LEARNING TO IDENTIFY PHYSICAL LAWS OF HAMILTONIAN SYSTEMS VIA META-LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Hamiltonian mechanics is an effective tool to represent many physical processes with concise yet well-generalized mathematical expressions. A well-modeled Hamiltonian makes it easy for researchers to analyze and forecast many related phenomena that are governed by the same physical law. However, in general, identifying a functional or shared expression of the Hamiltonian is very difficult. It requires carefully designed experiments and the researcher's insight that comes from years of experience. We propose that meta-learning algorithms can be potentially powerful data-driven tools for identifying the physical law governing Hamiltonian systems without any mathematical assumptions on the representation, but with observations from a set of systems governed by the same physical law. We show that a well meta-trained learner can identify the shared representation of the Hamiltonian by evaluating our method on several physical systems with various experimental settings.

1 INTRODUCTION

Hamiltonian mechanics, a reformulation of Newtonian mechanics, can be used to describe classical systems focusing on modeling continuous-time evolution of system dynamics with a conservative quantity called Hamiltonian (Goldstein et al., 2002). Interestingly, the formalism of the Hamiltonian provides both geometrically meaningful interpretation (Arnol'd et al., 2001) and efficient numerical schemes (Feng & Qin, 2010) representing the state of complex systems in phase space with symplectic structure. Although formalism was originally developed for classical mechanics, it has been applied to various fields of physics, such as fluid mechanics (Salmon, 1988), statistical mechanics (Reichl, 1999), and quantum mechanics (Sakurai & Commins, 1995).

While it has many useful mathematical properties, establishing an appropriate Hamiltonian of the unknown phenomena is a challenging problem. A Hamiltonian for a system can be modeled by a shared expression of the Hamiltonian and physical parameters. For instance, the Hamiltonian of an ideal pendulum is described as $H = \frac{p^2}{2ml^2} + mgl(1 - \cos q)$ (shared expression), with mass m, pendulum length l, and gravity constant g (physical parameters), whereas q and p are the angle of the pendulum and the corresponding conjugate momentum (state of the system), respectively. Once an appropriate functional of the Hamiltonian is established from observing several pendulums, a new pendulum-like system can be readily recognized by adapting new physical parameters on the expression. Therefore, identifying an appropriate expression of the Hamiltonian is an important yet extremely difficult problem in most science and engineering areas where there still remain numerous unknown processes where it is even uncertain whether a closed-form solution or mathematically clear expression exists.

In the recent era of deep learning, we can consider the use of learning-based algorithms with sufficient data to discover Hamiltonian. To determine the Hamiltonian underlying the unknown physical process, the Hamiltonian should satisfy two fundamental conditions: (1) it should fit well on previously observed data or motions, (2) it should generalize well on newly observed data from new systems if the systems share the same physical law with previous ones. The first condition has been mitigated by explicitly incorporating symplectic structure or conservation laws on neural networks, called Hamiltonian neural networks (HNN) (Greydanus et al., 2019) for learning Hamiltonian dynamics. HNN and its variants have been shown to be effective in learning many useful properties of the Hamiltonian (Toth et al., 2020; Chen et al., 2020; Zhong et al., 2020; Sanchez-Gonzalez et al.,

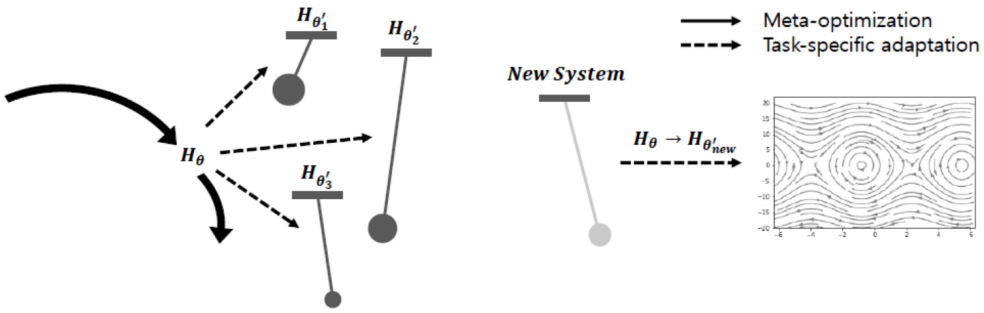


Figure 1: There is a resemblance between meta-learning and identifying the physical laws of Hamiltonian. A hypothesized governing equation of Hamiltonian could be corrected and established by evaluating many related systems, which could be learned using meta-learning as a data-driven method (left). Then, a well-established Hamiltonian can be utilized to predict new system dynamics, which could be viewed as a meta-transfer process by a well-trained meta-learner (right).

2019). In their experiments, it has been shown that HNN and its variants work well on learning conservation laws or continuous-time translational symmetry, enable the learning of complex systems stably by incorporating numerical integrators and generalize on multiple initial conditions or controls for the given system. However, there is limited work regarding a trained model that works well on totally new systems governed by the same physical law with novel physical parameters.

To consider the second condition, we propose that meta-learning, which aims to train a model well generalized on novel data from observing a few examples, can be potentially key to learning a functional of Hamiltonian as a data-driven method. There have been several representative categories of meta-learning algorithms, such as the metric-based method (Snell et al., 2017; Sung et al., 2018), black-box method (Santoro et al., 2016; Bertinetto et al., 2019), and gradient-based method (Rusu et al., 2019; Flennerhag et al., 2020). Among these methods, we especially focus on the gradient-based method, which is readily compatible with any differentiable model and flexibly applicable to a wide variety of learning problems (Finn et al., 2017; Xu et al., 2018; Hospedales et al., 2020).

One of the most successful algorithms of the gradient-based method is Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017), which consists of a task-specific adaptation process and a meta-optimization process. The key observations supporting its potential are the resemblance between these processes and the identification of the physical laws of the Hamiltonian. The schematic is shown in Figure 1. The task-adaptation process, which adapts the initial model parameters to a task-specific train set, resembles the process of adapting hypothesized governing equations to observations from several physical systems. The meta-optimization process, which updates the initial model parameters by validating each task-specific adapted parameters on a task-specific test set, is similar to correcting the hypothesized governing equations by validating each system-specific Hamiltonian on new data from the corresponding physical systems. In addition, (Raghu et al., 2020) proposed that the recent success behind these meta-learning algorithms was due to providing qualitative shared representation across tasks rather than learning initial model parameters that encourage rapid adaptation (Finn et al., 2017). This hypothesis may support our suggestion that a meta-learner can be efficient in identifying the shared representation of a Hamiltonian. From this point of view, we experiment on several physical systems to verify whether these meta-learning algorithms are beneficial to our desired learning problems. Our contributions are summarized as follows:

- We formulate the problem of identifying the shared representation of unknown Hamiltonian as a meta-learning problem.
- For learning to identify the Hamiltonian representations, we incorporate the HNN architecture on meta-learning algorithms.
- After meta-training the meta-learner, we adapt the model on new systems by learning the data of partial observations and predict the dynamics of the systems as a vector field in phase space.
- We evaluate our method on several physical systems to explore the efficiency of our methods with various experimental settings.

2 PRELIMINARIES

2.1 HAMILTONIAN NEURAL NETWORKS

In Hamiltonian mechanics, the state of a system can be described by the vector of canonical coordinates, $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p})$, which consist of position, $\boldsymbol{q} = (q_1, q_2, ..., q_n)$ and its conjugate momentum, $\boldsymbol{p} = (p_1, p_2, ..., p_n)$ in phase space, where n is the degree of freedom of the system. Then, the time evolution of the system is governed by Hamilton's equations $\dot{\boldsymbol{x}} = \left(\frac{\partial H}{\partial \boldsymbol{p}}, -\frac{\partial H}{\partial \boldsymbol{q}}\right) = \Omega \nabla_{\boldsymbol{x}} H(\boldsymbol{x})$, where $H(\boldsymbol{x}) : \mathbb{R}^{2n} \to \mathbb{R}$ is the Hamiltonian that is conservative during the process and $\Omega = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$ is a $2n \times 2n$ skew-symmetric matrix. From the Hamiltonian equations, the Hamiltonian vector field in phase space, which is interpreted as the time evolution of the system $\dot{\boldsymbol{x}}$, is the symplectic gradient of the Hamiltonian $\Omega \nabla_{\boldsymbol{x}} H(\boldsymbol{x})$, which is determined by the Hamiltonian function and the state of the system itself. Then, the trajectory of the state can be computed by integrating the symplectic gradient of the Hamiltonian. If the Hamiltonian does not depend on the time variable, the Hamiltonian remains constant during the time evolution, because moving along the direction of the symplectic gradient keeps the Hamiltonian constant (Arnol'd, 2013). In (Greydanus et al., 2019), the Hamiltonian function can be approximated by neural networks, H_θ, called HNN. To make the Hamiltonian function constant in motion, the loss of HNN is defined by the distance between the true vector field and the symplectic gradient of H_θ,

$$\mathcal{L}_{HNN} = \left\| \dot{\boldsymbol{x}} - \Omega \nabla_{\boldsymbol{x}} H_\theta(\boldsymbol{x}) \right\|_2^2 . \tag{1}$$

2.2 MODEL-AGNOSTIC META-LEARNING AND FEATURE REUSE HYPOTHESES

A key assumption behind MAML is that separately trained models for each task share meta-initial parameters θ that could be improved rapidly for any task (Finn et al., 2017). Suppose that each given task, $\mathcal{T}_i$, composed of $\mathcal{D}_i = \{\mathcal{D}_i^{tr}, \mathcal{D}_i^{te}\}$, is drawn from a task distribution, $\mathcal{T}_i \sim p(\mathcal{T})$. The learning algorithms usually consist of bi-level optimization processes; (inner-loop) the task-specific adaptation to each train set,

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i} \left(\mathcal{D}_i^{tr}; \theta \right) , \tag{2}$$

and (outer-loop) the meta-optimization on each test set,

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\mathcal{D}_i^{te}; \theta_i'), \tag{3}$$

where θ could be any differentiable model's parameters that are expected to learn the shared representations of various tasks, and α and β are the learning rates of the inner and outer loops, respectively.

Meanwhile, (Raghu et al., 2020) observed that during the inner-loop process, the task-specific distinction of the model parameters θ is mostly from the last layer of the networks, whereas the entire body of the model hardly changed. Therefore, they hypothesized that the body of the model behaves as a shared representation across the different tasks, whereas the head of the model behaves as a task-specific parameters, which is called *feature reuse* hypothesis. From this hypothesis, they proposed a gradient-based meta-learning algorithm called Almost No Inner Loop (ANIL) by slightly modifying MAML by freezing all but updating the last layer of the networks during the inner-loop process. They showed that ANIL performs on par or better than MAML on several benchmarks, and has a computational benefit compared to its counterpart. For the algorithm, when the meta-learner consists of l layers $\theta = (\theta^{(1)}, ..., \theta^{(l-1)}, \theta^{(l)})$, the inner-loop update is modified as

$$\theta_i' = (\theta^{(1)}, ..., \theta^{(l-1)}, \theta^{(l)} - \alpha \nabla_{\theta^{(l)}} \mathcal{L}_{\mathcal{T}_i} \left(\mathcal{D}_i^{tr}; \theta \right)). \tag{4}$$

3 METHOD

3.1 IDENTIFYING SHARED REPRESENTATION OF HAMILTONIAN VIA META-LEARNER

The main goal of our study is to train a model to identify the shared representation of the Hamiltonian using observations of dynamics from several systems that are assumed to be governed by the same

physical law with different physical parameters. From a meta-learning point of view, each system is regarded as a task $\mathcal{T}_i$, where the physical parameters of the system are drawn from the distribution of $p(\mathcal{T})$. The observations of the system $\mathcal{T}_i$ can be split into $\mathcal{D}_i = \{\mathcal{D}_i^{tr}, \mathcal{D}_i^{te}\}$, where $\mathcal{D}_i^{tr}$ and $\mathcal{D}_i^{te}$ denote the task-specific train and test sets, respectively. The observations of both $\mathcal{D}_i^{tr}$ and $\mathcal{D}_i^{te}$ are given by a set of tuples of canonical coordinates $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p})$ and their time derivatives $\dot{\boldsymbol{x}} = (\dot{\boldsymbol{q}}, \dot{\boldsymbol{p}})$ as the ground truth.

For each system, the task-specific model parameters are updated as Equation 2 or Equation 4 by computing the task-specific loss using Equation 1 on each train set $\mathcal{D}_i^{tr}$,

$$\mathcal{L}_{\mathcal{T}_i}(\mathcal{D}_i^{tr}; \theta) = \sum_{(\boldsymbol{x}, \dot{\boldsymbol{x}}) \sim \mathcal{D}_i^{tr}} \left\| \dot{\boldsymbol{x}} - \Omega \nabla_{\boldsymbol{x}} H_\theta(\boldsymbol{x}) \right\|_2^2, \tag{5}$$

and the meta-optimization can be operated on the batch of systems as Equation 3 by minimizing the loss over the batch of physical parameters sampled from $p(\mathcal{T})$. Each loss is computed by evaluating each task-specific adapted model parameters θ_i' to each test set $\mathcal{D}_i^{tr}$,

$$\sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\mathcal{D}_i^{te}; \theta_i') = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \sum_{(\boldsymbol{x}, \dot{\boldsymbol{x}}) \sim \mathcal{D}_i^{te}} \left\| \dot{\boldsymbol{x}} - \Omega \nabla_{\boldsymbol{x}} H_{\theta_i'}(\boldsymbol{x}) \right\|_2^2. \tag{6}$$

Depending on the inner-loop methods, we call the algorithms Hamiltonian Model-Agnostic Meta-Learning (HAMAML) when using Equation 2, and Hamiltonian Almost No Inner-Loop (HANIL) when using Equation 4.

3.2 Learning New System Dynamics from Partial Observations

It can be expected that if the learner is efficient in learning an unknown process, the model can appropriately predict the dynamics of novel systems from their partial observations. To evaluate whether a learner is efficient in identifying the representation of an unknown Hamiltonian, we set up the following validating scheme on several physical systems. After meta-training, novel systems are given as meta-test sets $\mathcal{D}_{new} = \{\mathcal{D}_{new}^{tr}, \mathcal{D}_{new}^{te}\}$ generated from the system's Hamiltonian with novel physical parameters $\mathcal{T}_{new} \sim p(\mathcal{T})$. The train set $\mathcal{D}_{new}^{tr}$ is used to adapt the learner on the new system dynamics ($\theta \rightarrow \theta_{new}'$). The new observed systems are given as K trajectories, which are reminiscent settings of the K-shot learning problem (Snell et al., 2017). Starting from the K initial states, each trajectory is obtained by integrating the time derivatives of the systems from 0s to Ts, with sampling rate $\frac{L}{T}$ which yields L rolled out sequence of states. The test set $\mathcal{D}_{new}^{te}$ is for validating the performance of the adapted model θ_{new}'.

4 Related Works

4.1 Learning Dynamics with Neural Networks

Several works on learning dynamical systems usually use various forms of graph neural networks (Battaglia et al., 2016; Santoro et al., 2017; Sanchez-Gonzalez et al., 2018; Kipf et al., 2018; Battaglia et al., 2018; Sanchez-Gonzalez et al., 2020), where each node usually represents the state of individual objects and each edge between two arbitrary nodes usually represents the relation of the nodes. (Chen et al., 2018) introduced a new type of deep architecture, which is considered as the differentiable Ordinary differential Equations (ODEs) solver parametrized by neural networks, called neural ODEs. Neural ODEs have been developed (Dupont et al., 2019; Finlay et al., 2020) and applied for normalizing flows (Grathwohl et al., 2019; Durkan et al., 2019). The most related to our work, HNN (Greydanus et al., 2019) is learning dynamics that incorporate the Hamiltonian with neural networks. Similarly, there are several works on explicitly learning the Lagrangian (Lutter et al., 2019; Cranmer et al., 2020a).

4.2 Identifying Physical Laws or Governing Equations from Data

In earlier works, symbolic regression was utilized to search the mathematical expressions of the governing equations automatically from observed data (Bongard & Lipson, 2007; Schmidt & Lipson, 2009). These methods have been developed by constructing a dictionary of candidate nonlinear

functions (Brunton et al., 2016; Mangan et al., 2017) or partial differentiations(Rudy et al., 2017; Schaeffer, 2017) by incorporating a sparsity-promoting algorithm to extract the governing equations. Recently, they have been refined through combining with neural networks (Udrescu & Tegmark, 2020; Both et al., 2019; Atkinson et al., 2019) or graph structures (Cranmer et al., 2019; 2020b). In contrast to our work, the existing methods of identifying the governing representations used the symbolic representation underlying the assumptions that the unknown physical laws are expressed by combinations of mathematically known expressions.

5 EXPERIMENTS

5.1 PHYSICAL SYSTEMS

Spring-Mass. Hamiltonian of the system is described by $H = \frac{p^2}{2m} + \frac{k(q-q_0)^2}{2}$, where the physical parameters m, k, and q_0 are the mass, spring constant, and equilibrium position, respectively. q and p are the position and conjugate momentum of the system, respectively.

Pendulum. Hamiltonian of the system is described by $H = \frac{p^2}{2ml^2} + mgl(1 - \cos(q - q_0))$, where the physical parameters m, l, and q_0 are the mass, pendulum length, and equilibrium angle from the vertical, respectively. q and p are the pendulum angle from the vertical and conjugate momentum of the system, respectively. g denotes the gravitational acceleration.

Kepler problem. The system consists of two objects with attracted each other by gravitational force. Hamiltonian of the system is described by $H = \dfrac{p_1^2 + p_2^2}{2m} - \dfrac{GMm}{\sqrt{q_1^2 + q_2^2}}$, where the physical parameters M and m are the mass of the two bodies and we choose the object M as the center of the coordinate system. The state is represented by $\boldsymbol{q} = (q_1, q_2)$ and $\boldsymbol{p} = (p_1, p_2)$ which denote the position and conjugate momentum of the object m in two-dimensional space, respectively. G denotes the gravitational constant.

Double Pendulums. The system consists of two pendulums with one attached to the other. Hamiltonian of the system is described by $H =$

$$\frac{m_2 l_2^2 p_1^2 + (m_1 + m_2) l_1^2 p_2^2 - 2m_2 l_1 l_2 p_1 p_2 \cos(q_1 - q_2)}{2m_2 l_1^2 l_2^2 (m_1 + m_2 \sin^2(q_1 - q_2))} - (m_1 + m_2) g l_1 \cos q_1 - m_2 g l_2 \cos q_2, \tag{7}$$

where the physical parameters, m_1, m_2, l_1, and l_2 denote the mass of two objects, and the length of two strings, respectively. The state is represented by $\boldsymbol{q} = (q_1, q_2)$ and $\boldsymbol{p} = (p_1, p_2)$, which denote the angle and angular momentum of m_1 and m_2 in two-dimensional space, respectively. g denotes the gravitational acceleration.

5.2 EXPERIMENTAL SETTINGS

Datasets. During the meta-training, we generate 10,000 tasks for meta-train sets of all systems. Each meta-train set consists of task-specific train set and test set given by 50 randomly sampled point states $\boldsymbol{x}$ and their time derivatives $\dot{\boldsymbol{x}}$ in phase space with task-specific physical parameters. The distributions of the sampled states and physical parameters for each physical process are described in Appendix A. During the meta-testing, the distributions of sampled states and physical parameters are the same as in the meta-training stage. To evaluate learners on the efficacy of identifying Hamiltonian of the new systems, meta-test sets are constructed as the following ways.

(1) Observing the new systems as point dynamics in phase space: $\mathcal{D}_{new}^{tr}$ consists of randomly sampled points in phase space with $K = \{25, 50\}$ and $L = 1$, and $\mathcal{D}_{new}^{te}$ consists of equally fine-spaced points in phase space.

(2) Observing the new systems as trajectories in phase space: $\mathcal{D}_{new}^{tr}$ consists of randomly sampled trajectories and $\mathcal{D}_{new}^{te}$ consists of equally fine-spaced points in phase space. The number of sampled trajectories with $K = \{5, 10\}$ and $L = 5$ sequences during $T = 1s$.

The trajectories are obtained by integrating the symplectic gradient of the Hamiltonian from initial states using the adaptive step-size Runge-Kutta method. Fine-spaced test sets consist of 50 equally spaced grids for each coordinate in the region of the phase space where we sampled the point states.

Table 1: Averaged MSEs across the test sets of 10 new systems from adapting to the corresponding train sets with 10 epochs.

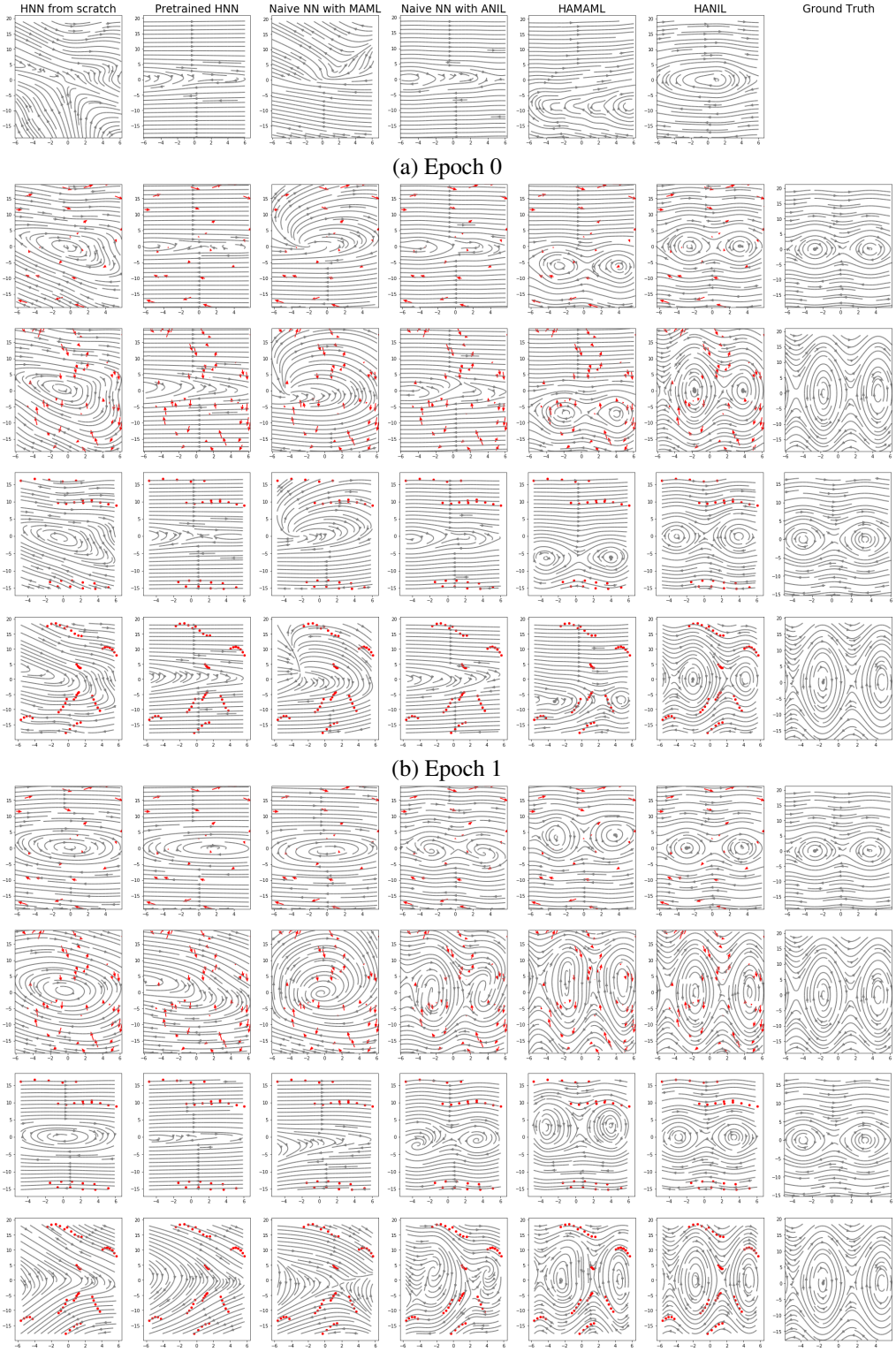
Learner	Spring-Mass				Pendulum			
	Point dynamics		Trajectories		Point dynamics		Trajectories	
	25-shot	50-shot	5-shot	10-shot	25-shot	50-shot	5-shot	10-shot
HNN from Scratch	82.28	82.08	84.73	82.39	6.46	6.08	6.70	6.55
Pretrained HNN	8.21	4.94	11.93	6.60	5.38	5.11	5.89	5.84
Naive NN with MAML	159.81	155.77	195.65	162.32	5.78	5.29	10.29	8.29
Navie NN with ANIL	16.64	16.24	36.02	19.56	2.32	2.00	3.30	2.79
HAMAML	2.94	2.35	4.21	1.57	1.46	1.37	4.651	2.91
HANIL	**0.015**	**0.010**	**0.42**	**0.11**	**0.022**	**0.0033**	**0.59**	**0.035**

Evaluation metric. The evaluation metric for the learners is the average of the mean squared errors (MSEs) between the symplectic gradients of predicted Hamiltonian $\Omega \nabla_{\boldsymbol{x}} H_\theta(\boldsymbol{x})$ and true vector fields $\dot{\boldsymbol{x}}$ across the test sets of new systems $\mathcal{D}^{te}_{new}$ from adapting the learners to the corresponding train set of the new systems $\mathcal{D}^{tr}_{new}$. In all tasks, the averaged MSEs of randomly sampled 10 new systems are averaged for evaluating the learners after 10 epochs adapting to each $\mathcal{D}^{tr}_{new}$. We also predict the state trajectories and energies by integrating the output vector fields of the learners that are adapted on a new system by observing samples of state dynamics. Following (Greydanus et al., 2019), we evaluate MSEs of the predicted trajectories and energies.

Baselines. We took several learners as baselines to assess the efficacy of our proposed methods, such as (1) training HNN on $\mathcal{D}^{tr}_{new}$ from scratch (random initialization), (2) pretrained HNN on the meta-train set, (3) meta-trained naive fully connected neural networks (Naive NN), which are given the inputs $\boldsymbol{x}$ and the outputs $\dot{\boldsymbol{x}}$ with MAML, and (4) with ANIL.

6 RESULTS

The quantitative results of meta-testing performance are shown in Table 1. It can be verified that HANIL outperforms the others for all experimental settings in all physical systems. The case of observing 25-shot and 5-shot trajectories have the same number of given samples in the phase space, and so on the case of the observing 50-shot and 10-shot trajectories. Comparing the cases of observation from point dynamics and trajectories with the same number of given samples, learned dynamics from observing point dynamics are more accurate than those from observing trajectories. In Figure 2, predicted new pendulum systems dynamics by adapting the learners from observing partial observations are represented as phase portraits by epochs. Note that those of spring-mass systems are shown in Figure 4 in Appendix C. The first rows are the initial output vectors in the phase space from the learners. During the adaptation on the observations, the output vectors are evolved to fit on the observations with keeping their own prior belief or representation learned from the meta-train set. HNN from scratch fails to predict the new systems dynamics from partial observations because the numbers of samples are too small to train HNN without any inductive bias. Pretrained HNN also fails to predict the new systems dynamics because it dose not aware the exact shared representation of the Hamiltonian. Naive NN with MAML and ANIL also fails to predict the dynamics because naive neural networks are hard to grasp the continuous and conservative structure of the vector fields. Thus, the phase portraits of them are discontinuous or dissipative-like shape. HANIL can accurately predict the new systems dynamics from partial observations with few gradient steps, while HAMAML is slower than HANIL to adapt the true vector fields because of the larger number of parameters to update in adaptation process. In Figure 3, we also evaluate the learners adapted on the new systems through their predictions of state and energy trajectories from a initial states of the new systems. Adapted HANIL on new systems predicts the state and energy trajectories almost the same as the ground truth with relative small errors from the ground truth, whereas the others fail to predict the right trajectories and energies of the system.

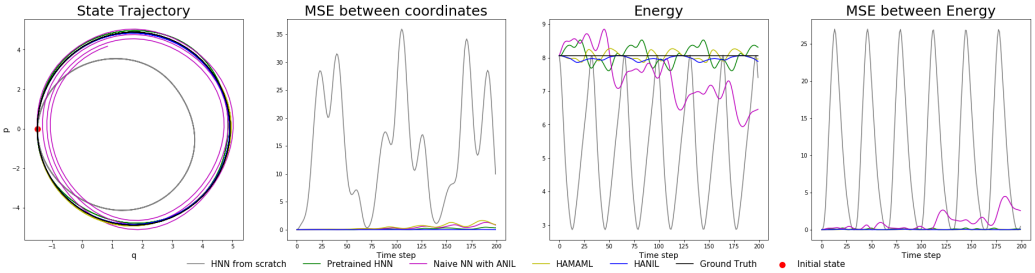


(a) Epoch 0

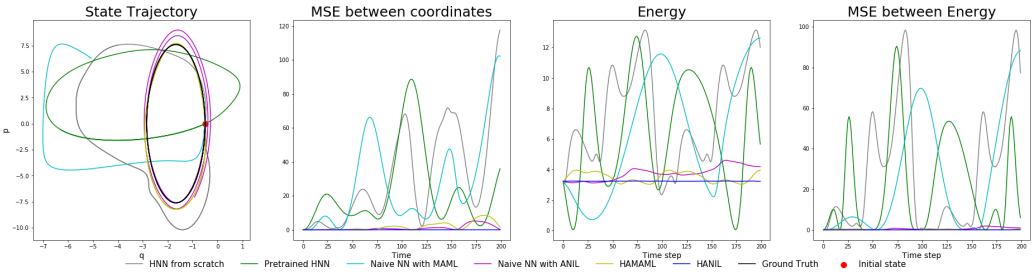
(b) Epoch 1

(c) Epoch 10

Figure 2: Predicted vector fields by adapting the learners on new pendulum systems by epochs. Red arrows and points denote the given point dynamics and trajectories, respectively. Gray streamlines represent the learned vector fields in phase space. The x-axis and y-axis denote q and p, respectively. At each epoch, observing the new systems as 25 and 50 points dynamics, and 5 and 10 trajectories are represented in order. Note that the first and third rows represent the same newly observed systems, and the second and fourth rows represent the same newly observed systems at each epoch.



(a) Spring-Mass



(b) Pendulum

Figure 3: Predicted state trajectories, energies and their corresponding MSEs starting from an initial state over time. The predicted values are computed by the learners adapted to 50 randomly sampled point dynamics of a new system in phase space with 50 epochs.

7 Conclusions

Author Contributions

Acknowledgments

References

Vladimir Igorevich Arnol'd. *Mathematical methods of classical mechanics*, volume 60. Springer Science & Business Media, 2013.

Vladimir I Arnol'd, Sergeĭ Petrovich Novikov, et al. *Symplectic geometry*. Springer, 2001.

Steven Atkinson, Waad Subber, Liping Wang, Genghis Khan, Philippe Hawi, and Roger Ghanem. Data-driven discovery of free-form governing differential equations. *arXiv preprint arXiv:1910.05117*, 2019.

Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al. Interaction networks for learning about objects, relations and physics. In *Advances in Neural Information Processing Systems*, pp. 4502–4510, 2016.

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Luca Bertinetto, Joao F. Henriques, Philip Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *International Conference on Learning Representations*, 2019.

Josh Bongard and Hod Lipson. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104(24):9943–9948, 2007.

Gert-Jan Both, Subham Choudhury, Pierre Sens, and Remy Kusters. Deepmod: Deep learning for model discovery in noisy data. *arXiv preprint arXiv:1904.09406*, 2019.

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.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing systems*, pp. 6571–6583, 2018.

Zhengdao Chen, Jianyu Zhang, Martin Arjovsky, and Léon Bottou. Symplectic recurrent neural networks. In *International Conference on Learning Representations*, 2020.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020a.

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. *arXiv preprint arXiv:2006.11287*, 2020b.

Miles D Cranmer, Rui Xu, Peter Battaglia, and Shirley Ho. Learning symbolic physics with graph networks. *arXiv preprint arXiv:1909.05862*, 2019.

Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. In *Advances in Neural Information Processing Systems*, pp. 3140–3150, 2019.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In *Advances in Neural Information Processing Systems*, pp. 7511–7522, 2019.

Kang Feng and Mengzhao Qin. *Symplectic geometric algorithms for Hamiltonian systems*. Springer, 2010.

Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam M Oberman. How to train your neural ode. *arXiv preprint arXiv:2002.02798*, 2020.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pp. 1126–1135, 2017.

Sebastian Flennerhag, Andrei A. Rusu, Razvan Pascanu, Francesco Visin, Hujun Yin, and Raia Hadsell. Meta-learning with warped gradient descent. In *International Conference on Learning Representations*, 2020.

Herbert Goldstein, Charles Poole, and John Safko. *Classical mechanics*. American Association of Physics Teachers, 2002.

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, and David Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, 2019.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, pp. 15379–15389, 2019.

Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *arXiv preprint arXiv:2004.05439*, 2020.

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

Thomas N. Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard S. Zemel. Neural relational inference for interacting systems. In *International Conference on Machine Learning*, pp. 2693–2702, 2018.

Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks: Using physics as model prior for deep learning. In *International Conference on Learning Representations*, 2019.

Niall M Mangan, J Nathan Kutz, Steven L Brunton, and Joshua L Proctor. Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2204):20170009, 2017.

Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. In *International Conference on Learning Representations*, 2020.

Linda E Reichl. *A modern course in statistical physics*. American Association of Physics Teachers, 1999.

Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.

Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. In *International Conference on Learning Representations*, 2019.

Jun John Sakurai and Eugene D Commins. *Modern quantum mechanics, revised edition*. American Association of Physics Teachers, 1995.

Rick Salmon. Hamiltonian fluid mechanics. *Annual review of fluid mechanics*, 20(1):225–256, 1988.

Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin A. Riedmiller, Raia Hadsell, and Peter Battaglia. Graph networks as learnable physics engines for inference and control. In *International Conference on Machine Learning*, pp. 4467–4476, 2018.

Alvaro Sanchez-Gonzalez, Victor Bapst, Kyle Cranmer, and Peter Battaglia. Hamiltonian graph networks with ode integrators. *arXiv preprint arXiv:1909.12790*, 2019.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W Battaglia. Learning to simulate complex physics with graph networks. *arXiv preprint arXiv:2002.09405*, 2020.

Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International Conference on Machine Learning*, pp. 1842–1850, 2016.

Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. A simple neural network module for relational reasoning. In *Advances in Neural Information Processing systems*, pp. 4967–4976, 2017.

Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473 (2197):20160446, 2017.

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

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pp. 4077–4087, 2017.

Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1199–1208, 2018.

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

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

Zhongwen Xu, Hado P van Hasselt, and David Silver. Meta-gradient reinforcement learning. In *Advances in Neural Information Processing Systems*, pp. 2396–2407, 2018.

Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. Symplectic ode-net: Learning hamiltonian dynamics with control. In *International Conference on Learning Representations*, 2020.

A SYSTEMS DETAILS

A.1 SPRING-MASS

The physical parameters are randomly sampled from $(m, k, q_0) \in ([0.5, 5], [0.5, 5], [-5, 5])$. The initial states are randomly sampled from $(q, p) \in ([-10, 10]^2)$.

A.2 PENDULUM

The physical parameters are randomly sampled from $(m, l, q_0) \in ([0.5, 5], [0.5, 5], [-\pi, \pi])$. We fix the gravitational acceleration as $g = 1$. The initial states are randomly sampled from $(q, p) \in ([-2\pi, 2\pi], [-20, 20])$.

A.3 KEPLER PROBLEM

The masses are randomly sampled from $(M, m) \in ([0.5, 5]^2)$. We fix the gravitational constant as $G = 1$. The initial states are randomly sampled from $(\boldsymbol{q}, \boldsymbol{p}) \in ([0.5, 5.0]^4)$.

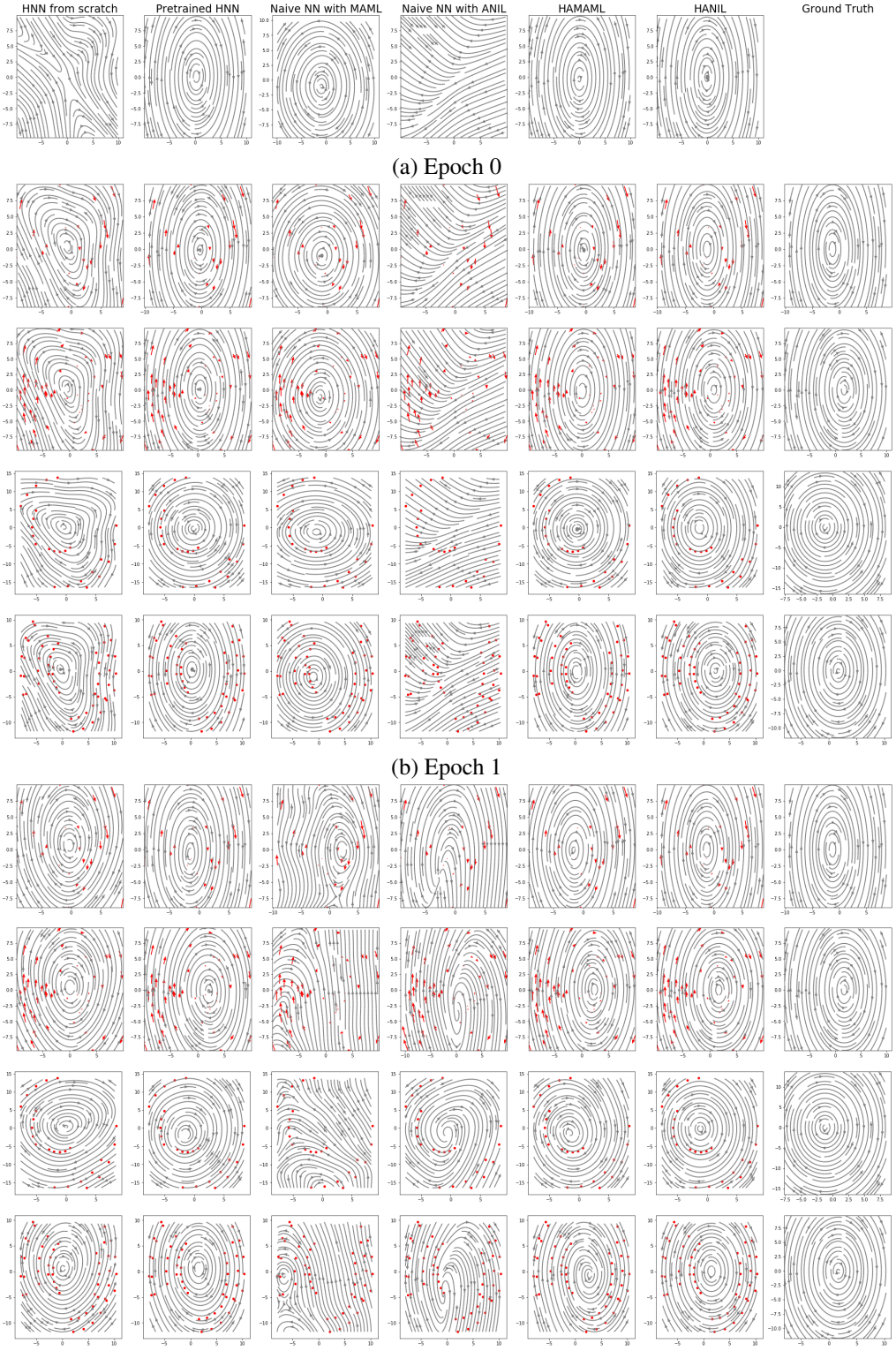
A.4 DOUBLE PENDULUM

Respective masses are randomly sampled from $(m_1, m_2, l_1, l_2) \in ([0.5, 2.0]^4)$. We fix the gravitational acceleration as $g = 1$. The initial states are randomly sampled from $(\boldsymbol{q}, \boldsymbol{p}) \in ([-\pi, \pi]^2, [-5, 5]^2)$.

B IMPLEMENTATION DETAILS

During meta-training or pretraining, we use the Adam optimizer (Kingma & Ba, 2015) on outer-loop with learning rate of 0.001 and use gradient descent on inner-loop with learning rate of 0.002. For all systems, we set the number of task batches of 10, inner gradient updates of 5, and episodes of outer loop of 100 for meta-optimization. During the meta-testing, we also use the Adam optimizer with a learning rate 0.002. There is no weight decay for all. We use a 4-fully connected layer of size 64 on both naive NN and HNN. We use softplus activations that outperform any other nonlinearity, for each layer, except the last layer.

C ADDITIONAL RESULTS



(a) Epoch 0

(b) Epoch 1

(c) Epoch 10

Figure 4: Predicted vector fields by adapting the learners on new spring-mass systems by epochs. Red arrows and points denote the given point dynamics and trajectories, respectively. Gray streamlines represent the learned vector fields in phase space. The x-axis and y-axis denote q and p, respectively. At each epoch, observing the new systems as 25 and 50 points dynamics, and 5 and 10 trajectories are represented in order. Note that the first and third rows represent the same newly observed systems, and the second and fourth rows represent the same newly observed systems at each epoch.