
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Haitsiukevich, Katsiaryna; Ilin, Alexander

Learning Trajectories of Hamiltonian Systems with Neural Networks

Published in:

Artificial Neural Networks and Machine Learning – ICANN 2022

DOI:

[10.1007/978-3-031-15919-0_47](https://doi.org/10.1007/978-3-031-15919-0_47)

Published: 07/09/2022

Document Version

Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:

Haitsiukevich, K., & Ilin, A. (2022). Learning Trajectories of Hamiltonian Systems with Neural Networks. In E. Pimenidis, M. Aydin, P. Angelov, C. Jayne, & A. Papaleonidas (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2022* (pp. 562–573). (Lecture Notes in Computer Science; Vol. 13529). Springer. https://doi.org/10.1007/978-3-031-15919-0_47

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Learning Trajectories of Hamiltonian Systems with Neural Networks^{*}

Katsiaryna Haitsiukevich[✉][0000–0001–8116–1292] and
Alexander Ilin^[0000–0001–6419–3006]

Aalto University, Espoo, Finland
`{firstname.lastname}@aalto.fi`

Abstract. Modeling of conservative systems with neural networks is an area of active research. A popular approach is to use Hamiltonian neural networks (HNNs) which rely on the assumptions that a conservative system is described with Hamilton’s equations of motion. Many recent works focus on improving the integration schemes used when training HNNs. In this work, we propose to enhance HNNs with an estimation of a continuous-time trajectory of the modeled system using an additional neural network, called a deep hidden physics model in the literature. We demonstrate that the proposed integration scheme works well for HNNs, especially with low sampling rates, noisy and irregular observations.

Keywords: Conservative systems · Deep Hidden Physics Models · Dynamical Systems · Hamiltonian Neural Networks · Physics-Informed Neural Networks.

1 Introduction

Many physical systems are modeled using (partial) differential equations which are derived from the laws of physics [9, 31, 32]. This modeling approach has the benefits that one can build a functional model using a small amount of data and the model may generalize well outside of the training data distribution (provided that the modeling assumptions are correct). However, building the model from the first principles requires deep understanding of the modeled process and often results in a tedious procedure when various modeling assumptions are tested in how well they can explain the data. The data-driven approach is therefore an attractive alternative: one can fit a generic model like neural networks [13] to training data without much effort on the model design and the derivations of the learning and inference procedures. The downside is, however,

^{*} This version of the contribution has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-031-15919-0_47. Use of this Accepted Version is subject to the publisher’s Accepted Manuscript terms of use <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

that the accuracy of this model depends greatly on the amount of available data: too little data may result in models that do not generalize well. Thus, there is clear demand for combining two modeling approaches: using neural networks models for better flexibility while also constraining the solutions with laws of physics can greatly improve sample efficiency. The laws of conservation (of energy/mass/momentum) are among very common modeling assumptions made for describing physical systems. Many physical systems can be modeled as closed and therefore conservative systems [27]. Combining the conservation laws with neural networks (see, e.g., [7, 14, 18, 19, 23]) is therefore a promising line of research with many potential applications.

One prominent research direction that emerged recently in the literature is modeling Hamiltonian systems with neural networks [14]. In Hamiltonian neural networks (HNNs), the law of the energy conservation is in-built in the structure of the dynamics model and therefore it is automatically satisfied. The idea of utilizing Hamilton’s equations was successfully used to predict the dynamics of Hamiltonian systems from pixel observations [17, 35, 40], to build representations of molecular data [25] and it was extended to control tasks [38, 40] and meta-learning [24]. The original HNN model [14] had the limitation of assuming the knowledge of the state derivatives with respect to time or approximating those using finite differences. Many recent works have used numerical integrators for modeling the evolution of the system state and several improvements of the integration procedure have been proposed [5, 8, 10, 20, 34, 37].

In this paper, we propose to model the evolution of the system state by adding another neural network instead of relying on traditional numerical integrators for ordinary differential equations (ODEs). This alternative to numerical integration is known in the literature under the name *deep hidden physics models* [29]. Our approach provides a continuous-time approximation of system states without relying on additional assumptions such as Hamiltonian separability. Hamiltonian preservation is encoded as a soft constraint through an extra loss term rather than being in-built in the architecture itself. We experimentally show that the proposed approach can improve the modeling accuracy in the presence of observation noise and for measurements with low sampling rates or irregularly-sampled observations.

2 Modeling Hamiltonian systems

2.1 Hamiltonian neural networks

The modeling assumption of Hamiltonian neural networks [14] is that the observed state $\mathbf{s} = (\mathbf{q}, \mathbf{p})$ of a dynamical system evolves according to Hamilton’s equations:

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

where \mathcal{H} is the Hamiltonian (total energy) of the system, \mathbf{q} is the position and \mathbf{p} is the momentum part of the system state $\mathbf{s} = (\mathbf{q}, \mathbf{p})$. Although Hamiltonian

$\mathcal{H}(\mathbf{q}, \mathbf{p})$ is modeled with a generic neural network with inputs \mathbf{q} and \mathbf{p} , using equation 1 to describe the system dynamics guarantees that the total energy is conserved:

$$\frac{\partial \mathcal{H}}{\partial t} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \frac{d\mathbf{p}}{dt} + \frac{\partial \mathcal{H}}{\partial \mathbf{q}} \frac{d\mathbf{q}}{dt} = 0. \quad (2)$$

The original HNN model [14] was trained by minimizing the loss

$$\mathcal{L}_{\text{HNN}} = \frac{1}{N} \sum_{i=1}^N \left(\frac{d\mathbf{q}_i}{dt} - \frac{\partial \mathcal{H}_i}{\partial \mathbf{p}} \right)^2 + \left(\frac{d\mathbf{p}_i}{dt} + \frac{\partial \mathcal{H}_i}{\partial \mathbf{q}} \right)^2 \quad (3)$$

where $\frac{\partial \mathcal{H}_i}{\partial \mathbf{p}}$, $\frac{d\mathbf{p}_i}{dt}$, $\frac{\partial \mathcal{H}_i}{\partial \mathbf{q}}$, $\frac{d\mathbf{q}_i}{dt}$ are partial derivatives computed at the locations of N training examples $\mathbf{q}(t_i), \mathbf{p}(t_i)$. The derivatives $\frac{\partial \mathcal{H}_i}{\partial \mathbf{p}}$, $\frac{\partial \mathcal{H}_i}{\partial \mathbf{q}}$ are calculated by differentiating the neural network which models the Hamiltonian, while the derivatives $\frac{d\mathbf{p}_i}{dt}$, $\frac{d\mathbf{q}_i}{dt}$ are either assumed to be known (from the simulator) or approximated with finite differences. Using finite differences to approximate the derivatives $\frac{d\mathbf{p}}{dt}$ and $\frac{d\mathbf{q}}{dt}$ is essentially equivalent to Euler integration with a time step being equal to the sampling interval, which limits the accuracy of the trained model [8].

Many extensions of HNNs [5, 10, 34, 35] use more advanced numerical integrators combined with the Neural ODE approach [3] to model the evolution of the system state in time. Several works [5, 10, 20, 34, 37] use symplectic integrators which preserve the conserved quantity and therefore are natural options for Hamiltonian systems. The analysis [41] of several numerical integrators when applied to HNNs shows that non-symplectic integrators cannot guarantee the recovery of true Hamiltonian \mathcal{H} and the prediction accuracy obtained with a symplectic integrator depends on the integrator accuracy order. In addition to the use of a symplectic integrator, SympNets [20] have the network architecture that guarantees zero energy loss by network design. Some symplectic integration schemes [5, 10, 34] make additional assumptions such as the separability of the Hamiltonian.¹ A recent model called Non-separable Symplectic Neural Networks (NSSNN) [37] releases this assumption by an improved symplectic integrator which works well for both separable and non-separable Hamiltonians.

One potential problem in applying model equations 1 to real-world data is the fact that the system dynamics in equation 1 is written for clean states \mathbf{p} and \mathbf{q} while in practice state measurements typically contain noise. Working with noisy states leads to inaccurate modeling due to the compounding error problem. Therefore the states should be denoised both at training and inference times. Many existing HNN models were trained with noisy observations but they do not have in-built techniques for handling observation noise in initial states at inference time. Some works directly address this issue: for example, [5] which proposes an initial state optimization procedure.

¹ Hamiltonian $\mathcal{H}(\mathbf{q}, \mathbf{p}) = V(\mathbf{q}) + K(\mathbf{p})$ where V and K are potential and kinetic energies is separable.

2.2 Physics-informed neural networks as an integrator for HNN

Physics-informed neural networks (PINNs) [22, 30] is a mesh-free method of solving given differential equations using neural networks. The method can be used to approximate a solution of an initial value problem defined by an ODE with a *known* function \mathbf{f}

$$\frac{d\mathbf{s}(t)}{dt} = \mathbf{f}(t, \mathbf{s}(t)) \quad (4)$$

and initial conditions $\mathbf{s}_0 = \mathbf{s}(t_0)$. The solution $\mathbf{s}(t)$ of the ODE is approximated by a neural network with time t as input trained by minimizing a composite loss function. The loss function for the network training can be represented as a weighted sum of the supervised learning loss for the initial conditions

$$\mathcal{L}_{\text{init}} = \|\mathbf{s}(t_0) - \mathbf{s}_0\|^2 \quad (5)$$

and the loss forcing the network to satisfy the ODE in equation 4:

$$\mathcal{L}_{\text{ode}} = \frac{1}{K} \sum_{k=1}^K \left(\frac{d\mathbf{s}_k}{dt} - \mathbf{f}(t_k) \right)^2 \quad (6)$$

where $\frac{d\mathbf{s}_k}{dt}$ denotes derivatives computed at locations t_k . The locations t_k can be sampled randomly on the interval on which the ODE is solved.

In case a sequence of observations $\{(t_0, \mathbf{s}_0), \dots, (t_N, \mathbf{s}_N)\}$ from the modelled system in equation 4 is available, PINNs allow to easily include the observations in the model training procedure in which case the loss in equation 5 is replaced with the following supervision loss:

$$\mathcal{L}_{\text{fit}} = \frac{1}{N} \sum_{i=1}^N \|\mathbf{s}(t_i) - \mathbf{s}_i\|^2. \quad (7)$$

The PINN method can be viewed as a supervised learning method with the ODE-based regularizer given in equation 6. Unlike traditional numerical solvers PINNs can handle ill-posed problems, e.g. with unknown initial conditions but with measurements for other time points.

Deep hidden physics models (DHPMs) [29] extend the PINNs approach to the case of an *unknown* function \mathbf{f} in equation 4. Function \mathbf{f} is approximated with another neural network which is trained by minimizing the loss in equation 6. Thus, DHPMs contain two neural networks: one defines the differential equation and the other one approximates its solution.

Deep Hidden Hamiltonian In this paper, we propose to use physics-informed neural networks as an integrator when learning HNNs. This means that we use an additional neural network $\mathbf{s}(t)$ to model the solution of Hamilton's equations 1 which describe the system dynamics. This additional neural network outputs the two components \mathbf{q}, \mathbf{p} of the system state as a function of input t .

The solution network $\mathbf{s}(t)$ is trained to fit the available observations $\mathbf{s}_i = (\mathbf{q}_i, \mathbf{p}_i)$ at time instances t_i by minimizing the loss in equation 7 and to satisfy Hamilton’s equations by minimizing the loss in equation 3. The value of Hamiltonian \mathcal{H} is approximated by an HNN which is jointly trained with the solution network $\mathbf{s}(t)$. Note that the derivatives $\frac{dq_i}{dt}$, $\frac{dp_i}{dt}$ required in the loss term in equation 3 can be computed by differentiating the neural network $\mathbf{s}(t)$ with respect to its input t . The locations of points t_i used to compute that loss do not have to coincide with the locations of the training samples: they are sampled randomly on the solution interval.

We found it beneficial to include in the loss an additional term which forces the energy values to stay constant throughout a trajectory:

$$\mathcal{L}_{\text{extra}} = \frac{1}{M} \sum_{ij} (\mathcal{H}(\mathbf{s}(t_i)) - \mathcal{H}(\mathbf{s}(t_j)))^2, \quad (8)$$

where pairs of points t_i, t_j are sampled randomly and M is the number of sampled pairs in a mini-batch. Thus, the total loss minimized during training is the weighted sum of the losses in equations 3, 7 and 8.

We call our algorithm Deep Hidden Hamiltonian (DHH) in analogy to DHPMs: we assume that Hamiltonian \mathcal{H} is unknown and should be learned from data.

3 Related work

Recently many improvements to the original HNN architecture have been proposed in the literature. The proposed improvements include the use of symplectic integrators [5, 10, 20, 34, 37], hard constraints on energy conservation [20] as well as modifications to the soft constraints, for example, by switching to the Cartesian coordinates [12]. More details on the comparison of different methods for Hamiltonian systems can be found in survey [39]. In contrast to many existing works, our method does not rely on traditional numerical integrators but instead utilizes an extra neural network for learning of the system trajectory. Another alternative to the approaches listed above is a method called generating function neural network or GFNN [2] that learns modified generating functions as a symplectic map representation instead of approximating a vector field directly. Thus, the method does not require finite difference approximations of the vector field. However, GFNN requires solving a system of non-linear equations for prediction of one step evolution which might be a computational bottleneck.

When modelling physical systems one should take into account the measurement noise in the observations. In general, neural network regularization techniques [16, 33] can be applied to prevent overfitting to the training data and to make the model more robust to noise. For modelling of physical systems in particular, previous works have a separate denoising step, for example, by basis functions approximations [36] or an optimization procedure for the initial step [5]. In comparison, the proposed algorithm relies on the solution network

that finds a continuous-time approximation of the system state trajectory from noisy observations together with the regularization loss term in equation 8.

Several recent works [11, 19, 26, 28] utilize physics-informed neural networks for modelling of conservative systems. The closest to ours is work [26] that applies PINNs for solving Hamilton’s differential equations with a *known* Hamiltonian \mathcal{H} . In contrast, we learn the Hamiltonian from the data following the DHPMs methodology.

4 Experiments

We test our method on the following four physical systems from [14]:

- mass-spring

$$\mathcal{H} = \frac{1}{2}kq^2 + \frac{p^2}{2m} \quad (9)$$

with $m = \frac{1}{2}$ and $k = 2$;

- pendulum

$$\mathcal{H} = 2mgl(1 - \cos q) + \frac{l^2p^2}{2m} \quad (10)$$

with $l = 1$, $m = \frac{1}{2}$ and $g = 3$;

- 2-body and 3-body systems

$$\mathcal{H} = \sum_{i=0}^N \frac{\|\mathbf{p}_i\|^2}{2m_i} - \sum_{1 \leq i < j \leq N} \frac{Gm_i m_j}{\|\mathbf{q}_i - \mathbf{q}_j\|} \quad (11)$$

with $G = 1$, $m_i = 1$, $i = 1, \dots, N$ and $N \in \{2, 3\}$.

We generate training data by randomly sampling the initial state and numerically solving differential equations 1 using the fourth-order Runge-Kutta method with \mathcal{H} given in equations 9 – 11 for a given number of time steps. We use 80% of sequential time steps from the generated time series as a training set and the remaining 20% as a test set. To test the robustness of the proposed method to noise in observations, the generated data points are corrupted with additive Gaussian noise with zero mean and 0.1 standard deviation.

In all the experiments, Hamiltonian \mathcal{H} , solutions $\mathbf{s}(t)$ and dynamics \mathbf{f} (for a DHPM baseline) are modeled with multi-layer perceptron networks. The models are optimized using the Adam optimizer [21] with learning rates of 0.0001 for HNN and dynamics \mathbf{f} and 0.01 for the solution network. We normalize the modeling interval of t to be $[-1, 1]$. Points t_k required for computing the loss terms in equation 3 and equation 6 for the DHPM baseline (described in Sec. 2.2) are sampled uniformly from the interval $[-1, 1]$ for each optimization step. Similarly, pairs of points (t_i, t_j) for the extra loss term in equation 8 are sampled at each optimization step such that $t_i \in [-1, 0]$ and $t_j \in [0, 1]$. In our experiments, the training loss is the weighted sum of the supervision loss in equation 7 (with weight 1), the loss in equation 3 (with weight 0.1 for the mass-spring system

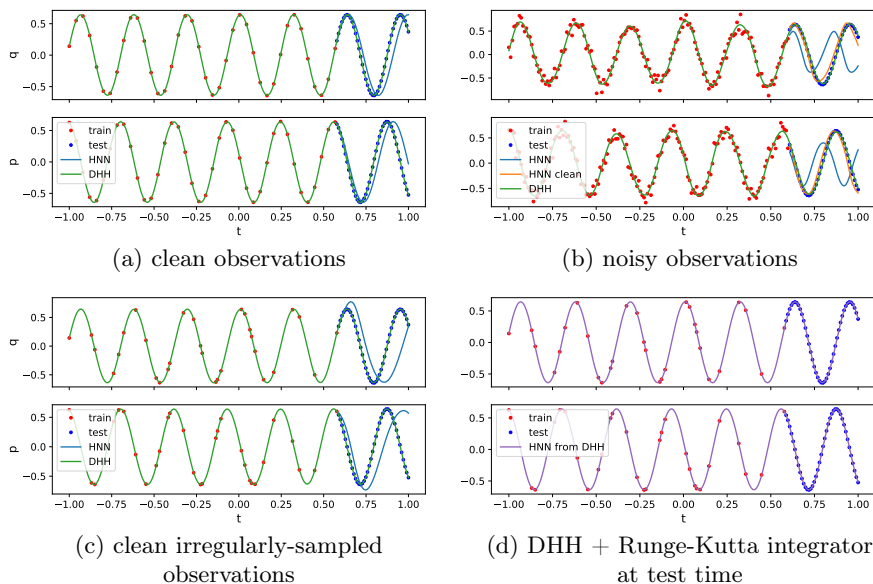


Fig. 1. (a)–(c): Estimated state trajectories for the mass-spring system with DHH (green) and HNN (blue). (d): The trajectory obtained by integrating the system equations 1 with the Runge-Kutta integrator using the Hamiltonian \mathcal{H} learned by DHH for the data from Fig. 1c.

and the pendulum and 1 for the 2-body and 3-body systems) and the loss in equation 8 (with weight 0.01).

In Fig. 1, we compare the trajectories for the mass-spring system estimated with the proposed DHH approach and with the HNN from [14] in which the derivatives are estimated using finite differences. Fig. 1a-c show the results for three settings: 1) clean regularly-sampled observations 2) noisy regularly-sampled observations 3) clean irregularly-sampled observations. In the experiment with noisy observations (Fig. 1b), we additionally show a trajectory obtained by the baseline HNN when we assume access to the clean state observation for the last time step in the training set. The results show that the proposed method performs well under low sampling rates and it tolerates noise in the data. Additionally, the system trajectory can be obtained by integrating equations 1 with, for example, the Runge-Kutta integrator using the Hamiltonian \mathcal{H} learned by DHH (see Fig. 1d). The results show that the Hamiltonian \mathcal{H} found by DHH is very accurate and the model works well even when changing the integration scheme at test time.

The proposed algorithm can also be used when some of the state variables are not observed (see Fig. 2). In this experiment, we model the mass-spring system using only noisy measurements of the position variable \mathbf{q} while \mathbf{p} is assumed unobserved. The only change in the training procedure is the calculation of the supervision loss in equation 7 just for \mathbf{q} measurements. As the results show, the

proposed method is able to reconstruct the missing coordinate (up to an additive constant).

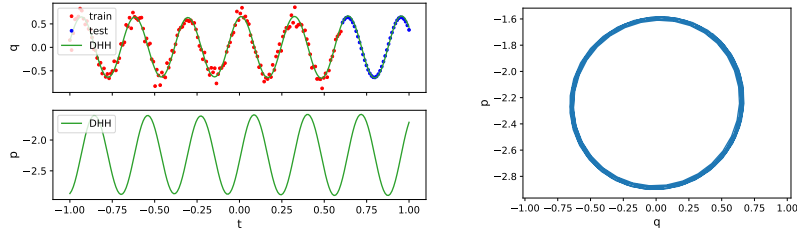


Fig. 2. Estimated trajectories and the corresponding vector field by DHH for the mass-spring system with noisy observations of q and no observations of p .

Next, we quantitatively compare the proposed approach against the following baselines: 1) HNN [14] with derivatives calculated as finite differences, 2) HNN [14] with derivatives provided by the simulator, 3) NSSNN [37], 4) Neural ODE [3] and 5) DHPMs [29]. The implementations of NSSNN and HNN are taken from the original papers. For Neural ODE, we use the same implementation as in [37] with the second-order Runge-Kutta integrator. DHPMs estimate the system dynamics in equation 4 by modeling f with a multi-layer perceptron and by minimizing the sum of the losses in equations 6 and 7.

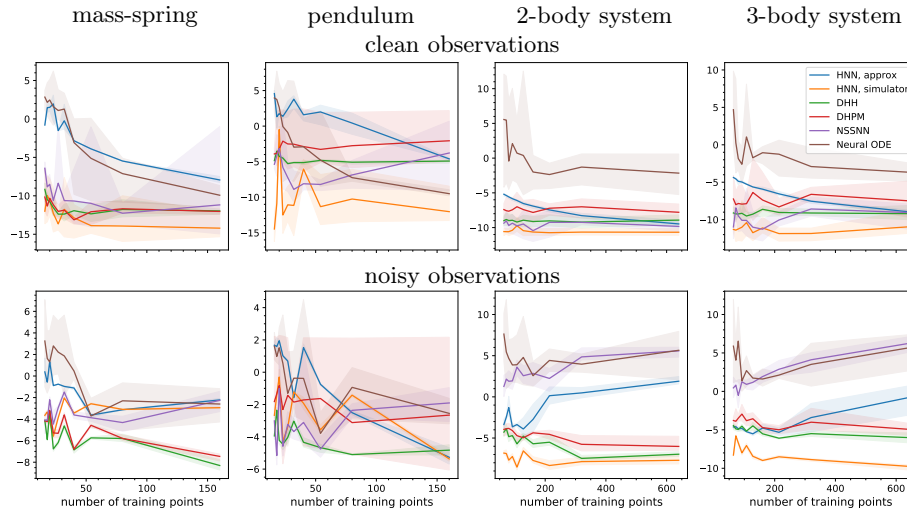


Fig. 3. Log-mean-squared errors between the estimated trajectories and the ground truth as a function of the sampling rate on clean (first row) and noisy (second row) observations.

In Fig. 3, we present the effect of the sampling rate on the log mean-squared error (log-MSE) of the estimated trajectories compared to the ground truth. We show the average log-MSE for each model across five runs with a solid line, while the shaded interval represent the minimum and maximum log-MSEs among the five runs.² The x -axis corresponds to the number of points in the training set: smaller number of points in the training set corresponds to a lower sampling rate. As can be seen from the results, the proposed method performs similarly to the baselines on clean observations and it outperforms the baselines on noisy observations. We attribute the improvement to the energy conservation bias in-built in the model and to the noise filtration capabilities of the solution network. The performance of DHH and DHPMs is more stable in the low sampling rate regime compared to the approaches with traditional numerical integrators due to computational challenges of numerical integration in this scenario. However, methods, such as Neural ODE and HNN perform better with increase of sampling rates of clean observations. Note that the HNN model that has access to the simulator derivatives is an unrealistic approach because those derivatives are not available in practical applications.

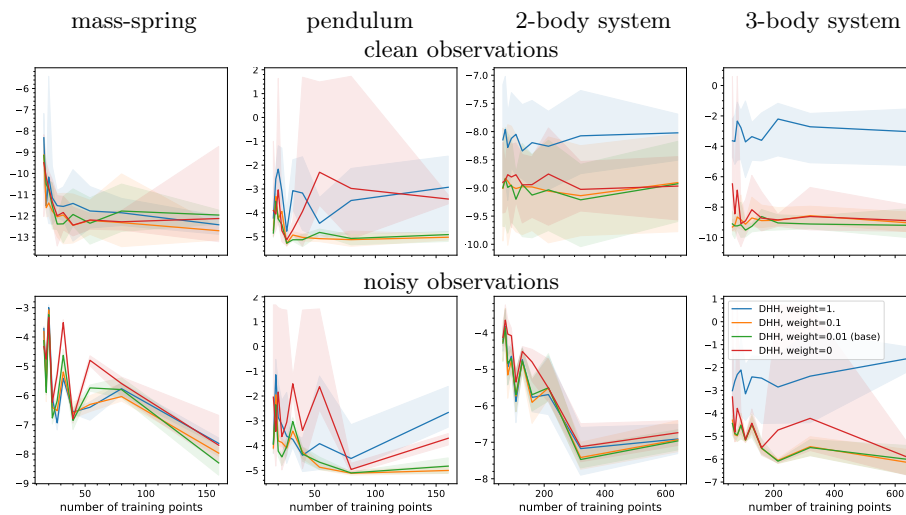


Fig. 4. Log-mean-squared errors between the estimated trajectories and the ground truth as a function of the sampling rate on clean (first row) and noisy (second row) observations for different weights for $\mathcal{L}_{\text{extra}}$ term added to the loss function.

Additionally, the use of the extra loss term in equation 8 has a positive effect on the training of the proposed model. In Fig. 4, we report the results for four studied systems trained with the term in equation 8 added to the loss function

² Minimum and maximum errors are used to emphasize extreme cases (e.g., failed runs).

with different weights. Even though the model is quite robust to small variations in the weight (e.g., orange and green lines), stronger regularization (e.g., with weight equal to 1 in blue) or no regularization (in red) hurts the performance.

5 Conclusion and future work

In this work, we proposed to learn a continuous-time trajectory of a modeled Hamiltonian system using an additional neural network. The time derivatives provided by this network can replace the finite-difference estimates of the derivatives used as the targets in the original HNN model. We showed experimentally that the proposed approach can outperform existing alternatives, especially in the case of low sampling rates and presence of noise in the state measurements.

A big limitations of the HNN methodology is its applicability only to conservative systems described by Hamilton’s equations and observed in the canonical coordinates. These assumptions may be restrictive in many practical applications. Addressing this limitation is an important line for future research with many promising results obtained recently (see, e.g., [4, 6, 7, 18]).

Another important research question is to identify the most efficient way to incorporate inductive biases from physics into modeling of dynamical systems, which is a topic of active debate at the moment [1, 15].

Acknowledgments We thank CSC (IT Center for Science, Finland) for computational resources and the Academy of Finland for the support within the Flagship programme: Finnish Center for Artificial Intelligence (FCAI).

References

1. Botev, A., Jaegle, A., Wirnsberger, P., Hennes, D., Higgins, I.: Which priors matter? benchmarking models for learning latent dynamics. arXiv preprint arXiv:2111.05458 (2021)
2. Chen, R., Tao, M.: Data-driven prediction of general hamiltonian dynamics via learning exactly-symplectic maps. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 1717–1727. PMLR (18–24 Jul 2021)
3. Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 31. Curran Associates, Inc. (2018)
4. Chen, Y., Matsubara, T., Yaguchi, T.: Neural symplectic form: Learning hamiltonian equations on general coordinate systems. In: Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems (2021)
5. Chen, Z., Zhang, J., Arjovsky, M., Bottou, L.: Symplectic recurrent neural networks. In: International Conference on Learning Representations (2020)
6. Choudhary, A., Lindner, J.F., Holliday, E.G., Miller, S.T., Sinha, S., Ditto, W.L.: Forecasting hamiltonian dynamics without canonical coordinates. *Nonlinear Dynamics* **103**(2), 1553–1562 (2021)

7. Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., Ho, S.: Lagrangian neural networks. arXiv preprint arXiv:2003.04630 (2020)
8. David, M., Méhats, F.: Symplectic learning for hamiltonian neural networks. arXiv preprint arXiv:2106.11753 (2021)
9. Davis, M.E., Davis, R.J.: Fundamentals of chemical reaction engineering. Courier Corporation (2012)
10. DiPietro, D., Xiong, S., Zhu, B.: Sparse symplectically integrated neural networks. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 6074–6085. Curran Associates, Inc. (2020)
11. Fang, Y., Wu, G.Z., Wen, X.K., Wang, Y.Y., Dai, C.Q.: Predicting certain vector optical solitons via the conservation-law deep-learning method. *Optics & Laser Technology* **155**, 108428 (2022)
12. Finzi, M., Wang, K.A., Wilson, A.G.: Simplifying hamiltonian and lagrangian neural networks via explicit constraints. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 13880–13889. Curran Associates, Inc. (2020)
13. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT Press (2016)
14. Greydanus, S., Dzamba, M., Yosinski, J.: Hamiltonian neural networks. In: Advances in Neural Information Processing Systems. vol. 32 (2019)
15. Gruver, N., Finzi, M.A., Stanton, S.D., Wilson, A.G.: Deconstructing the inductive biases of hamiltonian neural networks. In: International Conference on Learning Representations (2022)
16. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)
17. Hochlehnert, A., Terenin, A., Saemundsson, S., Deisenroth, M.: Learning contact dynamics using physically structured neural networks. In: Banerjee, A., Fukumizu, K. (eds.) Proceedings of The 24th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 130, pp. 2152–2160. PMLR (13–15 Apr 2021)
18. Hoedt, P., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G., Hochreiter, S., Klambauer, G.: MC-LSTM: mass-conserving LSTM. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning, ICML 2021. Proceedings of Machine Learning Research, vol. 139, pp. 4275–4286. PMLR (2021)
19. Jagtap, A.D., Kharazmi, E., Karniadakis, G.E.: Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering* **365**, 113028 (2020)
20. Jin, P., Zhang, Z., Zhu, A., Tang, Y., Karniadakis, G.E.: Sympnets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems. *Neural Networks* **132**, 166–179 (2020)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
22. Lagaris, I.E., Likas, A., Fotiadis, D.I.: Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks* **9**(5), 987–1000 (1998)

23. Lee, K., Carlberg, K.T.: Deep conservation: A latent-dynamics model for exact satisfaction of physical conservation laws. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**(1), 277–285 (May 2021)
24. Lee, S., Yang, H., Seong, W.: Identifying physical law of hamiltonian systems via meta-learning. In: *International Conference on Learning Representations* (2021)
25. Li, Z., Yang, S., Song, G., Cai, L.: Hamnet: Conformation-guided molecular representation with hamiltonian neural networks. In: *International Conference on Learning Representations* (2021)
26. Mattheakis, M., Sondak, D., Dogra, A.S., Protopapas, P.: Hamiltonian neural networks for solving differential equations. *arXiv preprint arXiv:2001.11107* (2020)
27. Noether, E.: Invariant variation problems. *Transport theory and statistical physics* **1**(3), 186–207 (1971)
28. Patel, R.G., Manickam, I., Trask, N.A., Wood, M.A., Lee, M., Tomas, I., Cyr, E.C.: Thermodynamically consistent physics-informed neural networks for hyperbolic systems. *Journal of Computational Physics* **449**, 110754 (2022)
29. Raissi, M.: Deep hidden physics models: Deep learning of nonlinear partial differential equations. *The Journal of Machine Learning Research* **19**(1), 932–955 (2018)
30. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* **378**, 686–707 (2019)
31. Rubinstein, I., Rubinstein, L.: *Partial Differential Equations in Classical Mathematical Physics*. Cambridge University Press (1994)
32. Taylor, J.R.: *Classical Mechanics*. University Science Books (2005)
33. Tibshirani, R.J.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288 (1996)
34. Tong, Y., Xiong, S., He, X., Pan, G., Zhu, B.: Symplectic neural networks in taylor series form for hamiltonian systems. *Journal of Computational Physics* **437**, 110325 (2021)
35. Toth, P., Rezende, D.J., Jaegle, A., Racanière, S., Botev, A., Higgins, I.: Hamiltonian generative networks. In: *International Conference on Learning Representations* (2020)
36. Wu, K., Qin, T., Xiu, D.: Structure-preserving method for reconstructing unknown hamiltonian systems from trajectory data. *SIAM Journal on Scientific Computing* **42**(6), A3704–A3729 (2020)
37. Xiong, S., Tong, Y., He, X., Yang, S., Yang, C., Zhu, B.: Nonseparable symplectic neural networks. In: *International Conference on Learning Representations* (2021)
38. Zhong, Y.D., Dey, B., Chakraborty, A.: Symplectic ode-net: Learning hamiltonian dynamics with control. In: *International Conference on Learning Representations* (2020)
39. Zhong, Y.D., Dey, B., Chakraborty, A.: Benchmarking energy-conserving neural networks for learning dynamics from data. In: *Proceedings of the 3rd Conference on Learning for Dynamics and Control*. *Proceedings of Machine Learning Research*, vol. 144, pp. 1218–1229. PMLR (07 – 08 June 2021)
40. Zhong, Y.D., Leonard, N.: Unsupervised learning of lagrangian dynamics from images for prediction and control. In: *Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems*, vol. 33, pp. 10741–10752. Curran Associates, Inc. (2020)
41. Zhu, A., Jin, P., Tang, Y.: Deep hamiltonian networks based on symplectic integrators. *arXiv preprint arXiv:2004.13830* (2020)