Symplectic Neural Flows

Davide Murari

Department of Applied Mathematics and Theoretical Physics, University of Cambridge.

dm2011@cam.ac.uk

In collaboration with Priscilla Canizares, Carola-Bibiane Schönlieb, Ferdia Sherry, Zakhar Shumaylov

Canonical Hamiltonian equations

- The equations of motion of canonical Hamiltonian systems write

$$\begin{cases} \dot{x} = \mathbb{J}\nabla H(x) = X_H(x) \in \mathbb{R}^{2n} \\ x(0) = x_0 \end{cases}, \quad \mathbb{J} = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix} \in \mathbb{R}^{2n\times 2n}. \tag{1}$$

Canonical Hamiltonian equations

- The equations of motion of canonical Hamiltonian systems write

$$\begin{cases} \dot{x} = \mathbb{J}\nabla H(x) = X_H(x) \in \mathbb{R}^{2n} \\ x(0) = x_0 \end{cases}, \quad \mathbb{J} = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix} \in \mathbb{R}^{2n\times 2n}. \tag{1}$$

- Denoted with $\phi_{H,t} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ the exact flow of (1), $\phi_{H,t}(x_0) = x(t)$, we have that
  - $$\frac{d}{dt}H(\phi_{H,t}(x_0)) = \nabla H(\phi_{H,t}(x_0))^\top \mathbb{J}\nabla H(\phi_{H,t}(x_0)) = 0,$$

Canonical Hamiltonian equations

- The equations of motion of canonical Hamiltonian systems write

$$\begin{cases} \dot{x} = \mathbb{J}\nabla H(x) = X_H(x) \in \mathbb{R}^{2n} \\ x(0) = x_0 \end{cases}, \quad \mathbb{J} = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix} \in \mathbb{R}^{2n\times 2n}. \tag{1}$$

- Denoted with $\phi_{H,t} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ the exact flow of (1), $\phi_{H,t}(x_0) = x(t)$, we have that
  - $$\frac{d}{dt} H(\phi_{H,t}(x_0)) = \nabla H(\phi_{H,t}(x_0))^\top \mathbb{J}\nabla H(\phi_{H,t}(x_0)) = 0,$$
  - $$\left(\frac{\partial \phi_{H,t}(x_0)}{\partial x_0}\right)^\top \mathbb{J} \left(\frac{\partial \phi_{H,t}(x_0)}{\partial x_0}\right) = \mathbb{J},$$

Canonical Hamiltonian equations

- The equations of motion of canonical Hamiltonian systems write

$$\begin{cases} \dot{x} = \mathbb{J}\nabla H(x) = X_H(x) \in \mathbb{R}^{2n} \\ x(0) = x_0 \end{cases}, \quad \mathbb{J} = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix} \in \mathbb{R}^{2n\times 2n}. \tag{1}$$

- Denoted with $\phi_{H,t} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ the exact flow of (1), $\phi_{H,t}(x_0) = x(t)$, we have that
  - $$\frac{d}{dt} H(\phi_{H,t}(x_0)) = \nabla H(\phi_{H,t}(x_0))^\top \mathbb{J} \nabla H(\phi_{H,t}(x_0)) = 0,$$
  - $$\left(\frac{\partial \phi_{H,t}(x_0)}{\partial x_0}\right)^\top \mathbb{J} \left(\frac{\partial \phi_{H,t}(x_0)}{\partial x_0}\right) = \mathbb{J},$$
  - the flow preserves the canonical volume form of $\mathbb{R}^{2n}$.

Forward invariant subset of the phase space

- Suppose $x(t) \in \Omega \subset \mathbb{R}^{2n}$, whenever $x(0) \in \Omega$, for any $t \geq 0$.

Forward invariant subset of the phase space

- Suppose $x(t) \in \Omega \subset \mathbb{R}^{2n}$, whenever $x(0) \in \Omega$, for any $t \geq 0$.
- By the group property of the flow map, we know that

$$\phi_{H,n\Delta t+\delta t} = \phi_{H,\delta t} \circ \underbrace{\phi_{H,\Delta t} \circ \dots \circ \phi_{H,\Delta t}}_{n \text{ times}}, \quad n \in \mathbb{N},\ \delta t \in (0, \Delta t).$$

As a consequence, to approximate $\phi_{H,t} : \Omega \to \Omega$ for any $t \geq 0$, we only have to approximate it for $t \in [0, \Delta t]$.

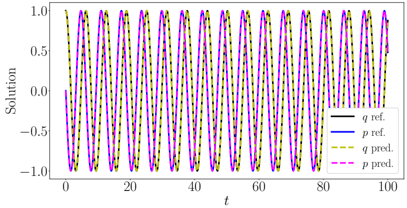


Figure 1: Neural network trained to approximate $\phi_{H,t}$ for $t \in [0, \Delta t = 1]$ and applied up to $T = 100$.

Two learning problems associated with Hamiltonian systems

Unsupervised solution of the Hamiltonian equations

Approximate the flow map $\phi_{H,t} : \Omega \to \Omega$, for any $t \geq 0$, on a compact forward invariant set $\Omega \subset \mathbb{R}^{2n}$, given the Hamiltonian energy $H : \mathbb{R}^{2n} \to \mathbb{R}$.

Two learning problems associated with Hamiltonian systems

Unsupervised solution of the Hamiltonian equations

Approximate the flow map $\phi_{H,t} : \Omega \to \Omega$, for any $t \geq 0$, on a compact forward invariant set $\Omega \subset \mathbb{R}^{2n}$, given the Hamiltonian energy $H : \mathbb{R}^{2n} \to \mathbb{R}$.

Supervised approximation of an unknown Hamiltonian flow map

Approximate the flow map $\phi_{H,t} : \Omega \to \Omega$, for any $t \geq 0$, on a compact forward invariant set $\Omega \subset \mathbb{R}^{2n}$, given trajectory segments $\{(x_0^n, y_1^n, \dots, y_M^n)\}_{n=1}^N$, $y_m^n \approx \phi_{H,t_m^n}(x_0^n)$.

Two learning problems associated with Hamiltonian systems

Unsupervised solution of the Hamiltonian equations

Approximate the flow map $\phi_{H,t} : \Omega \to \Omega$, for any $t \geq 0$, on a compact forward invariant set $\Omega \subset \mathbb{R}^{2n}$, given the Hamiltonian energy $H : \mathbb{R}^{2n} \to \mathbb{R}$.

Supervised approximation of an unknown Hamiltonian flow map

Approximate the flow map $\phi_{H,t} : \Omega \to \Omega$, for any $t \geq 0$, on a compact forward invariant set $\Omega \subset \mathbb{R}^{2n}$, given trajectory segments $\{(x_0^n, y_1^n, ..., y_M^n)\}_{n=1}^N$, $y_m^n \approx \phi_{H,t_m^n}(x_0^n)$.

Remark: Given the several known qualitative properties of $\phi_{H,t}$, we want to leverage them when designing the approximating map.

The SympFlow

- We now build a neural network that approximates $\phi_{H,t} : \Omega \to \Omega$ for a forward invariant set $\Omega \subset \mathbb{R}^{2n}$, and $t \in [0, \Delta t]$, while reproducing the qualitative properties of $\phi_{H,t}$.

The SympFlow

- We now build a neural network that approximates $\phi_{H,t} : \Omega \to \Omega$ for a forward invariant set $\Omega \subset \mathbb{R}^{2n}$, and $t \in [0, \Delta t]$, while reproducing the qualitative properties of $\phi_{H,t}$.
- We rely on two building blocks, which applied to $(q, p) \in \mathbb{R}^{2n}$ write:

$$\phi_{p,t}((q,p)) = \begin{bmatrix} q \\ p - (\nabla_q V(t,q) - \nabla_q V(0,q)) \end{bmatrix}, \ \phi_{q,t}((q,p)) = \begin{bmatrix} q + (\nabla_p K(t,p) - \nabla_p K(0,p)) \\ p \end{bmatrix}.$$

The SympFlow

- We now build a neural network that approximates $\phi_{H,t} : \Omega \to \Omega$ for a forward invariant set $\Omega \subset \mathbb{R}^{2n}$, and $t \in [0, \Delta t]$, while reproducing the qualitative properties of $\phi_{H,t}$.

- We rely on two building blocks, which applied to $(q, p) \in \mathbb{R}^{2n}$ write:

$$\phi_{p,t}((q,p)) = \begin{bmatrix} q \\ p - (\nabla_q V(t,q) - \nabla_q V(0,q)) \end{bmatrix}, \ \phi_{q,t}((q,p)) = \begin{bmatrix} q + (\nabla_p K(t,p) - \nabla_p K(0,p)) \\ p \end{bmatrix}.$$

- The SympFlow architecture is defined as

$$\mathcal{N}_\theta\,(t, (q_0, p_0)) = \phi^L_{p,t} \circ \phi^L_{q,t} \circ \cdots \circ \phi^1_{p,t} \circ \phi^1_{q,t}((q_0, p_0)),$$

with

$$V^i(t,q) = \ell_{\theta^i_3} \circ \sigma \circ \ell_{\theta^i_2} \circ \sigma \circ \ell_{\theta^i_1}\left(\begin{bmatrix} q \\ t \end{bmatrix}\right), \ K^i(t,p) = \ell_{\rho^i_3} \circ \sigma \circ \ell_{\rho^i_2} \circ \sigma \circ \ell_{\rho^i_1}\left(\begin{bmatrix} p \\ t \end{bmatrix}\right)$$

$$\ell_{\theta^i_k}(x) = A^i_k x + a^i_k, \ \ell_{\rho^i_k}(x) = B^i_k x + b^i_k, \ k = 1,2,3, \ i = 1,\dots,L.$$

Properties of the SympFlow

- The SympFlow is symplectic for every time $t \in \mathbb{R}$. The building blocks we compose are exact flows of time-dependent Hamiltonian systems:

$$\phi^i_{p,t}((q,p)) = \begin{bmatrix} q \\ p - \left(\nabla_q V^i(t,q) - \nabla_q V^i(0,q)\right) \end{bmatrix}$$
$$= \begin{bmatrix} q \\ p - \nabla_q \left(\int_0^t \partial_s V^i(s,q)ds\right) \end{bmatrix} = \phi_{\widetilde{V}^i,t}((q,p)),$$

with $\widetilde{V}^i(t,(q,p)) = \partial_t V^i(t,q)$.

Properties of the `SympFlow`

- The `SympFlow` is symplectic for every time $t \in \mathbb{R}$. The building blocks we compose are exact flows of time-dependent Hamiltonian systems:

$$\phi^i_{p,t}((q,p)) = \begin{bmatrix} q \\ p - \left(\nabla_q V^i(t,q) - \nabla_q V^i(0,q)\right) \end{bmatrix} = \begin{bmatrix} q \\ p - \nabla_q \left(\int_0^t \partial_s V^i(s,q) ds\right) \end{bmatrix} = \phi_{\widetilde{V}^i,t}((q,p)),$$

  with $\widetilde{V}^i(t,(q,p)) = \partial_t V^i(t,q)$.

- The `SympFlow` is volume preserving.

Properties of the `SympFlow`

- The `SympFlow` is symplectic for every time $t \in \mathbb{R}$. The building blocks we compose are exact flows of time-dependent Hamiltonian systems:

$$\phi^i_{p,t}((q,p)) = \begin{bmatrix} q \\ p - \left(\nabla_q V^i(t,q) - \nabla_q V^i(0,q)\right) \end{bmatrix}$$
$$= \begin{bmatrix} q \\ p - \nabla_q \left(\int_0^t \partial_s V^i(s,q)ds\right) \end{bmatrix} = \phi_{\widetilde{V}^i,t}((q,p)),$$

  with $\widetilde{V}^i(t,(q,p)) = \partial_t V^i(t,q)$.

- The `SympFlow` is volume preserving.
- The `SympFlow` is the exact solution of a time-dependent Hamiltonian system.

Composition of Hamiltonian flows

Theorem (The Hamiltonian flows are closed under composition)

Let $H^1, H^2 : \mathbb{R} \times \mathbb{R}^{2n} \to \mathbb{R}$ be twice-continuously differentiable functions. Then, the map $\phi_{H^2,t} \circ \phi_{H^1,t} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is the exact flow of the time-dependent Hamiltonian system defined by the Hamiltonian function

$$H^3(t,x) = H^2(t,x) + H^1\left(t, \phi^{-1}_{H^2,t}(x)\right).$$

- This theorem implies that there is a Hamiltonian function $\mathcal{H}(\mathcal{N}_\theta) : \mathbb{R} \times \mathbb{R}^{2n} \to \mathbb{R}$ such that

$$\mathcal{N}_\theta\,(t,x) = \phi_{\mathcal{H}(\mathcal{N}_\theta),t}(x)$$

for every $t \geq 0$ and $x \in \mathbb{R}^{2n}$.

Extension of the SympFlow outside of $[0, \Delta t]$

- Once we have trained $\mathcal{N}_\theta$ to be reliable for $t \in [0, \Delta t]$, we extend it for longer times as

$$\psi(t, x_0) := \bar{\psi}_{t-\Delta t \lfloor t/\Delta t \rfloor} \circ \left(\bar{\psi}_{\Delta t}\right)^{\lfloor t/\Delta t \rfloor}(x_0),$$

for $t \in [0, +\infty)$ and $x_0 \in \Omega \subset \mathbb{R}^{2n}$, where

$$\bar{\psi}_s(x_0) := \mathcal{N}_\theta(s, x_0), \ s \in [0, \Delta t),$$
$$\left(\bar{\psi}_{\Delta t}\right)^k := \underbrace{\bar{\psi}_{\Delta t} \circ \cdots \circ \bar{\psi}_{\Delta t}}_{k \text{ times}}, \ k \in \mathbb{N}.$$

Extension of the SympFlow outside of $[0, \Delta t]$

- Once we have trained $\mathcal{N}_\theta$ to be reliable for $t \in [0, \Delta t]$, we extend it for longer times as

$$\psi(t, x_0) := \bar{\psi}_{t-\Delta t\lfloor t/\Delta t\rfloor} \circ \left(\bar{\psi}_{\Delta t}\right)^{\lfloor t/\Delta t\rfloor}(x_0),$$

for $t \in [0, +\infty)$ and $x_0 \in \Omega \subset \mathbb{R}^{2n}$, where

$$\bar{\psi}_s(x_0) := \mathcal{N}_\theta(s, x_0), \;\; s \in [0, \Delta t),$$
$$\left(\bar{\psi}_{\Delta t}\right)^k := \underbrace{\bar{\psi}_{\Delta t} \circ \cdots \circ \bar{\psi}_{\Delta t}}_{k \text{ times}}, \;\; k \in \mathbb{N}.$$

- $\psi(t, \cdot) = \phi_{\widetilde{H},t}$ for the piecewise continuous Hamiltonian

$$\widetilde{H}(t, x) := \mathcal{H}(\mathcal{N}_\theta)(t - \Delta t\lfloor t/\Delta t\rfloor, x).$$

Universal Approximation Theorem

Theorem

Let $H : \mathbb{R} \times \mathbb{R}^{2d} \to \mathbb{R}$ be twice-continuously differentiable, and $\Omega \subset \mathbb{R}^{2n}$ a compact and forward invariant set. For any $\varepsilon > 0$, there is a `SympFlow` $\bar{\psi}_t$ *such that*

$$\sup_{\substack{t\in[0,\Delta t]\\ x\in\Omega}} \left\|\bar{\psi}_t(x) - \phi_{H,t}(x)\right\|_\infty < \varepsilon.$$

Training of the `SympFlow` to solve $\dot{x}(t) = X_H(x(t))$

- To train the overall model $\mathcal{N}_\theta : \mathbb{R} \times \mathbb{R}^{2d} \to \mathbb{R}^{2d}$, which could be a `SympFlow` or a generic neural network, we minimise the loss function

$$\mathcal{L}(\theta) = \frac{1}{N_r} \sum_{i=1}^{N_r} \left\| \frac{d}{dt} \mathcal{N}_\theta \left(t, x_0^i\right) \Big|_{t=t_i} - \mathbb{J} \nabla H \left(\mathcal{N}_\theta \left(t_i, x_0^i\right)\right) \right\|_2^2$$

where we sample $t_i \in [0, \Delta t]$, and $x_0^i \in \Omega \subset \mathbb{R}^{2n}$.

Equations of motion

$$\dot{x} = p, \ \ \dot{p} = -x.$$

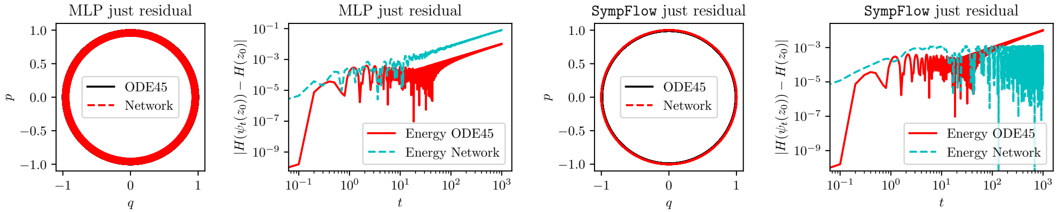


Figure 2: **Unsupervised experiment — Simple Harmonic Oscillator.** Comparison of the orbits and the energy behaviour up to time $T = 1000$.

Hénon–Heiles (unsupervised)

Equations of motion

$$\dot{x} = p_x, \ \dot{y} = p_y, \ \dot{p}_x = -x - 2xy, \ \dot{p}_y = -y - (x^2 - y^2).$$

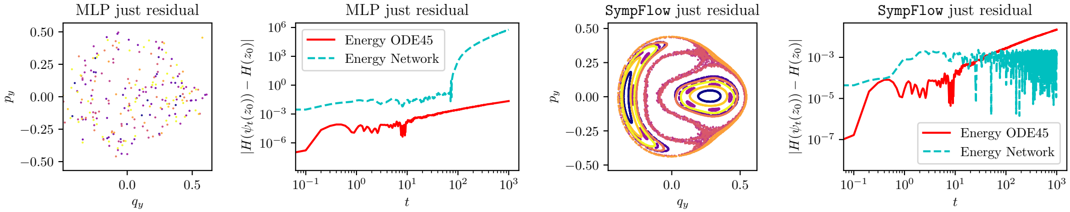


Figure 3: **Unsupervised experiment — Hénon–Heiles:** Comparison of the Poincaré sections and the energy behaviour up to time $T = 1000$.

Future extensions

- Improve the efficiency of the method by replacing gradients of MLPs with some other alternatives (Topic of a Summer Project that will start in a couple of weeks).
- Extend the approach to capture parametric dependencies, and apply this procedure for parameter identification.
- Improve our theoretical understanding of the dynamics exactly solved by the `SympFlow`.
- Apply the method to higher dimensional systems.

THANK YOU FOR THE ATTENTION

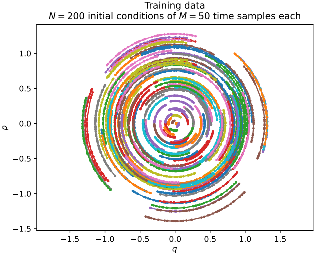
davidemurari.com/sympflow to read the paper

Supervised training of the `SympFlow` to approximate $\phi_{H,t}$

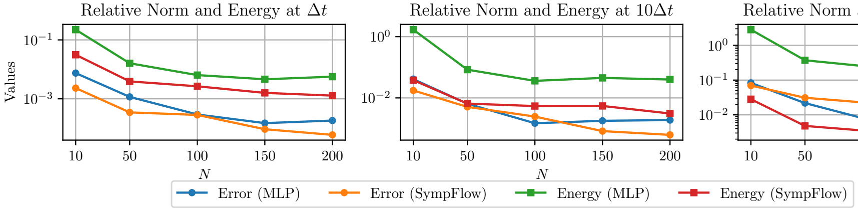
- To train the overall model $\mathcal{N}_\theta$ we minimise the loss function

$$\mathcal{L}(\theta) = \frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M} \left\|\mathcal{N}_\theta\left(t_m^n, x_0^n\right) - y_m^n\right\|_2^2 ,$$

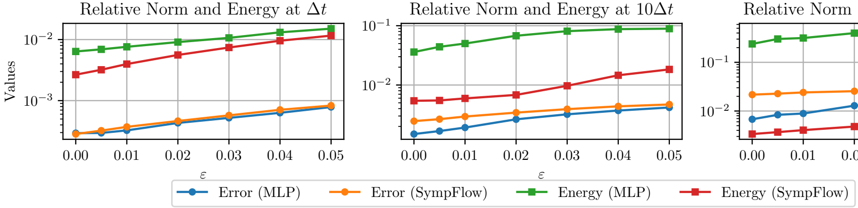
where $x_0^n \in \Omega \subset \mathbb{R}^{2n}$, and $y_m^n \approx \phi_{H,t_m^n}(x_0^n)$.



Simple Harmonic Oscillator (supervised)



(a) Fixed $M = 50$ and $\varepsilon = 0$.



(b) Fixed $N = 100$ and $M = 50$.

Physics-informed neural networks

- We introduce a parametric map $\mathcal{N}_\theta\left(\cdot, x_0\right) : [0, T] \to \mathbb{R}^d$ such that $\mathcal{N}_\theta\left(0, x_0\right) = x_0$, and choose its weights so that

$$\mathcal{L}(\theta) := \frac{1}{C} \sum_{c=1}^{C} \left\| \frac{d}{dt} \mathcal{N}_\theta\left(t, x_0\right)\Big|_{t=t_c} - \mathcal{F}\left(\mathcal{N}_\theta\left(t_c, x_0\right)\right) \right\|_2^2 \to \min$$

for some collocation points $t_1, \ldots, t_C \in [0, T]$.

Physics-informed neural networks

- We introduce a parametric map $\mathcal{N}_\theta(\cdot, x_0) : [0, T] \to \mathbb{R}^d$ such that $\mathcal{N}_\theta(0, x_0) = x_0$, and choose its weights so that

$$\mathcal{L}(\theta) := \frac{1}{C} \sum_{c=1}^{C} \left\| \frac{d}{dt} \mathcal{N}_\theta(t, x_0) \Big|_{t=t_c} - \mathcal{F}(\mathcal{N}_\theta(t_c, x_0)) \right\|_2^2 \to \min$$

for some collocation points $t_1, \dots, t_C \in [0, T]$.

- Then, $t \mapsto \mathcal{N}_\theta(t, x_0)$ will solve a different IVP

$$\begin{cases} \dot{y}(t) = \mathcal{F}(y(t)) + \left(\frac{d}{dt} \mathcal{N}_\theta(t, x_0) \Big|_{t=t} - \mathcal{F}(y(t)) \right) \in \mathbb{R}^d, \\ y(0) = x_0 \in \mathbb{R}^d, \end{cases}$$

where hopefully the residual $\frac{d}{dt} \mathcal{N}_\theta(t, x_0) \Big|_{t=t} - \mathcal{F}(y(t))$ is small in some sense.

Training issues with neural network

- Solving a single IVP on $[0, T]$ with a neural network can take long training time.
- The obtained solution can not be used to solve the same ordinary differential equation with a different initial condition.

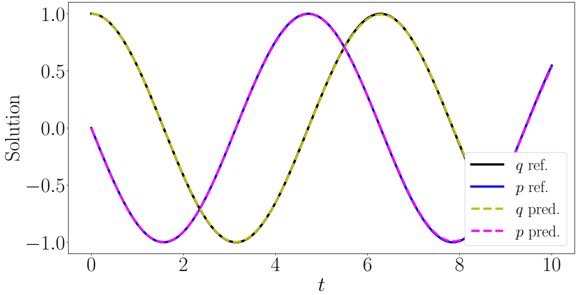
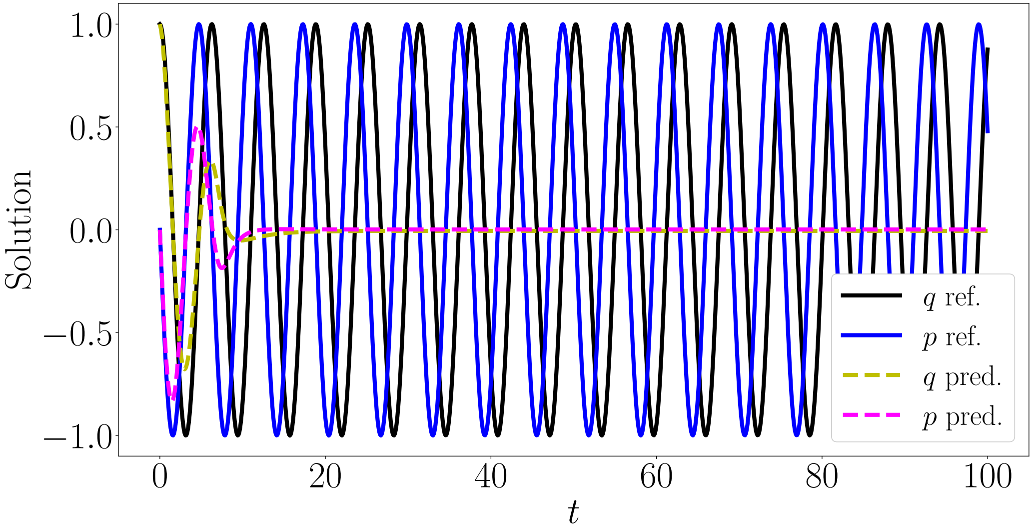


Figure 5: Solution comparison after reaching a loss value of 10^{-5}. The training time is of 87 seconds (7500 epochs with 1000 new collocation points randomly sampled at each of them).

Training issues with neural network

- It is hard to solve initial value problems over long time intervals.



Symplectic numerical methods

A one-step numerical method $\varphi^h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is symplectic if and only if when applied to a Hamiltonian system the map φ^h is symplectic, i.e.,

$$\left(\frac{\partial \varphi^h(x)}{\partial x}\right)^\top \mathbb{J} \left(\frac{\partial \varphi^h(x)}{\partial x}\right) = \mathbb{J}.$$

Symplectic numerical methods

A one-step numerical method $\varphi^h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is symplectic if and only if when applied to a Hamiltonian system the map φ^h is symplectic, i.e.,

$$\left(\frac{\partial \varphi^h(x)}{\partial x}\right)^\top \mathbb{J} \left(\frac{\partial \varphi^h(x)}{\partial x}\right) = \mathbb{J}.$$

Symplectic and energy preserving methods

Let $\dot{x} = \mathbb{J}\nabla H(x)$ be a Hamiltonian system with Hamiltonian H and no conserved quantities other than H. Let φ^h be a symplectic and energy-preserving method for the Hamiltonian system. Then φ^h reproduces the exact solution up to a time re-parametrisation.

Symplectic numerical methods

A one-step numerical method $\varphi^h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is symplectic if and only if when applied to a Hamiltonian system the map φ^h is symplectic, i.e.,

$$\left(\frac{\partial \varphi^h(x)}{\partial x}\right)^\top \mathbb{J} \left(\frac{\partial \varphi^h(x)}{\partial x}\right) = \mathbb{J}.$$

Symplectic and energy preserving methods

Let $\dot{x} = \mathbb{J}\nabla H(x)$ be a Hamiltonian system with Hamiltonian H and no conserved quantities other than H. Let φ^h be a symplectic and energy-preserving method for the Hamiltonian system. Then φ^h reproduces the exact solution up to a time re-parametrisation.

Informal theorem

A symplectic method almost conserves the Hamiltonian for an exponentially long time.

Example: simple harmonic oscillator