections is crossed with five optimizer seeds, giving 50 fits per condition. A spatial soft-argmax encoder is followed by a learned cubic polynomial head. It is jointly fitted with $z'' = c_0(z) + c_1(z)z' + c_d(z)z'|z'|$, where the polynomial degrees of (c_0, c_1, c_d) are $(5, 3, 5)$. Derivatives use a 31-frame, degree-two local-polynomial fit. Table 14 gives the optimization settings; evaluation uses the final checkpoint.

Table 14: **Coverage training settings.** Adam; each coverage condition uses ten IC collections and five optimizer seeds per collection.

IC collections	Seeds per collection	Updates	$\mathrm{lr}_{\mathrm{enc}}/\mathrm{lr}_{\mathrm{ode}}$
10	5	5,000	$0.0005/0.002$

M.2.4 Evaluation

Fit $z \simeq \lambda q + \tau$ on training references, then use $x = q$ and $\widetilde{x} = (z - \tau)/\lambda$ in Eq. (55) to compute e_{map}. Each clip contributes its frames with $q \in [-0.2, 0.2]$; all six conditions use the same four test clips within each IC collection.

Substitution of $z = \lambda q + \tau$ into the physical law gives

$$c_0(z) = -k(z - \tau), \qquad c_1(z) = -\delta, \qquad c_d(z) = -\frac{\kappa}{|\lambda|}. \tag{58}$$

Let $\bar{\kappa} = -\operatorname{mean} c_d$ be the mean learned drag magnitude, evaluated on 101 equally spaced learned-coordinate values over the encoded training interval $q \in [-0.2, 0.2]$. The parameter comparison uses the scale-corrected mean $\widehat{\kappa}_{\mathrm{aligned}} = |\lambda|\bar{\kappa}$ against $\kappa = 0.22$ in Eq. (56), with the same training-fitted λ as the coordinate comparison. Figure 8 shows both errors for all six designs.

The feature rank concerns $[1, v, v|v|, v^2]$ (Section 4.1). It is evaluated at 21 shared positions in $[-0.2, 0.2]$, using the distinct velocities at each position; Figure 8A reports the minimum rank. Rank four separates the declared features from the curvature column. A smaller rank fails this sufficient condition; it does not by itself establish nonidentifiability of every parameter.

M.2.5 Results and discussion

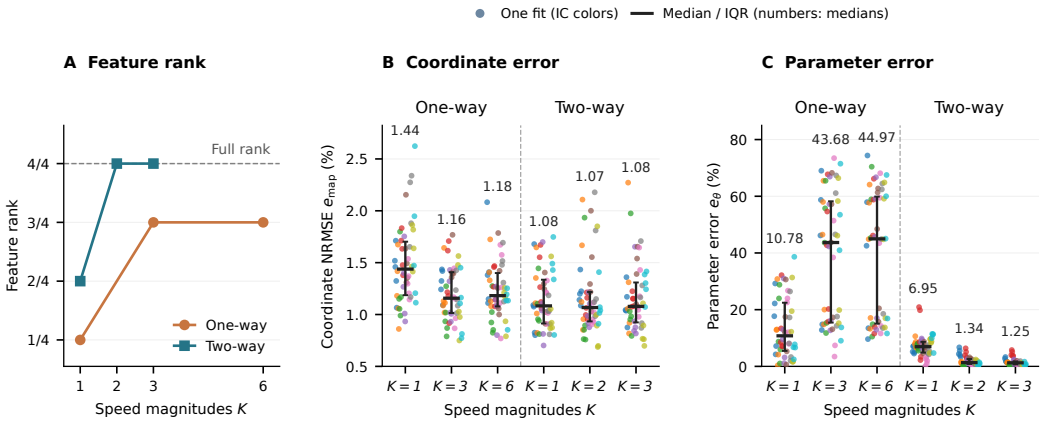


Figure 8: **Velocity coverage, parameter recovery, and coordinate accuracy.** A: minimum rank of $[1, v, v|v|, v^2]$ over 21 shared positions in $[-0.2, 0.2]$; all ten IC collections give the displayed rank for each condition. The dashed line marks full rank. B: coordinate NRMSE e_{map} (Eq. (55)) on the same four test clips within each IC collection. C: parameter relative error e_θ of the scale-corrected mean drag coefficient (Eq. (56)). B/C show all 50 fits per condition, colored by IC collection, with matching horizontal offsets. Black bars show the median and interquartile range over the 50 fits, using the convention of Figures 4 and 5; numbers give medians (%). All axes are linear. K counts speed magnitudes.

One-way motion gives $v|v| = v^2$, so adding speeds does not separate the drag feature from the curvature term introduced by a nonlinear coordinate map. Median drag-parameter errors are 43.68% and 44.97% for one-way $K = 3, 6$, compared with 1.34% and 1.25% for two-way $K = 2, 3$, which have full feature rank. Median coordinate NRMSE is 1.07–1.44% across all six conditions. Accurate coordinates therefore coexist with substantial errors in the recovered mean drag coefficient when velocity coverage is insufficient.

M.3 Canonical normalization

M.3.1 Observations and data

To test whether canonical normalization corrects nonlinear coordinate distortion, we simulate a pendulum with angular dynamics $q'' + 0.15q' + 4 \sin q = 0$, released from rest at different initial angles. An orthographic camera looks along the vertical axis and records $y = \sin q$ on $|q| < \pi/2$, discarding the depth coordinate (Figure 9). The projected rod and bob are rendered as 96×96 grayscale frames. Thus the observation depends nonlinearly on the physical angle even before learning.

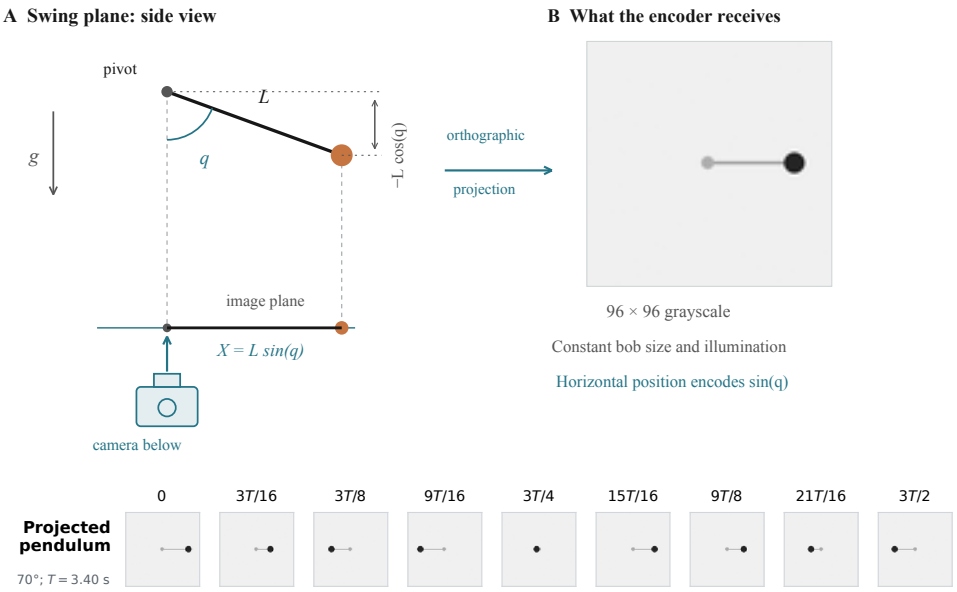


Figure 9: **Projected pendulum observations.** An orthographic camera looks along the vertical axis, retaining $X = L \sin q$ and discarding the depth coordinate $-L \cos q$. The projected rod and constant-size bob produce the image coordinate $y = \sin q$. Bottom: nine 96×96 observations from 0 to $3T/2$ for a $70°$ release; T is the first full oscillation duration. The illustrative strip continues the original ODE and renderer beyond the 4.8 s training-clip duration.

We train on three videos released at approximately $45°, 60°, 70°$, and test on held-out releases near $40°, 55°, 65°$. Each training angle has one rendering; each test angle has four, giving 3 training and 12 test videos. All videos last 4.8 s at 30 Hz (145 frames). We repeat training from three random initializations on the same video data. For each fitted model, Raw and Canonical are compared on exactly the same test videos.

M.3.2 Theoretical predictions

Differentiating $y = \sin q$ introduces the squared-velocity coefficient $-y/(1-y^2)$ into the projected law. We fit the corresponding structured family in a learned coordinate with free scale $s > 0$ and offset b:

$$z'' = -\delta z' - \omega^2 (z-b)\sqrt{1-u^2} - \rho \frac{z-b}{s^2-(z-b)^2}(z')^2, \qquad u = \frac{z-b}{s}. \tag{59}$$

Here ρ controls the squared-velocity channel. The exact projected law $z = s \sin q + b$ has $(\delta, \omega^2, \rho) = (0.15, 4, 1)$. For optimization, the fitted-law parameters are initialized at $(\delta, \omega^2, \rho, s, b) = (0.12, 3.2, 0.8, 0.82, 0)$ in all configurations.

This experiment tests Section 4.2: a squared-velocity channel permits nonlinear coordinate changes, so the raw coordinate need not be affinely related to the physical angle. Affine alignment alone therefore cannot in general remove this distortion. We consequently apply the normalizer in Eq. (60), which removes the squared-velocity channel and becomes $s \arcsin((z-b)/s)$ when $\rho = 1$.

The physical reference is the projected law induced by $y = \sin q$, whose normalizer gives $\psi_\theta(y) = \arcsin y = q$. Theorem 4.2 therefore predicts a shared affine relation $\psi_\eta(z) \simeq \lambda q + \tau$ under its coverage assumptions, which we test through canonical coordinate recovery and the transformed restoring force and damping. The theorem applies on the regular chart $|z-b| < s$ under the sufficient coverage condition of at least three distinct physical velocities at each covered position for this degree-two velocity family.

M.3.3 MODEL AND TRAINING

We jointly fit the spatial soft-argmax encoder and the law in Eq. (59), without physical-coordinate targets. Each fit uses 6,000 Adam updates and four 97-frame windows per step; learning rates are 2×10^{-4} for the encoder and 2×10^{-3} for the law. Derivatives use 21-frame, degree-four local polynomials. The standard-deviation threshold is $s_{\mathrm{floor}} = 0.03$ with the weight $\lambda_{\mathrm{var}} = 100$.

M.3.4 EVALUATION

For the main training trajectories, a numerical first-passage check on 161 positions in $[-0.4, 0.4]$ rad finds three distinct velocities and rank three for $[1, v, v^2]$. The full-test coordinate evaluation and wider force grid also examine behavior beyond this checked interval.

After freezing the encoder and law, solve $\psi'' + c_2\psi' = 0$ with $\psi(b) = 0$ and $\psi'(b) = 1$:

$$\psi(z) = s \int_0^{(z-b)/s} (1 - u^2)^{-\rho/2}\, du. \tag{60}$$

This map uses the learned (ρ, s, b) and requires no physical anchors. Cumulative trapezoidal quadrature uses 40,001 nodes spanning the encoded training/test values and b. A defined map requires $s^2 - (z - b)^2 > 0$ throughout that interval; all three main fits pass this check. For $h(z) = z$ and $h(z) = \psi(z)$ separately, fit $h(z) \simeq a_h q + b_h$ on training references, then freeze $\widetilde{q} = [h(z) - b_h]/a_h$ for testing. Both readouts have identical affine freedom. To assess coordinate accuracy at release angles unseen in training, we evaluate coordinate NRMSE e_{map} (Appendix L) on held-out videos near $40^\circ, 55^\circ, 65^\circ$. Compute it within each test clip using the original test frames, then take the clip median. Binning is used only to display residuals.

Report coefficient relative errors against $(\delta, \omega^2, \rho) = (0.15, 4, 1)$. Let $A(z)$ be the velocity-independent term of Eq. (59). With the frozen canonical readout, evaluate

$$z_q = \psi^{-1}(a_\psi q + b_\psi), \qquad \widetilde{F}(q) = \psi'(z_q)A(z_q)/a_\psi.$$

Compare $\widetilde{F}$ with $-4 \sin q$ on 361 points in $[-0.9, 0.9]$ rad, using the common dynamics relative RMSE e_{dyn} from Appendix L. This measures the restoring-force channel. The damping channel remains $-\delta$, so its relative error is the coefficient error of δ. All errors are percentages, summarized as median [Q25,Q75] across fits.

M.3.5 RESULTS AND DISCUSSION

Table 3 shows that normalization reduces coordinate NRMSE e_{map} from 5.633% to 0.305% and recovers the restoring force to 0.193% relative RMSE. Figure 10 shows the corrected residuals and their variation across fits; Table 15 reports the corresponding coefficient errors.

Table 15: **Fitted-law coefficient errors.** Parameter relative errors $e_{\theta,j}$ against $(\delta, \omega^2, \rho) = (0.15, 4, 1)$; median [Q25,Q75] over three fits on the same IC collection.

Damping $e_{\theta,\delta}$	Restoring coefficient e_{θ,ω^2}	Quadratic channel $e_{\theta,\rho}$
0.64% [0.53%, 0.68%]	0.21% [0.20%, 0.23%]	1.93% [1.83%, 1.95%]

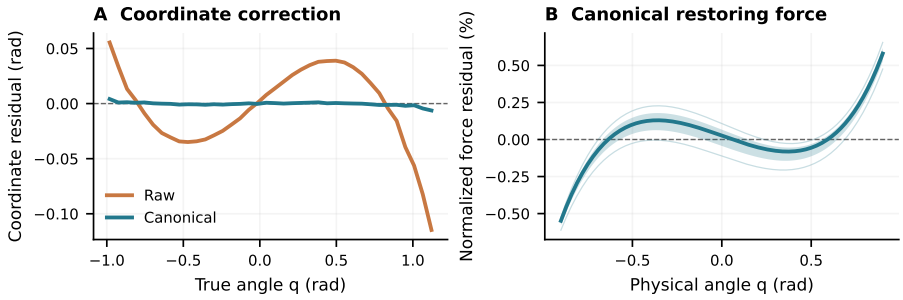


Figure 10: **Coordinate and force recovery across three fits.** A: held-out angle residuals after training-fitted affine alignment. B: canonical force residual $(\widetilde{F} - F)/\mathrm{RMS}(F)$ in percent, with $F = -4\sin q$. Thick curves: pointwise medians; shading: Q25–Q75; thin curves: individual fits. Coordinates are binned for display only; e_{map} uses original test frames.

M.3.6 Additional analysis: trajectory coverage and repeated observations

Table 16 compares one $60°$ video with one to three passages, and first-passage videos at $50°, 60°, 70°$ with one or two renderings per angle. Each condition uses 6,000 updates and five paired initializations. Held-out releases at $45°, 57.5°, 75°$ are evaluated on $U = [-0.4, 0.4]$ rad for both e_{map} and force e_{dyn}. Additional renderings vary noise and brightness; the trajectory and training references for affine calibration remain fixed.

Table 16: **Trajectory coverage and repeated observations.** Errors are median [Q25,Q75] over five fits per row; all 25 maps are defined. Three-angle rows use one passage per video. Velocity counts give the number of distinct velocities per state on $U = [-0.4, 0.4]$ rad. $\times 2$ denotes two renderings per angle with unchanged trajectories.

Training observations	Videos	Distinct velocities	Raw e_{map} (%)	Canonical e_{map} (%)	Force e_{dyn} (%)
$60°$; 1 passage	1	1	3.137 [3.092,3.140]	0.898 [0.886,0.899]	19.21 [19.19,20.02]
$60°$; 2 passages	1	2	8.441 [8.434,8.725]	0.823 [0.560,1.192]	3.49 [3.38,4.61]
$60°$; 3 passages	1	3	8.129 [8.114,8.196]	0.691 [0.657,0.719]	2.65 [1.91,3.48]
$50°, 60°, 70°$	3	3	2.853 [2.839,2.943]	0.748 [0.672,0.762]	9.55 [8.59,10.05]
$(50°, 60°, 70°) \times 2$	6	3	3.007 [2.956,3.038]	0.539 [0.527,0.546]	3.83 [3.22,5.43]

Extending the single video from one to two and three passages reduces force error from 19.21% to 3.49% and 2.65%, respectively. Three first-passage videos at $50°, 60°, 70°$ still give 9.55% force error. Doubling the renderings from three to six videos reduces this error to 3.83% with unchanged velocity coverage. Thus, three distinct velocities alone do not guarantee accurate finite-data recovery; repeated observations also affect force accuracy. The passage comparison additionally changes duration and state coverage.

N Real-video methods and additional results

We study three real-video experiments: pendulum recordings from the IRIS benchmark (Khanbayov et al., 2026), and side-view and overhead free-fall recordings collected by us. Figure 11 compares the falling-object settings in IRIS with our recordings. Although IRIS also provides similar side-view and overhead free-fall videos, these collections do not strictly match the shared-observation setting of our theory. The IRIS side-view settings use balls of different sizes and appearances, and their camera framing also differs. In the overhead recordings, the viewpoint and framing change across release-height settings. Consequently, pooling these settings changes the physical-state-to-image relation as well as the trajectories: the same physical position or height need not produce the same visual observation. Applying our shared-map analysis to such a collection would require additional treatment of the object and camera differences. For the two fall experiments, we therefore record

multiple releases while keeping the ball and camera configuration fixed within each view and varying the release height.

Real recordings still depart from ideal observations through finite frame rates, image noise, motion blur, and imperfect projection and calibration. The three cases examine whether the parameter information characterized by our theory, combined with physical anchors, supports accurate parameter estimates under these conditions. Appendix N.1 studies pendulum invariants and length recovery, including a matched LTI comparison; Appendix N.2 studies the affine acceleration orbit and diameter calibration in side-view fall; and Appendix N.3 studies canonical normalization and height calibration in overhead fall. Each case presents observations and data, theoretical predictions, model and training, evaluation, and results and physical calibration, in that order. Additional analyses follow the main results where applicable.

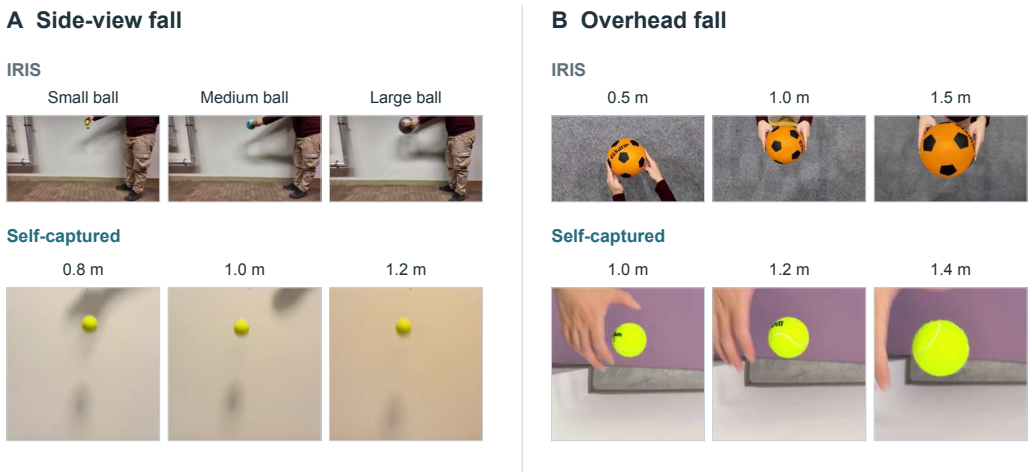


Figure 11: **Falling-object recording settings in IRIS and our collection.** IRIS examples are from Khanbayov et al. (2026). Left: side view; right: overhead view. Each half contains three IRIS settings above three self-captured release heights, with one recorded frame per setting. IRIS side-view examples vary the ball and framing; the overhead examples show different viewpoints and framing. Our examples use the same ball and camera configuration within each view. Images preserve aspect ratio; self-captured images use a common fixed crop within each view. These are acquisition examples, not matched physical states or performance comparisons.

N.1 Pendulum: invariant coefficients and length calibration

N.1.1 Observations and data

We use 30 IRIS recordings: ten repeated trials at each release angle $20°$, $45°$, and $90°$, with reference pendulum length $L = 0.50$ m (Khanbayov et al., 2026). Matching recording indices 01–05 across angles define five training groups, each containing one video at each angle. A model trains on one three-video group and is evaluated on the other 27 recordings. Native indices $60, 62, \ldots, 538$ provide 240 frames at 30 Hz per video; encoder inputs are grayscale images resized to 96×54. Figure 12 illustrates the three observation settings.

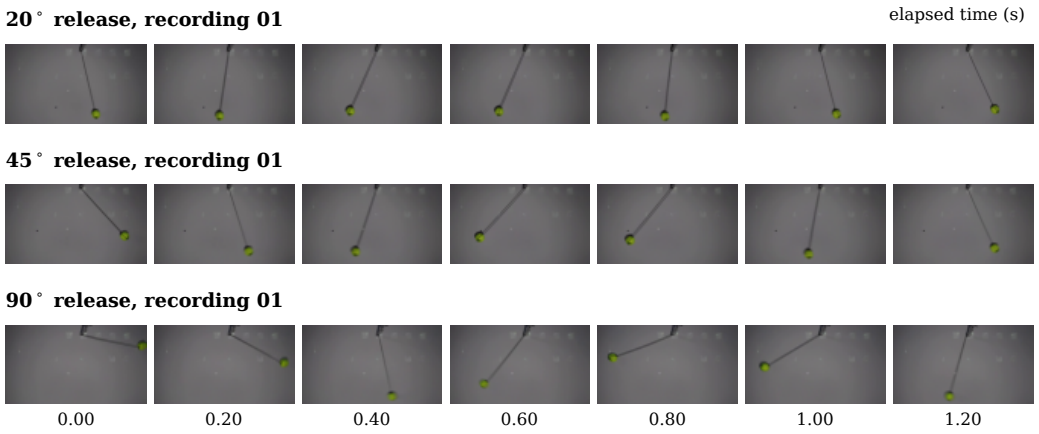


Figure 12: **Pendulum observations on IRIS** (Khanbayov et al., 2026). Each row shows recording 01 from one release-angle setting. The displayed RGB frames illustrate the observed motion; elapsed time is measured from the first selected frame.

N.1.2 Theoretical predictions

For physical angle q, the damped pendulum obeys

$$q'' + \delta q' + \kappa \sin q = 0, \qquad \kappa = g/L. \tag{61}$$

Under the shared-state and coverage assumptions of Theorem 3.1, the compatibility identities in (6) restrict the angular coordinate to $z = sq + 2\pi n$, with $s \in \{-1, +1\}$ and $n \in \mathbb{Z}$. Both δ and κ are invariant under this coordinate orbit (Table 1). Thus the theoretical prediction is that compatible fits preserve these coefficients despite sign/period ambiguity, and that supplying gravity determines length through $L = g/\kappa$. For small angles, $\sin q \simeq q$ gives an LTI approximation; the approximation becomes inaccurate at the larger angles in this collection. We therefore also train a matched LTI model to examine the effect of ODE-family misspecification on parameter estimates from real videos.

N.1.3 Model and training

A shared per-frame CNN with group normalization and an unbounded scalar head produces z. We jointly fit the encoder and coefficients of $z'' + \delta z' + \kappa \sin z = 0$; the matched LTI model replaces $\kappa \sin z$ by αz. Each training group has five paired nonlinear/LTI fits, giving 25 fits per model. Each pair starts from the same network initialization and uses the same training triplet. Damping is nonnegative and restoring coefficients are positive. Both models use 8,000 Adam steps, encoder/ODE learning rates $10^{-3}/10^{-2}$, and gradient clipping at 1. The objective uses local standard-deviation normalization and a non-collapse penalty with standard-deviation floor $s_{\mathrm{floor}} = 0.10$ and weight $\lambda_{\mathrm{var}} = 100$.

N.1.4 Evaluation

We evaluate coordinate NRMSE e_{map} from (55) under two alignment rules, each fitted on the three training recordings and frozen on the other 27 videos. The reference q is the optically measured angle. For the *affine* comparison, fit $z \simeq \lambda q + \tau$ and use $\widetilde{q} = (z - \tau)/\lambda$. This gives both models the same scale and offset freedom and measures coordinate transfer after those corrections. For the *sign/period* comparison, choose $s \in \{-1, +1\}$ and $n \in \mathbb{Z}$ using $z \simeq sq + 2\pi n$, and use $\widetilde{q} = (z - 2\pi n)/s$. This stricter rule tests angular recovery within the nonlinear family's theoretical coordinate orbit, without freely correcting scale or offset. For LTI, it is a fixed-angle diagnostic because LTI permits arbitrary coordinate scale. Reporting both rules separates agreement with the physical angular scale from transfer under a common affine readout.

N.1.5 Results and physical calibration

The nonlinear model has lower median test coordinate NRMSE under both rules (Table 17): 5.04% versus 7.13% after affine alignment, and 13.22% versus 39.04% after sign/period alignment. The affine comparison favors the nonlinear model in 20 of the 25 paired fits. Figure 13A–B shows the individual fits and median/IQR summaries within each training group. The remaining sign/period error shows that the nonlinear encoder only approximately recovers the theoretical angular gauge; its smaller affine error alone would not establish recovery of the physical angular scale.

Table 17: **Coordinate accuracy under two common alignment rules.** Coordinate NRMSE e_{map} from (55) (%), median [Q25, Q75] over 25 paired fits per model. Both alignments are fitted on training recordings and evaluated on the same 27 held-out recordings per pair.

Model	Affine: $z \simeq \lambda q + \tau$	Sign/period: $z \simeq sq + 2\pi n$
Nonlinear pendulum	5.04 [4.38, 5.83]	13.22 [12.20, 13.97]
LTI	7.13 [5.46, 7.45]	39.04 [28.65, 53.30]

The learned nonlinear restoring coefficient is $\hat{\kappa} = 19.050 \pm 0.044\,\mathrm{s}^{-2}$, while LTI gives $\hat{\alpha} = 17.376 \pm 0.015\,\mathrm{s}^{-2}$; the corresponding damping estimates are 0.07805 ± 0.00114 and $0.08270 \pm 0.00240\,\mathrm{s}^{-1}$. Here $\kappa = g/L$ in (61); it does not determine g and L separately.

Anchor. Using the gravity anchor $g = 9.81\,\mathrm{m/s}^2$ after fitting gives

$$\hat{L} = \frac{9.81}{\hat{\kappa}}, \qquad \hat{L}_{\mathrm{eff}} = \frac{9.81}{\hat{\alpha}}. \tag{62}$$

The nonlinear estimate is 0.5150 ± 0.0012 m, compared with 0.5646 ± 0.0005 m for the LTI effective length. Against the IRIS reference $L = 0.50$ m, median parameter relative errors e_θ from (56) are 2.92% [2.82, 3.17] and 12.92% [12.83, 12.97], respectively. Figure 13C displays the calibrated lengths and their mean/SD by training group, with the reference length dashed. The more accurate nonlinear length estimates show why retaining the nonlinear restoring law matters when calibrating large-angle motion: the same gravity anchor cannot compensate for the LTI approximation.

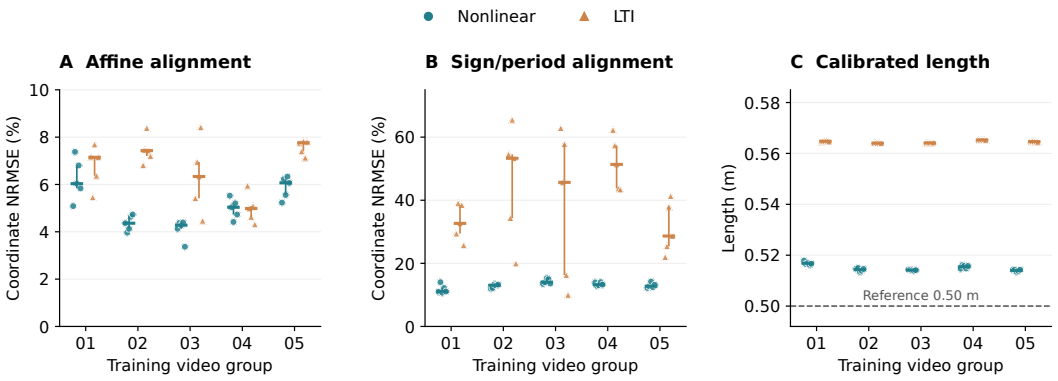


Figure 13: **Coordinate recovery and calibrated length by training group.** Groups 01–05 each contain five nonlinear fits (teal circles) and five matched LTI fits (orange triangles). A–B: test coordinate NRMSE e_{map} from (55) after affine and sign/period alignment, respectively; bars show median and IQR. C: length after gravity calibration in (62); bars show mean $\pm$ sample SD, and the dashed line marks the IRIS reference 0.50 m. All vertical axes are linear.

N.2 Side-view free fall: affine ambiguity and acceleration calibration

N.2.1 Observations and data

We record a falling ball with the same camera position, camera settings, and ball across all release heights. A fixed native crop $(x, y, w, h) = (390, 760, 300, 800)$ is resized to 72×192; each clip

contains 5–11 frames at approximately 60 Hz, with recorded timestamps retained. The camera captures a fixed vertical interval, so the observed windows need not include the release instant. Figure 14 shows three example windows and the motion trails present in these recordings.

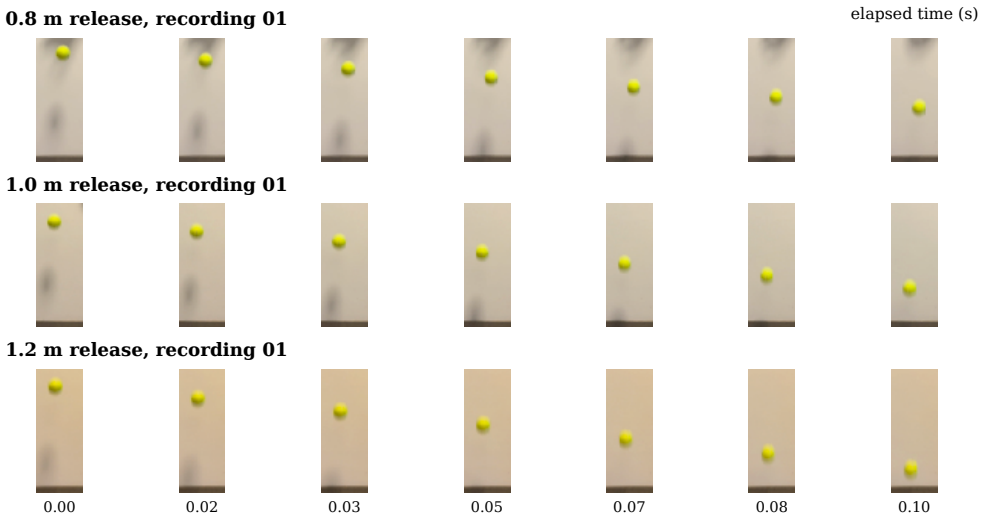


Figure 14: **Side-view observations.** Three release heights (0.8, 1.0, and 1.2 m), one row each. The fixed camera captures an intermediate vertical interval; the release instant is unavailable in these windows. The first seven frames of recording 01 are shown from each selected window; shared elapsed times are relative to the start of each window. Fixed-crop RGB images are 72×192.

We form five training IC collections, IC1–IC5, using different combinations of repeated recordings. Each contains two videos at each of 0.8, 1.0, and 1.2 m, giving six training videos per collection. The same five encoder initializations are used for each collection, giving 25 fits. All collections are evaluated on the same 14 test videos, none of which appears in any training collection.

N.2.2 Theoretical predictions

For downward physical displacement q, ideal free fall obeys $q'' = g$. Under the shared-state and three-slope coverage assumptions, Theorem 3.1 restricts a compatible learned coordinate to $z = \lambda q + \tau$, with $z'' = \lambda g$. Different release heights provide different velocities through the same observed height interval. With a locally constant image scale, the theory therefore predicts the coordinate and parameter relations

$$z \simeq \lambda_{\mathrm{px}} y + \tau, \qquad A \simeq \lambda_{\mathrm{px}} a_{\mathrm{px}}, \tag{63}$$

where y is vertical pixel position and a_{px} its acceleration. Thus fitted A can vary with λ_{px} while the ratio A/λ_{px} remains approximately invariant. We test this affine orbit across fits and use an external length anchor to convert the scale-corrected acceleration to physical gravity.

N.2.3 Model and training

Four convolutions with 16/24/32/48 channels, GroupNorm and SiLU, followed by fixed spatial pooling and a 128-unit head, produce scalar z from saturation-based images. At every training step, least squares profiles clip-specific intercepts and velocities and a common acceleration A in $z'' = A$. Gradients through this profiled fit update the encoder; A has no separate optimizer learning rate. The objective uses dynamics weight 10 and standard-deviation floor $s_{\mathrm{floor}} = 0.10$. Its non-collapse term is $\lambda_{\mathrm{var}} M^{-1} \sum_m (s_{\mathrm{floor}} - s_m)_+^2$ with $\lambda_{\mathrm{var}} = 100$, where s_m is the within-clip latent standard deviation and $(u)_+ = \max(u, 0)$. Horizontal-shift and brightness consistency have weight 0.1; weak upper-scale and centering penalties have weights 0.01 and 0.001. AdamW updates the encoder for 8,000 steps at learning rate 10^{-3}, with gradient clipping at 1 and zero weight decay. Random convolutional and linear weights and biases are scaled by 0.1 at initialization; normalization parameters are unchanged.

N.2.4 EVALUATION

The post-fit reference pool contains eight videos, four each at 0.8 and 1.2 m; it can overlap training and is disjoint from the test set. On this pool, fit one global $z \simeq \lambda_{\mathrm{px}} y + \tau$ and common pixel acceleration a_{px} with clip-specific initial conditions, then freeze these quantities. Coordinate NRMSE e_{map} in (55) uses $x = y$ and $\widetilde{x} = (z - \tau)/\lambda_{\mathrm{px}}$ on the 14 held-out videos, including all seven releases at the unseen training height 1.4 m. We compute each fit's median error over these videos, then summarize across fits. Parameter error e_θ in (56) compares A/λ_{px} with a_{px} to assess the affine acceleration orbit; after diameter calibration, it compares $\hat{g}$ with $9.81\,\mathrm{m/s^2}$. The reference position y, pixel acceleration a_{px}, and calibration width d_{px} are expressed in native-image pixel units.

N.2.5 RESULTS AND PHYSICAL CALIBRATION

The fitted $(\lambda_{\mathrm{px}}, A)$ pairs follow the predicted affine orbit in (63) (Figure 16A), with median scale-corrected acceleration error 8.69% [3.68, 11.63]. The same frozen affine readout gives median test coordinate NRMSE 3.75% [3.22, 5.14]; panel B shows its variation across the five IC collections. These results support approximate agreement with the orbit while retaining measurable acceleration error.

Anchor. To set the physical length scale, the stationary-ball image in Figure 15 provides horizontal ball width $d_{\mathrm{px}} = 88$ native pixels. With diameter reference $D = 0.067$ m, the pixel-to-metre conversion gives

$$\hat{g} = \frac{A}{\lambda_{\mathrm{px}}} \frac{D}{d_{\mathrm{px}}}. \tag{64}$$

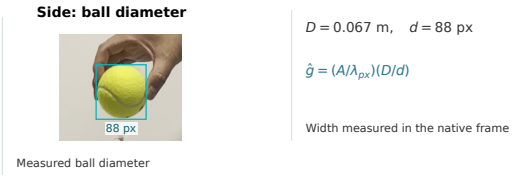


Figure 15: **Side-view physical calibration.** The measured ball width and diameter reference define the spatial conversion used to report gravity.

The calibrated gravity estimate is $9.370 \pm 0.771\,\mathrm{m/s^2}$, with parameter relative error 5.10% [3.21, 7.42] against $9.81\,\mathrm{m/s^2}$ (median [Q25, Q75]). Figure 16C reports the five gravity estimates and their mean/SD within each IC collection. The diameter anchor therefore turns the scale-corrected pixel acceleration into an estimate in physical units despite variation in the learned coordinate scale.

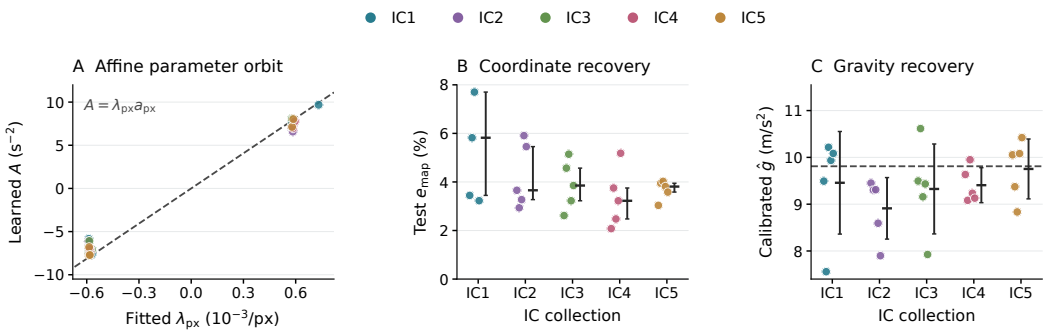


Figure 16: **Affine parameter orbit, coordinate recovery, and calibrated gravity.** Colors identify the five training IC collections, IC1–IC5; each point is one of five fits per collection. A: learned acceleration A versus coordinate scale λ_{px}; the dashed line is the theoretical relation $A = \lambda_{\mathrm{px}} a_{\mathrm{px}}$ from (63), using the independently fitted pixel-acceleration reference. B: held-out coordinate NRMSE e_{map} from (55); black bars show median and IQR. C: gravity after diameter calibration in (64); black bars show mean $\pm$ sample SD, and the dashed line marks $9.81\,\mathrm{m/s^2}$.

The small coordinate error does not guarantee an equally accurate acceleration estimate. Converting pixel acceleration to gravity using the ball diameter assumes that one pixel represents the same physical distance in the calibration image and throughout the falling trajectory. If the ball's distance from the camera changes, or its pixel diameter is measured inaccurately, the resulting gravity estimate can be biased. The IC collections and initialization scale were selected during development using this dataset. The comparisons therefore show how estimates vary with the training videos and encoder initialization within this dataset.

N.3 Overhead free fall: connection normalization and physical calibration

N.3.1 Observations and data

We fix the camera approximately 2 m above the ground, looking downward, and keep its position and camera settings unchanged throughout recording. The videos have resolution 1080×1920 in portrait orientation and are recorded at approximately 60 fps. A fixed 960×960 crop with native coordinates $(x, y, w, h) = (60, 480, 960, 960)$ is downsampled to 64×64 for the encoder. We release the same ball from heights 0.8, 1.0, 1.2, and 1.4 m above the ground and use five videos at each height, giving 20 dynamic videos in total. Figure 17 shows example observations from three release heights.

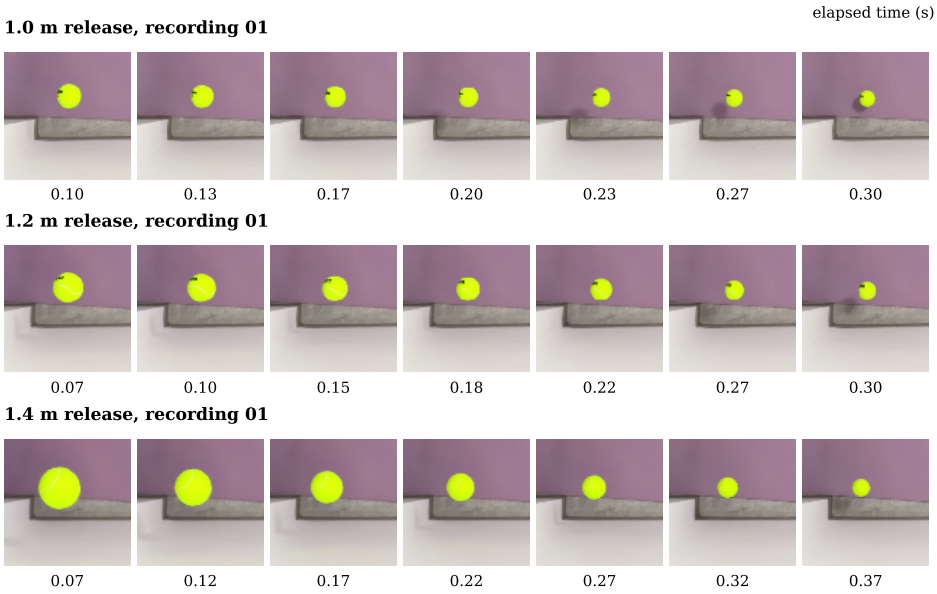


Figure 17: **Overhead observations.** Fixed-crop RGB frames at release heights 1.0, 1.2, and 1.4 m, one row each. Times are relative to the start of the corresponding original clip.

We form five training initial-condition (IC) collections, IC1–IC5, from different combinations of repeated releases at the four nominal release heights (Table 18). These nested collections contain 5, 7, 9, 11, and 14 videos, respectively, drawn from the same 14-video training pool; all share the other six videos for testing. Each collection includes at least one recording from every release height. Under ideal free fall, different release heights give distinct velocities at the same lower height; on their common height interval, this provides the three distinct velocities required by Theorem 4.2. The retained training segments contain only 13–23 frames each, and the test windows contain 7–18 frames. Because each clip provides only a short trajectory, we pool repeated recordings at the same release heights for joint training and evaluate on a separate pool of held-out recordings from those heights. This reduces reliance on any single short clip while keeping the training and test videos distinct. Recorded timestamps are retained throughout.

Table 18: **Overhead training and test recordings.** IC1–IC5 are nested training collections of repeated releases; each covers all four heights and shares the same six test videos. Recording IDs are local to each release height. Frames are counted within the training segments or test windows actually used.

Collection	0.8 m	1.0 m	1.2 m	1.4 m	Videos	Frames
IC1	01	01	01,02	01	5	95
IC2	01	01,02	01,02,03	01	7	132
IC3	01,02	01,02	01,02,03	01,02	9	166
IC4	01,02,03	01,02,03	01,02,03	01,02	11	199
IC5	01,02,03,04	01,02,03,04	01,02,03,04	01,02	14	247
Test	05	05	05	03,04,05	6	78

N.3.2 Theoretical predictions

The overhead view encodes height through the ball's apparent size. Let h be upward physical height, H the fixed camera height, and R the image radius. For a fixed ball under ideal pinhole projection, $R = C/(H - h)$, where $C > 0$ combines focal length and ball radius. Although physical free fall has constant acceleration $h'' = -g$, differentiating this nonlinear observation map gives

$$R'' = 2\frac{(R')^2}{R} - \frac{g}{C}R^2. \tag{65}$$

The squared-velocity term therefore arises from perspective projection. This motivates the learned family $z'' = \rho(z')^2/z - kz^2$, with $z > 0$; the raw latent coordinate need not be affine in physical height.

To remove this term, we use the law-dependent normalization of Section 4.2. Substituting $c_2(z) = \rho/z$ into (14), with $\psi_\rho(1) = 0$ and $\psi'_\rho(1) = 1$, gives

$$r = \psi_\rho(z) = \begin{cases} (z^{1-\rho} - 1)/(1-\rho), & \rho \neq 1, \\ \log z, & \rho = 1, \end{cases} \qquad r'' = -kz^{2-\rho}. \tag{66}$$

Indeed, $\psi'_\rho(z) = z^{-\rho}$, so the chain-rule terms involving $(z')^2$ cancel in $r'' = \psi'_\rho(z)z'' + \psi''_\rho(z)(z')^2$. The map is increasing and regular on $z > 0$. At $\rho = 2$, it becomes $\psi_2(z) = 1 - 1/z$. Applied to the ideal image radius, this gives $\psi_2(R) = h/C + 1 - H/C$, an affine function of height, with constant acceleration $-g/C$.

More generally, under the shared, non-collapsed state-map and exact compatibility assumptions, three distinct velocities at each covered height allow Theorem 4.2 to apply. Since height already has no squared-velocity channel, (16) requires $r = \lambda h + \tau$ with one affine map shared by all clips. Hence $r'' = -\lambda g$ is constant. For $k \neq 0$ and nonconstant z, the canonical law in (66) can satisfy this only when $\rho = 2$. The theory thus predicts recovery of an affine height coordinate after normalization and an invariant coefficient $\rho = 2$ in the ideal compatible setting.

The coefficient k still depends on coordinate scale. For example, rescaling the raw output to $\tilde{z} = cz$, $c > 0$, preserves ρ but changes k to k/c. In canonical coordinates, an affine change $\tilde{r} = \alpha r + \beta$ rescales acceleration by α. Normalization therefore removes the nonlinear distortion while leaving the conversion to physical units undetermined. Height anchors determine $h = ar + b_h$; at $\rho = 2$, the calibrated acceleration is $h'' = -ak$, so $g = ak$. For fitted ρ near 2, the corresponding acceleration is $-akz^{2-\rho}$, whose variation also tests departure from constant free-fall acceleration. The calibration and error checks below assess these predictions on the real recordings.

N.3.3 Model and training

We resize RGB frames to 64×64 and apply color-based preprocessing to enhance the ball's contrast against the background and reduce background interference. The encoder uses two 3×3 stride-one convolutions, each with two output channels, GroupNorm and SiLU, followed by global average pooling and a linear scalar head. Its output o is mapped to the positive latent $z = 0.05 + \mathrm{softplus}(o)$. Network initializations are generated by Gaussian perturbations of one untrained reference network: each convolutional and linear weight or bias tensor receives a random perturbation, while GroupNorm parameters retain their initial values.

The short overhead clips contain only 13–23 training frames, so numerical second derivatives can amplify frame-level noise. We therefore replace the finite-difference dynamics term in (3) with an integral residual of the same canonical law in (66). This uses the full clip at its recorded timestamps without numerically differentiating the encoder outputs. Writing $r_j = \psi_\rho(z_j)$ and integrating $r''_j = -kz_j^{2-\rho}$ twice from the start of clip j gives

$$r_j(t) = r_{j,0} + v_{j,0}t - k\int_0^t (t-s)z_j(s)^{2-\rho}\,ds. \tag{67}$$

The clip-specific $(r_{j,0}, v_{j,0})$ accommodate different initial conditions, while the encoder, ρ, and k remain shared across clips.

For smooth trajectories with $z_j > 0$, satisfying (67) for all t is equivalent to satisfying the canonical ODE, with $r_{j,0} = r_j(0)$ and $v_{j,0} = r'_j(0)$. Since ψ_ρ is invertible on this domain, this is also equivalent to the original learned ODE. The integral formulation thus preserves the continuous-time dynamics and parameter-compatibility target of Theorem 4.2. On sampled, noisy frames, the integral and finite-difference losses are different numerical objectives; a small integral loss alone does not establish the theorem's continuous-time compatibility or coverage assumptions.

Let $\varepsilon_{j,n}$ be the left-hand side minus the right-hand side of (67) at timestamp $t_{j,n}$, and let $\bar{r}_j$ be the mean encoded canonical state within clip j. The dynamics loss is

$$\mathcal{L}_{\mathrm{int}}(\phi, \rho) = \min_{k, \{r_{j,0}, v_{j,0}\}} \frac{\sum_{j,n} \varepsilon_{j,n}^2}{\sum_{j,n} \big(r_j(t_{j,n}) - \bar{r}_j\big)^2}. \tag{68}$$

The denominator expresses residual error relative to within-clip signal variation and uses a small numerical floor to prevent division by zero. We evaluate the integral using piecewise-linear interpolation of $z^{2-\rho}$ at the recorded timestamps. At every update, linear least squares computes the shared k and each clip's $(r_{j,0}, v_{j,0})$; these quantities have no separate optimizer learning rate. Adam jointly updates the encoder and ρ for 4,000 steps, with initial learning rates $\ell_{\mathrm{enc},0} = 10^{-3}$ and $\ell_{\rho,0} = 6 \times 10^{-4}$. Both rates follow

$$\ell_j(t) = \ell_{j,0}\left[0.05 + 0.475\left(1 + \cos\frac{\pi t}{8000}\right)\right], \qquad 0 \le t \le 4000, \tag{69}$$

so the final rates are 5.25×10^{-4} and 3.15×10^{-4}. Gradient norms are clipped at 20. Let $s_0 = \max(\mathrm{std}(z_0), 0.08)$, fixed from the initial training outputs. In addition to $\mathcal{L}_{\mathrm{int}}$, the objective includes the mean over clips of $[\max(0, 0.6s_0 - \mathrm{std}(z_j))/s_0]^2$ and a positivity penalty below 0.08, each with weight 10. The first discourages near-constant encodings, serving the same anti-collapse role as (22); the second adds a margin from the singular boundary $z = 0$ of the law and normalizer. These penalties support the non-collapse and domain conditions used by the theory without supplying physical coordinate or parameter labels.

N.3.4 Evaluation

Let $u = 1/R$ denote inverse tracked ball radius. For raw z and canonical $r = \psi_\rho(z)$, fit separate global affine readouts $h(z) \simeq \lambda_h u + \tau_h$ on pooled training frames, with $h(z) = z$ or $\psi_\rho(z)$. Use the frozen inverse $\widetilde{u} = [h(z) - \tau_h]/\lambda_h$ on the six test clips. The coordinate reference is $x = u$. Residual plots average test-frame residuals in twelve state bins to show how the distortion varies with ball size; error metrics use the original frames.

N.3.5 Results and physical calibration

The main results in Table 4 use IC5, the largest collection with 14 training videos, and $\rho_0 = 2.0$. Table 19 summarizes the effect of canonical normalization across its five fits. The coordinate comparison uses the inverse-radius reference and training-fitted affine readouts; it requires no physical calibration. Normalization reduces median test NRMSE from 15.85% to 5.57%, while the learned ρ is close to the ideal value 2.

Table 19: **Law-derived normalization in overhead fall.** IC5 with $\rho_0 = 2.0$, five fits. Coordinate errors e_{map} use (55) and are median [Q25, Q75]; ρ is mean $\pm$ sample SD.

Raw e_{map} (%)	Canonical e_{map} (%)	Learned ρ
15.85 [15.74, 16.01]	5.57 [5.57, 5.60]	2.072 ± 0.024

The predicted canonical acceleration $-kz^{2-\rho}$ has a median coefficient of variation of 0.73% over test frames: its standard deviation divided by the absolute value of its mean, in percent. Multiplication by a constant calibration scale leaves this quantity unchanged. Together with the coordinate errors, this supports an approximately affine height coordinate and nearly constant canonical acceleration. Figure 18 shows how normalization reduces the remaining coordinate distortion across ball sizes.

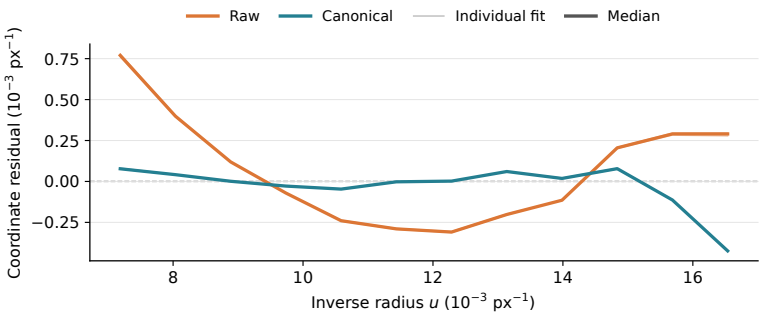


Figure 18: **Coordinate residuals for IC5.** All five fits use $\rho_0 = 2.0$. Test-coordinate residuals after the training-fitted raw or canonical affine readout are averaged in twelve inverse-radius bins. Thin curves show individual fits; thick curves show pointwise medians. Zero indicates agreement with the inverse-radius reference. Physical anchors are not used in these readouts.

Anchor. We use the nominal release heights recorded during data acquisition as external physical calibration information to map the learned canonical coordinate to height in metres. These height labels are not used to train the encoder or the dynamical model. After training, we associate each of the seven original first frames in Figure 19 with its nominal release height $h_i \in \{1.0, 1.2, 1.4\}$ m. The frozen encoder and learned normalizer produce the corresponding canonical coordinate $r_i = \psi_{\hat{\rho}}(z_i)$. Each pair (r_i, h_i) therefore provides a physical anchor linking the learned coordinate to a height in metric units. We use these anchors to determine the remaining affine scale and offset, $h = ar + b_h$:

$$(a, b_h) = \operatorname*{argmin}_{\alpha,\beta} \sum_i \frac{[h_i - \alpha r_i - \beta]^2}{n_{h_i}}, \qquad \hat{g} = a\hat{k}\left\langle z^{2-\hat{\rho}} \right\rangle_{\text{train}}. \tag{70}$$

Here n_{h_i} is the number of calibration frames assigned to height h_i, so each distinct height receives equal total weight. The fitted scale a converts canonical acceleration to physical acceleration; the offset b_h does not affect acceleration. We report $\hat{g}$ as the mean over all training frames.

This protocol treats each nominal release height as an approximation to the ball-center height in the corresponding first frame, rather than as an independently measured frame-level height. Any discrepancy therefore contributes to calibration error. The 1.4 m recording 04 contributes its first frame to calibration and a later window to coordinate testing. Thus, the two uses involve different frames, but this recording is not held out from physical calibration.

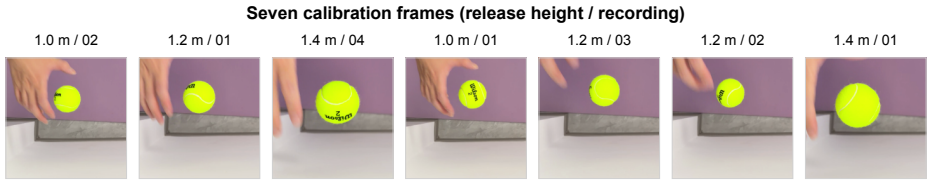


Figure 19: **Release-height anchors.** Original RGB frame zero from seven recordings, shown before masked-saturation preprocessing. Labels give nominal release height and recording ID in Table 18. Calibration applies the same preprocessing and frozen encoder used for the dynamic observations.

This calibration gives $\hat{g} = 10.146 \pm 0.029\,\text{m/s}^2$, with median parameter relative error $e_\theta = 3.35\%$ [3.26, 3.39] against $9.81\,\text{m/s}^2$, as reported in the main Table 4. The calibrated acceleration on held-out test frames has dynamics relative RMSE $e_{\text{dyn}} = 3.43\%$ [3.33, 3.65]. Gravity is reported as mean $\pm$ sample SD across the five fits; errors are median [Q25, Q75]. For (57), we use $\widetilde{F}_t = -akz_t^{2-\rho}$, $F_t = -9.81$, and equal test-frame weights. Thus $\hat{g}$ summarizes training-frame acceleration, whereas e_{dyn} checks the calibrated law on held-out trajectories.

N.3.6 Additional analysis: IC collections and coefficient initialization

We examine sensitivity to the training IC collection and the initial coefficient ρ_0 using IC1–IC5 from Table 18. Each collection is fitted with five initial networks at each $\rho_0 \in \{1.0, 2.0, 3.0\}$, giving $5 \times 5 \times 3 = 75$ fits. The initial networks are paired across collections and coefficient initializations; all settings use the same test videos and release-height anchors. Figure 20 compares the IC collections at $\rho_0 = 2.0$, and Table 20 gives all initialization results.

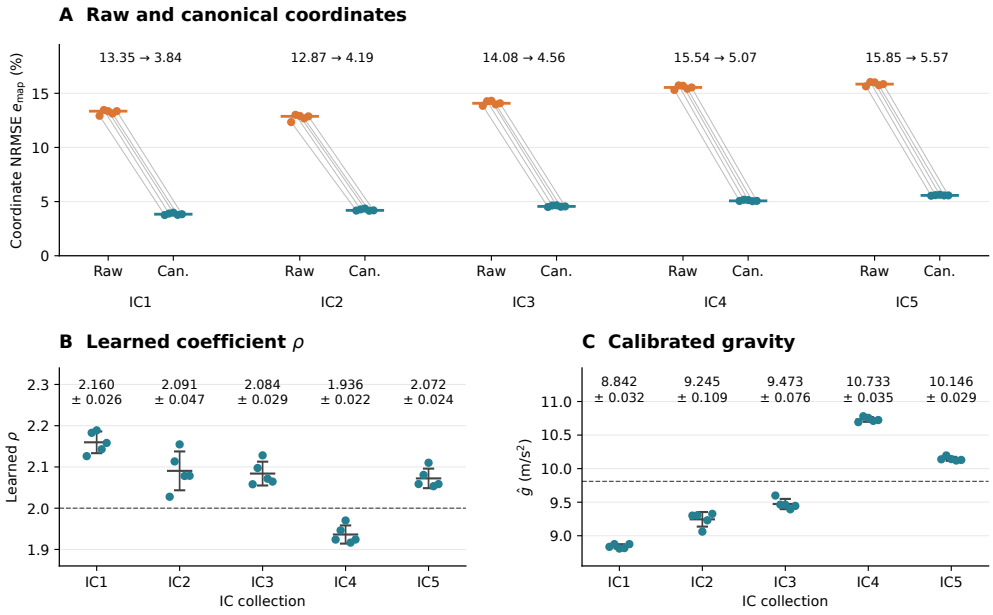


Figure 20: **Sensitivity to the training IC collection.** IC1–IC5 use the trajectory collections in Table 18, with $\rho_0 = 2.0$ and five paired local network initializations. A: paired raw/canonical test NRMSE on the same six test videos; horizontal bars and labels give medians; vertical bars show the IQR. B: learned ρ, with reference 2. C: gravity from release-height calibration, with reference $9.81\,\mathrm{m/s^2}$. Bars and labels in B/C show mean $\pm$ sample SD.

Canonical normalization reduces coordinate error for every collection and coefficient initialization. Across collections, however, coordinate accuracy and calibrated gravity accuracy do not vary together: IC5 has larger coordinate error than the smaller collections but a gravity estimate closer to the reference than IC1 and IC2. The estimated coefficient also varies with the training collection despite the shared physical setup. Because the nested IC collections differ in both trajectory composition and clip count, these differences reflect the two factors jointly.

The initialization comparison examines the estimates reached after 4,000 updates. For IC5, $\rho_0 = 1.0$ and 2.0 give $\rho = 2.070 \pm 0.019$ and 2.072 ± 0.024, respectively, whereas $\rho_0 = 3.0$ gives 2.259 ± 0.013. The first two starts reach similar estimates near 2, but the third retains a larger offset; the learned coefficient therefore remains sensitive to initialization. For the main paper, we use IC5, which includes all 14 training videos, with $\rho_0 = 2.0$. The other collections and initializations provide the sensitivity analysis reported here.

Table 20: **IC collections and coefficient initialization in overhead fall.** Five fits per row, 75 fits in total. Coefficient and gravity estimates are mean ± sample SD; coordinate errors are median [Q25, Q75] NRMSE (%). IC1–IC5 are defined in Table 18. All gravity estimates use release-height calibration. Settings share six test clips and seven calibration frames. The main experiment uses IC5 (14 videos) and $\rho_0 = 2.0$.

IC collection	ρ_0	Learned ρ	$\hat{g}$ $(\mathrm{m/s^2})$	Raw	Canonical
IC1	1.0	2.037 ± 0.012	9.037 ± 0.040	12.81 [12.78, 12.83]	3.86 [3.70, 3.88]
IC1	2.0	2.160 ± 0.026	8.842 ± 0.032	13.35 [13.13, 13.35]	3.84 [3.78, 3.90]
IC1	3.0	2.233 ± 0.079	8.674 ± 0.054	13.42 [13.29, 13.70]	3.85 [3.82, 3.94]
IC2	1.0	2.040 ± 0.025	9.380 ± 0.128	12.58 [12.57, 12.80]	4.17 [4.15, 4.24]
IC2	2.0	2.091 ± 0.047	9.245 ± 0.109	12.87 [12.68, 12.91]	4.19 [4.17, 4.28]
IC2	3.0	2.196 ± 0.095	8.960 ± 0.067	12.95 [12.84, 13.29]	4.26 [4.25, 4.39]
IC3	1.0	2.096 ± 0.019	9.423 ± 0.039	14.09 [14.01, 14.24]	4.57 [4.54, 4.64]
IC3	2.0	2.084 ± 0.029	9.473 ± 0.076	14.08 [13.97, 14.27]	4.56 [4.52, 4.64]
IC3	3.0	2.293 ± 0.027	9.177 ± 0.075	15.05 [14.92, 15.19]	4.69 [4.67, 4.77]
IC4	1.0	1.935 ± 0.019	10.718 ± 0.040	15.52 [15.39, 15.65]	5.06 [5.06, 5.13]
IC4	2.0	1.936 ± 0.022	10.733 ± 0.035	15.54 [15.42, 15.69]	5.07 [5.06, 5.14]
IC4	3.0	2.171 ± 0.011	10.292 ± 0.058	16.59 [16.48, 16.63]	5.17 [5.16, 5.24]
IC5	1.0	2.070 ± 0.019	10.142 ± 0.035	15.83 [15.72, 15.95]	5.58 [5.57, 5.60]
IC5	2.0	2.072 ± 0.024	10.146 ± 0.029	15.85 [15.74, 16.01]	5.57 [5.57, 5.60]
IC5	3.0	2.259 ± 0.013	9.834 ± 0.048	16.69 [16.60, 16.75]	5.69 [5.68, 5.69]

O SUPPLEMENTARY IDENTIFIABILITY TOOL

The supplementary offline browser tool takes a scalar ODE and parameter restrictions as input. For a supported theorem route, it reports parameter invariants under the admissible coordinate maps, calibration requirements, and the velocity feature rank required at each physical state. For squared-velocity models, it indicates when canonical normalization must precede parameter comparison. Figures 21–23 show three paper examples. The supplementary material includes the offline tool and a README with setup and usage instructions.

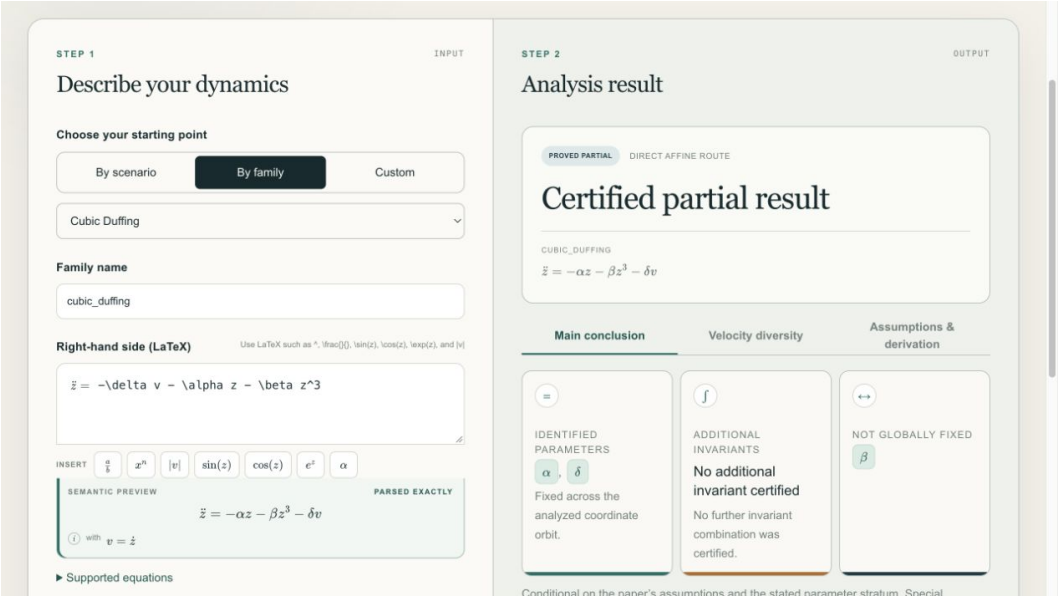


Figure 21: **Cubic Duffing example in the supplementary tool.** For the displayed nonzero-β family, the tool reports α and δ as fixed across the analyzed coordinate orbit, while β varies with coordinate scale.

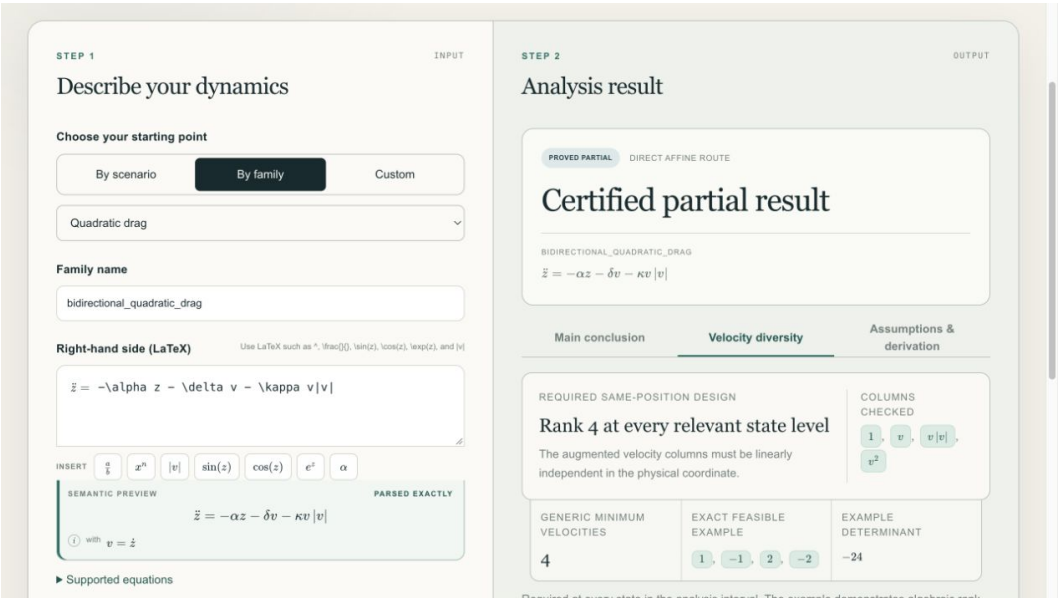


Figure 22: **Velocity coverage for quadratic drag.** The tool lists the required rank-four feature set $[1, v, v|v|, v^2]$ at each physical state and an algebraically feasible velocity example $\{1, -1, 2, -2\}$.

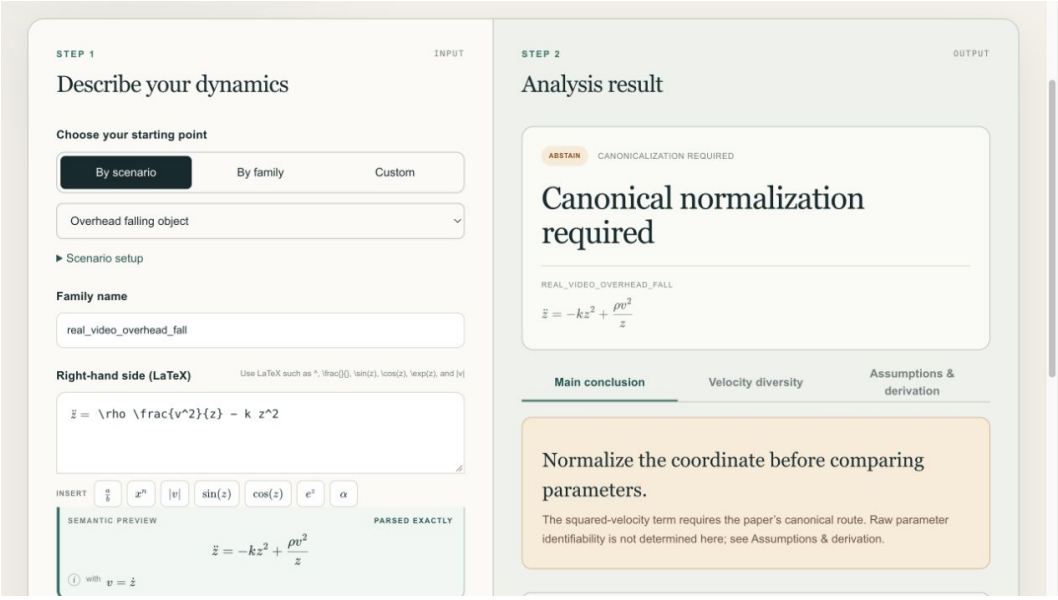


Figure 23: **Canonical route for overhead fall.** For $z'' = \rho(z')^2/z - kz^2$, the tool directs the comparison to the canonical coordinate before assessing parameter compatibility.