TDHNN: Time-Dependent Hamiltonian Neural Networks

Shaan A. Desai[1,2]
Machine Learning Research Group
University of Oxford [1]

David Sondak[2]
Institute of Applied Computational Science
Harvard University [2]

Marios Mattheakis[2]

Abstract

Deep neural-networks embedded with physically-informed priors demonstrate remarkable performance over vanilla neural-networks in accurately learning and predicting non-linear dynamical systems. Recently, it was shown that Hamiltonian Neural Networks (HNNs) - neural networks designed to conserve energy and exploit the Hamiltonian formalism, show strong and consistent performance in learning autonomous systems that depend implicitly on time. Here, we extend this framework to include an explicit time-dependence that generalizes the learning to non-autonomous dynamical systems. We achieve this generalisation 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 while maintaining comparable performace to HNNs in settings with no explicit time-dependence.

1 Introduction

Neural networks (NNs), as universal function approximators [6], have shown resounding success across a host of domains. However, their performance in learning physical systems has often been limited. New research in *scientific machine learning* - a branch 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 prior theoretical information embedded in networks,

such as Hamiltonian mechanics [5], Lagrangians [3, 7], Ordinary Differential Equations (ODEs) [2], physics equations [8] and Graph Neural Networks [1, 10] can significantly improve learning in complex physical domains. Furthermore, it has been shown that such systems can be adapted to learn trajectories of controlled dynamics [7, 13]. Despite this extensive research, there is still no method that accounts for explicit-time dependence, a crucial component in forced dynamical systems. Additionally, many of the existing approaches present no direct way to account for dissipation. Inspired by [13], we show that a Port-Hamiltonian - a method designed to account for forcing and dissipation, can be used to 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. Our network consistently outperforms other approaches while actively learning both the driving force and damping coefficient. We also show that our network can visually recover the Poincaré section of a duffing equation in the chaotic regime Fig..

2 Background

2.1 Hamiltonian Neural Networks

Recently, [5] demonstrated that dynamic predictions through time can be improved using Hamiltonian Neural Networks (HNNs) which endow models with a Hamiltonian constraint. The Hamiltonian is an important representation of a dynamical system because it is one of two well-known approaches that generalizes classical mechanics. The Hamiltonian $\mathcal{H}$ is a scalar function of position $\mathbf{q} = (q_1, q_2, ..., q_M)$ and momentum $\mathbf{p} = (p_1, p_2, ..., p_M)$. It is a powerful representation because it allows us to obtain the time derivatives of the inputs $(\dot{\mathbf{q}}, \dot{\mathbf{p}})$ by simply differentiating the Hamiltonian with respect to its inputs (see Eqn.1.)

$$\frac{d\mathbf{q}}{dt} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} \qquad (1)$$

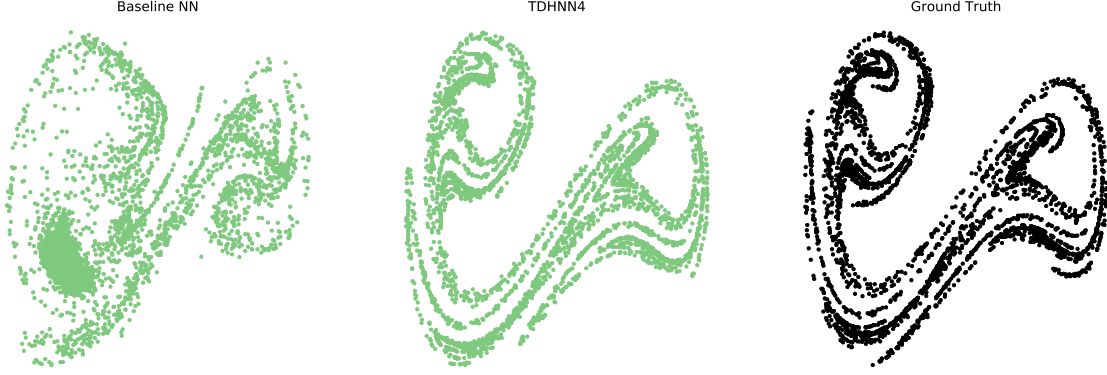As a consequence, it is noted in [5] that by parametrizing the Hamiltonian with a neural network e.g.

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.

$H_\theta(\mathbf{q}, \mathbf{p})$ where θ represents a deep neural network, one can easily obtain the system's dynamics by differentiating (via autograd) the Hamiltonian with its inputs. This information allows us to build two 1st-order differential equations which can be used to update the state space, $(\mathbf{q}, \mathbf{p})$. Equation 2 shows this integral, in which we define the symplectic gradient $\mathbf{S} = \left[\frac{\partial \mathcal{H}}{\partial \mathbf{p}}, -\frac{\partial \mathcal{H}}{\partial \mathbf{q}}\right]$:

$$(\mathbf{q}, \mathbf{p})_{t+1} = (\mathbf{q}, \mathbf{p})_t + \int_t^{t+1} \mathbf{S}(\mathbf{q}, \mathbf{p}) \mathrm{d}t \qquad (2)$$

However, this is not the only benefit in learning a Hamiltonian. Another key attribute of the Hamiltonian is that the vector field $\mathbf{S}$ is a symplectic gradient meaning $\mathcal{H}$ remains constant as long as state vectors are integrated along $\mathbf{S}$. This result links the Hamiltonian with the total energy of the system such that $\mathcal{H}(\mathbf{q}, \mathbf{p}) = E_{tot}$ for many physical systems. Therefore, the Hamiltonian is a powerful inductive bias that can be utilised to evolve a physical state while maintaining energy conservation.

Although this formalism is compact and powerful, it does not readily generalize to damped or forced system. As such, we refer to port-Hamiltonian systems.

2.2 Port-Hamiltonians

Port-Hamiltonians are a formalism that allow us to correct Hamilton's equations to incorporate damping and forcing terms. The first method to embed the port-Hamiltonian in a neural network [13] shows how to represent a general form for such a system. In [13] the Port-Hamiltonian is represented as:

$$\begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{p}} \end{bmatrix} = \left(\begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{bmatrix} + \mathbf{D}_{\theta_2}(\mathbf{q}) \right) \begin{bmatrix} \frac{\partial \mathcal{H}_{\theta_1}}{\partial \mathbf{q}} \\ \frac{\partial \mathcal{H}_{\theta_1}}{\partial \mathbf{p}} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{g}(\mathbf{q}) \end{bmatrix} u \quad (3)$$

where $\mathbf{D}_{\theta_2}(\mathbf{q})$ is the damping term and $g(q)u$ is the external constant control term.

This general formalism allows us to collapse to classical Hamiltonians by simply setting $\mathbf{D}$ and u to zero. Note here that the authors of [13] assume the vector $[\mathbf{q}, \mathbf{p}, u]$ is provided as input to the system. The matrix $\mathbf{D}_{\theta_2}(\mathbf{q})$ is assumed to be semi-positive definite and $g(q)$ is typically a non-linear scaling of q to account for polynomial scaling of forces, for example, as would be found in the Hill-Schroedinger equation.

3 Related Work

Numerous recent methods show how to learn dynamics via physically informed priors.

3.1 Functional Priors

Functional priors typically embed the full functional form of an equation, into the neural network. Physics-informed Neural Networks (PINNs) [8, 9] and Hamiltonian Nets (Marios) look at directly embedding the equations into the loss function. PINNs furthermore exploit autograd to compute partial-derivatives and show that such an approach is more powerful than symbolic regression.

3.2 Integrator Priors

Many methods seek to use a NN to parametrize the derivatives $\dot{S}$ of a state S and then use this learned parametrization in a scientific integrator such as a Runge-Kutta method. The authors of NeuralODE [2] show that the embedding of the integrator in the learning process induces a continuous depth neural network for large time-step predictions. As such, the number of

Shaan A. Desai[1,2], **David Sondak**[2], **Marios Mattheakis**[2]

gradient computations scales exponentially. The paper re-introduces the adjoint method to ensure linear scaling of the gradient computations. Other work such as [14] theoretically proves that the embedded integrator in Hamiltonian Neural Networks significantly benefits from being symplectic.

3.3 Graph based methods

In [1], the authors show 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 [12, 11, 3].

3.4 Lagrangian/Hamiltonian Methods

While [3] and [5] propose the explicit learning of Hamiltonian and Lagrangian functions, [7] 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.

3.5 Latest Advances

CHNN [4], one of the most recent advances published as a spotlight in NeurIPS 2020, 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.

Our method investigates a more general Hamiltonian formalism, and similar to [4], explicitly encodes the holonomic constraints into the network.

4 Method

The architecture for our method can be seen in Fig.2. We modify the Port-Hamiltonian in Eqn.3 in two ways. First, we know that many physical systems have a damping matrix that only consists of a non-zero, state-independent term in the lower right quadrant, so we replace $\mathbf{D}(\mathbf{q})$ with a single weight parameter independent of $\mathbf{q}$ in the lower right half of the matrix. Secondly, we set $g(q) = 1$ and replace u with $F_{\theta_3}(t)$ as many physical systems of interest do not typically scale the force by non-linear q terms as can be found in the examples in [13]. The resulting equation is:

$$\begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{p}} \end{bmatrix} = \left(\begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{D}_{\theta_2} \end{bmatrix} \right) \begin{bmatrix} \frac{\partial \mathcal{H}_{\theta_1}}{\partial \mathbf{q}} \\ \frac{\partial \mathcal{H}_{\theta_1}}{\partial \mathbf{p}} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{F}_{\theta_3}(\mathbf{t}) \end{bmatrix} \quad (4)$$

Although we have made some simplifying assumptions, our method remains versatile at capturing the underlying form of many complex physical phenomena.
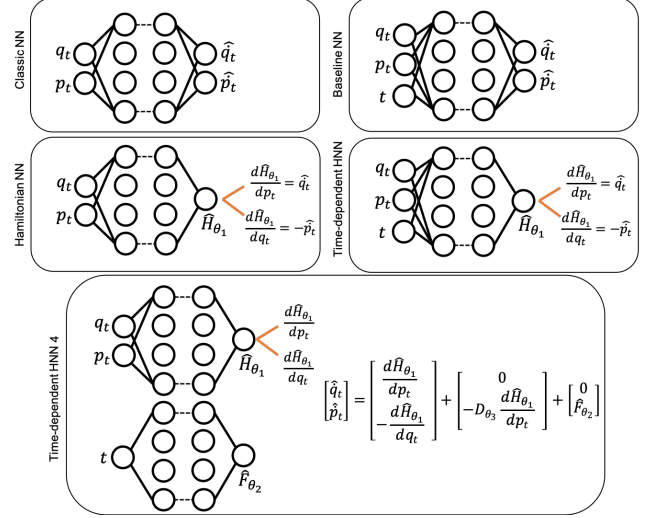

Figure 2: Architectures used to learn dynamics in this paper. The naive extension of classic NN and Hamiltonian NN (top left) is to incorporate time as an additional input variable (top right). Our innovation, which exploits Port-Hamiltonians, explicitly learns the force F_{θ_2} as well as the damping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

We define the loss function for our architecture as follows:

$$\mathcal{L} = \left\| \frac{\partial \mathcal{H}_{\theta_1}}{\partial \mathbf{q}} + \frac{\partial \mathbf{p}}{\partial t} \right\| + \left\| \frac{\partial \mathcal{H}_{\theta_1}}{\partial \mathbf{p}} - \frac{\partial \mathbf{q}}{\partial t} \right\| + \alpha_{reg}|\mathbf{F}_{\theta_3}(\mathbf{t})| + \beta_{reg}|\mathbf{D}_{\theta_2}| \quad (5)$$

To learn the dynamics, we feed in a state-vector $\mathbf{S}_t = [\mathbf{q}, \mathbf{p}, \mathbf{t}]$ to our model. The first neural-network $\mathcal{H}_{\theta_1}$ consists of 3 hidden layers designed to predict $\mathcal{H}$ from $[\mathbf{q}, \mathbf{p}]$ data. The second component $\mathbf{D}_{\theta_2}$ consists of a single weight parameter designed to learn the damping coefficient $\mathbf{D}$ and the third neural-network consists of 2 hidden layers designed to predict $\mathbf{F}(\mathbf{t})$. We use an L2-norm penalty for the predicted state-vectors and an L1-norm for force and damping to encourage sparsity. We do this because we would like our networks to identify classical autonomous systems (which may not have force or time) as well as non-autonomous systems. For our experiments we use 200 hidden layers and find that most activations, including tanh, sin and cos yield comparable results. To benchmark our method, as [5] do, we use a modified classic NN that takes in an additional time vector and predicts $[\dot{q}, \dot{p}]$. We also take the

straightforward extension of HNN to include time as a variable input. At inference, we use a Runge-Kutta 4th order integrator to rollout the initial conditions.

5 Results

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

5.1 Simple Mass Spring

The simple mass spring system is a well known system from classical physics that obeys Hooke's Law. The Hamiltonian for a simple mass spring system is written as:

$$\mathcal{H} = \frac{1}{2}k\mathbf{q}^2 + \frac{1}{2m}\mathbf{p}^2 \qquad (6)$$

For simplicity, we set k and m to 1 for our experiments. We sample 25 initial training conditions such that r in $q^2 + p^2 = r^2$ is between 0.5 and 1.5 as is done in [5]. The trajectories are sampled with $\delta t = 0.05$ and $T_{max} = 3.05$. The test error in fig.3 is measured as a Mean Squared Error (MSE) be across for 50 trajectories. We investigate this simple system to show that learning a separate, regularised forcing term results in much better state and energy rollout in comparison to TDHNN and baseline NN. The main reason for this is that while TDHNN and baseline NN will learn dynamics with time-steps within the training regime, they typically cannot generalize to unseen timesteps. Learning a separate forcing term and regularizing it keeps the time component independent of the Hamiltonian and therefore allows us to nearly match the performance of HNN.

5.2 Damped Mass Spring

Damping is generally not considered part of Hamiltonian systems. For example, let us take the following damped system:

$$\ddot{\mathbf{q}} = -\mathbf{q} - \delta\dot{\mathbf{q}} \qquad (7)$$

and 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:

$$\frac{\mathbf{q}^2}{2} + \delta\mathbf{q}\dot{\mathbf{q}} \qquad (8)$$

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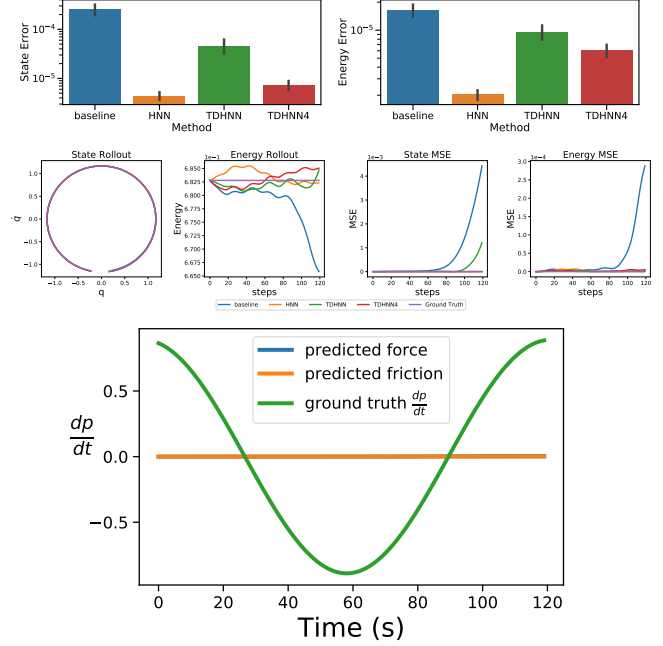

Figure 3: Simple Mass Spring system has no explicit time dependence. We see that TDHNN4 can almost recover the dynamics of HNN which doesn't depend on time. The result is achieved by regularising the force and friction terms as can be seen in the bottom figure. Baseline and TDHNN are unable to achieve the same state-error as their time-dependence

$$\mathcal{H} = \frac{\mathbf{p}^2}{2m} + \frac{\mathbf{q}^2}{2} + \delta\mathbf{q}\dot{\mathbf{q}} \qquad (9)$$

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 $\frac{\partial\mathcal{H}}{\mathrm{d}\mathbf{p}} = \dot{\mathbf{q}} + \delta\mathbf{q} \neq \dot{\mathbf{q}}$.

However, as we saw with Port-Hamiltonians, there is a straightforward way to account for the damping term without explicitly defining a full Hamiltonian. By combining a regular Hamiltonian with an additional term that learns the damping, we can disentangle the two terms for learning.

For training, we have 20 initial conditions, uniformly sampled in $[-1, 1]^2$. Each trajectory is evolved until $T_{max} = 30.1$ with a $\delta t = 0.1$. At inference, we compute the rollout of 25 trajectories, and report the average state and energy rollout MSE in fig.4.

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.

Shaan A. Desai[1,2], David Sondak[2], Marios Mattheakis[2]
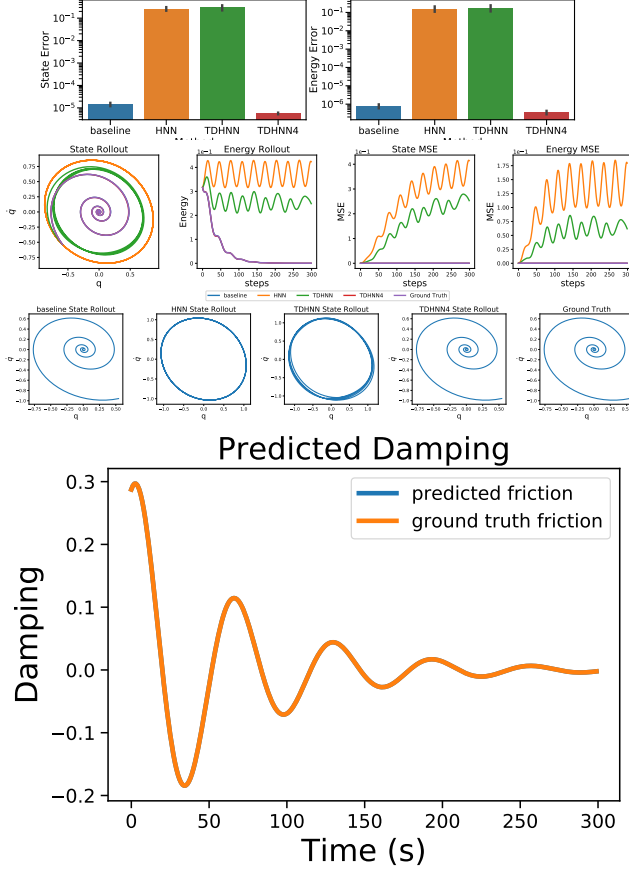
Predicted Damping

Figure 4: 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.

5.3 Forced Mass Spring

Typically, while we cannot write the Hamiltonian for a damped system, we can write one for a forced system. Here, we study 2 types of forced mass-spring systems. The first has the following Hamiltonian form:

$$\mathcal{H} = \frac{1}{2}k\mathbf{q}^2 + \frac{1}{2m}\mathbf{p}^2 - \mathbf{q}\mathbf{F_0}sin(\omega\mathbf{t}) \qquad (10)$$

The second has this Hamiltonian form:

$$\mathcal{H} = \frac{1}{2}k\mathbf{q}^2 + \frac{1}{2m}\mathbf{p}^2 - \mathbf{q}\mathbf{F_0}sin(\omega\mathbf{t})sin(2\omega\mathbf{t}) \qquad (11)$$

In both settings, we use 20 initial conditions, where the state is sampled from a radii between 0.5 and 1.5. The states are rolled out to $T_{max} = 10.01$ at a $\delta t = 0.01$. At inference, we compute the rollout of 25 trajectories, and report the average state and energy rollout MSE in the figures.
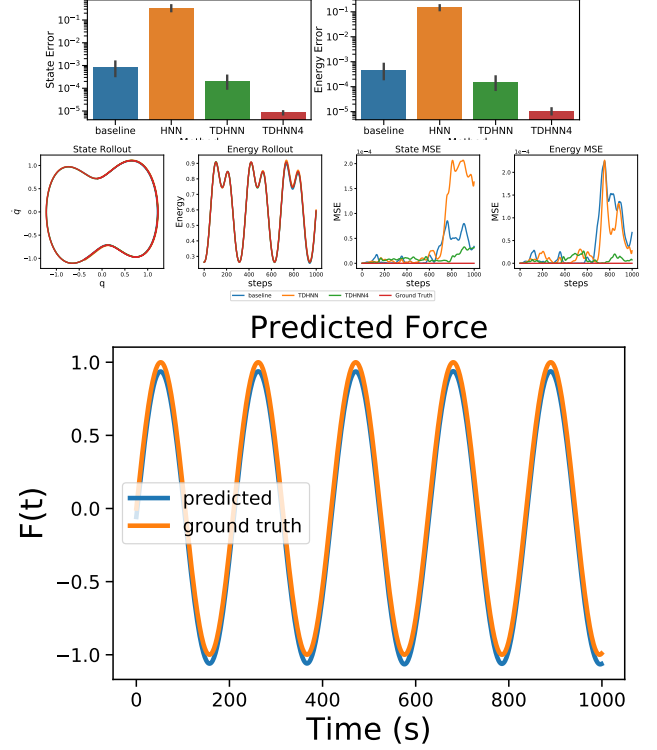
Predicted Force

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.

5.4 Duffing

The general duffing equation combines both the force and damping terms. Typically the equation is written as:

$$\ddot{\mathbf{q}} = -\delta\dot{\mathbf{q}} - \alpha\mathbf{q} - \beta\mathbf{q}^3 + \gamma\sin(\omega\mathbf{t}) \qquad (12)$$

One can see that the unforced and undamped regular Hamiltonian of the duffing system would be:

$$\mathcal{H}_{reg} = \frac{\mathbf{p}^2}{2m} + \alpha\frac{\mathbf{q}^2}{2} + \beta\frac{\mathbf{q}^4}{4} \qquad (13)$$

This Hamiltonian therefore has a potential that would either be a single or a double well. A combination of parameters $\alpha, \beta, \delta, \gamma, \omega$ will make the duffing system either chaotic or non-chaotic. We study both regimes.

5.4.1 Non-Chaotic

Given a set of initial parameters where $\alpha = -1, \beta = 1, \delta = 0.3, \gamma = 0.2, \omega = 1.2$ we can obtain training data for the non-chaotic regime of the duffing equation. We uniformly sample initial conditions in $[-1, 1]^2$. We use
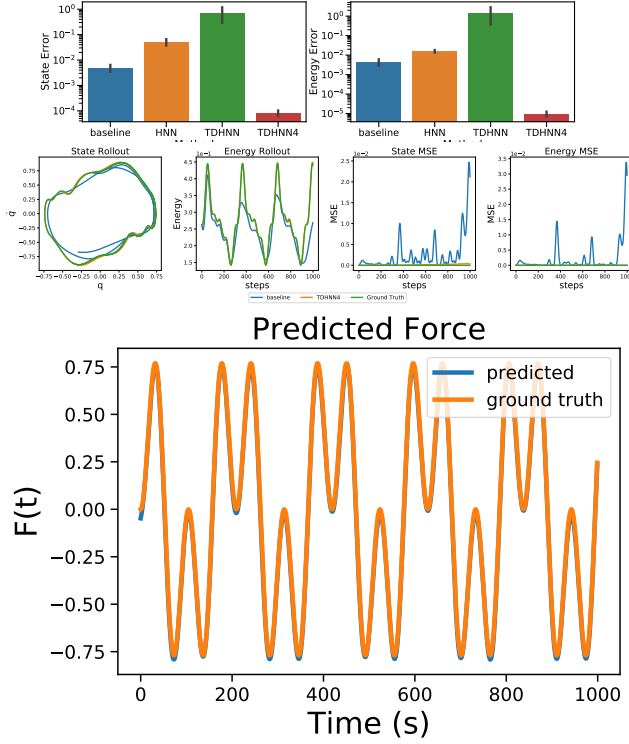
Predicted Force

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

25 initial conditions for training, rolled out to $T_{max} = 10.01$ and $\delta t = 0.01$ for the non-chaotice.

5.4.2 Chaotic

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

Training Data: 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.

Testing: we test the system on a single initial condition, rolled out to $T_{max} = 18,000$ with the same δt as training. One additional change we make at inference is a modification to the time variable which is fed in. 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 T. Using the unchanged time, our model predictions rapidly diverge from the ground truth. Our results are visually presented in Fig..

Predicted Force
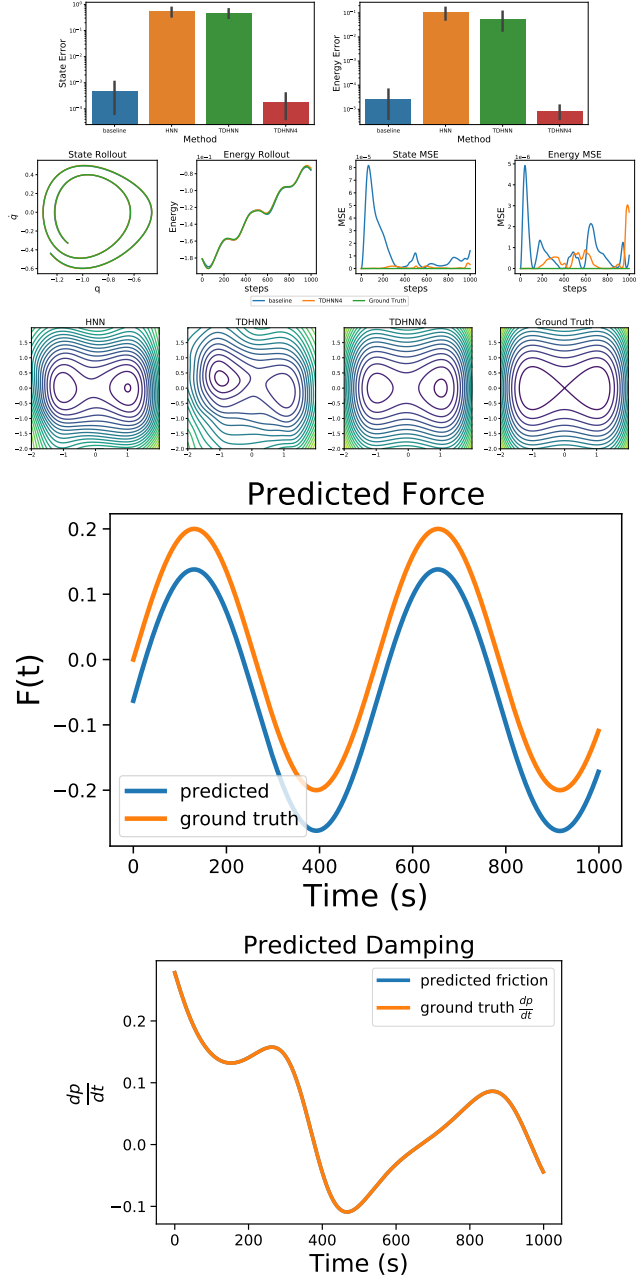

Predicted Damping

Figure 7: baseline NN and TDHNN4 both perform really well in this setting. With TDHNN4 we are also able to extract the ground truth force and damping coefficient. The predicted Hamiltonian also visually appears closer to the ground truth in comparison to HNN or TDHNN.

5.5 Relativity

We investigate the duffing equation with a relativistic kinetic energy term. The general Hamiltonian form is:

Shaan A. Desai[1,2], David Sondak[2], Marios Mattheakis[2]

$$\mathcal{H} = c\sqrt{\mathbf{p}^2 + m^2c^2} + \frac{\alpha}{2}\mathbf{q}^2 + \frac{\beta}{4}\mathbf{q}^4 - \mathbf{q}\mathbf{F_0}sin(\omega\mathbf{t}) \quad (14)$$

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

6   Discussion

7   Conclusion

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] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

Table 1: Test Rollout MSE

Method	Ideal Mass Spring		Damped Mass Spring		Forced Mass Spring (1)		Forced Mass Spring (2)		Duffing Non-Chaotic	
	State	Energy	State	Energy	State	Energy	State	Energy	State	Energy
Baseline NN										
HNN										
TDHNN										
TDHNN4										

[11] 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.

[12] 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.

[13] 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.

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