Symplectic Momentum Neural Networks – Using Discrete Variational Principles as a prior in Deep Learning

Saul Santos, Monica Ekal, Rodrigo Ventura
Instituto Superior Técnico,
Instituto de Sistemas e Robótica, Instituto Superior Técnico

Introduction

The application of the state-of-the-art black box models has frequently met with limited success in scientific domains due to:

- ❑ Large data requirements.
- ❑ Inability to produce physically consistent results.
- ❑ Lack of generalizability to out-of-sample scenarios.

One way to deal with the problems, consists in combining physical priors with deep learning techniques. In this work we are interested in using priors of the underlying equations of motion of mechanical systems with deep learning, resulting in a gray-box approach. The main contributions are as follows:

- ❑ Combination of continuous variational principles with and without prior knowledge of the underlying equations of motion with discretization schemes by Neural Ordinary Differential Equations (NODE) [1] through conventional integrators.
- ❑ Development of Symplectic Momentum Neural Networks based on a class of geometric integrations known by Variational Integrators.
- ❑ Development of End-to-End Symplectic Momentum Integrators by developing an implicit layer that accommodates root-finding procedures.
- ❑ Simulation results for toy mechanical systems.

Background

Continuous Lagrangian Mechanics

Euler-Lagrange equations

$$\frac{\partial \mathcal{L}(q,\dot q)}{\partial q} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}(q,\dot q)}{\partial \dot q}\right) + F\big(q(t),\dot q(t),u(t)\big) = 0$$

Discrete Lagrangian Mechanics [2]

Discrete Euler-Lagrange equations (DEL)

$$D_1\mathcal{L}_d(q_k,q_{k+1}) + D_2\mathcal{L}_d(q_{k-1},q_k) + f_d^+(q_{k-1},q_k,u_{k-1}) + f_d^-(q_k,q_{k+1},u_k) = 0$$

Baselines

- ❑ Geometric Neural Ordinary Differential equation (NODE): We consider observation of the form of (x_t, u, x_{t+h}). However, we rely on an embedding where we convert angles θ to the circle manifold $(\cos\theta, \sin\theta)$, hence the term geometric. The output is:

$$x_{t+h} = \underbrace{Integrator}_{\text{NODE}}\big(f(x,u;\theta)\big).$$

- ❑ Lagrangian Neural Ordinary Differential Equation (L-NODE): Parameterization of the potential energy $V(q)$ and the inertia matrix $H(q)$ and build the Euler-Lagrange equations.

Lagrangian for mechanical systems

$$\mathcal{L}(q,\dot q) = T - V = \frac{1}{2}\dot q^T H(q)\dot q - V(q)$$

⇩

Euler-Lagrange equations for mechanical systems

$$f^{(a)} = H^{-1}\big(F(q,\dot q,u) - C(q,\dot q) - g(q)\big)$$

⇩

Parameterize $H(q)$ and $V(q)$

Symplectic Momentum Neural Networks

- ❑ The input space can be defined by a set of two adjacent points in the configuration space and three adjacent discrete control inputs $x = (q_{k-1}, q_k, u_{k-1}, u_k, u_{k+1}) \in \mathcal{X}$ and the output space to be $y = (q_{k+1}) \in \mathcal{Y}$. Our goal is to build a neural network architecture that captures dependencies between the input and output space by using the Discrete Euler-Lagrange equation, $g(x, y;\theta)$, as a function that correlates both spaces.
- ❑ Inference consists in finding configurations of the variables q_{k+1}, obtained by implicitly solving the parameterized DEL, through the Newton's root finding algorithm. We designate this models by Symplectic Momentum Neural Networks (SyMo).

$$q_{k+1} = RootFind(g(q_{k+1}, x;\theta))$$

Discretization

Collocation points for the midpoint rule

$$q_{k-1/2} = 0.5q_{k-1} + 0.5q_k$$

$$q_{k+1/2} = 0.5q_k + 0.5q_{k+1}$$

Building discrete Lagrangians using the midpoint rule

$$\mathcal{L}_d(q_k,q_{k+1};\theta) = h\Big[\Big(\frac{q_{k+1}-q_k}{h}\Big)\hat H\Big(q_{k+\frac{1}{2}};\beta\Big)\Big(\frac{q_{k+1}-q_k}{h}\Big)^T - \hat V(q_{k+1/2};\varphi)\Big]$$

$$\mathcal{L}_d(q_{k-1},q_k;\theta) = h\Big[\Big(\frac{q_k-q_{k-1}}{h}\Big)\hat H(q_{k-1/2};\beta)\Big(\frac{q_k-q_{k-1}}{h}\Big)^T - \hat V(q_{k-1/2};\varphi)\Big]$$

Discrete Forces

Collocation points for the actuation

$$u_{k-1/2} = 0.5u_{k-1} + 0.5u_k$$

$$u_{k+1/2} = 0.5u_k + 0.5u_{k+1}$$

Left and right discrete force for control affine systems

$$f_d^-(q_k,q_{k+1},u_k,u_{k+1}) = \frac{h}{2}F\left(q_{k+1/2},\frac{q_{k+1}-q_k}{h},u_{k+1/2}\right) = \frac{h}{4}(u_k+u_{k+1})$$

$$f_d^+(q_{k-1},q_k,u_{k-1},u_k) = \frac{h}{2}F\left(q_{k-1/2},\frac{q_k-q_{k-1}}{h},u_{k-1/2}\right) = \frac{h}{4}(u_{k-1}+u_k)$$

⇩

Parameterized discrete Euler-Lagrange equations

$$g(x,y;\theta) = \frac{\partial}{\partial q_k}\left[\mathcal{L}_d(q_{k-1},q_k;\theta) + \mathcal{L}_d(q_k,q_{k+1};\theta)\right] + f_d^+(q_{k-1},q_k,u_{k-1},u_k) + f_d^-(q_k,q_{k+1},u_k,u_{k+1}) = 0$$

E2E-SyMo

- ❑ End-to-End Symplectic Momentum Neural Networks.
- ❑ End to end learning by including the root finding procedure into the learning framework.
- ❑ Forward Pass:

$$q_{k+1} = RootFind(g(q_{k+1}, x;\theta))$$

- ❑ Backward Pass:

Theorem. ***Gradient of the RootFind solution.*** *Let $q_{k+1} \in \mathbb{R}^n$ be the solution to the physical constrained parameterized RootFind procedure based on the implicit DEL mapping $(q_{k-1},q_k) \rightarrow (q_k,q_{k+1})$, defined by $g(q_{k+1},x;\theta) \in \mathbb{R}^n$. The gradients of a scalar loss function $\mathbb{L}(q_{k+1},x;\theta)$ with respect to the parameters θ are obtained by vector-Matrix products as follows:*

$$\frac{\partial \mathbb{L}}{\partial \theta} = -\frac{\partial \mathbb{L}}{\partial q_{k+1}}\left[\frac{\partial \mathcal{L}_d(q_k,q_{k+1};\theta)}{\partial q_{k+1}\partial q_k} + \frac{\partial f_d^-(q_k,q_{k+1},u_k,u_{k+1})}{\partial q_{k+1}}\right]^{-1}\frac{\partial g(q_{k+1},x;\theta)}{\partial \theta}$$

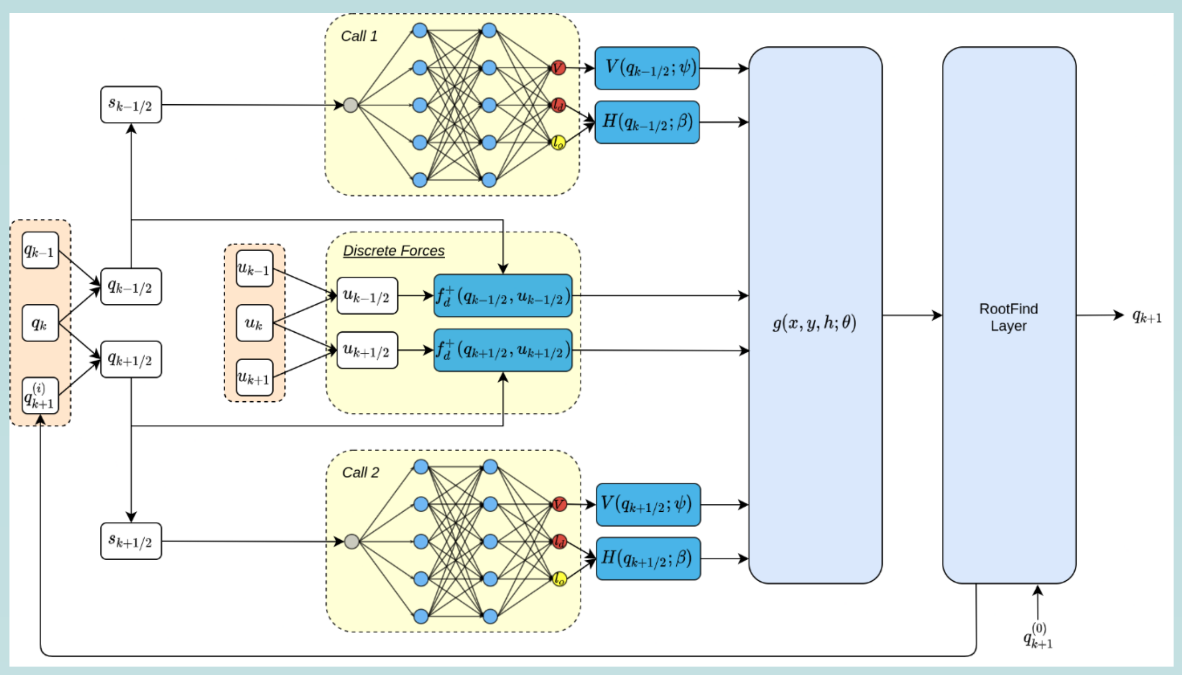

Figure 1: Given the embeddings s applied to rotational coordinates, we perform two calls to the neural network to get the inertia and potential energy correspondent to the two adjacent time-steps. These terms, alongside the discrete forces form then the discrete Euler-Lagrange equations, $g(x, y;\theta)$ as a function of the learning parameters. The DEL are then used by the implicit layer defined by the root finding procedure. Given an initial estimate $q_{k+1}^{(0)}$ the rootfind layer iterates over the DEL equations to obtain q_{k+1}.

Results

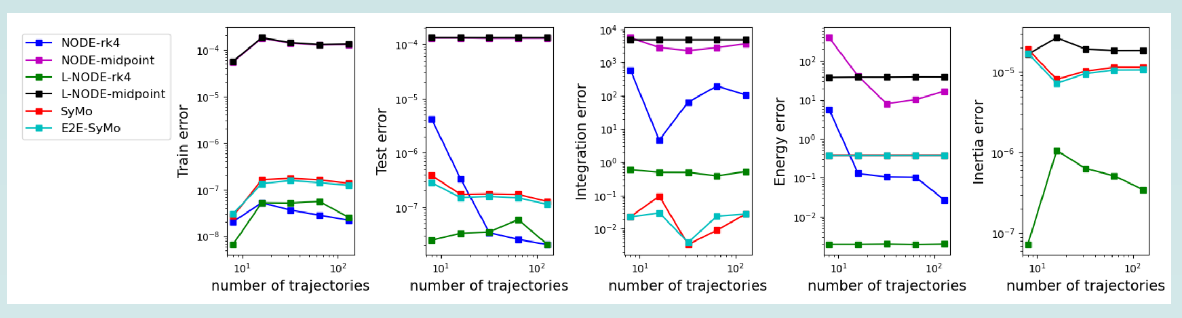

Figure 2: The figure shows the variation in train error, test error, integrator error, energy error and inertia error with changes in the number of initial state conditions in the training set. We can see that the incorporation of prior knowledge generally yields better train and test losses, specially for a small number of trajectories. For instance, the NODE-RK4 requires more trajectories to achieve similar results with the SyMos and L-NODE-RK4. However, in terms of integration, NODE-RK4 fails to keep up with those same models. SyMo and E2E-SyMo have a lower integration error which emphasizes their good long-term behaviour.

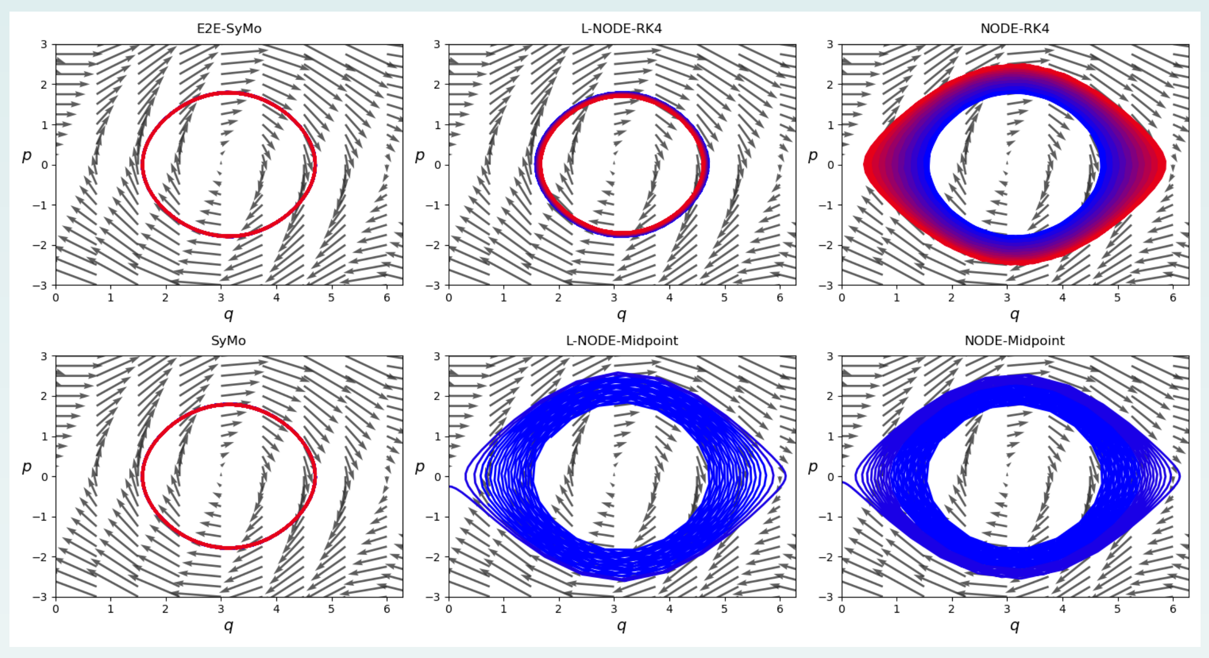

Figure 3: Phase Space of the simulated models for the pendulum system alongside the ground truth fields. Integration for 4000 time-steps with time-step h = 0.1 SyMo and E2E-SyMo preserve the symplectic form while the other models drift away from the ground truth.

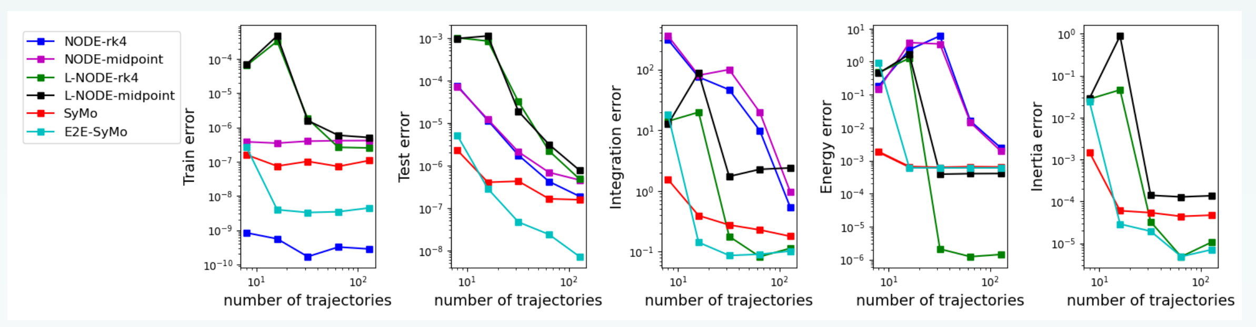

Figure 4: shows the results for the cartpole system. E2E-SyMo present the lowest test, integration and inertia error. Even outperforming the models with fourth-order integrators which shows how important is the preservation of geometric structures when integrating these models. In terms of energy conservation L-NODE-RK4 outperform the remainder methods. NODE-RK4 beats all the other models in terms of train loss but it's outperformed in the other metrics. This shows that the models without a prior are prone to overfitting.

Conclusion

- ❑ In this work we introduced Symplectic Momentum Neural Networks. We compared these models against the Neural ODEs with and without prior knowledge about the underlying equations of motion. Our models showed better long-term behaviour and conservation of properties, proving that these models can be used for data driven numerical integration in an efficient and interpretable way.

Acknowledgments

This work was supported by the LARSyS - FCT Project UIDB/50009/2020 and FCT PhD grant ref. PD/BD/150632/2020

References

[1] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. *Neural ordinary differential equations*. Advances in Neural Information Processing Systems, 2018.
[2] J. E. Marsden and M. West. Discrete mechanics and variational integrators. Acta Numerica, 10:357–514, 2001. doi: 10.1017/s096249290100006x.