Learning Hamiltonian neural Koopman operator and simultaneously sustaining and discovering conservation laws

Jingdong Zhang [1,2] Qunxi Zhu [2,3,4,*] and Wei Lin [1,2,3,4,†]

[1]*School of Mathematical Sciences, SCMS, and SCAM, Fudan University, Shanghai 200433, China*

[2]*Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China*

[3]*Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China*

[4]*MOE Frontiers Center for Brain Science and State Key Laboratory of Medical Neurobiology,
Fudan University, Shanghai 200032, China*

Accurately finding and predicting dynamics based on the observational data with noise perturbations is of paramount significance but still a major challenge presently. Here, for the Hamiltonian mechanics, we propose the Hamiltonian neural Koopman operator (HNKO), integrating the knowledge of mathematical physics in learning the Koopman operator, and making it automatically sustain and even discover the conservation laws. We demonstrate the outperformance of the HNKO and its extension using a number of representative physical systems even with hundreds or thousands of freedoms. 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 machine learning in solving physical problems.

Introduction. Accurate reconstruction of nonlinear dynamical systems solely based on the observational data with noise perturbations is a focal challenge in many fields of physics and engineering. The neural networks (NNs) equipped 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 NNs [1,2], the Lagrangian NNs [3], the neural differential equations [4–6], the physics-informed NNs [7–15], and the reservoir computing [16–18]. These frameworks have been applied successfully to many tasks (e.g., the generative tasks [19], the dynamics reconstruction [20–23], the intelligent control problems [24,25], and the tipping point detection [26,27]), sharing the common idea in design utilization of an appropriate loss function enforcing the model to nearly obey the physical principles. Although progresses have been outstandingly achieved, these frameworks, which either enlarge the network complexity or overfit the noisy data during the training stage to decrease the loss, are suffered from the poor generalization abilities. Recently, endowing NNs with natural physics priors becomes an effective approach to promote the sample efficiency, robustness, and generalization ability of NNs [28–31].

Recent advances in the Koopman operator theory pave a direct way to identify intrinsic kinetics using infinite-dimensional linear representations for strongly nonlinear systems [32–37]. Several algorithms using observational data have been developed for approximating such an operator, including the dynamic mode decomposition (DMD) [34,38,39] and the extending dynamic mode decomposition (EDMD) [40,41]. Although all these algorithms try to obtain parsimonious models and maintain accurate reconstructions of the unknown systems, either conservation properties or the accurate prediction cannot be obtained surely.

In this Letter, inspired by the advances of physics-informed learning and the Koopman operator theory, we articulate a framework to efficiently and robustly learn the Hamiltonian dynamics based solely on the observational data even with noise perturbations. Noticing the unitary property of the Koopman operator for the Hamiltonian dynamics, we use an orthogonal neural network to approximate the Koopman operator, so that the learned operator naturally sustains the conservation laws. Also, we include an autoencoder structure in the NNs to identify the nonlinear coordinate transformation, mapping the original data to a low-dimensional manifold on a sphere. We test the proposed HNKO framework on a group of representative Hamiltonian systems and show its advantages over many mainstream methods in several aspects including robust preservation of the conservation laws and accurate prediction of the dynamical behaviors.

Method. To begin with, we consider a dynamical system whose state vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathcal{M} \subset \mathbb{R}^n$ evolves along some smooth symplectic vector field $f(\boldsymbol{x})$. The flow mappings produced by this Hamiltonian dynamical system form an operator group, denoted by $\{F_t : F_t(\boldsymbol{x}) = \boldsymbol{x} + \int_0^t f[\boldsymbol{x}(s)]\mathrm{d}s, \boldsymbol{x} \in \mathcal{M}, t > 0\}$. The Koopman operator $\mathcal{K}_t$ with regard to any flow F_t is an infinite-dimensional linear operator, acting on the function space $\mathcal{F} = \{g : \mathcal{M} \to \mathbb{R}\}$ and satisfying $\mathcal{K}_t g = g \circ F_t$ for $g \in \mathcal{F}$. Specifically, if the time interval

*qxzhu16@fudan.edu.cn

†wlin@fudan.edu.cn

Δt and the initial state $\boldsymbol{x}_0 \in \mathcal{M}$ are given, a state trajectory is generated by this flow, denoted as $\{\boldsymbol{x}_k : \boldsymbol{x}_k = F_{k\Delta t}\boldsymbol{x}_0\}_{k=0}^m$. Thus, $\mathcal{K}_{\Delta t}g(\boldsymbol{x}_k) = g[F_{\Delta t}(\boldsymbol{x}_k)] = g(\boldsymbol{x}_{k+1})$. For the sake of simplicity, we use $\mathcal{K}$ for $\mathcal{K}_{\Delta t}$ if Δt is given. Since, in practice, the observational data $\{\boldsymbol{x}_i\}_{i=0}^m$ often contain noise perturbations, it is really difficult for the existing methods to achieve a robust and accurate approximation of $\mathcal{K}$ and simultaneously preserve the energylike quantity in the considered Hamiltonian dynamics. It thus motivates us to find a framework owning all these capabilities.

The DMD algorithm [34,38,39] constructs two data matrices $\boldsymbol{X}$ and $\boldsymbol{X}'$ from the observational data by $\boldsymbol{X} = (\boldsymbol{x}_0, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{m-1})$ and $\boldsymbol{X}' = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$ for $\boldsymbol{X}, \boldsymbol{X}' \in \mathbb{R}^{n \times m}$. Then, according to Ref. [38], the optimal linear operator $\boldsymbol{K} = \boldsymbol{X}'\boldsymbol{X}^+ \in \mathbb{R}^{n \times n}$ satisfying $\boldsymbol{K}\boldsymbol{X} \approx \boldsymbol{X}'$ is regarded as an approximation for $\mathcal{K}$, where $\boldsymbol{X}^+ = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ is the Moore-Penrose pseudoinverse. Particularly, since the Koopman operator for the conservative Hamiltonian dynamics is unitary, we should restrict the candidate $\boldsymbol{K}$ in the special orthogonal group $\mathbf{SO}(n) = \{\boldsymbol{B} \in \mathbb{R}^{n \times n} \mid \boldsymbol{B}\boldsymbol{B}^\top = \boldsymbol{I}, \det(\boldsymbol{B}) = 1\}$. Then, the DMD actually solves the vanilla optimal problem: $\arg\min_{\boldsymbol{K} \in \mathbf{O}(n)} \|\boldsymbol{K}\boldsymbol{X} - \boldsymbol{X}'\|_{\mathrm{F}}$, where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. The vanilla surrogate of the DMD has two major weaknesses: (i) the size of $\boldsymbol{K}$ is limited by the system's dimension n, which is not large enough to approximate the intrinsically infinite-dimensional operator $\mathcal{K}$, and (ii) the orthogonal transformation preserves the norm of the state and hence induces its dynamics $\{\boldsymbol{K}^i\boldsymbol{x}_0\}_{i=0}^m$ embedded on a sphere, while the conserved orbit of the original Hamiltonian dynamics may not be on some n-dimensional sphere.

To overcome the first weakness, the EDMD was developed in Refs. [40,41] to lift the dimension of $\boldsymbol{K}$ by introducing a dictionary of nonlinear observational functions $\{g_i\}_{i=1}^p$ and obtaining the augmented state $\boldsymbol{y} = [g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_p(\boldsymbol{x})]^\top \in \mathbb{R}^p$, $p > n$. Analogous to the DMD, the two data matrices are constructed as $\boldsymbol{Y} = (\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{m-1})$ and $\boldsymbol{Y}' = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$, which gives an approximated Koopman operator as $\boldsymbol{K} = \boldsymbol{Y}'\boldsymbol{Y}^+ \in \mathbb{R}^{p \times p}$. However, the EDMD uses the dictionary $\{g_i\}_{i=1}^p$ as a basis of $\mathcal{F}$ and, in practice, it requires an extremely large dictionary to approximate the coordinate function, which makes the accurate approximation inefficient and even impossible. To reduce the computational cost and promote the representation ability, we adopt an autoencoder NN to encode $\boldsymbol{y} = [g_1(\boldsymbol{x}), \ldots, g_p(\boldsymbol{x})]^\top$ as $\boldsymbol{y} = \phi_{\boldsymbol{\theta}_1}(\boldsymbol{x})$, and decode the coordinates as $\boldsymbol{x} = \phi_{\boldsymbol{\theta}_2}^{-1}(\boldsymbol{y})$, as shown in Fig. 1(a). We train the weights $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in this autoencoder using the loss function as $\mathcal{L}_{\mathrm{dict}}(\boldsymbol{\theta}) = \sum_{i=0}^m \|\boldsymbol{x}_i - \phi_{\boldsymbol{\theta}_2}^{-1}[\phi_{\boldsymbol{\theta}_1}(\boldsymbol{x}_i)]\|^2$ with the data.

To conquer the second weakness, we in this work embed the augmented state $\boldsymbol{y}$ to some higher-dimensional sphere, denoted by $S^p(r) = \{\boldsymbol{y} \in \mathbb{R}^p \mid \|\boldsymbol{y}\|^2 = r^2\}$, and obtain a constrained optimization problem as $\arg\min_{\boldsymbol{K}} \|\boldsymbol{K}\boldsymbol{Y} - \boldsymbol{Y}'\|_{\mathrm{F}}$ such that $\boldsymbol{K} \in \mathbf{SO}(p)$, $r^2 \geqslant \max_{0 \leqslant i \leqslant m} \|\boldsymbol{x}_i\|^2$, and $\boldsymbol{y}_i = \phi_{\boldsymbol{\theta}_1}(\boldsymbol{x}_i) \in S^p(r)$, $0 \leqslant i \leqslant m$. Actually, it is still difficult to solve this optimization problem because of its highly nonconvex property induced by the orthogonality constraint.

To solve this problem efficiently, we first notice the fact that the Lie exponent map $\boldsymbol{A} \mapsto \exp(\boldsymbol{A}) = \boldsymbol{I} + \boldsymbol{A} + \frac{1}{2}\boldsymbol{A}^2 + \ldots$ from the skew-symmetric group $\mathfrak{so}(n) = \{\boldsymbol{A} \in \mathbb{R}^{n \times n} : \boldsymbol{A} + \boldsymbol{A}^\top = \boldsymbol{0}\}$ to $\mathbf{SO}(n)$ is surjective [30]. There exists an
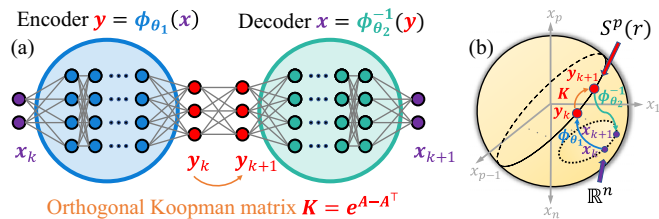

FIG. 1. A sketch for the HNKO framework. (a) A combination of the autoencoder with the orthogonal Koopman matrix $\boldsymbol{K}$ using NNs. (b) Geometrically, the encoder embeds the original data in $\mathbb{R}^n$ to some low-dimensional manifold on p-dimensional sphere, and the decoder reverses this process, where $\boldsymbol{K}$ maps the trajectory on the embedded manifold.

isomorphism α from $\mathbb{R}^{n(n-1)/2}$ to $\mathfrak{so}(n)$ as $\alpha(\boldsymbol{A}) = \boldsymbol{A} - \boldsymbol{A}^\top$ where $\boldsymbol{A} \in \mathbb{R}^{n(n-1)/2}$ is identified as an upper triangular matrix with the zero diagonal elements. Thus, in our framework, we represent the orthogonal Koopman operator approximately as a parameterized form $\boldsymbol{K} = \exp[\alpha(\boldsymbol{A})]$, where $\boldsymbol{A}$ owns $n(n-1)/2$ learnable parameters. Hence, we train $\boldsymbol{K}$ using the loss function as $\mathcal{L}_{\mathrm{koop}}(\boldsymbol{K}) = \sum_{i=0}^{m-1} \|\boldsymbol{K}\boldsymbol{y}_i - \boldsymbol{y}_{i+1}\|^2 = \sum_{i=0}^{m-1} \|\exp[\alpha(\boldsymbol{A})]\phi_{\boldsymbol{\theta}_1}(\boldsymbol{x}_i) - \phi_{\boldsymbol{\theta}_1}(\boldsymbol{x}_{i+1})\|^2$. Significantly, our framework does not require this uninterpretable regularization term in the loss function while all the conventional methods always use it [1,3], and, indeed, the above configurations of mathematical physics automatically guarantee the orthogonality of the operator $\boldsymbol{K}$ during its training procedure.

To further ensure the augmented state trajectory $\{\boldsymbol{y}_i\}_{i=0}^m$ on some p-dimensional sphere, we set r, the radius of the embedded sphere, as a learnable parameter and simply set the distance to the origin as another loss function $\mathcal{L}_{\mathrm{sphere}}(\boldsymbol{\theta}, r) = \sum_{i=0}^m (\|\boldsymbol{y}_i\|^2 - r^2)^2 = \sum_{i=0}^m (\|\phi_{\boldsymbol{\theta}_1}(\boldsymbol{x}_i)\|^2 - r^2)^2$. We prove that the trajectory generated by $\boldsymbol{K} \in \mathbf{SO}(p)$ belongs to a manifold of dimension at most $\lfloor p/2 \rfloor$ (see Supplemental Material (SM) [42]). Hence, the freedom degree of the trajectory $\{\boldsymbol{y}_i\}_{i=0}^m$ is lower than $\lfloor p/2 \rfloor$. Notice that once the hyperplane equation $\langle \boldsymbol{v}, \boldsymbol{y} \rangle = 0$ is satisfied for any nonzero vector $\boldsymbol{v}$, the freedom degree for $\boldsymbol{y}$ decreases by one order. Thus, to restrict the freedom degree of the augmented states, we introduce a loss as: $\mathcal{L}_{\mathrm{deg}}(q) = \sum_{k=1}^q \sum_{i=0}^m \langle \frac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|}, \boldsymbol{y}_i \rangle^2 = \sum_{k=1}^q \sum_{i=0}^m \langle \frac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|}, \phi_{\boldsymbol{\theta}_1}(\boldsymbol{x}_i) \rangle^2$, where $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_q)$ are learnable parameters with $\boldsymbol{v}_k \in \mathbb{R}^p$, $k = 1, 2, \ldots, q \leqslant p - 2$, and $q \in \mathbb{Z}^+$. To guarantee the linear independence of the column vectors in $\boldsymbol{V}$, we introduce an orthogonal regularization term as $\mathcal{L}_{\mathrm{ind}} = \sum_{k \neq j} \langle \boldsymbol{v}_k, \boldsymbol{v}_j \rangle^2$.

Finally, we train the parameters $\{\boldsymbol{\theta}, \boldsymbol{K}, r, \boldsymbol{V}\}$ in a delicately designed framework of NNs, integrating the prior knowledge of mathematical physics and using a comprehensive loss function as: $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{K}, r, \boldsymbol{V}) = \mathcal{L}_{\mathrm{dict}}(\boldsymbol{\theta}) + \mathcal{L}_{\mathrm{koop}}(\boldsymbol{K}) + \mathcal{L}_{\mathrm{sphere}}(\boldsymbol{\theta}, r) + \mathcal{L}_{\mathrm{deg}}(q, \boldsymbol{\theta}, \boldsymbol{V}) + \mathcal{L}_{\mathrm{ind}}(\boldsymbol{V})$, where $\mathcal{L}_{\mathrm{sphere}}$ with $\mathcal{L}_{\mathrm{deg}}$ embeds the original n-dimensional data to the $(p - q - 1)$-dimensional manifold on $S^p(r)$, and the adjustable hyperparameter q satisfies $p - \lfloor p/2 \rfloor - 1 \leqslant q \leqslant p - 2$. Significantly, the extracted features $\{g_i(\boldsymbol{x})\}_{i=1}^p$ from the encoder span the Koopman invariant subspace. The eigenvector $\boldsymbol{c} = (c_1, \ldots, c_p)^\top$ associated with the eigenvalue 1 of the orthogonal Koopman matirx $\boldsymbol{K}$

induces the analytical Hamiltonian of the original system, $g_c(\boldsymbol{x}) = \sum_{i=1}^p c_i g_i(\boldsymbol{x})$ (see SM [42]).

Scalability for high-dimensional systems. To reduce the computational complexity of the Lie exponent operation when applying the current HNKO to any high-dimensional system, we approximate the p-dimensional Koopman matrix $\boldsymbol{K}$ via $\boldsymbol{K}_1 \otimes \boldsymbol{K}_2$, where $(\boldsymbol{K}_i)_{p_i \times p_i}$ $(i = 1, 2)$ are the two orthogonal matrices with $p_1 p_2 = p$, implying the orthogonality of $\boldsymbol{K}$ [43], and $\otimes$ is the Kronecker product. We name this extension as HNKO*. As such, by applying the HNKO* with $p_{1,2} = \sqrt{p}$, we reduce the computational cost of the learnable parameters from $\mathcal{O}(p^2)$ to $\mathcal{O}(p)$ (see Appendix for more illustrations).

Next, we numerically show several advantages of our framework, and validate the natural existence of the conservation in our operator. This makes the framework extremely suitable for dealing with the noise-perturbed data produced by the Hamiltonian dynamics. Indeed, we take a number of representative nonlinear Hamiltonian systems of physical significance, including the many-body problem, the stiff-spring oscillator, and the Korteweg-de Vries (KdV) equation. Throughout, the noise-perturbed data are set as $\{\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i + \boldsymbol{\xi}_i\}_{i=0}^m$, where the information of the used noise $\{\boldsymbol{\xi}_i\}$ is provided in SM [42].

Many-body problem. We consider the classic n-body problem. First, we focus on the case of $n = 3$, where the canonical Hamiltonian dynamics [44]: $\dot{\boldsymbol{q}}_i = H_{\boldsymbol{p}_i}$, $\dot{\boldsymbol{p}}_i = -H_{\boldsymbol{q}_i}$ with $\boldsymbol{p}_i, \boldsymbol{q}_i \in \mathbb{R}^2$ $(i = 1, 2, 3)$ representing the space coordinates and the momenta, respectively. Here, $H = \sum_{i=1}^3 \frac{m_i}{2}\|\boldsymbol{p}_i\|_2^2 - g\sum_{i<j} m_i m_j (\|\boldsymbol{q}_i - \boldsymbol{q}_j\|_2)^{-1}$ is the total conserved energy with mass m_i and gravitational constant g. The three bodies interact with the others through an attractive force from gravity, and the force tends to infinity when the two particles get close to each other. As shown in Fig. 2(a), the orbits of the three bodies evolve in a higher-dimensional spatiotemporal space, which results in a challenge for accurate prediction based on the noise-perturbed observational data.

We show the prediction performance using the HNKO framework. Particularly, we train the NNs with the noise-perturbed data on the time interval [0,5], less than a period, and produce the trajectories using the trained model on the interval [0,50]. As shown in Fig. 2(b), the reconstructed and the predicted dynamics are preserved in high fidelity and for a fairly long time.[1] We also test the methods of the EDMD with a dictionary of at most second-order Hermite polynomials [40], the HNN [1], the SympNets [45], the deep-learning Koopman operator (DLKO) [35], the CNN-LSTM [46,47], the Hamiltonian ODE graph networks (HOGN) [21],

---

[1]The predicted trajectories by the HNKO framework seem to be a bit jagged, which reflects a peculiar ability of our framework in sustaining the feature of noise perturbations. In fact, using the technique of embedding the original data to the manifold on $S^p(r)$ not only extracts maximally the conservation law rooted in the original data, but also withstands the noise effect efficiently. All these indicate the necessity and importance of integrating the priors of mathematical physics in reconstructing and predicting a system.
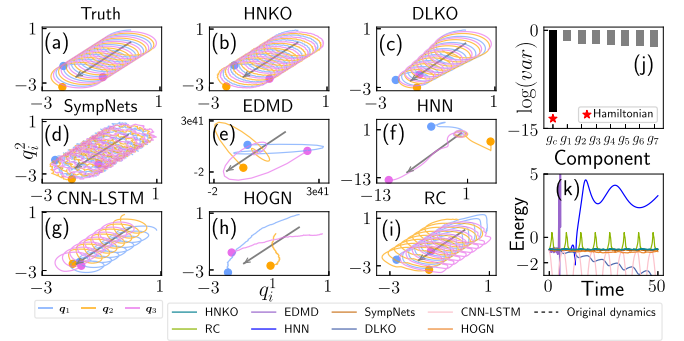

FIG. 2. Comparison studies on the three-body problem. (a) The original and noise-free dynamics for the interacted bodies. Here, the motion $\boldsymbol{q}_i = (q_i^1, q_i^2)$, and the trajectories are the projections from the original spatiotemporal space $q_i^1 - q_i^2 - t$ to the phase plane with a normal vector as $[\sin(-\frac{\pi}{50})\cos\frac{\pi}{4}, \sin(-\frac{\pi}{50})\sin\frac{\pi}{4}, \cos(-\frac{\pi}{50})]$. The gray direct line indicates the time direction and the terminal positions of the three bodies are highlighted by blue, purple, and orange colors, respectively. The reconstructed and the predicted dynamics using the HNKO (b) and the other the most advanced machine learning techniques (c)–(i) are shown, respectively. (j) The temporal variance in the logarithm scale [$\log(var)$] of the features and the discovered Hamiltonian. (k) The system's energies using different methods change over the time. Here, we set $m_{1,2,3} = g = 1$.

and the reservoir computing (RC) [48,49] on the same generated noise-perturbed data. The produced trajectories [see Figs. 2(c)–2(i)]either cannot sustain the conservation laws or diverge extremely fast. Figure 2(j) displays a successful discovery of the Hamiltonian using the HNKO. Additional comparisons in Fig. 2(k) suggest that the HNKO framework attains the least prediction errors and only has the ability of maintaining the conservation law. In SM [42], we further show successful examples using our framework when the length of the training data within a period is even shorter, and also show the reconstruction ability of the HNKO in dealing with the chaotic dynamics of the three-body system.

Next, we test our framework for a specific case: $n = 1$, the classic Kepler problem as: $\dot{\boldsymbol{q}} = H_{\boldsymbol{p}}$, $\dot{\boldsymbol{p}} = H_{\boldsymbol{q}}$ with $H = \frac{m}{2}\|\boldsymbol{p}\|_2^2 - gm^2\|\boldsymbol{q}\|_2^{-1}$ [50]. We also compare this framework
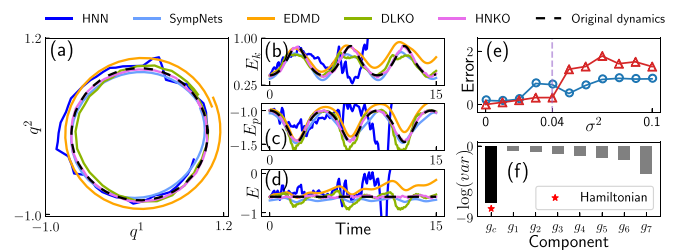

FIG. 3. Comparison studies on the Kepler problem. (a) The original dynamics and the predicted phase orbits using different methods and the coordinate $\boldsymbol{q} = (q^1, q^2)$. The changes of the kinetic energy E_k (b), the potential energy E_p (c), and the total energy E (d) over the time for different methods. (e) Prediction errors of the state and the energy change with the noise variance σ^2, using HNKO. Here, $m = g = 1$. (f) $\log(var)$ of the features and the discovered Hamiltonian.
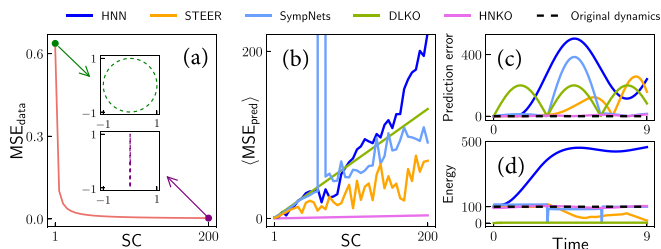
FIG. 4. Comparison studies on the stiff mass-spring system. (a) The mean-square error ($\mathrm{MSE}_{\mathrm{data}}$) between the normalized trajectories and the vertical line segment under different SCs. The subfigures show the trajectories in the least and the stiffest cases. (b) The mean prediction MSE on time interval [0,9] of different methods over SCs. The prediction error (c) and the energy (d) over the time for different methods.
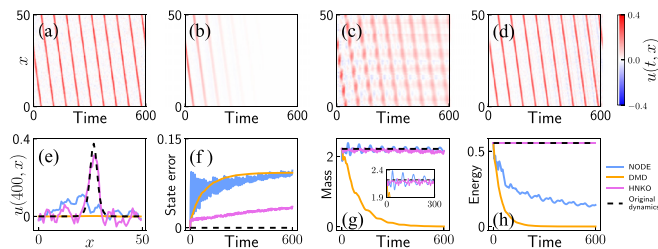

FIG. 5. Comparison studies on the KdV equation. (a) The dynamics produced by the KdV equation and perturbed with the Gaussian noise $\mathcal{N}(0, 0.03\boldsymbol{I})$. The predicted trajectories using the DMD (b), the NODE (c), and the HNKO (d). Different solitary waves produced by using different methods at $t = 400$ (e), the prediction error for the state (f), the total mass (g), and the total energy (h) over the time using different methods. The inset in (g) zoom in the changes of mass using the NODE and the HNKO. Here, we introduce the discretization to $\boldsymbol{u}(t, \cdot)$ on [0,50] by 64 predefined grid points.

with the other methods. As shown in Figs. 3(a)–3(d), the HNKO framework obtains the best prediction result and maintains the conservative law perfectly, while the corresponding results obtained by the EDMD and the HNN are far away from the original dynamics and energies. Notice that, for this case, the results by the EDMD seem to be better than those in Fig. 2(c) obtained for the above case $n = 3$. This suggests that the EDMD method using a second-order dictionary could provide acceptable prediction results for complex systems of lower dimensions. However, due to the curse of dimensionality, it cannot directly use a dictionary of higher-order polynomials to deal with the data produced from even a three-body problem, which thus impacts its practical usefulness. In SM [42], we further show the efficacy of the HNKO framework in reconstruction and prediction for the chaotic three-body case and for the n-body problem with $n \gg 3$ as well. Particularly, to show the advantages of the HNKO over the existing methods, we reconstruct the periodic solution, previously-found in Ref. [51], of the 24-body problem in the three-dimensional space (see Appendix, Table I). In addition, the robustness of the HNKO against the noise perturbation is demonstrated from the view of state and energy predictions, as shown in Fig. 3(e). Significantly, Fig. 3(f) shows a compelling and successful identification of the Hamiltonian using the HNKO.

Stiff mass-spring system. We consider the friction-free mass-spring system: $\dot{q} = p/m$, $\dot{p} = -kq$, where $(q, p) \in \mathbb{R}^2$ are the canonical coordinates representing the position and the momentum, m is the mass, and k is the elastic coefficient [52]. With large k and m, the system becomes a typical slow-fast system with the stiffness coefficient (SC) as $\sqrt{km}$ [53]. The system's conservative total energy $\frac{1}{2}kq^2 + \frac{1}{2m}p^2$ corresponds to the elliptic phase orbits with an eccentricity going to 1 as the SC goes to infinity. In Fig. 4(a), under a high SC, the trajectory looks like noise series wandering along the line segment, which leads to a failure of reconstruction and prediction using the ODE solver-based learners [53].

We compare the prediction performance of HNKO under different SCs, with existing methods including the STEER, a better method for solving stiff ODEs problems [54]. Figures 4(b)–4(d) show the HNKO alleviates the effects of

stiffness while the other methods perform unstably and diverge fast as SC grows.

KDV equation. We finally consider the KdV equation $\boldsymbol{u}_t + \boldsymbol{u}_{xxx} - 6\boldsymbol{u}\boldsymbol{u}_x = 0$, a Hamiltonian partial differential equation with infinitely many integrals of motion (IOM) [55]. This equation is used to describe the behavior of shallow water waves with periodic solitary wave phenomenon [56], as shown in Fig. 5(a). Two low-order IOMs of the system—$\int \boldsymbol{u}\mathrm{d}x$ and $\int \boldsymbol{u}^2\mathrm{d}x$, corresponding to the mass and the energy conservation—are selected as the indexes to evaluate the prediction performance.

We conduct comparison studies using respective methods: the HNKO, the DMD, and the neural ordinary differential equation (NODE), a recently developed and widely used framework [4]. The HNN and EDMD cannot work here.[2] As shown in Figs. 5(a)–5(d), the HNKO robustly keeps the periodicity and the consistency of solitary wave solutions for a long time, while the DMD only holds a short-term forecast, showing a rapid decay and the trained NODE shows high fluctuations due to its less robustness against noise interference. After a sufficiently long time evolution, the required solitary wave behavior is only observed in the trained model using the HNKO, as shown in Fig. 5(e). More significantly, due to the orthogonality rooted in the HNKO, the conservation laws of both mass and energy are sustained only in the trained model using the HNKO [see Figs. 5(f)–5(h)]. In Appendix, Table II, we further successfully apply the HNKO* to cope effectively with the high-dimensional tasks with the number of the grid points even as 1024.

Concluding remarks. We have proposed a machine learning framework to approximate the Koopman operator for the Hamiltonian dynamics. Different from the mainstream NN

---

[2]Since the KdV equation cannot be formulated in a canonical form with space coordinates and momenta, the symplectic structure-based methods including the HNN are not applicable to prediction task. The EDMD is not applicable because n^{64}, the size of the nth-order Hermite polynomials dictionary, incurs unbearable computational cost.

TABLE I. Comparison studies of different methods on the 24-body problem using full (144-dimensional) and partial (72-dimensional) observational data. Here, the performance (Perf.) is shown in terms of the 2-Wasserstein distance (2WD) and the time average of the MSE, denoted by $\langle$MSE$\rangle$, between the real data and the prediction data.

Data	Perf. $\downarrow$	Model				
		HNKO	DLKO	DMD	SympNets	HNN
Full	$\langle$MSE$\rangle$	**3.63**	12.89	10.13	1341.94	48.95
	2WD	**3.66**	26.34	9.48	1080.75	49.17
Partial	$\langle$MSE$\rangle$	**0.54**	0.97	0.97	–	–
	2WD	**0.48**	0.94	0.98	–	–

methods [1,3,57], our framework integrates typical mathematical physical structures of orthogonality and flexibility and thus has natural advantages in accurately reconstructing the Hamiltonian dynamics from the noise-perturbed data, and achieving accurate prediction simultaneously and perfectly with energy conservation [58,59]. Therefore, it can be applied to reconstruction and prediction problems in the real-world scenarios where the conservation laws are persistent but the collected data are more or less shuffled and contaminated. More importantly, based on the Koopman invariant space spanned by the features in the encoder and the orthogonal spectral decomposition of the Koopman matrix, we show that our HNKO owns an ability of discovering the conservation laws with analytical expression, which makes an essential step towards solving the critical problem of discovering conservation laws [29,57,60,61].

Appendix A: Reducing computational cost by Kronecker product. Since the Kronecker product of two orthogonal matrices is still an orthogonal matrix [43], one can reduce the computational complexity for computing the p-dimensional orthogonal matrix $\boldsymbol{K}$ by replacing it with the Kronecker product of two low-order orthogonal matrices $(\boldsymbol{K}_1)_{p_1 \times p_1}$ and $(\boldsymbol{K}_2)_{p_2 \times p_2}$. Originally, obtaining the orthogonal matrix $\boldsymbol{K}$ involved calculating the Lie exponent $\exp(\boldsymbol{A})$ of a skew-symmetric matrix $\boldsymbol{A}$, which results in a computational complexity of $\mathcal{O}(p^3)$ due to the high power of $\boldsymbol{A}$ in the expansion. However, with the Kronecker method, the computational cost can be reduced to $\mathcal{O}(p_1^3 + p_2^3)$. Since $p_1 p_2 = p$ and it can be easily verified that the minimum of the function $p_1^3 + p^3/p_1^3$ is $2p^{3/2}$. Thus, the Kronecker product operation can reduce the original computational cost to at least $\mathcal{O}(p^{3/2})$. Additionally, for extremely high-dimensional tasks, the order of the computational cost can be further reduced by repeating the Kronecker operation. For instance, by designing $\boldsymbol{K} = (\boldsymbol{K}_1)_{p_1 \times p_1} \otimes \ldots \otimes (\boldsymbol{K}_m)_{p_m \times p_m}$ in an order of $\mathcal{O}(\sum_{i=1}^{m} p_i^3)$ with $p_1 \times \ldots \times p_m = p$ and $p_1 \approx \ldots \approx p_m \approx p^{1/m}$, the computational cost of the matrix product in calculating the Lie exponent is reduced to $\mathcal{O}(m \cdot p^{3/m})$. For brevity, we only consider the most simple case of the Kronecker operation $\boldsymbol{K} = (\boldsymbol{K}_1)_{p_1 \times p_1} \otimes (\boldsymbol{K}_2)_{p_2 \times p_2}$, with $p_1 \approx p_2 \approx \sqrt{p}$, in this work. Notably, this condition is easily satisfied by choosing p as a square number.

Appendix B: Comparison studies. We present the comparison studies on different reconstruction and prediction tasks in Tables I and II, using the HNKO, its extension HNKO*, and other mainstream methods.

TABLE II. Comparison studies of different methods on the KdV equation. Here, s is the uniform discretization in the space. The prediction MSE and the training time (Tt) are shown, respectively, illustrating the high performance using the HNKO and its extension.

Model	$\langle$MSE$\rangle$ $\downarrow$					Tt (No. of Seconds) $\downarrow$				
	$s = 64$	144	256	576	1024	64	144	256	576	1024
DMD	0.52	7.3×10^4	2.93	2.55	1.68	–	–	–	–	–
DLKO	10.73	33.07	2.22	4.55	8.40	**37**	87	254	866	2449
NODE	0.52	1.18	14.34	4.9×10^6	1.8×10^8	595	8943	1.4×10^4	2.9×10^4	1.8×10^5
HNKO	**0.42**	**0.50**	**1.71**	2.79	10.56	62	262	646	8280	3.7×10^4
HNKO*	1.09	2.43	6.21	**1.07**	**0.93**	46	**59**	**64**	**175**	**426**

[1] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks, in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada* (Curran Associates, Inc., Vancouver, 2019), Vol. 32.

[2] C.-D. Han, B. Glaz, M. Haile, and Y.-C. Lai, Adaptable Hamiltonian neural networks, Phys. Rev. Res. **3**, 023156 (2021).

[3] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho, Lagrangian neural networks, in *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations* (OpenReview, Addis Ababa, 2019).

[4] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, Neural ordinary differential equations, in *Advances in Neural Information Processing Systems* (Curran Associates, New York, 2019), Vol. 31.

[5] Q. Zhu, Y. Guo, and W. Lin, Neural delay differential equations, in *International Conference on Learning Representations* (OpenReview, Virtual, 2021).

[6] Q. Zhu, Y. Shen, D. Li, and W. Lin, Neural piecewise-constant delay differential equations, in *Proceedings of the AAAI Conference on Artificial Intelligence* (OpenReview, Virtual, 2022), Vol. 36, pp. 9242–9250.

[7] 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, J. Comput. Phys. **378**, 686 (2019).

[8] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, Nature Rev. Phys. **3**, 422 (2021).

[9] F. de Avila Belbute-Peres, Y.-f. Chen, and F. Sha, HyperPINN: Learning parameterized differential equations with physics-informed hypernetworks, in *The Symbiosis of Deep Learning and Differential Equations* (OpenReview, Sydney, 2021).

[10] L. Lu, R. Pestourie, W. Yao, Z. Wang, F. Verdugo, and S. G. Johnson, Physics-informed neural networks with hard constraints for inverse design, SIAM J. Sci. Comput. **43**, B1105 (2021).

[11] W. Chen, Q. Wang, J. S. Hesthaven, and C. Zhang, Physics-informed machine learning for reduced-order modeling of nonlinear problems, J. Comput. Phys. **446**, 110666 (2021).

[12] H. Eivazi, M. Tahani, P. Schlatter, and R. Vinuesa, Physics-informed neural networks for solving reynolds-averaged navier–stokes equations, Phys. Fluids **34**, 075117 (2022).

[13] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, Physics-informed neural networks (PINNs) for fluid mechanics: A review, Acta Mech. Sin. **37**, 1727 (2021).

[14] Z. Mao, A. D. Jagtap, and G. E. Karniadakis, Physics-informed neural networks for high-speed flows, Comput. Methods Appl. Mech. Eng. **360**, 112789 (2020).

[15] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, Physics-informed neural networks for power systems, in *IEEE PES GM* (IEEE, New York, 2020), pp. 1–5.

[16] Q. Zhu, H. Ma, and W. Lin, Detecting unstable periodic orbits based only on time series: When adaptive delayed feedback control meets reservoir computing, Chaos **29**, 093125 (2019).

[17] Y. Tang, J. Kurths, W. Lin, E. Ott, and L. Kocarev, When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics, Chaos **30**, 063151 (2020).

[18] X.-Y. Duan, X. Ying, S.-Y. Leng, J. Kurths, W. Lin, and H.-F. Ma, Embedding theory of reservoir computing and

reducing reservoir network using time delays, Phys. Rev. Res. **5**, L022041 (2023).

[19] P. Toth, D. J. Rezende, A. Jaegle, S. Racanière, A. Botev, and I. Higgins, Hamiltonian generative networks, in *International Conference on Learning Representations* (OpenReview, Addis Ababa, 2020).

[20] Y. D. Zhong, B. Dey, and A. Chakraborty, Symplectic ODE-Net: Learning Hamiltonian dynamics with control, in *International Conference on Learning Representations* (OpenReview, Addis Ababa, 2020).

[21] A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, and P. Battaglia, Hamiltonian graph networks with ode integrators, in *NeurIPS Workshop on Machine Learning and the Physical Sciences* (OpenReview, Vancouver, 2019).

[22] H. Ma, W. Lin, and Y.-C. Lai, Detecting unstable periodic orbits in high-dimensional chaotic systems from time series: Reconstruction meeting with adaptation, Phys. Rev. E **87**, 050901(R) (2013).

[23] Q. Zhu, X. Li, and W. Lin, Leveraging neural differential equations and adaptive delayed feedback to detect unstable periodic orbits based on irregularly sampled time series, Chaos **33**, 031101 (2023).

[24] J. Zhang, Q. Zhu, and W. Lin, Neural stochastic control, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., New Orleans, 2022), Vol. 35, pp. 9098–9110.

[25] J. Zhang, Q. Zhu, W. Yang, and W. Lin, Sync: Safety-aware neural control for stabilizing stochastic delay-differential equations, in *International Conference on Learning Representations* (OpenReview, Kigali, 2023).

[26] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown *et al.*, Tackling climate change with machine learning, ACM Comput. Surv. **55**, 1 (2023).

[27] X. Li, Q. Zhu, C. Zhao, X. Qian, X. Zhang, X. Duan, and W. Lin, Tipping point detection using reservoir computing, Research **6**, 0174 (2023).

[28] P. Zhang, H. Shen, and H. Zhai, Machine learning topological invariants with neural networks, Phys. Rev. Lett. **120**, 066401 (2018).

[29] Z. Liu and M. Tegmark, Machine learning conservation laws from trajectories, Phys. Rev. Lett. **126**, 180604 (2021).

[30] M. Lezcano-Casado and D. Martınez-Rubio, Cheap orthogonal constraints in neural networks: A simple parametrization of the orthogonal and unitary group, in *Proceedings of the 36th International Conference on Machine Learning* (PMLR, Long Beach, 2019), Vol. 97.

[31] M. Lezcano-Casado, Trivializations for gradient-based optimization on manifolds, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Vancouver, 2019), Vol. 32.

[32] S. L. Brunton, M. Budisic, E. Kaiser, and J. N. Kutz, Modern Koopman theory for dynamical systems, SIAM Rev. **64**, 229 (2022).

[33] I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, Nonlinear Dyn. **41**, 309 (2005).

[34] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, Spectral analysis of nonlinear flows, J. Fluid Mech. **641**, 115 (2009).

[35] B. Lusch, J. N. Kutz, and S. L. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics, Nature Commun. **9**, 4950 (2018).

[36] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. **52**, 477 (2020).

[37] A. Otto, D. Müller, and G. Radons, Universal dichotomy for dynamical systems with variable delay, Phys. Rev. Lett. **118**, 044104 (2017).

[38] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. **656**, 5 (2010).

[39] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, On dynamic mode decomposition: Theory and applications, J. Comput. Dynam. **1**, 391 (2014).

[40] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, A data–driven approximation of the koopman operator: Extending dynamic mode decomposition, J. Nonlinear Sci. **25**, 1307 (2015).

[41] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, 2022).

[42] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.L012031 for the theories of the manifold dimension and the orthogonal projection of the Koopman operator, the reduced computational cost, the discovered Hamiltonian, the mechanism of filtering the noise, the illustrations on the simulation configurations, and the limitations and perspectives.

[43] C. F. Loan, The ubiquitous kronecker product, J. Comput. Appl. Math. **123**, 85 (2000).

[44] M. Valtonen and H. Karttunen, *The Three-Body Problem* (Cambridge University Press, Cambridge, 2006).

[45] P. Jin, Z. Zhang, A. Zhu, Y. Tang, and G. E. Karniadakis, Sympnets: Intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems, Neural Networks **132**, 166 (2020).

[46] R. Mutegeki and D. S. Han, A cnn-lstm approach to human activity recognition, in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (IEEE, New York, 2020), pp. 362–366.

[47] A. Vidal and W. Kristjanpoller, Gold volatility prediction using a cnn-lstm approach, Exp. Syst. Appl. **157**, 113481 (2020).

[48] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, Phys. Rev. Lett. **120**, 024102 (2018).

[49] H. Zhang, H. Fan, L. Wang, and X. Wang, Learning Hamiltonian dynamics with reservoir computing, Phys. Rev. E **104**, 024205 (2021).

[50] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 2013), Vol. 60.

[51] G. Fusco, G. F. Gronchi, and P. Negrini, Platonic polyhedra, topological constraints and periodic solutions of the classical n-body problem, Invent. Math. **185**, 283 (2011).

[52] G. R. Fowles, G. L. Cassiday *et al.*, *Analytical Mechanics* (Saunders College, Rochester, 1999).

[53] S. Kim, W. Ji, S. Deng, Y. Ma, and C. Rackauckas, Stiff neural ordinary differential equations, Chaos **31**, 093122 (2021).

[54] A. Ghosh, H. Behl, E. Dupont, P. Torr, and V. Namboodiri, STEER: Simple temporal regularization for neural ode, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Vancouver, 2020), Vol. 33.

[55] R. M. Miura, C. S. Gardner, and M. D. Kruskal, Korteweg-de vries equation and generalizations. II. Existence of conservation laws and constants of motion, J. Math. Phys. **9**, 1204 (1968).

[56] N. Zabusky and C. Galvin, Shallow-water waves, the korteweg-devries equation and solitons, J. Fluid Mech. **47**, 811 (1971).

[57] E. Kaiser, J. N. Kutz, and S. L. Brunton, Discovering conservation laws from data for control, in *IEEE Conference on Decision and Control* (IEEE, New York, 2018), pp. 6415–6421.

[58] S. Liang, Z. Huang, and H. Zhang, Stiffness-aware neural network for learning Hamiltonian systems, in *International Conference on Learning Representations* (OpenReview, Virtual, 2021).

[59] Z. Chen, J. Zhang, M. Arjovsky, and L. Bottou, Symplectic recurrent neural networks, in *International Conference on Learning Representations* (OpenReview, New Orleans, 2019).

[60] P. Y. Lu, R. Dangovski, and M. Soljačić, Discovering conservation laws using optimal transport and manifold learning, Nature Commun. **14**, 4744 (2023).

[61] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, Discovering physical concepts with neural networks, Phys. Rev. Lett. **124**, 010508 (2020).