


Machine learning of discrete field theories with guaranteed convergence and uncertainty quantification

Christian Offen 

Paderborn University, Department of Mathematics

Warburger Str. 100, 33098 Paderborn, Germany

christian.offen@uni-paderborn.de

July 10, 2024

We introduce a method based on Gaussian process regression to identify discrete variational principles from observed solutions of a field theory. The method is based on the data-based identification of a discrete Lagrangian density. It is a geometric machine learning technique in the sense that the variational structure of the true field theory is reflected in the data-driven model by design. We provide a rigorous convergence statement of the method. The proof circumvents challenges posed by the ambiguity of discrete Lagrangian densities in the inverse problem of variational calculus. Moreover, our method can be used to quantify model uncertainty in the equations of motions and any linear observable of the discrete field theory. This is illustrated on the example of the discrete wave equation and Schrödinger equation. The article constitutes an extension of our previous article for the data-driven identification of (discrete) Lagrangians for variational dynamics from an ode setting to the setting of discrete pdes.

1. Introduction

Partial differential equations that can be derived from variational principles (field theories) take a prominent role in physics, molecular biology, and engineering, for instance, as they describe wave phenomena, evolution of plasma dynamics, electro-magnetism, fluid dynamics, and quantum mechanical processes. Examples include the Korteweg–De Vries (KdV) equation, the shallow wave equation, and the Schrödinger equation.

In this context, data-driven technology has been used to obtain solutions to field theories [13] as well as to discover governing equations from observed solutions [29, 26]. This

article falls into the latter category. We propose to identify an action functional for a field theory by identifying a discrete Lagrangian density based on Gaussian process regression. We do not assume any prior knowledge of the specific form of the Lagrangian. Central novelty of the article is the provision of a rigorous convergence theory for the proposed method. Moreover, Lagrangian descriptions of field theories are highly ambiguous [12, 17, 16]. This has practical implications for our machine learning framework that we discuss systematically and account for in the convergence theory. Furthermore, we provide systematic uncertainty quantification of linear observables of the data-driven system. The article can be seen as an extension of our article [24] from the context of ordinary differential equations to a partial differential equations' context.

Continuous Lagrangian data-driven models A dynamical system is governed by a *variational principle* or a *least action principle*, if motions constitute critical points of an action functional. In case of an autonomous first-order field theory in the space of \mathbb{R}^n , the action functional is of the form

$$S(u) = \int L(u(t), u_{t_1}, \dots, u_{t_n}) dt_1 \dots dt_n, \quad (1)$$

where u is a scalar field and $u_{t_k} = \frac{\partial u}{\partial t_k}$ denote its derivatives ($k = 1, \dots, n$). The function L is referred to as a Lagrangian density of the field theory. In the special case of $n = 1$ the function L is called Lagrangian. A function $u: \mathbb{R}^n \rightarrow \mathbb{R}$ is a solution of the field theory if for any bounded, open domain $\mathcal{M} \subset \mathbb{R}^n$ with Lebesgue measurable boundary $\partial\mathcal{M}$ the action S is stationary at $u|_{\mathcal{M}}$ for all variations $\delta u: \mathcal{M} \rightarrow \mathbb{R}$ that fix $\partial\mathcal{M}$. This is (under regularity assumptions) equivalent to the condition that u fulfils the Euler-Lagrange equations $\text{EL}(u) = 0$, where

$$\text{EL}(u) = \sum_{k=1}^n \frac{\partial}{\partial t_k} \left(\frac{\partial L}{\partial u_{t_k}} \right) - \frac{\partial L}{\partial u} = \sum_{k,l=1}^n \left(\frac{\partial^2 L}{\partial u_{t_k} \partial u_{t_l}} u_{t_k, t_l} \right) - \frac{\partial L}{\partial u}. \quad (2)$$

Here, $u_{t_k, t_l} = \frac{\partial^2 u}{\partial t_k \partial t_l}$. Details may be found in [11, 32], for instance.

In the data-driven context, L is sought as a function of u and its gradient such that (2) is fulfilled at observed data points

$$\mathcal{D} = \left\{ (u(t^{(j)}), (u_{t_k}(t^{(j)}))_{k=1}^n, (u_{t_k, t_l}(t^{(j)}))_{k,l=1}^n) \right\}_{j=1}^M$$

where $t^{(j)} \in \mathbb{R}^n$ for $1 \leq j \leq M$. Once L is known, $\text{EL}(u) = 0$ can be solved with a numerical method such as a variational integrator [20].

In the special case of Lagrangian odes ($n = 1$), several methods have been developed for the data-driven identification of a Lagrangian [9, 23, 10, 14, 24].

Discrete Lagrangian data-driven models Instead of learning continuous variational principles, in [29] Qin proposes to learn discrete field theories by approximating discrete Lagrangian densities. Here training data consists of solutions of the field theory

observed over discrete spacetime locations, i.e. $\mathcal{D} = \{u(t^{(j)})\}_j$, where $t^{(j)}$ takes values in a temporal-spatial mesh. This needs to be contrasted with the identification of a continuous model for a Lagrangian density which requires observations of derivatives of first and second order of solutions.

In case of a temporal-spatial mesh on \mathbb{R}^2 ($n = 2$) with coordinates (t, x) , functions over the mesh (discrete fields) can be represented by the collection of their values u_j^i at mesh points with indices (i, j) . An example of a discrete Lagrangian density is the three-point Lagrangian $L_d(u_j^i, u_j^{i+1}, u_{j+1}^i)$. A corresponding discrete action is defined as the (formal) sum

$$S_\Delta(\{u_j^i\}) = \sum_{i,j} L_d(u_j^i, u_j^{i+1}, u_{j+1}^i) \Delta t \Delta x, \quad (3)$$

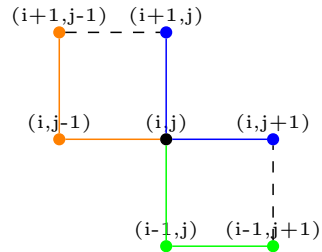


Figure 1: Stencil of (4)

where $(\Delta t, \Delta x)$ are the discretisation parameters of the mesh. The application of a discrete variational principle yields the discrete Euler–Lagrange equation

$$\frac{\partial}{\partial u_j^i} \left(L_d(u_j^i, u_j^{i+1}, u_{j+1}^i) + L_d(u_j^{i-1}, u_j^i, u_{j+1}^{i-1}) + L_d(u_{j-1}^i, u_{j-1}^{i+1}, u_j^i) \right) = 0 \quad (4)$$

which corresponds to $\frac{\partial}{\partial u_j^i} S_\Delta(\{u_j^i\}) = 0$ for all interior meshpoints u_j^i . The discrete Euler–Lagrange equation constitutes a seven-point stencil which relates values of a discrete field over seven mesh-points as visualised in Figure 1. Another example is the four-point Lagrangian density $L_d(u_j^i, u_j^{i+1}, u_{j+1}^i, u_{j+1}^{i+1})$. It yields the discrete Euler–Lagrange equation

$$\begin{aligned} \frac{\partial}{\partial u_j^i} \left(L_d(u_j^i, u_j^{i+1}, u_{j+1}^i, u_{j+1}^{i+1}) + L_d(u_j^{i-1}, u_j^i, u_{j+1}^{i-1}, u_{j+1}^i) \right. \\ \left. + L_d(u_{j-1}^i, u_{j-1}^{i+1}, u_j^i, u_j^{i+1}) + L_d(u_{j-1}^{i-1}, u_{j-1}^i, u_j^{i-1}, u_j^i) \right) = 0. \end{aligned} \quad (5)$$

This constitutes a nine-point stencil (Figure 2). We refer to [19] for details on discrete variational calculus. Three-point and four-point Lagrangian densities for discrete field theories over a temporal-spatial mesh will be the main examples in this article.

Ambiguity of Lagrangian densities and data-driven identification

The data-driven identification of a continuous or a discrete Lagrangian density is an ill-defined inverse problem as many different Lagrangian densities can yield equations of motions with the same set of solutions [12, 17, 16]. In the context of neural network based learning methods, the author develops regularisation strategies that optimise numerical conditioning of the learnt theory in [26, 25]. The present article will provide normalisation conditions suitable for the

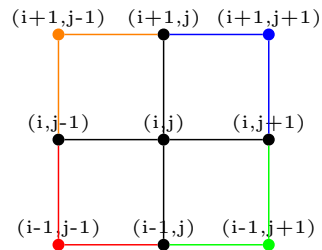


Figure 2: Stencil of (5)

context of learning Lagrangian densities with Gaussian processes.

The proposed method can be contrasted to operator learning (e.g. [15]), which aims to learn an operator that can reconstruct solutions of pdes from boundary data. Other approaches aim to identify an explicit, analytic form of the governing pde via sparse regression based on a dictionary of possible expressions [35, 33]. Furthermore, the method can be contrasted to methods based on model order reduction that seek a latent space on which the dynamics can be described by an ode that can be identified from data or derived from additional knowledge about a full order model. Structure-preserving approaches include [39, 37, 38, 1, 21].

Our method exploits the temporal-spatial locality of partial differential equations by learning a comparatively low dimensional object, a discrete Lagrangian density, that can recover a discrete pde. The discrete pde can subsequently be solved for unseen boundary conditions by propagating data via the stencil defined by the discrete Euler–Lagrange equation. Moreover, our method preserves the temporal-spatial variational structure of the problem, which is related to multisymplectic conservation laws. In combination with variational symmetries variational structure is related to conservation laws via Noether’s theorem [18]. Furthermore, preservation of local variational structure in machine learning approaches can be beneficial for the detection of travelling wave solutions [26, 25].

Novelty of article The article constitutes an extension of [24] from ordinary differential equations to partial differential equations.

1. We introduce a method based on Gaussian process regression to learn discrete Lagrangian densities from temporal-spatial data.
2. We formulate and prove a convergence theorem as the distance of data points tends to zero.
3. The article systematically discusses the ambiguity of Lagrangian densities and normalisation strategies for kernel-based machine learning methods.
4. Moreover, a statistical framework for a quantification of model uncertainty of any linear observable of the learnt field theory is provided.

Our method may be contrasted to the aforementioned approaches in the literature for learning Lagrangian densities, which mainly focus on ordinary differential equations (potentially in combination with model order reduction) or on less geometric approaches such as operator learning. Moreover, ambiguity of sought Lagrangians is often circumvented in the literature by making a less general ansatz for the form of the Lagrangian (e.g. [2]). Additionally, convergence guarantees are typically not provided and model uncertainty quantification of linear observables is not discussed systematically, especially in the presence of model ambiguity (e.g. [9]).

Interpreting a discrete Euler–Lagrange equation, such as (4) or (5), as a partial differential equation for L_d , the methodology of our method can be viewed in the context of meshless collocation methods [34] for linear partial differential equations. However, in contrast to approaches such as [6], solutions for L_d of discrete Euler–Lagrange equations are highly ambiguous, which is a major technical difficulty that the article overcomes. Indeed, for this, we employ techniques presented by Owhadi and Scovel in [28] to relate posterior means of Gaussian processes to constraint optimisation problems in reproducing kernel Hilbert spaces.

Outline In Section 2 the article continues the review of variational principles for discrete field theories and discusses the ambiguity of discrete Lagrangian densities for the description of field theories. The analogous discussion for continuous theories is provided in Appendix A, for comparison. Section 2, furthermore, provides some observations that justify that our normalisation strategy does not restrict the generality of the ansatz. After a brief review of the notion of reproducing kernel Hilbert spaces and Gaussian processes, Section 3 introduces our method to identify Lagrangian densities from discrete temporal-spatial data with the possibility to quantify model uncertainty of linear observables. Section 4 contains numerical experiments including the identification of a discrete Lagrangian density for the discrete wave equation and the Schrödinger equation. Section 5 contains a statement of convergence guarantees for the method. The article concludes with a summary and concluding remarks in Section 6.

As additional material, Appendix B relates the convergence statement of Section 5 to the convergence analysis for temporal discrete Lagrangians presented in [24].

2. Lagrangian dynamics

In the following we continue our review of discrete Lagrangian dynamics for field theories. Here we focus our discussion on the three-point Lagrangian and four-point Lagrangian on a temporal-spatial mesh as introduced in Section 1. The discussion can be expanded to other discrete Lagrangian densities that are obtained by applying quadrature formulas to exact discrete Lagrangian densities as defined in Section II.B of [26]. For details on discrete variational calculus we refer to [19]. An analogous discussion for continuous theories can be found in Appendix A for comparison.

2.1. Discrete Euler–Lagrange equation

As motivated in Section 1 (see (4) and (5)), we define the discrete Euler–Lagrange operator for three-point Lagrangians by

$$\text{DEL}(L_d)(\mathbf{u}) = \nabla_1 L_d(u, u^+, u_+) + \nabla_2 L_d(u^-, u, u_+) + \nabla_3 L_d(u_-, u_+, u) \quad (6)$$

and for four-point Lagrangians by

$$\begin{aligned} \text{DEL}(L_d)(\mathbf{u}) = & \nabla_1 L_d(u, u^+, u_+, u_+) + \nabla_2 L_d(u^-, u, u_+, u_+) \\ & + \nabla_3 L_d(u_-, u_+, u, u^+) + \nabla_4 L_d(u_-, u_-, u^-, u). \end{aligned} \quad (7)$$

Here $\mathbf{u} = (u, u^+, u_+, u^-, u_-, u^+) \in (\mathbb{R}^d)^7$ for three-point Lagrangians and $\mathbf{u} = (u, u^+, u_+, u_+^+, u^-, u_-, u_-^+, u_-^-) \in (\mathbb{R}^d)^9$ for four-point Lagrangians. The expression $\nabla_j L_d$ denotes the gradient of L_d with respect to the j th input argument. The discrete Euler–Lagrange equation is

$$\text{DEL}(L_d)(\mathbf{u}) = 0 \quad (8)$$

for all stencils \mathbf{u} that are contained in the mesh.

2.2. Ambiguity of discrete Lagrangian densities

The remainder of Section 2 prepares a justification that our normalisation strategy in the data-driven framework of Section 3 does not restrict the generality of the ansatz. Readers mostly interested in the data-driven aspects may skip directly to Section 3.

Discrete Lagrangian densities can be ambiguous in two different ways:

1. Discrete Lagrangian densities L_d and \tilde{L}_d yield the same discrete Euler–Lagrange operator up to rescaling, i.e.

$$\text{DEL}(L_d) = \rho \cdot \text{DEL}(\tilde{L}_d), \quad \rho \in \mathbb{R} \setminus \{0\} \quad (9)$$

if

$$L_d - \rho \tilde{L}_d - c = \text{div}_t F,$$

for $c \in \mathbb{R}$ and a discrete divergence $\text{div}_t F$. For three-point Lagrangians $L_d(a, b, c)$ discrete divergence takes the form

$$\text{div}_t F(a, b, c) = F_1(a) - F_1(b) + F_2(a) - F_2(c) + F_3(b) - F_3(c)$$

with continuously differentiable function $F = (F_1, F_2, F_3)$. These terms are telescopic in the action S_d (3). For four-point Lagrangians $L_d(a, b, c, d)$ a discrete divergence takes the form

$$\begin{aligned} \text{div}_t F(a, b, c, d) = & F_1(a) - F_1(b) + F_2(a) - F_2(c) + F_3(a) - F_3(d) \\ & + F_4(b) - F_4(c) + F_5(b) - F_5(d) + F_6(c) - F_6(d) \end{aligned}$$

with continuously continuously differentiable function $F = (F_1, F_2, F_3, F_4, F_5, F_6)$. Two Lagrangians L and \tilde{L} with (9) are called (*gauge-*) *equivalent*.

2. Even when two discrete Lagrangian densities L_d and \tilde{L}_d are not equivalent, they can yield the same set of solutions, i.e.

$$\text{DEL}(L_d)(\mathbf{u}) = 0 \iff \text{DEL}(\tilde{L}_d)(\mathbf{u}) = 0,$$

where \mathbf{u} denotes stencil data. In that case \tilde{L}_d is called an *alternative Lagrangian density* to L_d .

2.3. Normalisation of discrete Lagrangian densities

In the data-driven framework for learning discrete Lagrangian densities that we will introduce in a subsequent section, we will employ strategies to prevent degenerate solutions of discrete Lagrangians that fail to discriminate non-motions. An extreme example are Null-Lagrangians, for which $\text{DEL}(L_d) \equiv 0$.

To a four-point discrete Lagrangian density L_d define $\text{Mm}^+(L_d)$ by

$$\begin{aligned}\text{Mm}^+(L_d)(\mathbf{u}) &= \nabla_2 L_d(u^-, u, u_+, u_+) + \nabla_4 L_d(u_-, u_-, u^-, u) \\ \text{Mm}^-(L_d)(\mathbf{u}) &= -\nabla_1 L_d(u, u^+, u_+, u_+) - \nabla_3 L_d(u_-, u_-, u, u^+).\end{aligned}\tag{10}$$

for elements of nine-point stencils $\mathbf{u} = (u, u^+, u_+, u^-, u_-, u_-, u_-, u_+^+) \in (\mathbb{R}^d)^9$. Considering discrete three-point Lagrangian densities as four-point Lagrangian densities with trivial dependence on the fourth component, the definitions extend to three-point Lagrangian densities.

Remark 1 (Interpretation) The expression $\text{Mm}^+(L_d)$ and $\text{Mm}^-(L_d)$ can be viewed as a component of a conjugate momentum: to a discrete four-point Lagrangian density L_d consider the Lagrangian

$$L_{d,\Delta x}(U^i, U^{i+1}) = \sum_j L_d(u_j^i, u_j^{i+1}, u_{j+1}^i, u_{j+1}^{i+1}),\tag{11}$$

where $U^i = (u_j^i)_j$. The discrete Lagrangian describes the temporal motion of the system via the discrete Euler–Lagrangian equation for ordinary differential equations

$$0 = \nabla_{U^i} (L_{d,\Delta x}(U^{i-1}, U^i) + L_{d,\Delta x}(U^i, U^{i+1})).$$

Associated discrete conjugate momenta are defined by $P_+^i = \nabla_{U^i} L_{d,\Delta x}(U^{i-1}, U^i)$ and $P_-^i = -\nabla_{U^i} L_{d,\Delta x}(U^i, U^{i+1})$. Indeed, the j th component is given as

$$\begin{aligned}(P_+^i)_j &= \nabla_2 L_d(u_j^{i-1}, u_j^i, u_{j+1}^{i-1}, u_{j+1}^i) + \nabla_4 L_d(u_{j-1}^{i-1}, u_{j-1}^i, u_j^{i-1}, u_j^i) \\ (P_-^i)_j &= -\nabla_1 L_d(u_j^i, u_j^{i+1}, u_{j+1}^i, u_{j+1}^{i+1}) - \nabla_3 L_d(u_{j-1}^i, u_{j-1}^{i+1}, u_j^i, u_j^{i+1}).\end{aligned}\quad \square$$

Proposition 1 *Let $\mathbf{u}_b = (u, u^+, u_+, u^-, u_-, u_-, u_-, u_+^+) \in (\mathbb{R}^d)^9$ be a stencil and \mathring{L}_d be a discrete four-point Lagrangian. To any $c_b \in \mathbb{R}$, $p_b \in \mathbb{R}^d$ there exists an equivalent discrete Lagrangian L_d with*

$$L_d(u, u^+, u_+, u_+^+) = c_b, \quad \text{Mm}^-(L_d)(\mathbf{u}_b) = p_b.\quad \square$$

PROOF Let $\mathring{c}_b = \mathring{L}_d(u, u^+, u_+, u_+^+)$, $\mathring{p}_b = \text{Mm}^-(\mathring{L}_d)(\mathbf{u}_b)$, $F = (F_1, 0, 0, 0, 0, 0)$ with $F_1(u) = (\mathring{p}_b - p_b)^\top u$, and $c = c_b - \mathring{c}_b + (\mathring{p}_b - p_b)(u^+ - u)$. The Lagrangian

$$L_d = \mathring{L}_d + \text{div}_t F + c$$

is equivalent to \mathring{L}_d and fulfils the conditions (1). ■

Remark 2 An analogous statement holds with Mm^+ replacing Mm^- in Proposition 1. □

3. Data-driven framework

Consider an open and bounded domain $\Omega \subset (\mathbb{R}^d)^q$ ($q \in \mathbb{N}$) on which an (a priori unknown) discrete Lagrangian density $L_d^{\text{ref}}: \Omega \rightarrow \mathbb{R}$ defines true motions via $\text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 0$. Based on the observation of stencil data \mathbf{u} for which $\text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 0$, our goal is to approximate a Lagrangian density $L_d: \Omega \rightarrow \mathbb{R}$ such that $\text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 0 \iff \text{DEL}(L_d)(\mathbf{u}) = 0$ for all stencils $\mathbf{u} \in \Omega$.

For this, we will model L_d as a Gaussian Process over Ω and compute the posterior process of L_d conditioned on $\text{DEL}(L_d)(\mathbf{u}) = 0$ for all observed stencil data \mathbf{u} . To prevent learning a degenerate L_d (such as Null-Lagrangians) we will, furthermore, condition on additional normalisation conditions as motivated by Proposition 1.

For this, we briefly introduce our set-up making use of a reproducing kernel Hilbert-space (RKHS) setting. We will then proceed to the computation of the posterior process. We refer the reader to [8, 27] for an introduction and further information.

3.1. Reproducing kernel Hilbert spaces – Set-up

Consider a continuously differentiable function $K: \Omega \times \Omega \rightarrow \mathbb{R}$. Assume that K is a positive-definite kernel, i.e. $K(x, y) = K(y, x)$ for $x, y \in \Omega$ (symmetry) and the matrix $(K(x^{(i)}, x^{(j)}))_{i,j=1}^M$ is positive definite for all finite subsets $\{x^{(i)}\}_{i=1}^M \subset \Omega$.

To define the reproducing kernel Hilbert space U to K , consider the set $\{K(x^{(j)}, \cdot) \mid x^{(j)} \in \Omega\}$ and its linear span \mathring{U} . An inner product on \mathring{U} is provided by the linear extension of $\langle \cdot, \cdot \rangle$ defined by

$$\langle K(x, \cdot), K(y, \cdot) \rangle = K(x, y).$$

The reproducing kernel Hilbert space U is now obtained as the completion of \mathring{U} with respect to $\langle \cdot, \cdot \rangle$.

An identification of the dual space U^* and U is given by the bijective, linear, symmetric map $\mathcal{K}: U^* \rightarrow U$ defined by $\mathcal{K}(\Phi)(x) = \Phi(K(x, \cdot))$.

Let $\xi \in \mathcal{N}(0, \mathcal{K})$ be the *canonical Gaussian process* on U , i.e. $\xi: \mathcal{A} \rightarrow U$ is a random variable (where \mathcal{A} is a probability space) with $\mathbb{E}(\xi) = 0$ (zero mean) and covariance operator \mathcal{K} as introduced above. This means that for $\Phi = (\Phi_1, \dots, \Phi_m) \in (U^*)^m$ ($m \in \mathbb{N}$) the random variable $\Phi(\xi)$ on \mathbb{R}^m is multivariate-normally distributed with covariance matrix given by $\kappa = (\Phi_i(\mathcal{K}(\Phi_j)))_{i,j=1}^m$, i.e. $\Phi(\xi) \in \mathcal{N}(0, \kappa)$.

3.2. Computation of the posterior Gaussian Process

In case we are seeking a three-point discrete Lagrangian density, let us assume we are given stencil data

$$\mathbf{u}^{(k)} = (u^{(k)}, u^{(k)+}, u^{(k)+}, u^{(k)-}, u^{(k)-}, u^{(k)-}, u^{(k)-}u^{(k)+}) \in (\mathbb{R}^d)^7$$

for $k = 1, \dots, M$ such that the sub-triples are elements of Ω , i.e. $(u^{(k)}, u^{(k)+}, u^{(k)+})$, $(u^{(k)-}, u^{(k)-}, u^{(k)-})$, $(u^{(k)-}, u^{(k)+}, u^{(k)+}) \in \Omega$.

In case a four-point discrete Lagrangian density is sought, we consider stencil data of the form

$$\mathbf{u}^{(k)} = (u^{(k)}, u^{(k)+}, u^{(k)+}, u^{(k)+}, u^{(k)-}, u^{(k)-}, u^{(k)-}, u^{(k)-}, u^{(k)+}, u^{(k)-}) \in (\mathbb{R}^d)^9$$

for $k = 1, \dots, M$ such that the sub-quadruples are elements of Ω , i.e. $(u^{(k)}, u^{(k)+}, u^{(k)+}, u^{(k)+}), (u^{(k)-}, u^{(k)-}, u^{(k)-}, u^{(k)-}), (u^{(k)+}, u^{(k)+}, u^{(k)+}, u^{(k)-}), (u^{(k)-}, u^{(k)-}, u^{(k)-}, u^{(k)+}) \in \Omega$.

Moreover, consider an additional stencil $\mathbf{u}_b \in (\mathbb{R}^d)^7$ or $\mathbf{u}_b \in (\mathbb{R}^d)^9$, respectively, for which the sub-triples (sub-quadruples) are elements of Ω and values $p_b \in \mathbb{R}^d$, $c_b \in \mathbb{R}$ to which we define $y_b^M := (0, \dots, 0, p_b, c_b) \in (\mathbb{R}^d)^M \times \mathbb{R}^d \times \mathbb{R}$.

Moreover, let $\Phi_b^M: U \rightarrow (\mathbb{R}^d)^M \times \mathbb{R}^d \times \mathbb{R}$ be defined as

$$\Phi_b^M = (\text{DEL}_{\mathbf{u}^{(1)}}, \dots, \text{DEL}_{\mathbf{u}^{(M)}}, \text{Mm}_{\mathbf{u}_b}^-, \text{ev}_{\mathbf{u}_b}), \quad (12)$$

where for $L_d \in U$

$$\begin{aligned} \text{DEL}_{\mathbf{u}^{(k)}}(L_d) &:= \text{DEL}(L_d)(\mathbf{u}^{(k)}), \quad 1 \leq k \leq M \\ \text{Mm}_{\mathbf{u}_b}^-(L_d) &:= \text{Mm}^-(L_d)(\mathbf{u}_b) \\ \text{ev}_{\mathbf{u}_b}(L_d) &:= L_d(u_b, u_b^+, u_b^+, u_b^+) \text{ or } \text{ev}_{\mathbf{u}_b}(L_d) := L_d(u_b, u_b^+, u_b^+) \end{aligned}$$

with Mm^- as defined in (10).

The conditional process $\xi^M := \xi | \{\Phi_b^M(\xi) = y_b^M\}$ is again a Gaussian process $\xi^M \in \mathcal{N}(L_d^M, \mathcal{K}_{\Phi_b^M})$ since Φ_b^M is linear. Thus, ξ^M is fully defined by the conditional mean $L_d^M \in U$ and conditional covariance operator $\mathcal{K}_{\Phi_b^M}: U^* \rightarrow U$. These can be computed using general theory [27, Cor. 17.12] as follows.

Define the symmetric, matrix $\Theta \in \mathbb{R}^{(M+1)d \times (M+1)d}$ as

$$\Theta = \begin{pmatrix} (\text{DEL}_{\mathbf{u}^{(j)}}^1 \text{DEL}_{\mathbf{u}^{(i)}}^2 K)_{ij} & (\text{DEL}_{\mathbf{u}^{(j)}}^1 \text{Mm}_{\mathbf{u}_b}^{-2} K)_j & (\text{DEL}_{\mathbf{u}^{(j)}}^1 \text{ev}_{\mathbf{u}_b}^2 K)_j \\ (\text{Mm}_{\mathbf{u}_b}^{-1} \text{DEL}_{\mathbf{u}^{(i)}}^2 K)_i & \text{Mm}_{\mathbf{u}_b}^{-1} \text{Mm}_{\mathbf{u}_b}^{-2} K & \text{Mm}_{\mathbf{u}_b}^{-1} \text{ev}_{\mathbf{u}_b}^2 K \\ (\text{ev}_{\mathbf{u}_b}^1 \text{DEL}_{\mathbf{u}^{(i)}}^2 K)_i & \text{ev}_{\mathbf{u}_b}^1 \text{Mm}_{\mathbf{u}_b}^{-2} K & \text{ev}_{\mathbf{u}_b}^1 \text{ev}_{\mathbf{u}_b}^2 K. \end{pmatrix} \quad (13)$$

The upper index 1, 2 of the operators DEL , Mm^- , ev denote on which input element of K they act, i.e.

$$\text{DEL}_{\mathbf{u}^{(j)}}^1 \text{DEL}_{\mathbf{u}^{(i)}}^2 K = \text{DEL}_{\mathbf{u}^{(j)}}(a \mapsto \text{DEL}_{\mathbf{u}^{(i)}}(b \mapsto K(a, b))),$$

for instance. Moreover, we have used the convention that operators are applied component-wise to vectors of functions.

The Gaussian process $\xi^M \in \mathcal{N}(L_d, \mathcal{K}_{\Phi_b^M})$ has conditional mean

$$L_d = y_b^{M\top} \theta^{-1} \mathcal{K}_{\Phi_b^M}. \quad (14)$$

The conditional covariance operator $\mathcal{K}_{\Phi_b^M}: U^* \rightarrow U$ is defined for any $\psi, \phi \in U^*$ by

$$\psi \mathcal{K}_{\Phi_b^M} \phi = \psi \mathcal{K} \phi - (\psi \mathcal{K}_{\Phi_b^M}^{\top}) \theta^{-1} (\Phi_b^M \mathcal{K} \phi). \quad (15)$$

Here

$$\begin{aligned}\psi\mathcal{K}_{\Phi_b^M}\phi &= \psi^1\phi^2K \\ \psi\mathcal{K}_{\Phi_b^M}^\top &= \left(\psi^1\text{DEL}_{u^{(2)}}^2K, \dots, \psi^1\text{DEL}_{u^{(M)}}^2K, \psi^1\text{Mm}_{u_b}^{-2}K, \psi^1K(\cdot, u_b)\right) \\ \Phi_b^M\mathcal{K}\phi &= \left(\text{DEL}_{u^{(2)}}^1\phi^2K \dots \text{DEL}_{u^{(M)}}^1\phi^2K \text{Mm}_{u_b}^{-1}\phi^2K \phi^2K(u_b, \cdot)\right)^\top.\end{aligned}$$

Remark 1 (Normalisation) We would like that the discrete Euler–Lagrange equation (8) to a data-driven discrete Lagrangian density uniquely defines a solution over the temporal-spatial mesh for suitable boundary data. This is referred to as non-degeneracy of the Lagrangian density. (The exact notion depends on the type of boundary value problem.) The normalisation condition $(\text{Mm}_{u_b}^-(\xi), \text{ev}_{u_b}(\xi)) = (p_b, c_b)$ is compatible with the true motions defined by L_d^{ref} as the normalisation condition is fully covered by gauge freedom as proved in Proposition 1. Moreover, as by [28, Thm 12.5], the conditional mean L_d can be characterised as the minimiser of the convex optimisation problem

$$L_d = \underset{\Phi_b^M(\tilde{L}_d) = y_b^M}{\text{argmin}} \|\tilde{L}_d\|_U, \quad (16)$$

where U is the reproducing kernel Hilbert space to the kernel K . Even though the normalisation does not guarantee any non-degeneracy of the conditional mean L_d , we will see in the numerical examples that the condition $p_b \neq 0, c_b \neq 0$ causes well-behaved discrete Lagrangian densities for the prediction of new solutions to the field theory from unseen training data. Intuitively, discrete Lagrangians L_d with artificial degeneracies do not happen to constitute minimisers of (16) in practise unless the degeneracies are forced by the true underlying field theory. \square

Remark 2 (Computational aspects) The evaluation of the posterior mean L_d or covariance operator $\mathcal{K}_{\Phi_b^M}$ involves solving large linear systems of equations, see (14),(15). While we will employ standard libraries in the numerical experiments, we refer to the literature on computational methods for Gaussian process regression for various approaches such as [30, 36, 7]. \square

4. Numerical experiments

In this section, we apply the method of Section 3 to observational data of the discrete wave equation and a discrete Schrödinger equation. Similar experiments have been performed in [26] for neural network based methods.

4.1. Discrete wave equation

4.1.1. True model and training data generation

On the temporal-spatial domain $\Omega = [0, 0.5] \times [0, 1]$ with coordinates (t, x) and periodic boundary conditions in space, consider a uniform mesh with discretisation parameters

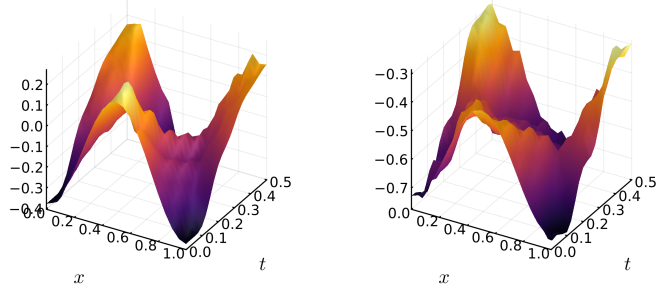


Figure 3: Complete training data set for wave equation experiment. (Only these two samples.)

$\Delta t = 1/40$, $\Delta x = 1/20$ and the following discrete Lagrangian

$$L_d^{\text{ref}}(u, u^+, u_+) = \frac{1}{2} \left(\frac{u^+ - u}{\Delta t} \right)^2 - \frac{1}{2} \left(\frac{u_+ - u}{\Delta x} \right)^2 - V(u), \quad \text{with } V(u) = \frac{1}{2}u^2.$$

The discrete Euler–Lagrange equations

$$0 = \text{DEL}(L_d^{\text{ref}})(u) = \frac{u^- - 2u + u^+}{\Delta t^2} - \frac{u_- - 2u + u_+}{\Delta x^2} + \nabla V(u) \quad (17)$$

is the discrete wave equation and can be viewed as a discrete version of the continuous wave equation $\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} + \nabla V(u) = 0$. To obtain training data, we sample $N_0 = 2$ initial conditions consisting of discrete spatial data $U^0 = (u_j^0)_j$ and discrete conjugate momentum data $P_-^0 = (p_j^0)_j$ at $t = 0$. The initial data are then propagated in time by forming the temporal Lagrangian density $L_{d,\Delta x}$ defined in (11). Then standard methods in variational integration [20] are used: in the first step, $P_-^0 = -\nabla_{U^0} L_{d,\Delta x}(U^0, U^1)$ is solved for U^1 . Then U^{i+1} is computed from U^i and U^{i-1} by solving the discrete Euler–Lagrange equation $\text{DEL}(L_{d,\Delta x}^{\text{ref}})(U^{i-1}, U^i, U^{i+1}) = 0$ numerically for U^{i+1} ($i \geq 1$). The discrete Euler–Lagrange operator DEL for temporal Lagrangians is given as

$$\text{DEL}(L_{d,\Delta x})(U^{i-1}, U^i, U^{i+1}) = \frac{\partial}{\partial U^i} (L_{d,\Delta x}(U^{i-1}, U^i) + L_{d,\Delta x}(U^i, U^{i+1})). \quad (18)$$

The two samples are plotted in Figure 3. These two solutions contain $M = 760$ seven-point stencils $\mathfrak{U} = (\mathbf{u}^{(k)})_{k=1}^M$ (N_0 times number of interior mesh points).

4.1.2. Predictions from unseen initial data (Extrapolation)

With the base stencil $\mathbf{u}_b = \mathbf{0} \in \mathbb{R}^7$, normalisation constants $c_b = 1$, $p_b = 1$, and kernel $K(x, y) = \exp(-1/2\|x - y\|^2)$, the stencil data \mathfrak{U} defines a posterior Gaussian process $\xi^M \in \mathcal{N}(L_d^M, \mathcal{K}_{\Phi_b^M})$ (Section 3.2).

Figure 4 shows that the mean L_d^M of the posterior process can successfully be used to predict solutions of the discrete wave equation from the unseen initial data $u(0, x) =$

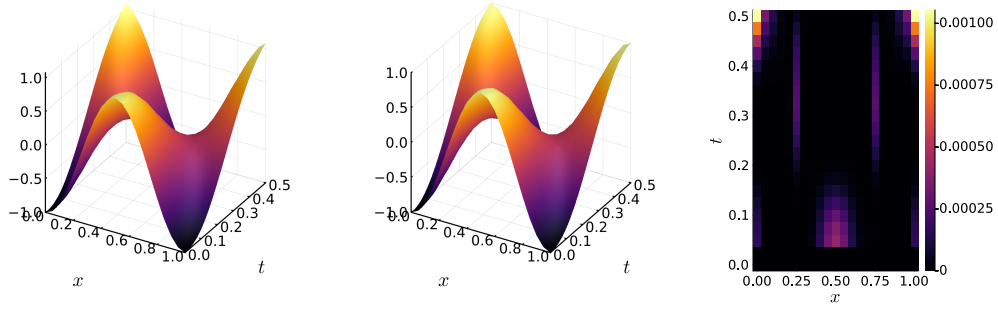


Figure 4: Wave equation experiment. Predicted solution from initial data at $t = 0$ with L_d^M (left) matches reference (centre) to high accuracy. Right: Standard deviation of the Gaussian variable $\text{DEL}(\xi^M)$ along predicted solution.

$-\cos(2\pi x)$, $u(\Delta t, x) = -\cos(2\pi x)$ via forward propagation with $\text{DEL}(L_{d,\Delta x}^M) = 0$. The discrete l_2 -norm of the error is smaller than 0.0222.

To quantify model uncertainty along the predicted solution, we consider at each interior mesh point $(i\Delta t, j\Delta x)$ the stencil $\mathbf{u}^{(i,j)}$ centred at $(i\Delta t, j\Delta x)$. The standard deviation σ of the Gaussian random variable $\text{DEL}_{\mathbf{u}^{(i,j)}}(\xi^M)$ can be computed using the conditional covariance operator $\mathcal{K}_{\Phi_b^M}$ as follows

$$\sigma(\text{DEL}_{\mathbf{u}^{(i,j)}}(\xi^M)) = \sqrt{\psi \mathcal{K}_{\Phi_b^M} \psi}, \text{ with } \psi = \phi = \text{DEL}_{\mathbf{u}^{(i,j)}}. \quad (19)$$

Here $\psi \mathcal{K}_{\Phi_b^M} \psi$ is computed as explained in (15). The plot to the right of Figure 4 displays the values of $\sigma(\text{DEL}_{\mathbf{u}^{(i,j)}}(\xi^M))$ at all interior mesh points (i, j) . This shows areas of very low and slightly increased model uncertainty and can be interpreted as an estimation of the local error.

4.1.3. Prediction of travelling waves (Extrapolation)

The functional equation

$$\frac{u(t - \Delta t, x) - 2u(t, x) + u(t + \Delta t, x)}{\Delta t^2} - \frac{u(t, x - \Delta x) - 2u(t, x) + u(t, x + \Delta x)}{\Delta x^2} + u(t, x) = 0$$

to a continuous interpolation of (17) admits travelling wave solutions $u_{\text{TW}}(t, x) = f_{\text{TW}}(s) = a_1 \sin(\kappa s) + a_2 \cos(\kappa s)$, $s = x - ct$, $\kappa = 2\pi k$ ($k \in \mathbb{Z}$, $a_1, a_2 \in \mathbb{R}$), where the wave speed c is a real solution to

$$\cos(\kappa c \Delta t) = 1 - \frac{\Delta t^2}{2} + \frac{\Delta t^2}{\Delta x^2} (\cos(k \Delta x) - 1)$$

(see [26, Example 11]). Figure 5 shows a repetition of the experiment of Figure 4 for the initial data of a travelling wave $u_{\text{TW}}(0, x)$, $u_{\text{TW}}(\Delta t, x)$. An upward trend of the ridge

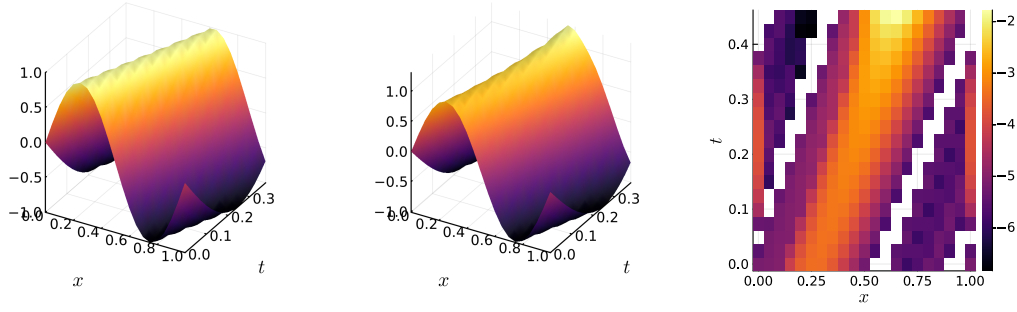


Figure 5: Wave equation experiment. Predicted solution from travelling wave initial data at $t = 0$ with L_d^M (left) matches reference (centre) to high accuracy but with an upward trend of the ridge. Right: Standard deviation of the Gaussian variable $\text{DEL}(\xi^M)$ along predicted solution. For better visualisation, \log_{10} is applied before plotting a heat map.

can be observed. The discrete l_2 -norm of the error is smaller than 0.067 (time interval $[0, 0.375]$). In the heatmap of the standard deviation (right plot), for better visualisation, the logarithm (basis 10) is applied to the data before plotting. The upward trend of the ridge is reflected in a higher model uncertainty at later times.

The experiments demonstrate that the learned model can extrapolate to unseen initial data and that enough information is contained in the stencil data of the two observed fields (Figure 3) to predict travelling waves to high accuracy. This is remarkable as the training data-set does not contain travelling waves. Indeed, the global behaviour of the two fields in the training data set (3) is quite different from the global behaviour of travelling waves. However, as the considered technique for identification of the dynamical system and the method to predict solutions relies on local dependencies, extrapolation to solutions with different global behaviour is possible.

4.2. Discrete Schrödinger equation

4.2.1. True model and training data generation

Let $\Omega = [0, 0.14] \times [0, 1]$ be a temporal-spatial domain with coordinates (t, x) and periodic boundary conditions in x . On the domain consider a uniform mesh that is periodic in space with mesh parameters $(\Delta t, \Delta x)$. To a smooth potential $V: \mathbb{R} \rightarrow \mathbb{R}$ consider the discrete Lagrangian

$$L_d^{\text{ref}}(u, u^+, u_+, u_+^+) = u_{\text{cntr}}^\top J u_{\Delta t} - \|u_{\Delta x}\|^2 - V(\|u_{\text{cntr}}\|^2)$$

with

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad u_{\Delta t} = \frac{1}{2\Delta t}((u_+^+ - u_+) + (u^+ - u)),$$

$$u_{\text{cntr}} = \frac{1}{4}(u + u^+ + u_+ + u_+^+), \quad u_{\Delta x} = \frac{1}{2\Delta x}((u_+^+ - u^+) + (u_+ - u)).$$

The discrete Lagrangian yields the following discrete Euler–Lagrange equation

$$\begin{aligned} 0 &= \text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 2Ju_{\Delta t, \text{cntr}} + 2u_{\Delta x^2} - 2(V'(\|u_{\text{cntr}}\|^2)u_{\text{cntr}})_{\text{cntr}} \\ \iff Ju_{\Delta t, \text{cntr}} &= u_{\Delta x^2} - (V'(\|u_{\text{cntr}}\|^2)u_{\text{cntr}})_{\text{cntr}} \end{aligned} \quad (20)$$

where we have used the notation

$$\begin{aligned} u_{\Delta t, \text{cntr}} &= \frac{1}{8\Delta t}((u_+^+ - u_-^+) + 2(u_+ - u_-) + (u_+^- - u_-^-)) \\ u_{\Delta x^2} &= \frac{1}{4\Delta x^2}((u_+^+ - 2u^+ + u_-^+) + 2(u_+ - 2u + u_-) + (u_+^- - 2u^- + u_-^-)) \\ (g(u, u^+, -u_+, u_+^+))_{\text{cntr}} &= \frac{1}{4}(\text{evaluate} + \text{shift}^- + \text{shift}_- + \text{shift}_-) (g, (u, u^+, u_+, u_+^+)). \end{aligned}$$

Above, the operator shift evaluates g on shifted input arguments. The shift occurs in the temporal or spatial direction as indicated by its indices. The discrete Euler–Lagrange equation (20) constitutes a discrete version of the continuous Schrödinger equation

$$\hbar i \frac{\partial}{\partial t} \Psi = \left(-\frac{\partial^2}{\partial x^2} + V'(|\Psi|^2) \right) \Psi, \quad (21)$$

which in real coordinates $u = (\text{Re}(\Psi), \text{Im}(\Psi))^\top$ reads

$$\hbar J \frac{\partial}{\partial t} u = \left(-\frac{\partial^2}{\partial x^2} + V'(\|u\|^2) \right) u. \quad (22)$$

(The Plank constant is set to $\hbar = 1$.) Equation (22) may be interpreted as a continuous analogue to (20). For the purpose of the numerical experiments, the discrete equation (20) is regarded as the true field theory. We may refer to the first component of u as real part, and to the second component as imaginary part.

In the experiments, $V(r) = r$, $\Delta t = 7/400$, $\Delta x = 1/10$ (9×10 mesh points). To obtain training data, we sample $N_0 = 30$ initial spatial data at time $t = 0$ and integrate forward in time using variational integration techniques as explained in Section 4.1. However, the momentum P^0 is not sampled but computed from U^0 exploiting the special structure of L_d^{ref} (refer to [26, II B.3] for details).

The first and third plot of Figure 6 display real and imaginary part of an instance of a training sample. These solutions constitute training data $\mathfrak{U} = (\mathbf{u}^{(k)})_{k=1}^M$ consisting of $M = 2100$ nine-point stencils (one stencil per interior mesh point per sample). With the base stencil $\mathbf{u}_b = 0 \in (\mathbb{R}^2)^9$, normalisation constants $c_b = 1$, $p_b = (1, -1)^\top$, and kernel $K(x, y) = \exp(-1/2\|x - y\|^2)$, the stencil data \mathfrak{U} defines a posterior Gaussian process $\xi^M \in \mathcal{N}(L_d^M, \mathcal{K}_{\Phi_b^M})$ (see Section 3.2).

4.2.2. Recovery of training data from initial values

Neglecting effects of finite-precision arithmetic and assuming non-degeneracy, by construction, training data can be reproduced exactly from initial data at $t = 0$, Δt by forming the temporal Lagrangian $L_{d, \Delta x}^M$ and forward propagation in time using $\text{DEL}(L_{d, \Delta x}^M) =$

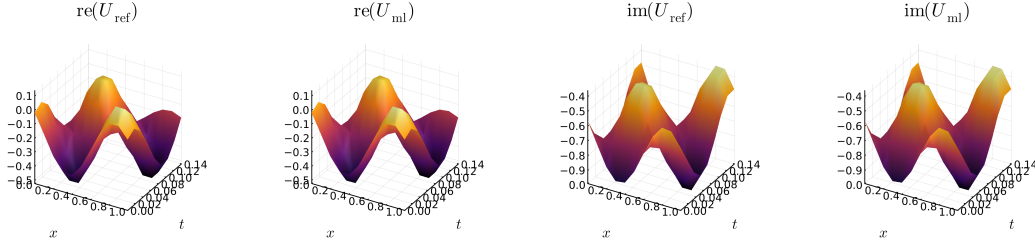


Figure 6: Schrödinger equation experiment. Recovery of training data from initial values at $t = 0, \Delta t$ by the learned model. From left to right: real part of training data, real part of predicted field, imaginary part of training data, imaginary part of predicted field.

0. Moreover, the model uncertainty $\sigma(\text{DEL}_{\mathbf{u}}(\xi^M))$ at stencils \mathbf{u} in the training data \mathcal{U} is zero in exact arithmetic. This is confirmed in the numerical experiment in Figure 6. The discrete l_2 norm of reference and predicted field is smaller than $1.1e - 9$. The maximal value of each component of $\sigma(\text{DEL}_{\mathbf{u}}(\xi^M))$ is smaller than $1.45e - 7$ for $\mathbf{u} \in \mathcal{U}$.

4.2.3. Extrapolation

Moreover, Figure 7 shows that the trained model can successfully predict motions from unseen initial data that was sampled from the same distribution as the training data. The discrete l_2 norm of the error is smaller $3.9e - 3$. Model uncertainty along the computed field is quantified as the component-wise standard deviation of $\text{DEL}_{\mathbf{u}(i,j)}(\xi^M)$ and visualised in Figure 7 displaying an increase in uncertainty as the solution is propagated forward in time.

Reproducibility

Computations have been performed in Julia [3] using packages NLSolve.jl (part of the Optim.jl project [22]) and ForwardDiff.jl [31], among others. The source code of the experiments can be found at https://github.com/Christian-Offen/Lagrangian_GP_PDE.

5. Convergence Analysis

In this section, we will state and prove convergence results for the considered method to learn discrete Lagrangian densities from temporal-spatial data. For notational convenience, the statement (Theorem 1) is formulated for three-point Lagrangian densities related to a 1+1-dimensional temporal-spatial dynamical systems. However, Theorem 1 and its proof extend in a straight forward manner to other types of discrete Lagrangian densities such as the four-point Lagrangian densities employed in the numerical experiment of Section 4.2 or, more generally, for the class of discrete Lagrangian densities

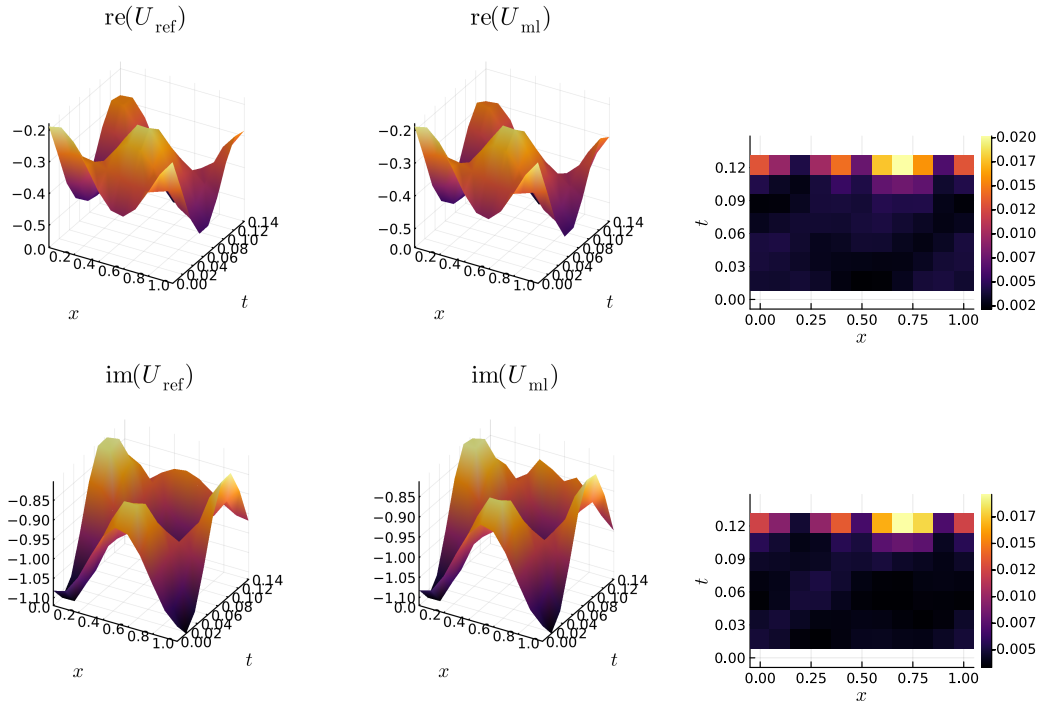


Figure 7: Schrödinger equation experiment. From left to right: true solution (not part of training data), predicted solution, standard deviation of $\text{DEL}(\xi^M)$ at stencils of the predicted solution. First row refers to real part, second row to imaginary parts of the fields.

described in Section II.B of [26].

Additionally, Appendix B extracts a statement about temporal discrete Lagrangians as a special case of the presented theory and relates the statement to the discussion in our article on (temporal) Lagrangian identification [24].

5.1. Statement of convergence theorem

Theorem 1 *Let $\Omega \subset \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ be open, bounded, non-empty domains. For $\mathbf{u} = (u, u^+, u_+, u^-, u_-, u_+^-, u_-^+) \in (\mathbb{R}^d)^7$ consider the projections $\text{pr}_1(\mathbf{u}) = (u, u^+, u_+)$, $\text{pr}_2(\mathbf{u}) = (u^-, u, u_+^-)$, $\text{pr}_3(\mathbf{u}) = (u_-, u_+^-, u)$. Consider a sequence of observations*

$$\Omega_0 = \{\mathbf{u}^{(j)} = (u^{(j)}, u^{+(j)}, u_+^{(j)}, u^{-(j)}, u_+^{-(j)}, u_-^{(j)}, u_-^{+(j)})\}_{j=1}^\infty \subset \bigcap_{l=1}^3 \text{pr}_l^{-1}(\Omega)$$

of a dynamical system governed by the discrete Euler–Lagrange equation of an (unknown) discrete Lagrangian density $L_d^{\text{ref}} \in \mathcal{C}^1(\overline{\Omega})$, i.e. $\text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 0$ for all $\mathbf{u} \in \Omega_0$. Let $K \in \mathcal{C}^1(\overline{\Omega} \times \overline{\Omega})$ be a continuously differentiable kernel on Ω , $\Gamma_b \in \Omega$, $r_b \in \mathbb{R}$, $p_b \in \mathbb{R}$ and assume that L_d^{ref} is contained in the reproducing kernel Hilbert space B to K and fulfils the normalisation condition

$$\Phi_b(L_d^{\text{ref}}) = (p_b, r_b) \quad \text{with} \quad \Phi_b(L_d) = (\nabla_2 L_d(\Gamma_b), L_d(\Gamma_b)) \quad (23)$$

and that B embeds continuously into $\mathcal{C}^1(\overline{\Omega})$. Let $\xi \in \mathcal{N}(0, \mathcal{K})$ be a canonical Gaussian random variable over B . Then the sequence of conditional means $L_{d,(j)}$ of ξ conditioned on the first j observations and the normalisation conditions

$$\text{DEL}(\xi)(\mathbf{u}^{(i)}) = 0 \quad (\forall i \leq j), \quad \Phi_N(\xi) = (p_b, r_b) \quad (24)$$

converges in the reproducing kernel Hilbert space norm $\|\cdot\|_B$ and in the uniform norm of the space of differentiable functions $\|\cdot\|_{\mathcal{C}^1(\overline{\Omega})}$ to a Lagrangian $L_{d,(\infty)} \in B$ that is

- *consistent with the normalisation $\Phi_b(L_{d,(\infty)}) = (p_b, r_b)$*
- *consistent with the dynamics, i.e. $\text{DEL}(L_{d,(\infty)})(\mathbf{u}) = 0$ for all $\mathbf{u} \in \overline{\Omega}_0$*
- *and $L_{d,(\infty)}$ minimizes $\|\cdot\|_B$ among all discrete Lagrangian densities with the properties above. □*

Remark 3 *By addition of null-Lagrangian densities (see Proposition 1) L_d^{ref} can be assumed to fulfil the normalisation condition (23) without loss of generality. However, it is stated in this form to assume compatibility with the reproducing kernel Hilbert space B . □*

Remark 4 *It makes sense to consider observations Ω_0 in Theorem 1 that densely fill a sufficiently large space, i.e. $\Omega \subset \text{pr}_k(\overline{\Omega}_0)$ for $k = 1, 2, 3$, such that the dynamics on all of Ω is covered. □*

5.2. Setting and proof of Theorem 1

Let $\Omega \subset \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ be an open, bounded, non-empty domain. For data $\mathbf{u} = (u, u^+, u_+, u^-, u_-, u_-) \in (\mathbb{R}^d)^7$ define the projections $\text{pr}_1(\mathbf{u}) = (u, u^+, u_+)$, $\text{pr}_2(\mathbf{u}) = (u^-, u, u_-)$, $\text{pr}_3(\mathbf{u}) = (u_-, u_+, u)$. To a discrete Lagrangian $L_d^{\text{ref}} \in \mathcal{C}^1(\bar{\Omega})$ consider the set of admissible stencil data

$$\hat{\Omega}_{\text{pre}} = \left\{ \mathbf{u} = (u, u^+, u_+, u^-, u_-, u_-) \mid \text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 0, \text{pr}_k(\mathbf{u}) \in \bar{\Omega}, k = 1, 2, 3 \right\}.$$

The set $\hat{\Omega}_{\text{pre}}$ is closed in $(\mathbb{R}^d)^7$.

Remark 3 In degenerate/symmetric situations, $\hat{\Omega}_{\text{pre}}$ can contain additional stencils that do not relate to discrete fields: for instance, for $L_d^{\text{ref}}(u, u^+, u_+) = \frac{(u^+ - u)^2}{2\Delta t^2} - \frac{(u_+ - u)^2}{2\Delta x^2}$ the discrete Euler–Lagrange equation $\text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 0$ is a five-point stencil only relating u, u^+, u^-, u_+, u_- rather than a seven-point stencil relating $u, u^+, u_+, u^-, u_-, u_-$. Thus $(u, u^+, u_+, u^-, r_1, u_-, r_2) \in \hat{\Omega}_{\text{pre}}$ for all $r_1, r_2 \in \mathbb{R}$ whenever

$$\frac{u^- - 2u + u^+}{\Delta t^2} - \frac{u_- - 2u + u_+}{\Delta x^2} = 0.$$

To cover the case where training data stencils are collected from observed discrete fields, we will, therefore, proof the statement for observations that are dense in some closed subset $\hat{\Omega}$ of $\hat{\Omega}_{\text{pre}}$. \square

Let $\Omega_0 = \{\mathbf{u}^{(j)}\}_{j=1}^\infty \subset \hat{\Omega}_{\text{pre}}$ be a sequence and $\hat{\Omega} := \bar{\Omega}_0 \subset \hat{\Omega}_{\text{pre}}$ (topological closures in $\hat{\Omega}_{\text{pre}}$ and $(\mathbb{R}^d)^7$ coincide). The discrete Lagrangian operator DEL constitutes a bounded linear operator

$$\Phi^\infty: \mathcal{C}^1(\bar{\Omega}) \rightarrow \mathcal{C}^0(\hat{\Omega}, \mathbb{R}^d), \quad L_d \mapsto \text{DEL}(L_d).$$

Since for each \mathbf{u} the evaluation functional $\text{ev}_{\mathbf{u}}: f \mapsto f(\mathbf{u})$ on $\mathcal{C}^0(\hat{\Omega}, \mathbb{R}^d)$ is bounded, the following functions constitute bounded linear functionals for $j \in \mathbb{N}$:

$$\begin{aligned} \Phi_j: \mathcal{C}^1(\bar{\Omega}) &\rightarrow \mathbb{R}^d, & \Phi_j(L_d) &= \Phi^{(\infty)}(L_d)(\mathbf{u}^{(j)}) \\ \Phi^{(j)}: \mathcal{C}^1(\bar{\Omega}) &\rightarrow (\mathbb{R}^d)^j, & \Phi^{(j)} &= (\Phi_1, \dots, \Phi_j). \end{aligned}$$

For a reference point $\square_b \in \Omega$ and for $p_b \in \mathbb{R}^d$, $r_b \in \mathbb{R}$ we define the bounded linear functional

$$\Phi_b: \mathcal{C}^1(\bar{\Omega}) \rightarrow \mathbb{R}^{d+1}, \quad \Phi_b(L) = (\nabla_2 L_d(\square_b), L_d(\square_b)), \quad (25)$$

related to our normalisation condition on the local momentum Mm^+ for seven-point Lagrangians (Section 2.3). Here $\nabla_2 L_d$ denotes the derivative with respect to the second component of L_d . We will further use the shorthands

$$\Phi_b^{(k)} = (\Phi_1, \dots, \Phi_k, \Phi_b) \quad \text{and} \quad \Phi_b^{(\infty)} = (\Phi^{(\infty)}, \Phi_b),$$

and define

$$\begin{aligned} y^{(k)} &= (0, \dots, 0, p_b, r_b) \in (\mathbb{R}^d)^k \times \mathbb{R}^d \times \mathbb{R} \\ y^{(\infty)} &= (0, p_b, r_b) \in \mathcal{C}^0(\overline{\Omega}, \mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}. \end{aligned}$$

Consider the following assumption.

Assumption 1 *Assume that there is a reflexive, uniformly convex Banach space B with continuous embedding $B \hookrightarrow \mathcal{C}^1(\overline{\Omega})$ such that*

$$\{L_d \in \mathcal{C}^1(\overline{\Omega}) \mid \Phi_b^{(\infty)}(L_d) = y^{(\infty)}\} \cap B \neq \emptyset$$

In other words, B is assumed to contain a Lagrangian consistent with the normalisation and the underlying dynamics.

The affine linear subspace

$$\begin{aligned} A^{(j)} &= \{L_d \in B \mid \Phi_b^{(j)}(L_d) = y^{(j)}\} \quad (j \in \mathbb{N}) \\ A^{(\infty)} &= \{L_d \in B \mid \Phi_b^{(\infty)}(L_d) = y^{(\infty)}\} \end{aligned}$$

are closed in B and not empty by Assumption 1. Therefore, the following minimisations constitute convex optimisation problems on B with unique minima in $A^{(j)}$ or $A^{(\infty)}$, respectively:

$$\begin{aligned} L_{d(j)} &= \arg \min_{L_d \in A^{(j)}} \|L_d\|_B \\ L_{d(\infty)} &= \arg \min_{L_d \in A^{(\infty)}} \|L_d\|_B \end{aligned} \tag{26}$$

Here $\|\cdot\|_B$ denotes the norm in B .

Proposition 2 *The minima $L_{d(j)}$ converge to $L_{d(\infty)}$ in the norm $\|\cdot\|_B$ and, thus, in $\|\cdot\|_{\mathcal{C}^1(\overline{\Omega})}$. \square*

PROOF The proof consists of the following steps and follows the same strategy as in the case of temporal Lagrangians [24]. We will prove:

1. The sequence $(L_{d(j)})_j$ has a weakly convergent subsequence.
2. The weak limit $L_{d(\infty)}^\dagger$ is contained in $A^{(\infty)}$.
3. $L_{d(\infty)}^\dagger$ coincides with $L_{d(\infty)}$.
4. The whole sequence $(L_{d(j)})_j$ converges weakly against $L_{d(\infty)}$.
5. The norms $\|L_{d(j)}\|_B$ converge to $\|L_{d(\infty)}\|_B$.
6. Step 4 and 5 imply that the sequence $(L_{d(j)})_j$ converges strongly against $L_{d(\infty)}$.

Step 1. The sequence of affine spaces $A^{(1)} \supseteq A^{(2)} \supseteq A^{(3)} \supseteq \dots$ decreases monotonously. Moreover, $\emptyset \neq A^{(\infty)} \subseteq \bigcap_{j=1}^{\infty} A^{(j)}$. It follows that the norms $\|L_{d(j)}\|_B$ increase monotonously. The sequence is bounded by $\|L_{d(\infty)}\|_B$. By reflexivity of B , we find a weakly convergent subsequence $(L_{d(j_i)})_{i \in \mathbb{N}}$ (see [4, Theorem 3.18]). Let $L_{j_i} \rightharpoonup L_{d(\infty)}^\dagger$ denote the weak limit in B . We have

$$\|L_{d(\infty)}^\dagger\|_B \leq \liminf_{i \rightarrow \infty} \|L_{d(j_i)}\|_B \leq \|L_{d(\infty)}\|_B \quad (27)$$

by the weak lower semi-continuity of the norm.

Step 2. We show $L_{d(\infty)}^\dagger \in A^{(\infty)}$.

Let $\mathbf{u} \in \hat{\Omega}$. As the sequence $\Omega_0 = (\mathbf{u}^{(m)})_{m=1}^{\infty}$ is dense in $\hat{\Omega}$, there exists a subsequence $(\mathbf{u}^{(m_l)})_{l=1}^{\infty}$ converging to \mathbf{u} .

For each l there is N such that $j_N \geq m_l$. For $i \geq N$ it holds that $\Phi_b^{(\infty)}(L_{d(j_i)})(\mathbf{u}^{(m_l)}) = 0$. Therefore, for each l

$$\lim_{i \rightarrow \infty} \Phi_b^{(\infty)}(L_{d(j_i)})(\mathbf{u}^{(m_l)}) = 0.$$

For each l the linear operator $\Phi_b^{(\infty)}(\cdot)(\mathbf{u}^{(m_l)}): B \rightarrow \mathbb{R}^d \times \mathbb{R}^{d+1}$ into the finite-dimensional space $\mathbb{R}^d \times \mathbb{R}^{d+1}$ is bounded. By the weak convergence $L_{d(j_i)} \rightharpoonup L_{d(\infty)}^\dagger$ we have

$$\Phi_b^{(\infty)}(L_{d(\infty)}^\dagger)(\mathbf{u}^{(m_l)}) = \lim_{i \rightarrow \infty} \Phi_b^{(\infty)}(L_{d(j_i)})(\mathbf{u}^{(m_l)}) = 0.$$

By continuity $\Phi_b^{(\infty)}(L_{d(\infty)}^\dagger) \in \mathcal{C}^0(\bar{\Omega})$, we have

$$\Phi_b^{(\infty)}(L_{d(\infty)}^\dagger)(\mathbf{u}) = \lim_{l \rightarrow \infty} \Phi_b^{(\infty)}(L_{d(\infty)}^\dagger)(\mathbf{u}^{(m_l)}) = 0.$$

As this holds for all $\mathbf{u} \in \hat{\Omega}$ we conclude $L_{d(\infty)}^\dagger \in A^{(\infty)}$.

Step 3. As $L_{d(\infty)}^\dagger \in A^{(\infty)}$ by Step 2, we have $\|L_{d(\infty)}^\dagger\|_B \geq \|L_{d(\infty)}\|_B$ since $L_{d(\infty)}$ is the unique minimiser of the minimisation problem of (26). Together with (27) we conclude $\|L_{d(\infty)}^\dagger\|_B = \|L_{d(\infty)}\|_B$. By uniqueness of the minimiser $L_{d(\infty)}$ this implies $L_{d(\infty)}^\dagger = L_{d(\infty)}$. It follows that $L_{d(j_i)}$ converges weakly to $L_{d(\infty)}$.

Step 4. The choice of a weakly convergent subsequence of $L_{d(j)}$ in Step 1 was arbitrary. Thus any subsequence of $L_{d(j)}$ has a weakly convergent subsequence and any weakly convergent subsequence has weak limit $L_{d(\infty)}$. It follows that the whole series $L_{d(j)}$ converges weakly to the unique weak limit $L_{d(\infty)}$.

Step 5. As $L_{d(j)} \rightharpoonup L_{d(\infty)}$, and $\|L_{d(j)}\|_B \leq \|L_{d(\infty)}\|_B$, and by the weak lower semi-continuity of the norm

$$\|L_{d(\infty)}\|_B \leq \liminf_{j \rightarrow \infty} \|L_{d(j)}\|_B \leq \limsup_{j \rightarrow \infty} \|L_{d(j)}\|_B \leq \|L_{d(\infty)}\|_B.$$

Therefore

$$\lim_{j \rightarrow \infty} \|L_{d(j)}\|_B = \|L_{d(\infty)}\|_B. \quad (28)$$

Step 6. As B is uniformly convex, by [4, Prop. 3.32] weak convergence $L_{d(j)} \rightharpoonup L_{d(\infty)}^\dagger$ together with (28) implies strong convergence $\lim_{j \rightarrow \infty} L_{d(j)} = L_{d(\infty)}$. \blacksquare

PROOF (THEOREM 1) By [28, Thm 12.5] the unique minimisers $L_{d(j)}$ in (26) coincide with the conditional means considered in Theorem 1. Moreover, reproducing kernel Hilbert spaces are uniformly convex, reflexive Banach spaces. Thus the statement follows directly from Proposition 2. \blacksquare

6. Summary and further directions

We have introduced a method based on Gaussian Process regression to obtain data-driven models of discrete Lagrangian densities that define discrete field theories. The data consists of discrete observations of fields of the true field theory, which we decompose into stencil data \mathfrak{U} for the discrete Euler–Lagrange equation of a discrete Lagrangian density. A canonical Gaussian process ξ modelling a discrete Lagrangian density is conditioned to fulfil the discrete Euler–Lagrange equation $\text{DEL}_{\mathfrak{u}}(\xi) = 0$ on all observed stencils $\mathfrak{u} \in \mathfrak{U}$. As these conditions are linear in ξ , the posterior process is Gaussian as well and the conditional mean and the conditional covariance operator can be explicitly computed. To avoid degeneracies of the posterior processes, we further condition the process on non-triviality conditions. We show that this does not restrict the generality of our ansatz. Next to providing a method to predict solutions to field theories, we use the statistical framework of Gaussian process regression to obtain a quantification of model uncertainty of the learned discrete Euler–Lagrange equations and any linear observable in the conditioned Gaussian process. Our method is structure-preserving in the sense that the learned discrete field theories are guaranteed to fulfil a discrete variational principle by construction. To illustrate the method, we provide numerical examples based on the wave equation and the Schrödinger equation.

The method can be interpreted as a meshless collocation method [34] for solving the discrete Euler–Lagrange equation for the discrete Lagrangian L_d . Overcoming the ambiguity of discrete Lagrangian densities (gauge freedom), we provide a statement that guarantees convergence of the posterior means to a true discrete Lagrangian density as the distance of data points converges to zero. For this, the proof exploits a characterisation of posterior means as minimisers of a reproducing kernel Hilbert space norm constrained by the observations [28].

In future work, it is of interest to prove convergence rates for the provided method and to develop efficient computational methods for evaluations of the machine-learned theory in high-dimensional, large data-regimes. Moreover, techniques of this article can be combined with Lie group based methods to learn symmetric representations of (discrete) temporal Lagrangians (such as [14]) which may be extended to Lagrangian densities and temporal-spatial symmetries. These are of interest for the data-driven identification of conservation laws and the detection of structurally simple solutions such as travelling [25, 26]. Furthermore, rather than learning of one specific variational formulation to a field theory, learning of alternative (i.e. non-gauge equivalent) variational formulations is of interest for system identification as alternative variational formulations allow for the derivation of a series of conservation laws [12, 17, 16, 5]. Finally and more generally, it is of interest to obtain a fundamental understanding of the interplay between geometric

prior knowledge, data requirements, and model uncertainty to clarify the role of geometry in machine learning.

Acknowledgments

The author would like to thank Konstantin Sonntag for a helpful discussion that has advanced the convergence proof. Moreover, the author acknowledges the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen and computing time provided by the Paderborn Center for Parallel Computing (PC2).

Data availability

The data that support the findings of this study are openly available in the GitHub repository Christian-Offen/Lagrangian_GP_PDE at https://github.com/Christian-Offen/Lagrangian_GP_PDE.

Appendices

A. Continuous Lagrangian theories

Let us review of continuous Euler–Lagrange dynamics. For details and a more formal treatment we refer to Section 5 and the literature on variational calculus [11, 32].

A.1. The Euler–Lagrange equations for continuous theories

The Euler–Lagrange operator for a first order field theory on \mathbb{R}^n or subsets of \mathbb{R}^n with coordinate $t = (t_1, \dots, t_n)$ and \mathbb{R}^d -valued field $u = (u^1, \dots, u^n)$ is given as

$$\begin{aligned} \text{EL}(L)^r((u^s)_s, (u_{t_k}^s)_{s,k}, (u_{t_l, t_k}^s)_{s,l,k}) \\ &= \sum_{k=1}^n \frac{\partial}{\partial t_k} \left(\frac{\partial L}{\partial u_{t_k}^r} \right) - \frac{\partial L}{\partial u^r} \\ &= \sum_{k,l=1}^n \sum_{s=1}^d \left(\frac{\partial^2 L}{\partial u_{t_k}^r \partial u_{t_l}^s} u_{t_k, t_l}^s \right) + \sum_{k=1}^n \sum_{s=1}^d \left(\frac{\partial^2 L}{\partial u_{t_k}^r \partial u^s} u_{t_l}^s \right) - \frac{\partial L}{\partial u^r} \end{aligned} \quad (29)$$

for $r = 1, \dots, d$. The Euler–Lagrange equation

$$\text{EL}(L) \left(u(t), \frac{\partial u}{\partial t_k}(t), \frac{\partial u}{\partial t_k \partial t_l}(t) \right) = 0 \quad (30)$$

constitutes a system of d partial differential equations. Here we use the convention $\text{EL}(L) = (\text{EL}(L)^1, \dots, \text{EL}(L)^d)$. Unless degenerate, the equations are of second order.

In case $n = 1$ the equation simplifies to the following set of second order ordinary differential equation:

$$0 = \text{EL}(L)(u, u_t, u_{tt}) = \frac{\partial^2 L}{\partial u \partial u_t} u_t + \frac{\partial^2 L}{\partial u_t \partial u_t} u_{tt} - \frac{\partial L}{\partial u}. \quad (31)$$

Here the first and second summand denote matrix vector multiplications and $\frac{\partial L}{\partial u}$ denotes the gradient of L with respect to u .

A.2. Ambiguity of Lagrangian densities

Lagrangian densities can be ambiguous in two different ways:

1. Lagrangians L and \tilde{L} can yield the same Euler–Lagrange operator up to rescaling, i.e.

$$\text{EL}(L) = \rho \cdot \text{EL}(\tilde{L}), \quad \rho \in \mathbb{R} \setminus \{0\}$$

and, therefore, the same Euler–Lagrange equations up to rescaling. We call L and \tilde{L} (*gauge-*) *equivalent*. For equivalent Lagrangians L, \tilde{L} there exists $\rho \in \mathbb{R} \setminus \{0\}$, $c \in \mathbb{R}$ such that $L - \rho\tilde{L} - c$ is a total divergence

$$L - \rho\tilde{L} - c = \text{div}_t F$$

for a continuously differentiable function $F = (F^1, \dots, F^n)$, where

$$\text{div}_t F(u, (u_{t_k})_k) = \sum_{s=1}^d \sum_{k=1}^n u_{t_k}^s \frac{\partial F^j}{\partial u^s}(u). \quad (32)$$

(We have restricted ourselves to autonomous Lagrangians.) This may be seen by noting that for the corresponding action functionals S and \tilde{S} the difference $S(u) - \rho\tilde{S}(u)$ is an integral over the boundary $\partial\mathcal{M}$ if $u: \mathcal{M} \rightarrow \mathbb{R}^d$. However, the considered variations fix $\partial\mathcal{M}$ and, thus, do not influence the stationarity of a field u [11].

2. Even when two Lagrangians L and \tilde{L} are not equivalent, they can yield the same set of solutions u , i.e.

$$\forall u: \mathcal{M} \rightarrow \mathbb{R}^d : \text{EL}(L)(u(t)) = 0 \iff \text{EL}(\tilde{L})(u(t)) = 0,$$

where $\mathcal{M} \subseteq \mathbb{R}^n$ is the domain of the field theory. Here

$$\mathbf{u}(t) = \left(u(t), \left(\frac{\partial u(t)}{\partial t_k} \right)_k, \left(\frac{\partial^2 u(t)}{\partial t_k \partial t_l} \right)_{k,l} \right)$$

denotes the second jet extension of a field u . In that case \tilde{L} is called an *alternative Lagrangian density* to L .

Example 1 (Affine linear motions) For any twice differentiable $g: \mathbb{R}^d \rightarrow \mathbb{R}$ with nowhere degenerate Hessian matrix $\text{Hess}(g)$, the Lagrangian $L(u, u_t) = g(u_t)$ describes affine linear motions in \mathbb{R}^d :

$$0 = \text{EL}(L) = \text{Hess}(g)(u_t)u_{tt}. \quad \square$$

In general, the existence of alternative Lagrangian densities is related to additional geometric structure and conserved quantities of the system [12, 17, 16, 5]. Gauge equivalence, however, is exhibited by all variational systems.

B. Alternative convergence statement for discrete Lagrangians

Theorem 1 provides a blueprint for convergence results for various types of discrete Lagrangians. In Theorem 2 below we reformulate Theorem 1 for discrete temporal Lagrangians. Discrete temporal Lagrangians have already been considered separately in [24]. We will then relate the statement of [24] and Theorem 2.

Theorem 2 *Let $\Omega \subset \mathbb{R}^d \times \mathbb{R}^d$ be open, bounded, non-empty domains. Consider a sequence of observations*

$$\Omega_0 = \{\mathbf{u}^{(j)} = (u^{-(j)}, u^{(j)}, u^{+(j)})\}_{j=1}^{\infty}$$

of a dynamical system governed by the discrete Euler–Lagrange equation of an (unknown) discrete Lagrangian density $L_d^{\text{ref}} \in \mathcal{C}^1(\bar{\Omega})$, i.e.

$$\Omega_0 \subset \hat{\Omega} = \left\{ \mathbf{u} = (u^-, u, u^+) \mid \text{DEL}(L_d^{\text{ref}})(\mathbf{u}) = 0, \begin{array}{l} (u, u^+) \in \bar{\Omega} \\ (u^-, u) \in \bar{\Omega} \end{array} \right\}.$$

Assume $\Omega \subset \text{pr}_-\bar{\Omega}_0$ and $\Omega \subset \text{pr}_+\bar{\Omega}_0$ with $\text{pr}_-(u^-, u, u^+) = (u^-, u)$ and $\text{pr}_+(u^-, u, u^+) = (u, u^+)$. Let $K \in \mathcal{C}^1(\bar{\Omega} \times \bar{\Omega})$ be a kernel on Ω , $u_b \in \Omega$, $r_b \in \mathbb{R}$, $p_b \in \mathbb{R}$ and assume that L_d^{ref} is contained in the reproducing kernel Hilbert space B to K and fulfils the normalisation condition

$$\Phi_b(L_d^{\text{ref}}) = (p_b, r_b) \quad \text{with} \quad \Phi_b(L_d) = (-\nabla_2 L_d(u_b), L_d(u_b)) \quad (33)$$

and that B embeds continuously into $\mathcal{C}^1(\bar{\Omega})$. Let $\xi \in \mathcal{N}(0, \mathcal{K})$ be a canonical Gaussian random variable over B . Then the sequence of posterior means $L_{d,(j)}$ of the Gaussian field ξ conditioned on the first j observations and the normalisation conditions

$$\text{DEL}(\xi)(\mathbf{u}^{(i)}) = 0 \ (\forall i \leq j), \quad \Phi_b(\xi) = (p_b, r_b) \quad (34)$$

converges in the reproducing kernel Hilbert space norm $\|\cdot\|_B$ and in the uniform norm of the space of differentiable functions $\|\cdot\|_{\mathcal{C}^1(\bar{\Omega})}$ to a Lagrangian $L_{d,(\infty)} \in B$ that is

- *consistent with the normalisation $\Phi_b(L_{d,(\infty)}) = (p_b, r_b)$*

- consistent with the dynamics, i.e. $\text{DEL}(L_{d,(\infty)})(\mathbf{u}) = 0$ for all $\mathbf{u} \in \overline{\Omega_0}$
- and it minimizes $\|\cdot\|_B$ among discrete Lagrangians with the above properties. \square

PROOF The proof of Theorem 1 can be adapted easily. \blacksquare

In comparison to [24, Theorem 2], Theorem 2 requires $\Omega = \Omega_a = \Omega_b$ (slightly more restrictive). Under the assumptions that a globally Lipschitz continuous discrete flow map g exists governing the observed dynamical system, any dense set $\{(x_0^{(j)}, x_1^{(j)})\}_{j=1}^{\infty} \subset \Omega_a$ gives rise to a dense, countable set of observations

$$\Omega_0 = \{(x_0^{(j)}, g(x_0^{(j)}, x_1^{(j)}))\} \subset \hat{\Omega}$$

where $\hat{\Omega} = \{(x_0, g(x_0, x_1)) \mid (x_0, x_1) \in \overline{\Omega}\}$. Hence, [24, Theorem 2] follows from Theorem 2 in case $\Omega_a = \Omega_b$.

References

- [1] Christine Allen-Blanchette, Sushant Veer, Anirudha Majumdar, and Naomi Ehrich Leonard. LagNetViP: A Lagrangian neural network for video prediction (AAAI 2020 symposium on physics guided ai), 2020. doi:10.48550/ARXIV.2010.12932.
- [2] Takehiro Aoshima, Takashi Matsubara, and Takaharu Yaguchi. Deep discrete-time lagrangian mechanics. *ICLR SimDL*, 5 2021. URL: <https://simdl.github.io/files/49.pdf>.
- [3] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017. doi:10.1137/141000671.
- [4] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York, New York, NY, 2011. doi:10.1007/978-0-387-70914-7.
- [5] J F Carinena and L A Ibort. Non-noether constants of motion. *Journal of Physics A: Mathematical and General*, 16(1):1, 1 1983. doi:10.1088/0305-4470/16/1/010.
- [6] Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M. Stuart. Solving and learning nonlinear pdes with Gaussian processes. *Journal of Computational Physics*, 447:110668, 2021. doi:10.1016/j.jcp.2021.110668.
- [7] Yifan Chen, Houman Owhadi, and Florian Schäfer. Sparse Cholesky factorization for solving nonlinear pdes via Gaussian processes. *Mathematics of Computation*, June 2024. doi:10.1090/mcom/3992.

- [8] Andreas Christmann and Ingo Steinwart. *Kernels and Reproducing Kernel Hilbert Spaces*, pages 110–163. Springer New York, New York, NY, 2008. doi:10.1007/978-0-387-77242-4_4.
- [9] Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks, 2020. doi:10.48550/ARXIV.2003.04630.
- [10] Giulio Evangelisti and Sandra Hirche. Physically consistent learning of conservative lagrangian systems with gaussian processes. In *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022. doi:10.1109/CDC51059.2022.9993123.
- [11] I.M. Gelfand, S.V. Fomin, and R.A. Silverman. *Calculus of Variations*. Dover Books on Mathematics. Dover Publications, 2000.
- [12] Marc Henneaux. Equations of motion, commutation relations and ambiguities in the Lagrangian formalism. *Annals of Physics*, 140(1):45–64, 1982. doi:10.1016/0003-4916(82)90334-7.
- [13] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, Jun 2021. doi:10.1038/s42254-021-00314-5.
- [14] Yana Lishkova, Paul Scherer, Steffen Ridderbusch, Mateja Jamnik, Pietro Liò, Sina Ober-Blöbaum, and Christian Offen. Discrete Lagrangian neural networks with automatic symmetry discovery. *IFAC-PapersOnLine*, 56(2):3203–3210, 2023. 22nd IFAC World Congress. doi:10.1016/j.ifacol.2023.10.1457.
- [15] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, mar 2021. URL: <https://doi.org/10.1038/s42256-021-00302-5>, doi:10.1038/s42256-021-00302-5.
- [16] Giuseppe Marmo and G. Morandi. On the inverse problem with symmetries, and the appearance of cohomologies in classical Lagrangian dynamics. *Reports on Mathematical Physics*, 28(3):389–410, 1989. doi:10.1016/0034-4877(89)90071-2.
- [17] Giuseppe Marmo and C. Rubano. On the uniqueness of the Lagrangian description for charged particles in external magnetic field. *Il Nuovo Cimento A*, 98(4):387–399, 10 1987. doi:10.1007/bf02902083.
- [18] Jerrold E. Marsden, George W. Patrick, and Steve Shkoller. Multisymplectic geometry, variational integrators, and nonlinear pdes. *Communications in Mathematical Physics*, 199(2):351–395, 12 1998. doi:10.1007/s002200050505.
- [19] Jerrold E. Marsden, Sergey Pekarsky, Steve Shkoller, and Matthew West. Variational methods, multisymplectic geometry and continuum mechanics. *Journal*

- of Geometry and Physics*, 38(3):253–284, 2001. doi:10.1016/S0393-0440(00)00066-8.
- [20] Jerrold E. Marsden and Matthew West. Discrete mechanics and variational integrators. *Acta Numerica*, 10:357–514, 2001. doi:10.1017/S096249290100006X.
- [21] Justice Mason, Christine Allen-Blanchette, Nicholas Zolman, Elizabeth Davison, and Naomi Leonard. Learning interpretable dynamics from images of a freely rotating 3d rigid body, 2022. doi:10.48550/ARXIV.2209.11355.
- [22] Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615, 2018. doi:10.21105/joss.00615.
- [23] Sina Ober-Blöbaum and Christian Offen. Variational learning of Euler–Lagrange dynamics from data. *Journal of Computational and Applied Mathematics*, 421:114780, 2023. doi:10.1016/j.cam.2022.114780.
- [24] Christian Offen. Machine learning of continuous and discrete variational odes with convergence guarantee and uncertainty quantification, 2024. arXiv:2404.19626.
- [25] Christian Offen and Sina Ober-Blöbaum. Learning discrete lagrangians for variational pdes from data and detection of travelling waves. In Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information*, volume 14071, pages 569–579, Cham, 2023. Springer Nature Switzerland. doi:10.1007/978-3-031-38271-0_57.
- [26] Christian Offen and Sina Ober-Blöbaum. Learning of discrete models of variational PDEs from data. *Chaos*, 34:013104, 1 2024. doi:10.1063/5.0172287.
- [27] Houman Owhadi and Clint Scovel. *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2019. doi:10.1017/9781108594967.
- [28] Houman Owhadi and Clint Scovel. *Optimal Recovery Splines*, page 154–159. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2019. doi:10.1017/9781108594967.017.
- [29] Hong Qin. Machine learning and serving of discrete field theories. *Scientific Reports*, 10(1), 11 2020. doi:10.1038/s41598-020-76301-0.
- [30] Joaquin Quiñonero-Candela and Carl Edward Rasmussen. A unifying view of sparse approximate gaussian process regression. *Journal of Machine Learning Research*, 6(65):1939–1959, 2005. URL: <http://jmlr.org/papers/v6/quinonero-candela05a.html>.

- [31] Jarrett Revels, Miles Lubin, and Theodore Papamarkou. Forward-mode automatic differentiation in Julia, 2016. [arXiv:1607.07892](https://arxiv.org/abs/1607.07892).
- [32] Tomáš Roubíček. *Calculus of Variations*, pages 1–38. John Wiley & Sons, Ltd, 2015. [doi:10.1002/3527600434.eap735](https://doi.org/10.1002/3527600434.eap735).
- [33] Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017. [doi:10.1126/sciadv.1602614](https://doi.org/10.1126/sciadv.1602614).
- [34] Robert Schaback and Holger Wendland. Kernel techniques: From machine learning to meshless methods. *Acta Numerica*, 15:543–639, 2006. [doi:10.1017/S0962492906270016](https://doi.org/10.1017/S0962492906270016).
- [35] Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2197):20160446, 2017. [doi:10.1098/rspa.2016.0446](https://doi.org/10.1098/rspa.2016.0446).
- [36] Florian Schäfer, Matthias Katzfuss, and Houman Owhadi. Sparse Cholesky factorization by Kullback–Leibler minimization. *SIAM Journal on Scientific Computing*, 43(3):A2019–A2046, 2021. [doi:10.1137/20M1336254](https://doi.org/10.1137/20M1336254).
- [37] Harsh Sharma and Boris Kramer. Preserving Lagrangian structure in data-driven reduced-order modeling of large-scale dynamical systems, 2022. [doi:10.48550/ARXIV.2203.06361](https://doi.org/10.48550/ARXIV.2203.06361).
- [38] Harsh Sharma, Hongliang Mu, Patrick Buchfink, Rudy Geelen, Silke Glas, and Boris Kramer. Symplectic model reduction of Hamiltonian systems using data-driven quadratic manifolds, 2023. [arXiv:2305.15490](https://arxiv.org/abs/2305.15490).
- [39] Harsh Sharma, Zhu Wang, and Boris Kramer. Hamiltonian operator inference: Physics-preserving learning of reduced-order models for canonical Hamiltonian systems. *Physica D: Nonlinear Phenomena*, 431:133122, 2022. [doi:10.1016/j.physd.2021.133122](https://doi.org/10.1016/j.physd.2021.133122).