TDHNN: Time-Dependent Hamiltonian Neural Networks

Shaan A. Desai[1,2] **Marios Mattheakis**[2] **David Sondak**[2] **Pavlos Protopapas**[2] **Stephen Roberts**[1]
Machine Learning Research Group
University of Oxford [1]
School of Engineering and Applied Science
Harvard University [2]

Abstract

Learning the dynamics of complex physical systems requires models with well chosen inductive biases. Recently, it was shown that neural networks designed to learn a Hamiltonian and exploit Hamilton's equations significantly outperform existing approaches in predicting trajectories of autonomous systems that depend implicitly on time. Despite this success, many real world physical systems are non-autonomous and depend explicitly on time. In this paper, we address this challenge by embedding a modified Port-Hamiltonian into our neural networks that extends the general Hamiltonian to capture damped and forced systems alike. We show that our network can learn systems as complex as the Duffing equation in the chaotic regime and recover the underlying Hamiltonian, force and damping terms. For such a system, we are also able to visually recover its Poincaré section - a crucial map in identifying chaotic trajectories.

1 Introduction

Neural networks (NNs), as universal function approximators [8], have shown resounding success across a host of domains including image segmentation, machine translation and material property predictions. [7, 4, 17, 18]. However, their performance in learning physical systems has often been limited [6, 11]. New research in *scientific machine learning* - a field that tackles scientific problems with domain-specific machine learning methods, is paving a way to address these challenges. For example, it has been shown that physically informed inductive biases embedded in networks, such as Hamiltonian mechanics [10, 6], Lagrangians [3, 9], Ordinary Differential Equations (ODEs) [2], physics-informed networks [12] and Graph Neural Networks [1, 14] can significantly improve learning over vanilla neural networks in complex physical domains. Furthermore, it has also been shown that such systems can be adapted to learn trajectories of controlled dynamics [9, 19]. Despite this extensive research, there is no method that accounts for explicit-time dependence, a crucial component in forced dynamical systems. Additionally, many of the existing approaches do not outline how to account for energy dissipation. Inspired by [19], we show that by embedding a Port-Hamiltonian into a neural network we can learn the dynamics of explicit time-dependent physical systems. We benchmark our network on a range of tasks from the simple mass-spring system to a chaotic Duffing equation for a relativistic particle. The proposed network consistently outperforms other approaches while actively learning both the driving force and the damping coefficient.

Proceedings of the 24[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

2 Background

2.1 Hamiltonian Neural Networks

Recently, the authors of [6] demonstrate that the dynamics of an energy conserving autonomous system can be accurately learned by guiding a neural network to predict a Hamiltonian - an important representation of a dynamical system that generalizes classical mechanics. Considering dynamical systems of M objects, the Hamiltonian $\mathcal{H}$ is a scalar function of a position vector $\mathbf{q} = (q_1, q_2, ..., q_M)$ and momentum vector $\mathbf{p} = (p_1, p_2, ..., p_M)$ of a given system that obeys Hamilton's equations (1).

$$\dot{\mathbf{q}} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} \tag{1}$$

As such, it is noted in [6] that by parametrizing the Hamiltonian with a neural network e.g. $H_\theta(\mathbf{q}, \mathbf{p})$ where θ represents the parameters of a deep neural

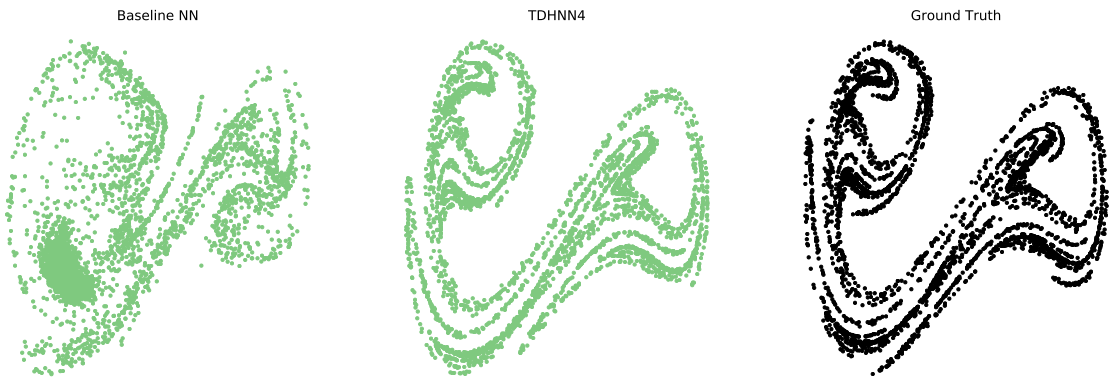


Figure 1: Poincaré Sections of a Duffing oscillator in a chaotic regime. Both Baseline NN and TDHNN4 are trained for 20000 iterations with 2000 data points. TDHNN4 significantly outperforms Baseline NN at recovering the ground truth Poincaré section of a test point not in the training set.

network and $\mathbf{q}, \mathbf{p}$ are the inputs to the network, one can learn the system's dynamics by differentiating (via autograd) the Hamiltonian with respect to its inputs. The derivatives of the Hamiltonian $\left[\frac{\partial \mathcal{H}}{\partial \mathbf{p}}, -\frac{\partial \mathcal{H}}{\partial \mathbf{q}}\right]$ provide us with time derivatives of the input state vector $[\dot{\mathbf{q}}, \dot{\mathbf{p}}]$. Once trained, this network therefore yields an approximate time derivative generator for the state $[\mathbf{q}, \mathbf{p}]$ and as such can be used in a numerical integrator to evolve an initial condition $[\mathbf{q}_t, \mathbf{p}_t]$, as in Eq.2.

$$(\mathbf{q}, \mathbf{p})_{t+1} = (\mathbf{q}, \mathbf{p})_t + \int_t^{t+1} \left[\frac{\partial \mathcal{H}}{\partial \mathbf{p}}, -\frac{\partial \mathcal{H}}{\partial \mathbf{q}}\right] \mathrm{d}t. \quad (2)$$

In addition, the vector field $\mathbf{S} = \left[\frac{\partial \mathcal{H}}{\partial \mathbf{p}}, -\frac{\partial \mathcal{H}}{\partial \mathbf{q}}\right]$ is a symplectic gradient. This means that $\mathcal{H}$ remains constant as long as the state vectors are integrated along $\mathbf{S}$. This result links the Hamiltonian with the total energy of the system such that $\mathcal{H}(\mathbf{q}, \mathbf{p}) = E_{\text{tot}}$ for many autonomous physical systems. Therefore, the Hamiltonian is a powerful inductive bias that can be utilized to evolve a physical state while maintaining energy conservation. However, this formulation does not readily generalize to damped or forced system.

2.2 Port-Hamiltonians

The Port-Hamiltonian is a formalism that allows us to generalize Hamilton's equations to incorporate damping and forcing terms. A well known method that embeds the Port-Hamiltonian in a neural network [19] shows how to represent a general form for such a system. In [19] the Port-Hamiltonian is represented as:

$$\begin{bmatrix}\dot{\mathbf{q}} \\ \dot{\mathbf{p}}\end{bmatrix} = \left(\begin{bmatrix}0 & \mathbf{I} \\ -\mathbf{I} & 0\end{bmatrix} + \mathbf{D}(\mathbf{q})\right)\begin{bmatrix}\frac{\partial \mathcal{H}}{\partial \mathbf{q}} \\ \frac{\partial \mathcal{H}}{\partial \mathbf{p}}\end{bmatrix} + \begin{bmatrix}\mathbf{0} \\ \mathbf{G}(\mathbf{q})\end{bmatrix}\mathbf{u}, \quad (3)$$

where $\mathbf{D}(\mathbf{q})$ is a matrix to account for damping and $\mathbf{G}(\mathbf{q})\mathbf{u}$ is an external control term where $\mathbf{u}$ is a temporal control input and $\mathbf{G}(\mathbf{q})$ is a spatial matrix.

This general formalism allows us to collapse to a classical Hamiltonian system by simply setting $\mathbf{D}$ and $\mathbf{u}$ to zero. In this setting, the vector $[\mathbf{q}, \mathbf{p}, \mathbf{u}]$ is provided as input to the system. The matrix $\mathbf{D}(\mathbf{q})$ is assumed to be semi-positive definite and $\mathbf{G}(\mathbf{q})$ is typically a non-linear scaling of $\mathbf{q}$.

3 Related Work

Numerous recent methods highlight how incorporating physically-informed inductive biases into neural networks.

Functional Priors

Functional priors, for example, embed the full functional form of an equation into the neural network. Physics-informed Neural Networks (PINNs) [12, 13] and Hamiltonian Nets [10] look at directly embedding the equations of motion into the loss function. While PINNs are data-driven approaches that rely on autograd to compute partial-derivatives of a hidden state, Hamiltonian Nets are data-independent approaches that pre-specify the full functional form of a system-specific Hamiltonian in the loss function.

Integrator Priors

Many recent methods seek to use a NN to parametrize the time derivatives of a state vector and then use this learned parametrization in a numerical integrator such as a Runge-Kutta method. The authors of NeuralODE [2] show that embedding the integrator into the learning process induces a continuous depth neural network for large time-step predictions. As such, the number of gradient computations scales exponen-

tially with the number of integrative time steps. This study re-introduces the adjoint method to ensure linear scaling of the gradient computations. Other work such as [20], theoretically shows the importance of using symplectic integrators over Runge-Kutta methods to evolve Hamiltonian systems.

Graph based methods

In [1], the authors detail how a Graph Neural Network, designed to capture the relational structure of systems, can be used to learn interacting particle systems. This work has been exploited in numerous advances [16, 15, 3].

Lagrangian/Hamiltonian Methods

While [3] and [6] propose to explicitly learn Hamiltonian and Lagrangian functions, [9] shows how one can exploit the Lagrangian equation to learn both a function which can predict the generalized forces as well as use generalized forces and coordinate information to predict the next state. The authors show that their inverse model can accurately predict a controlled double pendulum, which is pertinent for controlled robots.

Latest Advances

Recent work, studies Constrained HNN [5], an approach that looks at enforcing Cartesian coordinates in Hamiltonians and Lagrangians with holonomic constraints. The paper emphasises that such a framework works significantly better at learning deeply complex physical systems such as the 5-pendulum.

Despite these significant breakthroughs, there is no existing method that investigates explicit time-dependence and damping in such systems, elements that are often found in real world problems. As such, we outline a novel technique to address this challenge.

4 Method

Theory

Our method, TDHNN4, takes the form of the Port-Hamiltonian described earlier with two modifications. First, our approach exploits the fact that most damped systems consist of a non-zero, state-independent term in the lower right quadrant, so we replace the damping matrix $\mathbf{D}(\mathbf{q})$ with a single parameter independent of $\mathbf{q}$ in the lower right half of the matrix. Secondly, in order to generalize to time dependent forcing, we replace $\mathbf{G}(\mathbf{q})\mathbf{u}$ with $\mathbf{F}(t)$. The resulting representation, as highlighted in Eq.(4), is now general enough to handle many well-known forced systems but also specific enough to tackle learning in complex physical domains.

$$\begin{bmatrix}\dot{\mathbf{q}}\\ \dot{\mathbf{p}}\end{bmatrix} = \left(\begin{bmatrix}0 & \mathbf{I}\\ -\mathbf{I} & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & \mathbf{D}\end{bmatrix}\right)\begin{bmatrix}\frac{\partial \mathcal{H}}{\partial \mathbf{q}}\\ \frac{\partial \mathcal{H}}{\partial \mathbf{p}}\end{bmatrix} + \begin{bmatrix}0\\ \mathbf{F}(t)\end{bmatrix}. \quad (4)$$

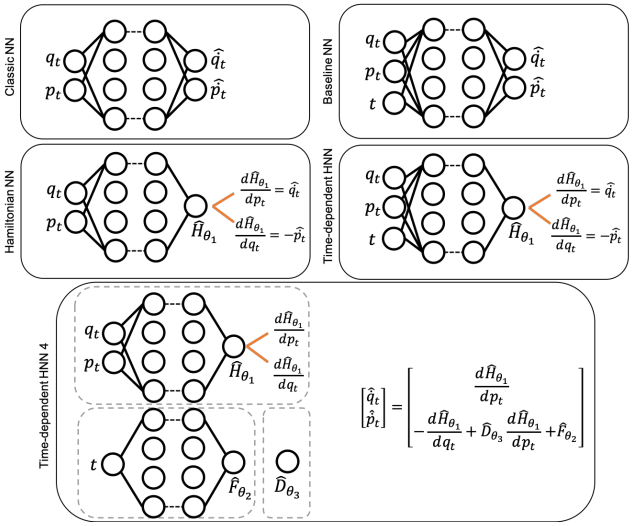


Figure 2: Architectures used to learn dynamics in this paper. The naive extension of classic NN and Hamiltonian NN (top left) is to incorporate time as an additional input variable (top right). Our innovation, which exploits Port-Hamiltonians, explicitly learns the force F_{θ_2} as well as the damping coefficient D_{θ_3}.

The architecture for our method is shown in Figure 2. We embed the Port-Hamiltonian by learning 3 main components: the Hamiltonian $\mathcal{H}$, the driving force $\mathbf{F}$ and the damping term $\mathbf{D}$ each parametrized by a neural network.

Training

To train the network, we feed in a state-vector $[\mathbf{q}, \mathbf{p}, t]$ to our model. The first component, $\mathcal{H}_{\theta_1}$, consists of three hidden layers designed to predict $\mathcal{H}$ from $[\mathbf{q}, \mathbf{p}]$ input data. The second component, $\mathbf{D}_{\theta_3}$, consists of a single weight parameter (i.e. θ_3 is a single node) designed to learn the damping $\mathbf{D}$. The third neural-network consists of three hidden layers designed to predict $\mathbf{F}(t)$. The output of each component is combined as per the equations in Fig.2 to obtain predictions of the state time derivatives $[\hat{\dot{\mathbf{q}}}, \hat{\dot{\mathbf{p}}}]$. Note that the Hamiltonian learned is a single value for each input that is differentiated (via autograd) with respect to the input $[\mathbf{q}, \mathbf{p}]$ to obtain components of the predicted time derivative. We can think of these as being the state

time derivatives in the unforced and undamped setting. Using these predictions, we define our loss function as:

$$\mathcal{L} = ||\hat{\dot{\mathbf{q}}}_t - \dot{\mathbf{q}}_t||_2^2 + ||\hat{\dot{\mathbf{p}}}_t - \dot{\mathbf{p}}_t||_2^2 + \lambda_F ||\mathbf{F}_{\theta_2}(t)||_2^2 + \lambda_D ||\mathbf{D}_{\theta_3}||_2^2 \tag{5}$$

The choice of hyper-parameters λ_F and λ_D is made via a grid search across $[10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}]$, for each parameter, that generates the lowest validation loss (details in Appendix D). For our experiments we use 200 nodes per layer and find that most activations, including tanh, sin and cos yield comparable results. To benchmark our method, we use a simple MLP with position,momentum and time as input which is designed to predict the state time derivatives (see Fig.2). We also take the straightforward extension of HNN to include time as variable input.

Training Dataset

The training dataset for each of the systems later described is carried out by sampling initial conditions $\mathbf{q}, \mathbf{p}$ and using a 4th order Runge-Kutta integrator to evolve the system from time $t = 0$ to $t = T_{\text{max}}$. At each iteration, the next state as well as the state derivatives are recorded. This is typically done for multiple initial conditions and then passed to the network for training. Details for how initial conditions are sampled are described independently for each system in the results section.

Testing

Once trained, each of the networks in Fig.2 can approximate $[\dot{\mathbf{q}}, \dot{\mathbf{p}}]$. As such, they can be used in a scientific integrator such as a Runge-Kutta 4th order method to integrate initial conditions not in the training dataset from $t = 0$ to $t = T_{\text{max}}$ which we refer to as a state rollout. We measure the performance of the network by comparing the predicted state rollout with the ground truth (measured using a 4th order Runge-Kutta integrator and the ground truth time derivatives). We do this by computing a mean squared error across the state variables and across time. Therefore:

$$\text{MSE}_{\text{state}} = \frac{1}{2N}\left(\sum_{i=1}^{N}(\mathbf{q}_i - \hat{\mathbf{q}}_i)^2 + \sum_{i=1}^{N}(\mathbf{p}_i - \hat{\mathbf{p}}_i)^2\right), \tag{6}$$

$$\text{MSE}_{\text{energy}} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathcal{H}(\mathbf{q}_i, \mathbf{p}_i) - \mathcal{H}(\hat{\mathbf{q}}_i, \hat{\mathbf{p}}_i)\right)^2, \tag{7}$$

where $N = T_{\text{max}}/\Delta t$. We typically compute these terms for multiple different initial conditions during inference and average across them (see Fig. 3 c).

Procedure

1. Obtain state variable data $[\mathbf{q}, \mathbf{p}, t]$ and time derivatives $[\dot{\mathbf{q}}, \dot{\mathbf{p}}]$ from trajectories of a given system.
2. Provide state-variable information to TDHNN4 which learns a Hamiltonian, force and damping term to predict $[\hat{\dot{\mathbf{q}}}, \hat{\dot{\mathbf{p}}}]$.
3. Minimize predictions with ground truth time-derivatives as in Eqn.(5).
4. Once trained, use the network as a time-derivative generator in a scientific integrator to integrate a set of initial conditions $[q_0, p_0]$ to $[q_T, p_T]$.

5 Results

We benchmark the performance of our models on numerous datasets that cover time-independent systems to complex chaotic forced systems.

5.1 Simple Mass-Spring System

We begin our analysis with a simple mass-spring system, obeying Hooke's Law from classical physics. The Hamiltonian for this is written as:

$$\mathcal{H} = \frac{1}{2}kq^2 + \frac{1}{2m}p^2 \tag{8}$$

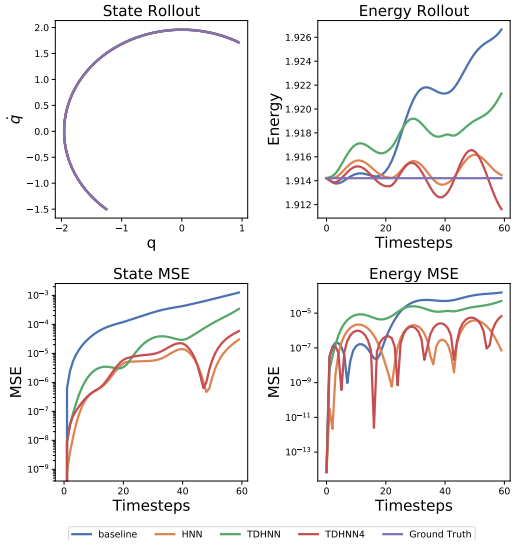
where k is the spring constant and m is the mass of the object.

Training: Without loss of generalization, we set k and m to 1 for our experiments. We randomly sample 25 initial training conditions $[q_0, p_0]$ that satisfy $q_0^2 + p_0^2 = r_0^2$ where $1 \leq r_0 \leq 4.5$ similar to [6] and integrate each using a 4th order RK method with $\Delta t = 0.05$ and $T_{\text{max}} = 3.05$. The initial conditions are sampled in a disc with radii between 1 and 4.5 to sample energies within a range.

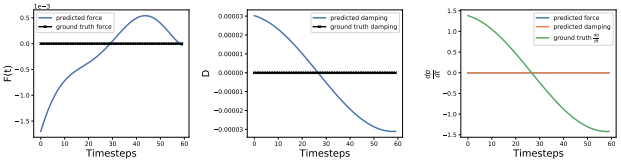
Test: At test time, we sample 25 initial conditions, not in the training set, and evolve them to $T_{\text{max}} = 6.1$. The test error in Fig.3 is measured as the MSE between the ground truth and predicted state and energy values for a given trajectory. We also show the average state and energy MSE across 25 trajectories.

We investigate this simple system to show that learning a separate, regularized forcing term results in much better state and energy predictions in comparison to TDHNN and the baseline NN. The main reason for this is that, while TDHNN and the baseline NN learn

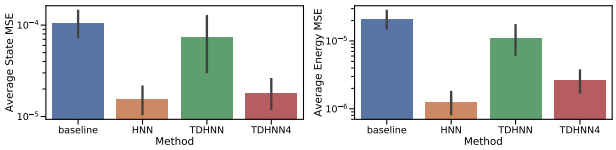
dynamics with time-steps within the training regime, they typically cannot generalize to unseen time-steps. Learning a separate forcing term and regularizing it keeps the time component independent of the Hamiltonian and therefore allows us to closely match the performance of classical time-implicit HNN.



(a) State and energy rollout/MSE of an initial condition in the test set.



(b) Learnt force and damping terms and their contribution to dp/dt



(c) Rollout state and energy MSE averaged across 25 initial test states

Figure 3: The simple mass spring system has no explicit time dependence. We see that TDHNN4 can almost recover the dynamics as well as HNN. Baseline and TDHNN are unable to achieve the same test state error as they are only reliable for time steps that are within the training regime.

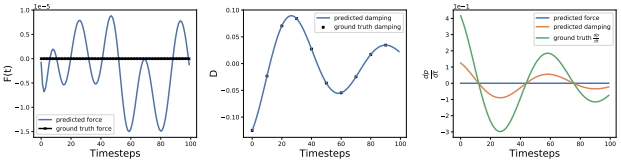
5.2 Damped Mass-Spring System

We extend the simple mass-spring system to include a damping term that reduces the initial energy of the system over time. Of course, the inclusion of this term violates energy conservation and therefore we cannot write a Hamiltonian for such a system (see Appendix A.1 for details).

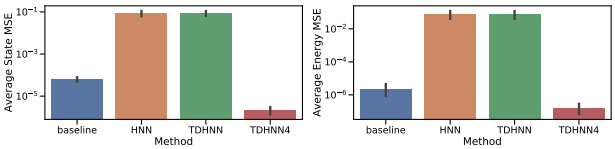
However, using Port-Hamiltonians, it is straightforward to account for the damping term without explicitly defining a full Hamiltonian. By combining a regular mass-spring Hamiltonian $\mathcal{H}_{\text{reg}}$ with an additional term that learns the damping, we can disentangle the two terms for learning. Indeed, there are alternative methods to do this, including generalised Hamiltonian decomposition which we do not explore here.

Training: We have 20 initial training conditions, with position and momentum uniformly sampled in $[-1, 1]^2$. Each trajectory is evolved until $T_{\text{max}} = 30.1$ with a $\Delta t = 0.1$. We fix the damping coefficient $\delta = 0.3$.

Test: At inference, we compute the average rollout MSE of 25 unseen initial conditions sampled and evaluated the same way as training, and report the average state and energy rollout MSE for initial conditions that are not in the training set (see Fig.4).



(a) Learnt Force and Damping terms and their contribution to dp/dt



(b) Rollout state and energy MSE averaged across 25 initial test states

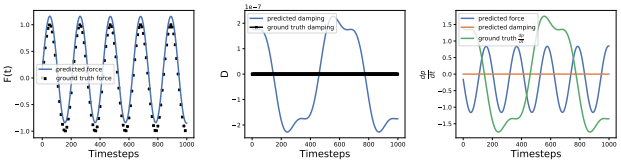
Figure 4: Damped mass spring setting: Baseline and TDHNN4 recover the underlying dynamics. TDHNN4 is also able to accurately learn the damping coefficient since the predicted damping is indistinguishable from the ground truth.

We can see that both baseline NN and TDHNN4 recover the dynamics well, whereas HNN and TDHNN struggle to learn the damping as there is no direct way of learning a Hamiltonian with damping.

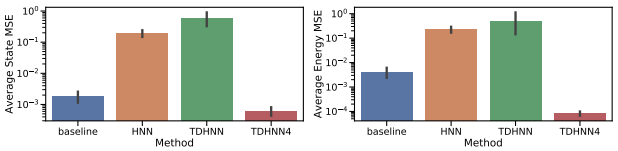
5.3 Forced Mass-Spring System

Typically, while we cannot write the Hamiltonian for a damped system, we can write one for a forced system. Here, we study two types of forced mass-spring systems. The first has the following Hamiltonian form:

$$\mathcal{H} = \frac{1}{2}kq^2 + \frac{1}{2m}p^2 - qF_0\sin(\omega t) \tag{9}$$



(a) Learnt Force and Damping terms and their contribution to dp/dt



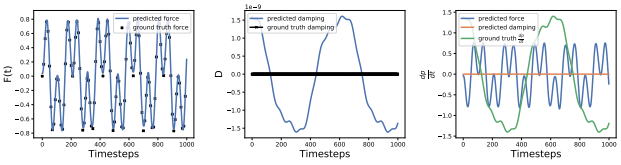
(b) Rollout state and energy MSE averaged across 25 initial test states

Figure 5: Forced mass spring setting: HNN cannot learn the underlying dynamics as it has no explicit-time dependence. TDHNN4 shows the best performance as it explicitly learns a time-dependent force.

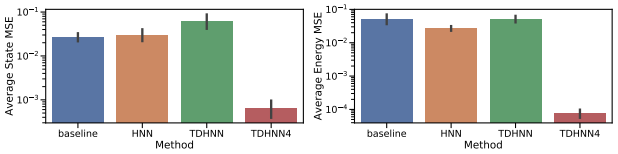
The second has the Hamiltonian:

$$\mathcal{H} = \frac{1}{2}kq^2 + \frac{1}{2m}p^2 - qF_0\sin(\omega t)\sin(2\omega t) \qquad (10)$$

Training: In both settings, we use 20 initial condi-



(a) Learnt Force and Damping terms and their contribution to dp/dt



(b) Rollout state and energy MSE averaged across 25 initial test states

Figure 6: The time dependent force here is non-trivial, but TDHNN4 shows it can recover it.

tions, where the initial state $[q_0, p_0]$ is sampled such that $q_0^2 + p_0^2 = r_0^2$ where $1 \leq r_0 \leq 4.5$. The states are rolled out to $T_{max} = 10.01$ at a $\Delta t = 0.01$. k, m and F_0 are all set to 1 without loss of generality. The forcing coefficient term ω is set to 3.

Test: At inference, we compute the rollout of 25 unseen initial conditions, and report the average state and energy rollout MSE in the figures 5 and 6.

5.4 Duffing Equation

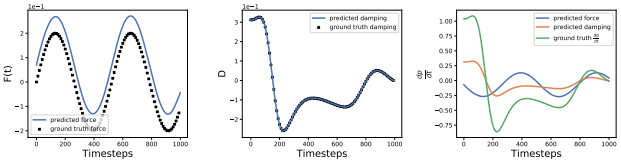
The general Duffing equation combines both the force and damping terms discussed previously and has a non-linear potential function $V(q)$. Typically the equation is written as:

$$\ddot{q} = -\delta\dot{q} - \alpha q - \beta q^3 + \gamma\sin(\omega t) \qquad (11)$$

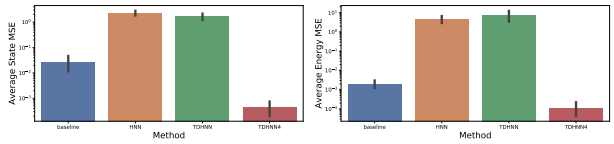
We note that the unforced and undamped regular Hamiltonian $\mathcal{H}_{reg}$ of the Duffing system is:

$$\mathcal{H}_{reg} = \frac{p^2}{2m} + \alpha\frac{q^2}{2} + \beta\frac{q^4}{4} \qquad (12)$$

This Hamiltonian therefore has a potential that can either be a single or a double well depending on the coefficients α and β. A combination of parameters $\alpha, \beta, \delta, \gamma, \omega$ will make the Duffing system either chaotic or non-chaotic. We study both regimes.



(a) Learnt Force and Damping terms and their contribution to dp/dt



(b) Rollout state and energy MSE averaged across 25 initial test states

Figure 7: Baseline NN and TDHNN4 both perform well in this setting. TDHNN4 is also able to extract the ground truth force and damping coefficient.

5.4.1 Non-Chaotic Regime

Given a set of initial parameters for the Duffing equation: $\alpha = -1, \beta = 1, \delta = 0.3, \gamma = 0.2, \omega = 1.2$ we can obtain training data for the non-chaotic regime of the Duffing equation.

Training: We uniformly sample initial conditions in $[-1, 1]^2$. We use 25 initial conditions for training, rolled out to $T_{\max} = 10.01$ and $\Delta t = 0.01$ for the non-chaotic regime.

Test: We integrate 25 unseen initial conditions at inference using the same conditions as above and evaluate both energy and state MSE (see Fig.7).

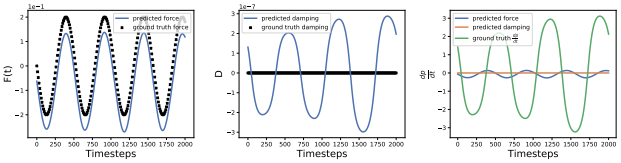
In addition to learning the force and damping terms accurately, inspecting the predicted Hamiltonian over a 2-D grid of position and momentum values shows that TDHNN4 can also accurately learn the $\mathcal{H}_{reg}$ term (see Fig.9).

5.4.2 Chaotic Regime

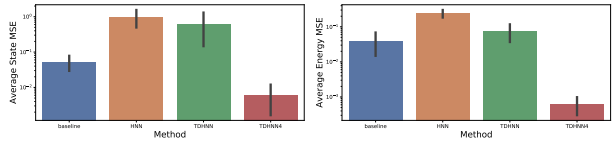
In the chaotic regime we use the following parameters: $\alpha = 1, \beta = 1, \delta = 0.1, \gamma = 0.39, \omega = 1.4$.

Training: 20 initial conditions, sampled uniformly in $[-1, 1]^2$ each rolled out for one period $T = 2\pi/\omega$ where $\Delta t = T/100$. This results in 2000 training points.

Test: We test our system by assessing whether it is visually able to recover the ground truth Poincaré section of an initial condition. The Poincaré map of a trajectory is measured by plotting the position and momentum values at regular intervals governed by the period of the forcing term. For example, a simple mass-spring system will generate a single point in phase space when measured at regular intervals. To visually asses the performance of our network through a Poincare map, we test the system on a single initial condition not in the training set rolled out to $T_{\max} = 18,000$ with the same Δt as training. In order to integrate our system to such a large $T_{\max}$ we work under the assumption that we have explicit knowledge of the period, and as such, we modulo the time variable with the period. This is necessary as the models are not explicitly trained on time steps beyond $2\pi/\omega$. Our results are visually presented in Fig.1. The outcome suggests that TDHNN4 can indeed be used to identify chaos even after being trained with only a few data points from a chaotic trajectory. We believe this is a deeply powerful result in helping us identify chaotic trajectories.



(a) Learnt Force and Damping terms and their contribution to dp/dt



(b) Rollout state and energy MSE averaged across 25 initial test states

Figure 8: Learned dynamics of a relativistic Duffing system.

5.5 Relativistic System

We investigate the Duffing equation in a special relativistic framework. The Hamiltonian under consideration is:

$$\mathcal{H} = c\sqrt{p^2 + m_0^2 c^2} + \frac{\alpha}{2}q^2 + \frac{\beta}{4}q^4 - q\gamma\sin(\omega t) \quad (13)$$

where c, the speed of light is typically set to 1. For simplicity, we also set the rest mass $m_0 = 1$ though our framework naturally accounts for other values.

Training: We train on 25 initial conditions, sampled in $[0, 2]^2$ uniformly. $T_{\max} = 20.01$ and $\Delta t = 0.01$. We set parameters such that $\alpha = 1, \beta = 1, \delta = 0, \gamma = 0.2, \omega = 1.2$.

Test: Using the same parameters as training, we rollout 25 unseen initial conditions and compute their state/energy MSE.

6 Discussion

We have shown that TDHNN4 can not only outperform other approaches in learning complex physical systems but can also help us recover the underlying force and damping terms of non-autonomous systems. One challenge in achieving this result is fine-tuning the λ_F and λ_D regularisation coefficients for the force and damping. In the Duffing setting where we have both terms, it is possible to learn a shifted force i.e. $F = F_0 + \epsilon$. The reason this is possible is because the state-vectors $\mathbf{q}, \mathbf{p}$ do not provide enough information to simultaneously identify both the Hamiltonian and the force. As such it is possible that our hyperparameter selection routine will pick coefficients that maintain a force or damping term in settings where they might not exist. However, we do find that in general a reasonable force and damping term are learnt so as to reveal the underlying dynamics.

In many settings we find that the classic Hamiltonian Neural Network approach or its time-dependent variant simply do not perform as well as TDHNN4 or baseline. We believe this is predominantly due to the fact that HNN and TDHNN attempt to enforce Hamilton's equations in settings where they cannot be satisfied.

While it may be argued that the method is constrained to learning and prediction within the training time horizon, we believe our method is still versatile at informing us of a periodic forcing (since we can inspect the force over time). This in turn can readily be used to renormalize the time variable at periodic intervals to integrate the system beyond the training time.

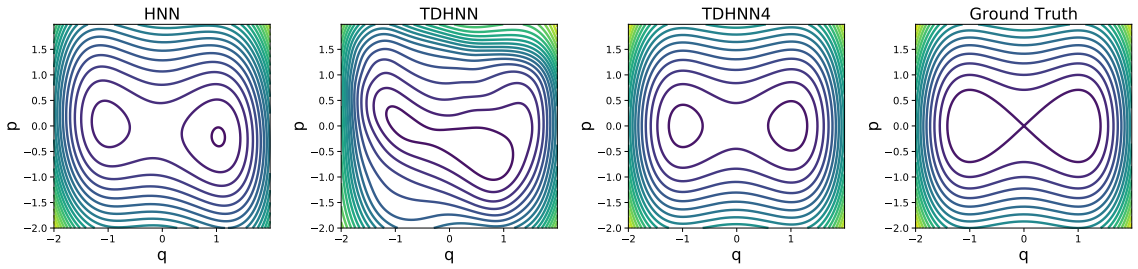


Figure 9: Learnt $\mathcal{H}_{reg}$ components across methods in the non-chaotic Duffing setting. HNN and TDHNN learn distorted Hamiltonians that strongly depend on the input time-variable.

7 Conclusion

We have shown that learning the dynamics of time-dependent non-autonomous systems can be achieved with TDHNN4, a versatile neural network embedded with a Port-Hamiltonian designed to recover the underlying Hamiltonian, damping and forcing terms. We show that such an approach, unique in its own right, outperforms naive extensions of existing methods. We believe this work forms a strong basis for further advances in learning complex systems including, but not limited to, chemical bond forces and controlled dynamics.

References

[1] P. W. Battaglia, R. Pascanu, M. Lai, D. Rezende, and K. Kavukcuoglu. Interaction Networks for Learning about Objects, Relations and Physics. *arXiv:1612.00222 [cs]*, Dec. 2016. arXiv: 1612.00222.

[2] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural Ordinary Differential Equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 6571–6583. Curran Associates, Inc., 2018.

[3] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho. Lagrangian Neural Networks. *arXiv:2003.04630 [physics, stat]*, Mar. 2020. arXiv: 2003.04630.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*, May 2019. arXiv: 1810.04805.

[5] M. Finzi, S. Stanton, P. Izmailov, and A. G. Wilson. Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data. *arXiv:2002.12880 [cs, stat]*, May 2020. arXiv: 2002.12880.

[6] S. Greydanus, M. Dzamba, and J. Yosinski. Hamiltonian Neural Networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d. Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 15379–15389. Curran Associates, Inc., 2019.

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. *arXiv:1703.06870 [cs]*, Jan. 2018. arXiv: 1703.06870.

[8] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, Jan. 1989.

[9] M. Lutter, C. Ritter, and J. Peters. Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning. *arXiv:1907.04490 [cs, eess, stat]*, July 2019. arXiv: 1907.04490.

[10] M. Mattheakis, D. Sondak, A. S. Dogra, and P. Protopapas. Hamiltonian Neural Networks for solving differential equations. *arXiv:2001.11107 [physics]*, Feb. 2020. arXiv: 2001.11107.

[11] A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum, and R. Komanduri. Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks. *The Journal of Chemical Physics*, 130(13):134101, Apr. 2009.

[12] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations. *arXiv:1711.10561 [cs, math, stat]*, Nov. 2017. arXiv: 1711.10561.

[13] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, Feb. 2019.

[14] A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, and P. Battaglia. Hamiltonian Graph Networks with ODE Integrators. *arXiv:1909.12790 [physics]*, Sept. 2019. arXiv: 1909.12790.

[15] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. W. Battaglia. Learning to Simulate Complex Physics with Graph Networks. *arXiv:2002.09405 [physics, stat]*, Feb. 2020. arXiv: 2002.09405.

[16] A. Sanchez-Gonzalez, N. Heess, J. T. Springenberg, J. Merel, M. Riedmiller, R. Hadsell, and P. Battaglia. Graph networks as learnable physics engines for inference and control. *arXiv:1806.01242 [cs, stat]*, June 2018. arXiv: 1806.01242.

[17] M. Toussaint, K. Allen, K. Smith, and J. Tenenbaum. Differentiable Physics and Stable Modes for Tool-Use and Manipulation Planning. In *Robotics: Science and Systems XIV*. Robotics: Science and Systems Foundation, June 2018.

[18] K. Yao, J. E. Herr, D. Toth, R. Mckintyre, and J. Parkhill. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chemical Science*, 9(8):2261–2269, 2018.

[19] Y. D. Zhong, B. Dey, and A. Chakraborty. Dissipative SymODEN: Encoding Hamiltonian Dynamics with Dissipation and Control into Deep Learning. *arXiv:2002.08860 [cs, eess, stat]*, Apr. 2020. arXiv: 2002.08860.

[20] A. Zhu, P. Jin, and Y. Tang. Deep Hamiltonian networks based on symplectic integrators. *arXiv:2004.13830 [cs, math]*, Apr. 2020. arXiv: 2004.13830.

Appendix

A.1

Why can we not including damping into the Hamiltonian? Let us take the following damped system:

$$\ddot{\mathbf{q}} = -\mathbf{q} - \delta\dot{\mathbf{q}} \tag{14}$$

where we know $\mathbf{p} = m\dot{\mathbf{q}}$ which implies $\dot{\mathbf{q}} = m^{-1}\mathbf{p}$.

Then, the integral of the right hand side with respect to q will give us:

$$\frac{\mathbf{q}^2}{2} + \delta\mathbf{q}\dot{\mathbf{q}} \tag{15}$$

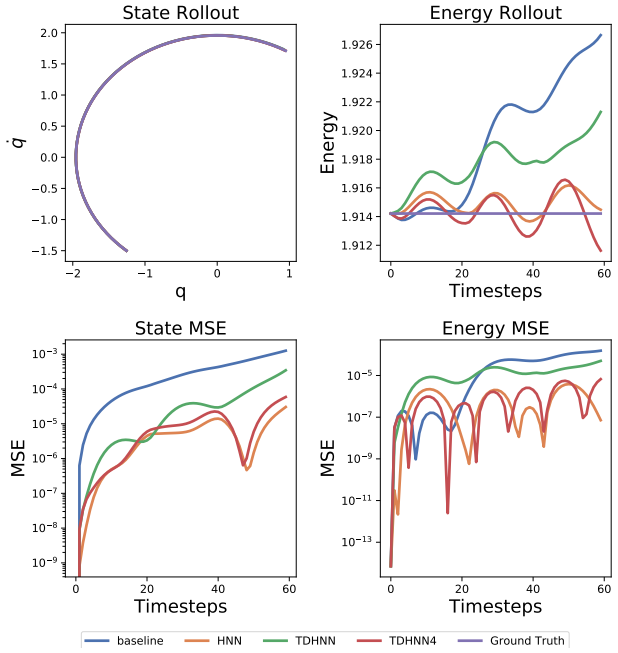
The equation above looks like a modified potential function which can be combined with a kinetic energy term to give a Hamiltonian s.t.:

$$\mathcal{H} = \frac{\mathbf{p}^2}{2m} + \frac{\mathbf{q}^2}{2} + \delta\mathbf{q}\dot{\mathbf{q}} \tag{16}$$

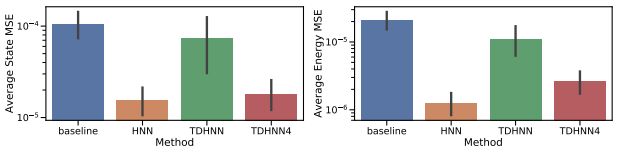
However, although we can recover the differential equation for $\ddot{\mathbf{q}}$ by $-\frac{\partial\mathcal{H}}{\mathrm{d}\mathbf{q}} = -\mathbf{q} - \delta\dot{\mathbf{q}} = \ddot{\mathbf{q}}$, we violate the rule that $\dot{\mathbf{q}} = m^{-1}\mathbf{p}$ since $\frac{\partial\mathcal{H}}{\mathrm{d}\mathbf{p}} = \dot{\mathbf{q}} + \delta\mathbf{q} \neq \dot{\mathbf{q}}$.

Indeed an alternative formulation that uses exponentiated time can allow us to represent a Hamiltonian for the damped system but this is not within the scope of our study as the formalism does not explain how to include external forcing.

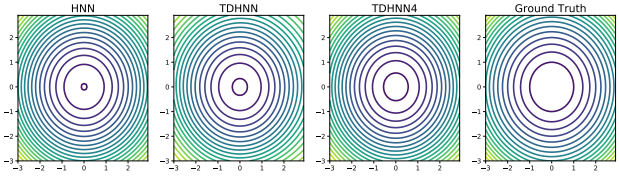
B



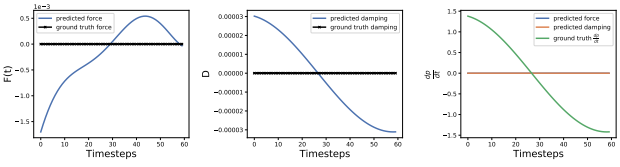
(a) State and energy rollout of an initial condition from the test set



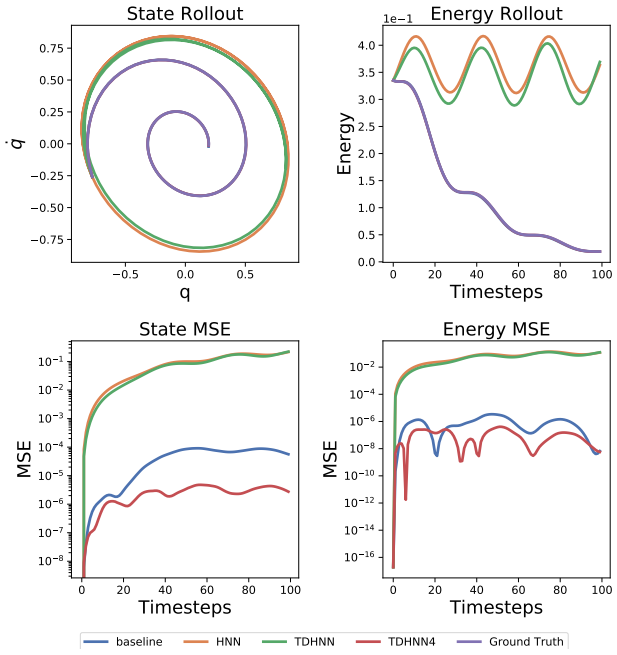
(b) The average state and energy MSE across 25 test points



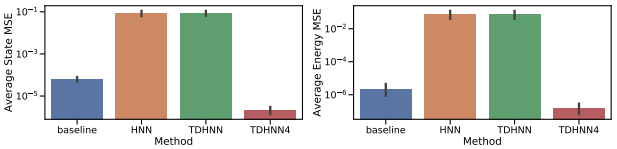
(c) The learnt Hamiltonian across methods



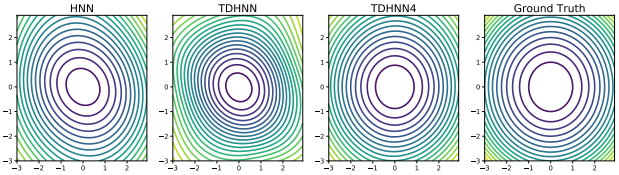
(d) The learnt force and damping of TDHNN4



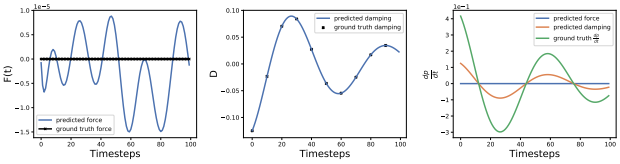
(a) State and energy rollout of an initial condition from the test set



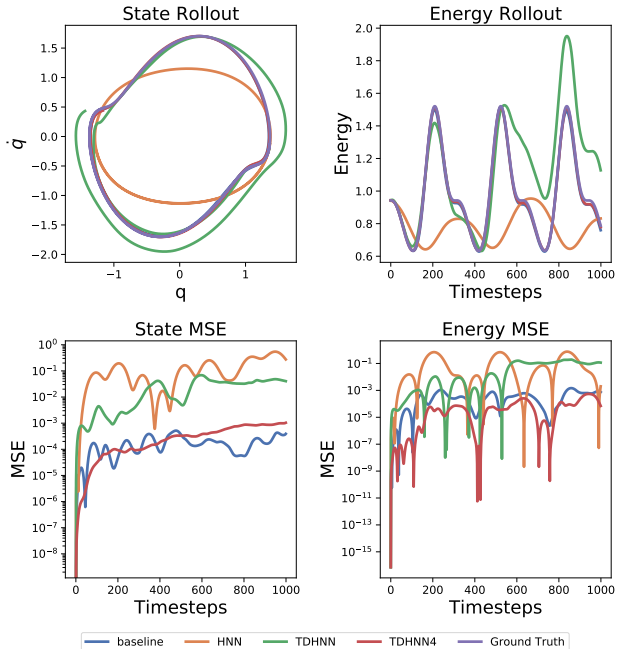
(b) The average state and energy MSE across 25 test points



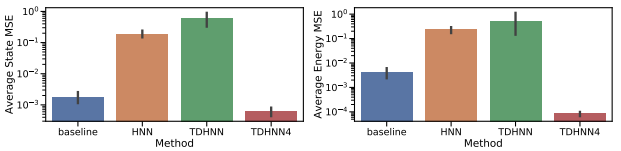
(c) The learnt Hamiltonian across methods



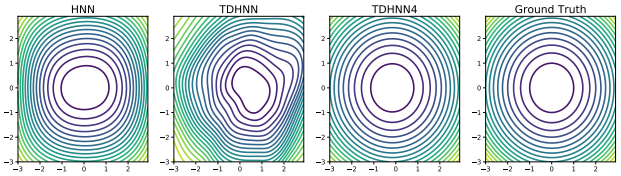
(d) The learnt force and damping of TDHNN4



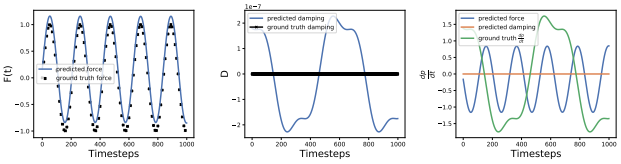
(a) State and energy rollout of an initial condition from the test set



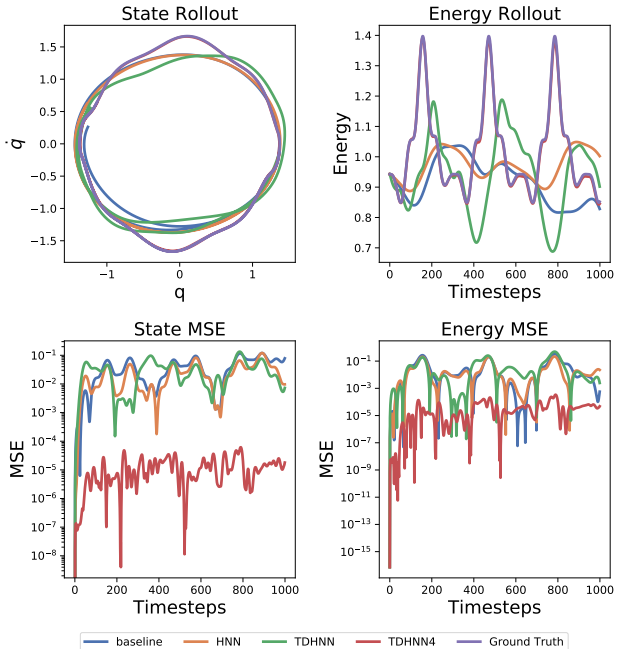
(b) The average state and energy MSE across 25 test points



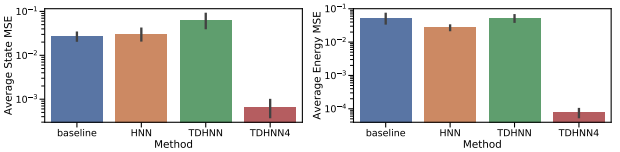
(c) The learnt Hamiltonian across methods



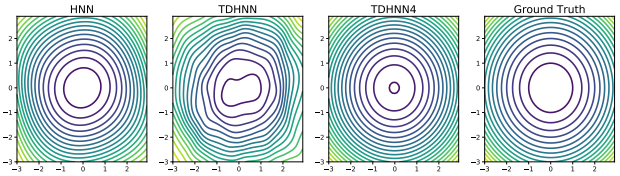
(d) The learnt force and damping of TDHNN4



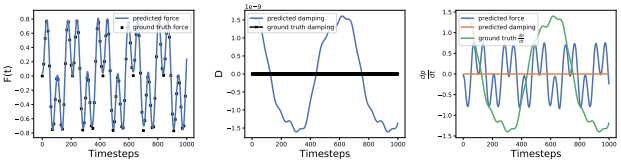
(a) State and energy rollout of an initial condition from the test set



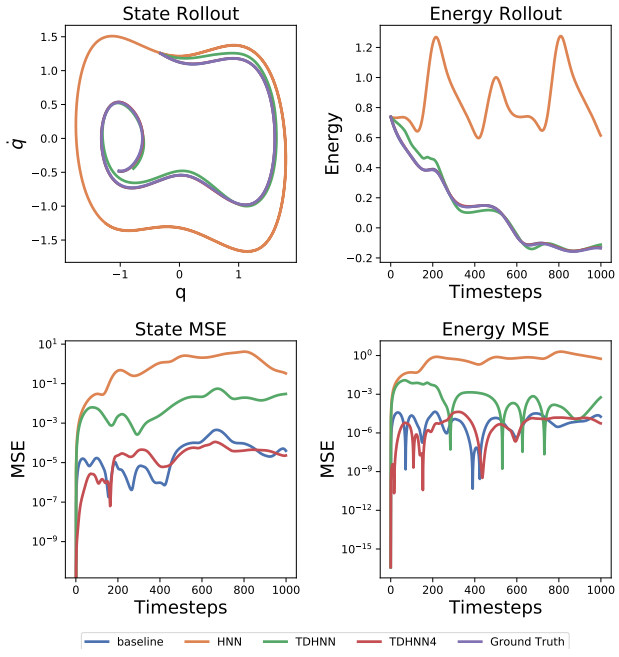
(b) The average state and energy MSE across 25 test points



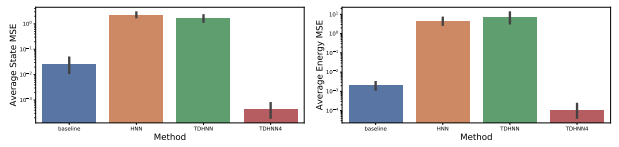
(c) The learnt Hamiltonian across methods



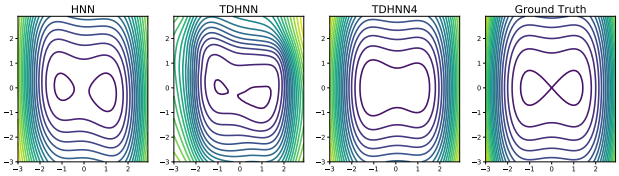
(d) The learnt force and damping of TDHNN4



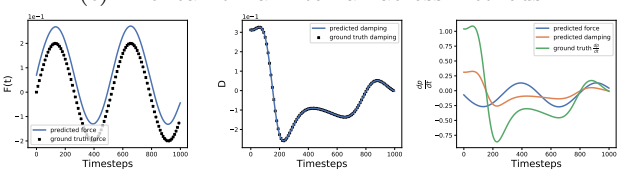
(a) State and energy rollout of an initial condition from the test set



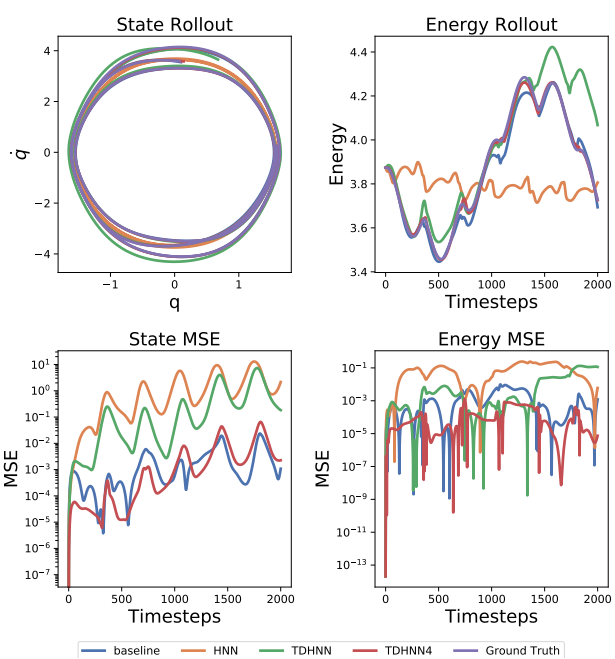
(b) The average state and energy MSE across 25 test points



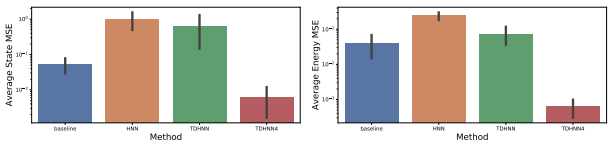
(c) The learnt Hamiltonian across methods



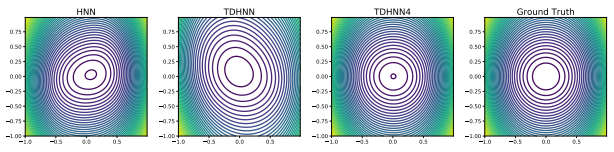
(d) The learnt force and damping of TDHNN4



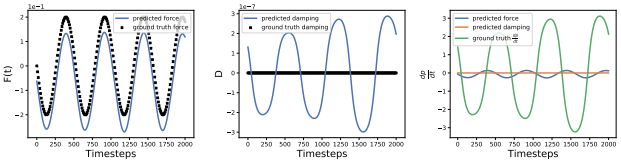
(a) State and energy rollout of an initial condition from the test set



(b) The average state and energy MSE across 25 test points



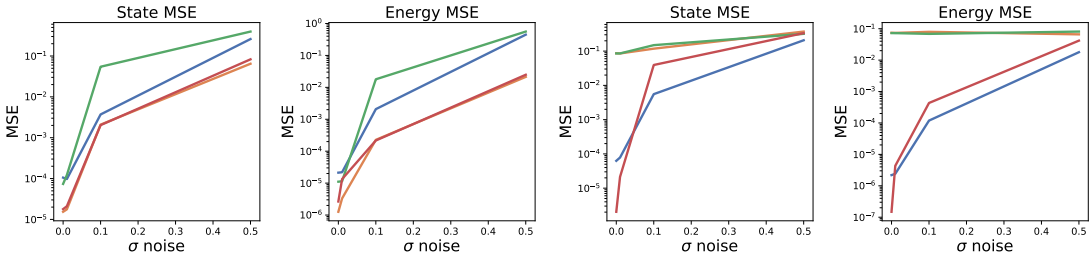
(c) The learnt Hamiltonian across methods



(d) The learnt force and damping of TDHNN4

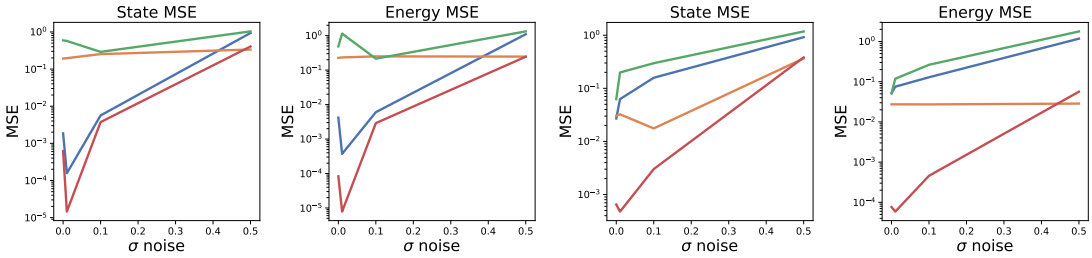
C

During the training of HNN, the authors add gaussian noise with a standard deviation $\sigma = 0.1$ to the input state vector data. The reason this is done is to ensure the model is robustly trained. We run a set of experiments to test the robustness to this 'noisy' input.



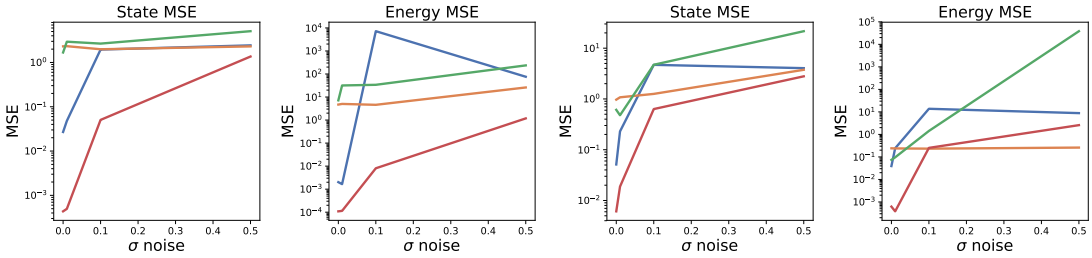
(a) mass spring

(b) damped



(c) forced mass spring (I)

(d) forced mass spring (II)



(e) duffing

(f) relativity

D

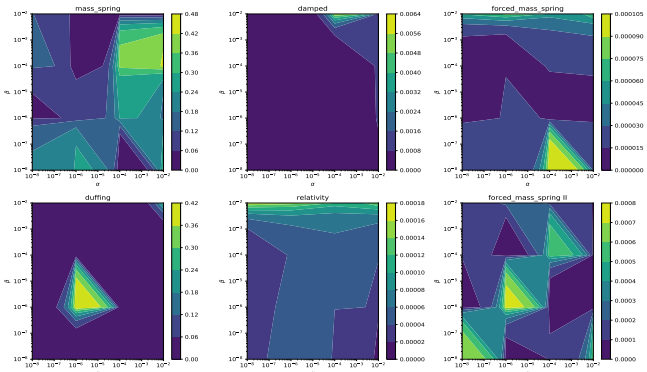


Figure 17: Hyperparameter Optimization for TDHNN4. We plot, for each system, the validation loss as a function of the α and β parameters from the loss in eqn. 5