Hamiltonian Neural Koopman Operator

Jingdong Zhang\(^1\,2\) Qunxi Zhu\(^3\,*\) Wei Lin\(^{1,2,3,4,5,\ast}\)

\(^1\)School of Mathematical Sciences and Shanghai Centre for Mathematical Sciences, Fudan University, China
\(^2\)Research Institute of Intelligent Complex Systems, Fudan University, China
\(^3\)MOE Frontiers Center for Brain Science and Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, China
\(^4\)State Key Laboratory of Medical Neurobiology, Institutes of Brain Science, Fudan University
\(^5\)Shanghai Artificial Intelligence Laboratory, China

\{zhangjd20,qxzhu16,wlin\}@fudan.edu.cn

Abstract

Recently, physics-informed learning, a class of deep learning framework that incorporates the physics priors and the observational noise-perturbed data into the neural network models, has shown outstanding performances in learning physical principles with higher accuracy, faster training speed, and better generalization ability. Here, for the Hamiltonian mechanics and using the Koopman operator theory, we propose a typical physics-informed learning framework, named as Hamiltonian Neural Koopman Operator (HNKO) to learn the corresponding Koopman operator automatically satisfying the conservation laws. We analytically investigate the dimension of the manifold induced by the orthogonal transformation, and use a modified auto-encoder to identify the nonlinear coordinate transformation that is required for approximating the Koopman operator. Taking the Kepler problem as an example, we demonstrate that the proposed HNKO in robustly learning the Hamiltonian dynamics outperforms the representative methods developed in the literature. Our results suggest that feeding the prior knowledge of the underlying system and the mathematical theory appropriately to the learning framework can reinforce the capability of the deep learning.

1 Introduction

Reconstructing nonlinear dynamical systems solely using the observational noise-perturbed data is a focal challenge in various fields. The neural networks embedded with the induced biases have remarkable abilities in learning and generalizing the intrinsic kinetics of the underlying systems from the noisy data, such as the Hamiltonian neural networks [7, 8], the Lagrangian neural networks [5], and the physics-informed neural networks [13, 9, 6]. These methods are successfully applied to a variety of tasks, such as the generative tasks [17] and the dynamics reconstruction [20, 15], sharing the similar design idea where utilisation of an appropriate loss function enforces the model to nearly obey the physical principles. Progress has been achieved; however, these methods have poor generalization abilities, either enlarging the network complexity or over-fitting the noisy data during the training stage for decreasing the loss. To overcome these drawbacks, endowing the neural networks with natural physical priors becomes one of the mainstream approaches to improving the sample efficiency, the robustness, and the generalization ability [11, 10].

\(*\)To whom correspondence should be addressed: Q.Z. and W.L.,
https://faculty.fudan.edu.cn/wlin/zh-CN/index.htm

Recent advances in the Koopman operator theory provide a new way to identify intrinsic kinetics using linear representations for strongly nonlinear systems. There are several algorithms for approximating such an infinite-dimensional operator in a data-driven manner, including the dynamic mode decomposition (DMD) and the extending dynamic mode decomposition (EDMD). In these works, they all try to obtain parsimonious models, while maintaining an accurate reconstruction of the unknown system in a data-driven manner.

In this article, inspired by the advances of physics-informed learning and the Koopman operator theory, we aim to develop a typical framework to efficiently and robustly learn the Hamiltonian dynamics solely using the observational noisy data. Specifically, based on the fact that the Koopman operator of the Hamiltonian dynamics is unitary, we use an orthogonal neural network to learn the approximated Koopman operator that naturally keeps the conservation laws, and also use an auto-encoder to identify the nonlinear coordinate transformation, mapping the original data to a low-dimensional manifold on a high-dimensional sphere. Taking the Kepler problem as an example, we demonstrate the efficacy of the proposed framework.

2 Methodology

We consider a dynamical system whose state vector \( \mathbf{x} = (x_1, \ldots, x_n)^T \in \mathcal{M} \subset \mathbb{R}^n \) evolves along some smooth symplectic vector field \( f(x) \). Actually, the flow mappings of this autonomous dynamical system form an operator group, denoted by \( \{F_t : F_t(x) = x + \int_0^t f(x(s)) ds, x \in \mathcal{M}, t > 0 \} \). The Koopman operator \( \mathcal{K} \) with respect to the flow \( F_t \) is an infinite-dimensional linear operator, acting on the function space \( \mathcal{F} \). In practice, the data \( \{x_k \} \) often contains noise perturbations due to the nature of data measurement, which therefore motivates us to robustly approximate the Koopman operator, and preserve the energy-like quantity in the original Hamiltonian dynamics. For a sake of simplicity, we denote by \( \mathcal{K}_{\Delta t} \) for \( \mathcal{K} \) when \( \Delta t \) is given.

Connections with DMD and EDMD The DMD constructs two data matrices \( \mathbf{X} \) and \( \mathbf{X}' \) from the observational sequential data as \( \mathbf{X} = (x_0, x_2, \ldots, x_{m-1}), \mathbf{X}' = (x_1, x_2, \ldots, x_m) \), \( \mathbf{X}, \mathbf{X}' \in \mathbb{R}^{n \times m} \). Then, obtain the optimal linear operator \( \mathbf{K} = \mathbf{X}' \mathbf{X}^+ \in \mathbb{R}^{n \times n} \) for \( \mathbf{KX} \approx \mathbf{X}' \) as the approximation of \( \mathcal{K} \). Particularly, the Koopman operator for the conservative dynamical systems governed by the Hamiltonian dynamics is unitary. This property implies that we should restrict the candidate \( \mathbf{K} \) in the orthogonal group \( \text{SO}(n) \). The vanilla surrogate of the DMD has two major weaknesses:

1. The size of \( \mathbf{K} \) is limited by the dimension \( n \) of the original system, which is not sufficiently large enough to approximate the intrinsic infinite-dimensional operator \( \mathcal{K} \);
2. The orthogonal transformation preserves the norm of the state and hence induces its dynamics \( \{ \mathbf{K}^i \mathbf{x}_0 \} \) embedded on a sphere, while the conserved orbit of the Hamiltonian dynamics may not be on some \( n \)-dimensional sphere.

For the first weakness, the EDMD lifts the dimension of \( \mathbf{K} \) by introducing a dictionary of nonlinear observation functions \( \{g_i\} \) to obtain the augmented state \( \mathbf{y} = (g_1(x), g_2(x), \ldots, g_p(x))^T \in \mathbb{R}^p, p > n \). Analogous to the DMD, the two data matrices are constructed as \( \mathbf{Y} = (y_0, \ldots, y_{m-1}) \) and \( \mathbf{Y}' = (y_1, \ldots, y_m) \), yielding an approximated Koopman operator as \( \mathbf{K} = \mathbf{Y}' \mathbf{Y}^+ \in \mathbb{R}^{p \times p} \). For the second weakness, in this work, we embed the augmented state \( \mathbf{y} \) to some higher-dimensional sphere, denoted by \( S^p(r) = \{ \mathbf{y} \in \mathbb{R}^p : \|\mathbf{y}\|^2 = r^2 \} \). Then, we

\[ \mathbf{X}^+ = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \] is the Moore-Penrose pseudo-inverse.
obtain the following constrained optimization problem:

\[
\begin{align*}
\arg \min_K & \quad \left\| K Y - Y^\top \right\|_F, \quad \text{s.t. } K \in \text{SO}(p), \quad r^2 \geq \max_{0 \leq i \leq m} \| x_i \|^2, \quad (1)
\end{align*}
\]

and \( y_i = (g_1(x_i), \ldots, g_p(x_i))^\top \in S^p(r), \) \( i = 1, \ldots, m. \)

In general, it is difficult to solve the optimization problem (1) with the classical optimization methods due to the highly nonconvex property. Hence, we transform the above problem into a neural network framework to efficiently address it.

### Auto-encoder structure of dictionary functions

In the EDMD, the dictionary \( \{ g_i \}_{i=1}^p \) is regarded as a basis of \( \mathcal{F} \) and the coordinate function is reconstructed as the optimal linear combination of these functions. In practice, it always requires a large dictionary to approximate a set of basis functions. To reduce the computational cost and sustain the representation ability, we adopt an auto-encoder structure to encode these functions. In practice, it always requires a large dictionary to approximate a set of basis functions. In practice, it always requires a large dictionary to approximate a set of basis functions. In practice, it always requires a large dictionary to approximate a set of basis functions. In practice, it always requires a large dictionary to approximate a set of basis functions.

### Introduction of the neural orthogonal method

The Lie exponent map \( A \to \exp(A) := I + A + \frac{1}{2} A^2 + \cdots \) from the skew-symmetric group \( \mathfrak{so}(n) = \{ A \in \mathbb{R}^{n \times n} : A + A^\top = 0 \} \) to \( \text{SO}(n) \) is surjective \( [1] \), and there exists an isomorphism \( \alpha : \mathbb{R}^{n(n-1)/2} \to \mathfrak{so}(n) \) as \( \alpha(A) = A - A^\top \), where \( A \in \mathbb{R}^{n(n-1)/2} \) is identified as an upper triangular matrix with the zero diagonal elements. Thus, we can represent the approximative orthogonal Koopman operator as a parameterized form \( K = \exp(\alpha(A)) \), where \( A \) owns \( n(n-1)/2 \) learnable parameters. Then we train the matrix \( K \) using the loss function as

\[
\mathcal{L}_{\text{aut}}(\theta) = \sum_{i=0}^m \| x_i - \phi_{\theta_1}(x_i) \|^2.
\]

Notice that, in our framework, the orthogonality of the operator \( K \) is automatically guaranteed by its construction procedure. Indeed, this framework does not require an injection of the regularization terms into the loss function as those usually required in the literature \([7, 5]\).

### Constrain the dimension of the embedded manifold

To ensure the augmented states \( \{ y_i \}_{i=0}^m \) on some \( p \)-dimensional sphere, we set the radius \( r \) of the embedded sphere as a learnable parameter and simply set the distance to the origin as the loss function

\[
\mathcal{L}_{\text{sphere}}(\theta, r) = \sum_{i=0}^m (\| y_i \|^2 - r^2)^2 = \sum_{i=0}^m (\| \phi_{\theta_1}(x_i) \|^2 - r^2)^2.
\]

From Theorem \([3, 1]\), it follows that the trajectory generated by \( K \in \text{SO}(p) \) belongs to an at most \( |p/2| \)-dimensional manifold. Hence, the freedom degree of the augmented state trajectory \( \{ y_i \}_{i=0}^m \) should be lower than \( |p/2| \). Notice that, once the hyperplane equation \( \langle v, y \rangle = 0 \) is satisfied for any nonzero vector \( v \), the freedom degree for \( y \) decreases by one order. Thus, to restrict the freedom degree of the augmented states, we introduce the loss as follows:

\[
\mathcal{L}_{\text{deg}}(q) = \sum_{k=1}^q \sum_{i=0}^m \left( \left\langle \frac{v_k}{\| v_k \|}, y_i \right\rangle \right)^2 = \sum_{k=1}^q \sum_{i=0}^m \left( \left\langle \frac{v_k}{\| v_k \|}, \phi_{\theta_1}(x_i) \right\rangle \right)^2,
\]

where \( \{ v_k \in \mathbb{R}^p, \ k = 1, \ldots, q \} \) are learnable parameters and \( q \in \mathbb{Z}^+ \) holds \( q \leq p - 2 \). To guarantee that \( \{ v_k \}_{k=1}^q \) are linearly independent, we design an orthogonal regularization term as \( \mathcal{L}_{\text{ind}} = \sum_{k \neq j} \langle v_k, v_j \rangle^2 \). Finally, we train the parameters \( \{ \theta, K, r, V = (v_1, \ldots, v_q) \} \) using the following integrated loss function

\[
\mathcal{L}(K, \theta, r, V) = \mathcal{L}_{\text{dict}}(\theta) + \mathcal{L}_{\text{aut}}(K) + \mathcal{L}_{\text{sphere}}(\theta, r) + \mathcal{L}_{\text{deg}}(q, \theta, V) + \mathcal{L}_{\text{ind}}(V), \quad (2)
\]

where \( q \) is an adjustable hyperparameter. We use \( \mathcal{L}_{\text{sphere}} \) and \( \mathcal{L}_{\text{deg}} \) to embed the original \( n \)-dimensional data to the \( (p - q - 1) \)-dimensional manifold on \( S^p(r) \). Here we select \( q \) such that \( p - |p/2| - 1 \leq q \leq p - 2 \) with \( p - q - 1 \leq |p/2| \).
Robustness learning from noisy data The proposed framework can be directly applied to the noise perturbed case, i.e., \( \{ \tilde{x}_i = x_i + \xi_i \}_{i=0}^m \), where \( \{ \xi_i \} \) are the samples of independent random variables. Generally, the existing models, such as the DMD, the EDMD and the other learning-based methods, likely to overfit the noise perturbed data due to a lack of prior knowledge. For example, it may break the conservative law in the Hamiltonian dynamics as the noise perturbed data act as an effect of a diffusion dynamics. On the contrary, our framework can identify the noisy samples effectively, because the conservation is naturally rooted in the orthogonal transform. We numerically validate the natural conservation in Appendix A.2.2.

3 Theoretical Results

In this section, we focus on the lowest dimension of manifold \( \mathcal{M} \subset \mathbb{R}^n \) that covers the trajectory \( \Gamma = \{ x, Ax, A^2x, \cdots \} \) generated by the orthogonal transform \( A \in SO(n) \). Although the orthogonal transform restricts the trajectory \( \Gamma \) to some \( n \)-dimensional sphere, this constraint is too rough to describe the manifold \( \mathcal{M} \). In particular, \( \Gamma \) may have degrees of freedom lower than \( n - 1 \). The following theorem precisely provides the admitted dimension for \( \mathcal{M} \). The proof is presented in Appendix A.1.1.

Theorem 3.1 For each \( A \in SO(n) \), if the orthogonal normal form of \( A \) has \( s \) diagonal blocks, denoted by \( A_i, i = 1, \ldots, s \) with \( A_i = \pm1, \{(a_i, -b_i)^\top, (b_i, a_i)^\top\} \), and \( b_i \neq 0 \). Then, for any initial state \( x \), the trajectory \( \Gamma = \{ A^jx \}_{j=0}^\infty \) belongs to an \( (n-s) \)-dimensional manifold. Moreover, \( \Gamma \) belongs to an at most \( \lceil \frac{n}{2} \rceil \)-dimensional manifold, where \( \lfloor \cdot \rfloor \) represents the floor function.

Using the result in Theorem 3.1, we can determine all the admitted dimensions for \( \mathcal{M} \) as \( 1, 2, \cdots, \lfloor n/2 \rfloor \) according to the number of diagonal blocks in \( A \)’s Jordan normal form. This theorem suggests that the degrees of freedom in the embedding procedure should be restricted such that the augmented states are distributed on some at most \( \lfloor p/2 \rfloor \)-dimensional manifold.

4 Experiments

We consider the planar Kepler problem with the Hamiltonian quantity \( H = \frac{m}{2}(\dot{x}^2 + \dot{y}^2) - \frac{gm^2}{\sqrt{x^2+y^2}} \), which is a special case for the classical \( n \)-body Hamiltonian system with \( n = 1 \) [2]. Here, \( (x, y) \) and \( (\dot{x}, \dot{y}) \) represent the coordinate and the velocity, respectively. Here, we set \( m = g = 1 \) and sample the trajectory of this 4-D equation with the odeint package [4] and add the Gaussian noise \( \mathcal{N}(0, 0.03) \) to each data point. We compare our HNKO with the currently mainstream methods, the HNN [7] and the EDMD [19], in terms of the trajectory prediction and the energy conservation. As shown in Figure 1, our HNKO outperforms the other methods in terms of robustness and accuracy.

Figure 1: Results of Kepler problem. The first row and the second row show the qualitative comparison of trajectories and different kinds of energy, respectively. The right column displays the mean square error (MSE) for trajectories and energy prediction.
5 Conclusion

We propose a learning-based framework to approximate the Koopman operator for the Hamiltonian dynamics. We provide theoretical analysis for the dimension of manifold induced by the orthogonal transform and using an auto-encoder to encode the original data to some low-dimensional manifold restricted in some high-dimensional sphere, then decode it back to the original data space. We show the proposed framework outperforms the existing methods through the numerical comparison on the Kepler problem.

6 Acknowledgments

We thank the anonymous reviewers for their valuable and constructive comments that helped us to improve the work. Q.Z. is supported by the Shanghai Postdoctoral Excellence Program (No. 2021091) and by the STCSM (Nos. 21511100200 and 22ZR1407300). W.L. is supported by the National Natural Science Foundation of China (No. 19511101404, 22JC1402500, 22JC1401402, and 2021SHZDZX0103).

References


A Appendix

A.1 Proofs

A.1.1 Proof of Theorem 3.1

First, denote the orthogonal normal form of $A$ as $J\text{diag}(A_1, \cdots, A_s)$, and there exists transform matrix $P \in SO(n)$ such that $A = P^TJP$. Then the transform $P$ induce an isomorphism between $\Gamma = \{A^jx\}_{j=0}^{\infty}$ and $\tilde{\Gamma} = \{J^jy\}_{j=0}^{\infty}$, where $y = Px$. Hence, we only need to identify the degree of freedom of $\tilde{\Gamma}$.

Next, split the state $y$ as $y = (y_1^T, \cdots, y_s^T)^T$ such that the number of variables of $y_s$ is the same as the order of $A_s$. Then we have $Jy = ((A_1y_1)^T, \cdots, (A_sy_s)^T)$. Since $A_i, i = 1, \cdots, s$ are also orthogonal matrix, then preserve the norm of the corresponding vector, i.e.

$$\|A_iy_i\| = \|y_i\|, \quad A_i \in SO(2), \quad \text{or} \quad |A_iy_i| = |y_i|, \quad A_i \in SO(1).$$

Each of the above equality constraint will remove one degree from $\tilde{\Gamma}$. Then these $s$ equations will restrict $\tilde{\Gamma}$ on some $(n - s)$-dimensional manifold, which completes the proof.

A.2 Experimental Configurations

**Numerical Implementation** For parameters learning of the constructed orthogonal Koopman operator $K$, we use the GeoTorch package for constraints on the manifold [10]. All the necessary models can be trained within several minutes on a computing device with a single i7-10870 CPU with 16GB memory.

A.2.1 Kepler Problem

Mathematically, the kepler problem is written as

$$\ddot{x} = g m^2 \frac{x}{(x^2 + y^2)^{3/2}},$$

$$\ddot{y} = g m^2 \frac{y}{(x^2 + y^2)^{3/2}}.$$  

(3)
We can see that HNKO learn the exact transform $\text{HNKO}$ learn the exact transform
We demonstrate the above analysis in Figure 2. To guarantee the predictive stability of HNKO, we
We compare the vanilla version of our HNKO and the classic DMD method in an planar rotation
We uniformly sample $50$ Tanh with $4$ functions, and this dictionary corresponds to a
The corresponding physical quantities are

\[
\begin{align*}
\text{Kinetic energy: } & E_k = \frac{m}{2}(x^2 + y^2), \\
\text{Potential energy: } & E_p = -gm\frac{1}{\sqrt{x^2 + y^2}}, \\
\text{Total energy: } & E = E_k + E_p
\end{align*}
\]

(4)

We uniformly sample 50 observables $\{x_i\}_{i=1}^{49}$ from the initial point $(1.0, 0.0, 0.0, 0.9)^T$ on time interval $[0, 5]$, then we add Gaussian noise $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 0.03$ to the samples.

In the learning procedure, we use a $4 \times 40 \times 40 \times 40 \times 13$ fully connected neural network (FNN) with Tanh activation to encode the original 4-D data to 13-D sphere and use a $13 \times 40 \times 40 \times 4$ FNN as decoder. The Koopman operator $K$ is constructed as a linear layer in FNN without bias under constraint $\text{geotorch.orthogonal}$. We train the parameters simultaneously with the loss function $\mathcal{L}^2$.

In the predicting procedure, we select any point $x$ on the original continuous trajectory as initial point and encode it to $y = \phi(x)$, then we use $K$ to generate the high-dimensional predictive trajectory $\{K^jy\}_{j=0}^3$. Finally, we use decoder to obtain the predictive trajectory in original space $\{\hat{x}_t = \phi^{-1}(K^jy)\}_{j=0}^3$. Specifically, we choose $x = x_0$ and $l = 50$ to test the performance of long term prediction of learned $K$.

For HNN, we use $4 \times 100 \times 100 \times 1$ FNN with Tanh activation and without the last layer bias as proposed by [7]. For EDMD, we use Hermite polynomial functions up to order 3 as dictionary functions, and this dictionary corresponds to a 256 order Koopman matrix, which is far greater than our HNKO. The predictive procedure is the same as above.

### A.2.2 Planar Rotation

We compare the vanilla version of our HNKO and the classic DMD method in an planar rotation model to demonstrate the robustness of our HNKO.

First, we define a rotation matrix as $A = ((\cos(a), \sin(a))^T, (-\sin(a), \cos(a))^T)$. Then we obtain $m$ observables as $\{A^jx\}_{j=0}^{m-1}$. We add Gaussian noise $\mathcal{N}(0, \sigma^2)$ to this dataset. Now we test the performance of HNKO and DMD. We select $a = 0.2$, $m = 99$, $\sigma = 0.0$, $0.1$. Notice that $A$ itself is the unique solution of the following optimal problem

\[
\begin{align*}
\underset{K}{\arg\min} & \quad ||KX - X'||, \\
X &= (x, Ax, \cdots, A^{m-1}x), \\
X' &= (Ax, A^2x, \cdots, A^mx).
\end{align*}
\]

(5)

Hence DMD can exactly learn the dynamic $A$ in noise-free case. However, DMD solution $K^+ = X'X^+$ generally cannot equal to $A$ in the noisy case because $A$ cannot minimize the objective function of the following optimal problem

\[
\begin{align*}
\underset{K}{\arg\min} & \quad ||K\hat{X} - \hat{X}'||, \\
\hat{X} &= (x + \xi_0, Ax + \xi_1, \cdots, A^{m-1}x + \xi_{m-1}), \\
\hat{X}' &= (Ax + \xi_1, A^2x + \xi_2, \cdots, A^mx + \xi_m), \\
\xi_i &\sim \mathcal{N}(0, \sigma^2), \quad i = 1, \cdots, m.
\end{align*}
\]

(6)

We demonstrate the above analysis in Figure.2 To guarantee the predictive stability of HNKO, we modify $\mathcal{L}_{\text{koop}}$ as

\[
\mathcal{L}_{\text{koop}} = \sum_{i=1}^{m-2} ||K^iy_i - y_{i+1}\|^2 + \sum_{i=1}^{m-2} ||K^iy_i - y_{i+2}\|^2.
\]

We can see that HNKO learn the exact transform $A$ in both noisy and noise-free case, while DMD fail to learn the $A$ in the noisy case due to overfit on the noisy data.
Figure 2: Results of planar rotation. The prediction performance of HNKO and DMD in noise-free case (left) and noisy case (right).