STRUCTURAL UNCERTAINTY IN PHYSICS-INFORMED INVERSE PROBLEMS: THE HIDDEN RISK IN SCIENTIFIC AI

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine learning is increasingly used to infer physical laws from data, yet a model that reproduces observations accurately can still encode a physically non-identifiable coefficient function. This danger is particularly acute in physics-informed machine learning (PIML), where coefficient functions of governing equations—such as Hamiltonians or Lagrangians—are learned by minimizing physics-based predictive losses. We show that this failure can originate in the governing equations themselves: multiple, physically distinct coefficient functions can satisfy the same differential equations on the same observations, a situation that additional data cannot resolve when the non-uniqueness is structural. We introduce a compatibility-aware rank criterion that screens structural ambiguity before training and evaluates whether candidate physical constraints restore identifiability. The criterion distinguishes the algebraic null space of the equation matrix $\mathbf{M}$ from coefficient-feature perturbations that can actually be realized by an admissible coefficient function. For neural parameterizations, we further intersect the numerical null space with the DNN tangent space to select the strength of a physical constraint without ground truth, based on the reduction of realizable structural uncertainty. Experiments on Hamiltonian, Lagrangian, diffusion, and wave-kinetic systems demonstrate that predictive accuracy, structural identifiability, and physical-model recovery are distinct quantities.

1 INTRODUCTION

There is growing interest in elucidating physical laws from observational data using machine learning (Karniadakis et al., 2021; Hao et al., 2022). When candidate basis functions for the governing laws are partially known, governing equations can be extracted from time-series data using linear regression (Brunton et al., 2016) or symbolic regression (Udrescu & Tegmark, 2020). When only limited prior information is available, expressive models such as deep neural networks (DNNs) can be combined with explainable artificial intelligence (XAI) techniques (Barredo Arrieta et al., 2020; Love et al., 2023) to extract interpretable physical information, including interaction laws (Cranmer et al., 2020) and conservation laws (Kaiser et al., 2018; Wetzel et al., 2020; Liu & Tegmark, 2021; Ha & Jeong, 2021; Liu & Tegmark, 2022; Mototake, 2021; Zhang et al., 2021; Lu et al., 2023).

Whereas machine learning models are typically evaluated primarily in terms of predictive performance, physics does not rely on predictive accuracy alone to assess the validity of a model. Multiple prediction-equivalent physical models may reproduce the same observations while admitting different physical interpretations. A canonical illustration is the description of planetary motion in heliocentric and geocentric coordinate systems. Even with arbitrarily accurate and abundant observations, the same celestial motion can be represented differently depending on the chosen reference frame. Predictive performance alone therefore cannot determine which coordinate-dependent representation a learning algorithm should recover.

This representational non-uniqueness also suggests a counterfactual interpretation of the historical transition from geocentric to heliocentric theories (Kuhn, 1957; Principe, 2011; Weinberg, 2015). Had modern machine learning been available in Kepler's time, a learning algorithm might have happened to learn a geocentric representation from among several prediction-equivalent descrip-

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.

tions. If that coordinate-dependent representation had then been mistaken for evidence that the geocentric theory was uniquely correct, without first evaluating the non-uniqueness of the learned physical model, the resulting interpretation could conceivably have altered the course of scientific development (Fig. 1). This counterfactual example does not imply that the historical geocentric and heliocentric theories were physically equivalent; rather, it illustrates the danger of interpreting whichever representation happens to be learned as the uniquely correct physical law. The present study therefore focuses not on XAI techniques themselves, but on the uncertainty inherent in the physical model that must be characterized before physical interpretation.

Partial differential equations (PDEs) are among the most common representations of physical models, and their coefficient functions play a central role in determining physical behavior. In Hamiltonian systems, for example, the Hamiltonian function plays an analogous role by determining the dynamics of observable variables. Physics-informed machine learning (PIML) approaches, including physics-informed neural networks (PINNs) (Raissi et al., 2019; Adams-Tew et al., 2024; Depina et al., 2022; Yang et al., 2021) and Hamiltonian neural networks (HNNs) (Schmidt & Lipson, 2009; Greydanus et al., 2019; Toth et al., 2020), estimate such unknown functions by incorporating governing differential equations or physical structures into learning. Although these approaches have been successfully applied to forward simulation, spatiotemporal forecasting (Karniadakis et al., 2021), and inverse problems such as parameter estimation (Raissi et al., 2018), the learned physical model need not be unique. Distinct coefficient functions may reproduce the same observations and satisfy the governing equations equally well, and recent studies have highlighted the lack of general identifiability guarantees in PINNs (Yang et al., 2021). Because this non-uniqueness is determined by the governing PDE, the observed variables and derivatives, and the imposed physical constraints, structural uncertainty should be diagnosed at the model level before a machine-learning solution is physically interpreted.

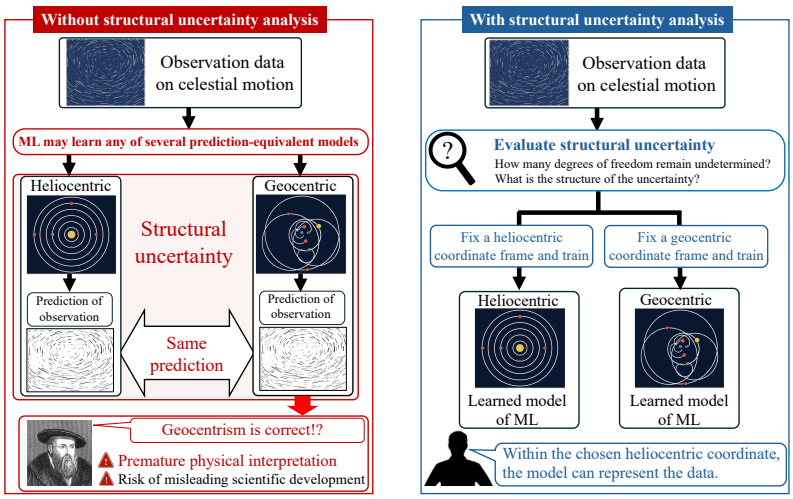


Figure 1: Structural uncertainty in machine-learning-based scientific inference. Evaluating structural uncertainty before learning helps prevent physical interpretation of an arbitrary prediction-equivalent model.

These observations motivate a model-level framework for evaluating structural uncertainty before learning and physical interpretation. Once the dimension and functional structure of the non-identifiable directions have been characterized, physically motivated constraints can be introduced explicitly to select a representative from the family of prediction-equivalent models. The resulting learned model can then be interpreted conditionally on those stated constraints, rather than being regarded as the uniquely determined physical law. The purpose of this study is to develop such a framework for quantitatively diagnosing structural uncertainty in the data-driven estimation of PDE coefficient functions. Our approach analyzes the null structure of the linear system induced by the governing PDE and characterizes the dimension and form of the coefficient-function directions that cannot be determined from the available information. This separates structural identifiability from the effects of the learning algorithm, neural-network parametrization, optimization, and finite or noisy data. The formal rank criterion developed below applies to systems that, conditional on the observed state, are linear in a finite collection of values and/or partial derivatives of the unknown coefficient function. Nonlinear dependence on these coefficient features is outside the scope of the present theorem.

2 Related Works

2.1 PIML and Inverse Problem

PIML integrates physical laws, such as PDEs, into the training process of machine learning models such as DNNs. Prominent instantiations of this idea include PINNs (Raissi et al., 2019) and HNNs (Greydanus et al., 2019), where a neural network is trained not only to fit observational data but also to satisfy a given PDE.

In the PINN framework, the governing PDE is typically of the form $\mathcal{N}[u, a](z) = 0, \quad z = (\boldsymbol{\xi}, t) \in \Omega \times [0, T]$, where $\mathcal{N}$ is a nonlinear differential operator derived from physical laws, $u : \Omega \times [0, T] \to \mathbb{R}$ is a sufficiently smooth state field, and $a : \Omega \times [0, T] \to \mathbb{R}$ is a sufficiently smooth coefficient function in the PDE. Here $\boldsymbol{\xi}$ denotes the spatial coordinate and z denotes the combined spatiotemporal coordinate. For example, this class includes equations of the form $\mathcal{N}[u, a] := u(z) - \nabla_{\boldsymbol{\xi}} \cdot (a(z)\nabla_{\boldsymbol{\xi}} u(z)) = 0$. To enforce this PDE constraint in machine learning, the loss function used in training is augmented by a physics-informed term:

$$\mathcal{L}(\theta_u, \theta_a) = \mathcal{L}_{\text{data}}(\theta_u) + \lambda_{\text{PDE}} \mathcal{L}_{\text{PDE}}(\theta_u, \theta_a), \tag{1}$$

where $\mathcal{L}_{\text{data}}(\theta_u) = \frac{1}{N} \sum_{i=1}^{N} \|u_\theta(z_i) - u_i\|^2, \quad \mathcal{L}_{\text{PDE}}(\theta_u, \theta_a) = \frac{1}{N_r} \sum_{r=1}^{N_r} \|\mathcal{N}[u_{\theta_u}, a_{\theta_a}](z_r)\|^2$, where $u_{\theta_u}(z)$ and $a_{\theta_a}(z)$ are functions parameterized by neural networks. Here, $\{(z_i, u_i)\}_{i=1}^{N}$ are supervised data points, whereas $\{(z_r)\}_{r=1}^{N_r}$ are residual points where the PDE is enforced. The parameter λ_{PDE} balances the relative importance of data fidelity and physics conformity. In this framework, the coefficient function $a(z)$ or its partial derivatives can also be estimated by minimizing $\mathcal{L}(\theta_u, \theta_a)$. In the HNN-type framework, the network directly parameterizes the unknown physical function, while the supervised information is supplied by measured trajectories and their time derivatives. A generic residual objective can be written as $\mathcal{L}_{\text{HNN}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathcal{N}[a_\theta; q_i, p_i, \dot{q}_i^{\text{obs}}, \dot{p}_i^{\text{obs}}] \right\|^2$, where the notation emphasizes that trajectory samples, rather than a tuple-valued observation function, provide the measured quantities entering the residual. For example, if the governing equations are canonical equations of motion, the loss function is

$$\mathcal{L}_{\text{HNN}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left\| \begin{bmatrix} \frac{\partial H_\theta}{\partial p}(q_i, p_i) \\ -\frac{\partial H_\theta}{\partial q}(q_i, p_i) \end{bmatrix} - \begin{bmatrix} \dot{q}_i^{\text{obs}} \\ \dot{p}_i^{\text{obs}} \end{bmatrix} \right\|^2, \tag{2}$$

where the supervised observations are trajectory samples $(q_i, p_i, \dot{q}_i^{\text{obs}}, \dot{p}_i^{\text{obs}})$, and the unknown coefficient function is the Hamiltonian $H(q, p)$. In HNNs, the Hamiltonian is therefore represented as a function parameterized by a neural network and is estimated by minimizing the dynamical residual.

2.2 Types of Uncertainty and Our Focus

Uncertainties in data-driven physical modeling can be classified into three types (Pelz et al., 2021). *Structural uncertainty* persists even with infinite, noiseless observations when multiple physical models remain observationally indistinguishable under the available variables and constraints. Examples include coordinate-dependent representations when the reference frame has not been fixed, as well as gauge freedoms in physical theories. *Model-form uncertainty* arises when the true physical law lies outside the assumed model class. *Data uncertainty* arises from finite, noisy, or insufficiently informative observations, even when the underlying physical model is structurally identifiable. These three types of uncertainty require fundamentally different remedies. Structural uncertainty requires additional physical constraints or complementary observables; model-form uncertainty requires enrichment or revision of the model class; and data uncertainty can be reduced through broader or more precise measurements and quantified using statistical approaches such as Bayesian inference (Pelz et al., 2021; René & Longtin,

Table 1: Types of uncertainty in inverse estimation of physical models.

Type	Source	Typical remedy
Structural	Multiple models fit the same data	Add physical constraints or complementary observables
Model-form	True law lies outside the assumed model class	Enrich or revise the model class
Data	Finite, noisy, or insufficiently informative observations	Increase amount, precision, and coverage of data

2025) (Table 1). When the goal is physical interpretation, structural uncertainty must be evaluated first, because neither additional data nor posterior concentration can select a unique model from an intrinsically non-identifiable family.

2.3 Structural Uncertainty in PIML

Structural uncertainty is closely related to the classical notion of structural identifiability, which has been extensively studied for scalar parameters in ordinary differential equation (ODE)-based dynamical systems using differential-algebraic, reparametrization, and algebraic methods (Ljung & Glad, 1994; Miao et al., 2011; Meshkat & Sullivant, 2014; Meshkat et al., 2018; Meshkat & Shiu, 2025). The setting considered here is different. The unknown is not a finite-dimensional parameter vector, but a coefficient *function*, together with its partial derivatives, defined over the domain of a PDE. Consequently, non-uniqueness is governed by the structure of the PDE, the available derivative information, and the physical constraints imposed on the coefficient function.

Classical PDE inverse theory provides powerful uniqueness and stability results for particular classes of problems. Examples include Calderón-type inverse problems (Calderón, 1980; Sylvester & Uhlmann, 1987) and hyperbolic inverse problems based on Carleman estimates (Yamamoto, 2009; Bellassoued & Yamamoto, 2017). However, these results rely on problem-specific measurement, regularity, and geometric assumptions and do not directly provide a general pre-learning diagnostic for coefficient-function estimation in PIML. Classical uniqueness results often formulate the inverse problem through boundary observations, whereas the PIML setting considered here assumes that the state and the derivatives required to assemble the equation matrix are available, or can be reliably estimated, at sufficiently dense interior or scattered observation points. We therefore do not claim uniqueness from boundary data alone. In the machine-learning context, PINNs can converge to physically incorrect solutions despite achieving low training loss (Krishnapriyan et al., 2021). Bayesian extensions such as Bayesian PINNs quantify uncertainty through posterior or predictive distributions (Yang et al., 2021; Mishra et al., 2022), but these distributions generally combine structural non-identifiability with data, model-form, approximation, and computational uncertainties. Moreover, approximate posterior inference may fail to reveal unbounded non-identifiable directions in coefficient-function space.

3 Theoretical Preliminaries for Uncertainty Evaluation

We first derive a rank criterion for structural uncertainty in PDE coefficient functions, which represent the physical models to be inferred.

Definition 1 (Structural Uncertainty Modulo Gauge)**.** Let $a : \Omega \to \mathbb{R}$ denote the true coefficient function and let $u^\star$ satisfy $\mathcal{F}(u^\star, a) = 0$ on Ω. Let $\mathcal{G}$ denote the class of unavoidable gauge perturbations that are not regarded as physically distinct in the inverse problem. For the derivative-only Hamiltonian examples below, $\mathcal{G}$ contains additive constants. We say that *nontrivial structural uncertainty* arises if there exists another admissible function $\hat{a}$ such that

$$\mathcal{F}(u^\star, \hat{a}) = 0, \qquad \hat{a} - a \notin \mathcal{G}.$$

Thus identifiability is always understood modulo the explicitly stated gauge freedom and conditional on the specified observation operator.

Theorem 1 (Compatibility-Aware Rank Criterion for PDE Coefficient Functions)**.** Fix the observed state $u^\star$ and a finite evaluation grid. Suppose that, conditional on these observations, the governing equations are linear in a finite vector of coefficient features—values and/or partial derivatives of the unknown coefficient function—so that the discretized inverse problem can be written as

$$\mathbf{M}\mathbf{d}(a) = \mathbf{c}, \tag{3}$$

where $\mathbf{d}(a) \in \mathbb{R}^m$ is the stacked coefficient-feature vector, $\mathbf{M} \in \mathbb{R}^{n\times m}$ is determined by $u^\star$ and its observed derivatives, and $\mathbf{c}$ is independent of the candidate coefficient function a.

Let $\boldsymbol{\xi} \in \mathbb{R}^p$ denote a discrete representation of admissible perturbations of the coefficient function after explicitly quotiented gauge coordinates have been removed, and let

$$\delta\mathbf{d} = \mathbf{T}\,\delta\boldsymbol{\xi}$$

be the linear compatibility map from coefficient perturbations to perturbations of the PDE coefficient features. The column space $\mathcal{C} = \mathrm{Range}(\mathbf{T})$ therefore contains only feature perturbations that can be generated by a common admissible coefficient function. In particular, repeated mixed partials of a sufficiently smooth scalar coefficient are represented only once through this common map, so relations such as $\partial_{xy}a = \partial_{yx}a$ are respected automatically.

Then the following statements are equivalent:

(i) No nonzero compatible coefficient-feature perturbation, modulo the removed gauge, leaves the discretized governing equations unchanged.

(ii)
$$\ker(\mathbf{M}) \cap \mathcal{C} = \{\mathbf{0}\}.$$

(iii)
$$\ker(\mathbf{MT}) = \ker(\mathbf{T}),$$
or equivalently
$$\mathrm{rank}(\mathbf{MT}) = \mathrm{rank}(\mathbf{T}).$$

If the coefficient features are deliberately treated as unrestricted algebraic variables, $\mathbf{T} = \mathbf{I}$ and the criterion reduces to the ordinary full-column-rank condition $\mathrm{rank}(\mathbf{M}) = m$ for a consistent system. Conversely, if $\mathbf{M}$ is rank deficient but its null directions do not intersect $\mathcal{C}$, the underlying coefficient function can still be identifiable. Thus the raw algebraic null space of $\mathbf{M}$ is, in general, an upper bound on the realizable functional ambiguity.

The proof of Theorem 1 is provided in Appendix A.

Remark 1 (Discrete scope and continuous interpretation)**.** Theorem 1 is a statement about the chosen discretized coefficient representation and does not infer exact continuous-domain uniqueness from a single finite grid. When the discrete differential operators are consistent and the grid is refined, the compatibility map $\mathbf{T}$ approximates the corresponding continuous differential relations under the stated smoothness assumptions. Continuous-domain uniqueness therefore requires the corresponding limiting and regularity assumptions, rather than following solely from uniqueness on one finite grid.

3.1 Canonical Examples

The compatibility-aware criterion separates the algebraic equation matrix $\mathbf{M}$ from the relations imposed by a common underlying coefficient function through $\mathbf{T}$. The following examples illustrate full rank, structural ambiguity under a specified observation operator, and data-dependent degeneracy.

Hamiltonian system. For canonical variables $(\boldsymbol{q}, \boldsymbol{p})$, Hamilton's equations give $\dot{\boldsymbol{q}} = \partial_{\boldsymbol{p}} H$ and $\dot{\boldsymbol{p}} = -\partial_{\boldsymbol{q}} H$. On an N-point grid, the coefficient-feature vector contains the $2Nd$ components $(\partial_{\boldsymbol{q}} H, \partial_{\boldsymbol{p}} H)$, and the equation matrix is the block symplectic matrix $\mathbf{M} = \left(\begin{smallmatrix} \mathbf{0} & I_{Nd} \\ -I_{Nd} & \mathbf{0} \end{smallmatrix}\right)$ with $\mathrm{rank}(\mathbf{M}) = 2Nd$. Hence the derivative field is uniquely fixed by the observed motion. The remaining additive constant lies in the kernel of the derivative representation and is therefore an explicitly quotiented gauge rather than nontrivial structural uncertainty.

Lagrangian system. We use an augmented observation setting in which the generalized momentum $\boldsymbol{p}^{\mathrm{obs}}$ and its time derivative $\dot{\boldsymbol{p}}^{\mathrm{obs}}$ are available as measured physical quantities independently of the candidate Lagrangian. The first-order Lagrangian relations are

$$\dot{\boldsymbol{p}}^{\mathrm{obs}} = \partial_{\boldsymbol{q}} L, \qquad \boldsymbol{p}^{\mathrm{obs}} = \partial_{\dot{\boldsymbol{q}}} L.$$

We deliberately use the first relation as the baseline governing equation and the second as an additional physical constraint. The coefficient-feature vector contains $2Nd$ components $(\partial_{\boldsymbol{q}} L, \partial_{\dot{\boldsymbol{q}}} L)$, whereas the baseline relation constrains only the first Nd components. Thus

$$\mathbf{M} = (I_{Nd}\ \mathbf{0}), \qquad \mathrm{rank}(\mathbf{M}) = Nd < 2Nd.$$

This algebraic null space contains realizable perturbations: for example, adding a nonconstant function of $\dot{\boldsymbol{q}}$ alone changes $\partial_{\dot{\boldsymbol{q}}} L$ while leaving $\partial_{\boldsymbol{q}} L$ unchanged. Hence the baseline observation operator is genuinely structurally underdetermined. Adding the independently observed momentum relation $\partial_{\dot{\boldsymbol{q}}} L = \boldsymbol{p}^{\mathrm{obs}}$ removes these null directions. This canonical example is therefore conditional on an augmented momentum observation channel; it is not a claim that $\boldsymbol{p}$ is directly observed in every inverse-Lagrangian problem.

One-dimensional diffusion system. For $\partial_t u = \partial_x(a\,\partial_x u) = \partial_x a\,\partial_x u + a\,\partial_{xx} u$, represent the unknown coefficient by its nodal values $\mathbf{a} = (a(x_1), \ldots, a(x_N))^\top$ and let $\mathbf{D}_x$ be the discrete first-derivative operator. The compatible coefficient-feature vector is

$$\mathbf{d}(a) = \begin{bmatrix} \partial_x a \\ a \end{bmatrix} \simeq \underbrace{\begin{bmatrix} \mathbf{D}_x \\ \mathbf{I} \end{bmatrix}}_{\mathbf{T}} \mathbf{a}.$$

At time t_r, the effective inverse operator is

$$\mathbf{A}_r = \mathrm{diag}(u_x(t_r))\mathbf{D}_x + \mathrm{diag}(u_{xx}(t_r)), \qquad \mathbf{A}_r \mathbf{a} = \mathbf{u}_t(t_r). \tag{4}$$

Stacking all observed times gives $\mathbf{A} = [\mathbf{A}_1^\top \cdots \mathbf{A}_R^\top]^\top$. The coefficient is identifiable on the discrete representation when $\mathrm{rank}(\mathbf{A}) = N$. This formulation handles the mixed derivative orders through the compatibility map rather than treating a and a_x as independent algebraic variables. For the zero-state control $u \equiv 0$, $\mathbf{A} = \mathbf{0}$; this is a data-induced degeneracy caused by an uninformative observation, not structural uncertainty of the diffusion PDE itself.

Together, these examples make the evaluation procedure operational: assemble the algebraic equation matrix $\mathbf{M}$, encode known compatibility relations through $\mathbf{T}$ when available, and test whether $\ker(\mathbf{M})$ intersects the realizable coefficient-feature subspace. In the numerical DNN procedure below, the local DNN tangent provides a representation-specific approximation to this realizable feature subspace, preventing algebraic null directions that the chosen neural representation cannot realize from being counted as model uncertainty.

Detailed derivations and numerical demonstrations are provided in Appendices B–D.

4 Proposed Framework

When estimating a coefficient function $a(x)$ using PIML, structural uncertainty in the underlying physical system may lead to a low-loss but physically arbitrary learned solution. Building on the rank criterion above, we use the following three-step framework.

Step 1: Diagnose structural uncertainty. Before implementing a machine-learning model, we construct the equation matrix $\mathbf{M}$ in Theorem 1 from the governing equation and observed state derivatives. Its numerical rank and algebraic null space are obtained from an SVD using a prescribed relative singular-value threshold. When an explicit compatibility map $\mathbf{T}$ is available, we test the realizable ambiguity through $\ker(\mathbf{M}) \cap \mathrm{Range}(\mathbf{T})$, or equivalently through the rank of $\mathbf{MT}$. Without this compatibility restriction, rank deficiency of $\mathbf{M}$ is interpreted conservatively as an upper bound on functional structural uncertainty.

Step 2: Introduce a physically motivated constraint. The null structure of $\mathbf{M}$ indicates what information is missing. We introduce a domain-grounded constraint that removes those directions and add its residual to the learning objective. The rank test is applied again to the constrained system to verify that the proposed constraint restores identifiability. Importantly, the framework does not infer which physical constraint is "correct" from data alone; identifiability is conditional on the stated physical hypothesis.

Step 3: Select the strength of a soft constraint from the numerical null–tangent overlap. We use the normalized objective $\mathcal{L}_\lambda(\theta) = \bar{\mathcal{L}}_{\mathrm{phys}}(\theta) + \lambda \bar{\mathcal{L}}_{\mathrm{con}}(\theta)$, with the numerical normalizations fixed before the λ sweep. Predictive validation loss alone is not sufficient to determine λ, because a low-loss solution may still move along coefficient-function directions left undetermined by the governing equation.

The structural directions are obtained numerically rather than supplied analytically. Let $\mathbf{a}_\theta$ be the discretized vector of coefficient-function derivatives in $\mathbf{Ma} = \mathbf{c}$, and let $\mathbf{J}_a = \partial \mathbf{a}_\theta / \partial \theta$. We compute the numerical SVD $\mathbf{M} = \mathbf{U}_M \mathbf{S}_M \mathbf{V}_M^\top$ and retain the right-singular vectors above a fixed relative threshold. If $\mathbf{Q}_R$ denotes the retained orthonormal row-space basis, the numerical null projector is $\mathbf{P}_N = \mathbf{I} - \mathbf{Q}_R \mathbf{Q}_R^\top$. Thus no problem-specific analytic expression for $\ker(\mathbf{M})$ is required.

We next remove DNN parameter redundancies in the same coefficient-derivative space. From the SVD $\mathbf{J}_a = \mathbf{U}_a \mathbf{S}_a \mathbf{V}_a^\top$, we retain the active singular directions and define $\mathbf{Q}_T = \mathbf{U}_{a,r}$ and $\mathbf{B} = \mathbf{V}_{a,r} \mathbf{S}_{a,r}^{-1}$, so that $\mathbf{J}_a \mathbf{B} = \mathbf{Q}_T$. Gauge directions that do not change the derivative vector are therefore removed automatically. The numerical overlap between the equation-level null space and the DNN-representable tangent is characterized by $\mathbf{K} = \mathbf{Q}_T^\top \mathbf{P}_N \mathbf{Q}_T$. If $\mathbf{K}\mathbf{v}_i = \gamma_i \mathbf{v}_i$, then $\gamma_i = \cos^2 \phi_i \in [0, 1]$ is the squared cosine of the principal angle between the tangent direction $\mathbf{Q}_T \mathbf{v}_i$ and the numerical null space of $\mathbf{M}$. Exact intersections have $\gamma_i = 1$, while near-unit values identify numerical near-intersections caused by discretization, noise, or a restricted neural representation.

In the derivative-normalized active coordinates defined by $\mathbf{B}$, we use the generalized Gauss–Newton (GGN) curvature $\mathbf{G}_\lambda = \mathbf{G}_{\text{phys}} + \lambda \mathbf{G}_{\text{con}}$ and the local Laplace covariance proxy $\mathbf{\Sigma}_\lambda \propto (\mathbf{G}_\lambda + \eta \mathbf{I})^{-1}$. The variance along principal direction i is $s_i^2(\lambda) = \mathbf{v}_i^\top \mathbf{\Sigma}_\lambda \mathbf{v}_i$. Using the continuous null-overlap weights γ_i, we define

$$U_{\text{struct}}(\lambda) = \frac{\sum_i \gamma_i s_i^2(\lambda)}{\sum_i \gamma_i}, \qquad U_{\text{id}}(\lambda) = \frac{\sum_i (1 - \gamma_i) s_i^2(\lambda)}{\sum_i (1 - \gamma_i)}, \qquad R_{\text{NT}}(\lambda) = \frac{U_{\text{struct}}(\lambda)}{U_{\text{id}}(\lambda)}. \tag{5}$$

In the exact noiseless limit, $\gamma_i \in \{0, 1\}$ and this reduces to the average covariance in $\ker(\mathbf{M}) \cap \text{Range}(\mathbf{J}_a)$ and its tangent-space complement.

For positive logarithmically spaced values λ_j, we compute the decade reduction $\Delta_j = R_{\text{NT}}(\lambda_{j-1}) - R_{\text{NT}}(\lambda_j)$ and, when $\Delta_j, \Delta_{j+1} > 0$, the diminishing-return score $E_j = 1 - \Delta_{j+1}/\Delta_j$, following the general knee/elbow principle of diminishing returns (Satopää et al., 2011). We select $\lambda^* = \lambda_{j^*}$ with $j^* = \arg\max_j E_j$. Thus Step 1 and Step 3 use the same numerical equation matrix $\mathbf{M}$, while the DNN tangent intersection prevents equation-level null directions that the chosen neural representation cannot realize locally from being counted as model uncertainty.

Algorithmic pseudocode for the complete numerical null–tangent procedure is given in Appendix G.

Appendix G reports two ground-truth-free implementation checks: finite-difference realization of the DNN tangent modes and the rank of the constraint Jacobian on the structural tangent. These diagnostics validate local realizability and constraint lifting but do not change λ^*.

5 Demonstration

We demonstrate the complete three-step framework on two structurally non-identifiable systems: a controlled Lagrangian problem and a wave-kinetic Hamiltonian inverse problem. Additional controlled Hamiltonian and diffusion experiments that validate the rank criterion in the full-rank and data-dependent cases are reported in Appendices B and D.

5.1 Lagrangian System

We first consider the simple-pendulum Lagrangian problem and training setup described in Appendices C and E, respectively. The synthetic ground-truth Lagrangian is used only for post-hoc evaluation and is not used in the structural diagnosis or in the selection of the constraint strength.

Step 1: Diagnose Structural Uncertainty. The observation operator for this controlled example supplies p^{obs} and $\dot{p}^{\text{obs}}$ independently of the candidate Lagrangian. We use the baseline relation $\partial_\theta L = \dot{p}^{\text{obs}}$, so for the coefficient-feature vector $\mathbf{a} = (\partial_\theta L, \partial_{\dot{\theta}} L)$ the numerically assembled equation matrix is algebraically equivalent to $\mathbf{M} = (\mathbf{I}_N \ \mathbf{0})$ and has $\text{rank}(\mathbf{M}) = N < 2N$. The null directions are realizable because functions of $\dot{\theta}$ alone can change $\partial_{\dot{\theta}} L$ without changing $\partial_\theta L$. Hence the baseline observation operator leaves genuine structural ambiguity. Consistent with this diagnosis, the unconstrained model attains a low prediction MSE of 0.076747 ± 0.021611 while its post-hoc cosine similarity to the ground-truth Lagrangian is only 0.647355 ± 0.096927.

Step 2: Introduce a Physical Constraint. To remove the undetermined derivative directions, we introduce the independently observed generalized-momentum relation $\partial_{\dot\theta} L = p^{\mathrm{obs}}$ as an additional physical constraint. For the normalized synthetic pendulum ($m = l = 1$), the observation channels satisfy $p^{\mathrm{obs}} = \dot\theta$ and $\dot p^{\mathrm{obs}} = \ddot\theta$ numerically, but they are treated as externally supplied observations rather than quantities computed from the learned L.

Step 3: Select the Constraint Strength. We impose the generalized-momentum relation softly through

$$\mathcal{L}_\lambda = \mathbb{E}\left[(\partial_\theta L_\vartheta - \dot p^{\mathrm{obs}})^2 + \lambda(\partial_{\dot\theta} L_\vartheta - p^{\mathrm{obs}})^2\right], \tag{6}$$

where ϑ denotes the neural-network parameters and is distinct from the state coordinate θ. Because $p^{\mathrm{obs}} = \dot\theta$ and $\dot p^{\mathrm{obs}} = \ddot\theta$ in the normalized synthetic system, this is numerically identical to the objective used to produce the reported results. Applying the ground-truth-free numerical null–tangent procedure of Sec. 4 gives $\lambda^*_{\mathrm{Lagrangian}} = 10^{-3}$. At this value, the post-hoc cosine similarity increases to 0.992389 ± 0.003381 while the prediction MSE remains essentially unchanged at 0.076911 ± 0.021777. The selected value lies at the onset of a broad high-recovery regime that contains the numerical maximum of the post-hoc cosine similarity [Fig. 2(c,d)]. The full sweep and numerical diagnostics are given in Appendix Table 2.

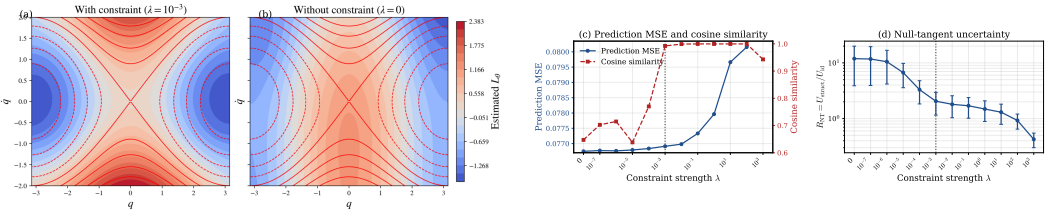


Figure 2: Lagrangian demonstration. (a,b) Learned Lagrangians with the selected generalized-momentum constraint and without the constraint; red contours show the ground truth. (c) Prediction MSE (baseline equation-of-motion residual without the momentum-constraint term) and post-hoc cosine similarity versus λ. (d) R_{NT}. Dashed lines mark $\lambda^* = 10^{-3}$.

5.2 Wave-Kinetic Equation

We next consider Hamiltonian inference in a wave-kinetic model of magnetized-plasma turbulence, a reduced description used to study turbulence and zonal-flow dynamics in fusion plasmas (Diamond et al., 2005; Gürcan & Diamond, 2015; Kaw et al., 2001; Sasaki et al., 2017; 2018; Garbet et al., 2021; Sasaki et al., 2021). The turbulence-intensity distribution $I(x, k_x, t)$ evolves according to

$$\partial_t I(z) + \partial_{k_x} H(z)\partial_x I(z) - \partial_x H(z)\partial_{k_x} I(z) = C(z), \tag{7}$$

where $z = (x, k_x, t)$, H is the Hamiltonian to be inferred, and C is a known generation–damping term. The numerical data are generated from a coupled wave-kinetic/geodesic-acoustic-mode simulation; the ground-truth Hamiltonian, source model, simulation parameters, and derivative construction are given in Appendix F. At each time slice $t = \tau$, we estimate $H(x, k_x, \tau)$ from I and its numerically evaluated derivatives.

Step 1: Diagnose Structural Uncertainty. Discretizing the wave-kinetic equation on an $N_x \times N_{k_x}$ grid yields $\mathbf{M} \in \mathbb{R}^{N_x N_{k_x} \times 2N_x N_{k_x}}$; its numerical SVD gives $\mathrm{rank}(\mathbf{M}) = N_x N_{k_x} < 2N_x N_{k_x}$. Hence the equation has an algebraic null space in the two Hamiltonian-derivative fields (Appendix Eqs. 24 and 25); the null–tangent analysis below confirms that substantial portions of this ambiguity are realizable by the chosen Hamiltonian network.

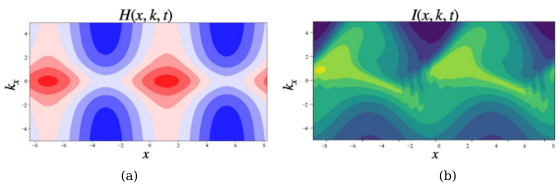


Figure 3: Wave-kinetic simulation fields. (a) Hamiltonian $H(x, k_x, t)$. (b) Turbulence intensity $I(x, k_x, t)$.

Step 2: Introduce a Physical Constraint. We impose the symmetry $H(x, -k_x, t) = H(x, k_x, t)$, implying that $\partial_x H$ is even and $\partial_{k_x} H$ is odd in k_x. This reduces the independent Hamiltonian derivatives to $N_x N_{k_x}$. For a paired point $(x, \pm k_x)$, the corresponding 2×2 block is invertible precisely when $(\partial_{k_x} I^{(+)})(\partial_x I^{(-)}) + (\partial_x I^{(+)})(\partial_{k_x} I^{(-)}) \neq 0$. All 49,000 retained paired interior blocks in the ten analyzed time slices satisfy this determinant condition, making the symmetry-constrained equation locally identifiable up to the additive-constant gauge (Appendix Eqs. 27 and 28).

Step 3: Select the Constraint Strength. To impose the symmetry softly during learning, we use

$$\mathrm{Loss}(\theta_{\mathrm{nn}}) = \bar{\mathcal{L}}_{\mathrm{PDE}} + \lambda \bar{\mathcal{L}}_{\mathrm{sym}}, \tag{8}$$

$$\bar{\mathcal{L}}_{\mathrm{PDE}} = 10^8\, \mathbb{E}\left[\left\| \partial_t I^{(i,j)} - C^{(i,j)} + \partial_{k_x} H^{(i,j)}_{\theta_{\mathrm{nn}}} \partial_x I^{(i,j)} - \partial_x H^{(i,j)}_{\theta_{\mathrm{nn}}} \partial_{k_x} I^{(i,j)} \right\|^2\right], \tag{9}$$

$$\bar{\mathcal{L}}_{\mathrm{sym}} = 10^4\, \mathbb{E}\left[\left\| H_{\theta_{\mathrm{nn}}}(x_i, k_j) - H_{\theta_{\mathrm{nn}}}(x_i, -k_j) \right\|^2\right]. \tag{10}$$

The fixed factors 10^8 and 10^4 normalize the residual magnitudes and are held fixed throughout the sweep. The ground-truth-free numerical null–tangent procedure selects $\lambda^*_{\mathrm{wave}} = 10^1$, within the broad high-recovery regime identified only post hoc [Fig. 4(c,d)]. In addition to mean-centered cosine similarity, we report the mean-centered relative L_2 error to retain sensitivity to the recovered Hamiltonian amplitude: it is 0.6907 ± 0.0268 at the selected value and reaches its numerical minimum 0.6238 ± 0.0289 at $\lambda = 10^{-1}$. Thus the selector identifies the same broad recovery regime rather than the exact post-hoc optimum. Implementation details, animations, and the full sweep are given in Appendices G and H and Table 6.

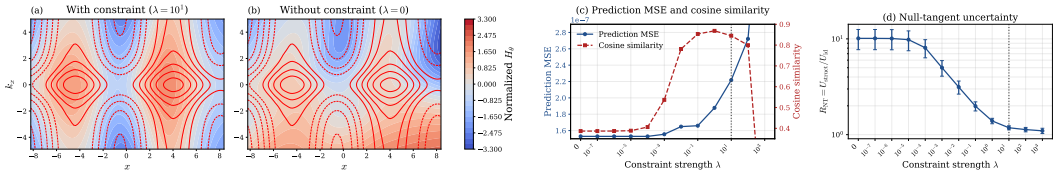


Figure 4: Wave-kinetic demonstration. (a,b) Hamiltonian estimates with the selected symmetry constraint and without the constraint; red contours show the ground truth. (c) Prediction MSE (held-out PDE residual only, excluding the symmetry term) and post-hoc mean-centered cosine similarity versus λ. (d) R_{NT}. Dashed lines mark $\lambda^* = 10^1$.

6 Discussion

We introduced a compatibility-aware rank criterion that characterizes structural uncertainty in PDE coefficient-function estimation and determines whether additional physical constraints restore identifiability. Hamiltonian, Lagrangian, diffusion, and wave-kinetic examples demonstrate its applicability across qualitatively different physical systems. The compatibility map distinguishes algebraic null directions from perturbations that can actually arise from a common coefficient function; consequently, a rank-deficient $\mathbf{M}$ alone is interpreted as an upper bound on functional ambiguity unless its null space intersects the compatible subspace. Candidate constraints remain hypotheses grounded in domain knowledge, and identifiability is always conditional on the stated observation operator. The criterion complements Bayesian uncertainty quantification: in the wave-kinetic example, a B-PINN posterior neither revealed the structural ambiguity diagnosed by the rank criterion nor separated structural from finite-sample effects (Appendix K), whereas the rank criterion identifies this ambiguity directly from the PDE and the observed state.

For soft constraints, we further introduced a ground-truth-free procedure for selecting λ by comparing the numerical null space of $\mathbf{M}$ with the DNN-realizable coefficient-function tangent. In the Lagrangian and wave-kinetic examples, it identifies a broad high-recovery regime without using the ground-truth coefficient function. In the wave-kinetic case, the scale-sensitive relative error further confirms that the selected value need not coincide with the exact post-hoc optimum. More broadly, high predictive accuracy does not guarantee a uniquely identifiable physical model. The present theorem assumes linearity in the chosen coefficient features and sufficiently informative interior or scattered observations to assemble the equation matrix; nonlinear coefficient dependence and boundary-only inverse settings require separate analysis.

REPRODUCIBILITY STATEMENT

The assumptions and complete proof of the rank criterion are provided in Appendix A. Controlled Hamiltonian, Lagrangian, and diffusion experiments are documented in Appendices B–D, and model architectures and training settings in Appendix E. Wave-kinetic rank analysis and data details are given in Appendix F; numerical null-space, DNN-tangent, generalized Gauss–Newton/Laplace, finite-difference, and constraint-lifting procedures are specified in Appendix G. Time-series animations are provided in Appendix H and as separate GIF files, while the full λ sweeps and evaluation metric are reported in Appendices I and J. The accompanying anonymized reproduction bundle contains source code, numerical tables, validation outputs, figure-generation scripts, and media required to reproduce the reported results.

AI USE STATEMENT

In this work, we used generative AI tools (OpenAI ChatGPT) for the following tasks for which disclosure is required: helping develop the methodological framework; proposing and refining research ideas; formulating mathematical claims and assisting in the writing and checking of derivations and proof arguments; implementing methods (code generation, debugging, and execution support for the computational experiments); generating the synthetic data for the Lagrangian experiments; and assisting with translation of manuscript text into English. We did not use generative AI tools for interpreting the experimental results. Additionally, we used generative AI tools for tasks for which disclosure is recommended: literature retrieval and identification of related work, and language editing and drafting of portions of the manuscript. We have reviewed all AI-assisted work. AI-generated suggestions were not accepted automatically: the authors independently inspected the relevant source literature, verified citations, checked mathematical claims and code, reproduced and validated the numerical results, and revised the manuscript. We take full responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Samuel I. Adams-Tew, Henrik Odéen, Dennis L. Parker, Cheng-Chieh Cheng, Bruno Madore, Allison Payne, and Sarang Joshi. Physics informed neural networks for estimation of tissue properties from multi-echo configuration state MRI. In Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pp. 502–511, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-72120-5.

Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, 2020. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2019.12.012. URL `https://www.sciencedirect.com/science/article/pii/S1566253519308103`.

Mourad Bellassoued and Masahiro Yamamoto. *Carleman Estimates and Inverse Problems for Hyperbolic Equations*. Springer Monographs in Mathematics. Springer Tokyo, Tokyo, 2017. ISBN 978-4-431-56598-7. doi: 10.1007/978-4-431-56600-7.

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.

Alberto P Calderón. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics*, pp. 65–73, Rio de Janeiro, 1980. Sociedade Brasileira de Matemática.

Miles Cranmer, Alvaro Sanchez Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. *Advances in neural information processing systems*, 33:17429–17442, 2020.

Ivan Depina, Saket Jain, Sigurdur Mar Valsson, and Hrvoje Gotovac. Application of physics-informed neural networks to inverse problems in unsaturated groundwater flow. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 16(1):21–36, 2022.

Patrick H Diamond, SI Itoh, K Itoh, and TS Hahm. Zonal flows in plasma—a review. *Plasma Physics and Controlled Fusion*, 47(5):R35, 2005.

X. Garbet, O. Panico, R. Varennes, C. Gillot, G. Dif-Pradalier, Y. Sarazin, E. Bourne, V. Grandgirard, P. Ghendrih, D. Zarzoso, and L. Vermare. Zonal instability and wave trapping. *Journal of Physics: Conference Series*, 1785(1):012002, feb 2021. doi: 10.1088/1742-6596/1785/1/012002. URL https://doi.org/10.1088/1742-6596/1785/1/012002.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Ö D Gürcan and PH Diamond. Zonal flows and pattern formation. *Journal of Physics A: Mathematical and Theoretical*, 48(29):293001, 2015.

Seungwoong Ha and Hawoong Jeong. Discovering invariants via machine learning. *Physical Review Research*, 3(4):L042035, 2021.

Zhongkai Hao, Songming Liu, Yichi Zhang, Chengyang Ying, Yao Feng, Hang Su, and Jun Zhu. Physics-informed machine learning: A survey on problems, methods and applications. *arXiv preprint arXiv:2211.08064*, 2022.

Eurika Kaiser, J Nathan Kutz, and Steven L Brunton. Discovering conservation laws from data for control. In *2018 IEEE Conference on Decision and Control (CDC)*, pp. 6415–6421. IEEE, 2018.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Predhiman Kaw, Raghvendra Singh, and PH Diamond. Coherent nonlinear structures of drift wave turbulence modulated by zonal flows. *Plasma Physics and Controlled Fusion*, 44(1):51, 2001.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 26548–26560, 2021.

Thomas S. Kuhn. *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Harvard University Press, Cambridge, MA, 1957. ISBN 978-0674171039. See especially Chapter 5, pp. 168-172, for the discussion on the comparative accuracy and complexity of the Ptolemaic and Copernican systems.

Ziming Liu and Max Tegmark. Machine learning conservation laws from trajectories. *Physical Review Letters*, 126(18):180604, 2021.

Ziming Liu and Max Tegmark. Machine learning hidden symmetries. *Phys. Rev. Lett.*, 128:180201, May 2022. doi: 10.1103/PhysRevLett.128.180201. URL https://link.aps.org/doi/10.1103/PhysRevLett.128.180201.

Lennart Ljung and Torkel Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2):265–276, 1994.

Peter ED Love, Weili Fang, Jane Matthews, Stuart Porter, Hanbin Luo, and Lieyun Ding. Explainable artificial intelligence (xai): Precepts, models, and opportunities for research in construction. *Advanced Engineering Informatics*, 57:102024, 2023.

Peter Y Lu, Rumen Dangovski, and Marin Soljačić. Discovering conservation laws using optimal transport and manifold learning. *Nature Communications*, 14(1):4744, 2023.

Nicolette Meshkat and Anne Shiu. Structural identifiability of compartmental models: Recent progress and future directions. *Current Opinion in Systems Biology*, 42:100559, 2025. doi: 10.1016/j.coisb.2025.100559.

Nicolette Meshkat and Seth Sullivant. Identifiable reparametrizations of linear compartment models. *Journal of Symbolic Computation*, 63:46–67, 2014.

Nicolette Meshkat, Zachary Rosen, and Seth Sullivant. Algebraic tools for the analysis of state space models. In *The 50th Anniversary of Gröbner Bases*, volume 77 of *Advanced Studies in Pure Mathematics*, pp. 171–205. Mathematical Society of Japan, 2018.

Hongyu Miao, Xiaohua Xia, Alan S. Perelson, and Hulin Wu. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, 53(1):3–39, 2011.

Siddhartha Mishra, Raffaele Molinaro, and Christoph Schwab. Estimates of uncertainty in physics-informed learning of PDEs. *Mathematics of Computation*, 91(334):1361–1403, 2022.

Yoh-ichi Mototake. Interpretable conservation law estimation by deriving the symmetries of dynamics from trained deep neural networks. *Physical Review E*, 103(3):033303, 2021.

Peter F. Pelz, Marc E. Pfetsch, Sebastian Kersting, Michael Kohler, Alexander Matei, Tobias Melz, Roland Platz, Maximilian Schaeffner, and Stefan Ulbrich. *Types of Uncertainty*, pp. 25–42. Springer International Publishing, Cham, 2021. ISBN 978-3-030-78354-9. doi: 10.1007/978-3-030-78354-9_2. URL https://doi.org/10.1007/978-3-030-78354-9_2.

Lawrence Principe. *The scientific revolution: A very short introduction*. Oxford University Press, 2011.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Alexandre René and André Longtin. Selecting fitted models under epistemic uncertainty using a stochastic process on quantile functions. *Nature Communications*, 16(1):9393, 2025.

M Sasaki, K Itoh, Ben F McMillan, T Kobayashi, H Arakawa, and J Chowdhury. Formation of density corrugations due to zonal flow in wave-kinetic framework. *Physics of Plasmas*, 28(11), 2021.

Makoto Sasaki, K Itoh, K Hallatschek, Naohiro Kasuya, M Lesur, Yusuke Kosuga, and S-I Itoh. Enhancement and suppression of turbulence by energetic-particle-driven geodesic acoustic modes. *Scientific Reports*, 7(1):16767, 2017.

Makoto Sasaki, Tatsuya Kobayashi, Kimitaka Itoh, Naohiro Kasuya, Yusuke Kosuga, A Fujisawa, and S-I Itoh. Spatio-temporal dynamics of turbulence trapped in geodesic acoustic modes. *Physics of Plasmas*, 25(1), 2018.

Ville Satopää, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a Kneedle in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, pp. 166–171. IEEE, 2011. doi: 10.1109/ICDCSW.2011.20.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

John Sylvester and Gunther Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Annals of Mathematics*, 125(1):153–169, 1987.

Peter Toth, Danilo Jimenez Rezende, Andrew Jaegle, Sébastien Racanière, Aleksandar Botev, and Irina Higgins. Hamiltonian generative networks. In *International Conference on Learning Representations (ICLR)*, 2020.

Silviu-Marian Udrescu and Max Tegmark. Ai feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

Steven Weinberg. *To Explain the World: The Discovery of Modern Science*. Harper, New York, 2015. ISBN 978-0062346650. See Chapter 11 for the discussion on Kepler's 8 minutes of arc discrepancy (p. 170) and the comparison of predictive accuracy between Copernican and Ptolemaic models (p. 153).

Sebastian J Wetzel, Roger G Melko, Joseph Scott, Maysum Panju, and Vijay Ganesh. Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks. *Physical Review Research*, 2(3):033499, 2020.

Masahiro Yamamoto. Carleman estimates for parabolic equations and applications. *Inverse Problems*, 25(12):123013, 2009.

Liu Yang, Xuhui Meng, and George Em Karniadakis. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.

Han Zhang, Huawei Fan, Liang Wang, and Xingang Wang. Learning hamiltonian dynamics with reservoir computing. *Physical Review E*, 104(2):024205, 2021.

A Proof of Theorem 1

Let $\delta\boldsymbol{\xi}$ be a perturbation of the discrete coefficient representation and let $\delta\mathbf{d} = \mathbf{T}\delta\boldsymbol{\xi}$ be the induced perturbation of the coefficient-feature vector. Because the reference coefficient already satisfies Eq. (3), a perturbed coefficient leaves the discretized governing equations unchanged if and only if

$$\mathbf{MT}\,\delta\boldsymbol{\xi} = \mathbf{0}.$$

Two coefficient perturbations that induce the same feature vector differ by an element of $\ker(\mathbf{T})$; after the explicitly stated gauge coordinates have been removed, this is the residual discrete gauge of the chosen representation.

Therefore feature-level identifiability modulo gauge means that every PDE-invisible admissible perturbation is already feature-invisible:

$$\mathbf{MT}\,\delta\boldsymbol{\xi} = \mathbf{0} \quad \Longrightarrow \quad \mathbf{T}\delta\boldsymbol{\xi} = \mathbf{0}.$$

Equivalently,

$$\ker(\mathbf{MT}) \subseteq \ker(\mathbf{T}).$$

The reverse inclusion always holds because $\mathbf{T}\delta\boldsymbol{\xi} = \mathbf{0}$ implies $\mathbf{MT}\delta\boldsymbol{\xi} = \mathbf{0}$. Hence identifiability is equivalent to

$$\ker(\mathbf{MT}) = \ker(\mathbf{T}).$$

By the rank–nullity theorem, both operators have the same domain $\mathbb{R}^p$, so equality of their nullities is equivalent to

$$\operatorname{rank}(\mathbf{MT}) = \operatorname{rank}(\mathbf{T}).$$

Finally, $\mathbf{T}\delta\boldsymbol{\xi}$ ranges exactly over $\mathcal{C} = \operatorname{Range}(\mathbf{T})$. A nonzero realizable coefficient-feature perturbation is invisible to the governing equations precisely when it belongs to $\ker(\mathbf{M}) \cap \mathcal{C}$. Thus

$$\ker(\mathbf{MT}) = \ker(\mathbf{T}) \quad \Longleftrightarrow \quad \ker(\mathbf{M}) \cap \mathcal{C} = \{\mathbf{0}\}.$$

This proves the equivalence of (i)–(iii). If $\mathbf{T} = \mathbf{I}$, the compatible subspace is the full algebraic feature space and the result reduces to the ordinary full-column-rank criterion. □

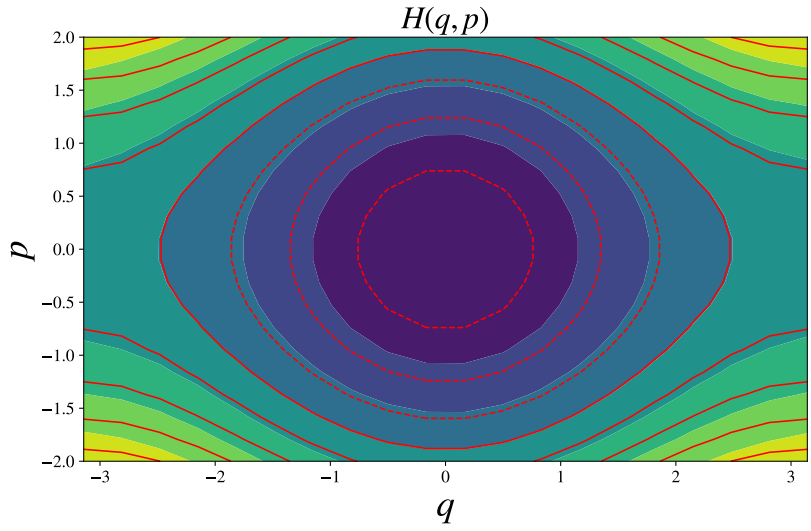


Figure 5: Hamiltonian example. The color map is the learned Hamiltonian and the red contour lines are the ground-truth Hamiltonian level sets.

B Controlled Hamiltonian Experiment

We consider a simple pendulum with $m = g = l = 1$,

$$H_{\text{true}}(q, p) = \frac{1}{2}p^2 + (1 - \cos q). \tag{11}$$

A 2–64–64–1 tanh network is trained using

$$\mathcal{L}_H = \frac{1}{N}\sum_i \left[(\partial_p H_\theta(q_i, p_i) - \dot{q}_i)^2 + (\partial_q H_\theta(q_i, p_i) + \dot{p}_i)^2\right]. \tag{12}$$

The experiment uses 50,000 training samples and 2,000 Adam iterations with learning rate 10^{-3} and mini-batch size 512.

As shown in Fig. 5, across five independent runs, the prediction MSE is 0.039460 ± 0.010941 and the mean-centered cosine similarity is 0.999776 ± 0.000056. The mean centering removes the unavoidable additive-constant gauge. This behavior is consistent with the full-rank structure of the canonical equations.

C Controlled Lagrangian Experiment and Constraint-Strength Selection

For the same pendulum,

$$L_{\text{true}}(\theta, \dot{\theta}) = \frac{1}{2}\dot{\theta}^2 - (1 - \cos\theta). \tag{13}$$

The observation operator for this controlled example supplies the generalized momentum p_i^{obs} and its time derivative $\dot{p}_i^{\text{obs}}$ independently of the candidate network L_ϑ. For the normalized pendulum used here, $m = l = 1$, so the synthetic observation channels satisfy

$$p_i^{\text{obs}} = \dot{\theta}_i, \qquad \dot{p}_i^{\text{obs}} = \ddot{\theta}_i.$$

Their numerical values therefore coincide with the velocity and acceleration channels in the generated dataset, but their role in the inverse problem is that of externally supplied momentum observations; they are never computed from L_ϑ.

Table 2: Lagrangian λ sweep. R_{NT} is computed by the generic numerical null–tangent algorithm; cosine is post-hoc only.

λ	Prediction MSE	Cosine	R_{NT}
0	0.076747 ± 0.021611	0.6474 ± 0.0969	11.900 ± 8.019
10^{-4}	0.076837 ± 0.021510	0.7700 ± 0.0209	3.305 ± 1.474
$\mathbf{10^{-3}}$	$\mathbf{0.076911 \pm 0.021777}$	0.9924 ± 0.0034	$\mathbf{2.051 \pm 0.897}$
10^{-2}	0.076980 ± 0.021704	0.99904 ± 0.00077	1.802 ± 0.766
10^{-1}	0.077331 ± 0.021907	$\mathbf{0.999750 \pm 0.000103}$	1.699 ± 0.682
10^{0}	0.077962 ± 0.022141	0.999688 ± 0.000085	1.484 ± 0.597
10^{1}	0.079662 ± 0.022388	0.999704 ± 0.000062	1.301 ± 0.512
10^{2}	0.080156 ± 0.022687	0.999732 ± 0.000048	0.925 ± 0.279
10^{3}	0.175400 ± 0.054382	0.9429 ± 0.0301	0.424 ± 0.124

We use the objective

$$\mathcal{L}_\lambda = \frac{1}{N} \sum_i \left[\left(\partial_\theta L_\vartheta(\theta_i, \dot{\theta}_i) - \dot{p}_i^{\mathrm{obs}} \right)^2 + \lambda \left(\partial_{\dot{\theta}} L_\vartheta(\theta_i, \dot{\theta}_i) - p_i^{\mathrm{obs}} \right)^2 \right]. \tag{14}$$

We use a 2–64–64–1 tanh architecture, 50,000 training samples, and 2,000 Adam iterations with learning rate 10^{-3} and mini-batch size 512. Because $p_i^{\mathrm{obs}} = \dot{\theta}_i$ and $\dot{p}_i^{\mathrm{obs}} = \ddot{\theta}_i$ in this normalized synthetic system, Eq. (14) is numerically identical to the objective used to produce the reported results.

The coefficient-derivative vector is ordered as

$$\mathbf{a} = (\partial_\theta L_1, \ldots, \partial_\theta L_N, \partial_{\dot{\theta}} L_1, \ldots, \partial_{\dot{\theta}} L_N)^\top.$$

The baseline residual is $\partial_\theta L_\vartheta - \dot{p}^{\mathrm{obs}}$. Accordingly, the numerically assembled equation matrix is algebraically equivalent to $\mathbf{M}_L = (\mathbf{I}_N \ \mathbf{0})$ and the right-hand side is independent of the candidate Lagrangian. The null space contains realizable perturbations, because adding a nonconstant function of $\dot{\theta}$ alone changes $\partial_{\dot{\theta}} L$ while leaving $\partial_\theta L$ unchanged. The code then computes $\mathbf{J}_a$ from the trained DNN and evaluates the generic numerical null–tangent overlap operator in Sec. G. The additional residual $\partial_{\dot{\theta}} L_\vartheta - p^{\mathrm{obs}}$ is the generalized-momentum constraint.

Across five runs, the tangent rank is 91.6 ± 5.1, the effective structural dimension $\sum_i \gamma_i$ is 52.48 ± 2.79, and the exact numerical intersection count is 27.6 ± 5.1. The continuous principal-angle weighting therefore retains the known structural geometry without relying on a Lagrangian-specific branch in the selection algorithm.

The mean ratio decreases from $R_{\mathrm{NT}} = 11.900 \pm 8.019$ at $\lambda = 0$ to 3.305 ± 1.474 at 10^{-4} and 2.051 ± 0.897 at 10^{-3}. The maximal positive diminishing-return score is $E = 0.801$, giving $\lambda^*_{\mathrm{Lagrangian}} = 10^{-3}$. At this value, prediction MSE is 0.076911 ± 0.021777 and post-hoc cosine is already 0.992389 ± 0.003381; the cosine remains close to unity from 10^{-3} through 10^2, a broad high-recovery regime that contains the numerical maximum.

D Controlled One-Dimensional Diffusion Experiment

We use the diffusion coefficient

$$a_{\mathrm{true}}(x) = 0.2 + 0.15\{1 + \sin(k\pi x)\}e^{-3x}, \qquad k \in \{1, 2, 3, 4, 5\}. \tag{15}$$

To handle the mixed occurrence of a and $\partial_x a$ consistently with Theorem 1, the unknown is represented by its nodal coefficient values $\mathbf{a}$. With the discrete first-derivative matrix $\mathbf{D}_x$, the coefficient-feature vector is

$$\mathbf{d}(a) = \begin{bmatrix} \partial_x a \\ a \end{bmatrix} \simeq \begin{bmatrix} \mathbf{D}_x \\ \mathbf{I} \end{bmatrix} \mathbf{a}.$$

At observed time t_r, the PDE therefore gives

$$[\mathrm{diag}(u_x(t_r))\mathbf{D}_x + \mathrm{diag}(u_{xx}(t_r))]\,\mathbf{a} = \mathbf{u}_t(t_r). \tag{16}$$

The numerical experiment uses multiple time levels, so these blocks are stacked into an overdetermined matrix. In the informative experiment the stacked matrix has full column rank on the retained points; no manual integration step or separately fitted integration constant is required.

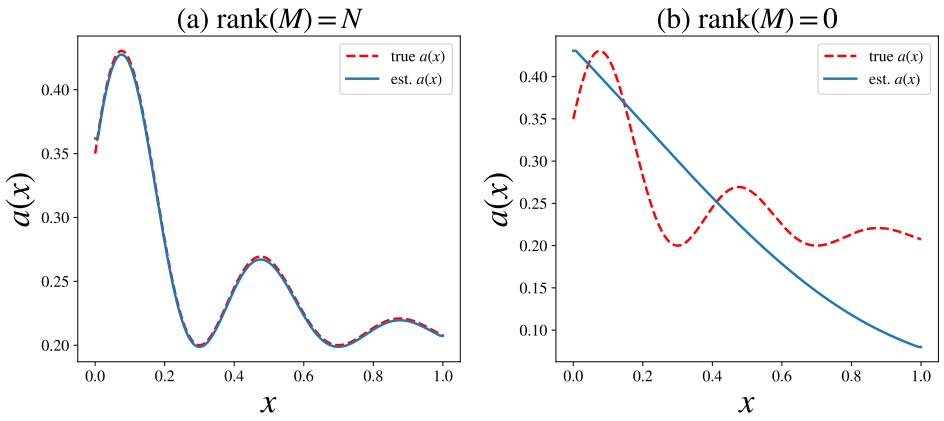


Figure 6: Diffusion example. Left: informative full-rank case. Right: zero-state data-degeneracy control. The red curve is the ground-truth diffusion coefficient.

We use a 1–64–64–1 tanh architecture and 4,000 Adam iterations with learning rate 2×10^{-3}. The full aggregated set of 160,801 retained spatiotemporal points is used in each optimization step. The loss is

$$\mathcal{L}_a = \frac{1}{N} \sum_i \left[\partial_t u - \{(\partial_x a_\theta)(\partial_x u) + a_\theta \partial_{xx} u\}\right]_i^2 . \tag{17}$$

For the informative initial condition $u(x, 0) = \sin(\pi x)$, the stacked operator has full rank on the retained informative points. For the control initial condition $u(x, 0) = 0$, the solution and all relevant derivatives vanish, so the loss is identically zero and the coefficient cannot be inferred. We classify this zero-state case as *data uncertainty caused by an uninformative observation*, not as structural uncertainty of the diffusion PDE.

As shown in Fig. 6, across the five k values, the informative case gives cosine similarity 0.999924 ± 0.000065 and prediction MSE $(7.03 \pm 13.47) \times 10^{-4}$. The zero-state control gives exactly zero prediction MSE but cosine similarity 0.0473 ± 0.9063. All dispersions in this paragraph are mean $\pm$ sample standard deviation over the five k values. This control illustrates that a vanishing physics residual does not imply coefficient recovery when the observations carry no information about the coefficient.

E DNN Model and Its Training Parameters

Here, we describe the DNN models and their training settings. In all cases we use fully connected multilayer neural networks. The controlled Hamiltonian, Lagrangian, and diffusion models consist of an input layer, two hidden layers, and an output layer; the wave-kinetic model uses the same fully connected form with the widths listed in Table 3. The activation function is the hyperbolic tangent,

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}. \tag{18}$$

Training uses Adam (Kingma & Ba, 2015). The controlled Hamiltonian and Lagrangian experiments use learning rate 10^{-3} and mini-batch size 512. The diffusion experiment uses learning rate 2×10^{-3} and the full aggregated training set in each optimization step. The training-set sizes, architectures, and iteration counts are summarized below.

The wave-kinetic architecture 2–100–10–1 intentionally uses a 10-unit second hidden layer; this is not a typographical abbreviation. For the wave-kinetic sweep, the spatial crop contains 10,000 points and is split 90/10 into train/test subsets with a fixed random state. Training uses mini-batches of size 8, initial learning rate 10^{-4}, weight decay 10^{-4}, and an adaptive learning-rate schedule. The learning rate is multiplied by 0.3 after three evaluation intervals without relative objective improvement of at least 10^{-3}, with a minimum value 10^{-6}; the objective is evaluated every 1,000 iterations. The same settings are used for all λ values and all ten time slices. Further implementation details are included in the accompanying anonymized reproduction bundle.

Table 3: DNN model and training parameters. Network widths are listed from input to output.

Case	Training size	Network	Iterations	Activation
Hamiltonian	50,000	2–64–64–1	2,000	tanh
Lagrangian	50,000	2–64–64–1	2,000	tanh
Diffusion	160,801	1–64–64–1	4,000	tanh
Wave-kinetic	10,000	2–100–10–1	40,000	tanh

F Details of the Uncertainty Evaluation in Wave-Kinetic Theory

This section gives the wave-kinetic Step 1 and Step 2 constructions in full detail. The numerical residual and matrix definitions below are those used in all reported experiments. The full source grid has 301 spatial points on $x \in [-8\pi, 8\pi]$ and 401 wavenumber points on $k_x \in [-20, 20]$, with spacings $\Delta x = 8\pi/150$ and $\Delta k_x = 0.1$. The learning data use a 100×100 crop with $x \in [-8.2100288, 8.3775804]$ and $k_x \in [-4.9, 5.0]$, evaluated at HNN time indices 158–167, corresponding to physical times 1590–1680.

The source simulation uses the generation–damping model

$$C(z) = \gamma_L(k_x) I(z) - \Delta\omega\, I(z)^2, \qquad \gamma_L(k_x) = \frac{k_y(k_x^2 + k_y^2)}{D(1 + k_x^2 + k_y^2)^3} \exp\left[-\left(\frac{k_x}{\Delta k} \right)^2 \right]. \tag{19}$$

The Hamiltonian in the coupled wave-kinetic/geodesic-acoustic-mode model is

$$H(z) = H_0 + \frac{k_y}{1 + k_x^2 + k_y^2} + k_y V_y(x, t), \tag{20}$$

and the geodesic-acoustic-mode field obeys (Sasaki et al., 2018)

$$\partial_t^2 V_y + \omega_G^2 V_y = \partial_t \partial_x^2 \int dk_x \, \frac{k_x k_y I(z)}{(1 + k_x^2 + k_y^2)^2} + \mu\, \partial_t \partial_x^2 V_y. \tag{21}$$

The simulation parameters are $k_y = 1$, $D = 3$, $\Delta k = 3$, $\omega_G = 0.1061$, $\mu = 0.05$, and the nonlinear damping coefficient used by the training code is $\Delta\omega = 0.4$. With these values, the training source term reduces exactly to

$$C = \gamma(k_x) I - 0.4 I^2, \qquad \gamma(k_x) = \frac{k_x^2 + 1}{3(2 + k_x^2)^3} \exp[-(k_x/3)^2]. \tag{22}$$

The turbulence intensity obeys $\partial_t I + \partial_{k_x} H\, \partial_x I - \partial_x H\, \partial_{k_x} I = C$. At a fixed time slice, the residual used in training is therefore

$$r_{ij} = \partial_t I^{(i,j)} - C^{(i,j)} + \partial_{k_x} H^{(i,j)} \partial_x I^{(i,j)} - \partial_x H^{(i,j)} \partial_{k_x} I^{(i,j)}. \tag{23}$$

The unconstrained discretization can be written as

$$\mathbf{c} = \mathbf{M}\mathbf{h}, \qquad \mathbf{h} = (\partial_x H^{(1,1)}, \partial_{k_x} H^{(1,1)}, \ldots)^\top, \tag{24}$$

with block-diagonal row structure

$$\mathbf{M} = \mathrm{diag}_{i,j} \begin{pmatrix} -\partial_{k_x} I^{(i,j)} & \partial_x I^{(i,j)} \end{pmatrix}, \tag{25}$$

so $\mathbf{M}$ has $N_x N_{k_x}$ rows and $2N_x N_{k_x}$ columns. The derivative vector $(\partial_x H, \partial_{k_x} H)$ therefore has twice as many components as the unconstrained scalar equations, giving the rank deficiency discussed in the main text.

Under the reflection constraint $H(x, -k_x) = H(x, k_x)$, paired points at $\pm k_x$ share the same $\partial_x H$ and opposite $\partial_{k_x} H$. The 2×2 block for each pair is

$$\begin{pmatrix} -\partial_{k_x} I^{(+)} & \partial_x I^{(+)} \\ -\partial_{k_x} I^{(-)} & -\partial_x I^{(-)} \end{pmatrix}. \tag{26}$$

Collecting the paired equations gives

$$\mathbf{c}_{\mathrm{pair}} = \mathbf{M}_{\mathrm{pair}} \mathbf{h}_{\mathrm{sym}}, \tag{27}$$

where the block-diagonal matrix is

$$\mathbf{M}_{\text{pair}} = \text{diag}_{i,j>0} \begin{pmatrix} -\partial_{k_x} I^{(+)} & \partial_x I^{(+)} \\ -\partial_{k_x} I^{(-)} & -\partial_x I^{(-)} \end{pmatrix}. \tag{28}$$

The determinant of each paired block is

$$D_{ij} = (\partial_{k_x} I^{(+)})(\partial_x I^{(-)}) + (\partial_x I^{(+)})(\partial_{k_x} I^{(-)}). \tag{29}$$

Thus the exact local rank condition is $D_{ij} \neq 0$; it is not the ordinary cross-product condition for linear independence of the two unmodified gradient vectors. Across the 49,000 retained paired interior blocks from the ten analyzed time slices, no determinant is zero. Full rank establishes local uniqueness but not uniform numerical stability: the spectral condition number has median 3.03, 95th percentile 23.99, 99th percentile 108.98, and maximum 2.84×10^4. Hence a small fraction of local blocks are strongly ill-conditioned. The learned Hamiltonian is therefore obtained from the global neural optimization rather than by independently inverting these local 2×2 systems.

F.1 Physical-time alignment and ground-truth Hamiltonian

The data-construction step drops the first grid index of the simulation output in time and space. Consequently, the time index i of the learning data corresponds to the physical time stored in the data array rather than to the raw simulation-grid index. All evaluations therefore match the learned and ground-truth fields by their physical time values.

The ground-truth Hamiltonian used for post-hoc comparison is evaluated on the same spatial and wavenumber grid as the learned model and is

$$H_{\text{true}}(x, k_x, t) = \frac{1}{2 + k_x^2} + V_y(x, t) + 0.02 + 0.02x. \tag{30}$$

The additive constant 0.02 is a gauge and does not affect the dynamics. Estimated and true Hamiltonians are mean-centered before both cosine similarity and the scale-sensitive relative L_2 error are computed.

F.2 Wave-kinetic λ selection by numerical null–tangent overlap

The wave-kinetic experiment evaluates ten consecutive time slices and $\lambda \in \{0, 10^{-7}, \ldots, 10^3\}$. For each slice, $\mathbf{M}$ is assembled numerically from the observed intensity derivatives on the evaluation grid; the selection code does not insert the analytic null vector of the local wave-kinetic row. The numerical null projector is intersected with the DNN derivative tangent as described in Sec. G.

The ratio decreases from $R_{\text{NT}}(0) = 10.125 \pm 2.410$ to 1.985 ± 0.216 at 10^{-1}, 1.391 ± 0.090 at 10^0, and 1.180 ± 0.057 at 10^1. The maximal positive diminishing-return score is $E = 0.760$, giving $\lambda^*_{\text{wave}} = 10^1$. Post-hoc mean-centered cosine similarity is 0.8450 ± 0.0256 at the selected value. Together with 0.8534 ± 0.0309 at 10^{-1} and 0.8685 ± 0.0303 at 10^0, this defines a broad high-recovery regime spanning approximately 10^{-1}–10^1. The scale-sensitive mean-centered relative L_2 error is 0.6907 ± 0.0268 at the selected 10^1, compared with 0.6238 ± 0.0289 at 10^{-1} and 0.6313 ± 0.0371 at 10^0. The ground-truth-free criterion therefore locates the same broad recovery regime without attempting to predict the exact post-hoc optimum of either ground-truth metric. Prediction MSE is the held-out PDE residual only, excluding the symmetry penalty, and has its numerical minimum at 10^{-4}.

G Numerical Null-Space and DNN-Tangent Implementation

This section gives the implementation used for both the Lagrangian and wave-kinetic λ selections. No analytic formula for a structural direction is passed to the code.

Algorithm 1 Ground-truth-free selection by numerical null–tangent overlap

Require: Equation matrix $\mathbf{M}$, DNN coefficient-derivative vector $\mathbf{a}_\theta$, physical and constraint residuals, log-spaced grid $\{\lambda_j\}$

1: Compute a numerical SVD of $\mathbf{M}$ and construct the numerical null projector $\mathbf{P}_N$.
2: Fit the reference network and compute $\mathbf{J}_a = \partial \mathbf{a}_\theta / \partial \theta$.
3: SVD $\mathbf{J}_a$ to obtain the derivative-active tangent $\mathbf{Q}_T$ and active parameter map $\mathbf{B}$.
4: Diagonalize $\mathbf{K} = \mathbf{Q}_T^\top \mathbf{P}_N \mathbf{Q}_T$ to obtain principal directions $\mathbf{v}_i$ and null-overlap weights γ_i.
5: Construct $\mathbf{G}_{\mathrm{phys}}$ and $\mathbf{G}_{\mathrm{con}}$ in the derivative-normalized active coordinates.
6: **for** each λ_j **do**
7: Compute $\mathbf{\Sigma}_{\lambda_j}$ and directional variances $s_i^2(\lambda_j)$.
8: Compute U_{struct}, U_{id}, and $R_{\mathrm{NT}}(\lambda_j)$ from the γ_i weights.
9: **end for**
10: Compute decade reductions Δ_j and elbow scores E_j; select λ^* by the diminishing-return rule above.

Numerical equation-level null space

On an evaluation grid, the coefficient derivatives are stacked into $\mathbf{a} \in \mathbb{R}^n$ and the governing equation into $\mathbf{M}\mathbf{a} = \mathbf{c}$. We compute a thin SVD

$$\mathbf{M} = \mathbf{U}_M \mathbf{S}_M \mathbf{V}_M^\top$$

and define the numerical rank using a relative singular-value threshold τ_M. If $\mathbf{Q}_R$ contains the retained right-singular vectors, the row and null projectors are

$$\mathbf{P}_R = \mathbf{Q}_R \mathbf{Q}_R^\top, \qquad \mathbf{P}_N = \mathbf{I} - \mathbf{P}_R.$$

The default experiments use $\tau_M = 10^{-8}$ relative to the largest singular value. This threshold is exposed in the code and can be changed for noisy applications.

Derivative-active DNN tangent

We compute the raw parameter Jacobian of the same coefficient-derivative vector,

$$\mathbf{J}_a = \frac{\partial \mathbf{a}_\theta}{\partial \theta},$$

and its SVD $\mathbf{J}_a = \mathbf{U}_a \mathbf{S}_a \mathbf{V}_a^\top$. Singular values below 10^{-3} of the leading singular value are discarded in the reported experiments. With retained rank r, $\mathbf{Q}_T = \mathbf{U}_{a,r}$ and $\mathbf{B} = \mathbf{V}_{a,r} \mathbf{S}_{a,r}^{-1}$, so that $\mathbf{J}_a \mathbf{B} = \mathbf{Q}_T$. This normalization removes raw parameter scaling and ensures that one unit of active-coordinate displacement corresponds to one unit in the local coefficient-derivative tangent norm. Additive constants in Hamiltonians or Lagrangians do not appear in $\mathbf{J}_a$ and are therefore removed automatically.

Numerical null–tangent overlap

Instead of assuming that every vector in $\ker(\mathbf{M})$ can be represented by the chosen DNN, we restrict the null projector to the DNN tangent:

$$\mathbf{K} = \mathbf{Q}_T^\top \mathbf{P}_N \mathbf{Q}_T. \tag{31}$$

Its eigendecomposition $\mathbf{K}\mathbf{v}_i = \gamma_i \mathbf{v}_i$ gives $0 \le \gamma_i \le 1$. The corresponding coefficient-derivative tangent direction is $\mathbf{q}_i = \mathbf{Q}_T \mathbf{v}_i$, and $\gamma_i = \|\mathbf{P}_N \mathbf{q}_i\|_2^2 = \cos^2 \phi_i$, where ϕ_i is the principal angle to the numerical null space. Therefore $\gamma_i = 1$ identifies an exact intersection direction, while intermediate values quantify near-null overlap without an extra binary cutoff.

The effective structural dimension is $d_{\mathrm{struct}}^{\mathrm{eff}} = \sum_i \gamma_i$ and the complementary identifiable dimension is $d_{\mathrm{id}}^{\mathrm{eff}} = r - d_{\mathrm{struct}}^{\mathrm{eff}}$. Figure 7 shows the resulting spectra. For wave-kinetic data the mean tangent rank is 93.3 ± 2.1, $d_{\mathrm{struct}}^{\mathrm{eff}} = 46.82 \pm 1.11$, and the numerical exact-intersection count is 29.3 ± 2.1. For the Lagrangian runs the corresponding values are 91.6 ± 5.1, 52.48 ± 2.79, and 27.6 ± 5.1.

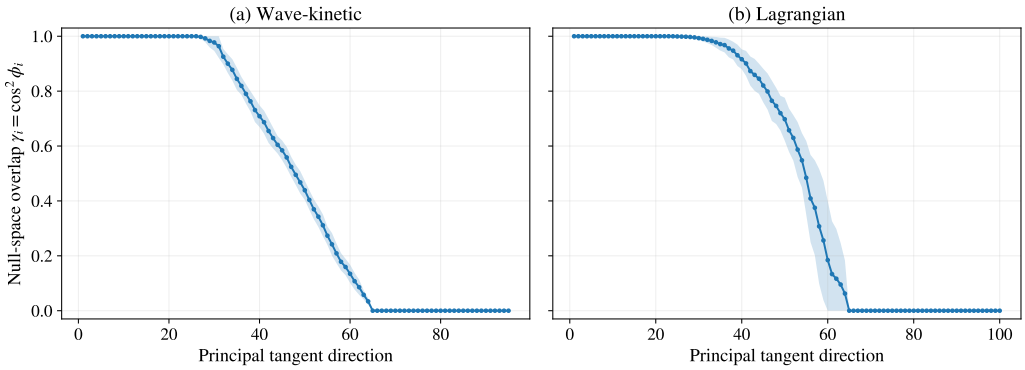


Figure 7: Numerical null–tangent overlap spectra. Each $\gamma_i = \cos^2 \phi_i$ measures how strongly a DNN coefficient-derivative tangent direction overlaps the numerical null space of $\mathbf{M}$. Curves show mean and standard deviation across ten wave-kinetic time slices and five Lagrangian runs.

Table 4: Finite-difference realization check for the five highest-overlap numerical null–tangent modes in the representative Lagrangian run (seed 85, $\epsilon = 10^{-3}$). The equation-row leakage is $\|\mathbf{Q}_R^\top \mathbf{q}_i\|_2^2$.

Mode	γ_i	FD cosine	Relative error	Row leakage^2
0	1.000000	0.99999829	1.85×10^{-3}	8.03×10^{-30}
1	1.000000	0.99999788	2.07×10^{-3}	5.28×10^{-30}
2	1.000000	0.99999959	9.03×10^{-4}	2.09×10^{-29}
3	1.000000	0.99999803	1.99×10^{-3}	4.97×10^{-30}
4	1.000000	0.99999831	1.84×10^{-3}	1.25×10^{-29}

FINITE-DIFFERENCE REALIZATION OF THE DNN TANGENT

The principal-angle construction is local, so we check that the corresponding tangent directions are realized by finite perturbations of the trained network rather than being artifacts of the raw Jacobian factorization. For the ith principal direction, define

$$\delta\theta_i = \mathbf{B}\mathbf{v}_i, \qquad \mathbf{q}_i = \mathbf{Q}_T \mathbf{v}_i,$$

and use the centered finite difference

$$\mathbf{q}_i^{\mathrm{FD}}(\epsilon) = \frac{\mathbf{a}_{\hat{\theta}+\epsilon\delta\theta_i} - \mathbf{a}_{\hat{\theta}-\epsilon\delta\theta_i}}{2\epsilon}. \tag{32}$$

We then compare $\mathbf{q}_i^{\mathrm{FD}}$ with the unit tangent vector $\mathbf{q}_i$ through their cosine similarity and relative error. This is a diagnostic only and is not used in the uncertainty ratio or in selecting λ.

For a representative Lagrangian run (seed 85), using $\epsilon = 10^{-3}$ in derivative-normalized active coordinates, the first five highest-overlap modes all have finite-difference cosine above 0.99999. Their relative linearization errors are of order 10^{-3} (Table 4). This confirms that the numerical null–tangent modes are locally realizable by the DNN parametrization.

CONSTRAINT-LIFTING DIAGNOSTIC ON THE STRUCTURAL TANGENT

A physical constraint should remove the ambiguity identified by the governing equation, rather than merely add curvature in unrelated DNN directions. Let $\mathbf{J}_{\mathrm{con}}$ denote the residual Jacobian of the proposed constraint and

$$\mathbf{K}_{\mathrm{con}} = \mathbf{J}_{\mathrm{con}}\mathbf{B}$$

its representation in derivative-normalized active coordinates. Rotating into the principal null–tangent coordinates gives $\mathbf{K}_{\mathrm{con}}\mathbf{V}$. For the hard-angle sensitivity set

$$\mathcal{S}_{0.99} = \{i : \cos\phi_i \geq 0.99\},$$

we report

$$\operatorname{rank}\left(\mathbf{K}_{\mathrm{con}}\mathbf{V}_{\mathcal{S}_{0.99}}\right) \tag{33}$$

relative to $|\mathcal{S}_{0.99}|$. Full rank means that the proposed constraint has local sensitivity to every structural mode in this thresholded numerical intersection. This diagnostic does not determine which physical constraint is scientifically appropriate; it only tests whether a stated constraint acts on the ambiguity that was diagnosed.

For Lagrangian seed 85, $|\mathcal{S}_{0.99}| = 33$ and the generalized-momentum constraint has rank $33/33$ on this structural tangent. The smallest and largest singular values of the restricted constraint map are 1.72×10^{-2} and 2.76, respectively. Thus the constraint locally lifts all thresholded structural modes in this run, consistent with the rank-restoration interpretation in the main text. The continuous R_{NT} calculation itself continues to use all γ_i weights rather than this hard threshold.

GGN/Laplace readout and λ selection

The physical and constraint residual Jacobians are projected with the same active parameter map $\mathbf{B}$, producing the generalized Gauss–Newton (GGN) curvature contributions $\mathbf{G}_{\mathrm{phys}}$ and $\mathbf{G}_{\mathrm{con}}$. We use the unconstrained optimized solution as the common local reference. For each tested λ,

$$\Sigma_\lambda = (G_{\mathrm{phys}} + \lambda G_{\mathrm{con}} + \eta I)^{-1}, \tag{34}$$

with an eigenvalue floor used only for numerical stabilization. For the wave-kinetic experiment the same fixed residual scales as in training are retained: 10^8 for the PDE term and 10^4 for the symmetry term.

Rotating the covariance into the principal null–tangent coordinates gives

$$\sigma_i^2(\lambda) = \mathbf{v}_i^\top \Sigma_\lambda \mathbf{v}_i.$$

We compute

$$U_{\mathrm{struct}}(\lambda) = \frac{\sum_i \gamma_i \sigma_i^2(\lambda)}{\sum_i \gamma_i}, \qquad U_{\mathrm{id}}(\lambda) = \frac{\sum_i (1-\gamma_i)\sigma_i^2(\lambda)}{\sum_i (1-\gamma_i)},$$

and $R_{\mathrm{NT}} = U_{\mathrm{struct}}/U_{\mathrm{id}}$. In the ideal exact-subspace limit, γ_i are binary and the formula becomes an ordinary mean over the exact intersection and its tangent complement. The continuous weighting is used in the paper because it remains stable when structural directions are only approximately represented.

For each logarithmic λ grid we use the diminishing-return score defined in the main text. Ground-truth cosine similarity is excluded from selection. As a sensitivity analysis, the code also reports a hard-intersection ratio using $\cos\phi_i \geq 0.99$; it selects the same $\lambda^* = 10^{-3}$ for the Lagrangian system and $\lambda^* = 10^1$ for the wave-kinetic system.

H Video of Hamiltonian Estimation Results

Time-series animations of the wave-kinetic Hamiltonian estimates are included as separate GIF files. The file `appendix_media/with_constraint.gif` shows the estimates obtained with the symmetry constraint, and `appendix_media/without_constraint.gif` shows the corresponding estimates without the constraint. These animations provide the time-resolved counterpart of the representative snapshots shown in Fig. 4(a,b).

I Detailed λ-Sweep Tables

For compactness, the main text keeps the λ-selection figures, while the full numerical sweep values are reported in Tables 5 and 6.

Table 5: Lagrangian constraint-strength sweep (mean ± standard deviation over five runs). R_{NT} uses the numerical null space of $\mathbf{M}$ and its principal-angle overlap with the DNN coefficient-derivative tangent. Cosine similarity is post-hoc only.

λ	Prediction MSE	Cosine similarity	$R_{\mathrm{NT}}(\lambda)$
0	0.076747 ± 0.021611	0.6474 ± 0.0969	11.900 ± 8.019
10^{-7}	0.076771 ± 0.021612	0.7026 ± 0.0645	11.722 ± 7.790
10^{-6}	0.076762 ± 0.021704	0.7148 ± 0.0212	10.488 ± 6.312
10^{-5}	0.076790 ± 0.021530	0.6382 ± 0.0620	6.684 ± 3.103
10^{-4}	0.076837 ± 0.021510	0.7700 ± 0.0209	3.305 ± 1.474
$\mathbf{10^{-3}}$	$\mathbf{0.076911 \pm 0.021777}$	0.9924 ± 0.0034	$\mathbf{2.051 \pm 0.897}$
10^{-2}	0.076980 ± 0.021704	0.999040 ± 0.000769	1.802 ± 0.766
10^{-1}	0.077331 ± 0.021907	$\mathbf{0.999750 \pm 0.000103}$	1.699 ± 0.682
10^{0}	0.077962 ± 0.022141	0.999688 ± 0.000085	1.484 ± 0.597
10^{1}	0.079662 ± 0.022388	0.999704 ± 0.000062	1.301 ± 0.512
10^{2}	0.080156 ± 0.022687	0.999732 ± 0.000048	0.925 ± 0.279
10^{3}	0.175400 ± 0.054382	0.9429 ± 0.0301	0.424 ± 0.124

Table 6: Wave-kinetic constraint-strength sweep (mean ± standard deviation over ten time slices). R_{NT} is computed from the numerical null space of $\mathbf{M}$ intersected with the DNN coefficient-derivative tangent. Cosine similarity and relative L_2 error are post-hoc and are not used by the selection algorithm.

λ	Prediction MSE	Cosine similarity	Relative L_2 error	$R_{\mathrm{NT}}(\lambda)$
0	$(1.5291 \pm 0.0990) \times 10^{-7}$	0.3868 ± 0.0672	0.9741 ± 0.0842	10.125 ± 2.410
10^{-7}	$(1.5291 \pm 0.0990) \times 10^{-7}$	0.3868 ± 0.0672	0.9741 ± 0.0842	10.122 ± 2.409
10^{-6}	$(1.5291 \pm 0.0990) \times 10^{-7}$	0.3870 ± 0.0671	0.9740 ± 0.0841	10.092 ± 2.398
10^{-5}	$(1.5291 \pm 0.0989) \times 10^{-7}$	0.3888 ± 0.0669	0.9724 ± 0.0834	9.815 ± 2.298
10^{-4}	$\mathbf{(1.5289 \pm 0.0981) \times 10^{-7}}$	0.4061 ± 0.0649	0.9568 ± 0.0766	8.076 ± 1.735
10^{-3}	$(1.5550 \pm 0.1110) \times 10^{-7}$	0.5369 ± 0.0756	0.8534 ± 0.0483	5.006 ± 0.927
10^{-2}	$(1.6484 \pm 0.1684) \times 10^{-7}$	0.7813 ± 0.0227	0.6895 ± 0.0348	3.140 ± 0.489
10^{-1}	$(1.6595 \pm 0.1053) \times 10^{-7}$	0.8534 ± 0.0309	$\mathbf{0.6238 \pm 0.0289}$	1.985 ± 0.216
10^{0}	$(1.8779 \pm 0.2220) \times 10^{-7}$	$\mathbf{0.8685 \pm 0.0303}$	0.6313 ± 0.0371	1.391 ± 0.090
$\mathbf{10^{1}}$	$(2.2173 \pm 0.1441) \times 10^{-7}$	0.8450 ± 0.0256	0.6907 ± 0.0268	$\mathbf{1.180 \pm 0.057}$
10^{2}	$(2.7218 \pm 0.2191) \times 10^{-7}$	0.7996 ± 0.0349	0.7694 ± 0.0310	1.130 ± 0.060
10^{3}	$(5.0133 \pm 0.4247) \times 10^{-7}$	-0.2160 ± 0.2435	1.1128 ± 0.0427	1.097 ± 0.069

J Evaluation Metric

For coefficient functions with an additive gauge, cosine similarity is evaluated after mean centering. For a grid-evaluated estimate f and ground truth f_{true},

$$\cos(f, f_{\mathrm{true}}) = \frac{\langle f - \bar{f},\, f_{\mathrm{true}} - \bar{f}_{\mathrm{true}} \rangle}{\|f - \bar{f}\|_2 \, \|f_{\mathrm{true}} - \bar{f}_{\mathrm{true}}\|_2}. \tag{35}$$

This metric is used only for post-hoc synthetic-data diagnostics and never for choosing λ^*.

For amplitude-sensitive recovery, we additionally report the mean-centered relative L_2 error

$$E_{\mathrm{rel}} = \frac{\|(f - \bar{f}) - (f_{\mathrm{true}} - \bar{f}_{\mathrm{true}})\|_2}{\|f_{\mathrm{true}} - \bar{f}_{\mathrm{true}}\|_2}. \tag{36}$$

Unlike cosine similarity, E_{rel} penalizes an incorrect multiplicative scale while remaining invariant to the additive gauge.

For the wave-kinetic experiment, "prediction MSE" denotes the mean squared PDE residual on the held-out 10% test subset. It excludes the symmetry penalty and is reported before multiplication by the fixed 10^8 training normalization factor. Thus the predictive metric is not mechanically reduced by decreasing the soft-constraint weight.

K Comparison with Bayesian PINNs on the Wave-Kinetic System

Posterior-based uncertainty quantification, as in Bayesian PINNs (B-PINNs) (Yang et al., 2021), does not by itself evaluate structural uncertainty. The posterior spread combines structural, data,

model-form, approximation, and sampling contributions without separating them, and a structurally non-identifiable direction appears only as a flat direction of the likelihood, whose extent is set by the prior and by how far the sampler explores it. Functionally extended degrees of freedom such as an arbitrary function $F(I)$ are therefore difficult to represent with finitely many posterior samples. In contrast, the rank criterion of Theorem 1 identifies such directions directly from the PDE and the observed state through $\ker(\mathbf{M}) \cap \mathcal{C}$.

To illustrate this, we applied a B-PINN to the wave-kinetic system of Sec. 5.2 for six time slices ($t = 170$–175), sampling the posterior with Hamiltonian Monte Carlo (2,000 samples after 1,000 burn-in steps, 20 leapfrog steps per sample, initialized from the trained deterministic network); the step size was tuned so that the average acceptance rate remained between 0.7 and 0.95. The credible intervals of the estimated Hamiltonian remained finite and did not reveal the known structural ambiguity $H = H_{\mathrm{true}} + F(I)$, which leaves the unconstrained residual unchanged for any F because $\partial_{k_x} F(I)\, \partial_x I - \partial_x F(I)\, \partial_{k_x} I = 0$.

We estimated the number of structurally uncertain directions from the B-PINN posterior by computing the posterior covariance of $\big(H(x^{(1)}, k_x^{(1)}), \ldots, H(x^{(N)}, k_x^{(N)})\big)$ and counting singular values above a threshold. This count refers to the degrees of freedom of H itself, whereas the proposed criterion counts those of $(\partial_x H, \partial_{k_x} H)$, so the comparison is qualitative (Table 7). For $I \equiv 0$, $\mathbf{M}$ is the zero matrix and H is unconstrained, so the expected count for H is $N_x N_{k_x} = 10^4$; 10,000 posterior samples were used in this case.

Table 7: Number of structurally uncertain directions diagnosed by the proposed rank criterion (derivative space) and estimated from the B-PINN posterior covariance of H (function values) for different singular-value thresholds. B-PINN entries are mean $\pm$ standard deviation over the six time slices; entries for $I \equiv 0$ are approximate.

Method	Without constraint	With symmetry constraint	$I \equiv 0$
Proposed (nullity of $\mathbf{M}$)	$N_x N_{k_x} = 10^4$	0	$2N_x N_{k_x} = 2 \times 10^4$
B-PINN (threshold $= \infty$)	0	0	0
B-PINN (threshold $= 1.0$)	0.67 ± 0.47	0.33 ± 0.47	≈ 9990
B-PINN (threshold $= 0.01$)	9.5 ± 0.96	2.0 ± 0.0	≈ 9999

Three observations follow. First, the proposed diagnosis is consistent with theory in all cases, whereas the B-PINN count depends strongly on the post-hoc threshold. Second, in the unconstrained case, where H has the functional ambiguity $F(I)$, the B-PINN posterior reports only a few uncertain directions, so the structural non-identifiability is not detected. Third, under the symmetry constraint a small nonzero count remains, and the posterior alone cannot tell whether it reflects structural ambiguity or finite-sample and approximation effects. B-PINNs and the proposed framework are therefore complementary: B-PINNs quantify posterior uncertainty, whereas the rank criterion diagnoses the structural origin and dimension of non-uniqueness from the PDE itself. From a Bayesian viewpoint, structural uncertainty induces highly extended flat directions that may require more advanced inference schemes. This comparison is exploratory; its code is not part of the reproduction bundle.