Joint Bayesian Inference on Lagrangian Physics and Trajectories

Michael Obermayr
Institute for Machine Learning and Neural Computation
Technical University Graz
Graz, Austria
michael.obermayr@tugraz.at

Robert Peharz
Institute for Machine Learning and Neural Computation
Graz, Austria
robert.peharz@tugraz.at

Abstract

Numerical integration and ODE discovery are two sides of the same coin—converse problems of finding trajectories from known physics versus inferring physics from observed trajectories. Although these problems have been extensively studied in isolation, they can be unified through the minimization of a common quantity: the Euler–Lagrange residual. In this paper, we build on this insight and introduce the Integrated Squared Action Residual (ISAR), which enables both tasks to be performed simultaneously. We formulate numerical integration and model discovery as a joint Bayesian inference problem, allowing for the systematic incorporation of physical prior knowledge and domain constraints in settings with sparse and noisy observations, where traditional approaches typically fail. While we demonstrate the performance on two mechanical toy problems, it can be readily extended towards multiphysics systems including dissipative dynamics.

1 Introduction

Physicists are primarily concerned with two activities: discovering models that describe nature and using those models to predict future events. The former is an active field with recent powerful data-driven approaches (Brunton et al., 2016; Raissi et al., 2017; M. Raissi, 2019; Greydanus et al., 2019), while the latter relies mainly on classical numerical integrators with notable learning-based extensions (Chen et al., 2018). Interestingly, these two activities are rarely recognized as two sides of the same coin.

Both can be seen as algorithms minimizing the residual of a differential equation evaluated over the observed or proposed trajectory. While equation discovery does this very explicitly, numerical integrators inherently minimize discretization error, which can also be seen as a physical residual. Variational integrators, for instance, are designed to minimize the local residual of the Euler-Lagrange equations, which vanishes only if the trajectory is perfectly explained by the underlying physics.

We propose using the global Euler-Lagrange residual as a unified loss function for both numerical integration (the forward problem) and equation discovery (the backward problem). While this loss function can be used for each problem in isolation, it is now possible to tackle both simultaneously. This becomes particularly useful when dealing with systems with partially known dynamics and sparse, noisy observations of trajectories. For such systems, numerical integration is inapplicable due

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

to incomplete physical equations, and standard discovery methods struggle because derivatives cannot be reliably estimated via finite differences. Moreover, solutions may be highly underdetermined, especially when the applied algorithm relies solely on observations and not all available partial knowledge of the dynamics is utilized.

To address these challenges, we employ a fully Bayesian treatment to infer the complete solution distribution rather than a single point estimate. We define a probabilistic model of the form Physics $\phi \rightarrow$ Trajectory $\theta \rightarrow$ Data, where ϕ and θ are linked by a physics likelihood derived from the global Euler-Lagrange residual. This framework allows for data-efficient inference over the full posterior of plausible physics and trajectories, given sparse noisy observations and weak priors, which are derived from partial system knowledge.

While preliminary results on mechanical toy problems are promising, the remaining challenge consist of reliably estimating the normalizing constant in the proposed physics likelihood to counteract the observed posterior distortion.

2 Background

2.1 Principle of Stationary Action

Our methodology is based on the Principle of Stationary Action, which is fundamental to all reversible non-dissipative systems. It states that for a physical trajectory $x(t)$ connecting two fixed points $x(t_0)$ and $x(t_1)$, the action functional $S[x(t)]$ must be stationary under infinitesimal variations of the path. It provides a rigorous method for solving boundary-value problems in classical mechanics.

Using variational calculus, one can derive the Euler-Lagrange Equations of the first kind. For a system governed by a Lagrangian function L, subject to m constraints f_j, weighted by multipliers λ_j these equations read:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = \sum_{j=1}^{m} \lambda_j \frac{\partial f_j}{\partial x} \tag{1}$$

This transformation of the Stationary Action Principle into a system of differential equations allows to address both boundary-value and initial-value problems.

In a "Natural Lagrangian System," where coordinates x are holonomic and the Lagrangian has no explicit time dependence, the right-hand side of the equation vanishes. For simplicity, we limit our examples to such systems, although the framework is readily extendable.

2.2 ISAR: Integrated-Squared-Action-Residual

We now propose to measure the global violation of the Principle of Stationary Action by integrating the squared residual of the Euler-Lagrange Equations, which we define in Eq. 2. It is a strong form residual loss for Lagrangian systems and similar losses are established in literature (Lutter et al., 2019; Cranmer et al., 2020; Kharazmi et al., 2019). The integral, which we will call **ISAR** is non-negative and only zero, if a trajectory $x(t)$ perfectly satisfies the equations of motion, derived from the Lagrangian function, or vice versa.

$$ISAR := \int_{t_0}^{t_1} \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x}\right)^2 dt \tag{2}$$

$$= \int_{t_0}^{t_1} \left(\ddot{x} \cdot \frac{\partial^2 L}{\partial \dot{x}^2} + \dot{x} \cdot \frac{\partial^2 L}{\partial \dot{x} \partial x} - \frac{\partial L}{\partial x}\right)^2 dt \tag{3}$$

Evaluation of *ISAR* is straightforward with modern autodifferentiation libraries, where differential operators can be readily evaluated on trajectories. This leaves us with a one-dimensional integral for each state-space dimension, which can be approximated numerically.

The unit of the resulting integral is $\frac{kg^2 m^2}{s^3} = N^2 s$, we can interpret it as the integrated squared discrepancy in predicted vs. observed force. Note, that Lagrangian dynamics are invariant under scaling and the addition of a gauge, leading to identifiability issues. To avoid trivial solutions (such as $L = 0$), we need either strong physical priors or measures to prevent a collpase of the phase space volume, such as normalizing *ISAR* with the generalized mass-matrix $\frac{\partial^2 L}{\partial \dot{x} \partial \dot{x}}$. This mass-normalized *ISAR* can be interpreted as the integrated discrepancy in acceleration (unit $\frac{m^2}{s^3}$) and is the same training objective used by Cranmer et al. (2020) to learn Lagrangians with neural networks.

The integral *ISAR* is differentiable with respect to both the trajectory parameters and the Lagrangian parameters. Thus, one may apply gradient-based optimization methods. If the Lagrangian function and boundary/initial conditions are given, one can minimizing *ISAR* with respect to trajectory parameters to find the unknown trajectory. This essentially corresponds to solving a boundary-value/initial-value problem via numerical integration. Conversely, if a trajectory is given, one can minimize with respect to the Lagrangian parameters to arrive at the governing Lagrangian, essentially performing equation discovery.

3 Methodology

3.1 Probabilistic Model

To infer jointly the trajectory parameters θ and physics parameters ϕ, we propose a physics likelihood based on *ISAR*:

$$p(\theta|\phi) = p(\theta) \cdot \exp(-\lambda \cdot ISAR(\phi, \theta)) \cdot \frac{1}{Z_\phi} \tag{4}$$

Here $p(\theta)$ is the prior over trajectories which may encourage smoothness or limit the class of possible functions. Z_ϕ is the normalizing constant, which is nontrivial to estimate and remains the main hurdle in the current work.

Together with a physics prior $p(\phi)$ and a gaussian observation likelihood $p(Data|\theta)$ we propose the following posterior distribution for trajectory parameters θ and Lagrangian parameters ϕ given the noisy observations $Data$.

$$p(\phi, \theta|Data) = \frac{p(\phi)p(\theta|\phi)p(Data|\theta)}{p(Data)} \tag{5}$$

The model hierarchy $\phi \rightarrow \theta \rightarrow Data$ resembles the conventional causal direction in physics: physical equations ϕ determine trajectories θ, whose measurement produces observed $Data$. By defining $p(\theta|\phi)$ we ensure that measurements remain conditionally independent of the underlying physics if the trajectory is given, while establishing a connection between the random variables ϕ and θ.

3.2 Parametrization and Inference

We parametrize the trajectories using third-order Hermite splines, as they are simple, and are continuously and analytically differentiable. For the Lagrangian, we begin with a physically motivated ansatz and learn its physical parameters. In scenarios where physical knowledge is minimal, this framework remains flexible: the Lagrangian may instead be represented by a structured neural network, similar to the approach by Cranmer et al. (2020).

For inference, we draw samples from the log-posterior, using the No-U-Turn Sampler (NUTS), a common implementation of Hamiltonian Monte Carlo (HMC). Additionally, we compute a MAP estimate from a single starting-point using the L-BFGS-B algorithm. This serves both as a quick sanity check and to highlight the loss of information when favoring a single point estimate over the full posterior distribution.

We can neglect the evidence term $log(p(Data))$ in the log-posterior, as it is constant. In our preliminary results, we further omit the normalizing constant of the physics likelihood Z_ϕ, due to difficulties in estimation. This results in an unnormalized physics likelihood, which distorts the posterior, as we observe in our results.

3.3 Experiments

3.3.1 Harmonic Oscillator

As an introductory example, we begin by studying the harmonic oscillator. First we simulate a single trajectory $q(t)$ with mass $m = 1.5$ and a spring constant $k = 5.0$, which we observe at 8 equally spaced points with a noise level of 1 %. We parametrize the Lagrangian function with the correct harmonic oscillator ansatz in Eq. 6 with learnable parameters $\phi = \{m, k\}$ and the trajectory with 125 3rd order Hermite splines, to allow for high-quality approximations of high-order frequencies. Our goal is to recover the posterior distribution of physics parameters ϕ and trajectory parameters θ from the noisy observations as well as broad priors over ϕ and a smoothness prior on the trajectory parameters θ.

$$L(q, \dot{q}) = \frac{1}{2} m \dot{q}^2 - \frac{1}{2} k q^2 \tag{6}$$

The solutions to the harmonic oscillator are sinusoidal functions with the frequency $\omega = \sqrt{\frac{k}{m}}$. Note, that the observations are far too sparse to estimate reliable derivatives from finite differences, many established Physics Discovery methods are doomed to fail. Also note, that there is an infinite set of frequencies and thus trajectories satisfying the Lagrangian for the given observations and only by utilizing further assumptions the infinite set collapses to a limited set of posterior modes. The key advantage of our method is to define these assumptions explicitly and systematically with the priors on $p(\phi)$ and $p(\theta)$.

Although the MAP Estimate can only identify one mode, posterior samples obtained via NUTS reveal a multitude of plausible frequencies, as can be seen in Fig. 1. Many solutions might be missed, if one does not perform proper Bayesian Inference.

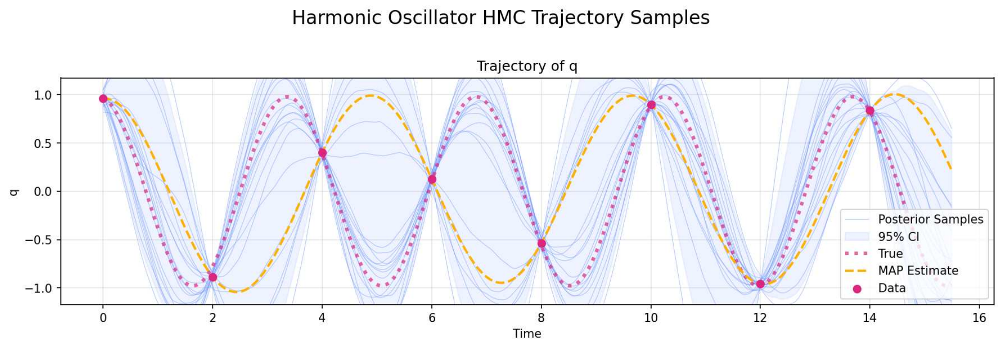

Figure 1: Trajectories generated from marginal posterior samples for the harmonic oscillator system

3.3.2 Double Pendulum

Next we address the chaotic double pendulum system, defined by the Lagrangian function in Eq. 7. Again, we provide this Lagrangian as the model ansatz with physical parameters $\phi = \{m_1, m_2, l_1, l_2\}$ and parametrize the trajectories with 301 3rd order Hermite splines. From noisy observations of a single simulated trajectory and priors on ϕ and θ we infer the joint posterior distribution over physical and trajectory parameters via NUTS.

$$\begin{aligned} \mathcal{L} =& \frac{1}{2}(m_1 + m_2) l_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 l_2^2 \dot{\theta}_2^2 + m_2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \cos(\theta_1 - \theta_2) \\ &+ (m_1 + m_2) g l_1 \cos\theta_1 + m_2 g l_2 \cos\theta_2 \end{aligned} \tag{7}$$

In Fig. 2 we plot the marginal posterior distribution for all physical parameters $\phi = \{m_1, m_2, l_1, l_2\}$, together with the initial prior distributions. The ground-truth parameters used for simulation are shown in dark magenta, the MAP estimate is shown in orange. We observe that the pendulum length parameters l_1, l_2 collapse precisely around the true values. On the other hand the posterior distribution over m_1 and m_2 barely contract, showcasing the scale invariance of the Lagrangian. Strikingly, we observe a distinct shift in the mass posteriors away from the prior, which was centered at the true value. Further experiments confirm that this bias stems from neglecting the normalizing constant of the physics likelihood. Therefor it is necessary to estimate Z_ϕ reliably and subsequently correct the log posterior.

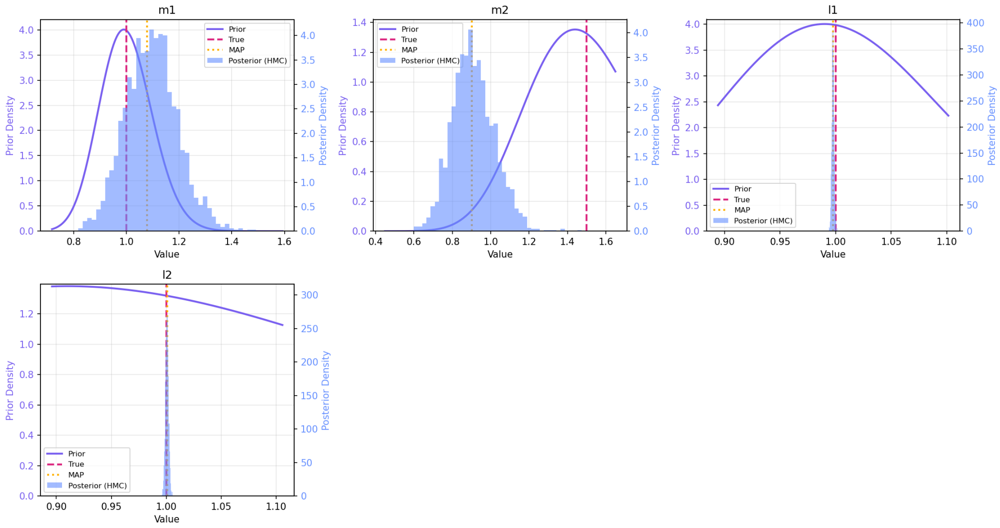

Figure 2: Marginal posteriors of physics parameters for the double pendulum system

4 Conclusion

The joint inference over Lagrangian physics and trajectories allows to infer distributions of plausible solutions in scenarios with sparse noisy data and limited physical knowledge. In such scenarios, finite-difference methods typically fail and deterministic approaches are inadequate as the inverse problem is heavily underdetermined.

Preliminary results on mechanical toy problems look promising, although the found posteriors remain distorted as an unnormalized physics likelihood is used. Therefor, the focus of our ongoing work is on reliably estimating the normalizing constant for correcting the log posterior, to obtain proper and unbiased posterior samples.

References

Brunton, S. L., Proctor, J. L., and Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, March 2016. ISSN 1091-6490. doi: 10.1073/pnas.1517384113. URL http://dx.doi.org/10.1073/pnas.1517384113.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., and Ho, S. Lagrangian Neural Networks, July 2020. URL http://arxiv.org/abs/2003.04630. arXiv:2003.04630 [cs].

Greydanus, S., Dzamba, M., and Yosinski, J. Hamiltonian neural networks. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/26cd8ecadce0d4efd6cc8a8725cbd1f8-Paper.pdf.

Kharazmi, E., Zhang, Z., and Karniadakis, G. E. Variational Physics-Informed Neural Networks For Solving Partial Differential Equations. *arXiv e-prints*, art. arXiv:1912.00873, November 2019. doi: 10.48550/arXiv.1912.00873.

Lutter, M., Ritter, C., and Peters, J. Deep lagrangian networks: Using physics as model prior for deep learning. In *International Conference on Learning Representations*, 2019. URL https://ui.adsabs.harvard.edu/abs/2019arXiv190704490L.

M. Raissi, P. Perdikaris, G. K. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. Machine learning of linear differential equations using gaussian processes. *Journal of Computational Physics*, 348:683–693, 2017. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2017.07.050. URL https://www.sciencedirect.com/science/article/pii/S0021999117305582.