Probabilistic Numerics for Hamiltonian Dynamics

Frederik De Ceuster *Leuven Gravity Institute, Department of Physics & Astronomy, KU Leuven, Belgium*
Tom Colemont *Leuven Gravity Institute, Department of Electrical Engineering (STADIUS), KU Leuven, Belgium*
Mathias Van Gompel *Leuven Gravity Institute, Department of Electrical Engineering (PSI), KU Leuven, Belgium*
Tjonnie G.-F. Li *Leuven Gravity Institute, Department of Physics & Astronomy, KU Leuven, Belgium*

Abstract

Many ordinary differential equations (ODEs) encountered in science originate from physical principles that contain substantially more structure than the ODE alone. In particular, Hamiltonian systems arise from variational principles, possess a symplectic structure, and exhibit conservation laws induced by symmetries. Standard probabilistic ODE solvers, that typically condition on the residual of the ODE, can overlook this additional physical information. In this paper, we focus on Hamiltonian dynamics, we revisit variational integrators and propose a probabilistic-numerical extension, based on a physics-informed prior. Furthermore, we discuss the symmetry-based Bayesian ODE framework of Wang et al. (2020), clarifying the role of integrability in Hamiltonian systems and the related Lie-algebra structure in defining Bayesian formulations.

Proceedings of the 2nd International Conference on Probabilistic Numerics (ProbNum) 2026,
Lappeenranta, Finland. PMLR Volume 341.

1 Introduction

Differential equations are a fundamental modelling tool in science and engineering. Many physical laws are most commonly expressed in differential form. Yet a differential equation is typically only one representation of an underlying physical principle. Classical mechanics provides a canonical example (see e.g. Arnold, 1989). Hamiltonian systems, such as the motion of a pendulum or celestial orbits, can be described either by Newton's second law, leading to a system of differential equations, or equivalently by Hamilton's principle of least action, formulated as a variational problem. Both formulations are equivalent but encode structure in different ways.

When solving such problems numerically, focusing solely on the differential equation can obscure the essential structure of the underlying physics. For example, applying standard numerical integrators to Newton's equations will generally fail to conserve energy. Structure-preserving numerical methods were developed precisely to address this issue: symplectic integrators enforce geometric invariants of Hamiltonian flows at the discrete level, while variational integrators discretise the action functional itself rather than the resulting equations of motion (see e.g. Blanes and Casas, 2025).

The goal of computation is rarely merely to solve a particular differential equation, but to faithfully represent the physical system it models. In probabilistic numerics (PN; Hennig et al., 2015, 2022; Cockayne et al., 2019), this perspective suggests that uncertainty quantification should not be tied exclusively to the residual of a chosen differential formulation. Instead, additional structural information, such as conservation laws, symmetries, or variational principles, may provide alternative notions of model discrepancy. From a Bayesian viewpoint, this opens the possibility of defining priors and likelihoods that reflect physical validity beyond the differential equation, thereby aligning inference more closely with the underlying scientific model.

Bosch et al. (2022) have shown that probabilistic solvers for ordinary differential equations (ODEs) can incorporate structural information beyond the vector field by expressing the likelihood through flexible information operators. In particular, they demonstrate that higher-order structure and conservation laws can be encoded directly in the Bayesian update, leading to more accurate and physically meaningful solutions. Their work highlights that additional physical constraints can be treated on equal footing with the ODE residual.

Wang et al. (2020) have pointed out that most existing probabilistic numerical methods (PNMs) for ODEs are not strictly Bayesian because evaluations of the gradient field generally do not correspond to exact information operators, violating the likelihood principle. As an alternative, they construct an exact Bayesian PNM for a restricted class of ODEs that admit a solvable Lie algebra. Using the Lie symmetry, the ODE is transformed into a sequence of integrals, on which exact Bayesian inference can be performed, and the posterior is then mapped back to the original solution space. The approach is mathematically rigorous but limited in scope, as the solvable Lie algebra condition is restrictive and requires structural knowledge about the ODE.

In this paper, we focus on Hamiltonian systems. Here, the physical context provides an intuitive way to interpret the ideas of Wang et al. (2020). Furthermore, we provide a PN extension to variational integrators, by using them as a physics-informed prior, and in addition conditioning them on the Hamiltonian constraint, following Bosch et al. (2022). We demonstrate our findings on the canonical examples of Hamiltonian systems: the single and double pendulum.

We use the following notation: Vectors are bold-italic ($\boldsymbol{a}$), matrices are sans-serif (A), and transposes are indicated by a $\dagger$. Overdots indicate time-derivatives ($\dot{a} \equiv \mathrm{d}a/\mathrm{d}t$), and probability distributions are denoted by $\mathbb{P}$.

2 Classical mechanics

As background for the rest of this paper, we provide a short introduction to the mathematical methods used in classical mechanics. For more details, we refer to the textbook by Arnold (1989).

2.1 Lagrangian mechanics

In Newtonian mechanics, dynamics is described in terms of Newton's second law of motion, i.e. a second-order system of ordinary differential equations that relates the acceleration of each component to the sum of all forces that act on it. In contrast, in Lagrangian mechanics, dynamics is described as an optimization problem over the space of possible trajectories. Consider the action, which is a functional that maps a trajectory, $\boldsymbol{q}(t)$, over a time interval, $t \in [0, T]$, to a real value, S. The action is defined as the time-integral over a Lagrangian, L,

$$S[\boldsymbol{q}] \;=\; \int_0^T \mathrm{d}t\; L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t). \tag{1}$$

The dynamics is entirely described by the Lagrangian, which is a function of $\boldsymbol{q}$, $\dot{\boldsymbol{q}}$, and t, where the dot is shorthand for a time-derivative. Equations of motion can be obtained by extremizing the action, i.e. by demanding that for any variation, $\delta\boldsymbol{q}(t)$, of the true path, $\boldsymbol{q}_\star(t)$, the resulting variation in the action vanishes, $\delta S[\boldsymbol{q}_\star] = 0$. The variation in the action of a trajectory reads,

$$\begin{aligned}\delta S \;&=\; \int_0^T \mathrm{d}t\; \left(\frac{\partial L}{\partial \boldsymbol{q}} \cdot \delta\boldsymbol{q} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \cdot \delta\dot{\boldsymbol{q}}\right) &(2)\\ &=\; \int_0^T \mathrm{d}t\; \left(\frac{\partial L}{\partial \boldsymbol{q}} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right)\right) \cdot \delta\boldsymbol{q}, &(3)\end{aligned}$$

where in the second line, we used integration by parts and assumed that the endpoints remaind fixed during the variation, such that $\delta\boldsymbol{q}(0) = \delta\boldsymbol{q}(T) = \boldsymbol{0}$. The requirement that δS vanishes for any variation, $\delta\boldsymbol{q}$, yields the Euler-Lagrange equations,

$$\frac{\partial L}{\partial \boldsymbol{q}} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right) \;=\; \boldsymbol{0}. \tag{4}$$

These are the equations of motion for the dynamics, and are often equivalent to Newton's second law of motion.

To make the transition to Hamiltonian mechanics, one starts by replacing velocities, $\dot{\boldsymbol{q}}$, by conjugate momenta, defined as,

$$\boldsymbol{p} \;\equiv\; \frac{\partial L}{\partial \dot{\boldsymbol{q}}}. \tag{5}$$

The Hamiltonian corresponding to the Lagrangian can then be constructed using a Legendre transform,

$$H(\boldsymbol{q}, \boldsymbol{p}, t) \;=\; \boldsymbol{p} \cdot \dot{\boldsymbol{q}} \;-\; L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t), \tag{6}$$

where, on the right-hand side, the velocities $\dot{\boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{p})$, are interpreted as functions of position and momentum. This provides yet another way to define mechanics.

2.2 Hamiltonian mechanics

Consider a dynamical system with states defined by phase space vectors $\boldsymbol{q}$ and $\boldsymbol{p}$, for which the dynamics is defined by Hamilton's equations,

$$\dot{\boldsymbol{q}} \;=\; \frac{\partial H}{\partial \boldsymbol{p}}, \qquad \dot{\boldsymbol{p}} \;=\; -\frac{\partial H}{\partial \boldsymbol{q}}, \tag{7}$$

where dots over variables still indicate time derivatives and $H(\boldsymbol{q}, \boldsymbol{p})$ is the Hamiltonian of the system.

Defining the phase space coordinate vector, $\boldsymbol{z} \equiv (\boldsymbol{q}, \boldsymbol{p})^\dagger$, and the symplectic matrix,

$$\mathsf{J} \;\equiv\; \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{8}$$

Hamilton's eqs. can be written as an ODE of the form,

$$\dot{\boldsymbol{z}} \;=\; \mathsf{J}\,\partial_{\boldsymbol{z}} H(\boldsymbol{z}). \tag{9}$$

When both the Lagrangian and Hamiltonian are well-defined, Hamilton's eqs. (7 or 9) are the first-order form of the second-order Euler-Lagrange eqs. (4).

2.3 Symplectic geometry

The phase space in Hamiltonian dynamics defines a symplectic manifold, i.e. a differentiable manifold with a closed, nondegenerate 2-form. Many results in classical mechanics can be ascribed to this rich mathematical structure. For this paper, we will limit ourselves to the definition of a symplectic map.

Define a flow map, $\boldsymbol{\varphi}_t$, parametrized by a time, t, which maps the initial condition, $\boldsymbol{z}_0$, onto its time-evolved state, $\boldsymbol{z}_t \equiv \boldsymbol{\varphi}_t(\boldsymbol{z}_0)$. Define the Jacobian of this flow map, $\boldsymbol{\Phi}_t \equiv \partial_{\boldsymbol{z}} \boldsymbol{\varphi}_t$. Poincaré's theorem states that the flow map of a twice-differentiable Hamiltonian system is a symplectic map, i.e. for all t throughout the evolution,

$$\boldsymbol{\Phi}_t^\dagger\, \mathsf{J}\, \boldsymbol{\Phi}_t \;=\; \mathsf{J}. \tag{10}$$

This is the central property of the symplectic geometry. It turns out that to retain the essential properties of Hamiltonian dynamics, for instance, after discretization, it is essential that also the discretized evolution map is symplectic (see e.g. Blanes and Casas, 2025).

2.4 Liouville's theorem

Liouville's theorem in Hamiltonian dynamics states that evolution in phase space is incompressible, $\partial_{\boldsymbol{z}} \cdot \dot{\boldsymbol{z}} = 0$. This follows by direct computation from Hamilton's eqs.

$$\begin{aligned}\partial_{\boldsymbol{z}} \cdot \dot{\boldsymbol{z}} \;&=\; \partial_{\boldsymbol{q}} \cdot \dot{\boldsymbol{q}} + \partial_{\boldsymbol{p}} \cdot \dot{\boldsymbol{p}} &(11)\\ &=\; \big(\partial_{\boldsymbol{q}} \cdot \partial_{\boldsymbol{p}} \;-\; \partial_{\boldsymbol{p}} \cdot \partial_{\boldsymbol{q}}\big) H \;=\; 0, &(12)\end{aligned}$$

since partial derivatives commute. This is an essential property when considering (probability) distributions over phase space, since it implies that given an initial (probability) distribution, its value remains constant along any trajectory in phase space. In other words, Hamiltonian dynamics will not concentrate or dilute probability mass, but merely rearrange it.

2.5 Symmetries and conservation laws

Consider a function, $A(\boldsymbol{q}, \boldsymbol{p})$, over phase space. The time-evolution of this function can be expressed as,

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{\partial A}{\partial \boldsymbol{q}} \cdot \dot{\boldsymbol{q}} + \frac{\partial A}{\partial \boldsymbol{p}} \cdot \dot{\boldsymbol{p}} \tag{13}$$

$$= \frac{\partial A}{\partial \boldsymbol{q}} \cdot \frac{\partial H}{\partial \boldsymbol{p}} - \frac{\partial A}{\partial \boldsymbol{p}} \cdot \frac{\partial H}{\partial \boldsymbol{q}} = \{A, H\} \tag{14}$$

where, in the last line, we introduced the Poisson bracket,

$$\{a, b\} \equiv \frac{\partial a}{\partial \boldsymbol{q}} \cdot \frac{\partial b}{\partial \boldsymbol{p}} - \frac{\partial a}{\partial \boldsymbol{p}} \cdot \frac{\partial b}{\partial \boldsymbol{q}}. \tag{15}$$

The time-evolution of a function in phase space can thus be expressed through the Poisson bracket of that function with the Hamiltonian. The vector space of infinitely differentiable functions on phase space forms a Lie algebra under point-wise multiplication, where the Poisson bracket plays the role of Lie bracket.

From eqs. (13) and (14), one can see that quantities, say $C(\boldsymbol{q}, \boldsymbol{p})$, that commute with the Hamiltonian are conserved, i.e. $\{C, H\} = 0$ implies that $\mathrm{d}C/\mathrm{d}t = 0$, and thus are constants of the motion.

It naturally follows that, if the Hamiltonian is time-independent, that it is conserved, $\mathrm{d}H/\mathrm{d}t = \{H, H\} = 0$.

2.6 Integrability

Consider a Hamiltonian system with N degrees of freedom (i.e. both $\boldsymbol{q}$ and $\boldsymbol{p}$ are each N-dimensional). This system is said to be Liouville-integrable if and only if:

1. there exist N constants of motion, $\{C_n\}_{n=1}^{N}$;
2. all these constants of motion are in involution, i.e. $\{C_i, C_j\} = 0$, for all i and j; and
3. their differentials, $\partial_{\boldsymbol{z}} C_i$, are linearly independent.

In other words, an Hamiltonian system is said to be Liouville-integrable if there are N constants of motion that form an Abelian Lie algebra. We note that every Abelian Lie algebra is also solvable.

The notion of integrability can be slightly generalized by relaxing requirement (2), and not demanding that the Lie algebra is strictly Abelian, but merely solvable. The dynamics is then said to be superintegrable.

The PNM defined by Wang et al. (2020), which requires a solvable Lie algebra, thus applies to (super)integrable Hamiltonian systems. The sequence of integrals they define are known as action-angle integrals in Hamiltonian dynamics. Intuitively, we can see that integrable systems evolve along the contours of constant C_i. Hence, given an initial condition, $\boldsymbol{z}_0$, we can evaluate the correctness of a solution with residuals $r_i \equiv |C_i(\boldsymbol{z}) - C_i(\boldsymbol{z}_0)|$. Given N such residuals for a system with N degrees of freedom, we can fully constrain the system and define an exact information operator in terms of these residuals. We can make this more explicit in an example.

2.7 Examples

We consider the two prototypical examples of classical mechanical systems: the single and double pendulum.

2.7.1 Single pendulum

For simplicity, we assume that the mass, arm length, and gravitational acceleration are all equal to one. The configuration space is given by the deflection angle of the pendulum. The Lagrangian can be obtained by subtracting the kinetic and potential energy, which yields,

$$L_{\mathrm{SP}} = \frac{1}{2}\dot{q}^2 + \cos(q). \tag{16}$$

Defining the conjugate momentum,

$$p \equiv \frac{\partial L_{\mathrm{SP}}}{\partial \dot{q}} = \dot{q} \tag{17}$$

yields the Hamiltonian for a (single) pendulum,

$$H_{\mathrm{SP}} = \frac{1}{2}p^2 - \cos(q). \tag{18}$$

The resulting Hamilton equations (7) then read,

$$\dot{q} = p \tag{19}$$

$$\dot{p} = -\sin(q) \tag{20}$$

and can be combined to obtain Newton's second law,

$$\ddot{q} = -\sin(q). \tag{21}$$

The single pendulum is integrable: there is one degree of freedom, q, and there is one constant of motion, H_{SP}. This means we can express the momentum in terms of the constant of motion, using eq. (18), say $H_{\mathrm{SP}} = E$,

$$p = \pm\sqrt{2\big(E + \cos(q)\big)} \tag{22}$$

and, using eq. (19), the dynamics becomes an integral,

$$t = \int \frac{\pm \mathrm{d}q}{\sqrt{2\big(E + \cos(q)\big)}}. \tag{23}$$

Note that this is an integral for $t(q)$, instead of $q(t)$. The trajectory in phase space is completely determined by the level-set of the Hamiltonian. The integral (23) computes the time corresponding to each location along trajectory.

2.7.2 Double pendulum

The natural next step after the (single) pendulum is the double pendulum. The corresponding Lagrangian can again be obtained by subtracting the kinetic and potential energy of the system,

$$L_{\mathrm{DP}} = L_{\mathrm{SP1}} + \mu\, L_{\mathrm{SP2}} + \lambda\mu\, \dot{q}_1 \dot{q}_2 \cos(q_1 - q_2). \tag{24}$$

We recognize the Lagrangians of two single pendulums and a coupling term. We introduced a parameter $\mu = m_2/(m_1 + m_2)$, which can be interpreted as the reduced

mass of the system. For $\mu = 1/2$, both pendulums have the same mass, while for $\mu \to 0$ or $\mu \to 1$, one mass is much heavier than the other. We also introduced a coupling constant λ. When $\lambda = 0$, the system describes the dynamics of two separate (uncoupled) pendulums, while when $\lambda = 1$, it describes a coupled system. The specific form of the coupling describes the situation where one pendulum is suspended from the other, i.e. the typical setup for a double pendulum.

We can again define the conjugate momenta,

$$p_1 \equiv \frac{\partial L_{\mathrm{DP}}}{\partial \dot{q}_1} = \dot{q}_1 + \lambda\mu\,\dot{q}_2\cos(q_1 - q_2) \tag{25}$$

$$p_2 \equiv \frac{\partial L_{\mathrm{DP}}}{\partial \dot{q}_2} = \mu\,\dot{q}_2 + \lambda\mu\,\dot{q}_1\cos(q_1 - q_2). \tag{26}$$

Note that $p_2 \propto \mu$. In the following, we will be interested in the limit $\mu \to 0$. In that limit, p_2 would not be a well-defined phase space coordinate, since it would not allow to recover $\dot{q}_2$. Instead, we will therefore use the rescaled momentum, $\bar{p}_2 \equiv p_2/\mu$, which remains well-defined, even in that limit. Using this new variable, we can solve the momentum equations for the velocities,

$$\dot{q}_1 = \frac{1}{a}\left(p_1 - \lambda\mu\,\bar{p}_2\cos(q_2 - q_1)\right) \tag{27}$$

$$\dot{q}_2 = \frac{1}{a}\left(\bar{p}_2 - \lambda\,p_1\cos(q_2 - q_1)\right) \tag{28}$$

where we used the auxiliary function,

$$a \equiv 1 - \lambda\mu\cos^2(q_2 - q_1). \tag{29}$$

The Hamiltonian for the double pendulum then reads,

$$H_{\mathrm{DP}} = H_1 + \mu\,H_2 \tag{30}$$

where we defined,

$$H_1 = \frac{1}{2a}\,p_1^2 - \cos(q_1) \tag{31}$$

$$H_2 = \frac{1}{2a}\left(\bar{p}_2^2 - \lambda p_1\bar{p}_2\cos(q_2 - q_1)\right) - \cos(q_2). \tag{32}$$

We roughly recognize the Hamiltonians of two single pendulums with a correction factor $(1/a)$ in the kinetic term, and with an additional coupling term between the pendulums which we absorbed in H_2. This yields the set of Hamilton equations,

$$\dot{q}_1 = \frac{1}{a}\left(p_1 - \lambda\mu\,\bar{p}_2\cos(q_2 - q_1)\right) \tag{33}$$

$$\dot{q}_2 = \frac{1}{a}\left(\bar{p}_2 - \lambda\,p_1\cos(q_2 - q_1)\right) \tag{34}$$

$$\dot{p}_1 = -\sin(q_1) - \lambda\mu\,b\sin(q_1 - q_2) \tag{35}$$

$$\dot{\bar{p}}_2 = -\sin(q_2) + \lambda\mu\,b\sin(q_1 - q_2) \tag{36}$$

where we defined the auxiliary function,

$$b = p_1^2 + \mu\,\bar{p}_2^2 - \big(1 - 2\lambda\mu\cos(q_2 - q_1)\big)p_1\bar{p}_2. \tag{37}$$

The double pendulum is not integrable: there are two degrees of freedom, q_1 and q_2, but there is still only one constant of motion, H_{DP}. The energies of the individual single pendulums (H_1 and H_2) are not conserved, since they can exchange energy through their coupling. One can indeed verify that,

$$\frac{\mathrm{d}H_1}{\mathrm{d}t} = \{H_1, H\} = \mu\{H_1, H_2\}, \tag{38}$$

$$\frac{\mathrm{d}H_2}{\mathrm{d}t} = \{H_2, H\} = \mu\{H_2, H_1\}. \tag{39}$$

In general, both are non-zero, since,

$$\{H_1, H_2\} = -\lambda\,c, \tag{40}$$

where c is a rather complicated function of q_1, q_2, p_1, and $\bar{p}_2$, that is in general non-zero (and well-behaved in the limits $\lambda \to 0$ and $\mu \to 0$). As a result, the Hamiltonians in general do not commute. However, we can distinguish two interesting limits.

Decoupling limit ($\lambda \to 0$) Considering the limit $\lambda \to 0$, we can effectively turn off the coupling between the two pendulums. From eq. (40), we can see that this implies that $\{H_1, H_2\} \to 0$. Then, from eqs. (38 and 39), it follows that both H_1 and H_2 become conserved quantities, and the system becomes integrable. This is to be expected, since, in this limit, the system effectively describes two separate single pendulums. One can verify that the Lagrangian, Hamiltonian, and Hamilton's eqs. all decouple. As expected, the conserved quantities also commute (i.e. $\{H_1, H_2\} \to 0$), and thus they form an Abelian Lie algebra, i.e. the simplest form of solvable Lie algebra. Following Wang et al. (2020), we can therefore formulate a PNM by Bayesian integration of the integral formulation (23) for each individual pendulum.

Extreme mass-ratio limit ($\mu \to 0$) Another interesting case is the limit where m_1 is much heavier than m_2, such that $\mu \to 0$. From eqs. (38 and 39), we can see that both H_1 and H_2 become conserved quantities, but that they do not commute, since, from eq. (40), we see that $\{H_1, H_2\} \neq 0$. As a result, the system is not strictly Liouville-integrable. However, the first pendulum alone is Liouville-integrable, since we have one conserved quantity (H_1) and one degree of freedom (q_1). Once the solution of the first pendulum is known, and we consider it a given for the dynamics of the second pendulum, also the second pendulum becomes integrable, since again we have one conserved quantity (H_2) and one degree of freedom (q_2). The conserved quantities for the combined system (H_1 and H_2) do not form an Abelian Lie algebra, but they still form a solvable Lie algebra. Again, following, Wang et al. (2020), we can formulate a PNM for the superintegrable system by Bayesian integration of the two separate systems. For the first pendulum this integral will again yield (23), but the second integral, which requires the substitution of the solution of the first, will be more involved.

An interesting aspect now is that by taking $\mu > 0$, we can "gently" turn off the solvability of the Lie algebra. In classical mechanics, this regime can be treated as a perturbation of the integrable system (Arnold, 1989). This might provide a route towards generalizing the PNM of Wang et al. (2020), or conversely, demonstrate the insensibility of a PNM far away from integrability.

3 Variational integrators

Hamiltonian systems have a rich geometric structure which can be leveraged when devising numerical integrators for these systems. There exists a wide variety of such integrators which are collectively referred to as geometric integrators (see e.g. Blanes and Casas, 2025). In this paper, we will focus on a specific constructive method known as variational integrators.

We discretize the time interval $[0, T]$ in $I+1$ points, $\{t_i\}_{i=0}^{I}$, where t_0 defines the initial condition, and to simplify notation, we define $\boldsymbol{q}_i \equiv \boldsymbol{q}(t_i)$. Now, we can define the (for now still exact) discrete Lagrangian,

$$\tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \;\equiv\; \int_{t_i}^{t_{i+1}} \mathrm{d}t\; L\big(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t\big), \tag{41}$$

and evaluate the corresponding discrete action,

$$\tilde{S}\left(\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_I\right) \;=\; \sum_{i=0}^{I-1} \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}), \tag{42}$$

in which we consider each $\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_I$ as independent variable. Note that all variables are positions now, instead of positions and velocities or momenta. The variation of the discrete action then yields,

$$\begin{aligned}
&\delta\tilde{S}\left(\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_I\right) \\
&= \sum_{i=0}^{I-1} \Big(\partial_1 \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \cdot \delta\boldsymbol{q}_i \;+\; \partial_2 \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \cdot \delta\boldsymbol{q}_{i+1} \Big) \\
&= \sum_{i=1}^{I-1} \Big(\partial_1 \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \;+\; \partial_2 \tilde{L}(\boldsymbol{q}_{i-1}, \boldsymbol{q}_i) \Big) \cdot \delta\boldsymbol{q}_i \\
&\quad + \partial_1 \tilde{L}(\boldsymbol{q}_0, \boldsymbol{q}_1) \cdot \delta\boldsymbol{q}_0 \;+\; \partial_2 \tilde{L}(\boldsymbol{q}_{I-1}, \boldsymbol{q}_I) \cdot \delta\boldsymbol{q}_I
\end{aligned}$$

where ∂_1 and ∂_2 denote partial derivatives with respect to the first and second argument of the discrete Lagrangian respectively. Since we keep the endpoints fixed, we have $\delta\boldsymbol{q}_0 = \delta\boldsymbol{q}_I = \boldsymbol{0}$. Now, demanding that $\delta\tilde{S}\left(\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_I\right) = \boldsymbol{0}$, for any variation $\delta\boldsymbol{q}_i$, yields the discrete Euler-Lagrange (DEL) equations,

$$\partial_1 \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \;+\; \partial_2 \tilde{L}(\boldsymbol{q}_{i-1}, \boldsymbol{q}_i) \;=\; \boldsymbol{0}. \tag{43}$$

The DEL equations can be cast into Hamiltonian or first-order form, by defining the discrete momenta,

$$\boldsymbol{p}_i \;\equiv\; \partial_2 \tilde{L}(\boldsymbol{q}_{i-1}, \boldsymbol{q}_i) \;=\; -\partial_1 \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}), \tag{44}$$

where, in the last equality, we used eq. (43). This yields the DEL update map,

$$\boldsymbol{p}_i \;=\; -\partial_1 \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \tag{45}$$

$$\boldsymbol{p}_{i+1} \;=\; \partial_2 \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \tag{46}$$

where the former can be implicitly solved for $\boldsymbol{q}_{i+1}$ and the latter explicitly yields $\boldsymbol{p}_{i+1}$. One can show by direct computation that the resulting discrete map is indeed symplectic (see Appendix A). However, we note that the discrete dynamics does not conserve the continuum Hamiltonian exactly, i.e. $H(\boldsymbol{q}_{i+1}, \boldsymbol{p}_{i+1}) \neq H(\boldsymbol{q}_i, \boldsymbol{p}_i)$. In fact, one can show that if a symplectic integrator would exactly conserve the (continuous) Hamiltonian, it would be exact (Zhong and Marsden, 1988).

3.1 Example

Consider a typical separable Lagrangian of the form,

$$L \;=\; \frac{1}{2}\,\dot{\boldsymbol{q}}^2(t) \;-\; V\big(\boldsymbol{q}(t)\big), \tag{47}$$

in which V is a general potential function that only depends on the positions, $\boldsymbol{q}$. Note that the Lagrangian of the single pendulum (16) can be written in this form, but the double pendulum Lagrangian (24) cannot. The discrete Lagrangian (41) corresponding to (47) reads,

$$\tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1}) \;=\; \int_{t_i}^{t_{i+1}} \mathrm{d}t\, \left(\frac{1}{2}\,\dot{\boldsymbol{q}}^2(t) \;-\; V\big(\boldsymbol{q}(t)\big) \right) \tag{48}$$

$$\approx\; \frac{\Delta t_i}{2} \left(\frac{\boldsymbol{q}_{i+1} - \boldsymbol{q}_i}{\Delta t_i} \right)^2 \;-\; \Delta t_i\, V\left(\boldsymbol{q}_i\right) \tag{49}$$

where, in the last line, we defined, $\Delta t_i \equiv t_{i+1} - t_i$, and we *chose* a specific approximation for the Lagrangian. From the DEL update equations (45 and 46), we find,

$$\boldsymbol{p}_i \;=\; \frac{\boldsymbol{q}_{i+1} - \boldsymbol{q}_i}{\Delta t_i} \;+\; \Delta t_i\, V'\left(\boldsymbol{q}_i\right), \tag{50}$$

$$\boldsymbol{p}_{i+1} \;=\; \frac{\boldsymbol{q}_{i+1} - \boldsymbol{q}_i}{\Delta t_i}. \tag{51}$$

Note that by discretizing the integral over the potential, V, using a left Riemann sum, which only depends on $\boldsymbol{q}_i$, we can explicitly solve equation (50) for $\boldsymbol{q}_{i+1}$. As a result, the DEL update equations can be written as,

$$\boldsymbol{p}_{i+1} \;=\; \boldsymbol{p}_i \;-\; \Delta t_i\, V'(\boldsymbol{q}_i) \tag{52}$$

$$\boldsymbol{q}_{i+1} \;=\; \boldsymbol{q}_i \;+\; \Delta t_i\, \boldsymbol{p}_{i+1}. \tag{53}$$

This provides an integration scheme for Lagrangians of the form (47) that is guaranteed to be be symplectic, and hence exponentially stable for long time evolutions (see e.g. Blanes and Casas, 2025).

In this example, we chose a separable Lagrangian (47) and a particular discretization (49) to get an *explicit* equation for $\boldsymbol{q}_{i+1}$ in eq. (53). However, we should note that variational integrators work equally well for general Lagrangians and discretizations, with the only possible complication that the DEL update (50) might result in an *implicit* equation for $\boldsymbol{q}_{i+1}$ such that we require a root finder to extract it.

4 PN variational integrator

Before introducing the PN variational integrator with a physics-informed prior, we first review the standard PN-ODE approach, following Hennig et al. (2022).

4.1 PN-ODE solvers

Consider the following (for the moment general) ODE,

$$\dot{\boldsymbol{z}} \;=\; \boldsymbol{f}(\boldsymbol{z}, t). \tag{54}$$

with initial condition $\boldsymbol{z}_0 \equiv \boldsymbol{z}(t_0)$. Instead of a point-estimate, a probabilistic numerical (PN-)ODE solver aims to approximate the posterior distribution,

$$p\Big(\boldsymbol{z}(t) \,\Big|\, \boldsymbol{z}(t_0) = \boldsymbol{z}_0,\; \big\{\dot{\boldsymbol{z}}(t_i) = \boldsymbol{f}\big(\boldsymbol{z}(t_i), t_i\big)\big\}_{i=0}^{I}\Big), \tag{55}$$

for a given time-discretisation $\{t_i\}_{i=0}^I$. This posterior can efficiently be estimated by treating it as a Bayesian state estimation problem, through (extended) Kálmán filtering and smoothing (Tronarp et al., 2019). In this framework, a numerical ODE integrator is viewed as a state-space model over states defined as,

$$\boldsymbol{x} \;\equiv\; \left(\boldsymbol{z}^{(0)}, \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(d)}\right)^\dagger, \tag{56}$$

where superscripts denote time-derivatives, such that $\boldsymbol{z}^{(0)} \equiv \boldsymbol{z}$, $\boldsymbol{z}^{(1)} \equiv \dot{\boldsymbol{z}}$, and so on. This yields the following state-space representation,

$$\begin{aligned} \boldsymbol{x}_{i+1} &= \boldsymbol{A}_i(\boldsymbol{x}_i) + \boldsymbol{q}_i & (57)\\ \boldsymbol{y}_i &= \boldsymbol{H}_i(\boldsymbol{x}_i) + \boldsymbol{r}_i & (58) \end{aligned}$$

where we assume to have Gaussian noise,

$$\begin{aligned} \boldsymbol{q}_i &\sim \mathcal{N}(\boldsymbol{0}, \mathsf{Q}_i), & (59)\\ \boldsymbol{r}_i &\sim \mathcal{N}(\boldsymbol{0}, \mathsf{R}_i). & (60) \end{aligned}$$

This results in the probabilistic state-space model,

$$\begin{aligned} \mathbb{P}\left(\boldsymbol{x}_{i+1} \,|\, \boldsymbol{x}_i\right) &= \mathcal{N}\big(\boldsymbol{A}_i(\boldsymbol{x}_i), \mathsf{Q}_i\big), & (61)\\ \mathbb{P}\left(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i\right) &= \mathcal{N}\big(\boldsymbol{H}_i(\boldsymbol{x}_i), \mathsf{R}_i\big). & (62) \end{aligned}$$

The prediction model (57, 61) defines prior knowledge about the solution, for instance about the smoothness, and can be chosen such that the posterior mean reproduces classical solvers (see e.g. Schober et al., 2014). The measurement model (58, 62) defines the likelihood, and allows to condition on observations of the (residual of) the ODE. Defining selection vectors, $\boldsymbol{e}^{(n)}$, such that, $\boldsymbol{e}^{(n)} \cdot \boldsymbol{x} = \boldsymbol{z}^{(n)}$, the measurement model for the residual of the ODE (54) then reads,

$$\boldsymbol{H}(\boldsymbol{x}) \;=\; \boldsymbol{e}^{(1)} \cdot \boldsymbol{x} \;-\; \boldsymbol{f}\left(\boldsymbol{e}^{(0)} \cdot \boldsymbol{x}, t\right). \tag{63}$$

Conditioning on the ODE then means observing $\boldsymbol{y}_i = \boldsymbol{0}$.

As pointed out by Wang et al. (2020), observations of (63) do not define an exact information operator if $\boldsymbol{x}$ is not the true solution. As a result, PN-ODE solvers defined in this way violate the likelihood principle.

Bosch et al. (2022) demonstrated the benefits of adding different information in the measurement model, for instance, the conservation of energy. Given a Hamiltonian system with initial energy, $H(\boldsymbol{z}_0) = E$, a measurement model can include,

$$\boldsymbol{H}(\boldsymbol{x}) \;=\; H(\boldsymbol{e}^{(0)} \cdot \boldsymbol{x}) \;-\; E. \tag{64}$$

such that conditioning on energy conservation again means observing $\boldsymbol{y}_i = \boldsymbol{0}$. Note that, in contrast to (63), this is an exact information operator, since, if we assume to exactly know the initial energy, it should strictly be conserved throughout the entire evolution.

Assuming a Gaussian initial distribution,

$$\boldsymbol{x}_0 \;\sim\; \mathcal{N}(\boldsymbol{m}_0^{\mathrm{F}}, \mathsf{P}_0^{\mathrm{F}}) \tag{65}$$

and propagating distributions through the possibly nonlinear functions $\boldsymbol{A}_i$ and $\boldsymbol{H}_i$ by linearisation, the Kálmán filter and smoother now yields an efficient algorithm to estimate the PN-ODE posterior distribution (55). The algorithm is detailed for reference in Appendix B.

4.2 Physics-informed prior

In the classical PN-ODE solvers described above, we noted that the priors typically only include information about smoothness, and that including the ODE in the measurement model (58) yields an approximate rather than exact information operator. Therefore, one could consider moving the dynamics to the prior. A "natural" (physics-informed) way to define a prior for dynamics defined by an action, S, is to chose a prior proportional to $\exp(-S)$. This is natural in the sense that it is the least informative prior that still respects the physics (see e.g. Jaynes, 1957). Defining this prior over the discretized phase space coordinates, $\boldsymbol{z} \equiv (\boldsymbol{q}, \boldsymbol{p})^\dagger$, yields,

$$\mathbb{P}\left(\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_I\right) \;\propto\; \exp\left(-\tilde{S}\left(\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_I\right)\right). \tag{66}$$

This however requires the discretized action in terms of the phase space coordinates[1].

For example, using (6), we can write the discretized action (42) in terms of the phase space coordinates,

$$\begin{aligned} &\tilde{S}\left(\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_I\right)\\ &= \sum_{i=0}^{I-1} \int_{t_i}^{t_{i+1}} \mathrm{d}t \left\{\boldsymbol{p} \cdot \dot{\boldsymbol{q}} \;-\; H(\boldsymbol{q}, \boldsymbol{p})\right\}\\ &\approx \sum_{i=0}^{I-1} \left\{\boldsymbol{p}_{i+1} \cdot \left(\boldsymbol{q}_{i+1} - \boldsymbol{q}_i\right) \;-\; \left(\frac{\boldsymbol{p}_{i+1}^2}{2} + V(\boldsymbol{q}_i)\right) \Delta t_i\right\}, \end{aligned}$$

where the specific discretization in the second line was specifically chosen such that the resulting dynamics is exactly equivalent to that of (49).

To make this prior (66) more practically suitable, we approximate it by a Gaussian via Laplace approximation. First, we expand the discretized action (42) around its mode. By construction of the DEL eqs. (45, 46), this mode corresponds to their solution, since $\delta\tilde{S}(\boldsymbol{z}_i^\star) = \boldsymbol{0}$. The second-order expansion of the discretized action is,

$$\tilde{S} \;\approx\; \tilde{S}_\star \;+\; \sum_{i=0}^{I-1} \sum_{j=0}^{I-1} \delta\boldsymbol{z}_i \cdot \frac{\partial^2 \tilde{S}}{\partial \boldsymbol{z}_i \partial \boldsymbol{z}_j} \cdot \delta\boldsymbol{z}_j \tag{67}$$

where the partial derivatives are evaluated at the DEL solution, $\boldsymbol{z}_\star$, and we defined $\delta\boldsymbol{z}_i \equiv \boldsymbol{z}_i - \boldsymbol{z}_i^\star$. From the form of the discretized action, we note that its Hessian,

$$\mathbb{H}_{ij} \;\equiv\; \frac{\partial^2 \tilde{S}}{\partial \boldsymbol{z}_i \partial \boldsymbol{z}_j} \tag{68}$$

will be tridiagonal in terms of the indices i and j, so we can rewrite the double sum in the expansion (67) as,

$$\tilde{S} \;\approx\; \tilde{S}_\star \;+\; \frac{1}{2} \sum_{i=0}^{I-1} \begin{bmatrix} \delta\boldsymbol{z}_i \\ \delta\boldsymbol{z}_{i+1} \end{bmatrix}^\dagger \begin{bmatrix} \mathbb{H}_{i\,i} & \mathbb{H}_{i\,i+1} \\ \mathbb{H}_{i+1\,i} & \mathbb{H}_{i+1\,i+1} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{z}_i \\ \delta\boldsymbol{z}_{i+1} \end{bmatrix}.$$

The Laplace approximation of (66) for the joint distribution of all $\{\boldsymbol{z}_i\}_{i=0}^I$, then reads,

$$\mathbb{P}\left(\begin{bmatrix} \boldsymbol{z}_i \\ \boldsymbol{z}_{i+1} \end{bmatrix}\right) \;=\; \mathcal{N}\left(\begin{bmatrix} \boldsymbol{z}_i^\star \\ \boldsymbol{z}_{i+1}^\star \end{bmatrix}, \begin{bmatrix} \mathbb{H}_{i\,i} & \mathbb{H}_{i\,i+1} \\ \mathbb{H}_{i+1\,i} & \mathbb{H}_{i+1\,i+1} \end{bmatrix}^{-1}\right),$$

[1] Here, we work in phase space to recover the variational integrator in the mean. However, one can define an equivalent prior over the trajectories (see Appendix C).

where, for simplicity, in the square brackets, we only wrote two consecutive phase-space coordinates, whereas we mean the entire vector of all $\{\boldsymbol{z}_i\}_{i=0}^{I}$.

By virtue of the conditioning properties of Gaussian distributions, the resulting transition probability (61) for the resulting prediction model conveniently yields,

$$\mathbb{P}\big(\boldsymbol{z}_{i+1} \,\big|\, \boldsymbol{z}_i = \boldsymbol{z}_i^\star\big) \;=\; \mathcal{N}\big(\boldsymbol{z}_{i+1}^\star, \mathbb{H}^{-1}_{i+1\,i+1}\big) \tag{69}$$

where the mean is given by the DEL update (45, 46) and the transition covariance (Q_i) can be computed from the inverse of the Hessian of the discretized action.

For example, for the discretization of the action (49), the DEL update is given by eqs. (52 and 53), and the covariance matrix reads,

$$\mathsf{Q}_i \;=\; \mathbb{H}^{-1}_{i+1\,i+1} \;=\; \begin{bmatrix} \Delta t_i + \lambda^2 & 1 \\ 1 & \lambda^2 \end{bmatrix}. \tag{70}$$

Note that without regularisation (λ^2) this matrix is not positive-definite, which is a reflection of the fact that the momentum equation should be exactly satisfied by definition of the momentum.

4.3 PN Variational Integrator

We define the PN variational integrator (PN-VI) by a particular choice of prediction and measurement model. As prediction model (57, 61), we chose the physics-informed prior from Section 4.2. As such, the prior already includes the dynamics. As measurement model (58, 62), we only consider the conservation of energy (64), resulting in an exact information operator.

To propagate the probability distributions through the (possibly nonlinear) transition and measurement functions, we also need their Jacobians. The Jacobian of the transition function (the DEL update map) was already computed in the symplecticity proof in Appendix A,

$$\mathsf{A}_i \;=\; -\begin{pmatrix} \tilde{L}_{12}^{-1}\tilde{L}_{11} & \tilde{L}_{12}^{-1} \\ \tilde{L}_{22}\,\tilde{L}_{12}^{-1}\tilde{L}_{11} - \tilde{L}_{21} & \tilde{L}_{22}\,\tilde{L}_{12}^{-1} \end{pmatrix}, \tag{71}$$

where we defined, $\tilde{L}_{ab} \equiv \partial_a \partial_b \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1})$. Similarly, the Jacobian of the measurement function (64) reads,

$$\mathsf{H}_i \;=\; \begin{pmatrix} \partial_{\boldsymbol{q}} H & \partial_{\boldsymbol{p}} H \end{pmatrix}. \tag{72}$$

For example, for the discrete Lagrangian (49), defined in the example variational integrator in Section 3.1, the transition and measurement Jacobians read,

$$\mathsf{A}_i \;=\; \begin{pmatrix} 1 - \Delta t_i^2\, V''(\boldsymbol{q}_i) & \Delta t_i \\ -\Delta t_i\, V''(\boldsymbol{q}_i) & 1 \end{pmatrix}, \tag{73}$$

$$\mathsf{H}_i \;=\; \begin{pmatrix} V'(\boldsymbol{q}_i) & \boldsymbol{p}_i \end{pmatrix}. \tag{74}$$

The resulting PNM can be thought of as a variational integrator (the prior) combined in a probabilistic way with a projection onto the constant-energy manifold (the measurement; see e.g. Hairer et al., 2006). Note that the prior mean is symplectic (Appendix A), but there are no guarantees for the symplecticity of the posterior mean.

5 Demonstration

We demonstrate the PN variational integrator on the single pendulum example from Section 2.7.1. The code for the demonstration is available online[2].

Figure 1 shows the resulting trajectories in phase space and Figure 2 shows the phase and energy errors for four different solvers and the ground truth. The four solvers are: the PN-VI posterior, the PN-VI prior (i.e. a classical variational integrator in the mean), a PN-ODE solver conditioned on the ODE and the Hamiltonian constraint (following Bosch et al., 2022), once for the first-order ODE (eqs. 19 and 20) with a once-integrated Wiener prior, and once for the second-order ODE (eq. 21) with a twice-integrated Wiener prior.

We see that the physics-informed PN-VI prior already provides an excellent estimate, but with an oscillatory error in the Hamiltonian constraint, as expected from a symplectic integrator (Blanes and Casas, 2025). Moreover, we see that the PN-VI posterior outperforms both the first-order and second-order PN-ODE solvers in satisfying the Hamiltonian constraint, with a penalty in the phase error that is comparable to the second-order PN-ODE solver. This can be understood since the PN-VI already starts from a better prior estimate, and in the measurement step is only conditioned on the Hamiltonian constraint and not also on the ODE.

Regarding computational cost, it should be noted that PN-VI evolves a $2N$-dimensional state-space, with N the dimension of $\boldsymbol{q}$ (i.e. $N = 1$ for the single pendulum), while the first-order and second-order PN-ODE solvers evolve a $4N$ and $3N$-dimensional state-space, respectively. Hence, PN-VI can be significantly faster for high-dimensional simulations, such as many-particle systems. However, in practice, the cost of PN-VI will probably be dominated the evaluation of the DEL map (45 and 46), especially if (45) has to be solved implicitly.

6 Conclusion

Many differential equations encountered in science originate from underlying physical principles that encode substantially richer structure than the equation alone. In particular, Hamiltonian dynamics arises from variational principles, possesses a symplectic structure, and exhibits conservation laws induced by symmetries. Exclusively focussing on the differential equation, as is common in standard PN-ODE solvers that condition on residuals, risks overlooking this additional structure.

In this paper, we revisited variational integrators and proposed a PN extension based on a physics-informed prior. The prior already captures the dynamics, such that we only still need to condition on the Hamiltonian constraint, which yields an exact information operator. Although the physics-informed prior is slightly more cumbersome to construct, the resulting solver outperforms the more ad hoc PN-ODE approach of adding Hamiltonian constraints from Bosch et al. (2022).

[2] github.com/Poincare-code/paper_ProbNum26_PN4HD

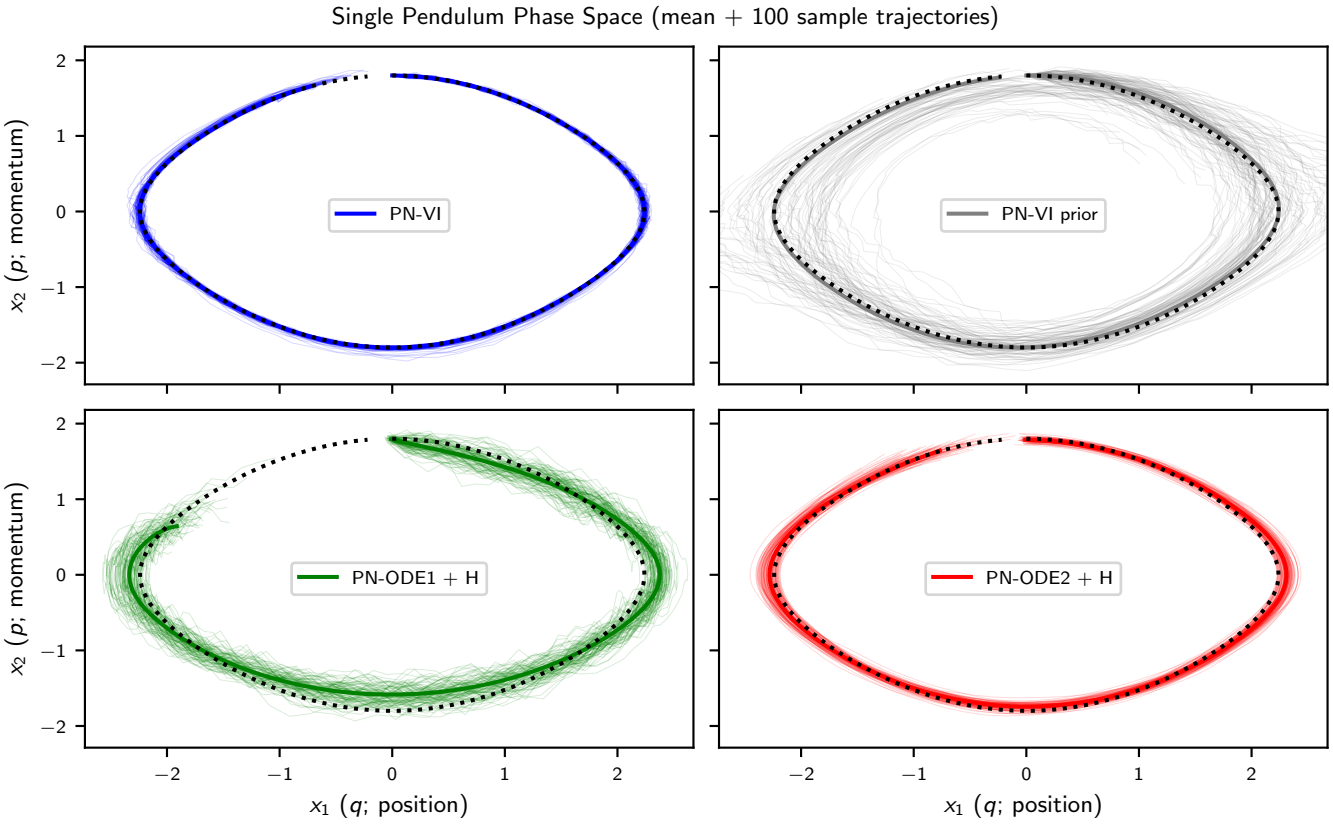

Figure 1: Phase space trajectories of the single pendulum, obtained using four different solvers. The thick solid lines shows the mean of the distribution, while the thin lines show 100 samples. The black dotted line indicates the ground truth in each plot.

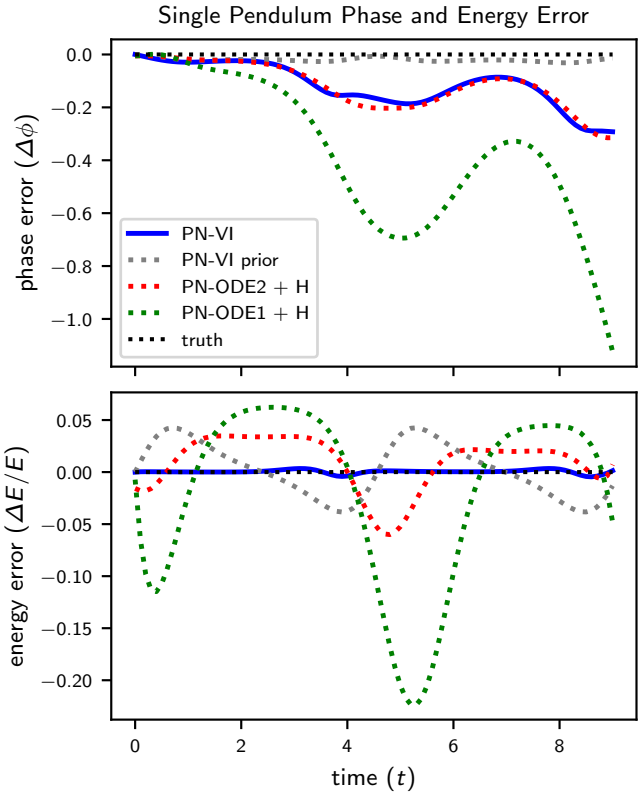

Figure 2: Phase and energy error for the single pendulum, obtained using four different solvers. The phase is defined as $\phi \equiv \arctan(q/p)$, and differences are taken with respect to the true solution.

In addition, we discussed the symmetry-based Bayesian ODE framework of Wang et al. (2020) for Hamiltonian dynamics, clarifying how their restriction to solvable associated Lie-algebras relates to Liouville-integrability. This connection highlights a broader theme: symmetry and structure are not merely properties to be preserved numerically, but can actively inform the construction of probabilistic numerical methods.

At the same time, several open questions remain. In particular, the precise role of variational structure in defining numerical uncertainty and the development of scalable inference procedures, for instance, for partial differential equations both warrant further research.

We only considered conservative Hamiltonian systems. However, PN variational integrator can also be applied to non-conservative systems. In that case, we need another measurement model (since we can no longer use energy conservation). We can use, for instance, another conservation law, such as the conservation of (angular) momentum, or, if energy is not conserved but still modelled, we can condition on that model.

Overall, this work advocates for a shift in probabilistic numerics: from residual-based formulations towards structure-aware approaches that are grounded in the physical origin of the problem. By combining ideas from variational mechanics, geometric integration, and Bayesian inference, this direction offers a promising pathway towards probabilistic solvers that are both physics-informed and statistically principled.

Acknowledgements

This research was partially supported by the Research Foundation – Flanders (FWO; Grant No. I000725N, I002123N, and 1178926N).

References

V. I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer, 1989.

S. Blanes and F. Casas. *A concise introduction to geometric numerical integration*. CRC press, 2025.

N. Bosch, F. Tronarp, and P. Hennig. Pick-and-Mix Information Operators for Probabilistic ODE Solvers . In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 10015–10027. PMLR, 28–30 Mar 2022.

J. Cockayne, C. J. Oates, T. J. Sullivan, and M. Girolami. Bayesian probabilistic numerical methods. *SIAM review*, 61(4):756–789, 2019.

E. Hairer, C. Lubich, and G. Wanner. Geometric numerical integration: structure-preserving algorithms for ordinary differential equations. *Springer series in computational mathematics*, 2006.

P. Hennig, M. A. Osborne, and M. Girolami. Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179), 2015.

P. Hennig, M. A. Osborne, and H. P. Kersting. *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press, 2022.

E. T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.*, 106:620–630, May 1957.

M. Schober, D. Duvenaud, and P. Hennig. Probabilistic ODE solvers with Runge-Kutta means. *Advances in neural information processing systems*, 27, 2014.

F. Tronarp, H. Kersting, S. Särkkä, and P. Hennig. Probabilistic solutions to ordinary differential equations as nonlinear Bayesian filtering: a new perspective. *Statistics and Computing*, 29(6):1297–1315, 2019.

J. Wang, J. Cockayne, and C. J. Oates. A role for symmetry in the Bayesian solution of differential equations. *Bayesian Analysis*, 15(4), 2020.

G. Zhong and J. E. Marsden. Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators. *Physics Letters A*, 133(3):134–139, 1988.

A Symplecticity of the DEL map

We show that the discrete Euler-Lagrange (DEL) map, resulting form a variational integrator as described in Section 3, and defined by equations (45 and 46), is a discretized symplectic map.

The discretized Jacobian is defined as,

$$\Phi_i \equiv \begin{pmatrix} \partial_{\boldsymbol{q}_i} \boldsymbol{q}_{i+1} & \partial_{\boldsymbol{p}_i} \boldsymbol{q}_{i+1} \\ \partial_{\boldsymbol{q}_i} \boldsymbol{p}_{i+1} & \partial_{\boldsymbol{p}_i} \boldsymbol{p}_{i+1} \end{pmatrix} = \begin{pmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{C} & \mathsf{D} \end{pmatrix}. \tag{75}$$

Using the newly introduced notation, the condition of symplecticity (i.e. $\Phi_i^{\mathrm{T}} \mathsf{J} \Phi_i = \mathsf{J}$, see 10), is equivalent to,

$$\mathsf{A}^\dagger \mathsf{C} - \mathsf{C}^\dagger \mathsf{A} = 0, \tag{76}$$
$$\mathsf{B}^\dagger \mathsf{D} - \mathsf{D}^\dagger \mathsf{B} = 0, \tag{77}$$
$$\mathsf{A}^\dagger \mathsf{D} - \mathsf{C}^\dagger \mathsf{B} = 1. \tag{78}$$

We will show that these three conditions indeed hold for the discretized Euler-Lagrange equations.

From the DEL update equations (45 and 46), we find,

$$\mathrm{d}\boldsymbol{p}_i = -\tilde{L}_{11}\,\mathrm{d}\boldsymbol{q}_i - \tilde{L}_{12}\,\mathrm{d}\boldsymbol{q}_{i+1} \tag{79}$$
$$\mathrm{d}\boldsymbol{p}_{i+1} = \tilde{L}_{21}\,\mathrm{d}\boldsymbol{q}_i + \tilde{L}_{22}\,\mathrm{d}\boldsymbol{q}_{i+1} \tag{80}$$

where we defined, $\tilde{L}_{ab} \equiv \partial_a \partial_b \tilde{L}(\boldsymbol{q}_i, \boldsymbol{q}_{i+1})$. Note that, since partial derivatives commute, $\tilde{L}_{ab} = \tilde{L}_{ba}^\dagger$. Hence, we can rearrange the differentials,

$$\mathrm{d}\boldsymbol{q}_{i+1} = -\tilde{L}_{12}^{-1}\tilde{L}_{11}\,\mathrm{d}\boldsymbol{q}_i - \tilde{L}_{12}^{-1}\,\mathrm{d}\boldsymbol{p}_i \tag{81}$$
$$\mathrm{d}\boldsymbol{p}_{i+1} = (\tilde{L}_{21} - \tilde{L}_{22}\tilde{L}_{12}^{-1}\tilde{L}_{11})\mathrm{d}\boldsymbol{q}_i - \tilde{L}_{22}\tilde{L}_{12}^{-1}\,\mathrm{d}\boldsymbol{p}_i \tag{82}$$

such that we can read the entries of the Jacobian,

$$\mathsf{A} \equiv \partial_{\boldsymbol{q}_i}\boldsymbol{q}_{i+1} = -\tilde{L}_{12}^{-1}\tilde{L}_{11} \tag{83}$$
$$\mathsf{B} \equiv \partial_{\boldsymbol{p}_i}\boldsymbol{q}_{i+1} = -\tilde{L}_{12}^{-1} \tag{84}$$
$$\mathsf{C} \equiv \partial_{\boldsymbol{q}_i}\boldsymbol{p}_{i+1} = \tilde{L}_{21} - \tilde{L}_{22}\tilde{L}_{12}^{-1}\tilde{L}_{11} \tag{85}$$
$$\mathsf{D} \equiv \partial_{\boldsymbol{p}_i}\boldsymbol{p}_{i+1} = -\tilde{L}_{22}\tilde{L}_{12}^{-1} \tag{86}$$

Now, by direct computation, we find that,

$$\mathsf{A}^\dagger\mathsf{C} = -\tilde{L}_{11}^\dagger - \tilde{L}_{11}^\dagger\tilde{L}_{12}^{-\dagger}\tilde{L}_{22}\tilde{L}_{12}^{-1}\tilde{L}_{11} = \mathsf{C}^\dagger\mathsf{A} \tag{87}$$
$$\mathsf{B}^\dagger\mathsf{D} = \tilde{L}_{12}^{-\dagger}\tilde{L}_{22}\tilde{L}_{12}^{-1} = \mathsf{D}^\dagger\mathsf{B} \tag{88}$$
$$\mathsf{A}^\dagger\mathsf{D} = \tilde{L}_{11}^\dagger\tilde{L}_{12}^{-\dagger}\tilde{L}_{22}\tilde{L}_{12}^{-1} = \mathsf{C}^\dagger\mathsf{B} + 1 \tag{89}$$

proving that the DEL map is indeed symplectic.

B PN-ODE solvers

We detail the Kálmán filter and smoother approach to efficiently solve the probabilistic state-space model (defined by eqs. 61 and 62) that is central to PN-ODE solvers. For more details, see e.g. Tronarp et al. (2019) or Hennig et al. (2022).

Kálmán predictor The Kálmán predictor defines the probability of the state, $\boldsymbol{x}_i$, based on all previous observations, $\boldsymbol{y}_{0:i-1}$, as a Gaussian,

$$\mathbb{P}\left(\boldsymbol{x}_i \,|\, \boldsymbol{y}_{0:i-1}\right) = \mathcal{N}\left(\boldsymbol{m}_i^{\mathrm{P}}, \mathsf{P}_i^{\mathrm{P}}\right). \tag{90}$$

The predictor mean, $\boldsymbol{m}^{\mathrm{P}}_{i+1}$, and covariance, $\mathsf{P}^{\mathrm{P}}_{i+1}$, can be expressed in terms of the previous-step filter mean, $\boldsymbol{m}^{\mathrm{F}}_{i}$, and covariance, $\mathsf{P}^{\mathrm{F}}_{i}$, under the transition function, $\boldsymbol{A}_i$, and the model noise covariance, Q_i,

$$\boldsymbol{m}^{\mathrm{P}}_{i+1} \;=\; \boldsymbol{A}_i\left(\boldsymbol{m}^{\mathrm{F}}_{i}\right) \tag{91}$$

$$\mathsf{P}^{\mathrm{P}}_{i+1} \;=\; \mathsf{A}_i\,\mathsf{P}^{\mathrm{F}}_{i}\,\mathsf{A}_i^{\dagger} \,+\, \mathsf{Q}_i \tag{92}$$

where we defined A_i as the Jacobian of the transition function, $\boldsymbol{A}_i$, evaluated at the filter mean, $\boldsymbol{m}^{\mathrm{F}}_{i}$.

Measurement model The measurement model defines the probability of the measurement, $\boldsymbol{y}_i$, based on all previous observations, $\boldsymbol{y}_{0:i-1}$, as a Gaussian,

$$\mathbb{P}\left(\boldsymbol{y}_i \,|\, \boldsymbol{y}_{0:i-1}\right) \;=\; \mathcal{N}\left(\boldsymbol{w}_i, \mathsf{S}_i\right). \tag{93}$$

The measurement mean, $\boldsymbol{w}_i$, and covariance, S_i, can be expressed in terms of the predictor mean, $\boldsymbol{m}^{\mathrm{P}}_{i}$, and covariance, $\mathsf{P}^{\mathrm{P}}_{i}$, under the measurement function, $\boldsymbol{H}_i$, and the measurement noise covariance, R_i,

$$\boldsymbol{w}_i \;=\; \boldsymbol{H}_i(\boldsymbol{m}^{\mathrm{P}}_{i}) \tag{94}$$

$$\mathsf{S}_i \;=\; \mathsf{H}_i\,\mathsf{P}^{\mathrm{P}}_{i}\,\mathsf{H}_i^{\dagger} \,+\, \mathsf{R}_i \tag{95}$$

where we defined H_i as the Jacobian of the measurement function, $\boldsymbol{H}_i$, evaluated at the predictor mean, $\boldsymbol{m}^{\mathrm{P}}_{i}$.

Kálmán filter The Kálmán filter defines the probability of the state, $\boldsymbol{x}_i$, based on all previous and the current observation, $\boldsymbol{y}_{0:i}$, as a Gaussian,

$$\mathbb{P}\left(\boldsymbol{x}_i \,|\, \boldsymbol{y}_{0:i}\right) \;=\; \mathcal{N}\left(\boldsymbol{m}^{\mathrm{F}}_{i}, \mathsf{P}^{\mathrm{F}}_{i}\right). \tag{96}$$

The filter mean, $\boldsymbol{m}^{\mathrm{F}}_{i}$, and covariance, $\mathsf{P}^{\mathrm{F}}_{i}$, can be expressed in terms of the current-step predictor mean, $\boldsymbol{m}^{\mathrm{P}}_{i}$, and covariance, $\mathsf{P}^{\mathrm{P}}_{i}$, updated by the measurement,

$$\mathsf{K}_i \;\equiv\; \mathsf{P}^{\mathrm{P}}_{i}\,\mathsf{H}_i^{\dagger}\,\mathsf{S}_i^{-1} \tag{97}$$

$$\boldsymbol{m}^{\mathrm{F}}_{i} \;=\; \boldsymbol{m}^{\mathrm{P}}_{i} \,+\, \mathsf{K}_i\left(\boldsymbol{y}_i - \boldsymbol{w}_i\right) \tag{98}$$

$$\mathsf{P}^{\mathrm{F}}_{i} \;=\; \mathsf{P}^{\mathrm{P}}_{i} \,-\, \mathsf{K}_i\,\mathsf{S}_i\,\mathsf{K}_i^{\dagger} \tag{99}$$

Rauch-Tung-Striebel smoother The Rauch-Tung-Striebel (RTS) smoother defines the probability of the state, $\boldsymbol{x}_i$, based on all observations, $\boldsymbol{y}_{0:I}$, as a Gaussian,

$$\mathbb{P}\left(\boldsymbol{x}_i \,|\, \boldsymbol{y}_{0:I}\right) \;=\; \mathcal{N}\left(\boldsymbol{m}^{\mathrm{S}}_{i}, \mathsf{P}^{\mathrm{S}}_{i}\right). \tag{100}$$

The smoother mean, $\boldsymbol{m}^{\mathrm{S}}_{i}$, and covariance, $\mathsf{P}^{\mathrm{S}}_{i}$, can be expressed in terms of the next-step smoother and predictor means, $\boldsymbol{m}^{\mathrm{S}}_{i+1}$ and $\boldsymbol{m}^{\mathrm{P}}_{i+1}$, and covariances, $\mathsf{P}^{\mathrm{S}}_{i+1}$ and $\mathsf{P}^{\mathrm{P}}_{i+1}$, and in terms of the current-step filter mean, $\boldsymbol{m}^{\mathrm{F}}_{i}$, and covariance, $\mathsf{P}^{\mathrm{F}}_{i}$,

$$\mathsf{G}_i \;\equiv\; \mathsf{P}^{\mathrm{F}}_{i}\,\mathsf{A}_i^{\dagger}\left(\mathsf{P}^{\mathrm{P}}_{i+1}\right)^{-1} \tag{101}$$

$$\boldsymbol{m}^{\mathrm{S}}_{i} \;=\; \boldsymbol{m}^{\mathrm{F}}_{i} \,+\, \mathsf{G}_i\left(\boldsymbol{m}^{\mathrm{S}}_{i+1} \,-\, \boldsymbol{m}^{\mathrm{P}}_{i+1}\right) \tag{102}$$

$$\mathsf{P}^{\mathrm{S}}_{i} \;=\; \mathsf{P}^{\mathrm{F}}_{i} \,+\, \mathsf{G}_i\left(\,\mathsf{P}^{\mathrm{S}}_{i+1} \,-\, \mathsf{P}^{\mathrm{P}}_{i+1}\right)\mathsf{G}_i^{\dagger} \tag{103}$$

Combining the prediction, measurement, and filter steps in a forward pass through the measurements and adding the smoother steps in a backward pass provides a linear-in-time solution algorithm for the (linearized) probabilistic state-space model (eqs. 61 and 62).

C Equivalent physics-informed prior over trajectories

In Section 4.2, the physics-informed prior (66) was taken over phase space (i.e. containing both positions and momenta) to recover the variational integrator from Section 3 in the mean. However, when considering Gaussian process regression in "batch mode" instead of the Kálmán filter and smoother approach, it is more natural to consider a prior over the trajectories, $\boldsymbol{q}(t)$,

$$\mathbb{P}[\boldsymbol{q}] \;\propto\; \exp\left(-S[\boldsymbol{q}]\right). \tag{104}$$

A discretized approximation to (104) can be made,

$$\mathbb{P}\left(\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_I\right) \;\propto\; \exp\Big(-\tilde{S}\left(\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_I\right)\Big), \tag{105}$$

where we used the discretized action (42). To make this suited for practical computations, we make a Laplace approximation around the mode, which coincides with solution of the DEL eqs. (45, 46), since $\delta\tilde{S}(\boldsymbol{q}^{\star}_i) = \boldsymbol{0}$,

$$\mathbb{P}(\boldsymbol{q}_i) \;\approx\; \mathcal{N}\left(\boldsymbol{q}^{\star}_i, \Lambda^{-1}_{ij}\right), \tag{106}$$

The precision matrix can be derived from the second-order expansion of the discretized action around the DEL solution,

$$\tilde{S} \;\approx\; \tilde{S}_{\star} + \sum_{i=0}^{I-1}\sum_{j=0}^{I-1}\left(\boldsymbol{q}_i - \boldsymbol{q}^{\star}_i\right)\cdot\frac{\partial^2\tilde{S}}{\partial\boldsymbol{q}_i\partial\boldsymbol{q}_j}\cdot\left(\boldsymbol{q}_j - \boldsymbol{q}^{\star}_j\right) \tag{107}$$

where partial derivatives of the discretized action are evaluated at the DEL solution, $\boldsymbol{q}_{\star}$. The precision matrix of the approximate Gaussian distribution (106) is then given by the Hessian of the discretized action,

$$\mathbb{H}_{ij} \;\equiv\; \frac{\partial^2\tilde{S}}{\partial\boldsymbol{z}_i\partial\boldsymbol{z}_j}. \tag{108}$$

For example, for the discrete Lagrangian (49), defined in the example variational integrator in Section 3.1, the Hessian of the discretized action is tridiagonal, where the diagonal elements are given by,

$$\Lambda_{i\,i} \;=\; \frac{\partial^2\tilde{S}}{\partial\boldsymbol{q}_i^2} \;=\; \frac{1}{\Delta t_i} + \frac{1}{\Delta t_{i-1}} - \Delta t_i\,V''\left(\boldsymbol{q}^{\star}_i\right), \tag{109}$$

and where the only non-zero off-diagonal elements are,

$$\Lambda_{i\,i+1} \;=\; \frac{\partial^2\tilde{S}}{\partial\boldsymbol{q}_i\partial\boldsymbol{q}_{i+1}} \;=\; -\frac{1}{\Delta t_i}. \tag{110}$$

This provides a physics-informed prior over trajectories. Although in general in classical mechanics the extremum of the action is a saddle point, it can be shown that its Hessian, and thus the precision matrix, is guaranteed to be positive semi-definite for sufficiently small Δt_i (Arnold, 1989).