

Herglotz Lagrangian neural networks: Learning dissipative dynamics via contact variational principles

Qi Huang ¹, Jack Murdoch Moore ^{1,2}, Ting-Ting Gao ^{3,*} and Gang Yan ^{1,2,†}

¹MOE Key Laboratory of Advanced Micro-Structured Materials, and School of Physical Science and Engineering, *Tongji University*, Shanghai 200092, People's Republic of China

²National Key Laboratory of Autonomous Intelligent Unmanned Systems, MOE Frontiers Science Center for Intelligent Autonomous Systems, *Tongji University*, Shanghai 200092, People's Republic of China

³Network Science Institute, *Northeastern University*, Boston, Massachusetts 02115, USA



(Received 12 January 2026; accepted 10 May 2026; published 29 May 2026)

Physics-informed machine learning has proven effective for learning conservative dynamical systems. However, most real-world systems are nonconservative, and current approaches typically handle dissipation through *ad hoc* modifications or require access to full-state information. Here, we introduce the discrete Herglotz Lagrangian neural network (HLNN), a structure-preserving framework for learning a class of smooth dissipative dynamics from data. Grounded in the Herglotz variational principle, HLNN extends classical variational learning to dissipative systems by modeling dissipation intrinsically within a contact-geometric framework, without prescribing an explicit empirical dissipation law, and can be trained from position-only observations without explicit velocity supervision. Across benchmarks spanning state-dependent damping, time-modulated damping, and external driving, HLNN reliably recovers long-term trajectories and dissipation rates, outperforming baseline models on the considered benchmarks. Our results establish a unified variational and geometric framework for learning smooth dissipative dynamics and open avenues for data-driven modeling of nonequilibrium and open systems.

DOI: [10.1103/9gnh-89jd](https://doi.org/10.1103/9gnh-89jd)

I. INTRODUCTION

Understanding the dynamical mechanisms underlying complex systems is a long-standing central problem in science and engineering [1–4]. Data-driven approaches have made substantial progress toward this goal. In particular, machine learning provides a highly flexible modeling paradigm with universal approximation capabilities [5], enabling the modeling of dynamical systems, even the discovery of explicit governing equations directly from measured trajectories or state observations [6–9]. However, despite their expressive power, unconstrained neural models often suffer from limited interpretability, less physical consistency, and degraded generalization, especially in data-sparse or extrapolative regimes [10]. Growing evidence suggests that embedding physical constraints, such as conservation laws and symmetries, serves as a powerful inductive bias that substantially enhances both interpretability and generalization [11–17]. Representative examples include Hamiltonian neural networks (HNNs) [18] and Lagrangian neural networks (LNNs) [19], which parametrize

the Hamiltonian or Lagrangian and recover system dynamics through Hamilton's equations or the Euler-Lagrange equations. By construction, these models preserve geometric structure and conservation laws, leading to improved long-term stability and physical interpretability compared to black-box architectures. Related hard-constraint frameworks, such as SympNets [20] and GFNNs [13], further demonstrate the advantages of incorporating variational and geometric principles directly into network design.

Despite these successes, most real-world dynamical systems are intrinsically nonconservative. Irreversible processes such as friction, plastic deformation, thermal dissipation, and external driving are ubiquitous in natural and engineered systems [21–24]. Extending structure-preserving learning frameworks beyond conservative dynamics, therefore, remains a fundamental challenge. Recent works have attempted to incorporate dissipation into Hamiltonian- or Lagrangian-based neural networks, including dissipative Hamiltonian neural networks (DHNNs) [25], port-Hamiltonian neural networks (PHNNs) [26], and generalized Lagrangian neural networks (GLNNs) [27]. These approaches typically introduce dissipation through Rayleigh dissipation potentials, damping matrices, or external forcing terms appended to otherwise conservative equations of motion [28]. While effective on specific benchmarks, such constructions involve important geometric and modeling trade-offs. First, dissipation is introduced in an *ad hoc* manner as external disturbances on the symplectic manifold, rather than emerging from a unified variational principle. These modifications generally

*Contact author: ti.gao@northeastern.edu

†Contact author: gyan@tongji.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

do not preserve the original symplectic structure; when the geometric structure itself is important to the modeling task, this can reduce the reliability of long-term predictions [29]. Although some studies maintain the symplectic structure by doubling the manifold's dimension to represent nonconservative dynamics, they introduce auxiliary variables without direct physical interpretation [30,31].

Second, many existing approaches rely on prescribed damping ansatzes; in particular, some formulations employ constant or Rayleigh-type damping [25,26], whereas more generalized formulations relax this assumption by learning nonconservative forces directly. When the damping law of the system is known and relatively simple, Rayleigh-type formulations can provide a natural and straightforward modeling choice. In contrast, when the dissipation law is unknown, state dependent, or explicitly time modulated, Herglotz-type formulations become more natural. This is especially true when the modeling objective is not only to fit trajectories, but also to provide an intrinsic variational and geometric description of smooth dissipation and to interpret the dissipative mechanism itself within the model.

Third, many existing formulations rely on explicit velocity or momentum information during training, which can limit applicability in sparse or partially observed settings [18,25]. Although recent work has explored learning Hamiltonians in generalized coordinates [32], velocity information remains implicitly or explicitly involved in many existing formulations.

To address these limitations, we propose a learning framework based on the Herglotz variational principle and contact geometry. This principle generalizes the classical variational formulation for conservative systems to the dissipative regime by naturally incorporating dissipation via the Lagrangian's dependence on the action variable [33,34]. Building on this theoretical foundation, we introduce discrete Herglotz Lagrangian neural networks (HLNNs). The model parametrizes the Herglotz Lagrangian with neural networks and is trained by minimizing the residual of the discrete Herglotz Euler-Lagrange equations [33,34]. Within this framework, dissipation emerges directly from the underlying contact geometry rather than being appended as a phenomenological penalty to a conservative model. Specifically, HLNN adopts the discrete Herglotz variational principle as its foundational structure. The training objective minimizes the residual of the resulting discrete Herglotz-Euler-Lagrange equations, which structurally embeds the correct dissipative dynamics into the learned model. This construction ensures structural consistency with contact geometry and is capable of learning a broad class of smooth dissipative behaviors. Importantly, in the present implementation, HLNN can be trained using only generalized-coordinate observations, without explicit velocity supervision. We validate HLNN on diverse dissipative systems, including oscillators with state-dependent damping, time-modulated damping, externally driven systems, and chaotic systems. Our results demonstrate that HLNN outperforms the tested baselines in predicting trajectories and energy dissipation rates, even when trained on limited position-only data. Furthermore, we conducted additional investigations of the model's robustness performance under degraded observation quality, including enlarged

sampling intervals, noisy position-only observations, and trajectories with missing points reconstructed by interpolation. The corresponding results provide evidence that HLNN remains competitive under these degraded observation settings and performs particularly well in reconstructing dissipative structure. In addition to recovering the system dynamics, the model also provides a transparent route for analyzing and diagnosing energy dissipation within the learned contact variational framework.

The main contributions of this work are as follows:

- (1) We bridge the gap between deep learning and contact geometry by introducing HLNNs, enabling intrinsic and geometrically consistent learning of smooth dissipative dynamics.
- (2) We demonstrate that HLNNs can learn smooth dissipative mechanisms from position-only data without explicit velocity supervision, and we further examine their behavior under enlarged sampling intervals, noisy observations, interpolation-recovered missing data, and controlled multi-scale decay.
- (3) By explicitly parametrizing the Herglotz Lagrangian, our model provides a transparent route to estimate the instantaneous energy dissipation rate. This offers a physically interpretable metric for diagnosing dissipative behavior in dynamical systems.

The theoretical foundations and HLNN architecture are presented in Sec. II and Appendix. In Sec. III, we report numerical experiments on a range of damped and forced oscillators, comparing HLNNs with existing baselines. Finally, we discuss the limitations of the proposed framework, its broader physical implications, and potential directions for future research.

II. METHODS

A. Theory

1. Contact geometry and dissipative dynamics

The contact-geometric formulation of dissipative mechanics provides the geometric setting for the Herglotz variational principle [33,35,36]. Specifically, contact geometry is defined on the extended phase space $M = T^*Q \times \mathbb{R}$, where Q denotes the configuration space with generalized coordinates q^i , T^*Q is the associated cotangent bundle with canonical momenta p_i conjugate to q^i , and the additional variable $z \in \mathbb{R}$ represents an actionlike scalar that encodes nonconservative effects such as dissipation or energy injection. Here, the index i labels the degrees of freedom of the system. In local coordinates (q^i, p_i, z) , the canonical contact form is $\eta = dz - p_i dq^i$. Given a contact Hamiltonian $H(q, p, z)$, the induced dynamics are governed by the contact Hamiltonian equations:

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i} - p_i \frac{\partial H}{\partial z}, \quad \dot{z} = p_i \frac{\partial H}{\partial p_i} - H. \quad (1)$$

Notably, the nonconserved Hamiltonian H evolves as

$$\dot{H} = -H \frac{\partial H}{\partial z}, \quad (2)$$

which provides a geometric mechanism for energy dissipation ($\partial H / \partial z > 0$) or gain ($\partial H / \partial z < 0$). While contact geometry provides a Hamiltonian and phase-space description of

dissipative dynamics, an equivalent formulation exists at the level of variational principles, given by the Herglotz variational principle.

2. Herglotz variational principle

Continuous principle. The Lagrangian counterpart of contact geometry is the Herglotz variational principle [33,34,37]. Instead of defining action as a time integral, the Herglotz action variable $z(t)$ evolves according to

$$\dot{z} = \mathcal{L}(t, q, \dot{q}, z), \quad z(0) = z_0. \quad (3)$$

Stationarity of the terminal value $z(T)$ yields the generalized Euler-Lagrange equations

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i} \right) - \frac{\partial \mathcal{L}}{\partial q^i} - \frac{\partial \mathcal{L}}{\partial z} \frac{\partial \mathcal{L}}{\partial \dot{q}^i} = 0. \quad (4)$$

Defining the generalized energy as $E = \frac{\partial \mathcal{L}}{\partial \dot{q}^i} \dot{q}^i - \mathcal{L}$, we derive the generalized-energy balance law

$$\dot{E} = \left(\frac{\partial \mathcal{L}}{\partial z} \right) E, \quad (5)$$

from Eq. (4), where $\partial \mathcal{L} / \partial z$ directly quantifies the instantaneous relative rate of generalized-energy change: Negative values correspond to dissipation, whereas positive values correspond to energy injection. This balance law can be viewed as a special case of a more general Noether-type picture for Herglotz systems, in which the symmetry-related quantity is typically not an ordinary strictly conserved quantity, but rather a suitably rescaled generalized quantity [33–35,37].

Discrete principle. To enable data-driven learning from discrete trajectories, the continuous Herglotz variational principle must be discretized in a structure-preserving manner. Here, we adopt the discrete Herglotz variational principle [33,38,39]. Given a discrete curve $\{q_k\}_{k=0}^N$ with step size h , the action sequence $\{z_k\}$ is defined by

$$z_{k+1} - z_k = L_d(q_k, q_{k+1}, z_k), \quad (6)$$

where $L_d(q_k, q_{k+1}, z_k)$ denotes the discrete Herglotz Lagrangian, namely, a one-step approximation of the continuous Herglotz Lagrangian $L(q, \dot{q}, z)$. In the present work, we use the midpoint discretization

$$L_d(q_k, q_{k+1}, z_k) = h L(\bar{q}_k, \bar{\dot{q}}_k, z_k), \quad (7)$$

$$\bar{q}_k = \frac{q_k + q_{k+1}}{2}, \quad \bar{\dot{q}}_k = \frac{q_{k+1} - q_k}{h}.$$

Thus, L_d depends on two neighboring discrete configurations and the action state z_k , rather than on the continuous variables $(q, \dot{q}, z)$.

Stationarity of the discrete action yields the discrete Herglotz-Euler-Lagrange (DHEL) equations [38], for $k = 1, \dots, N-1$:

$$D_1 L_d(q_k, q_{k+1}, z_k) + D_2 L_d(q_{k-1}, q_k, z_{k-1}) \times (1 + D_3 L_d(q_k, q_{k+1}, z_k)) = 0, \quad (8)$$

where $D_i L_d$ denotes the partial derivative of L_d with respect to its i th argument. Further details can be found in Sec. III of Supplemental Material [40]. From a numerical implementation perspective, introducing the action variable z leads, in

the present HLNN framework, to a contact variational update structure in which the action state must be propagated recursively along the trajectory, while the next configuration is obtained implicitly through the DHEL equations, making the method computationally heavier than explicit baseline models.

B. Model and training overview

The overall architecture and training pipeline of the proposed HLNN are illustrated in Fig. 1. The model takes sequences of observed generalized coordinates (q_{k-1}, q_k, q_{k+1}) as input and initializes the action variable with $z_0 = 0$. Using a midpoint approximation, the network internally constructs velocities and evaluates a parametrized Herglotz Lagrangian of the form

$$L(q, \dot{q}, z) = \frac{1}{2} \dot{q}^T M \dot{q} - V_\theta(q) + \gamma_\theta(q, \dot{q})z, \quad (9)$$

where the potential energy V_θ and dissipation term γ_θ are represented by neural networks, and $M = mI$, where m is a learnable positive mass parameter. Following Ref. [33], the Herglotz term is written as $+\gamma_\theta(q, \dot{q})z$. Accordingly, dissipative behavior corresponds to regions where $\gamma_\theta(q, \dot{q}) < 0$. To ensure physical interpretability and numerical stability, we avoid a black-box representation of the Herglotz Lagrangian. Instead, we introduce the structured parametrization in Eq. (9), which corresponds to a first-order Taylor expansion around the conservative limit $z = 0$. In this formulation, the action z is an internal state variable initialized at $z_0 = 0$, not a standard network input.

Computationally, z_k is explicitly evaluated from the preceding tuple (q_{k-1}, q_k, z_{k-1}) via Eq. (6), and the implicit numerical solve is strictly confined to the d -dimensional configuration space for q_{k+1} . The inherent algorithmic cost of HLNN therefore stems not from higher-dimensional root finding, but from the necessary sequential propagation of z , the calculation of the solvability factor $\Delta_k = 1 + \partial L_d / \partial z_k$, and the associated regularization constraints.

Training is performed end to end by minimizing a physics-informed objective derived from the DHEL equations. Specifically, the loss function takes the form

$$\mathcal{J}_{\text{total}} = \frac{w_{\text{phys}}}{N_\tau(L-1)} \sum_{\tau=1}^{N_\tau} \sum_{k=1}^{L-1} \mathcal{J}_{\text{phys}}(q_{k-1}, q_k, q_{k+1}, z_{k-1}, z_k) + \frac{w_{\text{reg}}}{N_\tau(L-1)} \sum_{\tau=1}^{N_\tau} \sum_{k=1}^{L-1} \mathcal{J}_{\text{reg}}(q_k, q_{k+1}, z_k), \quad (10)$$

where the physics loss $\mathcal{J}_{\text{phys}}$ is

$$\mathcal{J}_{\text{phys}}(q_{k-1}, q_k, q_{k+1}, z_{k-1}, z_k) = \|D_1 L_d(q_k, q_{k+1}, z_k) + D_2 L_d(q_{k-1}, q_k, z_{k-1}) \times (1 + D_3 L_d(q_k, q_{k+1}, z_k))\|^2, \quad (11)$$

and D_1 , D_2 , and D_3 denote partial derivatives of the learned discrete Lagrangian with respect to its arguments, which correspond to the DHEL equation (8). All derivatives are computed via automatic differentiation. The regularization term is

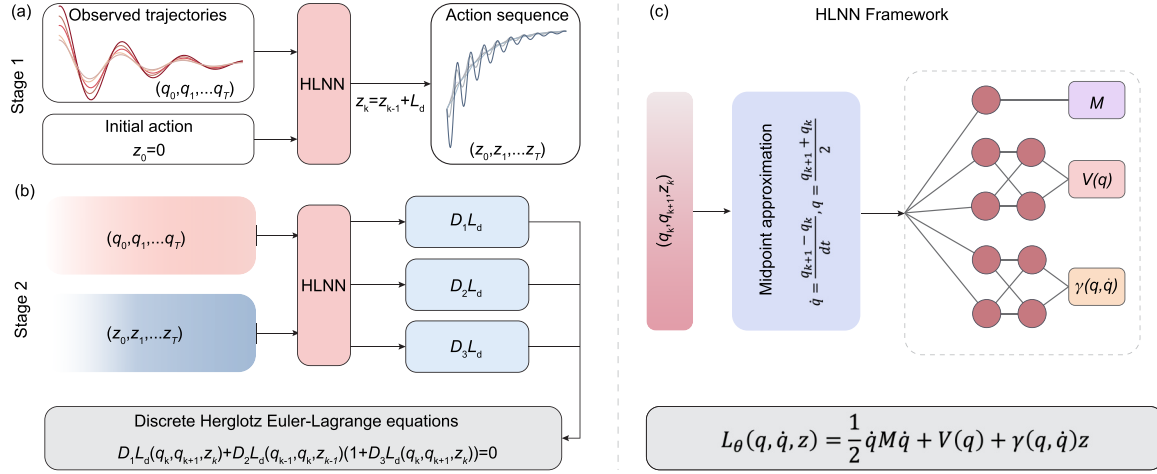


FIG. 1. Schematic architecture of the discrete Herglotz Lagrangian neural network. The workflow illustrates the data processing from observed coordinates to the physics-informed loss computation. (a) Stage 1: The model takes a sequence of observed coordinates $(q_0, q_1, \dots, q_T)$ as input and initializes the action variable $z_0 = 0$. The action sequence is then updated recursively according to Eq. (6). (b) Stage 2: Given the observed coordinates and the corresponding action sequence, HLNN computes the residual of the DHLE equations (bottom) as physics loss to update the model parameters. (c) HLNN framework: The HLNN core evaluates the learned continuous Herglotz Lagrangian L_θ , where θ denotes the trainable network parameters, and constructs its midpoint-discretized counterpart L_d . The learned Lagrangian is structurally decomposed into kinetic, potential, and dissipation-related components, while the action variable is initialized and propagated recursively through the discrete Herglotz update. The entire network is trained end to end by minimizing the residual of the DHLE (bottom), computed via automatic differentiation (D_1, D_2, D_3) of the learned discrete Lagrangian.

defined as

$$\mathcal{J}_{\text{reg}} = \beta(\tau_{\min} - \sigma_{\min}(H_k)) + \beta(\delta_{\min} - |\Delta_k|), \quad (12)$$

which enforces the regularity and solvability conditions required for the discrete Herglotz update to define a well-posed dynamical system. Here, $\beta(x)$ denotes the Softplus function, H_k is the mixed derivative matrix of the discrete Lagrangian, $\sigma_{\min}(H_k)$ denotes its smallest singular value, and Δ_k is the discrete contact factor. The constants $\tau_{\min}$ and $\delta_{\min}$ are safety thresholds used to penalize near-singular or ill-posed updates. The physics term enforces consistency with the DHLE equations, while the regularization term promotes the regularity and solvability conditions required for the implicit discrete update. Together, they ensure that the learned model remains structurally consistent and well posed. Further details of these quantities and their implementation are provided in Appendix 4; additional training details are given in Sec. VIII of Supplemental Material [40].

After training, the learned HLNN is used for prediction by recursively solving the implicit DHLE equations from two initial positions, enabling long-horizon rollouts using position-only data. In a system with d generalized coordinates, each rollout step requires solving a d -dimensional implicit nonlinear system for the next configuration. Our current implementation employs Newton iterations, where the computational cost naturally scales with the system dimension. Model performance is evaluated by comparing predicted trajectories and energy dissipation rates against ground truth. The trajectory prediction error is quantified by the mean squared error (MSE),

$$\text{MSE}_{\text{rollout}} = \frac{1}{N_\tau L} \sum_{\tau=1}^{N_\tau} \sum_{k=0}^{L-1} \|q_k^{(\tau)} - \hat{q}_k^{(\tau)}\|^2. \quad (13)$$

To assess whether the learned model captures dissipative behavior, we further compare the positive intrinsic dissipation rate induced by the learned Herglotz model. For HLNN, we define the generalized energy

$$E_\theta(t) = \dot{q}^i \frac{\partial \mathcal{L}_\theta}{\partial \dot{q}^i} - \mathcal{L}_\theta \quad (14)$$

and evaluate the corresponding positive intrinsic dissipation rate as

$$D_{\text{diss}}^\theta(t) = -\left(\frac{\partial \mathcal{L}_\theta}{\partial z}\right) E_\theta(t). \quad (15)$$

Here, dissipative behavior corresponds to $\partial \mathcal{L}_\theta / \partial z < 0$, so that D_{diss}^θ is positive on the dissipative benchmarks considered in this work. In practice, velocities are approximated via midpoint discretization. The dissipation error is measured by

$$\text{MSE}_{\text{diss}} = \frac{1}{N_\tau L} \sum_{\tau=1}^{N_\tau} \sum_{k=0}^{L-1} (D_{\text{diss}}^{\text{pred}}(\tau, k) - D_{\text{diss}}^{\text{true}}(\tau, k))^2, \quad (16)$$

where $D_{\text{diss}}^{\text{pred}}(\tau, k)$ and $D_{\text{diss}}^{\text{true}}(\tau, k)$ denote the predicted and ground-truth positive intrinsic dissipation rates evaluated at the k th time step along the τ th trajectory. All theoretical derivations, numerical solvers, regularization details, and evaluation protocols are provided in the Appendix and Secs. III–VIII of Supplemental Material [40].

III. RESULTS

We evaluate our model on a suite of dynamical systems spanning increasing levels of complexity, ranging from simple damped harmonic oscillators to nonlinear systems with time-modulated dissipation and externally driven dynamics. Model performance is assessed using two complementary metrics: the mean squared error of rollout trajectories and the MSE

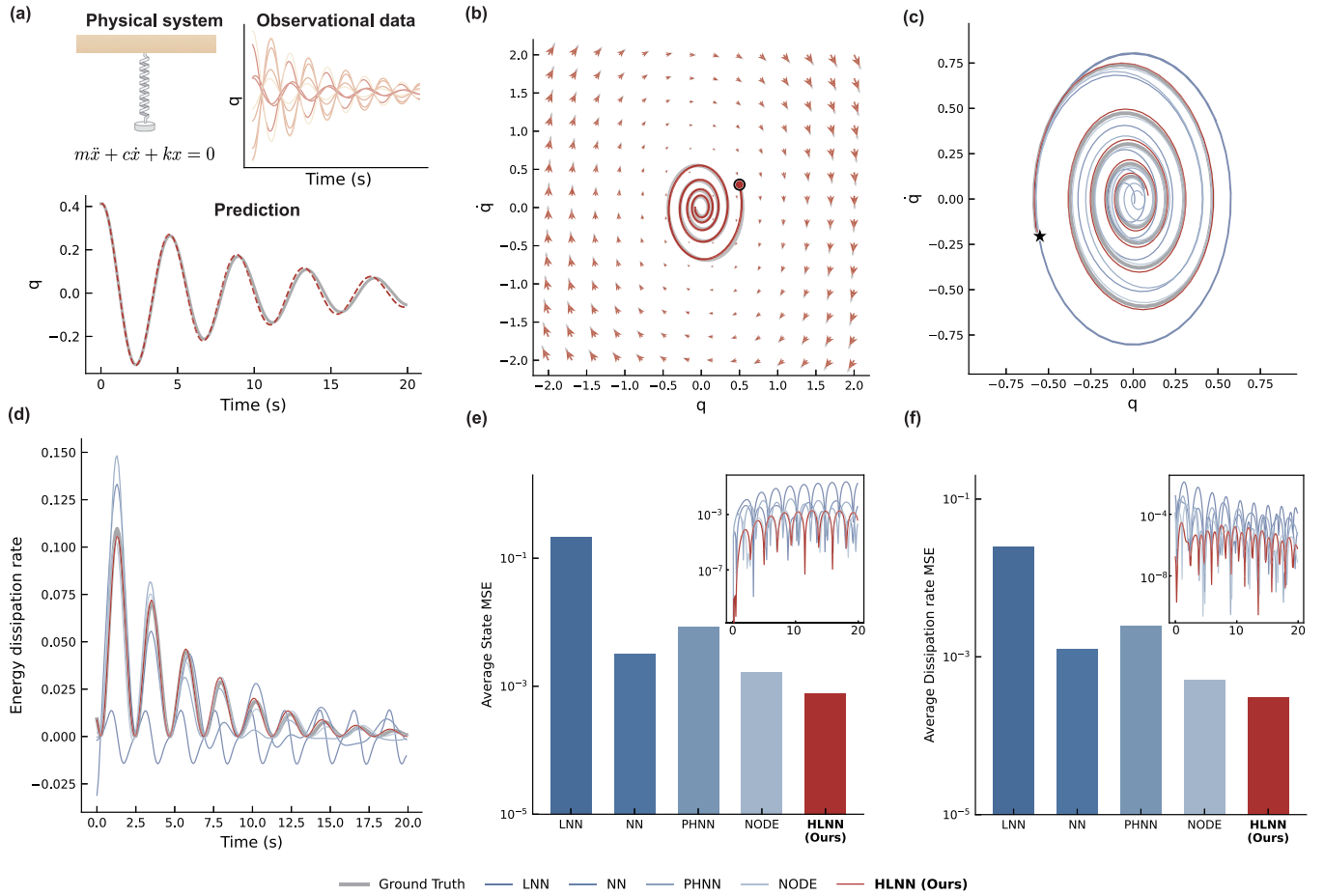


FIG. 2. Reconstruction of dynamics and energy dissipation for the damped harmonic oscillator. (a) Top: Schematic of the physical system and observed displacement time series $q(t)$. Bottom: Comparison of the predicted trajectory from the same initial condition, showing HLNN (red dashed) vs the ground truth (gray solid). (b) Comparison of the ground truth (gray) and learned (red) vector fields (arrows) and their corresponding phase-space trajectories; the red dot marks the initial condition. Comparison of reconstructed dynamics for a randomly selected trajectory from the test set: (c) phase-space trajectory (velocity vs position) and (d) time evolution of the instantaneous energy dissipation rate. The initial state is marked by a black star. HLNN predictions (red) closely match the ground truth (gray). Overall quantitative summary: (e) state MSE and (f) dissipation rate MSE, averaged over the full time horizon and across all test trajectories. Insets show the temporal evolution of the corresponding average errors.

of the energy dissipation rate, both averaged over 25 unseen initial conditions sampled in the same manner as the training data. Because absolute MSE can understate errors in the late low-amplitude regime of strongly decaying trajectories, we additionally performed a supplementary controlled multiscale decay analysis; the corresponding relative-error and low-energy diagnostics are reported in Sec. IX of Supplemental Material [40]. Unless stated otherwise, the available observations consist only of position measurements $q(t)$. For rollout comparisons, all methods are initialized from the same pair of consecutive observed positions, so that the midpoint-based initialization error is shared across methods. Notably, we compare our model against the following baseline architectures: standard feedforward neural networks (NNs), Lagrangian neural networks [19], port-Hamiltonian neural networks (PHNNs) [26], and neural ordinary differential equations (neural ODEs; NODEs) [41]. Additionally, we further examined HLNN on a strongly chaotic driven damped pendulum; the corresponding setup and results are provided in Sec. X of Supplemental Material [40].

A. Damped harmonic oscillator

We begin our analysis with a classical dissipative system, the damped harmonic oscillator. This obeys the equation of motion:

$$m\ddot{q} + c\dot{q} + kq = 0, \quad (17)$$

where m denotes the mass, c is the damping coefficient, and k is the spring constant.

Training. We set $k = 2$, $m = 1$, and $c = 0.2$ for our data generation. We randomly sampled initial training conditions $[q_0, \dot{q}_0]$ with $q_0 \in [-2, 2]$ and $\dot{q}_0 \in [-0.5, 0.5]$. We evolved each initial state using the fourth-order Runge-Kutta (RK4) integrator with a time step of $\Delta t = 0.1$ over a duration of $T_{\max} = 20.01$.

Testing. We initiate our evaluation with a classical one-degree-of-freedom damped harmonic oscillator. As summarized in Fig. 2, our method yields more accurate predictions for both state trajectories and energy evolution compared to the baseline architectures. Specifically, Figs. 2(c) and 2(d)

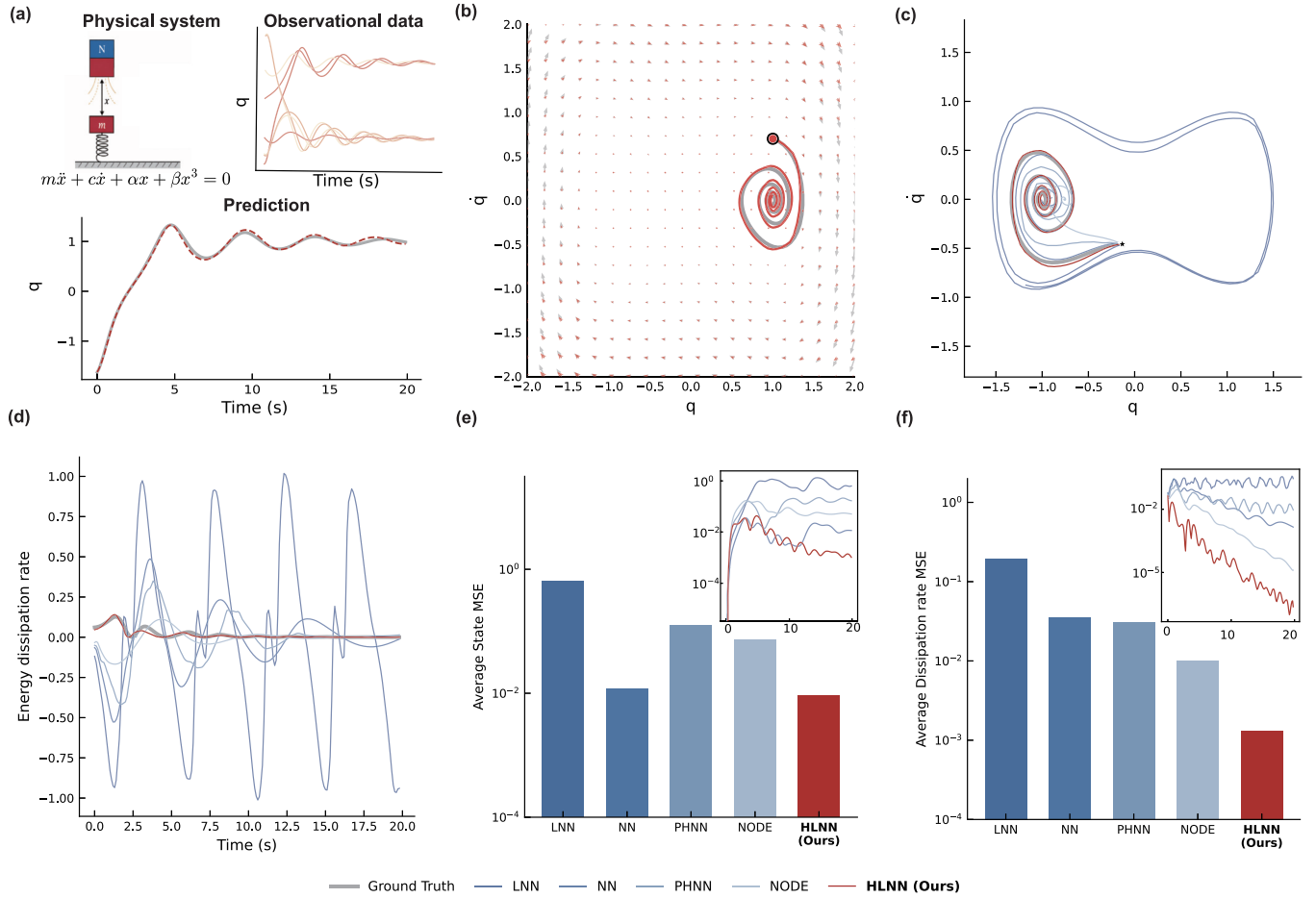


FIG. 3. Results for the damped double-well oscillator. (a) Top: Schematic of the physical system and observed displacement time series $q(t)$. Bottom: Rollout prediction of $q(t)$ from the same initial condition, comparing HLNN (red dashed) to ground truth (gray solid). (b) Comparison of ground truth (gray) vs learned (red) vector fields and their corresponding trajectories in phase space; the red dot marks the initial condition. Comparison of reconstructed dynamics for a randomly selected test trajectory: (c) phase-space trajectory, with the initial state marked by a black star, and (d) temporal evolution of the instantaneous energy dissipation rate. HLNN predictions (red) demonstrate close agreement with the ground truth (gray). Quantitative performance summary: (e) average state MSE and (f) average dissipation rate MSE, computed over the full time domain across all test trajectories. Insets display the temporal evolution of the corresponding average errors.

illustrate the reconstructed state and dissipation rate for a randomly selected initial condition from the test set. Figures 2(e) and 2(f) present the overall MSE of state and dissipation rate, while the insets show the time evolution of the corresponding MSE. These results indicate that, among all methods considered, HLNN attains the lowest errors on this benchmark, accurately recovering the energy dissipation rate with a mean squared error below 10^{-3} . As an additional structure-oriented diagnostic, we further examine, on this benchmark, the generalized Noether-type quantity associated with the Herglotz variational principle. The corresponding definitions are provided in Appendix 2 and Sec. V of Supplemental Material [40]. The compensated quantity remains close to its reference value throughout the rollout interval, indicating that HLNN preserves an internal evolution pattern approximately consistent with the learned discrete Herglotz structure.

B. Damped oscillator with a double-well potential

After validating HLNN on a damped harmonic oscillator with a purely quadratic potential, we further consider a system

with strongly nonlinear stiffness: a damped oscillator with a double-well potential. This system can be described by the equation of motion:

$$m\ddot{q} + c\dot{q} + \alpha q + \beta q^3 = 0, \quad (18)$$

where α and β control the linear and nonlinear stiffness, respectively, and c is the damping coefficient. A classical physical system that can be modeled by this equation is a bistable oscillator with magnetic interaction [42]. Compared with the damped harmonic oscillator, this system exhibits bistability and a much richer phase-space structure, allowing us to test whether the model can recover a nonlinear energy landscape and the associated dissipation law purely from data.

Training. We sample initial states uniformly with $q_0 \in [-2, 2]$ and $\dot{q}_0 \in [-0.5, 0.5]$ and fix the parameters $m = 1$, $\alpha = -1$, $\beta = 1$, $c = 0.3$. Each initial state is evolved forward in time using an RK4 integrator with a time step of $\Delta t = 0.1$ up to $T_{\max} = 20.01$.

Testing. We observe from Fig. 3 that our proposed model successfully reconstructs the underlying dynamics. Notably,

while NN performs similarly to HLNN in recovering dynamical states, HLNN is superior to NN in reproducing the energy dissipation rate. This discrepancy arises because the double-well potential features a quartic nonlinearity, whereas baseline models implicitly assume a purely quadratic potential when reconstructing the dissipation rate from trajectory data (see Sec. VI of Supplemental Material [40]). Consequently, although the NN is able to recover the state trajectories, it cannot simultaneously learn the correct dissipation law. While HLNN does not fully recover the exact energy dissipation rate, possibly due to the inherent identifiability challenge where the decomposition between conservative and dissipative forces is not unique from position-only data, it nonetheless accurately captures the evolution of the energy dissipation rate. Unlike the baseline models, within the Herglotz parametrization, HLNN can learn an interpretable internal dissipation mechanism from data without committing in advance to a specific empirical damping formula.

C. Nonlinear damped oscillator

Furthermore, we extend our evaluation to systems with state-dependent damping. In practice, the damping term is often not constant, but varies with the velocity or displacement. It is therefore crucial to assess whether a model can capture such state-dependent dissipation mechanisms. To this end, we consider an oscillator with nonlinear damping as a test system. Its equation of motion can be written as

$$m\ddot{q} + c_1\dot{q} + c_2|\dot{q}|\dot{q} + kq = 0, \quad (19)$$

where m denotes the mass, k is the spring stiffness, and c_1 and c_2 are the coefficients of linear and quadratic damping, respectively. This model can be regarded as a simplified description of dynamical systems subject to both air (or fluid) drag and internal friction, such as a mass-spring system with bearing friction operating in air. In our implementation, PHNN is constructed under the assumption of a constant damping coefficient [26]. As a result, its structural prior is inherently mismatched with such state-dependent dissipation, and we therefore expect PHNN to perform poorly on this system.

Training. Initial conditions are sampled uniformly with $q_0 \in [-2, 2]$ and $\dot{q}_0 \in [-1, 1]$. We set $k = m = 1$ and damping coefficients $c_1 = 0.3$, $c_2 = 0.1$. Trajectories are integrated using an RK4 scheme with time step $\Delta t = 0.1$ up to $T_{\max} = 20.01$.

Testing. Figure 4 indicates that when damping is nonconstant, the structural priors imposed by PHNN limit its ability to capture the underlying dynamics, as it enforces a constant damping coefficient that lacks the expressivity to model state-dependent dissipation. In contrast, HLNN is not constrained by a predetermined damping form. It naturally generalizes from fixed damping to state-dependent and time-varying dynamics, including both time-dependent damping and external forcing (see Secs. III D and III E). As shown in Fig. 4(d), HLNN recovers the system's energy dissipation rate solely from data, without relying on explicit damping assumptions.

D. Time-modulated damped oscillator

We consider a time-modulated damped oscillator as a further example. The classical equation of motion for this system can be written as

$$m\ddot{q} + c(t)\dot{q} + \alpha q = 0, \quad (20)$$

where m denotes the mass and α is the stiffness coefficient. In contrast to the previous examples, the dynamics of this system depend explicitly on time. Consequently, we include the time variable as part of the model input for both training and evaluation.

Training. We sample the initial states $[q_0, \dot{q}_0]$ randomly from the ranges $q_0 \in [-1.5, 1.5]$ and $\dot{q}_0 \in [-1, 1]$. We use $\alpha = 1$ and $m = 1$, and we set $c(t) = 0.2 + 0.5 \sin(0.2t)$. Each initial state is evolved forward in time using the RK4 integrator with a time step of $\Delta t = 0.1$ up to $T_{\max} = 20.01$. The position data $q(t)$ and time t are provided to model as inputs.

Testing. As shown in Fig. 5, HLNN exhibits higher accuracy and more stable long-term predictions than the other models. Notably, Fig. 5(d) further indicates that HLNN does not perfectly reproduce the detailed energy dissipation behavior. Consistent with the observations in the double-well oscillator system, HLNN exhibits minor deviations in the instantaneous dissipation rate. However, these local discrepancies do not compromise the model's ability to accurately capture the overall energy evolution trend and HLNN remains the lowest-error method among all methods considered on this benchmark.

E. Forced damped oscillator

For the final experiment, we investigate a forced dissipative system whose equation of motion is given by

$$m\ddot{q} + c\dot{q} + \alpha q = F(t), \quad (21)$$

where m is the mass, α is the stiffness coefficient, c denotes the damping coefficient, and $F(t)$ represents an external time-dependent driving force.

Training. We sample the initial states uniformly from the ranges $q_0 \in [-1.5, 1.5]$ and $\dot{q}_0 \in [-1, 1]$. We set $\alpha = 1$, $c = 0.2$, and $m = 1$. The external driving force is given by

$$F(t) = 0.5 \sin(0.8t).$$

Each initial state is evolved forward in time using the RK4 integrator with a time step of $\Delta t = 0.1$ up to $T_{\max} = 20.01$. Only the position $q(t)$ and the time t are provided to the model as inputs.

Testing. The comparative results, summarized in Fig. 6, show that HLNN attains the lowest trajectory and dissipation-rate errors among the tested baselines on this externally driven system. A notable observation is that while the HLNN does not yield a perfect pointwise reconstruction of the energy dissipation rate, it successfully captures the essential dynamic trends. Overall, the experiments confirm that the HLNN framework extends naturally to externally driven systems and accurately recovers the governing dissipative mechanisms.

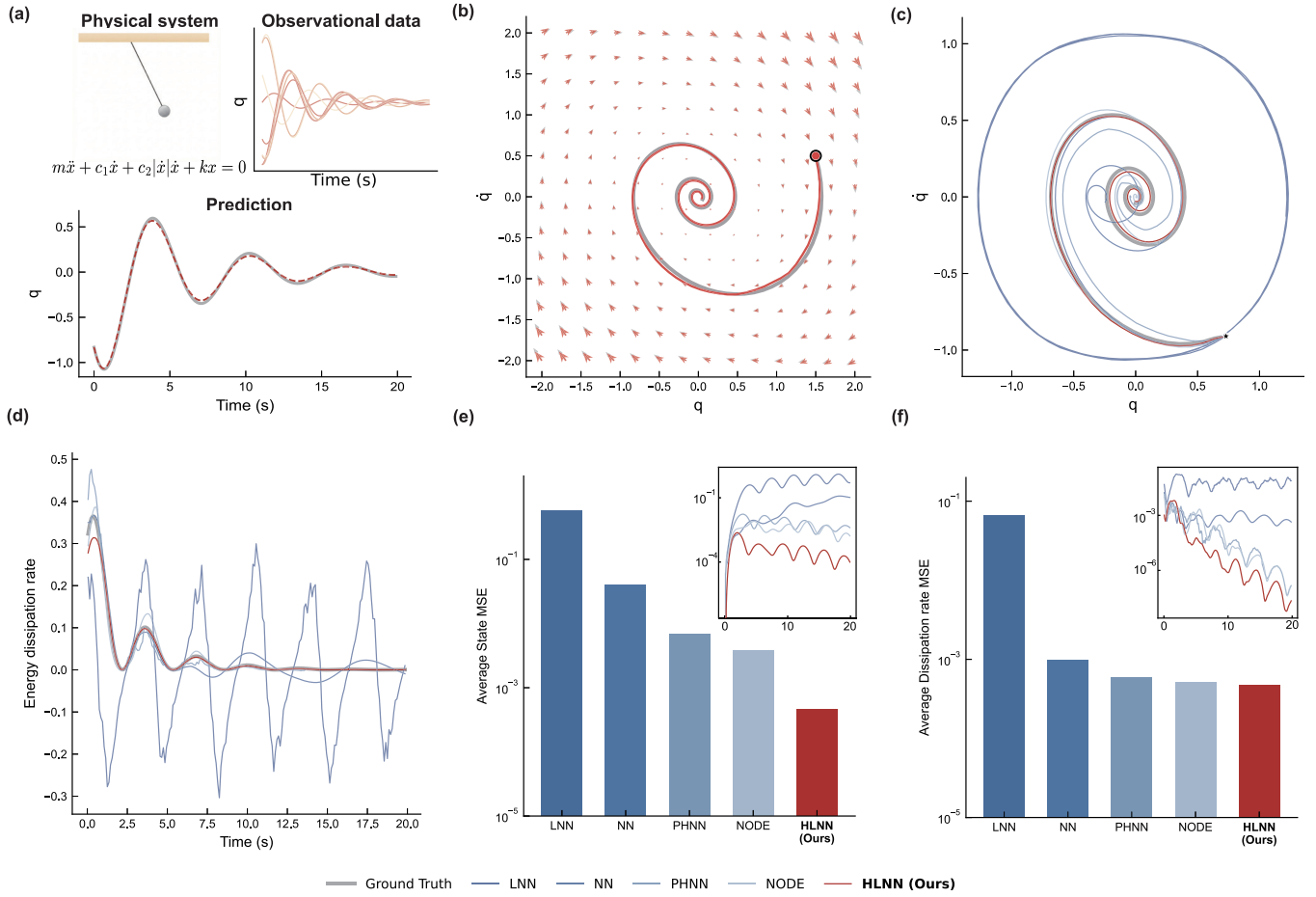


FIG. 4. Nonlinear damped oscillator. (a) Top: Schematic of the physical system and the observed displacement time series $q(t)$. Bottom: Rollout prediction of $q(t)$ from the same initial condition, comparing HLNN (red dashed) to ground truth (gray solid). (b) A visualization comparing the ground truth (gray) and learned (red) vector fields (arrows) alongside their respective trajectories; the red dot indicates the initial condition. Comparison of reconstructed dynamics for a randomly selected test trajectory: (c) phase-space trajectory, with the initial state marked by a black star, and (d) temporal evolution of the instantaneous energy dissipation rate. The HLNN predictions (red) demonstrate close agreement with ground truth (gray). Quantitative summary: (e) average state MSE and (f) average dissipation rate MSE, computed over the full time domain across all test trajectories. Insets display the temporal evolution of the corresponding average errors.

F. Robustness

To further assess the robustness of HLNN beyond the standard benchmark setting, we performed additional tests on the damped harmonic oscillator under three degraded observation conditions: enlarged sampling intervals, noisy position-only observations, and interpolation-recovered missing observations. The results are summarized in Fig. 7.

Large sampling intervals. To examine the effect of sparse temporal sampling, we generated additional datasets with enlarged sampling intervals and retrained all models under the corresponding settings. For comparison, we report results for

$$\Delta t \in \{0.1, 0.2, 0.3, 0.5, 0.7\}.$$

As shown in Figs. 7(a) and 7(d), HLNN attains the lowest errors among the tested methods across all considered step sizes. Its trajectory-rollout and dissipation-rate errors remain below those of the baselines considered and vary only mildly as Δt increases, indicating that the learned discrete Herglotz structure remains effective even under sparse temporal sampling.

Noisy position-only observations. To evaluate the effect of imperfect measurements, we added zero-mean Gaussian noise to the position observations in the training data:

$$q_k^{\text{obs}} = q_k + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma^2 I),$$

with noise levels

$$\sigma \in \{0.0001, 0.001, 0.005, 0.01, 0.05\}.$$

All methods were then trained using the corresponding noisy position observations under their original training pipelines. Figures 7(b) and 7(e) show that HLNN achieves the best or near-best performance in the low-to-moderate noise regime. At the highest tested noise level, however, HLNN is no longer uniformly optimal, and NODE attains lower errors on this benchmark.

A possible explanation is related to the discrete multipoint structure of HLNN. Rather than learning only a one-step map, HLNN is trained by enforcing the DHREL relation (8), so the shared discrete variational structure can act as a physically consistent regularizer when observations are sufficiently clean. At the same time, this structure also introduces

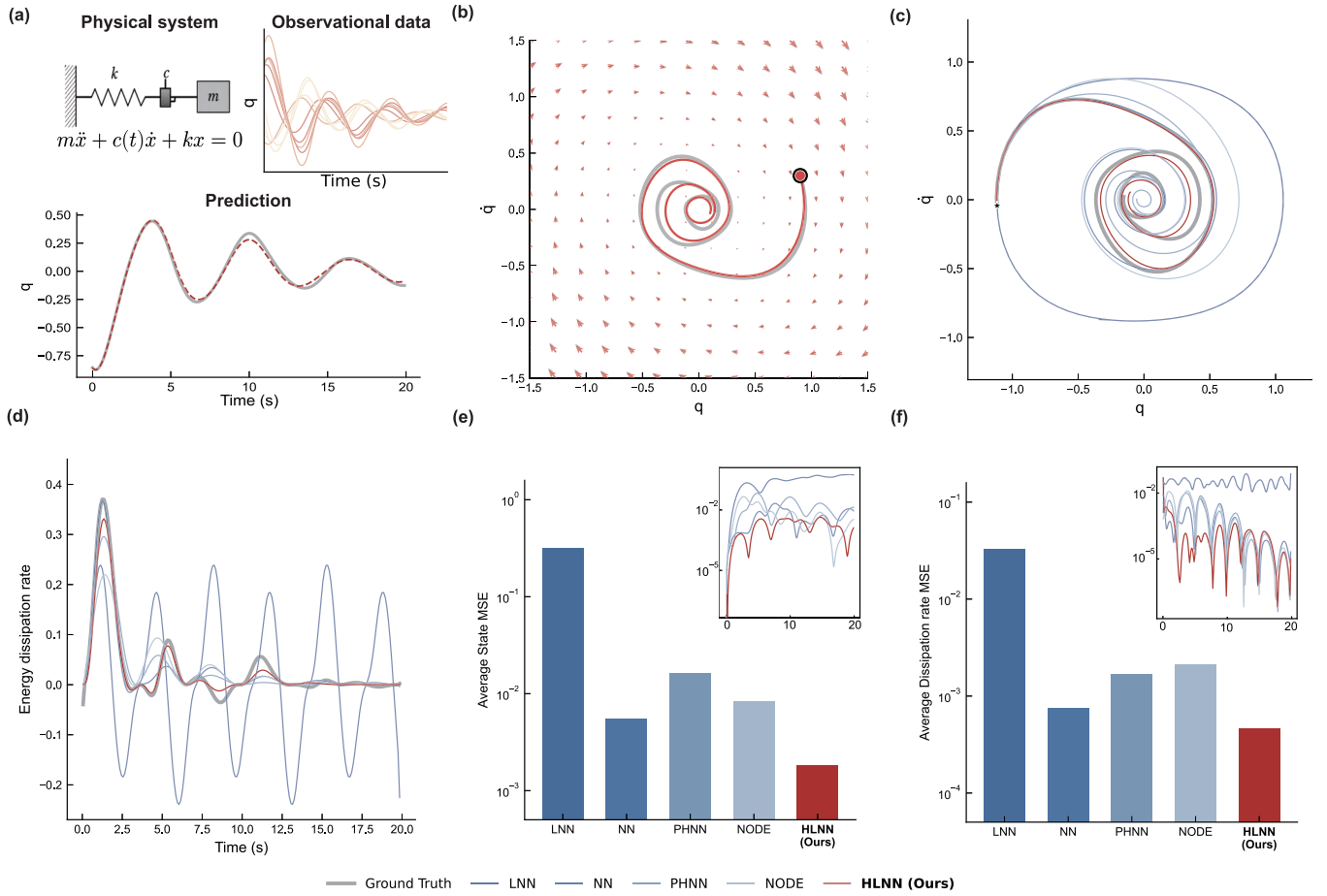


FIG. 5. Evaluation on a more complex dynamical system: the time-modulated damped oscillator. (a) Top: Schematic of the physical system and observed trajectories $q(t)$. Bottom: Rollout prediction of $q(t)$ from the same initial condition, comparing HLNN (red dashed) to ground truth (gray solid). (b) Ground truth (gray) and learned (red) vector fields (arrows) in phase space, overlaid with their respective trajectories; the red dot marks the initial condition. Comparison of reconstructed dynamics for a randomly selected test trajectory: (c) phase-space trajectory and (d) temporal evolution of the instantaneous energy dissipation rate. The initial state is marked by a black star. HLNN predictions (red) closely match ground truth (gray). Quantitative performance summary: (e) average state MSE and (f) average dissipation rate MSE, computed over the full time domain across all test trajectories. Insets display the temporal evolution of the corresponding average errors.

sensitivity under strong noise, because velocity is represented through local differences,

$$\dot{q}_k \approx \frac{q_{k+1} - q_k}{h},$$

which under noisy observations becomes

$$\dot{q}_k^{\text{obs}} \approx \dot{q}_k + \frac{\varepsilon_{k+1} - \varepsilon_k}{h}.$$

Accordingly,

$$\text{Var}\left(\frac{\varepsilon_{k+1} - \varepsilon_k}{h}\right) = \frac{2\sigma^2}{h^2},$$

so observation noise is further amplified through the local-difference representation used to encode discrete dynamics. This helps explain the degradation of trajectory prediction in the high-noise regime. Notably, despite this $\mathcal{O}(1/h^2)$ amplification, HLNN remains comparatively stable in dissipation-rate reconstruction, suggesting that the Herglotz structural constraints still preserve dissipative information to some extent.

Interpolation-recovered missing observations. We next examined the effect of incomplete observations. Each training trajectory was generated with a fixed time step and total length $T_1 = 200$. For each missing ratio

$$r \in \{0.1, 0.2, 0.4, 0.5, 0.7\},$$

we randomly removed approximately $T_1 r$ internal time points from each trajectory without replacement, while preserving the first two and the final time points. We then reconstructed the full position trajectory by applying one-dimensional piecewise linear interpolation independently to each coordinate, using only the remaining position observations. Only the position sequence $q(t)$ was interpolated; velocity and acceleration were not directly reconstructed. All methods were then trained on the same recovered position trajectories under their original training pipelines. Therefore, this experiment should be interpreted as evaluating model behavior under a unified interpolation-based recovery protocol for incomplete observations, rather than as assessing a native irregular-sampling capability of HLNN.

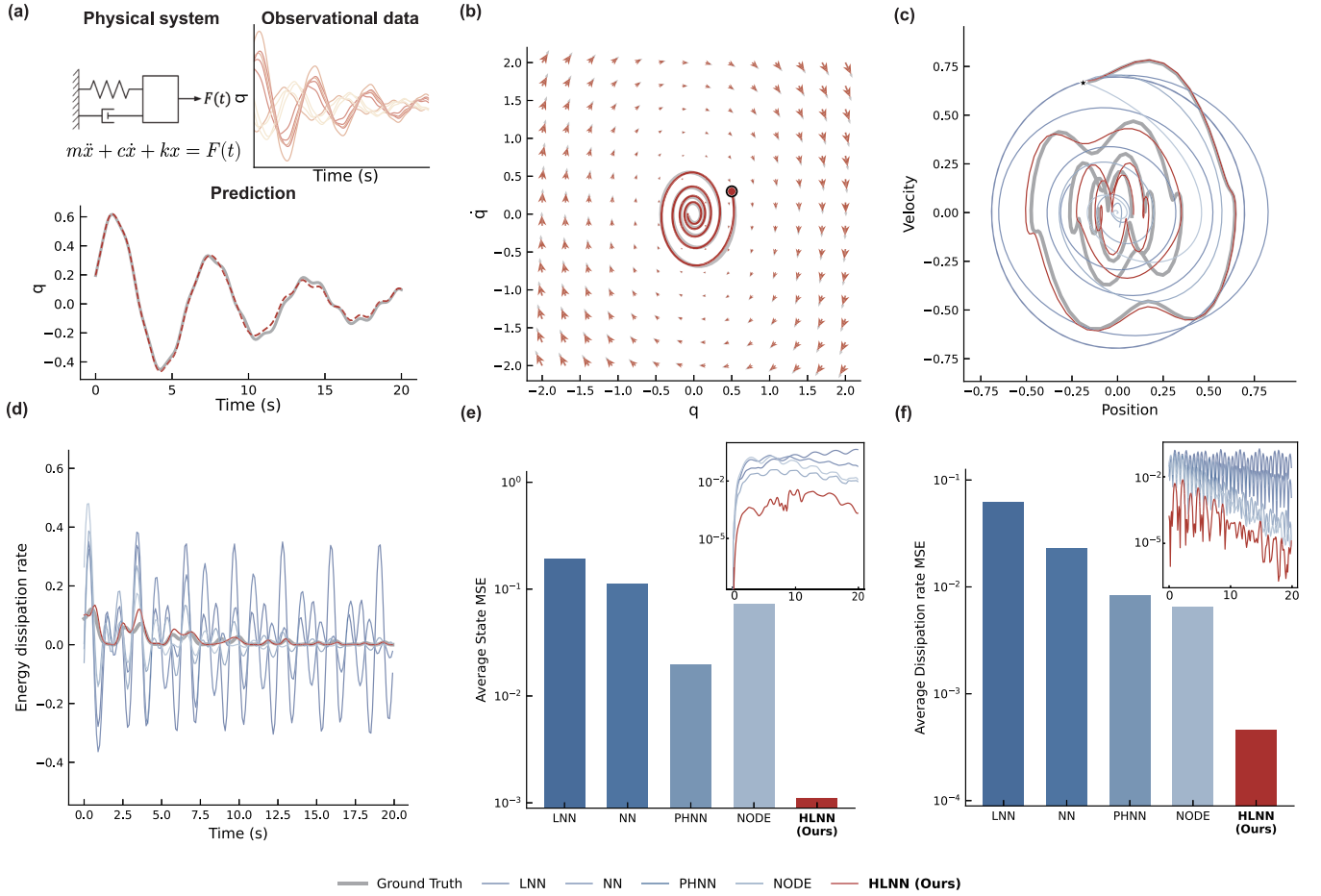


FIG. 6. Results for the forced damped oscillator. (a) Top: Schematic of the physical system and observed trajectories $q(t)$. Bottom: Rollout prediction of $q(t)$ from the same initial condition, comparing HLNN (red) with ground truth (gray). (b) Comparison of the true (gray) and learned (red) vector fields (arrows) and their corresponding phase-space trajectories; the red dot marks the initial condition. Comparison of reconstructed dynamics for a randomly selected test trajectory: (c) phase-space trajectory and (d) temporal evolution of the instantaneous energy dissipation rate. The initial state is marked by a black star. HLNN predictions (red) closely match GT (gray). Quantitative comparison: (e) average state MSE and (f) average dissipation rate MSE, computed over the full time domain across all test trajectories. Insets display the temporal evolution of the corresponding average errors.

As shown in Figs. 7(c) and 7(f), HLNN remains robust under this setting. Across all tested missing ratios, it achieves the lowest trajectory-rollout error and the most accurate dissipation-rate reconstruction. This indicates that, under a unified interpolation-recovery protocol, HLNN is relatively tolerant to incomplete observations and is particularly effective at preserving dissipative structure. Overall, these results show that HLNN is robust under sparse temporal sampling and incomplete observations, remains competitive under low-to-moderate observation noise, and also exhibits a clear limitation in the high-noise regime.

IV. DISCUSSION

Our results demonstrate that, on the systems considered here, HLNN outperforms the tested baselines in both long-horizon state prediction and energy dissipation-rate estimation while relying solely on position observations. The model recovers dissipative dynamics over extended time horizons without requiring a prescribed empirical damping law. This is because HLNN is built upon the Herglotz variational

principle, which incorporates dissipation intrinsically into the Lagrangian formalism rather than introducing phenomenological modifications to otherwise conservative frameworks. As a result, the learned dynamics remain structurally consistent with the contact formulation, which is associated with improved long-horizon qualitative fidelity on the benchmark systems studied here.

The primary conceptual contribution of HLNN lies in a fundamental shift of inductive bias: Rather than merely appending a new loss term, it transitions the framework from a conservative variational structure to its dissipative contact variational counterpart. Accordingly, HLNN should not be viewed simply as a conservative model with an additional penalty term, but as a framework whose underlying generating structure is changed at the variational level. At the same time, HLNN is not intended as a universally optimal framework for learning dissipative dynamics. The present implementation is most naturally suited to low-dimensional smooth dissipative systems with unknown, state-dependent, or time-modulated dissipation, especially when only position observations are available. In contrast, when the damping law is known and

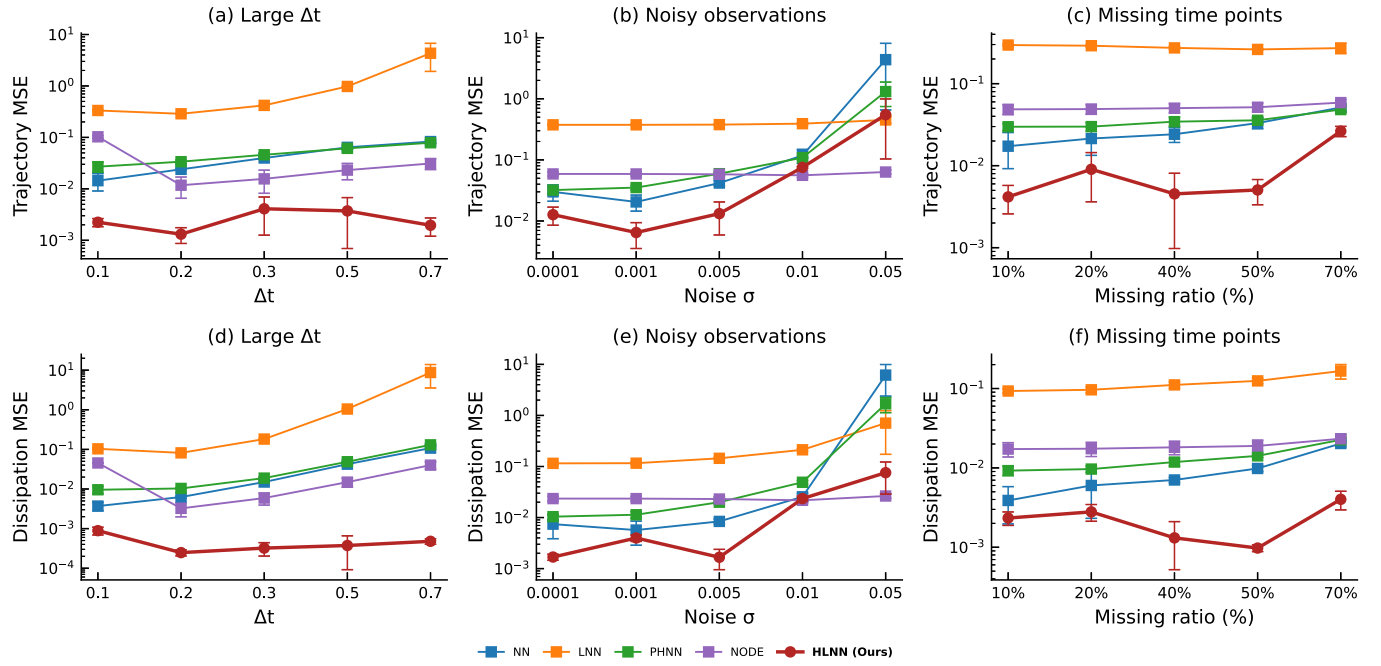


FIG. 7. Robustness evaluation of HLNN under degraded observation conditions. HLNN is compared with the baseline methods on the damped harmonic oscillator under three types of degraded observations: enlarged sampling intervals $\Delta t \in \{0.1, 0.2, 0.3, 0.5, 0.7\}$ in panels (a) and (d); position-only observations corrupted by zero-mean Gaussian noise with standard deviation $\sigma \in \{0.0001, 0.001, 0.005, 0.01, 0.05\}$ in panels (b) and (e); and trajectories with missing internal time points reconstructed by piecewise linear interpolation at missing ratios $r \in \{0.1, 0.2, 0.4, 0.5, 0.7\}$ in panels (c) and (f). For the missing-data setting, all methods are trained on the same interpolation-recovered position trajectories. The top row shows the trajectory prediction MSE, while the bottom row shows the dissipation-rate reconstruction MSE; error bars indicate the standard deviation over five random seeds. Panels (a)–(f) report the trajectory-prediction and dissipation-rate reconstruction errors of all methods under the tested sampling, noise, and missing-data conditions.

simple, more explicit Rayleigh-, PHNN-, or DHNN-style formulations can be more direct and parsimonious. When richer state supervision is available and the goal is an explicit decomposition of nonconservative effects, GLNN-like approaches may be more natural. Likewise, in severely noise-corrupted settings, explicit vector-field models such as NODE may be preferable when pure trajectory prediction is the primary objective.

We further examined HLNN under pronounced phase-space contraction. The supplementary controlled decay experiments show that HLNN is not entirely immune to the multiscale challenge associated with strongly decaying dynamics. However, compared with a matched one-step-MSE-trained control under the same Herglotz parametrization, HLNN exhibits weaker error amplification, particularly in the low-energy regime. Detailed settings and quantitative results are provided in Sec. IX of Supplemental Material [40].

Building upon this framework, extending HLNN to broader physical domains offers several exciting possibilities. Given that Herglotz contact mechanics offers a unified variational description for a wide range of dissipative processes, including continuum fields, nonequilibrium thermodynamics, and more general open systems [38,43–45], a particularly compelling direction is the application to thermodynamic systems. The variable z can be associated with specific thermodynamic potentials such as entropy in the contact geometry of thermodynamics, depending on the chosen representation [43,45,46]. Consequently, by adapting the HLNN architecture

to incorporate macroscopic observables such as temperature and pressure, the model could be used to learn effective thermodynamic potentials and, in turn, to infer state functions such as heat capacity.

While HLNN demonstrates strong empirical performance across a broad range of settings, the very structural rigor that allows it to capture the rich geometry of dissipative systems introduces inherent trade-offs. To maintain exact contact variational fidelity, the architecture actively incorporates recursive trajectory-wise propagation and the solving of implicit DHLE equations. The necessary consequence of this theoretical robustness is a higher computational overhead compared to explicit baselines. Because our current formulation preserves generality without imposing strong prior structural assumptions (such as sparsity), the Newton iterations required during rollout, specifically the repeated evaluations of the residual's Jacobian and associated linear solves, scale poorly as the spatial dimension d grows. Therefore, while manageable for the present proof-of-principle benchmarks, scaling to generic high-dimensional systems will require exploiting additional structures, such as local coupling or specialized quasi-Newton solvers [47,48]. Similarly, to ensure the direct physical interpretability of the learned dynamics, our current focus is restricted to settings where observations correspond directly to generalized coordinates. Extending this rigorous framework to ingest raw, high-dimensional observational data will necessitate integration with autoencoders [17,49–51], provided the learned latent space admits structured DHLE solves

to preserve computational viability [52]. A similar duality applies to the framework's data requirements and robustness. By deriving dynamics purely from the discrete Herglotz principle, HLNN can effectively infer dissipative mechanisms from position-only data. However, the inherent limitation of this approach is an identifiability challenge: Explicitly decoupling specific underlying potentials or external forces solely from position observations remains difficult. Furthermore, while its discrete multipoint structure provides stability against moderate noise and missing data, it naturally becomes sensitive to high-frequency perturbations under strong observation noise. In severely corrupted regimes where pure trajectory prediction is the primary objective, explicit vector-field models may remain more optimal. Addressing these regimes by integrating dedicated denoising modules or noise-aware discrete regularization is an important direction for future research.

Finally, the requirement to construct a tractable and physically meaningful discrete model necessitates certain constraints on the hypothesis space. By adopting the restricted, local first-order-in- z Herglotz ansatz [Eq. (9)], we promote its well posedness and interpretability, reduce the representational ambiguity of the learned Lagrangian, improve the regularity and numerical stability of the implicit DHGL update, and offer an interpretable route for estimating the intrinsic dissipation rate.

We also acknowledge that this decomposition itself constitutes a substantive physics-informed inductive bias: It restricts the hypothesis space to a function class consistent with these physical assumptions. Correspondingly, similar structural priors may also benefit other frameworks. Indeed, analogous decompositions have been introduced in the HNN [18] family by SRNN [53] and in the LNN [19] family by DeLaN [51], both of which adopt explicit kinetic-potential decompositions and report improved performance over their unified counterparts. Therefore, the reported performance of HLNN relative to the tested baselines should be interpreted as reflecting the combined effects of the explicit physics-informed decomposition and the Herglotz contact variational coupling through the action variable z .

However, this parametrization also introduces several inherent limitations. By restricting the kinetic energy to a quadratic form with a learnable scalar mass parameter and representing the conservative part as a position-dependent potential, the current ansatz may not faithfully capture dynamical systems that cannot be well approximated by this explicit mechanical structure—those with configuration-dependent mass matrices, strong inertial couplings, nonquadratic velocity dependence, or nonseparable position-velocity interactions. Typical examples include charged-particle dynamics in magnetic fields, rotating-frame dynamics dominated by Coriolis forces, and systems with relativistic kinetic-energy forms. In addition, the first-order-in- z approximation restricts the model from representing more general action-dependent Lagrangians, including higher-order contact effects, long-memory dissipation, and hysteretic processes. Relaxing the current Lagrangian representation and introducing higher-order contact terms are natural future directions for capturing richer classes of dynamics and broadening the applicability of the method.

Moreover, the framework's reliance on smooth activation functions—which are essential for continuous optimization—currently precludes the modeling of nonsmooth phenomena like dry friction. Incorporating recent theoretical advances in nonsmooth [54] and stochastic [55,56] Herglotz variational principles represents a promising pathway to further generalize this contact variational learning paradigm.

V. CONCLUSION

We have shown that HLNN, constructed upon the Herglotz variational principle, successfully learns a class of smooth dissipative systems. Our experimental results demonstrate that HLNN achieves strong overall performance relative to the baseline methods considered in this work across the benchmark dissipative systems studied here. Specifically, the model not only accurately recovers dissipative dynamics and achieves strong long-horizon predictive performance in damped harmonic oscillator systems, but also generalizes to more complex dissipative systems using position-only data. Furthermore, HLNN exhibits a distinct structural advantage: It can infer effective dissipation behavior without prescribing an explicit empirical damping law in advance and reconstructs the time evolution of the system's energy dissipation rate.

Overall, these results establish a promising framework for the data-driven modeling of smooth dissipative systems. While the current study focuses on smooth, low-dimensional dissipative dynamics, the variational structure of HLNN opens meaningful pathways for broader applications. Future work may investigate the extension of this formalism to nonsmooth systems and explore its potential in uncovering thermodynamic potentials from macroscopic observables. We believe that this study serves as a useful step toward the interpretable, data-driven identification of complex dissipative mechanisms.

ACKNOWLEDGMENTS

Q.H., J.M.M., and G.Y. acknowledge support from the National Natural Science Foundation of China (Grants No. T2225022, No. 12350710786, No. 62088101, No. T2541015, and No. 12375036), and the Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0100). T.-T.G. did not receive support from the funding sources listed above.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: THEORETICAL DERIVATIONS AND IMPLEMENTATION DETAILS

1. Connecting the Rayleigh and Herglotz formulations of the damped oscillator

To provide a direct connection with the familiar ODE formulation of dissipative dynamics, we consider the classical damped harmonic oscillator

$$m\ddot{q} + c\dot{q} + kq = 0, \quad (\text{A1})$$

where m is the mass, c is the damping coefficient, and k is the spring constant. In the Rayleigh formulation, one starts from the conservative Lagrangian

$$L_0(q, \dot{q}) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}kq^2, \quad (\text{A2})$$

together with the Rayleigh dissipation function

$$R(\dot{q}) = \frac{1}{2}c\dot{q}^2. \quad (\text{A3})$$

The damped equation of motion is then obtained from

$$\frac{d}{dt} \left(\frac{\partial L_0}{\partial \dot{q}} \right) - \frac{\partial L_0}{\partial q} + \frac{\partial R}{\partial \dot{q}} = 0, \quad (\text{A4})$$

which immediately gives

$$m\ddot{q} + c\dot{q} + kq = 0. \quad (\text{A5})$$

In the Herglotz formulation, the same ODE can be obtained from the action-dependent Lagrangian

$$\mathcal{L}(q, \dot{q}, z) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}kq^2 - \frac{c}{m}z, \quad (\text{A6})$$

together with the evolution equation

$$\dot{z} = \mathcal{L}. \quad (\text{A7})$$

Under the sign convention adopted in Eq. (9), this corresponds to the special case $\gamma = -c/m$. Using the Herglotz Euler-Lagrange equation

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) - \frac{\partial \mathcal{L}}{\partial q} - \frac{\partial \mathcal{L}}{\partial z} \frac{\partial \mathcal{L}}{\partial \dot{q}} = 0, \quad (\text{A8})$$

and noting that

$$\frac{\partial \mathcal{L}}{\partial \dot{q}} = m\dot{q}, \quad \frac{\partial \mathcal{L}}{\partial q} = -kq, \quad \frac{\partial \mathcal{L}}{\partial z} = -\frac{c}{m}, \quad (\text{A9})$$

we obtain

$$m\ddot{q} + kq + c\dot{q} = 0. \quad (\text{A10})$$

Therefore, for the standard damped harmonic oscillator, the Rayleigh and Herglotz formulations lead to the same ODE, but they introduce dissipation in different ways. In the Rayleigh description, dissipation appears as an additional generalized force term. In the Herglotz description, dissipation is incorporated intrinsically through the dependence of the Lagrangian on the action variable z .

2. Noether-type quantities in Herglotz systems

In conservative variational mechanics, Noether's theorem associates continuous symmetries with strictly conserved quantities. For Herglotz systems, however, the corresponding objects are generally not strictly conserved in the usual sense. Instead, one obtains generalized or dissipated quantities, or equivalently quantities that become constant after multiplication by a suitable integrating factor [34,37]. For an autonomous continuous Herglotz system, if the Lagrangian $L(q, \dot{q}, z)$ does not depend explicitly on time, then the generalized energy

$$E_L = \dot{q}^i \frac{\partial L}{\partial \dot{q}^i} - L \quad (\text{A11})$$

satisfies

$$\dot{E}_L = \left(\frac{\partial L}{\partial z} \right) E_L. \quad (\text{A12})$$

Therefore, the rescaled generalized quantity

$$I_E(t) = E_L(t) \exp \left(- \int_0^t \frac{\partial L}{\partial z}(\tau) d\tau \right) \quad (\text{A13})$$

remains constant along the exact trajectory. In this sense, time-translation symmetry in the Herglotz framework is associated not with an ordinary strictly conserved energy, but with a generalized energy-type quantity. Since HLNN is trained and rolled out through the discrete Herglotz principle and the discrete Herglotz-Euler-Lagrange equations, we do not directly use the continuous expression (A13) as a numerical diagnostic. Instead, motivated by the continuous generalized energy and by the conformal scaling structure of discrete contact systems [34], we introduce an energy-type diagnostic adapted to the actual discrete implementation employed in this work. Under the midpoint discretization adopted in HLNN, the learned discrete Herglotz Lagrangian is

$$L_d(q_k, q_{k+1}, z_k) = h L_\theta(\bar{q}_k, \bar{v}_k, z_k). \quad (\text{A14})$$

We then define the discrete energy-type quantity

$$J_k = \bar{v}_k \cdot \partial_{\bar{v}_k} L_\theta(\bar{q}_k, \bar{v}_k, z_k) - L_\theta(\bar{q}_k, \bar{v}_k, z_k), \quad (\text{A15})$$

together with the discrete contact multiplier

$$\Delta_k := 1 + D_3 L_d(q_k, q_{k+1}, z_k). \quad (\text{A16})$$

On this basis, we construct the compensated normalized diagnostic

$$\mathcal{I}_k := \frac{J_k}{J_0 \prod_{j=1}^k \Delta_j}. \quad (\text{A17})$$

Here, we do not interpret $\mathcal{I}_k$ as a theorem-guaranteed exact invariant. Rather, we use it as a structure-aware numerical diagnostic: If the learned model remains close to the underlying discrete Herglotz structure, then $\mathcal{I}_k$ is expected to stay close to the constant value 1 during rollout without visible long-time secular drift. The corresponding experimental evaluation on the damped harmonic oscillator is reported in Sec. V of Supplemental Material [40], where we present the rollout curves and summary statistics of this compensated diagnostic.

3. Model

The *discrete Herglotz Lagrangian neural networks*' details are shown below. Instead of directly learning the dynamical vector field [25–27,41], HLNNs learn a Herglotz-type Lagrangian L such that, after discretization into L_d , the resulting DHLE equations (8) reproduce the observed trajectories and their dissipative behavior. To obtain a tractable and interpretable subclass of Herglotz models, we motivate the model ansatz by a Taylor expansion of the Herglotz Lagrangian with respect to the action variable z around the conservative limit:

$$L(q, \dot{q}, z) = L(q, \dot{q}, 0) + \frac{\partial L}{\partial z} \Big|_{z=0} z + \frac{1}{2} \frac{\partial^2 L}{\partial z^2} \Big|_{z=0} z^2 + \dots \quad (\text{A18})$$

By adopting a first-order approximation around $z = 0$ and identifying the zeroth-order term with the standard classical form of kinetic energy minus potential energy [39], we use the structured ansatz

$$L(q, \dot{q}, z) = T(q, \dot{q}) - V_\theta(q) + \gamma_\theta(q, \dot{q})z. \quad (\text{A19})$$

A straightforward approach would be to parametrize all three components using separate neural networks. However, in practice, we find that such a flexible structure leads to an overparametrized model. Moreover, a well-known feature of Lagrangian mechanics is the nonuniqueness of the Lagrangian: Many different Lagrangians can generate the same equations of motion [39,57]. Without additional structure, the network may converge to an arbitrary representative within this equivalence class, making the learned Lagrangian harder to interpret. To mitigate these issues, we adopt the structured decomposition in Eq. (9). In particular, the kinetic term is fixed to the quadratic form

$$T(q, \dot{q}) = \frac{1}{2} \dot{q}^T M \dot{q}, \quad (\text{A20})$$

with $M = mI$, where m is a learnable positive scalar constrained by

$$m = \text{Softplus}(s) + \epsilon, \quad \epsilon = 10^{-3}. \quad (\text{A21})$$

The potential term $V_\theta(q)$ is represented by a neural network and captures the conservative part of the dynamics, while the dissipation coefficient $\gamma_\theta(q, \dot{q})$ is represented by another neural network and governs the action-dependent part of the Herglotz Lagrangian. Thus, the three terms in the model correspond, respectively, to the kinetic, potential, and dissipation-related parts of the learned variational structure. For forced systems with explicit time dependence, a natural Herglotz-type Lagrangian is [33]

$$L(t, q, \dot{q}, z) = \frac{1}{2} \dot{q}^T M \dot{q} - V_\theta(q) + \gamma_\theta(q, \dot{q})z + q^T F(t). \quad (\text{A22})$$

In principle, HLNN could incorporate an additional force network to parametrize $F(t)$. However, this would introduce excessive model flexibility and can hinder convergence. We therefore absorb the forcing term $F(t)$ into the potential network $V_\theta(q, t)$, interpreted as a generalized time-dependent potential, rather than parametrizing it separately. Since our model operates on position-only data, the velocity $\dot{q}$ is not given as an input but is approximated internally via midpoint discretization [33,34]:

$$\bar{q}_k = \frac{q_k + q_{k+1}}{2}, \quad \bar{\dot{q}}_k = \frac{q_{k+1} - q_k}{h}. \quad (\text{A23})$$

With the above definition, the discrete Lagrangian is

$$L_d(q_k, q_{k+1}, z_k) = h L(\bar{q}_k, \bar{\dot{q}}_k, z_k). \quad (\text{A24})$$

Importantly, in the present implementation, z is not treated as an ordinary observed input to the neural networks. Instead, it acts as an internal action state that is recursively updated through the discrete Herglotz principle. Although $D_3 L_d$ does not explicitly depend on z_k under the present ansatz, the action variable does not disappear from the discrete dynamics. Under midpoint discretization, the derivatives $D_1 L_d$ and $D_2 L_d$ still contain explicit z_k - and z_{k-1} -dependent contributions through

derivatives of γ_θ . Hence, even though $D_3 L_d$ itself is independent of z_k in the present ansatz, the action variables z_k and z_{k-1} still enter the DHEL residual and continue to constrain the learned dynamics through the $D_1 L_d$ and $D_2 L_d$ terms. Only in the degenerate case where γ_θ is constant, so that $\nabla_q \gamma_\theta = \nabla_{\dot{q}} \gamma_\theta = 0$, do these explicit z -dependent contributions vanish.

This structure captures a broad class of smooth dissipative mechanisms while restricting the search space to physically meaningful Lagrangians. Moreover, the same parametrization extends naturally to nonautonomous and forced systems by including time or external forcing signals as additional inputs to the neural networks.

4. Network optimization

With the network architecture defined, we now detail the optimization objective that enforces the physical constraints. In Fig. 1, we show the model architecture of the proposed HLNN and a schematic illustration of the flow of an observed trajectory of data points through the neural networks. The training procedure takes as input the state vector $\mathbf{q}$ and outputs the Lagrangian L_θ . The objective function combines a physics-informed loss and a regularization term, as shown in Eq. (10).

Optimizing the network solely by minimizing the physics residual can lead to degenerate solutions where the learned Lagrangian fails to define a valid discrete dynamical system [58–60]. To address this challenge for the Herglotz variational principle, we must satisfy two regularity conditions [33,34]:

(1) Regularity of the discrete Legendre transform, which requires the mixed partial derivative matrix H_k (with entries $H_{ij} = \frac{\partial^2 L_d}{\partial q_i^t \partial q_{k+1}^j}$) to be nonsingular, corresponding to the discrete regularity condition; see Sec. III of Supplemental Material [40].

(2) Solvability of the contact update, which requires the contact factor $\Delta_k = 1 + \frac{\partial L_d}{\partial z_k}$ to be nonvanishing [33].

Unlike existing regularization loss functions designed for discrete Lagrangian neural networks, we formulate our regularization term based on singular value decomposition. This approach effectively circumvents the numerical instability associated with the direct computation of the determinant of the mixed-derivative matrix [61]:

$$\mathcal{J}_{\text{reg}} = \beta(\tau_{\min} - \sigma_{\min}(H_k)) + \beta(\delta_{\min} - |\Delta_k|). \quad (\text{A25})$$

Here, $\sigma_{\min}(H_k)$ denotes the smallest singular value of the discrete mixed-derivative matrix at step k and $\beta(x)$ denotes the Softplus function

$$\beta(x) = \text{Softplus}(x) = \ln(1 + e^x), \quad (\text{A26})$$

which imposes a heavy penalty only when $\sigma_{\min}(H_k)$ and $|\Delta_k|$ fall below the safety thresholds $\tau_{\min} = 10^{-3}$ and $\delta_{\min} = 10^{-3}$.

These two terms are combined using hyperparameters w_{phys} and w_{reg} , which balance DHEL consistency against structural regularity. Empirically, we found that this ratio balanced the convergence speed and satisfaction of the regularity condition. They were determined via grid search (see Sec. VII of Supplemental Material [40]).

The networks were implemented as multilayer perceptrons with three hidden layers and 256 neurons per layer, using different activation functions for different tasks. We used the

Adam optimizer with a learning rate of 10^{-3} for most tasks. The specific training parameters and details can be found in Sec. VIII of Supplemental Material [40].

For data generation, we employed an RK4 integrator to simulate each dynamical system given various initial conditions. To ensure the accuracy of the reference data, we used a time step $\Delta t = 0.1$. The detailed sampling strategy for initial conditions and the parameter ranges of each system are described separately in Sec. III.

5. Testing

During the training phase, the model was fitted on observed trajectory segments by minimizing the total loss (10). In the prediction phase, rather than providing full trajectories as input, we supply two initial positions q_0, q_1 as discrete-time initial data. Using the learned Lagrangian L_θ , we then apply Newton's method (with a tolerance of $\epsilon = 10^{-6}$) to solve the implicit DHLE equation $\mathcal{R}(q_0, q_1, \hat{q}_2) = 0$ for the next-step prediction $\hat{q}_2$. This process is repeated recursively by feeding predictions as new inputs to generate

long-horizon rollouts. To quantify the extrapolation capability of the model, we define the extrapolation error as the mean squared error between the true trajectory and the predicted rollout. We averaged this error over all time steps k and all trajectories τ to obtain an overall prediction error metric, $\text{MSE}_{\text{rollout}}$.

Furthermore, we assessed whether the HLNN had learned an effective Herglotz Lagrangian by comparing the positive instantaneous energy dissipation rates of the prediction and the ground truth, using MSE_{diss} .

When the Herglotz Lagrangian takes the form (9), the positive intrinsic dissipation rate of the system is given by Eq. (15). For the HLNN, we computed the predicted dissipation rate $D_{\text{diss}}^{\text{pred}}$ by applying Eq. (15), using midpoint discretization to approximate velocities. In contrast, for the baseline models, we calculated the energy dissipation rate numerically by differentiating the empirically defined energy E (derived from the predicted states) with respect to time, i.e., $-\frac{dE}{dt}$. The ground-truth energy dissipation rate was obtained analytically. Further details of the testing procedure are given in Sec. VI of Supplemental Material [40].

-
- [1] S. L. Brunton and J. N. Kutz, Promising directions of machine learning for partial differential equations, *Nat. Comput. Sci.* **4**, 483 (2024).
 - [2] H. Levine and Y. Tu, Machine learning meets physics: A two-way street, *Proc. Natl. Acad. Sci. USA* **121**, e2403580121 (2024).
 - [3] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, *et al.*, Scientific discovery in the age of artificial intelligence, *Nature (London)* **620**, 47 (2023).
 - [4] R. Yu and R. Wang, Learning dynamical systems from data: An introduction to physics-guided deep learning, *Proc. Natl. Acad. Sci. USA* **121**, e2311808121 (2024).
 - [5] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359 (1989).
 - [6] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* **113**, 3932 (2016).
 - [7] M. Schmidt and H. Lipson, Distilling free-form natural laws from experimental data, *Science* **324**, 81 (2009).
 - [8] T.-T. Gao and G. Yan, Autonomous inference of complex network dynamics from incomplete and noisy data, *Nat. Comput. Sci.* **2**, 160 (2022).
 - [9] T.-T. Gao, B. Barzel, and G. Yan, Learning interpretable dynamics of stochastic complex systems from experimental data, *Nat. Commun.* **15**, 6029 (2024).
 - [10] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, *Nat. Rev. Phys.* **3**, 422 (2021).
 - [11] S.-M. Udrescu and M. Tegmark, AI Feynman: A physics-inspired method for symbolic regression, *Sci. Adv.* **6**, eaay2631 (2020).
 - [12] Z. Liu and M. Tegmark, Machine learning conservation laws from trajectories, *Phys. Rev. Lett.* **126**, 180604 (2021).
 - [13] Z. Zhang, Y. Shin, and G. E. Karniadakis, GFNNs: GENERIC formalism informed neural networks for deterministic and stochastic dynamical systems, *Philos. Trans. R. Soc. A* **380**, 20210207 (2022).
 - [14] H. Yu, X. Tian, W. E, and Q. Li, OnsagerNet: Learning stable and interpretable dynamics using a generalized Onsager principle, *Phys. Rev. Fluids* **6**, 114402 (2021).
 - [15] X. Chen, B. W. Soh, Z.-E. Ooi, E. Vissol-Gaudin, H. Yu, K. S. Novoselov, K. Hippalgaonkar, and Q. Li, Constructing custom thermodynamics using deep learning, *Nat. Comput. Sci.* **4**, 66 (2024).
 - [16] J. Yang, W. Rao, N. Dehmamy, R. Walters, and R. Yu, Symmetry-informed governing equation discovery, in *Advances in Neural Information Processing Systems*, Vol. 37 (NeurIPS, 2024), pp. 65297–65327.
 - [17] J. Zhang, Q. Zhu, and W. Lin, Learning Hamiltonian neural Koopman operator and simultaneously sustaining and discovering conservation laws, *Phys. Rev. Res.* **6**, L012031 (2024).
 - [18] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2019), pp. 15379–15389.
 - [19] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho, Lagrangian neural networks, *arXiv:2003.04630*.
 - [20] P. Jin, Z. Zhang, A. Zhu, Y. Tang, and G. E. Karniadakis, SympNet: Intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems, *Neural Networks* **132**, 166 (2020).
 - [21] H. C. Öttinger, Nonequilibrium thermodynamics for open systems, *Phys. Rev. E* **73**, 036126 (2006).
 - [22] M. te Vrugt and R. Wittkowski, Metareview: A survey of active matter reviews, *Eur. Phys. J. E* **48**, 12 (2025).
 - [23] A. Di Vita, *Non-Equilibrium Thermodynamics*, Lecture Notes in Physics (Springer, Cham, 2022), Vol. 1007.
 - [24] C. R. Galley, Classical mechanics of nonconservative systems, *Phys. Rev. Lett.* **110**, 174301 (2013).

- [25] A. Sosanya and S. Greydanus, Dissipative Hamiltonian neural networks: Learning dissipative and conservative dynamics separately, [arXiv:2201.10085](#).
- [26] S. A. Desai, M. Mattheakis, D. Sondak, P. Protopapas, and S. J. Roberts, Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems, *Phys. Rev. E* **104**, 034312 (2021).
- [27] S. Xiao, J. Zhang, and Y. Tang, Generalized Lagrangian neural networks, *Commun. Comput. Phys.* **39**, 59 (2026).
- [28] Z. Chen, M. Feng, J. Yan, and H. Zha, Learning neural Hamiltonian dynamics: A methodological overview, [arXiv:2203.00128](#).
- [29] A. Testa, S. Hauberg, T. Asfour, and L. Roza, Geometric contact flows: Contactomorphisms for dynamics and control, in *Proceedings of the 42nd International Conference on Machine Learning*, edited by A. Singh, M. Fazel, D. Hsu, S. Lacoste-Julien, F. Berkenkamp, T. Maharaj, K. Wagstaff, and J. Zhu, Proceedings of Machine Learning Research, Vol. 267 (PMLR, Vancouver, Canada, 2025), pp. 59259–59284.
- [30] P. Canizares, D. Murari, C.-B. Schönlieb, F. Sherry, and Z. Shumaylov, Symplectic neural flows for modeling and discovery, [arXiv:2412.16787](#).
- [31] V. Sundararaghavan, M. N. Shah, and J. P. Simmons, Lagrangian neural networks for reversible dissipative evolution, [arXiv:2405.14645](#).
- [32] Y. Chen, T. Matsubara, and T. Yaguchi, Neural symplectic form: Learning Hamiltonian equations on general coordinate systems, in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21 (Curran Associates Inc., Red Hook, NY, 2021).
- [33] M. Vermeeren, A. Bravetti, and M. Seri, Contact variational integrators, *J. Phys. A: Math. Theor.* **52**, 445206 (2019).
- [34] A. Anahory Simoes, D. M. de Diego, M. Lainz Valcázar, and M. de León, On the geometry of discrete contact mechanics, *J. Nonlinear Sci.* **31**, 53 (2021).
- [35] A. Bravetti, H. Cruz, and D. Tapias, Contact Hamiltonian mechanics, *Ann. Phys. (NY)* **376**, 17 (2017).
- [36] H. Geiges, *An Introduction to Contact Topology* (Cambridge University Press, Cambridge, 2008), Vol. 109.
- [37] B. Georgieva, R. Guenther, and T. Bodurov, Generalized variational principle of Herglotz for several independent variables. First Noether-type theorem, *J. Math. Phys.* **44**, 3911 (2003).
- [38] A. A. Simoes, M. De León, M. L. Valcázar, and D. M. De Diego, Contact geometry for simple thermodynamical systems with friction, *Proc. R. Soc. A* **476**, 20200244 (2020).
- [39] J. E. Marsden and M. West, Discrete mechanics and variational integrators, *Acta Numer.* **10**, 357 (2001).
- [40] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/9gnh-89jd> for additional theoretical derivations, implementation details, robustness analyses, and supplementary numerical results, which includes Refs. [62,63].
- [41] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, Neural ordinary differential equations, in *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2018), pp. 6572–6583.
- [42] P. Wulff, L. Lentz, and U. von Wagner, Determination of the polynomial restoring force of a one DoF bistable Duffing oscillator by linear regression, *Acta Mech.* **234**, 1973 (2023).
- [43] A. Bravetti, M. Á. García-Ariza, and D. Tapias, Thermodynamic entropy as a Noether invariant from contact geometry, *Entropy* **25**, 1082 (2023).
- [44] J. Gaset, M. Lainz, A. Mas, and X. Rivas, The Herglotz variational principle for dissipative field theories, *Geom. Mech.* **01**, 153 (2024).
- [45] A. Bravetti, Contact Hamiltonian dynamics: The concept and its use, *Entropy* **19**, 535 (2017).
- [46] R. Mrugała, Geometrical formulation of equilibrium phenomenological thermodynamics, *Rep. Math. Phys.* **14**, 419 (1978).
- [47] J. Lee, C. K. Liu, F. C. Park, and S. S. Srinivasa, A linear-time variational integrator for multibody systems, in *Algorithmic Foundations of Robotics XII: Proceedings of the Twelfth Workshop on the Algorithmic Foundations of Robotics*, edited by K. Goldberg, P. Abbeel, K. Bekris, and L. Miller, Springer Proceedings in Advanced Robotics, Vol. 13 (Springer, Cham, 2020), pp. 352–367.
- [48] J. Brüdigam, S. Sosnowski, Z. Manchester, and S. Hirche, Variational integrators and graph-based solvers for multibody dynamics in maximal coordinates, *Multibody Syst. Dyn.* **61**, 381 (2024).
- [49] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, *Proc. Natl. Acad. Sci. USA* **116**, 22445 (2019).
- [50] P. Toth, D. J. Rezende, A. Jaegle, S. Racanière, A. Botev, and I. Higgins, Hamiltonian generative networks, in *International Conference on Learning Representations* (OpenReview.net, 2020).
- [51] M. Lutter, C. Ritter, and J. Peters, Deep Lagrangian networks: Using physics as model prior for deep learning, in *International Conference on Learning Representations* (OpenReview.net, 2019).
- [52] S. Saemundsson, A. Terenin, K. Hofmann, and M. Deisenroth, Variational integrator networks for physically structured embeddings, in *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, edited by S. Chiappa and R. Calandra, Proceedings of Machine Learning Research, Vol. 108 (PMLR, Online, 2020), pp. 3078–3087.
- [53] Z. Chen, J. Zhang, M. Arjovsky, and L. Bottou, Symplectic recurrent neural networks, in *International Conference on Learning Representations* (OpenReview.net, 2020).
- [54] A. López-Gordón, L. Colombo, and M. De León, Nonsmooth Herglotz variational principle, in *Proceedings of the 2023 American Control Conference (ACC)* (IEEE, Piscataway, NJ, 2023), pp. 3376–3381.
- [55] Q. Zhan, J. Duan, X. Li, and Y. Li, Numerical integration of stochastic contact Hamiltonian systems via stochastic Herglotz variational principle, *Phys. Scr.* **98**, 055211 (2023).
- [56] Q. Zhan, J. Duan, and L. Wang, Numerical integrations of stochastic contact Hamiltonian systems via stochastic contact Hamilton–Jacobi equation, *Commun. Nonlinear Sci. Numer. Simul.* **150**, 108986 (2025).
- [57] E. Massa and E. Pagani, On the Herglotz variational problem, *J. Math. Phys.* **64**, 102902 (2023).
- [58] Y. Lishkova, P. Scherer, S. Ridderbusch, M. Jamnik, P. Liò, S. Ober-Blöbaum, and C. Offen, Discrete Lagrangian

- neural networks with automatic symmetry discovery, [IFAC-PapersOnLine](#) **56**, 3203 (2023).
- [59] S. Ober-Blöbaum and C. Offen, Variational learning of Euler–Lagrange dynamics from data, [J. Comput. Appl. Math.](#) **421**, 114780 (2023).
- [60] M. D. Hansen, E. Celledoni, and B. K. Tapley, Learning mechanical systems from real-world data using discrete forced Lagrangian dynamics, [arXiv:2505.20370](#).
- [61] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. (Johns Hopkins University Press, Baltimore, MD, 2013).
- [62] R. M. Santilli, *Foundations of Theoretical Mechanics I: The Inverse Problem in Newtonian Mechanics* (Springer, New York, 1978), reprinted 2013.
- [63] J. Brüdigam, M. Schuck, A. Capone, S. Sosnowski, and S. Hirche, Structure-preserving learning using Gaussian processes and variational integrators, in *Proceedings of the 4th Annual Learning for Dynamics and Control Conference*, edited by R. Firoozi, N. Mehr, E. Yel, R. Antonova, J. Bohg, M. Schwager, and M. Kochenderfer, Proceedings of Machine Learning Research, Vol. 168 (PMLR, Stanford, CA, 2022), pp. 1150–1162.