

Physics-Informed Learning Using Hamiltonian Neural Networks with Output Error Noise Models [★]

Sarvin Moradi ^{*,**} Nick Jaensson ^{***} Roland Tóth ^{*,****}
Maarten Schoukens ^{*,**}

^{*} Control System group, Eindhoven University of Technology,
Eindhoven, the Netherlands

^{**} Eindhoven Artificial Intelligence Systems Institute, Eindhoven, the
Netherlands

^{***} Processing and Performance of Materials Group, Eindhoven
University of Technology, Eindhoven, the Netherlands

^{****} Systems and Control Laboratory, Institute for Computer Science
and Control, Budapest, Hungary

Abstract: In order to make data-driven models of physical systems interpretable and reliable, it is essential to include prior physical knowledge in the modeling framework. Hamiltonian Neural Networks (HNNs) implement Hamiltonian theory in deep learning and form a comprehensive framework for modeling autonomous energy-conservative systems. Despite being suitable to estimate a wide range of physical system behavior from data, classical HNNs are restricted to systems without inputs and require noiseless state measurements and information on the derivative of the state to be available. To address these challenges, this paper introduces an Output Error Hamiltonian Neural Network (OE-HNN) modeling approach to address the modeling of physical systems with inputs and noisy state measurements. Furthermore, it does not require the state derivatives to be known. Instead, the OE-HNN utilizes an ODE-solver embedded in the training process, which enables the OE-HNN to learn the dynamics from noisy state measurements. In addition, extending HNNs based on the generalized Hamiltonian theory enables to include external inputs into the framework which are important for engineering applications. We demonstrate via simulation examples that the proposed OE-HNNs results in superior modeling performance compared to classical HNNs.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Hamiltonian Neural Networks, Nonlinear system identification, Machine learning.

1. INTRODUCTION

Despite significant developments in modeling and system identification methods, learning complex systems is still a challenging problem (Schoukens and Ljung (2019)). The development of classical physics-based (white-box) models requires detailed knowledge of the system dynamics which is often not sufficiently available or tedious and expensive to obtain. On the other hand, data-driven (black-box) methods are able to reconstruct complex unknown dynamics from data, but the estimated models often lack generalizability and are hard to interpret.

Hamiltonian neural networks (HNNs) (Greydanus et al., 2019) are a state-of-the-art grey-box modeling tool in which Hamiltonian mechanics is embedded as prior knowledge in the neural network. By means of Hamiltonian theory, various physical mechanisms from quantum (Verdon et al., 2019) to large-scale systems (Sienko et al., 2004) can be expressed in a unified approach. However, the application of classical HNNs is limited to physical

systems without inputs and it requires noiseless state measurements and the knowledge of the state derivatives for the training process. These are not readily available in engineering applications.

After the introduction of HNNs by Greydanus et al. (2019), many researchers from different fields studied its application and proposed extensions to the original idea. For instance, Choudhary et al. (2020) applied HNNs for modeling nonlinear systems with chaotic behaviour, and Zhong et al. (2020) generalized HNNs for Hamiltonian systems with external inputs. A review of the different approaches for application of HNNs is available in Chen et al. (2022). Nevertheless, there remains a need for an HNN approach that is capable of dealing with input, dissipation, and noisy data in engineering systems.

Hamilton's theory assumes the system is energy conservative. However, in practice, energy changes due to dissipation and external forces. Greydanus and Sosanya (2022) has proposed to model dissipativity in dynamic systems by decomposition of dissipative and conservative quantities and training two subnetworks for these quantities separately. They have assumed that the noiseless time

[★] This work is part of the DAMOCLES research project which received funding from the Eindhoven Artificial Intelligence Systems Institute, as part of the EMDAIR funding programme.

derivative of the states of the physical system is available. Zhong et al. (2020) tried to model systems with control inputs. They have applied generalized Hamiltonian theory to incorporate the control inputs into the HNN framework. In their method, the noiseless state signals and constant input levels have been considered as the inputs of the model. Desai et al. (2021) used the port-Hamiltonian theory to formulate neural networks based model learning capable to consider both dissipation and inputs of the physical systems. In their approach, it is assumed that the input and dissipative terms are unknown, but noiseless measured data on the state derivatives are available. Unlike previous studies, we will not rely on noiseless state/state derivatives for training the generalized HNNs, instead we focus on identification of the systems subject to inputs with noisy state measurements.

In order to improve the stability of the HNNs, Chen et al. (2020), and Toth et al. (2020) introduced an other physical constraint to the HNNs by applying symplectic integrator for deriving states from HNNs. Since the symplectic integrators conserve the Hamiltonian, they can help the HNN to conserve energy for longer time periods and improve the accuracy of the final results. Although by employing symplectic integrators, the assumption that the Hamiltonian is separable into the kinetic and potential energy components must hold, which significantly limits the applicability of this approach. Xiong et al. (2021) and Sharma et al. (2022) have tried to overcome this limitation for modeling systems with non-separable Hamiltonians, but their work has been limited to systems without inputs and dissipation. It is worth mentioning that the main objective with using the symplectic integrators has been to reduce the errors originating from discretization and numerical integration while handling measurements noise through an appropriate estimation error method has been not addressed.

Real-life engineering systems are, in most cases, subjected to inputs and their measured states are noisy. Since Hamiltonian systems are stiff (Zhong et al., 2020) and sensitive to noise, the application of HNNs for modeling systems with noisy measurements is challenging. Stiff systems need small integration steps to deliver stable solutions. In other words, these systems are sensitive to the discretization error and noisy measurements making the final results easily diverge in forward simulation of the models. Previous studies mainly focused on delivering stable results by introducing extra constraint via incorporating symplectic integrators into HNNs. In this paper, we show that the underlying problems can be successfully addressed by handling noisy measurements via a proper model structure (Ljung, 1999) which provides estimation of the deterministic part of the model under statistical guarantees.

In order to handle the noise in measurements properly together with external inputs, an output-error model structure for HNNs is introduced for the first time, which results in an *output-error Hamiltonian neural network* (OE-HNN) method. For the proposed OE-HNN approach, we consider to minimize the simulation error to cope with the effects of the measurement noise and the unavailability of the state derivatives in practice. In order to accommodate the simulation error as a training objective an ODE-solver is used in the training step. While previous studies mainly

focused on modeling energy-conservative systems without inputs, the current study addresses modeling physical systems with inputs. The inputs should fully capture the dynamics of the system under study, hence we propose the use of multisines as a general type of inputs which allows for a more systematic exploration of the system's response across different frequency bands, enabling a more comprehensive and accurate model of the system's behavior.

To summarise our main contributions in this paper are the following:

- Incorporating an output-error model structure into the HNN modeling concept which allows to properly handle noisy measurement data.
- Contrary to the classical HNN approach, there is no need for numerical approximation or direct noise-free measurements of the state derivatives for training the model.
- Introduction of general type of inputs into the HNN modeling approach. For gathering data from such systems for HNN modeling we propose the use of multisines, giving the user full control over the applied power spectrum on the system.

The paper is organized as follows: Section 2 briefly recapitulates the Hamiltonian theory. Section 3 introduces the considered system class, model structure, and the details of the proposed OE-HNN based identification approach. Next, the performance of the proposed OE-HNNs is evaluated via simulation examples in Section 4. Finally, the conclusion on the developed approach and the achieved results are drawn in Section 5.

2. HAMILTONIAN DYNAMICS

Hamiltonian mechanics is a reformulation of classical mechanics. It focuses on symplectic geometry and the conservation of energy. In Hamiltonian mechanics, the total energy of the system E_{tot} is conserved and defined as the Hamiltonian of the position $\mathbf{q}(t) \in \mathbb{R}^n$ and momenta $\mathbf{p}(t) \in \mathbb{R}^n$ vectors, i.e. $E_{\text{tot}} = H(\mathbf{q}(t), \mathbf{p}(t))$ where $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$. The Hamiltonian is defined to be a differentiable scalar function, satisfying

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}. \quad (1)$$

Equation (1) represents the time evolution of the system in which the right-hand side of the equation is the symplectic gradient of the Hamiltonian. Since

$$\dot{H} = \left(\frac{\partial H}{\partial \mathbf{q}} \right)^\top \dot{\mathbf{q}} + \left(\frac{\partial H}{\partial \mathbf{p}} \right)^\top \dot{\mathbf{p}} = 0, \quad (2)$$

one can make sure that by moving along the symplectic gradient of the Hamiltonian, the Hamiltonian does not change; in other words, as long as the system follows the trajectory introduced in (1), the total energy is conserved.

In the generalized version of Hamiltonian theory, inputs $\mathbf{u}(t) \in \mathbb{R}^m$ are added to the relation as

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} + \mathbf{G}\mathbf{u}, \quad (3)$$

where $\mathbf{G} \in \mathbb{R}^{n \times m}$ is the input matrix. Unlike in the classical Hamiltonian theory, inputs can add or remove

energy in the system. Hence, the total energy is no longer conserved.

3. IDENTIFICATION APPROACH

In this section, at first the considered system class and the details of the proposed model structure are discussed. Later, a brief overview of the classical HNN identification is provided and the proposed OE-HNN framework is introduced.

3.1 Considered system class

We consider systems that obey the Hamiltonian dynamics (1)-(3) and that can be represented in terms of continuous-time state-space equations as

$$\begin{aligned}\dot{\mathbf{x}}(t) &= f(\mathbf{x}(t), \mathbf{u}(t)) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t),\end{aligned}\quad (4)$$

where $\mathbf{x}(t) = (\mathbf{q}(t), \mathbf{p}(t))^T \in \mathbb{R}^{2n_q}$ is the state at time moment $t \in \mathbb{R}$ and f describes the evolution of \mathbf{x} , i.e., the dynamics of the system. The output measurements $\mathbf{y}(t) \in \mathbb{R}^{n_y}$ are contaminated by $\mathbf{v}(t) \in \mathbb{R}^{n_y}$, a zero-mean i.i.d. (white) noise with finite variance $\Sigma_v \in \mathbb{R}^{n_y} \times \mathbb{R}^{n_y}$. Here we consider $\mathbf{C} = \mathbf{I}$.

It is assumed that sampled measurements of the system are obtained with rate T_s . Hence, a data set of the measurements with \mathbf{x}_0 as the initial state is available:

$$\mathcal{D}_N = \{(\mathbf{y}(kT_s), \mathbf{u}(kT_s))\}_{k=0}^{N-1}$$

Since we need to integrate over time, the inter-sample behaviour of $\mathbf{u}(t)$ needs to be known. In this regard, the Zero-Order Hold (ZOH) assumption for the imposed inputs is necessary (i.e. $\mathbf{u}(kT_s + \alpha) = \mathbf{u}(kT_s), \forall \alpha \in [0, T_s)$) (Pintelon and Schoukens, 2012).

3.2 Model structure

In order to model the dynamics of the considered systems, an output-error model structure is chosen. A schematic description of the system and model structure is illustrated in Figure 1. Since the generalized Hamiltonian theory (3) is capable of fully representing dynamics of the above mentioned system class, one can define the parameterized model of the state dynamics, f_θ , as,

$$\begin{aligned}\dot{\hat{\mathbf{x}}}(t) &= f_\theta(\hat{\mathbf{x}}(t), \mathbf{u}(t)) = \mathbf{J} \frac{\partial H_\theta}{\partial \hat{\mathbf{x}}}(t) + \mathbf{G}\mathbf{u}(t) \\ \hat{\mathbf{y}}(t) &= \mathbf{C}\hat{\mathbf{x}}(t),\end{aligned}\quad (5)$$

where $\mathbf{J} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix}$, $\mathbf{G} = [\mathbf{0} \ \mathbf{I}]^T$, and $\mathbf{C} = \mathbf{I}$. Here, H_θ is the Hamiltonian of the system, considered to be a function parameterized with parameter vector θ .

3.3 Identification of classical HNN

Greydanus et al. (2019) proposed to learn the Hamiltonian, as an energy-like scalar value in an unsupervised manner. They formulated H as an *artificial neural network*

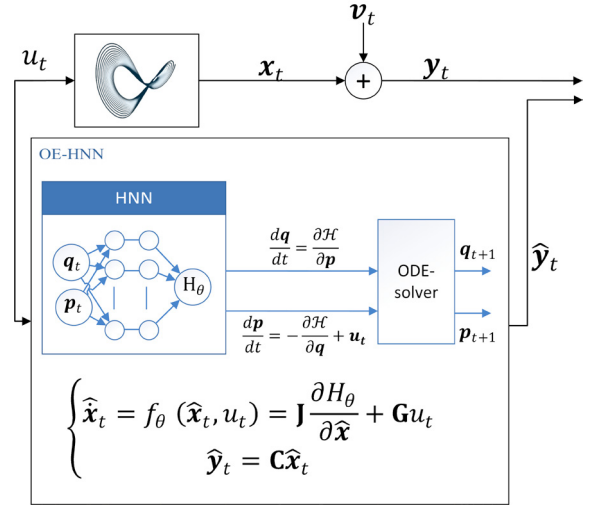


Fig. 1. Schematic description of the system and the proposed OE-HNN model structure.

(ANN), denoted as H_θ , and trained it by minimizing the loss function given below:

$$L_{\text{HNN}} = \left\| \frac{\partial H_\theta}{\partial \mathbf{p}} - \dot{\mathbf{q}} \right\|_2 + \left\| \frac{\partial H_\theta}{\partial \mathbf{q}} + \dot{\mathbf{p}} \right\|_2, \quad (6)$$

where $\|\cdot\|_2$ stands for the ℓ_2 norm of signals. In this way, the conservation law (1) is embedded in the trained ANN. The identified H fully defines a continuous-time motion model of the system, which can be calculated directly. However, as stated in (6), in order to train the HNNs, noiseless measurements of the position and momentum states are needed to obtain unbiased estimates using (6) and state derivatives are needed to be known as well.

3.4 Identification of OE-HNN

In most engineering systems, only noisy measurements of the position and momenta (states) are available. In order to surpass this issue, a proper treatment of the noise during the model estimation is required. To this end, the popular prediction error framework (Ljung, 1999) can be used. In this paper, we consider an output error noise setting during the identification of HNN models, resulting in the OE-HNN model structure. This corresponds of the presence of measurement noise on the states $\mathbf{x}(t) = (\mathbf{q}(t), \mathbf{p}(t))^T$ associated with the system under consideration.

Minimizing the prediction error in an OE setting results in minimizing the simulation error. Hence, similar to Chen et al. (2020) and Lee et al. (2021), an ODE solver is integrated in the HNN learning method. As illustrated in Figure 2, at each time step, the outputs of the HNN, the simulated state derivatives, are fed into the ODE-solver to predict the state in the next time step.

On top of that, unlike classical HNNs, the proposed approach is generalized to systems with inputs. The simulation loss function for the proposed OE-HNN approach is given by:

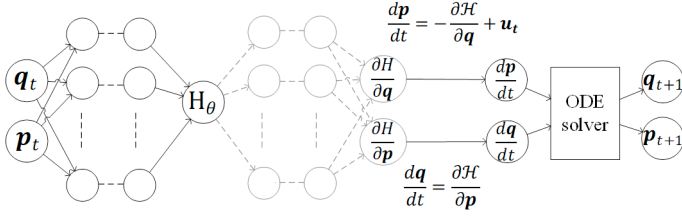


Fig. 2. Output-Error Hamiltonian Neural Network.

$$\begin{aligned} \min_{\theta} \sum_{k=1}^{N-1} \frac{1}{N} \|\mathbf{y}(kT_s) - \hat{\mathbf{y}}(kT_s)\|_2 \\ \text{s.t. } \dot{\hat{\mathbf{x}}}(t) = \mathbf{J} \frac{\partial H_{\theta}}{\partial \hat{\mathbf{x}}}(t) + \mathbf{G} \mathbf{u}(t) \\ \hat{\mathbf{y}} = \mathbf{C} \hat{\mathbf{x}}(t). \end{aligned} \quad (7)$$

To obtain the simulated model output, $\hat{\mathbf{y}}(kT_s)$, one can use an ODE-solver to solve (5) for a given initial state as

$$\begin{aligned} \hat{\mathbf{x}}((k+1)T_s) = \\ \text{ODE-solve} \left(\frac{\partial H_{\theta}}{\partial \hat{\mathbf{x}}}, \hat{\mathbf{x}}(kT_s), \mathbf{u}(kT_s), T_s \right). \end{aligned} \quad (8)$$

The ultimate goal of minimizing the loss function is to train the neural network which represents the parameterized Hamiltonian H_{θ} . This is depicted schematically in Figure 2. During training of the neural network, derivatives of the output of the neural network with respect to inputs are calculated via automatic differentiation. The derivatives of the Hamiltonian with respect to states and inputs are applied in (3) to calculate state time derivatives, which are fed into an ODE-solver to predict the states at the next time step. This procedure is repeated for the full length of the data trajectory starting from a known initial state. Here, the simulated outputs of the trajectory are compared to the output measurements (7) in order to train the model.

4. SIMULATION STUDY

4.1 Data-generating system

In order to illustrate performance of the proposed OE-HNN approach, data-driven modelling of a Duffing-type oscillator, depicted in Figure 3, is studied. Typically, these oscillators are used in passive vibration absorbers. In the oscillator, q is the position of the mass, and the momenta is defined by $p = m\dot{q}$ where m is the mass of the oscillator. The dynamic behaviour of the oscillator with a cubic nonlinear spring is expressed as

$$m\ddot{q} + kq - kq^3 = u, \quad (9)$$

where k is the stiffness of the spring and u is the imposed force. In this study, it is assumed that $k = m = 1$.

4.2 Data-set and training

To gather data, a multisine input is applied to the system. Multisines provide the user full control over the applied power spectrum (Pintelon and Schoukens, 2012). The input force u is defined as

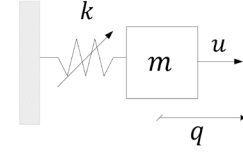
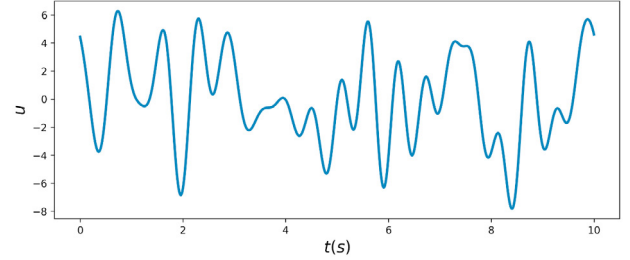
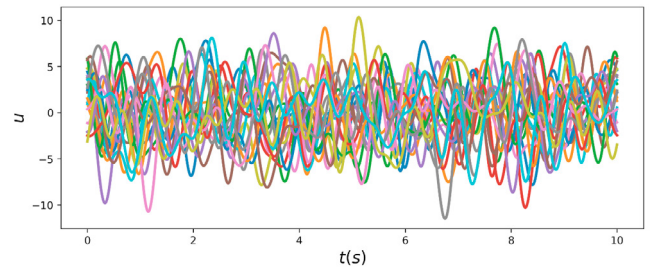


Fig. 3. Schematic representation of a Duffing-type oscillator.



(a) Multisine input force



(b) Various realizations of the input force

Fig. 4. Input force of the oscillator.

$$u = \sum_{k=1}^{k=20} \sin(2\pi k f_0 t + \phi_k), \quad (10)$$

where $f_0 = 0.1$ and the phase components ϕ_k are randomly chosen between $[0, 2\pi)$. The input force is designed to fully capture the behaviour of the considered oscillator. Figure 4a depicts a realization of the multisine input force.

In order to generate the data-set, (9) is numerically solved for 25 different realizations of the input (Figure 4b) with random initial states. The calculated states for each input are considered to be measured under an additive independent white noise of variance $\Sigma_v = 0.1$. Each trajectory contains 500 samples which are sampled from $t \in [5, 10]$ s with the sampling rate, $T_s = 0.01$. The response trajectories together with their inputs are collected into the data-set $\{(y_k, u_k)\}_{k=1}^N = \mathcal{D}_N$ for each realization. The training, validation and test data-sets consist of 15, 5, and 5 series of trajectories and input, respectively. The position trajectories of the training data-set are plotted in Figure 5.

A fully connected neural network with one hidden layer, 200 nodes and tanh activation function is used to formulate H_{θ} in the considered OE-HNN model. At each time step, the outputs of the neural network are fed into a fourth-order Runge-Kutta solver to predict the state at the next time step. The model is trained with a learning rate of 10^{-3} using the Adam optimizer method (Kingma and Ba, 2015).

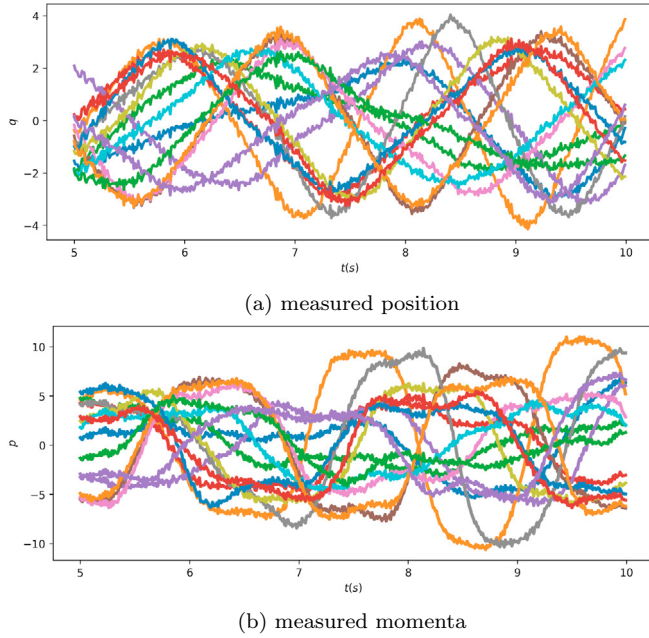


Fig. 5. Measured state trajectories in the training data-set.

4.3 Results

In order to assess the capability of the proposed OE-HNN approach to capture the behaviour of the oscillator, the simulated output response of the estimated OE-HNN model are compared on the test data set. Figures 6a and 6b show the resulting simulated position and momenta, respectively. By comparing the simulated states to the sampled states, it is evident that the OE-HNN can accurately model the behaviour of the oscillator.

A *multilayer perceptron* (MLP), fully connected feed forward neural network with q, p, u as inputs and \dot{q}, \dot{p} as outputs, and regular HNN model with the same architecture for H_θ as the OE-HNN are also used for modeling the oscillator in order to compare performance of the OE-HNN estimator to other methods. Unlike the OE-HNN, MLP and HNN methods need both states and state derivatives for training. Since the classical HNN does not take input into consideration, we had to slightly modify the loss function (6) to impose the input:

$$L_{\text{HNN}} = \left\| \frac{\partial H_\theta}{\partial p} - \dot{q} \right\|_2 + \left\| \frac{\partial H_\theta}{\partial q} + \dot{p} - Gu \right\|_2. \quad (11)$$

The same considerations are also taken into account during training of the MLP.

In Table 1, the root mean square error (RMSE) of the simulated position and momenta responses of the estimated models are cross compared. Based on the results, the proposed OE-HNN outperforms the MLP and the HNN. Note that the OE-HNN accumulates only a small error during simulation thanks to the Hamiltonian structure and the simulation error based objective function, which strongly penalizes the accumulation of errors over time. While the OE-HNN approach produces more accurate results, the training process requires more computation time. The same reasoning also applies to both the HNN and MLP approaches in which the MLP is trained faster.

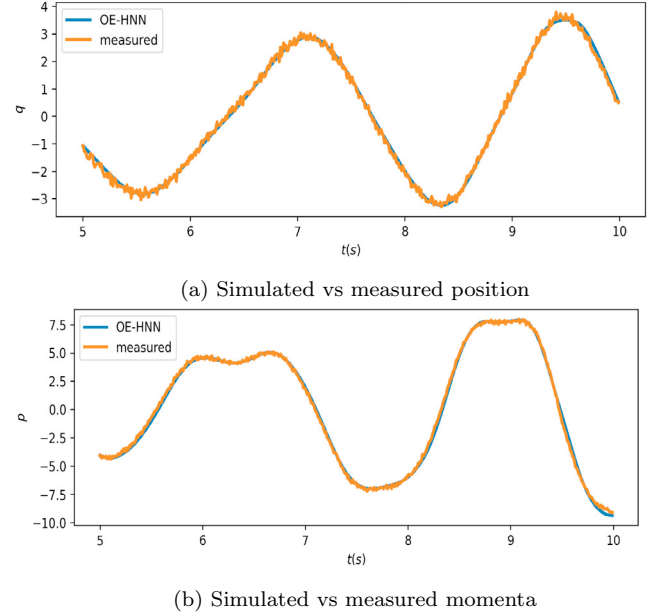


Fig. 6. Test results of the OE-HNN on the first realization of the test-set.

Table 1: RMSE of the simulated model responses for the nonlinear oscillator identification problem.

Method	\hat{q}	\hat{p}
MLP	0.071	0.135
HNN	0.048	0.084
OE-HNN	0.025	0.047

4.4 Connected nonlinear oscillators

In order to evaluate the generalizability of the proposed method to higher degrees of freedom, we also consider data-driven modelling of connected spring-mass systems, illustrated in Figure 7. For the sake of simplicity, it is assumed that $m_1 = m_2 = 0.5$, and $k_1 = k_2 = 1$. The same input force as in Section 4.2 is imposed to the second mass. Here the measured states are contaminated with a white noise of variance $\Sigma_v = 0.05$. For two connected oscillators, the considered state is defined as $\mathbf{x} = [q_1 \ q_2 \ p_1 \ p_2]^\top$.

The simulated response of the resulting model in comparison to the measured positions of the two masses are depicted in Figure 8. Based on the results, the obtained OE-HNN model can successfully capture the behaviour of two connected oscillators. With the same approach one can generalize the proposed method for systems with higher degrees of freedom.

Table 2 compares the accuracy of three ANNs for identifying the behaviour of two connected oscillators. According to Table 2, the HNN and MLP are outperformed by the OE-HNN in terms of accuracy on the test data set.

Table 2: RMSE of the simulated model responses for the connected two mass spring systems based identification problem.

Method	\hat{q}_1	\hat{q}_2	\hat{p}_1	\hat{p}_2
MLP	0.113	0.120	0.219	0.256
HNN	0.061	0.068	0.143	0.174
OE-HNN	0.028	0.035	0.074	0.088

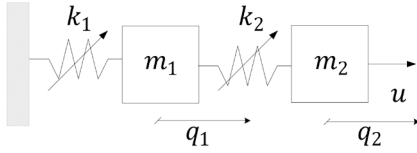
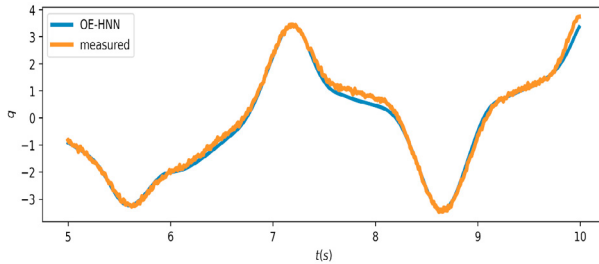
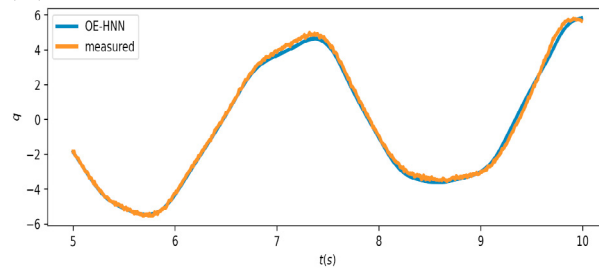


Fig. 7. Schematic illustration of the connected two mass-spring systems.



(a) Simulated model response vs measured momenta of the first mass (q_1)



(b) Simulated model response vs measured momenta of the second mass (q_2)

Fig. 8. Test results of the estimated OE-HNN model for two connected nonlinear oscillators on the first realization of the test-set.

5. CONCLUSION

In this paper, an OE-HNN approach is introduced to address the identification of physical systems subjected to the inputs and noisy state measurements. The OE-HNN method incorporates an output-error model structure which enables it to handle the noisy measurements properly by focusing the estimation on minimizing the simulation error instead of the prediction error in case of HNNs. To minimize the simulation error, an ODE-solver is implemented within the training step, which also enables to train the OE-HNN without the need for knowing the noiseless state time derivatives as is the case for the classical HNNs. In addition, by using the generalized Hamiltonian theory to formulate the OE-HNN, the approach is capable of modeling systems with inputs. The performance of the proposed OE-HNN is evaluated on a Duffing-type oscillator subject to multisine inputs. Based on the results, it is evident that the OE-HNN can successfully identify the oscillator given noisy state measurements. In addition, the generalizability of the proposed approach to systems with larger degrees of freedom is shown via a two mass-spring system example.

REFERENCES

Chen, Z., Zhang, J., Arjovsky, M., and Bottou, L. (2020). Symplectic recurrent neural networks. In *Proc. of the*

International Conference on Learning Representations, 1–23.

Chen, Z., Feng, M., Yan, J., and Zha, H. (2022). Learning neural hamiltonian dynamics: A methodological overview. *arXiv preprint arXiv:2203.00128*.

Choudhary, A., Lindner, J.F., Holliday, E.G., Miller, S.T., Sinha, S., and Ditto, W.L. (2020). Physics-enhanced neural networks learn order and chaos. *Physical Review E*, 101, 062207:1–8.

Desai, S.A., Mattheakis, M., Sondak, D., Protopapas, P., and Roberts, S.J. (2021). Port-hamiltonian neural networks for learning explicit time-dependent dynamical systems. *Physical Review E*, 104(3), 034312:1–10.

Greydanus, S. and Sosanya, A. (2022). Dissipative hamiltonian neural networks: Learning dissipative and conservative dynamics separately. *arXiv preprint arXiv:2201.10085*.

Greydanus, S., Dzamba, M., and Yosinski, J. (2019). Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 1–11. Curran Associates, Inc.

Kingma, D.P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *Proc. of the International Conference on Learning Representations (ICLR)*. Ithaca, NY.

Lee, K., Trask, N., and Stinis, P. (2021). Machine learning structure preserving brackets for forecasting irreversible processes. *Advances in Neural Information Processing Systems*, 34, 5696–5707.

Ljung, L. (1999). *System Identification: Theory for the User, Second Edition*. Prentice-Hall.

Pintelon, R. and Schoukens, J. (2012). *System identification: a frequency domain approach*. John Wiley & Sons.

Schoukens, J. and Ljung, L. (2019). Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*, 39(6), 28–99.

Sharma, H., Galioto, N., Gorodetsky, A.A., and Kramer, B. (2022). Bayesian identification of nonseparable hamiltonian systems using stochastic dynamic models. *arXiv preprint arXiv:2209.07646*.

Sienko, W., Citko, W., and Jakóbczak, D. (2004). Learning and system modeling via hamiltonian neural networks. In *Proc. of the International Conference on Artificial Intelligence and Soft Computing*, 266–271.

Toth, P., Rezende, D.J., Jaegle, A., Racanière, S., Botev, A., and Higgins, I. (2020). Hamiltonian generative networks. In *Proc. of the International Conference on Learning Representations*, 1–19.

Verdon, G., Marks, J., Nanda, S., Leichenauer, S., and Hidary, J. (2019). Quantum hamiltonian-based models and the variational quantum thermalizer algorithm. *arXiv preprint arXiv:1910.02071*.

Xiong, S., Tong, Y., He, X., Yang, S., Yang, C., and Zhu, B. (2021). Nonseparable symplectic neural networks. In *Proc. of the International Conference on Learning Representations*, 1–14.

Zhong, Y.D., Dey, B., and Chakraborty, A. (2020). Symplectic ode-net: Learning hamiltonian dynamics with control. In *Proc. of the International Conference on Learning Representations*, 1–17.