UC San Diego
UC San Diego Previously Published Works

Title
Approximation of nearly-periodic symplectic maps via structure-preserving neural networks.

Permalink
https://escholarship.org/uc/item/5qb2w7k5

Journal
Scientific reports, 13(1)

ISSN
2045-2322

Authors
Duruisseaux, Valentin
Burby, Joshua W
Tang, Qi

Publication Date
2023-05-01

DOI
10.1038/s41598-023-34862-w

Peer reviewed

scientific reports

OPEN

Approximation of nearly-periodic symplectic maps via structure-preserving neural networks

Valentin Duruisseaux[1]✉, Joshua W. Burby[2] & Qi Tang[2]

A continuous-time dynamical system with parameter ε is nearly-periodic if all its trajectories are periodic with nowhere-vanishing angular frequency as ε approaches 0. Nearly-periodic maps are discrete-time analogues of nearly-periodic systems, defined as parameter-dependent diffeomorphisms that limit to rotations along a circle action, and they admit formal $U(1)$ symmetries to all orders when the limiting rotation is non-resonant. For Hamiltonian nearly-periodic maps on exact presymplectic manifolds, the formal $U(1)$ symmetry gives rise to a discrete-time adiabatic invariant. In this paper, we construct a novel structure-preserving neural network to approximate nearly-periodic symplectic maps. This neural network architecture, which we call symplectic gyroceptron, ensures that the resulting surrogate map is nearly-periodic and symplectic, and that it gives rise to a discrete-time adiabatic invariant and a long-time stability. This new structure-preserving neural network provides a promising architecture for surrogate modeling of non-dissipative dynamical systems that automatically steps over short timescales without introducing spurious instabilities.

Dynamical systems evolve according to the laws of physics, which can usually be described using differential equations. By solving these differential equations, it is possible to predict the future states of a dynamical system. Identifying accurate and efficient dynamic models based on observed trajectories is thus critical for the analysis, simulation and control of dynamical systems. We consider here the problem of learning dynamics: given a dataset of trajectories followed by a dynamical system, we wish to infer the dynamical law responsible for these trajectories and then possibly use that law to predict the evolution of similar systems in different initial states. We are particularly interested in the surrogate modeling problem: the underlying dynamical system is known, but traditional simulations are either too slow or expensive for some optimization task. This problem can be addressed by learning a less expensive, but less accurate surrogate for the simulations.

Models obtained from first principles are extensively used across science and engineering. Unfortunately, due to incomplete knowledge, these models based on physical laws tend to over-simplify or incorrectly describe the underlying structure of the dynamical systems, and usually lead to high bias and modeling errors that cannot be corrected by optimizing over the few parameters in the models.

Deep learning architectures can provide very expressive models for function approximation, and have proven very effective in numerous contexts[1–3]. Unfortunately, standard non-structure-preserving neural networks struggle to learn the symmetries and conservation laws underlying dynamical systems, and as a result do not generalize well. Indeed, they tend to prefer certain representations of the dynamics where the symmetries and conservation laws of the system are not exactly enforced. As a result, these models do not generalize well as they are often not capable of producing physically plausible results when applied to new unseen states. Deep learning models capable of learning and generalizing dynamics effectively are typically over-parameterized, and as a consequence tend to have high variance and can be very difficult to interpret[4]. Also, training these models usually requires large datasets and a long computational time, which makes them prohibitively expensive for many applications.

A recent research direction is to consider a hybrid approach which combines knowledge of physics laws and deep learning architectures[2,3,5,6]. The idea is to encode physics laws and the conservation of geometric properties of the underlying systems in the design of the neural networks or in the learning process. Available physics prior knowledge can be used to construct physics-constrained neural networks with improved design and efficiency

[1]Department of Mathematics, University of California San Diego, La Jolla, CA 92093, USA. [2]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ✉email: vduruiss@ucsd.edu

and a better generalization capacity, which take advantage of the function approximation power of neural networks to deal with incomplete knowledge.

In this paper, we will consider the problem of learning dynamics for highly-oscillatory Hamiltonian systems. Examples include the Klein–Gordon equation in the weakly-relativistic regime, charged particles moving through a strong magnetic field, and the rotating inviscid Euler equations in quasi-geostrophic scaling[7]. More generally, *any* Hamiltonian system may be embedded as a normally-stable elliptic slow manifold in a nearly-periodic Hamiltonian system[8]. Highly-oscillatory Hamiltonian systems exhibit two basic structural properties whose interactions play a crucial role in their long-term dynamics. First is preservation of the symplectic form, as for all Hamiltonian systems. Second is timescale separation, corresponding to the relatively short timescale of oscillations compared with slower secular drifts. Coexistence of these two structural properties implies the existence of an adiabatic invariant[8–11]. Adiabatic invariants differ from true constants of motion, in particular energy invariants, which do not change at all over arbitrary time intervals. Instead adiabatic invariants are conserved with limited precision over very large time intervals. There are no learning frameworks available today that exactly preserve the two structural properties whose interplay gives rise to adiabatic invariants. This work addresses this challenge by exploiting a recently-developed theory of *nearly-periodic symplectic maps*[11], which can be thought of as discrete-time analogues of highly-oscillatory Hamiltonian systems[9].

As a result of being symplectic, a mapping assumes a number of special properties. In particular, symplectic mappings are closely related to Hamiltonian systems: any solution to a Hamiltonian system is a symplectic flow[12], and any symplectic flow corresponds locally to an appropriate Hamiltonian system[13]. It is well-known that preserving the symplecticity of a Hamiltonian system when constructing a discrete approximation of its flow map ensures the preservation of many aspects of the dynamical system such as energy conservation, and leads to physically well-behaved discrete solutions over exponentially-long time intervals[13–17]. It is thus important to have structure-preserving neural network architectures which can learn symplectic maps and ensure that the learnt surrogate map preserves symplecticity. Many physics-informed and structure-preserving machine learning approaches have recently been proposed to learn Hamiltonian dynamics and symplectic maps[2,3,18–35]. In particular, Hénon Neural Networks (HénonNets)[2] can approximate arbitrary well any symplectic map via compositions of simple yet expressive elementary symplectic mappings called Hénon-like mappings. In the numerical experiments conducted in this paper, HénonNets[2] will be our preferred choice of symplectic map approximator to use as building block in our framework for approximation of nearly-periodic symplectic maps, although some of the other approaches listed above for approximating symplectic mappings can be used within our framework as well.

As shown by Kruskal[9], every nearly-periodic system, Hamiltonian or not, admits an approximate $U(1)$-symmetry, determined to leading order by the unperturbed periodic dynamics. It is well-known that a Hamiltonian system which admits a continuous family of symmetries also admits a corresponding conserved quantity. It is thus not surprising that a nearly-periodic Hamiltonian system, which admits an approximate symmetry, must also have an approximate conservation law[11], and the approximately conserved quantity is referred to as an adiabatic invariant.

Nearly-periodic maps, first introduced by Burby et al.[11], are natural discrete-time analogues of nearly-periodic systems, and have important applications to numerical integration of nearly-periodic systems. Nearly-periodic maps may also be used as tools for structure-preserving simulation of non-canonical Hamiltonian systems on exact symplectic manifolds[11], which have numerous applications across the physical sciences. Noncanonical Hamiltonian systems play an especially important role in modeling weakly-dissipative plasma systems[36–42]. Similarly to the continuous-time case, nearly-periodic maps with a Hamiltonian structure (that is symplecticity) admit an approximate symmetry and as a result also possess an adiabatic invariant[11]. The adiabatic invariants that our networks target only arise in purely Hamiltonian systems. Just like dissipation breaks the link between symmetries and conservation laws in Hamiltonian systems, dissipation also breaks the link between approximate symmetries and approximate conservation laws in Hamiltonian systems. We are not considering systems with symmetries that are broken by dissipation or some other mechanism, but rather considering systems which possess approximate symmetries. This should be contrasted with other frameworks[43–45] which develop machine learning techniques for systems that explicitly include dissipation.

We note that neural network architectures designed for multi-scale dynamics and long-time dependencies are available[46], and that many authors have introduced numerical algorithms specifically designed to efficiently step over high-frequency oscillations[47–49]. However, the problem of developing surrogate models for dynamical systems that avoid resolving short oscillations remains open. Such surrogates would accelerate optimization algorithms that require querying the dynamics of an oscillatory system during the optimizer's "inner loop". The network architecture presented in this article represents a first important step toward a general solution of this problem. Some of its advantages are that it aims to learn a fast surrogate model that can resolve long-time dynamics using very short time data, and that it is guaranteed to enjoy symplectic universal approximation within the class of nearly periodic maps. As developed in this paper, our method applies to dynamical systems that exhibit a single fast mode of oscillation. In particular, when initial conditions for the surrogate model are selected on the zero level set of the learned adiabatic invariant, the network automatically integrates along the slow manifold[50–54]. While our network architecture generalizes in a straightforward manner to handle multiple non-resonant modes, it cannot be applied to dynamical systems that exhibit resonant surfaces.

Note that many of the approaches listed earlier for physics-based or structure-preserving learning of Hamiltonian dynamics focus on learning the vector field associated to the continuous-time Hamiltonian system, while others learn a discrete-time symplectic approximation to the flow map of the Hamiltonian system. In many contexts, we do not need to infer the continuous-time dynamics, and only need a surrogate model which can rapidly generate accurate predictions which remain physically consistent for a long time. Learning a discrete-time approximation to the evolution or flow map, instead of learning the continuous-time vector field, allows for fast prediction and simulation without the need to integrate differential equations or use neural ODEs and

adjoint techniques (which can be very expensive and can introduce additional errors due to discretization). In this paper, we will learn nearly-periodic symplectic approximations to the flow maps of nearly-periodic Hamiltonian systems, with the intention of obtaining algorithms which can generate accurate and physically-consistent simulations much faster than traditional integrators.

Outline. We first review briefly some background notions from differential geometry in Sect. "Differential geometry background". Then, we discuss how symplectic maps can be approximated using HénonNets in Sect. "Approximation of symplectic maps via hénon neural networks", before defining nearly-periodic systems and maps and reviewing their important properties in Sect. "Nearly-periodic systems and nearly-periodic maps". In Sect. "Novel structure-preserving neural network architectures", we introduce novel neural network architectures, *gyroceptrons* and *symplectic gyroceptrons*, to approximate symplectic and non-symplectic nearly-periodic maps. We then show in Sect. "Numerical confirmation of the existence of adiabatic invariants" that symplectic gyroceptrons admit adiabatic invariants regardless of the values of their weights. Finally, in Sect. "Numerical examples of learning surrogate maps", we demonstrate how the proposed architecture can be used to learn surrogate maps for the nearly-periodic symplectic flow maps associated to two different systems: a nearly-periodic Hamiltonian system composed of two nonlinearly coupled oscillators (in Sect. "Nonlinearly coupled oscillators"), and the nearly-periodic Hamiltonian system describing the evolution of a charged particle interacting with its self-generated electromagnetic field (in Sect. "Charged particle interacting with its self-generated electromagnetic field").

Preliminaries

Differential geometry background.

In this paper, we reserve the symbol M for a smooth manifold equipped with a smooth auxiliary Riemannian metric g, and $\mathcal{E}$ will always denote a vector space for the parameter ε. We will now briefly introduce some standard concepts from differential geometry that will be used throughout this paper (more details can be found in introductory differential geometry books[55–57]).

A smooth map $h : M_1 \to M_2$ between smooth manifolds M_1, M_2 is a **diffeomorphism** if it is bijective with a smooth inverse. We say that $f_\varepsilon : M_1 \to M_2$, $\varepsilon \in \mathcal{E}$, is a smooth ε-dependent mapping when the mapping $M_1 \times \mathbb{R} \to M_2 : (m, \varepsilon) \mapsto f_\varepsilon(m)$ is smooth.

A **vector field** on a manifold M is a map $X : M \to TM$ such that $X(m) \in T_m M$ for all $m \in M$, where $T_m M$ denotes the **tangent space** to M at m and $TM = \{(m, v) \mid m \in M, v \in T_m M\}$ is the **tangent bundle** TM of M. The vector space dual to $T_m M$ is the **cotangent space** $T_m^* M$, and the **cotangent bundle** of M is $T^* M = \{(m, p) \mid m \in M, p \in T_m^* M\}$. The integral curve at m of a vector field X is the smooth curve c on M such that $c(0) = m$ and $c'(t) = X(c(t))$. The **flow** of a vector field X is the collection of maps $\varphi_t : M \to M$ such that $\varphi_t(m)$ is the integral curve of X with initial condition $m \in M$.

A **k-form** on a manifold M is a map which assigns to every point $m \in M$ a skew-symmetric k-multilinear map on $T_m M$. Let α be a k-form and β be a s-form β on a manifold M. Their tensor product $\alpha \otimes \beta$ at $m \in M$ is defined via

$$(\alpha \otimes \beta)_m(v_1, \ldots, v_{k+s}) = \alpha_m(v_1, \ldots, v_k)\beta_m(v_{k+1}, \ldots, v_{k+s}).$$

The alternating operator Alt acts on a k-form α via

$$\mathrm{Alt}(\alpha)(v_1, \ldots, v_k) = \frac{1}{k!} \sum_{\pi \in S_k} \mathrm{sgn}(\pi)\alpha(v_{\pi(1)}, \ldots, v_{\pi(k)}),$$

where S_k is the group of all the permutations of $\{1, \ldots, k\}$ and $\mathrm{sgn}(\pi)$ is the sign of the permutation. The **wedge product** $\alpha \wedge \beta$ is then defined via

$$\alpha \wedge \beta = \frac{(k+s)!}{k!s!}\mathrm{Alt}(\alpha \otimes \beta).$$

The **exterior derivative** of a smooth function $f : M \to \mathbb{R}$ is its differential $\mathbf{d}f$, and the **exterior derivative** $\mathbf{d}\alpha$ of a k-form α with $k > 0$ is the $(k + 1)$-form defined by

$$\mathbf{d}\left(\sum_{i_1, \ldots, i_k} \alpha_{i_1 \ldots i_k}\mathbf{d}x^{i_1} \wedge \ldots \wedge \mathbf{d}x^{i_k}\right) = \sum_j \sum_{i_1, \ldots, i_k} \partial_j \alpha_{i_1 \ldots i_k}\mathbf{d}x^j \wedge \mathbf{d}x^{i_1} \wedge \ldots \wedge \mathbf{d}x^{i_k}.$$

The **interior product** $\iota_X \alpha$ where X is a vector field on M and α is a k-form is the $(k - 1)$-form defined via

$$(\iota_X \alpha)_m(v_2, \ldots, v_k) = \alpha_m(X(m), v_2, \ldots, v_k).$$

The **pull-back** $\psi^* \alpha$ of α by a smooth map $\psi : M \to N$ is the k-form defined by

$$(\psi^* \alpha)_m(v_1, \ldots, v_k) = \alpha_{\psi(m)}(\mathbf{d}\psi \cdot v_1, \ldots, \mathbf{d}\psi \cdot v_k).$$

The **Lie derivative** $\mathcal{L}_X \alpha$ of the k-form α along a vector field X with flow φ_t is $\mathcal{L}_X \alpha = \frac{d}{dt}\big|_{t=0} \varphi_t^* \alpha$, and for a smooth function $f : M \to \mathbb{R}$, $\mathcal{L}_X f$ is the directional derivative $\mathcal{L}_X f = \mathbf{d}f \cdot X$.

The **circle group** $U(1)$, also known as first unitary group, is the one-dimensional Lie group of complex numbers of unit modulus with the standard multiplication operation. It can be parametrized via $e^{i\theta}$ for $\theta \in [0, 2\pi)$, and is isomorphic to the special orthogonal group $SO(2)$ of rotations in the plane. A **circle action** on a manifold

M is a one-parameter family of smooth diffeomorphisms $\Phi_\theta : M \to M$ that satisfies the following three properties for any $\theta, \theta_1, \theta_2 \in U(1) \cong \mathbb{R} \bmod 2\pi$:

$$\Phi_{\theta+2\pi} = \Phi_\theta \quad \text{(periodicity)}, \qquad \Phi_0 = \text{Id}_M \quad \text{(identity)}, \qquad \Phi_{\theta_1+\theta_2} = \Phi_{\theta_1} \circ \Phi_{\theta_2} \quad \text{(additivity)}.$$

The **infinitesimal generator** of a circle action Φ_θ on M is the vector field on M defined by $m \mapsto \frac{d}{d\theta}\Big|_{\theta=0} \Phi_\theta(m)$.

Approximation of symplectic maps via Hénon neural networks.

Let $U \subset \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$ be an open set in an even-dimensional Euclidean space. Denote points in $\mathbb{R}^n \times \mathbb{R}^n$ using the notation (x, y), with $x, y \in \mathbb{R}^n$. A smooth mapping $\Phi : U \to \mathbb{R}^{2n}$ with components $\Phi(x, y) = (\bar{x}(x, y), \bar{y}(x, y))$ is symplectic if

$$\sum_{i=1}^n \mathbf{d}x^i \wedge \mathbf{d}y^i = \sum_{i=1}^n \mathbf{d}\bar{x}^i \wedge \mathbf{d}\bar{y}^i. \tag{2.1}$$

The symplectic condition (2.1) implies that the mapping Φ has a number of special properties. In particular, there is a close relation between Hamiltonian systems and symplecticity of flows: Poincaré's Theorem[12] states that any solution to a Hamiltonian system is a symplectic flow, and it can also be shown that any symplectic flow corresponds locally to an appropriate Hamiltonian system. Preserving the symplecticity of a Hamiltonian system when constructing a discrete approximation of its flow map ensures the preservation of many aspects of the dynamical system such as energy conservation, and leads to physically well-behaved discrete solutions[13–17]. It is thus important to have structure-preserving network architectures which can learn symplectic maps.

The space of all symplectic maps is infinite dimensional[58], so the problem of approximating an arbitrary symplectic map using compositions of simpler symplectic mappings is inherently interesting. Turaev[59] showed that every symplectic map may be approximated arbitrarily well by compositions of *Hénon-like maps*, which are special elementary symplectic maps.

Definition 2.1 Let $V : \mathbb{R}^n \to \mathbb{R}$ be a smooth function on $\mathbb{R}^n$ and let $\eta \in \mathbb{R}^n$ be a constant. We define the **Hénon-like map** $H[V, \eta] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ with potential V and shift η via

$$H[V, \eta] \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y + \eta \\ -x + \nabla V(y) \end{pmatrix}. \tag{2.2}$$

Theorem 2.1 (Turaev[59]) *Let $\Phi : U \to \mathbb{R}^n \times \mathbb{R}^n$ be a C^{r+1} symplectic mapping. For each compact set $C \subset U$ and $\delta > 0$ there is a smooth function $V : \mathbb{R}^n \to \mathbb{R}$, a constant η, and a positive integer N such that $H[V, \eta]^{4N}$ approximates the mapping Φ within δ in the C^r topology.*

Remark 2.1 The significance of the number 4 in this theorem follows from the fact that the fourth iterate of the Hénon-like map with trivial potential $V = 0$ is the identity map: $H[0, \eta]^4 = \text{Id}_{\mathbb{R}^n \times \mathbb{R}^n}$.

Turaev's result suggests a specific neural network architecture to approximate symplectic mappings using Hénon-like maps[2]. We review the construction of HénonNets[2], starting with the notion of a *Hénon layer*.

Definition 2.2 Let $\eta \in \mathbb{R}^n$ be a constant vector, and let V be a scalar feed-forward neural network on $\mathbb{R}^n$, that is, a smooth mapping $V : \mathcal{W} \times \mathbb{R}^n \to \mathbb{R}$, where $\mathcal{W}$ is a space of neural network weights. The **Hénon layer** with potential V, shift η, and weight W is the iterated Hénon-like map

$$L[V[W], \eta] = H[V[W], \eta]^4, \tag{2.3}$$

where we use the notation $V[W]$ to denote the mapping $V[W](y) = V(W, y)$, for any $y \in \mathbb{R}^n$, $W \in \mathcal{W}$.

There are various network architectures for the potential $V[W]$ that are capable of approximating any smooth function $V : \mathbb{R}^n \to \mathbb{R}$ with any desired level of accuracy. For example, a fully-connected neural network with a single hidden layer of sufficient width can approximate any smooth function. Therefore a corollary of Theorem 2.1 is that any symplectic map may be approximated arbitrarily well by the composition of sufficiently many Hénon layers with various potentials and shifts. This leads to the notion of a *Hénon Neural Network*.

Definition 2.3 Let N be a positive integer and

- $\boldsymbol{V} = \{V_k\}_{k \in \{1,\dots,N\}}$ be a family of scalar feed-forward neural networks on $\mathbb{R}^n$
- $\boldsymbol{W} = \{W_k\}_{k \in \{1,\dots,N\}}$ be a family of network weights for $\boldsymbol{V}$
- $\boldsymbol{\eta} = \{\eta_k\}_{k \in \{1,\dots,N\}}$ be a family of constants in $\mathbb{R}^n$

The **Hénon neural network (HénonNet)** with layer potentials $\boldsymbol{V}$, layer weights $\boldsymbol{W}$, and layer shifts $\boldsymbol{\eta}$ is the mapping

$$\mathcal{H}[\boldsymbol{V}[\boldsymbol{W}], \boldsymbol{\eta}] = L[V_N[W_N], \eta_N] \circ \dots \circ L[V_2[W_2], \eta_2] \circ L[V_1[W_1], \eta_1] \tag{2.4}$$

$$= H[V_N[W_N], \eta_N]^4 \ \circ \ \ldots \ \circ \ H[V_2[W_2], \eta_2]^4 \ \circ \ H[V_1[W_1], \eta_1]^4. \tag{2.5}$$

A composition of symplectic mappings is also symplectic, so every HénonNet is a symplectic mapping, regardless of the architectures for the networks V_k and of the weights W_k. Furthermore, Turaev's Theorem 2.1 implies that the family of HénonNets is sufficiently expressive to approximate any symplectic mapping:

Lemma 2.1 *Let $\Phi : U \to \mathbb{R}^n \times \mathbb{R}^n$ be a C^{r+1} symplectic mapping. For each compact set $C \subset U$ and $\delta > 0$ there is a HénonNet $\mathcal{H}$ that approximates Φ within δ in the C^r topology.*

Remark 2.2 Note that Hénon-like maps are easily invertible,

$$H[V, \eta]\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y + \eta \\ -x + \nabla V(y) \end{pmatrix} \quad \Rightarrow \quad H^{-1}[V, \eta]\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \nabla V(x - \eta) - y \\ x - \eta \end{pmatrix}, \tag{2.6}$$

so we can also easily invert Hénon networks by composing inverses of Hénon-like maps.

We also introduce here modified versions of Hénon-like maps and HénonNets to approximate symplectic maps possessing a near-identity property:

Definition 2.4 Let $V : \mathbb{R}^n \to \mathbb{R}$ be a smooth function and let $\eta \in \mathbb{R}^n$ be a constant. We define the **near-identity Hénon-like map** $H_\varepsilon[V, \eta] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ with potential V and shift η via

$$H_\varepsilon[V, \eta]\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y + \eta \\ -x + \varepsilon \nabla V(y) \end{pmatrix}. \tag{2.7}$$

Near-identity Hénon-like maps satisfy the near-identity property $H_0[V, \eta]^4 = \mathrm{Id}_{\mathbb{R}^n \times \mathbb{R}^n}$.

Definition 2.5 Let N be a positive integer and

- $\boldsymbol{V} = \{V_k\}_{k \in \{1, \ldots, N\}}$ be a family of scalar feed-forward neural networks on $\mathbb{R}^n$
- $\boldsymbol{W} = \{W_k\}_{k \in \{1, \ldots, N\}}$ be a family of network weights for $\boldsymbol{V}$
- $\boldsymbol{\eta} = \{\eta_k\}_{k \in \{1, \ldots, N\}}$ be a family of constants in $\mathbb{R}^n$

The **near-identity Hénon network** with layer potentials $\boldsymbol{V}$, layer weights $\boldsymbol{W}$, and layer shifts $\boldsymbol{\eta}$ is the mapping defined via

$$\mathcal{H}_\varepsilon[\boldsymbol{V}[\boldsymbol{W}], \boldsymbol{\eta}] \ = \ H_\varepsilon[V_N[W_N], \eta_N]^4 \ \circ \ \ldots \ \circ \ H_\varepsilon[V_2[W_2], \eta_2]^4 \ \circ \ H_\varepsilon[V_1[W_1], \eta_1]^4, \tag{2.8}$$

and it satisfies the near-identity property $\mathcal{H}_0[\boldsymbol{V}[\boldsymbol{W}], \boldsymbol{\eta}] = \mathrm{Id}_{\mathbb{R}^n \times \mathbb{R}^n}$.

Nearly-periodic systems and nearly-periodic maps. *Nearly-periodic systems.* Intuitively, a continuous-time dynamical system with parameter ε is *nearly-periodic* if all of its trajectories are periodic with nowhere-vanishing angular frequency in the limit $\varepsilon \to 0$. Such a system characteristically displays limiting short-timescale dynamics that ergodically cover circles in phase space. More precisely, a nearly-periodic systems can be defined as follows:

Definition 2.6 [Burby et al.[11]] A **nearly-periodic system** on a manifold M is a smooth ε-dependent vector field X_ε on M such that $X_0 = \omega_0 R_0$, where

- R_0 is the infinitesimal generator for a circle action $\Phi_\theta : M \to M, \theta \in U(1)$.
- $\omega_0 : M \to \mathbb{R}$ is strictly positive and its Lie derivative satisfies $\mathcal{L}_{R_0} \omega_0 = 0$.

The vector field R_0 is called the **limiting roto-rate**, and ω_0 is the **limiting angular frequency**.

Examples from physics include charged particle dynamics in a strong magnetic field, the weakly-relativistic Dirac equation, and any mechanical system subject to a high-frequency, time-periodic force. In the broader context of multi-scale dynamical systems, nearly-periodic systems play a special role because they display perhaps the simplest possible non-dissipative short-timescale dynamics. They therefore provide a useful proving ground for analytical and numerical methods aimed at more complex multi-scale models.

Remark 2.3 In a paper[9] on basic properties of continuous-time nearly-periodic systems, Kruskal assumed that R_0 is nowhere vanishing, in addition to requiring that ω_0 is sign-definite. This assumption is usually not essential and it is enough to require that ω_0 vanishes nowhere. This is an important restriction to lift since many interesting circle actions have fixed points.

It can be shown that every nearly-periodic system admits an approximate $U(1)$-symmetry[9], known as the *roto-rate*, that is determined to leading order by the unperturbed periodic dynamics:

Definition 2.7 A **roto-rate** for a nearly-periodic system X_ε on a manifold M is a formal power series $R_\varepsilon = R_0 + \varepsilon R_1 + \varepsilon^2 R_2 + \ldots$ with vector field coefficients such that R_0 is equal to the limiting roto-rate and the following equalities hold in the sense of formal series:

$$\exp(2\pi \mathcal{L}_{R_\varepsilon}) = 1 \qquad \text{and} \qquad [X_\varepsilon, R_\varepsilon] = 0.$$

Proposition 2.1 (Kruskal[9]) *Every nearly-periodic system admits a unique roto-rate R_ε.*

A subtle argument allows to upgrade leading-order $U(1)$-invariance to all-orders $U(1)$-invariance for integral invariants:

Proposition 2.2 (Burby et al.[11]) *Let α_ε be a smooth ε-dependent differential form on a manifold M. Suppose α_ε is an absolute integral invariant for a smooth nearly-periodic system X_ε on M. If $\mathcal{L}_{R_0}\alpha_0 = 0$ then $\mathcal{L}_{R_\varepsilon}\alpha_\varepsilon = 0$, where R_ε is the roto-rate for X_ε.*

Nearly-periodic maps. Nearly-periodic maps are natural discrete-time analogues of nearly-periodic systems, which were first introduced in[11]. The following provides a precise definition.

Definition 2.8 A **nearly-periodic map** on a manifold M with parameter vector space $\mathcal{E}$ is a smooth mapping $F : M \times \mathcal{E} \to M$ such that $F_\varepsilon : M \to M : m \mapsto F(m, \varepsilon)$ has the following properties:

- F_ε is a diffeomorphism for each $\varepsilon \in \mathcal{E}$.
- There exists a $U(1)$-action $\Phi_\theta : M \to M$ and a constant $\theta_0 \in U(1)$ such that $F_0 = \Phi_{\theta_0}$.

We say F is **resonant** if θ_0 is a rational multiple of 2π, otherwise F is **non-resonant**. The infinitesimal generator of Φ_θ, R_0, is the **limiting roto-rate**.
Example 2.1 Let X_ε be a nearly-periodic system on a manifold M with limiting roto-rate R_0 and limiting angular frequency ω_0. Assume that ω_0 is constant. For each $\varepsilon \in \mathbb{R}$ let $\mathcal{F}_t^\varepsilon$ denote the time-t flow for X_ε. The mapping $F(m, \varepsilon) = \mathcal{F}_{t_0}^\varepsilon(m)$ is nearly-periodic for each t_0. To see why, first note that the flow of the limiting vector field $X_0 = \omega_0 R_0$ is given by $\mathcal{F}_t^0(m) = \Phi_{\omega_0 t}(m)$, where Φ_θ denotes the $U(1)$-action generated by R_0. It follows that $F(m, 0) = \Phi_{\omega_0 t_0}(m) = \Phi_{\theta_0}(m)$, where $\theta_0 = \omega_0 t_0 \mod 2\pi$. This example is more general than it first appears since any nearly-periodic system can be rescaled to have a constant limiting angular frequency. Indeed if the nearly-periodic system X_ε has non-constant limiting angular frequency ω_0 then $X_\varepsilon' = X_\varepsilon / \omega_0$ is a nearly-periodic system with limiting angular frequency 1. The integral curves of X_ε' are merely time reparameterizations of integrals curves of X_ε.

Let X be a vector field on a manifold M with time-t flow map $\mathcal{F}_t$. A $U(1)$-action Φ_θ is a $U(1)$-*symmetry* for X if $\mathcal{F}_t \circ \Phi_\theta = \Phi_\theta \circ \mathcal{F}_t$, for each $t \in \mathbb{R}$ and $\theta \in U(1)$. Differentiating this condition with respect to θ at the identity implies and is implied by $\mathcal{F}_t^* R = R$, where R denotes the infinitesimal generator for the $U(1)$-action. Since we would like to think of nearly-periodic maps as playing the part of a nearly-periodic system's flow map, the latter characterization of symmetry allows us to naturally extend Kruskal's notion of roto-rate to our discrete-time setting.

Definition 2.9 A **roto-rate** for a nearly-periodic map $F : M \times \mathcal{E} \to M$ is a formal power series $R_\varepsilon = R_0 + R_1\varepsilon + R_2\varepsilon^2 + \ldots$ whose coefficients are vector fields on M such that R_0 is the limiting roto-rate and the following equalities hold in the sense of formal power series: $F_\varepsilon^* R_\varepsilon = R_\varepsilon$ and $\exp(2\pi \mathcal{L}_{R_\varepsilon}) = 1$.

A first fundamental result concerning nearly-periodic maps establishes the existence and uniqueness of the roto-rate in the non-resonant case. Like the corresponding result in continuous time, this result holds to all orders in perturbation theory.

Theorem 2.2 (Burby et al.[11]) *Each non-resonant nearly-periodic map admits a unique roto-rate.*

Thus, non-resonant nearly-periodic maps formally reduce to mappings on the space of $U(1)$-orbits, corresponding to the elimination of a single dimension in phase space.

Nearly-periodic systems and maps with a hamiltonian structure.
Definition 2.10 A ε-**dependent presymplectic manifold** is a manifold M equipped with a smooth ε-dependent 2-form Ω_ε such that $\mathbf{d}\Omega_\varepsilon = 0$ for each $\varepsilon \in \mathcal{E}$. We say (M, Ω_ε) is **exact** when there is a smooth ε-dependent 1-form ϑ_ε such that $\Omega_\varepsilon = -\mathbf{d}\vartheta_\varepsilon$.
Definition 2.11 A **nearly-periodic Hamiltonian system** on an exact presymplectic manifold (M, Ω_ε) is a nearly-periodic system X_ε on M such that $\iota_{X_\varepsilon}\Omega_\varepsilon = \mathbf{d}H_\varepsilon$, for some smooth ε-dependent function $H_\varepsilon : M \to \mathbb{R}$.

We already know from Proposition 2.1 that every nearly-periodic system admits a unique roto-rate R_ε. In the Hamiltonian setting, it can be shown that both the dynamics and the Hamiltonian structure are $U(1)$-invariant to all orders in ε.

Proposition 2.3 (Kruskal[9], Burby et al.[11]) *The roto-rate R_ε for a nearly-periodic Hamiltonian system X_ε on an exact presymplectic manifold (M, Ω_ε) with Hamiltonian H_ε satisfies $\mathcal{L}_{R_\varepsilon} H_\varepsilon = 0$, and $\mathcal{L}_{R_\varepsilon} \Omega_\varepsilon = 0$ in the sense of formal power series.*

According to Noether's celebrated theorem, a Hamiltonian system that admits a continuous family of symmetries also admits a corresponding conserved quantity[57,60,61]. Therefore one might expect that a Hamiltonian system with an approximate symmetry must also have an approximate conservation law. This is indeed the case for nearly-periodic Hamiltonian systems:

Proposition 2.4 (Burby et al.[11]) *Let X_ε be a nearly-periodic Hamiltonian system on the exact presymplectic manifold (M, Ω_ε). Let R_ε be the associated roto-rate. There is a formal power series $\theta_\varepsilon = \theta_0 + \varepsilon\, \theta_1 + \dots$ with coefficients in $\Omega^1(M)$ such that $\Omega_\varepsilon = -\mathbf{d}\theta_\varepsilon$ and $\mathcal{L}_{R_\varepsilon} \theta_\varepsilon = 0$. Moreover, the formal power series $\mu_\varepsilon = \iota_{R_\varepsilon} \theta_\varepsilon$ is a constant of motion for X_ε to all orders in perturbation theory. In other words, $\mathcal{L}_{X_\varepsilon} \mu_\varepsilon = 0$, in the sense of formal power series. The formal constant of motion μ_ε is the **adiabatic invariant** associated with the nearly-periodic Hamiltonian system.*

Note that general expressions for the adiabatic invariant μ_ε can be obtained[62]. It can also be shown that the (formal) set of fixed points for the roto-rate is an elliptic almost invariant slow manifold whose normal stability is mediated by the adiabatic invariant associated with the nearly-periodic Hamiltonian system[8].

A similar theory can be established for nearly-periodic maps with a Hamiltonian structure.

Definition 2.12 A **presymplectic nearly-periodic map** on a ε-dependent presymplectic manifold (M, Ω_ε) is a nearly-periodic map F such that $F_\varepsilon^* \Omega_\varepsilon = \Omega_\varepsilon$ for each $\varepsilon \in \mathcal{E}$.

Theorem 2.3 (Burby et al.[11]) *If F is a non-resonant presymplectic nearly-periodic map on a ε-dependent presymplectic manifold (M, Ω_ε) with roto-rate R_ε then $\mathcal{L}_{R_\varepsilon} \Omega_\varepsilon = 0$.*

Definition 2.13 A **Hamiltonian nearly-periodic map** on a ε-dependent presymplectic manifold (M, Ω_ε) is a nearly-periodic map F such that there is a smooth (t, ε)-dependent vector field $Y_{t,\varepsilon}$ with $t \in \mathbb{R}$ such that the following properties hold true:

- $\iota_{Y_{t,\varepsilon}} \Omega_\varepsilon = \mathbf{d}H_{t,\varepsilon}$, for some smooth (t, ε)-dependent function $H_{t,\varepsilon}$.
- For each $\varepsilon \in \mathcal{E}$, F_ε is the $t = 1$ flow of $Y_{t,\varepsilon}$.

Lemma 2.2 *Each Hamiltonian nearly-periodic map is a presymplectic nearly-periodic map.*

Using presymplecticity of the roto-rate, Noether's theorem can be used to establish existence of adiabatic invariants for many interesting presymplectic nearly-periodic maps.

Theorem 2.4 (Burby et al.[11]) *Let F be a non-resonant presymplectic nearly-periodic map on the exact ε-dependent presymplectic manifold (M, Ω_ε) with roto-rate R_ε. Assume that F is Hamiltonian or that the manifold M is connected and the limiting roto rate R_0 has at least one zero. Then there exists a smooth ε-dependent 1-form θ_ε such that $\mathcal{L}_{R_\varepsilon} \theta_\varepsilon = 0$ and $-\mathbf{d}\theta_\varepsilon = \Omega_\varepsilon$ in the sense of formal power series. Moreover the quantity $\mu_\varepsilon = \iota_{R_\varepsilon} \theta_\varepsilon$ satisfies $F_\varepsilon^* \mu_\varepsilon = \mu_\varepsilon$ in the sense of formal power series, that is, μ_ε is an adiabatic invariant for F.*

When an adiabatic invariant exists, the phase-space dimension is formally reduced by two. On the slow manifold $\mu_\varepsilon = 0$ the reduction in dimensionality may be even more dramatic. For example, the slow manifold for the symplectic Lorentz system[8] has half the dimension of the full system.

Novel structure-preserving neural network architectures

Approximating nearly-periodic maps via gyroceptrons. We first consider the problem of approximating an arbitrary nearly-periodic map $P : M \times \mathcal{E} \to M$ on a manifold M. From Definition 2.8, there must be a corresponding circle action $\Phi_\theta : M \to M$ and $\theta_0 \in U(1)$ such that $P_0 = \Phi_{\theta_0}$. Consider the map $I_\varepsilon : M \to M$ given by

$$I_\varepsilon = P_\varepsilon \circ \Phi_{\theta_0}^{-1} \qquad \forall \varepsilon \in \mathcal{E}. \tag{3.1}$$

This defines a near-identity map on M satisfying $I_0 = \mathrm{Id}_M$. By composing both sides of Eq. (3.1) on the right by Φ_{θ_0}, we obtain a representation for any nearly-periodic map P as the composition of a near-identity map and a circle action,

$$P_\varepsilon = I_\varepsilon \circ \Phi_{\theta_0} \qquad \forall \varepsilon \in \mathcal{E}. \tag{3.2}$$

As a consequence, if we can approximate any near-identity map and any circle action, then by the above representation we can approximate any nearly-periodic map.

Different circle actions can act on manifolds in topologically different ways, so it would be very challenging, if not impossible, to construct a single strategy which allows to approximate any circle action to arbitrary accuracy. Here, we will consider the simpler case where we assume that we know a priori the topological type of action for the nearly-periodic system, and work within conjugation classes. Conjugation of a circle action $\Phi_\theta : M \to M$ with a diffeomorphism ψ results in the map $\psi \circ \Phi_\theta \circ \psi^{-1}$, and two circle actions belong to the same conjugation class if one can be written as the conjugation with a diffeomorphism of the other one. Note that although compositions of nearly-periodic maps are not necessarily nearly-periodic, the map obtained by conjugation of a nearly-periodic map with a diffeomorphism is nearly-periodic:

Lemma 3.1 *Let $P : M \times \mathcal{E} \to M$ be a nearly-periodic map on a manifold M, and let $\psi : M \to M$ be a diffeomorphism on M. Then the map $\tilde{P} : M \times \mathcal{E} \to M$ defined for any $\varepsilon \in \mathcal{E}$ via*

$$\tilde{P}_\varepsilon \equiv \psi \circ P_\varepsilon \circ \psi^{-1} \tag{3.3}$$

is a nearly-periodic map.

Proof ψ and P_ε are diffeomorphisms for any $\varepsilon \in \mathcal{E}$ so $\tilde{P}_\varepsilon$ is also a diffeomorphism for any $\varepsilon \in \mathcal{E}$. Now, from Definition 2.8, there is a circle action $\Phi_\theta : M \to M$ and $\theta_0 \in U(1)$ such that $P_0 = \Phi_{\theta_0}$. Define $\tilde{\Phi}_\theta : M \to M$ via $\tilde{\Phi}_\theta \equiv \psi \circ \Phi_\theta \circ \psi^{-1}$ for any $\theta \in U(1)$. Then, for any $\theta, \theta_1, \theta_2 \in U(1)$,

- $\tilde{\Phi}_{\theta+2\pi} = \psi \circ \Phi_{\theta+2\pi} \circ \psi^{-1} = \psi \circ \Phi_\theta \circ \psi^{-1} = \tilde{\Phi}_\theta$

- $\tilde{\Phi}_0 = \psi \circ \Phi_0 \circ \psi^{-1} = \psi \circ \mathrm{Id}_M \circ \psi^{-1} = \mathrm{Id}_M$

- $\tilde{\Phi}_{\theta_1} \circ \tilde{\Phi}_{\theta_2} = \psi \circ \Phi_{\theta_1} \circ \psi^{-1} \circ \psi \circ \Phi_{\theta_2} \circ \psi^{-1} = \psi \circ \Phi_{\theta_1} \circ \Phi_{\theta_2} \circ \psi^{-1}$

  $= \psi \circ \Phi_{\theta_1+\theta_2} \circ \psi^{-1} = \tilde{\Phi}_{\theta_1+\theta_2}$

Therefore, $\tilde{\Phi}_\theta$ is a circle action, and $\theta_0 \in U(1)$ is such that $\tilde{\Phi}_{\theta_0} = \psi \circ \Phi_{\theta_0} \circ \psi^{-1} = \psi \circ P_0 \circ \psi^{-1} = \tilde{P}_0$.

As a consequence, $\tilde{P}$ is a nearly-periodic map. □

We also have the following useful factorization result for nearly-periodic maps with limiting rotation within a given conjugacy class:

Lemma 3.2 *Let $\Phi_\theta : M \to M$ be a circle action on a manifold M. Every nearly-periodic map $P_\varepsilon : M \to M$ whose limiting rotation $\Phi'_{\theta_0} = P_0$ is conjugate to Φ_{θ_0} admits the decomposition*

$$P_\varepsilon = I_\varepsilon \circ \psi \circ \Phi_{\theta_0} \circ \psi^{-1}, \tag{3.4}$$

where $\psi : M \to M$ is a diffeomorphism and $I_\varepsilon : M \to M$ is a near-identity diffeomorphism.

We will thus assume that we know in advance the topological type of the circle action Φ_θ for the dynamics of interest, and then propose to learn the nearly-periodic map P_ε by learning each component map in the composition

$$P_\varepsilon = I_\varepsilon \circ \psi \circ \Phi_\theta \circ \psi^{-1}. \tag{3.5}$$

This formula may be interpreted intuitively as follows. The map ψ learns the mode structure of an oscillatory system's short timescale dynamics. The circle action Φ_θ provides an aliased phase advance for the learnt mode. Finally, I_ε captures the averaged dynamics that occurs on timescales much larger than the limiting oscillation period.

I_ε and ψ can be learnt using any standard neural network architecture, as long as the near-identity property is enforced in the representation for I_ε. It is however important to invert ψ exactly, and this strongly motivates using explicitly invertible neural network architectures for ψ. It has been shown that those coupling-based invertible neural networks are universal diffeomorphism approximators[63]. The parameter θ in the circle action Φ_θ can also be considered as a trainable parameter. We will refer to the resulting architecture as a **gyroceptron**, named after a combination of gyrations of phase with perceptron.

Definition 3.1 A **gyroceptron** is a feed-forward neural network

$$P_\varepsilon[W] = I_\varepsilon[W_I] \circ \psi[W_\psi] \circ \Phi_\theta \circ \psi[W_\psi]^{-1} \tag{3.6}$$

with weights $W = (W_I, W_\psi)$ and rotation parameter $\theta \in U(1)$, where

- $I_\varepsilon[W_I] : M \to M$ is a diffeomorphism for each (ε, W_I) such that $I_0[W_I] = \mathrm{Id}_M$ for each W_I
- $\psi[W_\psi] : M \to M$ is a diffeomorphism for each W_ψ
- $\Phi_\theta : M \to M$ is a circle action on M

Gyroceptrons enjoy the following universal approximation property.

Theorem 3.1 *Fix a circle action $\Phi_\theta : M \to M$ and a compact set $C \subset M$. Let $P_\varepsilon : M \to M$ be a nearly-periodic map whose limiting rotation is conjugate to Φ_θ. Let $\psi[W_\psi] : M \to M$ be a feed-forward network architecture that provides a universal approximation within the class of diffeomorphisms, and let $I_\varepsilon[W_I]$ be a feed-forward network architecture that provides a universal approximation within the class of ε-dependent diffeomorphisms with $I_0[W] = \mathrm{Id}_M$. For each $\delta > 0$, there exist weights W_ψ^* and W_I^* such that the gyroceptron $P_\varepsilon[W^*] = I_\varepsilon[W_I^*] \circ \psi[W_\psi^*] \circ \Phi_\theta \circ \psi[W_\psi^*]^{-1}$ approximates P_ε within δ on C.*

Approximating nearly-periodic symplectic maps via symplectic gyroceptrons. We now focus on approximating an arbitrary nearly-periodic symplectic map $P : M \times \mathcal{E} \to M$ on a manifold M. We will restrict our attention to symplectic manifolds with ε-independent symplectic forms (the ε-dependent case is more subtle and will not be pursued in the current study). From Definition 2.8, there must be a corresponding symplectic circle action $\Phi_\theta : M \to M$ and $\theta_0 \in U(1)$ such that $P_0 = \Phi_{\theta_0}$. As before, consider the map $I_\varepsilon : M \to M$ given by

$$I_\varepsilon = P_\varepsilon \circ \Phi_{\theta_0}^{-1}, \qquad \forall \varepsilon \in \mathcal{E}. \tag{3.7}$$

Now, the inverse of a symplectic map is symplectic and any composition of symplectic maps is also symplectic. Thus, the map $\Phi_{\theta_0}^{-1} = P_0^{-1}$ is symplectic, and as a result, I_ε is symplectic on M for any $\varepsilon \in \mathcal{E}$ and it satisfies the near-identity property $I_0 = \mathrm{Id}_M$. By composing both sides of Eq. (3.7) on the right by Φ_{θ_0}, we obtain a representation for any nearly-periodic symplectic map P as the composition of a near-identity symplectic map and a symplectic circle action:

$$P_\varepsilon = I_\varepsilon \circ \Phi_{\theta_0}, \qquad \forall \varepsilon \in \mathcal{E}. \tag{3.8}$$

Lemma 3.3 *Let $\Phi_\theta : M \to M$ be a symplectic circle action on a symplectic manifold (M, ω). Every nearly-periodic symplectic map $P_\varepsilon : M \to M$ whose limiting rotation $\Phi_{\theta_0}' = P_0$ is conjugate to Φ_{θ_0} admits the decomposition*

$$P_\varepsilon = I_\varepsilon \circ \psi \circ \Phi_{\theta_0} \circ \psi^{-1}, \tag{3.9}$$

where $\psi : M \to M$ is a symplectic diffeomorphism and $I_\varepsilon : M \to M$ is a near-identity symplectic diffeomorphism.

If we can approximate any near-identity symplectic map and any symplectic circle action, then by the above representation we can approximate any nearly-periodic symplectic map. As before, we will assume that we know a priori the topological type of the circle action Φ_θ for the nearly-periodic symplectic system of interest, and work within conjugation classes. Since compositions of symplectic maps are symplectic, Lemma 3.1 implies that the map $\psi \circ P \circ \psi^{-1}$, obtained by conjugating a nearly-periodic symplectic map P with a symplectomorphism ψ (i.e. a symplectic diffeomorphism), is also a nearly-periodic symplectic map. We will then learn the nearly-periodic symplectic map by learning each component map in the composition

$$P_\varepsilon = I_\varepsilon \circ \psi \circ \Phi_\theta \circ \psi^{-1}, \tag{3.10}$$

where I_ε is a near-identity symplectic map and ψ is symplectic.

The symplectic map ψ can be learnt using any neural network architecture which strongly enforces symplecticity. It is preferable however to choose an architecture which can easily be inverted, so that the computations involving ψ^{-1} can be conducted efficiently. The near-identity symplectic map I_ε can be learnt using any neural network architecture strongly enforcing symplecticity with the additional property that it limits to the identity as ε goes to 0. The parameter θ in the circle action Φ_θ can also be considered as a trainable parameter. We will refer to any such resulting composition of neural network architectures as a **symplectic gyroceptron**.

Definition 3.2 A **symplectic gyroceptron** is a feed-forward neural network

$$P_\varepsilon[W] = I_\varepsilon[W_I] \circ \psi[W_\psi] \circ \Phi_\theta \circ \psi[W_\psi]^{-1} \tag{3.11}$$

with weights $W = (W_I, W_\psi)$ and rotation parameter $\theta \in U(1)$, where

- $I_\varepsilon[W_I] : M \to M$ is a symplectic diffeomorphism for each (ε, W_I) such that $I_0[W_I] = \mathrm{Id}_M$ for each W_I
- $\psi[W_\psi] : M \to M$ is a symplectic diffeomorphism for each W_ψ
- $\Phi_\theta : M \to M$ is a symplectic circle action on M

Symplectic gyroceptrons enjoy a universal approximation property comparable to the non-symplectic case.

Theorem 3.2 *Fix a symplectic circle action $\Phi_\theta : M \to M$ on the symplectic manifold (M, ω) and a compact set $C \subset M$. Let $P_\varepsilon : M \to M$ be a nearly-periodic symplectic map whose limiting rotation is conjugate to Φ_θ. Let $\psi[W_\psi] : M \to M$ be a feed-forward network architecture that provides a universal approximation within the class of symplectic diffeomorphisms, and let $I_\varepsilon[W_I]$ be a feed-forward network architecture that provides a universal approximation within the class of ε-dependent symplectic diffeomorphisms with $I_0[W] = \mathrm{Id}_M$. For each $\delta > 0$, there*

exist weights W_ψ^* and W_I^* such that the symplectic gyroceptron $P_\varepsilon[W^*] = I_\varepsilon[W_I^*] \circ \psi[W_\psi^*] \circ \Phi_\theta \circ \psi[W_\psi^*]^{-1}$ approximates P_ε within δ on C.

In this paper, we will use HénonNets[2] as the main building blocks of our symplectic gyroceptrons. The symplectic map ψ will be learnt using a standard HénonNet (see Definition 2.3), its inverse ψ^{-1} can be obtained easily by composing inverses of Hénon-like maps (see Remark 2.2), and the near-identity symplectic map I_ε will be learnt using a near-identity HénonNet (see Definition 2.5). The neural network architectures considered in this paper are summarized in Figure 1.

We would like to emphasize that symplectic building blocks other than HénonNets could have been used as the basis for our symplectic gyroceptrons. For instance, a possible option would have been to use SympNets[3] since they also strongly ensure symplecticity and enjoy a universal approximation property for symplectic maps. However, numerical experiments conducted in the original HénonNet paper[2] suggested that HénonNets have a higher per layer expressive power than SympNets, and as a result SympNets are typically much deeper than HénonNets, and slower for prediction. This is consistent with the observations we will make later in Sect. "Nonlinearly coupled oscillators" where a SympNet takes 127 seconds to generate trajectories that were generated by a HénonNet of similar size in 3 seconds. Together with the fact that SympNets are not as easily invertible as HénonNets, the computational advantage of HénonNets makes them more desirable as building blocks than SympNets.

Numerical confirmation of the existence of adiabatic invariants

In this section, we will confirm numerically that for any random set of weights and bias, the dynamical system generated by the symplectic gyroceptron

$$I_\varepsilon \circ \psi \circ \Phi_\theta \circ \psi^{-1}, \tag{4.1}$$

introduced in Sect. "Approximating nearly-periodic symplectic maps via symplectic gyroceptrons", admits an adiabatic invariant.

In our numerical experiments, we take the circle action given by the clockwise rotation

$$\mathcal{R}_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{4.2}$$

The quantity $\mathfrak{I}_0(q,p) = \frac{1}{2}q^2 + \frac{1}{2}p^2$ is an invariant of the dynamics associated to the circle action (4.2), and as a result

$$\mu = \mathfrak{I}_0 \circ \psi^{-1} \tag{4.3}$$

is an invariant of the dynamics associated to the composition $\psi \circ \Phi_\theta \circ \psi^{-1}$, and an adiabatic invariant of the dynamics associated to the symplectic gyroceptron (4.1).

Figure 2 displays the evolution of the adiabatic invariant (4.3) over 10000 iterations of the dynamical system generated by the symplectic gyroceptron (4.1), for different values of ε. Here, ψ is a HénonNet and I_ε a near-identity HénonNet, both with 3 Hénon layers, each of which has 8 neurons in its single-hidden-layer fully-connected neural networks layer potential. We can clearly see that the conservation of the adiabatic invariant gets significantly better as ε gets closer to 0, going from chaotic oscillations of large amplitude when $\varepsilon = 0.1$ to very regular oscillations of minute amplitude when $\varepsilon = 10^{-8}$.

We investigated further by obtaining the number of iterations needed for the adiabatic invariant μ to deviate significantly from its original value μ_0 as ε is varied. More precisely, given a value of ε, we search for the smallest integer $N(\varepsilon)$ such that
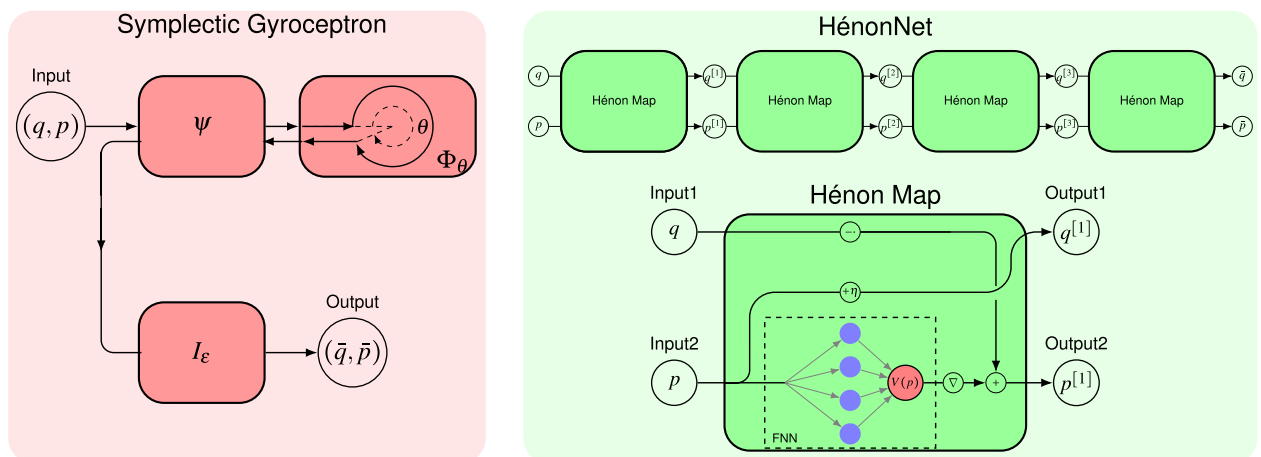

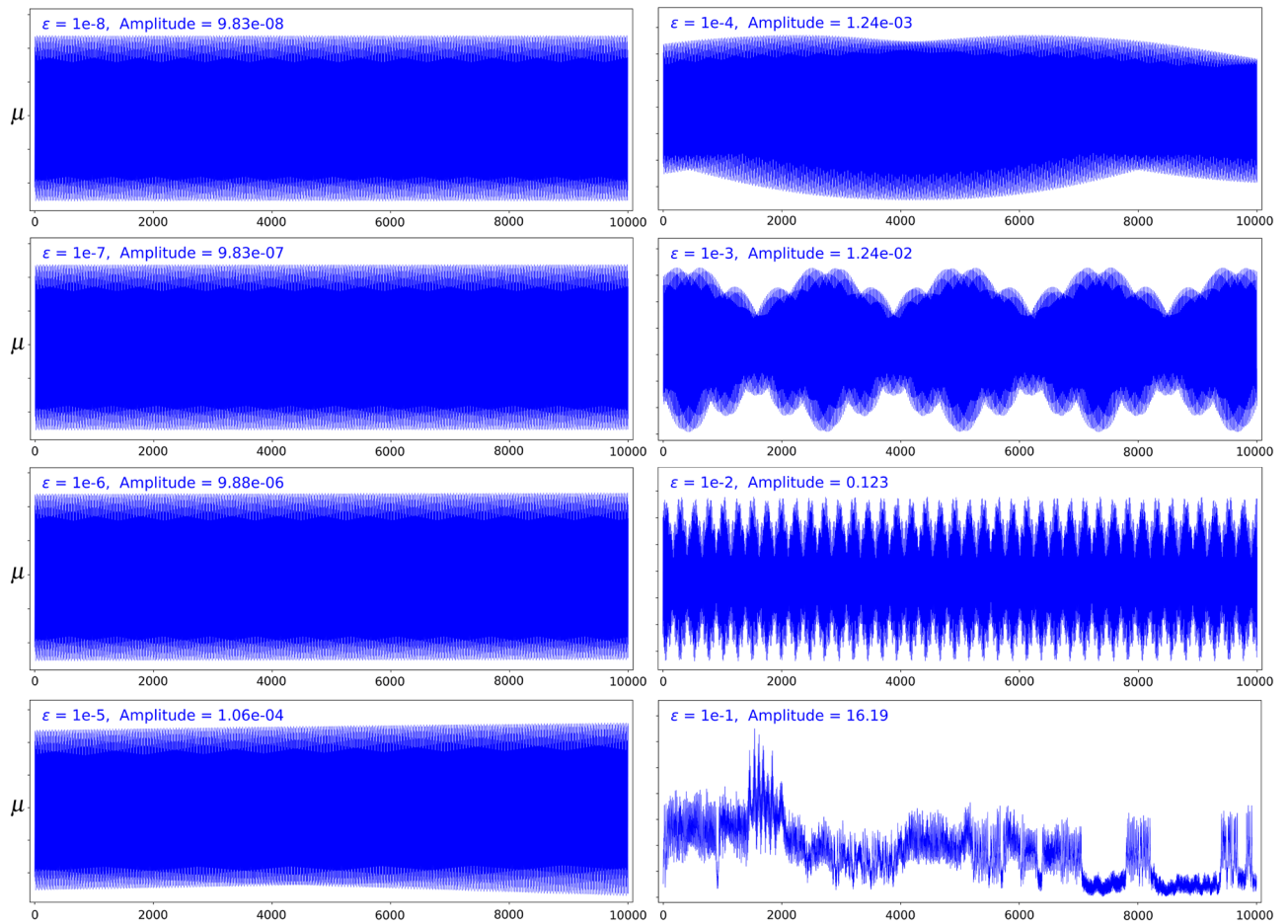**Figure 1.** Network diagrams. Left: Symplectic Gyroceptron. Right: Hénon Network.

Figure 2. Conservation of the adiabatic invariant (4.3) over 10000 iterations for the map generated by the symplectic gyroceptron (4.1) as ε is increased.

$$|\mu_{N(\varepsilon)} - \mu_0| > \rho \max_{k=0,\dots,K(\varepsilon)} |\mu_k - \mu_0|, \quad \text{where } K(\varepsilon) = \lfloor 10 + \varepsilon^{-1/4} \rfloor. \tag{4.4}$$

In other words, we record the first iteration where the value of the adiabatic invariant μ deviates from its original value μ_0 by more than some constant factor $\rho > 1$ of the maximum deviations experienced in the first few $K(\varepsilon)$ iterations. The results are plotted in Figure 3 for $\rho = 1.1$.

We can clearly see from Figure 3 that $N(\varepsilon)$, the number of iterations needed for the adiabatic invariant μ to deviate from its original value μ_0 by more than $\rho = 1.1$ times the maximum deviations experienced in the first few iterations, increases sharply as ε gets closer to 0. This is consistent with theoretical expectations. Note that using higher values of ρ and smaller values of ε would probably generate more interesting and meaningful results. Unfortunately, this is not computationally realizable since $N(\varepsilon)$ becomes very large when ρ is increased beyond 1.2. Even for larger values of ε, computing a single point would take several days.
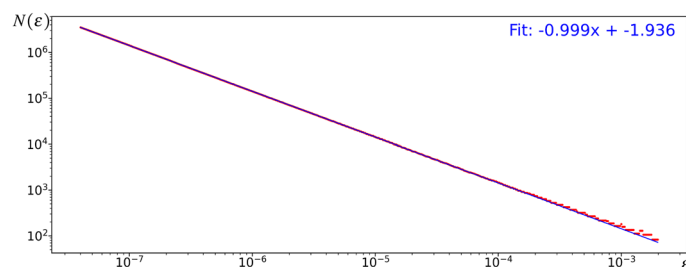
Figure 3. $N(\varepsilon)$ as a function of ε for $\rho = 1.1$ and a random set of weights.

Numerical examples of learning surrogate maps

Nonlinearly coupled oscillators. In this section, we use the symplectic gyroceptron architecture introduced in Sect. "Approximating nearly-periodic symplectic maps via symplectic gyroceptrons" to learn a surrogate map for the nearly-periodic symplectic flow map associated to a nearly-periodic Hamiltonian system composed of two nonlinearly coupled oscillators, where one of them oscillates significantly faster than the other:

$$\begin{cases} \dot{q}_1 = p_1 & \dot{p}_1 = -q_1 - \varepsilon \partial_{q_1} U(q_1, q_2) \\ \dot{q}_2 = \varepsilon p_2 & \dot{p}_2 = -\varepsilon q_2 - \varepsilon \partial_{q_2} U(q_1, q_2) \end{cases} \tag{5.1}$$

These equations of motion are the Hamilton's equations associated to the Hamiltonian

$$H_\varepsilon(q_1, q_2, p_1, p_2) = \frac{1}{2}(q_1^2 + p_1^2) + \frac{1}{2}\varepsilon(q_2^2 + p_2^2) + \varepsilon U(q_1, q_2). \tag{5.2}$$

The $\varepsilon = 0$ dynamics are decoupled, where the first oscillator, initialized at $\big(q_1(0), p_1(0)\big) = (\mathrm{II}, \sqrt{})$, follows a trajectory characterized by periodic clockwise circular rotation in phase space, while the second oscillator remains immobile:

$$q_1(t) = q \cos t + p \sin t, \quad p_1(t) = p \cos t - q \sin t. \tag{5.3}$$

Thus, this is a nearly-periodic Hamiltonian system on $\mathbb{R}^4$ with associated $\varepsilon = 0$ circle action given by the clockwise rotation

$$\mathcal{R}_\theta = \begin{pmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We will use the nonlinear coupling potential $U(q_1, q_2) = q_1 q_2 \sin(2q_1 + 2q_2)$ in our numerical experiments since the resulting nearly-periodic Hamiltonian system displays complicated dynamics as the value of ε is increased from 0. We have plotted in Figure 4 a few trajectories of this dynamical system corresponding to different values of ε.

To learn a surrogate map for the nearly-periodic symplectic flow map associated to this nearly-periodic Hamiltonian system, we use the symplectic gyroceptron $I_\varepsilon \circ \psi \circ \Phi_\theta \circ \psi^{-1}$ introduced in Sect. "Approximating nearly-periodic symplectic maps via symplectic gyroceptrons". In our first numerical experiments, $\varepsilon = 0.01$, θ is a trainable parameter, ψ is a HénonNet with 10 Hénon layers each of which has 8 neurons in its single-hidden-layer fully-connected neural networks layer potential, and I_ε is a near-identity HénonNet with 8 Hénon layers each of which has 6 neurons in its single-hidden-layer fully-connected neural network layer potential.

The resulting symplectic gyroceptron of 549 trainable parameters was trained for a few thousands epochs on a dataset of 20,000 updates $(q_1, q_2, p_1, p_2) \mapsto (\tilde{q}_1, \tilde{q}_2, \tilde{p}_1, \tilde{p}_2)$ of the time-0.05 flow map associated to the nearly-periodic Hamiltonian system (5.1). The training data was generated using the classical Runge–Kutta 4 integrator with very small time-steps, and the Mean Squared Error was used as the loss function in the training. Figure 5 shows the dynamics predicted by the symplectic gyroceptron for seven different initial conditions with the same initial values of (q_1, p_1) against the reference trajectories generated by the classical Runge–Kutta 4 integrator with very small time-steps. We only display the trajectories of the second oscillator since the motion of the first oscillator follows a simple nearly-circular curve.

We can see that the dynamics learnt by the symplectic gyroceptron match almost perfectly the reference trajectories and follow the level sets of the averaged Hamiltonian $\bar{H} = \frac{1}{2\pi} \int_0^{2\pi} \Phi_\theta^* H \, d\theta$, which is given by

$$\bar{H}(q_2, p_2) = \frac{1}{2}(q_2^2 + p_2^2) + \frac{1}{2\pi} \int_0^{2\pi} U(q_1(t), q_2) dt \tag{5.4}$$

$$= \frac{1}{2}(q_2^2 + p_2^2) + \frac{q_2}{2\pi} \int_0^{2\pi} \big(q \cos t + p \sin t\big) \sin\big(2[q \cos t + p \sin t] + 2q_2\big) dt \tag{5.5}$$

$$= \frac{1}{2}(q_2^2 + p_2^2) + q_2 \cos(2q_2) \sqrt{q^2 + p^2} \, \mathcal{J}_1\bigg(\sqrt{q^2 + p^2}\bigg) \tag{5.6}$$

where $\mathcal{J}_1(x)$ is the first-order Bessel function of the first kind, up to an unimportant constant. Using Kruskal's theory of nearly-periodic systems, it is straightforward to show that this averaged Hamiltonian is the leading-order approximation of the Hamiltonian for the formal $U(1)$-reduction of the two-oscillator system.

We also learned a surrogate map for the nearly-periodic symplectic time-5 flow map associated to the dynamical system (5.1), using a symplectic gyroceptron where $\varepsilon = 0.01$, θ is a trainable parameter, and ψ and I_ε both have 10 Hénon layers each of which has 8 neurons in its single-hidden-layer fully-connected neural network layer potential. This symplectic gyroceptron of 681 trainable parameters was trained for a few thousands epochs on a dataset of 60,000 updates $(q_1, q_2, p_1, p_2) \mapsto (\tilde{q}_1, \tilde{q}_2, \tilde{p}_1, \tilde{p}_2)$. For comparison, we also trained a HénonNet[2] and a SympNet[3] of similar sizes and ran simulations from the same seven different initial conditions. The HénonNet used has 16 layers each of which has 10 neurons in its single-hidden-layer fully-connected neural network layer
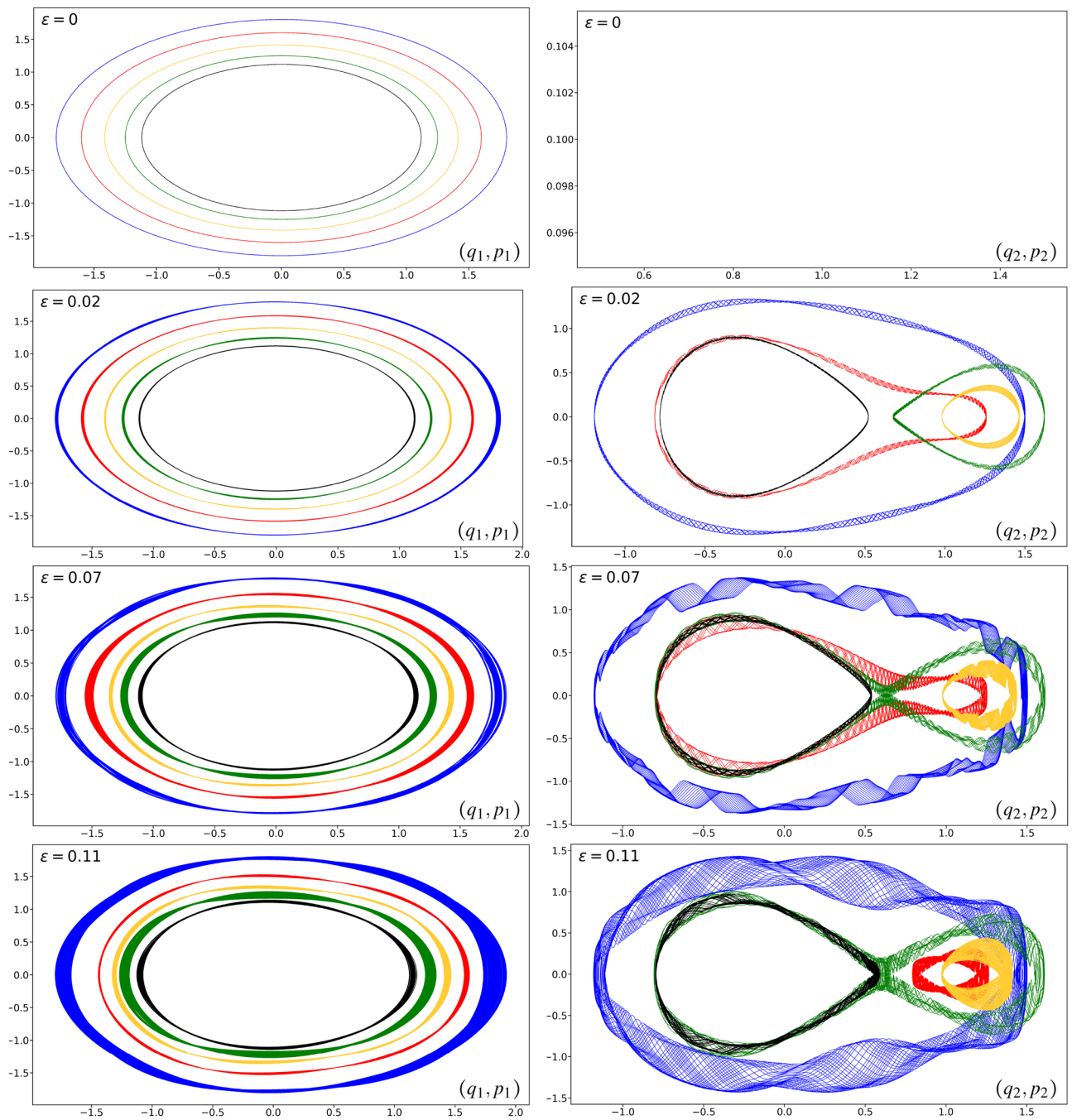
Figure 4. Sample trajectories in (q_1, p_1) and (q_2, p_2) phase spaces (left column: first oscillator, right column: second oscillator) for the nearly-periodic Hamiltonian system (Nonlinearly coupled oscillators) as the value of the parameter ε is increased.

potential, for a total of 672 trainable parameters. The SympNet[3] used has 652 trainable parameters in a network structure of the form $\mathcal{L}_n^{(k+1)} \circ (\mathcal{N}_{\text{up/low}} \circ \mathcal{L}_n^{(k)}) \circ \ldots \circ (\mathcal{N}_{\text{up/low}} \circ \mathcal{L}_n^{(1)})$, where each $\mathcal{L}_n^{(k)}$ is the composition of n trainable linear symplectic layers, and $\mathcal{N}_{\text{up/low}}$ is a non-trainable symplectic activation map.

Figure 6 shows the dynamics predicted by the symplectic gyroceptron, the HénonNet, and the SympNet, for seven different initial conditions with the same initial values of (q_1, p_1) against the reference trajectories generated by the Runge–Kutta 4 integrator (RK4) with small time-steps. As before, we only display the trajectories of the second oscillator. We can see that the dynamics predicted by the symplectic gyroceptron match the reference trajectories very well, although the predicted oscillations around the level sets of the averaged Hamiltonian are unsurprisingly larger than when learning the time-0.05 flow map. The crucial advantage that the symplectic gyroceptron offers over the other architectures considered, which only enforce the symplectic constraint, is provable existence of an adiabatic invariant. After training, the other architectures may empirically display preservation
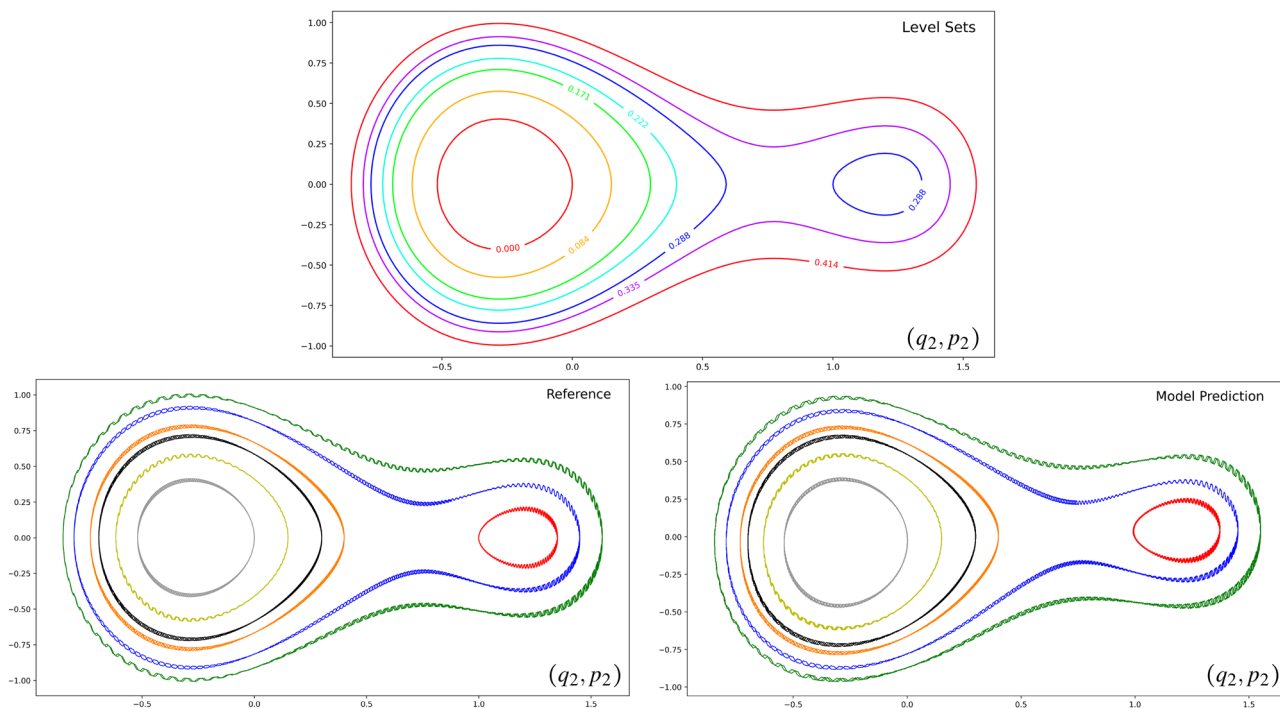
Figure 5. Level sets of the averaged Hamiltonian (5.6), and the symplectic gyroceptron predictions against the reference trajectories for the second oscillator in the nearly-periodic Hamiltonian system (5.1) with $\varepsilon = 0.01$ and a time-step of 0.05.
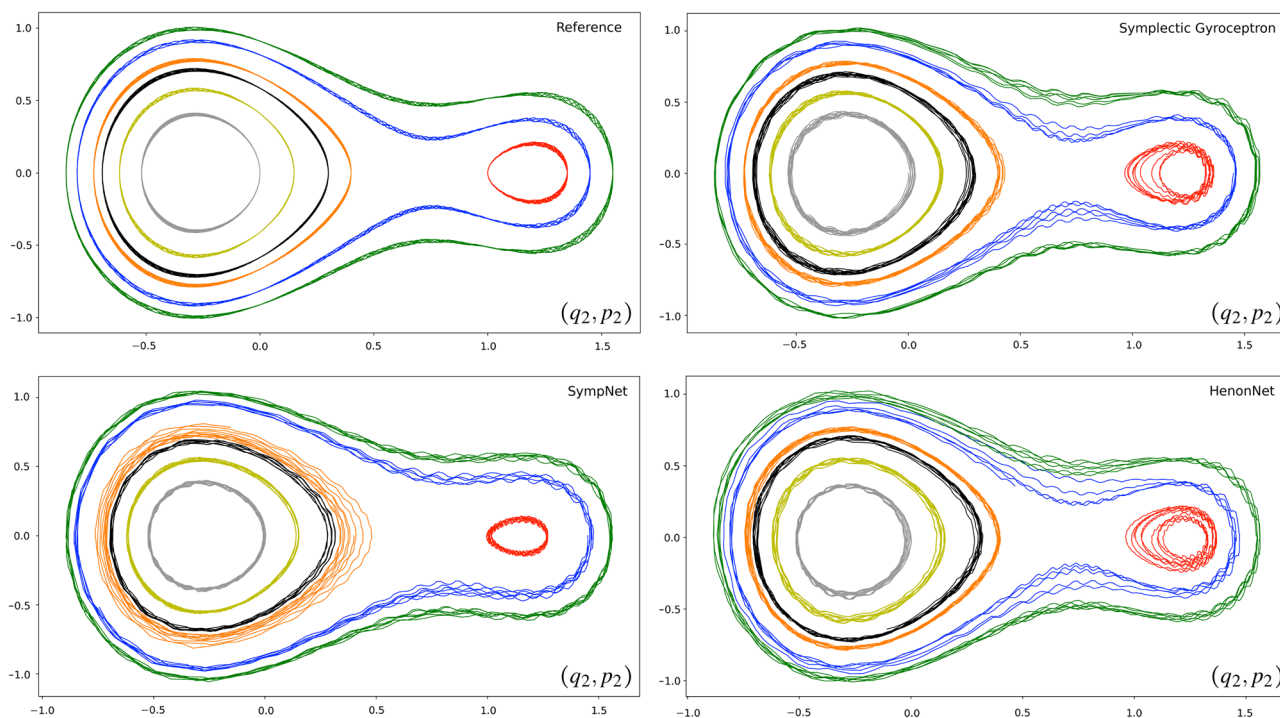

Figure 6. Predictions from a Symplectic Gyroceptron, a SympNet, and a HenonNet, against the reference trajectories for the second oscillator in the nearly-periodic Hamiltonian system (5.1) with $\varepsilon = 0.01$ and the larger time-step of 5.

of an adiabatic invariant, but this cannot be proved rigorously from first principles. In contrast, the symplectic gyroceptron enjoys provable existence of an adiabatic invariant before, during, and after training.

Note that the symplectic gyroceptron generated the seven trajectories in 5 seconds, which is several orders of magnitude faster than RK4 with small time-steps which took 6,055 seconds. The HénonNet allowed to simulate the dynamics slightly faster, in 3 seconds, while the SympNet was much slower with a running time of 127 seconds, consistently with the observations made in the original HénonNet paper[2] which motivated choosing HénonNets over SympNets in the symplectic gyroceptrons.

Charged particle interacting with its self-generated electromagnetic field.   *Problem formulation.*   Next we test the ability of symplectic gyroceptrons to function as surrogates for higher-dimension nearly-periodic systems, and for systems where the limiting circle action is not precisely known.

To formulate the ground-truth model, first fix a positive integer K and a sequence of single-variable functions $V_k : \mathbb{R} \to \mathbb{R}$, $k = 1, \ldots, K$. Consider the canonical Hamiltonian system on $\mathbb{R}^2 \times (\mathbb{R}^2)^K$ with coordinates $(q, p, Q_1, P_1, \ldots, Q_K, P_K)$, defined by the Hamiltonian

$$H_\epsilon = \frac{1}{2}\epsilon \left(p - \sum_{k=1}^{K} \sin(kq)\, Q_k \right)^2 + \frac{1}{2}\sum_{k=1}^{K} k\left([P_k - V_k(Q_k)]^2 + Q_k^2\right). \tag{5.7}$$

The equations of motion are

$$\dot{q} = \partial_p H_\epsilon = \epsilon \left(p - \sum_{\ell=1}^{K} \sin(\ell q)\, Q_\ell \right),$$

$$\dot{p} = -\partial_q H_\epsilon = \epsilon \left(p - \sum_{\ell=1}^{K} \sin(\ell q)\, Q_\ell \right) \sum_{m=1}^{K} m\, \cos(mq)\, Q_m,$$

$$\dot{Q}_k = \partial_{P_k} H_\epsilon = k\, (P_k - V_k(Q_k)),$$

$$\dot{P}_k = -\partial_{Q_k} H_\epsilon = -k\, Q_k + k\, (P_k - V_k(Q_k))\, V_k'(Q_k) + \epsilon \left(p - \sum_{\ell=1}^{K} \sin(\ell q)\, Q_\ell \right) \sin(kq).$$

These equations may be regarded as a simplified model of a charged particle (q, p) interacting with its self-generated electromagnetic field $(Q_1, P_1, \ldots, Q_K, P_K)$. We will describe the application of symplectic gyroceptrons to the development of a dynamical surrogate for this system when $\epsilon \ll 1$.

First, we verify that this Hamiltonian system is nearly-periodic, since this is the type of dynamical systems that symplectic gyroceptrons are designed to handle. So consider the limiting dynamics when $\epsilon = 0$. The equations of motion reduce to

$$\dot{q} = 0, \quad \dot{p} = 0, \quad \dot{Q}_k = \partial_{P_k} H = k(P_k - V_k(Q_k)), \quad \dot{P}_k = -\partial_{Q_k} H_\varepsilon = -kQ_k + k(P_k - V_k(Q_k))\, V_k'(Q_k). \tag{5.8}$$

While these equations of motion may appear impenetrable at first glance, the symplectic transformation of variables given by $\Lambda_0^{-1} : (q, p, Q_1, P_1, \ldots, Q_K, P_K) \mapsto (q, p, Q_1, \Pi_1, \ldots, Q_K, \Pi_K)$ where $\Pi_k = P_k - V_k(Q_k)$ simplifies them dramatically into

$$\dot{q} = 0, \qquad \dot{p} = 0, \qquad \dot{Q}_k = k\, \Pi_k, \qquad \dot{\Pi}_k = -k\, Q_k, \tag{5.9}$$

which correspond to a family (indexed by k) of harmonic oscillators with angular frequencies k. The solution map in these nice variables is therefore $\Phi_t^0(q, p, Q_1, \Pi_1, \ldots, Q_K, \Pi_K) = (q, p, Q_1(t), \Pi_1(t), \ldots, Q_K(t), \Pi_K(t))$, where

$$Q_k(t) = \cos(kt)\, Q_k + \sin(kt)\, \Pi_k, \qquad \Pi_k(t) = -\sin(kt)\, Q_k + \cos(kt)\, \Pi_k. \tag{5.10}$$

Note that Φ_t^0 is periodic with minimal period 2π. The solution map in terms of the original variables (Q_k, P_k) is therefore $\Phi_t = \Lambda_0 \circ \Phi_\theta^0 \circ \Lambda_0^{-1}$. Since Φ_t is periodic in t with minimal period 2π the ground-truth equations are Hamiltonian and nearly-periodic. The leading-order adiabatic invariant is

$$\mu_0 = \frac{1}{2}\sum_{k=1}^{K} k\, (\Pi_k^2 + Q_k^2) = \frac{1}{2}\sum_{k=1}^{K} k\, ([P_k - V_k(Q_k)]^2 + Q_k^2). \tag{5.11}$$

Symplectic gyroceptrons are therefore well-suited to surrogate modeling for this system.

Numerical experiments.   Here, we learn the nearly-periodic Hamiltonian system (5.7) in the 6-dimensional case (i.e., $K = 2$) with $V_1(Q_1) = \frac{1}{2}\sin(2Q_1)$ and $V_2(Q_2) = \frac{1}{2}\exp(-5Q_2^2)$. In our symplectic gyroceptron $I_\varepsilon \circ \psi \circ \Phi_\theta \circ \psi^{-1}$, the circle action Φ_θ is taken to be the rotation in Eq. (5.10) with θ treated as a trainable parameter, and the HénonNets ψ and I_ε both have 12 Hénon layers each of which has 8 neurons in its single-hidden-layer fully-connected neural network layer potential. The resulting architecture of 1,033 trainable parameters was trained for a few thousands epochs on a dataset of 60,000 updates $(q, p, Q_1, P_1, Q_2, P_2) \mapsto (\tilde{q}, \tilde{p}, \tilde{Q}_1, \tilde{P}_2, \tilde{Q}_1, \tilde{P}_2)$.

To verify visually that we have learnt the dynamics successfully, we select initial conditions on the zero level set of the adiabatic invariant μ_0. There, dynamics should remain on that slow manifold which is lower-dimensional and thus more easily portrayed. For the Hamiltonian system (5.7), the slow manifold is the zero level set of

$\mu_0 = 0$, which we can see from Eq. (5.11), is the set of points $(q, p, Q_1, P_1, Q_2, P_2)$ such that $Q_1 = Q_2 = 0$ and $P_1 = V_1(Q_1) = V_1(0)$, $P_2 = V_2(Q_2) = V_2(0)$.

On that slow manifold, the dynamics reduce to

$$\dot{q} = \varepsilon p, \quad \dot{p} = 0, \quad \dot{Q}_1 = 0, \quad \dot{Q}_2 = 0, \quad \dot{P}_1 = \varepsilon p \sin(q), \quad \dot{P}_2 = \varepsilon p \sin(2q), \tag{5.12}$$

where in particular the (q, p) dynamics are now independent of (Q_1, Q_2, P_1, P_2) and can easily be solved for explicitly, given some initial conditions $(q(0), p(0)) = (q, p)$:

$$q(t) = q + \varepsilon p t, \qquad p(t) = p. \tag{5.13}$$

Figures 7a,b show that the trained symplectic gyroceptron generates predictions for the evolution of q and p which remain very close to the true trajectories on the slow manifold when the initial conditions are selected on the zero level set of μ_0.

We also generate dynamics outside the zero level set of μ_0 and verify that the quantity $\mathfrak{I}_0 \circ \psi^{-1}$ matches the learnt adiabatic invariant μ_0^{learnt} along the trajectories generated by the symplectic gyroceptron $I_\varepsilon \circ \psi \circ \Phi_\theta \circ \psi^{-1}$, where

$$\mathfrak{I}_0(q, p, Q_1, \Pi_1, Q_2, \Pi_2) = \frac{1}{2} \sum_{k=1}^{K=2} k (\Pi_k^2 + Q_k^2),$$

$$\text{and} \quad \mu_0^{learnt}(q, p, Q_1, P_1, Q_2, P_2) = \frac{1}{2} \sum_{k=1}^{K=2} k ([P_k - V_k(Q_k)]^2 + Q_k^2). \tag{5.14}$$

More precisely, we check whether $\mathfrak{I}_0 \circ \psi^{-1} = \mu_0^{learnt}$ with both quantities being approximately constant along trajectories generated by the symplectic gyroceptron, where

$$\mathfrak{I}_0 \circ \psi^{-1}(q, p, Q_1, P_1, Q_2, P_2) = \frac{1}{2} \sum_{k=1}^{K=2} k (\tilde{\Pi}_k^2 + \tilde{Q}_k^2), \quad \text{with } (\tilde{q}, \tilde{p}, \tilde{Q}_1, \tilde{\Pi}_k, \tilde{Q}_2, \tilde{\Pi}_2) = \psi^{-1}(q, p, Q_1, P_1, Q_2, P_2).$$

$$\tag{5.15}$$

From Figure 7c), we see that along trajectories which are not started on the zero level set of μ_0, the value of $\mathfrak{I}_0 \circ \psi^{-1}$ remains close to the approximately constant quantity μ_0^{learnt}, although $\mathfrak{I}_0 \circ \psi^{-1}$ displays small oscillations. Since $\mathfrak{I}_0 \circ \psi^{-1}$ is an adiabatic invariant for the network, these oscillations remain bounded in amplitude for very large time intervals. The amplitude can in principle be reduced by finding a more optimal set of weights
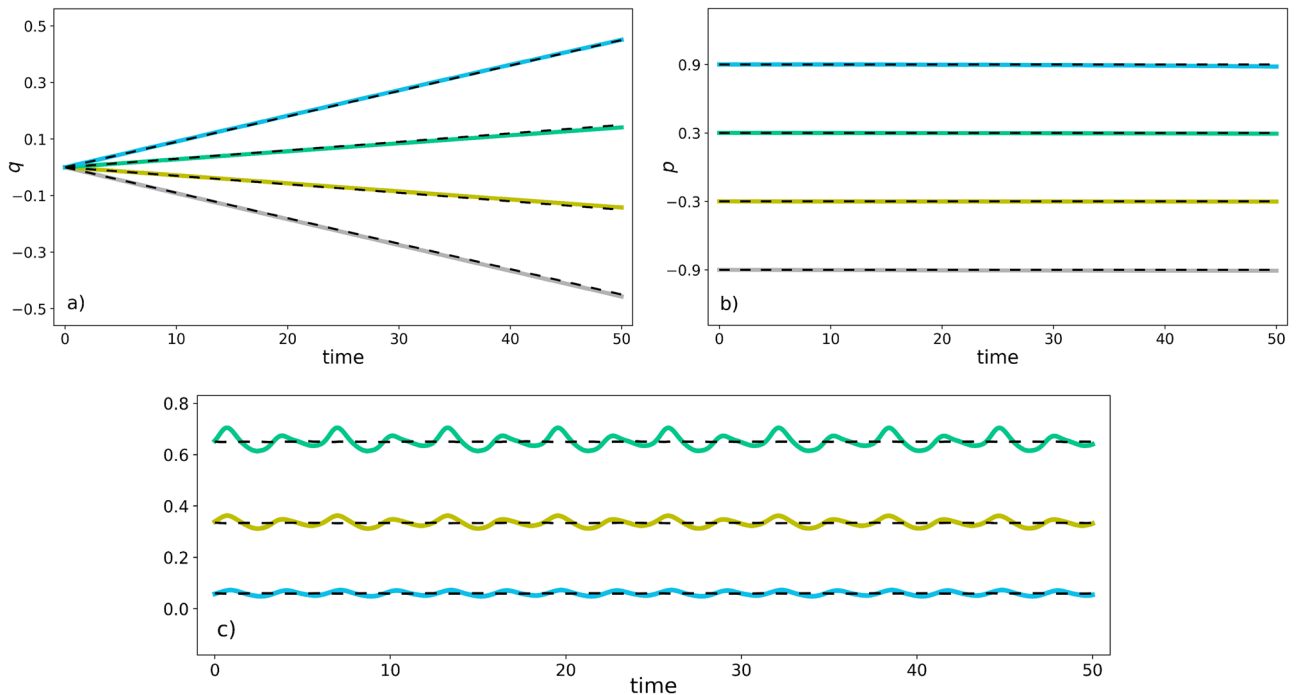

Figure 7. (**a**, **b**) Symplectic gyroceptron predictions (colors) against the true trajectories (dashed) with 4 different choices of initial conditions on the zero level set of the adiabatic invariant for the nearly-periodic Hamiltonian system (5.7) with $\varepsilon = 0.01$. **c**) Evolution of $\mathfrak{I}_0 \circ \psi^{-1}$ (colors) and μ_0^{learnt} (dashed lines) along trajectories generated by the symplectic gyroceptron with 3 different choices of initial conditions for the nearly-periodic Hamiltonian system (5.7) with $\varepsilon = 0.01$.

for the network, but it can never be reduced to zero since the true adiabatic invariant is not exactly conserved (oscillations in μ_0 are not visible at the scales displayed in the plot).

Discussion

In this paper, we have successfully constructed novel structure-preserving neural network architectures, gyroceptrons and symplectic gyroceptrons, to learn nearly-periodic maps and nearly-periodic symplectic maps, respectively. By construction, these proposed architectures define nearly-periodic maps, and symplectic gyroceptrons also preserve symplecticity. Furthermore, it was confirmed experimentally that in the symplectic case, the maps generated by the proposed symplectic gyroceptrons admit discrete-time adiabatic invariants, regardless of the values of their parameters and weights.

We also demonstrated that the proposed architectures can be effectively used in practice, by learning very precisely surrogate maps for the nearly-periodic symplectic flow maps associated to two different nearly-periodic Hamiltonian systems. Note that the hyperparameters in our architectures have not been optimized to maximize the quality of our training outcomes, and future applications of this architecture may benefit from further hyperparameter tuning.

Symplectic gyroceptrons provide a promising class of architectures for surrogate modeling of non-dissipative dynamical systems that automatically steps over short timescales without introducing spurious instabilities, and could have potential future applications for the Klein–Gordon equation in the weakly-relativistic regime, for charged particles moving through a strong magnetic field, and for the rotating inviscid Euler equations in quasi-geostrophic scaling[7]. Symplectic gyroceptrons could also be used for structure-preserving simulation of non-canonical Hamiltonian systems on exact symplectic manifolds[11], which have numerous applications across the physical sciences, for instance in modeling weakly-dissipative plasma systems[36–42].

The approach to symplectic gyroceptrons presented here targets surrogate modeling problems, where the dynamical system of interest is known but slow or expensive to simulate. In principle, symplectic gyroceptrons could also be used to discover dynamical models from observational data without detailed knowledge of the underlying dynamical system. However, in order to apply symplectic gyroceptrons effectively in this context data-mining methods must be developed for learning the topological conjugacy class of the limiting circle action. Given a topological classification of circle actions on the relevant state space (e.g. see[64] for the case of a 3-dimensional state space), a straightforward approach would be to test an ensemble of topologically-distinct circle actions for best results. A more nuanced approach would use the observed dynamics to estimate values for the classifying topological invariants of a circle action. This topological learning problem warrants further investigation.

Data availibility

References

1. Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
2. Burby, J. W., Tang, Q. & Maulik, R. Fast neural Poincaré maps for toroidal magnetic fields. *Plasma Phys. Control. Fusion* **63**, 024001. https://doi.org/10.1088/1361-6587/abcbaa (2020).
3. Jin, P., Zhang, Z., Zhu, A., Tang, Y. & Karniadakis, G. E. SympNets: Intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems. *Neural Netw.* **132**, 166–179. https://doi.org/10.1016/j.neunet.2020.08.017 (2020).
4. Willard, J. D., Jia, X., Xu, S. & Steinbach, M. S. & Kumar, V *A Survey, Integrating Physics-Based Modeling with Machine Learning.* (2020).
5. Lei, H., Wu, L. & Weinan, E. Machine-learning-based non-Newtonian fluid model with molecular fidelity. *Phys. Rev. E* **102**, 043309 (2020).
6. Qin, H. Machine learning and serving of discrete field theories. *Sci. Rep.* **10**, 1–15 (2020).
7. Cotter, C. J. & Reich, S. Adiabatic invariance and applications: From molecular dynamics to numerical weather prediction. *BIT Numer. Math.* **44**, 439 (2004).
8. Burby, J. W. & Hirvijoki, E. Normal stability of slow manifolds in nearly periodic Hamiltonian systems. *J. Math. Phys.* **62**, 093506. https://doi.org/10.1063/5.0054323 (2021).
9. Kruskal, M. Asymptotic theory of Hamiltonian and other systems with all solutions nearly periodic. *J. Math. Phys.* **3**, 806–828. https://doi.org/10.1063/1.1724285 (1962).
10. Burby, J. W. & Squire, J. General formulas for adiabatic invariants in nearly periodic hamiltonian systems. *J. Plasma Phys.* **86**, 835860601 (2020).
11. Burby, J. W., Hirvijoki, E. & Leok, M. Nearly-periodic maps and geometric integration of noncanonical Hamiltonian systems. *J. Nonlinear Sci.* **33**(2), 38 (2021).
12. Poincaré, H. *Les Methods Nouv. Mech Celeste* Vol. 3 (Gauthier-Villars, 1899).
13. Hairer, E., Lubich, C. & Wanner, G. *Geometric Numerical Integration. Springer Series in Computational Mathematics* Vol. 31 (Springer-Verlag, Berlin, 2006).
14. Iserles, A. & Quispel, G. *Why Geometric Numerical Integration?* (Springer International Publishing, 2018).
15. Blanes, S. & Casas, F. *A Concise Introduction to Geometric Numerical Integration* (University of Cambridge, 2017).
16. Leimkuhler, B. & Reich, S. *Simulating Hamiltonian Dynamics of Cambridge Monographs on Applied and Computational Mathematics* Vol. 14 (Cambridge University Press, 2004).
17. Holm, D., Schmah, T. & Stoica, C. *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions* (Oxford Texts in Applied and Engineering Mathematics (OUP Oxford, 2009).
18. Chen, Z., Zhang, J., Arjovsky, M. & Bottou, L. Symplectic recurrent neural networks. *Int. Conf. Learn. Rep.* **34**, 24048–24062 (2020).
19. Chen, Y., Matsubara, T. & Yaguchi, T. Neural symplectic form: Learning Hamiltonian equations on general coordinate systems. *Adv. Neural Inf. Process. Syst.* **34**, 16659–16670 (2021).

20. Cranmer, M. *et al.* Lagrangian neural networks. *ICLR Workshop on Integration of Deep Neural Models and Differential Equations* (2020).
21. Greydanus, S. l., Dzamba, M. & Yosinski, J. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, vol. 32 (2019).
22. Lutter, M., Ritter, C. & Peters, J. Deep Lagrangian networks: Using physics as model prior for deep learning. In *Int. Conf. Learn. Rep.* (2019).
23. Zhong, Y. D., Dey, B. & Chakraborty, A. Symplectic ODE-Net: Learning Hamiltonian dynamics with control. In *Int. Conf. Learn. Rep.* (2020).
24. Zhong, Y. D., Dey, B. & Chakraborty, A. Dissipative SymODEN: Encoding Hamiltonian dynamics with dissipation and control into deep learning. In *ICLR 2020 Work Integr. Deep Neural Models Differ. Eq.* (2020).
25. Zhong, Y. D., Dey, B. & Chakraborty, A. Benchmarking energy-conserving neural networks for learning dynamics from data. *Learn. Dyn. Control* **144**, 1218–1229 (2021).
26. Sæmundsson, S., Terenin, A., Hofmann, K. & Deisenroth, M. P. Variational integrator networks for physically structured embeddings. In *AISTATS*. PLMR (2020).
27. Havens, A. & Chowdhary, G. *Forced Variational Integrator Networks for Prediction and Control of Mechanical Systems* (PLMR, 2021).
28. Duruisseaux, V., Duong, T., Leok, M. & Atanasov, N. Lie group forced variational integrator networks for learning and control of robot systems, *5th Annual Learning for Dynamics and Control Conference (L4DC)*, 2023.
29. Santos, S., Ekal, M. & Ventura, R. Symplectic momentum neural networks - using discrete variational mechanics as a prior in deep learning. In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, vol. 168. *Proc. of Machine Learning Research*, 584–595 (2022).
30. Valperga, R., Webster, K., Turaev, D., Klein, V. & Lamb, J. Learning reversible symplectic dynamics. In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, vol. 168 of *Proc. of Machine Learning Research*, 906–916 (2022).
31. Bertalan, T., Dietrich, F., Mezić, I. & Kevrekidis, I. G. On learning Hamiltonian systems from data. *Chaos Interdiscip. J. Nonlinear Sci.* **29**, 121107. https://doi.org/10.1063/1.5128231 (2019).
32. Rath, K., Albert, C. G., Bischl, B. & von Toussaint, U. Symplectic gaussian process regression of maps in hamiltonian systems. *Chaos: An Interdiscip. J. Nonlinear Sci.* **31**, 053121. https://doi.org/10.1063/5.0048129 (2021).
33. Offen, C. & Ober-Blöbaum, S. Symplectic integration of learned Hamiltonian systems. *Chaos Interdiscip. J. Nonlinear Sci.* **32**, 013122. https://doi.org/10.1063/5.0065913 (2022).
34. Marco, D. & Méhats, F. Symplectic learning for Hamiltonian neural networks (2021).
35. Mathiesen, F. B., Yang, B. & Hu, J. Hyperverlet: A symplectic hypersolver for Hamiltonian systems. *Proc. AAAI Conf. Artif. Intell.* **36**, 4575–4582. https://doi.org/10.1609/aaai.v36i4.20381 (2022).
36. Morrison, P. The Maxwell-Vlasov equations as a continuous hamiltonian system. *J. Phys. Lett.* **80A**, 383 (1980).
37. Morrison, P. J. & Greene, J. M. Noncanonical Hamiltonian Density Formulation of Hydrodynamics and Ideal Magnetohydrodynamics. *Phys. Rev. Lett.* **45**, 790 (1980).
38. Morrison, P. J. Nonlinear stability of fluid and plasma equilibria. *Rev. Mod. Phys.* **70**, 467. https://doi.org/10.1103/RevModPhys.70.467 (1998).
39. Burby, J. W., Brizard, A. J., Morrison, P. J. & Qin, H. Hamiltonian gyrokinetic vlasov-maxwell system. *Phys. Lett. A* **379**, 2073. https://doi.org/10.1016/j.physleta.2015.06.051 (2015).
40. Morrison, P. J. & Vanneste, J. Weakly nonlinear dynamics in noncanonical hamiltonian systems with applications to fluids and plasmas. *Ann. Phys.* **368**, 117. https://doi.org/10.1016/j.aop.2016.02.003 (2016).
41. Morrison, P. J. & Kotschenreuther, M. *The free energy principle, negative energy modes, and stability* 9–22 (Texas University, 1989).
42. Burby, J. W. Slow manifold reduction as a systematic tool for revealing the geometry of phase space. *Phys. Plasmas* **29**, 042102. https://doi.org/10.1063/5.0084543 (2022).
43. Hernandez, Q., Badías, A., González, D., Chinesta, F. & Cueto, E. Deep learning of thermodynamics-aware reduced-order models from data. *Comput. Methods Appl. Mech. Eng.* **379**, 113763. https://doi.org/10.1016/j.cma.2021.113763 (2021).
44. Hernández, Q., Badias, A., Chinesta, F. & Cueto, E. Port-metriplectic neural networks: thermodynamics-informed machine learning of complex physical systems. *Comput. Mech.* (2023), 1–9 https://doi.org/10.1007/s00466-023-02296-w.
45. Huang, S., He, Z., Chem, B. & Reina, C. Variational Onsager Neural Networks (VONNs): A thermodynamics-based variational learning strategy for non-equilibrium PDEs. *J. Mech. Phys. Solids* **163**, 104856. https://doi.org/10.1016/j.jmps.2022.104856 (2022).
46. Rusch, T. K. & Mishra, S. Unicornn: A recurrent model for learning very long time dependencies. In *Proceedings of the 38th Int. Conf. on Machine Learning. Proc. of Machine Learning Research*, vol. 139, pp. 9168–9178 (2021).
47. Chen, G., Chacón, L. & Barnes, D. C. An energy-and charge-conserving, implicit, electrostatic particle-in-cell algorithm. *J. Comput. Phys.* **230**, 7018–7036 (2011).
48. Chen, G. & Chacón, L. A multi-dimensional, energy-and charge-conserving, nonlinearly implicit, electromagnetic Vlasov-Darwin particle-in-cell algorithm. *Comput. Phys. Commun.* **197**, 73–87 (2015).
49. Miller, S. T. *et al.* IMEX and exact sequence discretization of the multi-fluid plasma model. *J. Comput. Phys.* **397**, 108806 (2019).
50. Lorenz, E. N. The slow manifold—what is it?. *J. Atmos. Sci.* **49**, 2449–2451 (1992).
51. Lorenz, E. N. & Krishnamurthy, V. On the nonexistence of a slow manifold. *J. Atmos. Sci.* **44**, 2940–2950 (1987).
52. Lorenz, E. N. On the existence of a slow manifold. *J. Atmos. Sci.* **43**, 1547–1557 (1986).
53. MacKay, R. S. *Slow manifolds In Energy Localization and Transfer of Advanced Series in Nonlinear Dynamics* Vol. 22 (World Scientific, 2004).
54. Burby, J. W. & Klotz, T. J. Slow manifold reduction for plasma science. *Comm. Nonlin. Sci. Numer. Simul.* **89**, 105289 (2020).
55. McInerney, A. *First Steps in Differential Geometry: Riemannian, Contact, Symplectic*. Undergraduate Texts in Mathematics (Springer, 2013).
56. Lang, S. *Fundamentals of Differential Geometry of Graduate Texts in Mathematics* Vol. 191 (Springer-Verlag, 1999).
57. Marsden, J. & Ratiu, T. *Introduction to mechanics and symmetry of Texts in Applied Mathematics* 2nd edn, Vol. 17 (Springer-Verlag, New York, 1999).
58. Weinstein, A. Symplectic manifolds and their Lagrangian submanifolds. *Adv. Math.* **6**, 329–346. https://doi.org/10.1016/0001-8708(71)90020-X (1971).
59. Turaev, D. Polynomial approximations of symplectic dynamics and richness of chaos in non-hyperbolic area-preserving maps. *Nonlinearity* **16**, 123–135. https://doi.org/10.1088/0951-7715/16/1/308 (2002).
60. Abraham, R. & Marsden, J. E. *Foundations of Mechanics* (Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass, 1978).
61. Arnol'd, V. I. *Mathematical Methods of Classical Mechanics of Graduate Texts in Mathematics (Translated from the Russian)* 2nd edn, Vol. 60 (Springer-Verlag, New York, 1989).
62. Burby, J. W. & Squire, J. General formulas for adiabatic invariants in nearly periodic Hamiltonian systems. *J. Plasma Phys.* **86**, 835860601. https://doi.org/10.1017/S002237782000080X (2020).
63. Teshima, T. *et al.* Coupling-based invertible neural networks are universal diffeomorphism approximators. *Adv. Neural Inf. Process. Syst.* **33**, 3362–3373 (2020).
64. Raymond, F. Classification of the actions of the circle on 3-manifolds. *Transact. Am. Math. Soc.* **131**, 51–78 (1968).

65. Duruisseaux, V., Burby, J. W. & Tang, Q. *Code Demonstration: Approximation of Nearly-Periodic Symplectic Maps via Structure-Preserving Neural Networks.* (2023) https://doi.org/10.2172/1972078.

Acknowledgements

Author contributions

V.D. wrote the manuscript and conducted the numerical experiments. J.W.B. supervised the project on the theoretical side. Q.T. supervised the project on the computational side, and designed the diagrams in Figure 1. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to V.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.