

Learning discrete Lagrangians for variational PDEs from data and detection of travelling waves

Christian Offen  and Sina Ober-Blöbaum

Paderborn University, Warburger Str. 100, 33098 Germany
christian.offen@uni-paderborn.de
<https://www.uni-paderborn.de/en/person/85279>

Abstract. The article shows how to learn models of dynamical systems from data which are governed by an unknown variational PDE. Rather than employing reduction techniques, we learn a discrete field theory governed by a discrete Lagrangian density L_d that is modelled as a neural network. Careful regularisation of the loss function for training L_d is necessary to obtain a field theory that is suitable for numerical computations: we derive a regularisation term which optimises the solvability of the discrete Euler–Lagrange equations. Secondly, we develop a method to find solutions to machine learned discrete field theories which constitute travelling waves of the underlying continuous PDE.

Keywords: System identification · discrete Lagrangians · travelling waves

Publication information For the published version of the article refer to [15].

1 Introduction

In data-driven system identification, a model is fitted to observational data of a dynamical system. The quality of the learned model can greatly be improved when prior geometric knowledge about the dynamical system is taken into account such as conservation laws [16,14,5,9,2], symmetries [10,8,7], equilibrium points [19], or asymptotic behaviour of its motions.

One of the most fundamental principles in physics is the variational principle: it says that motions constitute stationary points of an action functional. The presence of variational structure is related to many qualitative features of the dynamics such as the validity of Noether’s theorem: symmetries of the action functional are in correspondence with conservation laws. To guarantee that these fundamental laws of physics hold true for learned models, Greydanus et al propose to learn the action functional from observational data [5] (Lagrangian neural network) and base prediction on numerical integrations of Euler–Lagrange equations. Quin proposes to learn a discrete action instead [18]. An ansatz of a discrete model has the advantage that it can be trained with position data of motions only. In contrast, learning a continuous theory typically requires information about higher derivatives (corresponding to velocity, acceleration, momenta, for instance) which are typically not observed but only approximated.

Moreover, the discrete action functional (once it is learned) can naturally be used to compute motions numerically.

However, in [14] the authors demonstrate that care needs to be taken when learned action functionals are used to compute motions: even if the data-driven action functional is perfectly consistent with the training data (i.e. the machine learning part of the job is successfully completed), minimal errors in the initialisation process of numerical computations get amplified. As a remedy, the authors develop Lagrangian Shadow Integrators which mitigate these amplified numerical errors based on a technique called backward error analysis. Moreover, using backward error analysis they relate the discrete quantities to their continuous analogues and show how to analyse qualitative aspects of the machine learned model. Action functionals are not uniquely determined by the motions of a dynamical system. Therefore, regularisation is needed to avoid learning degenerate theories. While in [14] the authors develop a regularisation strategy when the action functional is modelled as a Gaussian Process, Lishkova et al develop a corresponding regularisation technique for artificial neural networks in [10] in the context of ordinary differential equations (ODEs).

In this article we show how to learn a discrete action functional from discrete data which governs solutions to partial differential equations (PDEs) using artificial neural networks extending the regularisation strategy which we have developed in [10]. Our technique to learn (discrete) densities of action functionals can be contrasted to approaches where a spatial discretisation of the problem is considered first, followed by structure-preserving model reduction techniques (data-driven or analytical) [3,4] and then a model for the reduced system of ODEs is learned from data [20,12,1].

Travelling waves solutions of PDEs are of special interest due to their simple structure. When a discrete field theory for a continuous process described by a PDE is learned, they typically "get lost" because the mesh of the discrete theory is incompatible with certain wave speeds. In this article, we introduce a technique to find the solutions of data-driven discrete theories that correspond to travelling waves in the underlying continuous dynamics (shadow travelling waves). The article contains the following novelties:

- We transfer our Lagrangian ODE regularising strategy [10] to data-driven discrete field theories in a PDE setting and provide a justification using numerical analysis.
- The development of a technique to detect travelling waves in data-driven discrete field theories.

The article proceeds with a review of variational principles (Section 2), an introduction of our machine learning architecture and derivation of the regularisation strategy (Section 3). In Section 4, we define the notion of *shadow travelling waves* and show how to find them in data-driven models. The article concludes with numerical examples relating to the wave equation (Section 5).

2 Discrete and continuous variational principles

Continuous variational principles Many differential equations describing physical phenomena such as waves, the state of a quantum system, or the evolution of a relativistic fields are derived from a variational principle: solutions are characterised as critical points of a (non-linear) functional S defined on a suitable space of functions $u: X \rightarrow \mathbb{R}^d$ and has the form

$$S(u) = \int_X L(\mathbf{x}, u(\mathbf{x}), u_{x_0}(\mathbf{x}), u_{x_1}(\mathbf{x}), \dots, u_{x_n}(\mathbf{x})) d\mathbf{x}, \quad (1)$$

where $\mathbf{x} = (x_0, x_1, \dots, x_n) \in X$ and u_{x_j} denote partial derivatives of u . This variational principle can be referred to as a *first-order field theory*, since only derivatives to the first order of u appear. In many applications the free variable x_0 corresponds to time and is denoted by $x_0 = t$. The functional S is stationary at u with respect to all variations $\delta u: X \rightarrow \mathbb{R}$ vanishing at the boundary (or with the correct asymptotic behaviour) if and only if the Euler–Lagrange equations

$$0 = \text{EL}(L) = \frac{\partial L}{\partial u} - \sum_{j=0}^n \frac{d}{dx_j} \frac{\partial L}{\partial u_{x_j}} \quad (2)$$

are fulfilled on X .

Example 1. The wave equation

$$u_{tt}(t, x) - u_{xx}(t, x) + \nabla V(u(t, x)) = 0 \quad (3)$$

is the Euler–Lagrange equation $0 = \text{EL}(L)$ to the Lagrangian

$$L(u, u_t, u_x) = \frac{1}{2}(u_t^2 - u_x^2) - V(u). \quad (4)$$

Here ∇V denotes the gradient of a potential V .

Remark 1. Lagrangians are not uniquely determined by the motions of a dynamical system: two first order Lagrangians L and \tilde{L} yield equivalent Euler–Lagrange equations if $sL - \tilde{L}$ ($s \in \mathbb{R} \setminus \{0\}$) is a total divergence $\nabla_{\mathbf{x}} \cdot F(\mathbf{x}, u(\mathbf{x})) = \frac{\partial}{\partial x_1}(F^1(\mathbf{x}, u(\mathbf{x}))) + \dots + \frac{\partial}{\partial x_n} F^n(\mathbf{x}, u(\mathbf{x}))$ for $F = (F^1, \dots, F^n): X \times \mathbb{R}^d \rightarrow \mathbb{R}^n$.

Discrete variational principle For simplicity, we consider the two dimensional compact case: let $X = [0, T] \times [0, l] / \{0, l\}$ with $T, l > 0$. Here $[0, l] / \{0, l\}$ is the real interval $[0, l]$ with identified endpoints (periodic boundary conditions). Consider a uniform, rectangular mesh X_Δ on X with mesh widths $\Delta t = \frac{T}{N}$ and $\Delta x = \frac{l}{M}$ for $N, M \in \mathbb{N}$. A discrete version of the action functional (1) is

$$S_d: (\mathbb{R}^d)^{(N-1) \times M} \rightarrow \mathbb{R}, \quad S_d(U) = \Delta t \Delta x \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} L_d(u_j^i, u_j^{i+1}, u_{j+1}^i)$$

for a discrete Lagrangian density $L_d: (\mathbb{R}^d)^3 \rightarrow \mathbb{R}$ together with temporal boundary conditions for $u_j^0 \in \mathbb{R}^d$ and $u_j^N \in \mathbb{R}^d$ for $j = 0, \dots, M-1$. Above U denotes the values $(u_j^i)_{j=0, \dots, N-1}^{i=1, \dots, M-1}$ on inner mesh points. We have $u_M^i = u_0^i$ by the periodicity in space. Solutions of the variational principle are $U \in (\mathbb{R}^d)^{(N-1) \times M}$ such that U is a critical point of S_d . This is equivalent to the condition that for all $i = 1, \dots, N-1$ and $j = 0, \dots, M-1$ the discrete Euler–Lagrange equations

$$\frac{\partial}{\partial u_j^i} (L_d(u_j^i, u_j^{i+1}, u_{j+1}^i) + L_d(u_j^{i-1}, u_j^i, u_{j+1}^{i-1}) + L_d(u_{j-1}^i, u_{j-1}^{i+1}, u_j^i)) = 0 \quad (5)$$

are fulfilled. The expression on the left of (5) is abbreviated as $\text{DEL}(L_d)_j^i(U)$ in the following.

Remark 2. Instead of periodic boundary conditions in space, S_d can be adapted to other types of boundary conditions such as Dirichlet- or Neumann conditions.

Example 2. The discretised wave equation

$$\frac{(u_j^{i-1} - 2u_j^i + u_j^{i+1})}{\Delta t^2} - \frac{(u_{j-1}^i - 2u_j^i + u_{j+1}^i)}{\Delta x^2} + \nabla V(u_j^i) = 0 \quad (6)$$

is the discrete Euler–Lagrange equations to the discrete Lagrangian

$$L_d(u_j^i, u_j^{i+1}, u_{j+1}^i) = \frac{1}{2} \left(\frac{u_j^{i+1} - u_j^i}{\Delta t^2} \right)^2 - \frac{1}{2} \left(\frac{u_{j+1}^i - u_j^i}{\Delta x^2} \right)^2 - V(u_j^i).$$

Remark 3. In analogy to Remark 1, notice that L_d and \tilde{L}_d yield the same discrete Euler–Lagrange equations (5) if

$$L_d(a, b, c) - s\tilde{L}_d(a, b, c) = \chi_1(a) - \chi_1(b) + \chi_2(a) - \chi_2(c) + \chi_3(b) - \chi_3(c) \quad (7)$$

for differentiable functions $\chi_1, \chi_2, \chi_3: X \rightarrow \mathbb{R}$ and $s \in \mathbb{R} \setminus \{0\}$.

Remark 4. If $\text{DEL}(L_d)_j^i(U) = 0$ (see (5)) and if $\frac{\partial^2 L_d}{\partial u_j^i \partial u_{j+1}^{i+1}}(u_j^i, u_j^{i+1}, u_{j+1}^i)$ is of full rank, then (5) is solvable for u_j^{i+1} as a function of $u_j^i, u_{j+1}^i, u_j^{i-1}, u_{j+1}^{i-1}, u_{j-1}^i, u_{j-1}^{i+1}$ by the implicit function theorem locally around a solution of (5). All of these points correspond to mesh points that either lie to the left or below the point with indices (i, j) . If u_j^1 is known for $0 \leq j \leq M-1$, then utilising the boundary conditions $u_j^0 \in \mathbb{R}^d$ and $u_0^i = u_M^i$ we can compute U by subsequently solving (5). This corresponds to the computation of a time propagation.

The following Proposition analyses the convergence of Newton-Iterations when solving (5) for u_j^{i+1} , as is required to compute time propagations. It introduces a quantity ρ^* that relates to how well the iterations converge. We will make use of this quantity in the design of our machine learning framework.

Proposition 1. Let $u_j^i, u_j^{i+1}, u_{j+1}^i, u_j^{i-1}, u_{j+1}^{i-1}, u_{j-1}^i, u_{j-1}^{i+1}$ such that (5) holds. Let $O \subset \mathbb{R}^d$ be a convex, neighbourhood of $u^* = u_j^{i+1}$, $\|\cdot\|$ a norm of \mathbb{R}^d inducing an operator norm on $\mathbb{R}^{d \times d}$. Define $p(u) := \frac{\partial^2 L_d}{\partial u_j^i \partial u} (u_j^i, u, u_{j+1}^i)$ and let θ and $\bar{\theta}$ be Lipschitz constants on O for p and for $\text{inv} \circ p$, respectively, where inv denotes matrix inversion. Let

$$\rho^* := \|\text{inv}(p(u^*))\| = \left\| \left(\frac{\partial^2 L_d}{\partial u_j^i \partial u^*} (u_j^i, u^*, u_{j+1}^i) \right)^{-1} \right\| \quad (8)$$

and let $f(u^{(n)})$ denote the left hand side of (5) with u_j^{i+1} replaced by $u^{(n)}$. If $\|u^{(0)} - u^*\| \leq \min\left(\frac{\rho^*}{\theta}, \frac{1}{2\theta\rho^*}\right)$ for $u^{(0)} \in \mathcal{O}$, then the Newton Iterations $u^{(n+1)} := u^{(n)} - \text{inv}(p(u^{(n)}))f(u^{(n)})$ converge quadratically against u^* , i.e.

$$\|u^{(n+1)} - u^*\| \leq \rho^* \theta \|u^{(n)} - u^*\|^2. \quad (9)$$

Proof. The statement follows from an adaption of the standard estimates for Newton's method (see [6, §4], for instance) to the considered setting. A detailed proof of the Proposition is contained in the Appendix (Preprint/ArXiv version only).

Remark 5. The assumptions formulated in Proposition 1 are sufficient but not sharp. The main purpose of the proposition is to identify quantities that are related to the efficiency of our numerical solvers and to use this knowledge in the design of machine learning architectures.

3 Machine learning architecture for discrete field theories

We model a discrete Lagrangian L_d as a neural network and fit its parameters

- such that the discrete Euler–Lagrange equations (5) for the learned L_d are consistent with observed solutions $U = (u_j^i)$ of (5)
- and such that (5) is easily solvable for u_j^{i+1} using iterative numerical methods, so that we can use the discrete field theory to predict solutions via forward propagation of initial conditions (see Remark 4).

For given observations $U^{(1)}, \dots, U^{(K)}$ with $U^{(k)} = (u_j^{i(k)})$ on the interior mesh X_Δ , we consider the loss function $\ell = \ell_{\text{DEL}} + \ell_{\text{reg}}$ consisting of a data consistency term ℓ_{DEL} and a regularising term ℓ_{reg} . We have

$$\ell_{\text{DEL}} = \sum_{k=1}^K \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \text{DEL}(L_d)_j^i (U^{(k)})^2 \quad (10)$$

with $\text{DEL}(L_d)_j^i$ from (5). ℓ_{DEL} measures how well L_d fits to the training data.

Since a discrete Lagrangian L_d is not uniquely determined by the system's motions by Remark 3 (indeed, $L_d \equiv \text{const}$ is consistent with any observed dynamics), careful regularisation is required. Indeed, in [14] we demonstrate in an ode setting that if care is not taken, then machine learned models for L_d can be unsuitable for numerical purposes and amplify errors of numerical integration schemes. In view of Proposition 1, we aim to minimize ρ^* (see (8)) and define the regularisation term

$$\ell_{\text{reg}} = \sum_{k=1}^K \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \left\| \left(\frac{\partial^2}{\partial u_j^i \partial u_{j+1}^{i+1}} L_d \left(u_j^{i(k)}, u_{j+1}^{i+1(k)}, u_{j+1}^{i(k)} \right) \right)^{-1} \right\|^2. \quad (11)$$

In our experiments, we use the spectral norm in (11), which is the operator norm induced by the standard Euclidian vector norm on \mathbb{R}^d . Let $A_{\text{reg}}^{i,j,k} := \frac{\partial^2}{\partial u_j^i \partial u_{j+1}^{i+1}} L_d \left(u_j^{i(k)}, u_{j+1}^{i+1(k)}, u_{j+1}^{i(k)} \right)$, i.e. the summands of ℓ_{reg} are $\|(A_{\text{reg}}^{i,j,k})^{-1}\|^2 = \frac{1}{\lambda_{\min}^2}$, where λ_{\min} is the singular value λ_{\min} of $A_{\text{reg}}^{i,j,k}$ with the smallest absolute norm. If $u_j^i \in \mathbb{R}^d$ with $d = 1$, then $\|(A_{\text{reg}}^{i,j,k})^{-1}\|$ can be evaluated without problems. Otherwise, λ_{\min}^2 is computed as the smallest eigenvalue of the symmetric matrix $(A_{\text{reg}}^{i,j,k})^\top A_{\text{reg}}^{i,j,k}$. The eigenvalue can be approximated by inverse matrix vector iterations [6, §5] or computed exactly if the dimension d is small.

4 Periodic travelling waves

For simplicity, we continue within the two-dimensional space time domain $X = [0, T] \times [0, l] / \{0, l\}$ with periodic boundary conditions in space introduced in Section 2. A periodic travelling wave (TW) of a pde on X is a solution of the form $u(t, x) = f(x - ct)$ for $c \in \mathbb{R}$ and with $f: [0, l] / \{0, l\} \rightarrow \mathbb{R}^d$ defined on the periodic spatial domain. Due to their simple structure, TWs are important solutions to pdes. While the defining feature of a TW is its symmetry $u(t + s, x + sc) = u(t, x)$ for $s \in \mathbb{R}$, evaluated on a mesh X_Δ , no such structure is evident unless the quotient $c\Delta t / \Delta x$ is rational and T sufficiently large. However, after a discrete field theory is learned defined by its discrete Lagrangian L_d , it is of interest, whether the underlying continuous PDE has TWs. As in [13] we define *shadow travelling waves* (TWs) of (5) as solutions to the functional equation

$$\begin{aligned} 0 = & \partial_1 L_d(f(\xi), f(\xi - c\Delta t), f(\xi + \Delta x)) \\ & + \partial_2 L_d(f(\xi + c\Delta t), f(\xi), f(\xi + c\Delta t + \Delta x)) \\ & + \partial_3 L_d(f(\xi - \Delta x), f(\xi - c\Delta t - \Delta x), f(\xi)) \end{aligned} \quad (12)$$

where $\partial_j L_d$ denotes the partial derivative of L_d with respect to its j th slot.

Example 3. A Fourier series ansatz for f reveals that the discrete wave equation (6) with potential $V(u) = \frac{1}{2}u^2$ away from resonant cases TWs are $u(t, x) = f(x - c_n t)$ with

$$f(\xi) = \alpha \sin(\kappa_n \xi) + \beta \cos(\kappa_n \xi), \quad \kappa_n = \frac{2\pi n}{l}, \quad n \in \mathbb{Z}, \alpha, \beta \in \mathbb{R} \quad (13)$$

and with wave speed c_n a real solution of

$$\cos(\kappa_n c_n \Delta t) = 1 - \frac{\Delta t^2}{2} + \frac{\Delta t^2}{\Delta x^2} (\cos(\kappa_n \Delta x) - 1). \quad (14)$$

A contour plot for $n = 1$ is shown to the left of Fig. 2.

Remark 6. The TW equation (12) inherits variational structure from the underlying PDE: an application of Palais' principle of criticality [17] of the action of $(\mathbb{R}, +)$ on the Sobolev space $H^1(X_c, \mathbb{R})$ with $X_c = [0, l/c] \times [0, l]$ defined by $(s.u)(t, x) := u(t + s, x + cs)$ to the functional $S(u) = \int_{X_c} L_d(u(t, x), u(t + \Delta t, x), u(t, x + \Delta x)) dt$ reveals that (12) is governed by a formal 1st order variational principle. This is investigated more closely in [13].

To identify TWs in a machine learned model of a discrete field theory, we make an ansatz of a discrete Fourier series $f(\xi) = \sum_{m=-\frac{M}{2}}^{\frac{M}{2}} \hat{f}_{|m|} \exp(m \frac{2\pi i}{T} \xi)$, where bounds of the sum are rounded such that we have M summands. To locate a TW, the loss function $\ell_{\text{TW}}(c, \hat{\mathbf{f}}) + \ell_{\text{TW}}^{\text{reg}}(c, \hat{\mathbf{f}})$ is minimised with

$$\ell_{\text{TW}}(c, \hat{\mathbf{f}}) = \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \|\text{DEL}_j^i(U)\|^2, \quad U = (f(i\Delta t - cj\Delta x))_{\substack{0 \leq j \leq M-1 \\ 0 \leq i \leq N}} \quad (15)$$

and regularisation $\ell_{\text{TW}}^{\text{reg}} = \exp(-100\|U\|_{l^2}^2)$ with discrete l^2 -norm $\|\cdot\|_{l^2}$ to avoid trivial solutions. Here $\hat{\mathbf{f}} = (\hat{f}_m)_m$.

5 Experiment

Creation of training data We use the space-time domain X (Section 2) with $T = 0.5$, $l = 1$, $\Delta x = 0.05$, $\Delta t = 0.025$. To obtain training data that behaves like discretised smooth functions, we compute $K = 80$ solutions to the discrete wave equation (Example 2) with potential $V(u) = \frac{1}{2}u^2$ on the mesh X_Δ from initial data $\mathbf{u}^0 = (u_j^0)_{0 \leq j \leq M-1}$ and $\mathbf{u}^1 = (u_j^1)_{0 \leq j \leq M-1}$. To obtain \mathbf{u}^0 we sample r values from a standard normal distribution. Here r is the dimension of the output of a real discrete Fourier transformation of an M -dimensional vector. These are weighted by the function $m \mapsto M \exp(-2j^4)$, where $m = 0, \dots, r-1$ is the frequency number. The vector \mathbf{u}^0 is then obtained as the inverse real discrete Fourier transform of the weighted frequencies. To obtain \mathbf{u}^1 an initial velocity field $\mathbf{v}^0 = (v_j^0)_{0 \leq j \leq M-1}$ is sampled from a standard normal distribution. Then we proceed as in a variational discretisation scheme [11] applied to the Lagrangian density L of the continuous wave equation (Example 1): to compute conjugate momenta we set $L_\Sigma(\mathbf{u}, \mathbf{v}) = \sum_{j=0}^{M-1} \Delta x L(u_j, v_j)$ and compute $\mathbf{p}^0 = \frac{\partial L_\Sigma}{\partial \mathbf{v}^0}(\mathbf{u}^0, \mathbf{v}^0)$. Then $\mathbf{p}^0 = -L_\Sigma(\mathbf{u}^1, (\mathbf{u}^1 - \mathbf{u}^0)/\Delta t)$ is solved for \mathbf{u}^1 . A plot of an element of the training data set is displayed in Fig. 1. (TWs are *not* part of the training data.)

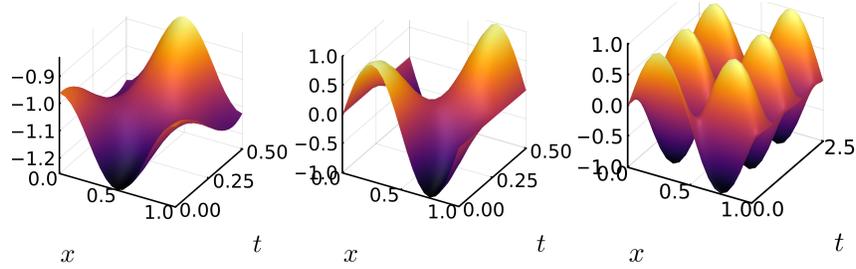


Fig. 1. Left: Element of training data set. Centre: Predicted solution to unseen initial values. Right: Continued solution from centre plot outside training domain

Training and Evaluation A discrete Lagrangian $L_d: \mathbb{R} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is modelled as a three layer feed-forward neural network, where the interior layer has 10 nodes (160 parameters in total). It is trained on the aforementioned training data set and loss function $\ell = \ell_{\text{DEL}} + \ell_{\text{reg}}$ using the optimiser `adam`. We perform 1320 epochs of batch training with batch size 10. For the trained model we have $\ell_{\text{DEL}} \approx 8.6 \cdot 10^{-8}$ and $\ell_{\text{reg}} \approx 1.4 \cdot 10^{-7}$. To evaluate the performance of the trained model for L_d , we compute solutions to initial data by forward propagation (Remark 4) and compare with solutions to the discrete wave equation (Example 2). For initial data $\mathbf{u}^0, \mathbf{u}^1$ not seen during training, the model recovers the exact solution up to an absolute error $\|U - U_{\text{ref}}\|_{\infty} < 0.012$ on X_{Δ} and up to $\|U - U_{\text{ref}}\|_{\infty} < 0.043$ on an extended grid with $T_{\text{ext}} = 2.5$ (Fig. 1).

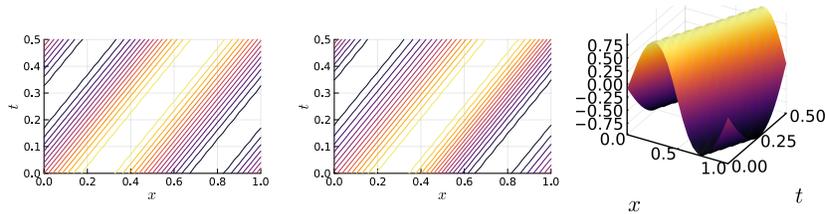


Fig. 2. Left: Reference TW. Centre and Right: Identified TW in learned model

We have $\max_{i,j} \|\text{DEL}(L_d)_{i,j}(U_{\text{ref}}^{\text{TW}})\| < 0.004$, where $U_{\text{ref}}^{\text{TW}}$ is the TW from Example 3 ($n = 1$). This shows that the exact TW is a solution of the learned discrete field theory. This is remarkable since TWs are not part of the training data. However, Remark 6 hints that the ansatz of an autonomous L_d favours TWs as it contains the right symmetries. Using the method of Section 4, a TW U^{TW} and speed c can be found numerically: with $(c_1, U_{\text{ref}}^{\text{TW}})$ as an initial guess with normally distributed random noise ($\sigma = 0.5$) added to the Fourier coefficients of $U_{\text{ref}}^{\text{TW}}$ and to c_1 , we find U^{TW} and c for the learned L_d with errors $\|U^{\text{TW}} - U_{\text{ref}}^{\text{TW}}\|_{\infty} < 0.12$ and $|c - c_1| < 0.001$ (using 10^4 epochs of `adam`). (Fig. 2)

6 Conclusion and future work

We present an approach to learn models of dynamical systems that are governed by an (a priori unknown) variational PDE from data. This is done by learning a neural-network model of a discrete Lagrangian such that the discrete Euler–Lagrange equations (DELs) are fulfilled on the training data. As DELs are local, the model can be efficiently trained and used in simulations. Even though the underlying system is infinite dimensional, model order reduction is not required. It would be interesting to relate the implicit locality assumption of our data-driven model to the more widely used approach to fit a dynamical system on a low-dimensional latent space that is identified using model order reduction techniques [1,3,4,12,20].

Our approach fits in the context of physics-informed machine learning because the data-driven model has (discrete) variational structure by design. However, our model is numerical analysis-informed as well: since our model is discrete by design, it can be used in simulations without an additional discretisation step. Based on an analysis of Newton’s method when used to solve DELs, we develop a regulariser that rewards numerical regularity of the model. The regulariser is employed during the training phase. It plays a crucial role to obtain a non-degenerate discrete Lagrangian.

Our work provides a proof of concept illustrated on the wave equation. It is partly tailored to the hyperbolic character of the underlying PDE. It is future work to adapt this approach to dynamical systems of fundamentally different character (such as parabolic or elliptic behaviour) by employing discrete Lagrangians and regularisers that are adapted to the information flow within such PDEs.

Finally, we clarify the notion of travelling waves (TWs) in discrete models and show how to locate TWs in data-driven models numerically. Indeed, in our numerical experiment the data-driven model contains the correct TWs even though the training data does not contain any TWs. In future work it would be interesting to develop techniques to identify more general highly symmetric solutions in data-driven models and use them to evaluate qualitative aspects of learned models of dynamical systems.

Source Code <https://github.com/Christian-Offen/LagrangianDensityML>

Acknowledgements C. Offen acknowledges the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen and computing time provided by the Paderborn Center for Parallel Computing (PC2).

A Proofs

Proof (Proposition 1). We adapt the standard estimates for Newton's method (see [6, §4], for instance) to the considered setting. Let $f: \mathcal{O} \rightarrow \mathbb{R}^d$ with

$$f(u) = \frac{\partial}{\partial u_j^i} (L_d(u_j^i, u, u_{j+1}^i) + L_d(u_j^{i-1}, u_j^i, u_{j+1}^{i-1}) + L_d(u_{j-1}^i, u_{j-1}^{i+1}, u_j^i)).$$

With this definition, $f(u^*) = 0$, θ is a Lipschitz constant for $Df: \mathcal{O} \rightarrow \mathbb{R}^{d \times d}$, $u \mapsto Df(u)$, $\bar{\theta}$ is a Lipschitz constant for $\text{inv} \circ Df: \mathcal{O} \rightarrow \mathbb{R}^{d \times d}$, $u \mapsto Df(u)^{-1}$, and $\rho^* = \|Df(u^*)^{-1}\|$. Here $Df(u)$ denotes the Jacobian matrix of f at $u \in \mathcal{O}$.

Assume that for $n \in \mathbb{N}$ an iterate $u^{(n)} \in \mathcal{O}$ fulfils $\|u^{(n)} - u^*\| \leq \min\left(\frac{\rho^*}{\theta}, \frac{1}{2\theta\rho^*}\right)$. Then

$$\begin{aligned} \|Df(u^{(n)})^{-1}\| &= \|Df(u^{(n)})^{-1} - Df(u^*)^{-1} + Df(u^*)^{-1}\| \\ &\leq \|Df(u^{(n)})^{-1} - Df(u^*)^{-1}\| + \rho^* \leq \underbrace{\bar{\theta}\|u^{(n)} - u^*\|}_{\leq \rho^*} + \rho^* \leq 2\rho^*. \end{aligned}$$

The next iterate $u^{(n+1)} = u^{(n)} - Df(u^{(n)})^{-1}f(u^{(n)})$ can be bounded:

$$\begin{aligned} \|u^{(n+1)} - u^*\| &= \|u^{(n)} - u^* - Df(u^{(n)})^{-1}(f(u^{(n)}) - \underbrace{f(u^*)}_{=0})\| \\ &\leq \underbrace{\|Df(u^{(n)})^{-1}\|}_{\leq 2\rho^*} \left\| \underbrace{Df(u^{(n)})(u^{(n)} - u^*)}_{= \int_0^1 Df(u^{(n)})(u^{(n)} - u^*) dt} - \underbrace{(f(u^{(n)}) - f(u^*))}_{= \int_0^1 Df(u^{(n)} + t(u^* - u^{(n)}))(u^{(n)} - u^*) dt} \right\| \\ &\leq 2\rho^* \left\| \int_0^1 (Df(u^{(n)}) - Df(u^{(n)} + t(u^* - u^{(n)})))(u^{(n)} - u^*) dt \right\| \\ &\leq 2\rho^* \int_0^1 \underbrace{\|Df(u^{(n)}) - Df(u^{(n)} + t(u^* - u^{(n)}))\|}_{\leq \theta t \|u^{(n)} - u^*\|} \|u^{(n)} - u^*\| dt \\ &\leq 2\rho^* \theta \|u^{(n)} - u^*\|^2 \int_0^1 t dt \\ &= \rho^* \theta \|u^{(n)} - u^*\|^2 \end{aligned}$$

Moreover, since $\|u^{(n)} - u^*\| \leq (2\theta\rho^*)^{-1}$, we have $\|u^{(n+1)} - u^*\| \leq \frac{1}{2}\|u^{(n)} - u^*\|$. By induction, the Newton Iterations converge against u^* and (9) holds true.

References

1. Allen-Blanchette, C., Veer, S., Majumdar, A., Leonard, N.E.: LagNetViP: A Lagrangian neural network for video prediction (AAAI 2020 symposium on physics guided ai) (2020). <https://doi.org/10.48550/ARXIV.2010.12932>

2. Bertalan, T., Dietrich, F., Mezić, I., Kevrekidis, I.G.: On learning Hamiltonian systems from data. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29**(12), 121107 (dec 2019). <https://doi.org/10.1063/1.5128231>
3. Buchfink, P., Glas, S., Haasdonk, B.: Symplectic model reduction of Hamiltonian systems on nonlinear manifolds and approximation with weakly symplectic autoencoder. *SIAM Journal on Scientific Computing* **45**(2), A289–A311 (2023). <https://doi.org/10.1137/21M1466657>
4. Carlberg, K., Tuminaro, R., Boggs, P.: Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics. *SIAM Journal on Scientific Computing* **37**(2), B153–B184 (2015). <https://doi.org/10.1137/140959602>
5. Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., Ho, S.: Lagrangian neural networks (2020). <https://doi.org/10.48550/ARXIV.2003.04630>
6. Deuffhard, P., Hohmann, A.: *Numerical Analysis in Modern Scientific Computing*. Springer New York (2003). <https://doi.org/10.1007/978-0-387-21584-6>
7. Dierkes, E., Flaßkamp, K.: Learning Hamiltonian systems considering system symmetries in neural networks. *IFAC-PapersOnLine* **54**(19), 210–216 (2021). <https://doi.org/10.1016/j.ifacol.2021.11.080>, 7th IFAC Workshop on Lagrangian and Hamiltonian Methods for Nonlinear Control LHMNC 2021
8. Dierkes, E., Offen, C., Ober-Blöbaum, S., Flaßkamp, K.: Hamiltonian neural networks with automatic symmetry detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **33**(6) (06 2023). <https://doi.org/10.1063/5.0142969>, 063115
9. Greydanus, S., Dzamba, M., Yosinski, J.: Hamiltonian Neural Networks. In: Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché Buc, F., Fox, E., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*. vol. 32. Curran Associates, Inc. (2019), <https://proceedings.neurips.cc/paper/2019/file/26cd8ecadce0d4efd6cc8a8725cbd1f8-Paper.pdf>
10. Lishkova, Y., Scherer, P., Ridderbusch, S., Jamnik, M., Liò, P., Ober-Blöbaum, S., Offen, C.: Discrete Lagrangian neural networks with automatic symmetry discovery (accepted). In: *Proceedings of IFAC World Congress 2023*, Yokohama, Japan (9–14/07/2023). *IFAC-PaperOnLine* (2023). <https://doi.org/10.48550/ARXIV.2211.10830>
11. Marsden, J.E., West, M.: Discrete mechanics and variational integrators. *Acta Numerica* **10**, 357–514 (2001). <https://doi.org/10.1017/S096249290100006X>
12. Mason, J., Allen-Blanchette, C., Zolman, N., Davison, E., Leonard, N.: Learning interpretable dynamics from images of a freely rotating 3d rigid body (2022). <https://doi.org/10.48550/ARXIV.2209.11355>
13. McLachlan, R.I., Offen, C.: Backward error analysis for variational discretisations of pdes. *Journal of Geometric Mechanics* **14**(3), 447–471 (2022). <https://doi.org/10.3934/jgm.2022014>
14. Ober-Blöbaum, S., Offen, C.: Variational learning of Euler–Lagrange dynamics from data. *Journal of Computational and Applied Mathematics* **421**, 114780 (2023). <https://doi.org/10.1016/j.cam.2022.114780>
15. Offen, C., Ober-Blöbaum, S.: Learning discrete Lagrangians for variational pdes from data and detection of travelling waves. In: Nielsen, F., Barbaresco, F. (eds.) *Geometric Science of Information. Lecture Notes in Computer Science*, vol. 14071, pp. 569–579. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-38271-0_57
16. Offen, C., Ober-Blöbaum, S.: Symplectic integration of learned Hamiltonian systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **32**(1), 013122 (1 2022). <https://doi.org/10.1063/5.0065913>

17. Palais, R.S.: The principle of symmetric criticality. *Comm. Math. Phys.* **69**(1), 19–30 (1979), <https://projecteuclid.org:443/euclid.cmp/1103905401>
18. Qin, H.: Machine learning and serving of discrete field theories. *Scientific Reports* **10**(1) (11 2020). <https://doi.org/10.1038/s41598-020-76301-0>
19. Ridderbusch, S., Offen, C., Ober-Blöbaum, S., Goulart, P.: Learning ODE models with qualitative structure using Gaussian Processes. In: 2021 60th IEEE Conference on Decision and Control (CDC). IEEE (12 2021). <https://doi.org/10.1109/cdc45484.2021.9683426>
20. Sharma, H., Kramer, B.: Preserving Lagrangian structure in data-driven reduced-order modeling of large-scale dynamical systems (2022). <https://doi.org/10.48550/ARXIV.2203.06361>