Physics-Constrained Learning of PDE Systems
with Uncertainty Quantified Port-Hamiltonian Models

Kaiyuan Tan   KAIYUAN.TAN@VANDERBILT.EDU
Peilun Li   PEILUN.LI@VANDERBILT.EDU
Thomas Beckers   THOMAS.BECKERS@VANDERBILT.EDU
Department of Computer Science, Vanderbilt University, Nashville, TN 37212

Editors: A. Abate, M. Cannon, K. Margellos, A. Papachristodoulou

Abstract

Modeling the dynamics of flexible objects has become an emerging topic in the community as these objects become more present in many applications, e.g., soft robotics. Due to the properties of flexible materials, the movements of soft objects are often highly nonlinear and, thus, complex to predict. Data-driven approaches seem promising for modeling those complex dynamics but often neglect basic physical principles, which consequently makes them untrustworthy and limits generalization. To address this problem, we propose a physics-constrained learning method that combines powerful learning tools and reliable physical models. Our method leverages the data collected from observations by sending them into a Gaussian process that is physically constrained by a distributed Port-Hamiltonian model. Based on the Bayesian nature of the Gaussian process, we not only learn the dynamics of the system, but also enable uncertainty quantification. Furthermore, the proposed approach preserves the compositional nature of Port-Hamiltonian systems.

Keywords: physics-constrained learning, scientific machine learning, Gaussian processes

1. Introduction

Modeling the dynamics of partial differential equation (PDE) systems plays a pivotal role in making safe and efficient control decisions, as they can forecast future states of the desired systems. The applicability of such models extends across various domains, including but not limited to flexible robots (Zheng and Lin, 2022), heat transfer processes (Zobeiry and Humfeld, 2021), and electrical systems (Cieślik, 2021). However, accurately modeling nonlinear PDE systems remains challenging, particularly due to their inherent complexity. A salient aspect of this complexity is the intricate modeling of nonlinearities within these systems, a task that is challenging and critical for accurate models. Traditional first-principle modeling methods often struggle with highly nonlinear effects, leading to solutions that may not fully capture the underlying physical reality. To address the lack of knowledge of complex systems, data-driven methods have been proposed to address state forecasting tasks, e.g., Long et al. (2018, 2019); Bar-Sinai et al. (2019); Wu and Xiu (2020); Stephany and Earls (2022). However, those pure data-driven learning approaches pose new questions regarding the efficiency, safety, and physical correctness of learned models (Hou and Wang, 2013). Due to the neglect of physical priors, the pure-learning approach suffers from many disadvantages, e.g., limited trustworthiness (Feng and Lavaei, 2021). Hence, learning methods that are both trustworthy and powerful gradually arouse researchers' attention. This paper proposes a novel method that leverages the principles of physically constrained learning, aiming to infuse the model with a deeper under-

standing of the physical laws governing the system. Unlike conventional pure-learning approaches, which may overlook the conservation laws, our method utilizes distributed Port-Hamiltonian systems (dPHS) as a framework to incorporate these physical priors as inductive bias directly into the learning process.

The Port-Hamiltonian framework is especially suited for modeling physical systems (Nageshrao et al., 2015), as it provides a systematic representation of the energetic and dissipative characteristics intrinsic to the system under consideration. Further, Port-Hamiltonian systems have many beneficial properties such as being suitable for modeling multi-physics, inherently respecting energy conservation/dissipation, and being closed under composition. As an extension for PDEs, distributed Port-Hamiltonian systems have been introduced that enable representing a class of PDEs within the Port-Hamiltonian framework. The term distributed refers to the capability of these systems to operate across different variable domains, such as temporal and spatial. We capitalize on the data obtained from the observations of the PDE system. This data is then subsequently assimilated into a Gaussian Process, which aims to learn the Hamiltonian function of the system, thereby constructing a Bayesian data-driven distributed Port-Hamiltonian framework tailored for the PDE, see Fig. 1. Incorporating these physical constraints, our methodology not only conforms to established physical laws but also amplifies the model's capacity for generalization from sparse data, which enhances the desired learning efficiency. Furthermore, our proposed model incorporates mechanisms for uncertainty quantification. This feature is pivotal in ensuring the reliability of the model and can be exploited for safe model-based control approaches. In summary, we propose a physics-constrained learning approach that combines dPHS and Gaussian processes (GP-dPHS) to enable physically
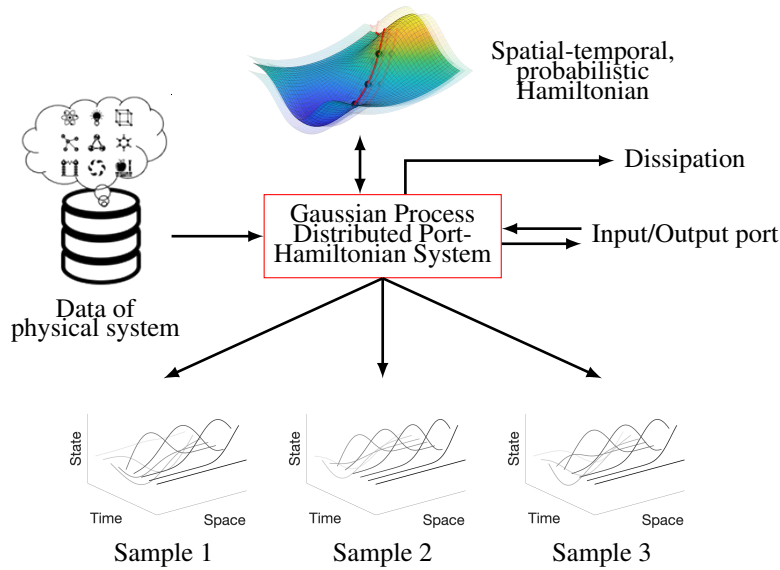

Figure 1: Gaussian Process Distributed Port-Hamiltonian (GP-dPHS) systems allow to learn a probabilistic spatial-temporal Hamiltonian based on collected data. The Port-Hamiltonian structure enables the inclusion of control inputs as well as dissipation. All sampled models of the GP-dPHS model are physically correct in terms of the evolution of the total energy in the system.

correct modeling for a class of PDEs with uncertainty quantification based on data. **Related Work:** Prior research has explored the realm of learning the characteristics of Partial Differential Equation (PDE) systems, as evidenced by works such as Raissi et al. (2019), Long et al. (2018) and Chen et al. (2021). However, these methodologies are predominantly learning-centric and tend to overlook the inherent physical properties of the systems in question. Reliably forecasting the evolution of the physics system requires uncertainty quantification due to the properties of physics, data, and the model itself. However, the approaches mentioned above ignore this point by offering merely predictions. Studies detailed in Greydanus et al. (2019) and Bertalan et al. (2019) have initiated the incorporation of the Hamiltonian framework to align the learning process with physical laws, a concept further extended to the domain of Port-Hamiltonian systems as demonstrated in Desai et al. (2021) and Nageshrao et al. (2015). Despite these advancements, a notable deficiency within these methodologies is the absence of uncertainty quantification, a metric vital for assessing the learning's reliability. The Gaussian process (GP), distinguished by its Bayesian underpinnings, has been introduced to the learning of generating functions within Hamiltonian systems as delineated in Karniadakis et al. (2021), and has been applied to Port-Hamiltonian Systems in Beckers (2023). These approaches, however, are not able to accommodate the learning of PDE systems, signifying a gap in the current state-of-the-art that presents an opportunity for significant contributions to the field.

Contribution: We propose a physics-constrained learning approach, GP-dPHS, that combines dPHS and Gaussian processes. This approach is beneficial in multiple ways: i) it enables physically correct modeling for a class of PDEs in terms of the evolution of energy, ii) it is highly expressive due to the nonparametric nature of the GP and iii) it allows us to quantify the uncertainty of the model. The proposed model is in particular suitable for modeling low-dimensional but highly nonlinear PDEs and opens the door for the design of robust controllers based on the provided uncertainty of the proposed model. The detailed reasoning can be found in the following parts.

This paper details the theoretical underpinnings of our approach in 2, elucidates the modeling process in 3, and presents empirical results in 4 that showcase the efficacy of incorporating physical priors in learning nonlinear PDE systems.

2. Background and Problem Formulation

In this section, we briefly describe the class of distributed Port-Hamiltonian systems, revisit GP regression, and introduce the problem setting.

2.1. Distributed Port-Hamiltonian Systems

Composing Hamiltonian systems with input/output ports leads to a so-called Port-Hamiltonian system, which is a dynamical system with ports that specify the interactions of its components. In Macchelli et al. (2004), the classical finite-dimensional Port-Hamiltonian formulation of a dynamical system is generalized in order to include the distributed parameter and multivariable case. In contrast to finite-dimensional PHS, the interconnection, damping, and input/output matrices are replaced by matrix differential operators that are assumed to be constant, i.e., no explicit dependence on the state (energy) variables is considered. As in finite dimensions, given the Stokes–Dirac structure, the model of the system easily follows once the Hamiltonian function is specified. The resulting class of infinite-dimensional systems in Port-Hamiltonian form is quite general, thus allowing the

interpretation of many classical PDEs within this framework and the description of several physical phenomena, such as the heat conduction, piezoelectricity, and elasticity. In the following, we recall the definition of dPHS from Macchelli et al. (2004).

Denote by $\mathcal{Z}$ a compact subset of $\mathbb{R}^n$ representing the spatial domain of the distributed parameter system. Then, denote by $\mathcal{U}$ and $\mathcal{V}$ two sets of smooth functions from $\mathcal{Z}$ to $\mathbb{R}^{n_u}$ and $\mathbb{R}^{n_v}$, respectively.

Definition 1 *A constant matrix differential operator of order N is a map L from $\mathcal{U} \to \mathcal{V}$ such that, given $u = (u^1, \ldots, u^{n_u}) \in \mathcal{U}$ and $v = (v^1, \ldots, v^{n_v}) \in \mathcal{V}$*

$$v = Lu \iff v^b := \sum_{\#\alpha=0}^{N} P_{a,b}^{\alpha} D^{\alpha} u^{\alpha} \tag{1}$$

where $\alpha := \{\alpha_1, \ldots, \alpha_d\}$ is a multi-index of order $\#\alpha := \sum_{i=1}^{d} \alpha_i$, P^{α} are a set of constant $n_u \times n_v$ matrices and $D^{\alpha} := \partial_{z_1}^{\alpha_1} \cdots \partial_{z_d}^{\alpha_d}$ is an operator resulting from a combination of spatial derivatives. Note that, in (6), the sum is intended over all the possible multi-indices α with order 0 to N and implicitly on a from 1 to q.

Definition 2 *Denote by J a constant matrix differential operator. Then, J is skew-adjoint if and only if its formal adjoint J^* is the map from $\mathcal{V}$ to $\mathcal{U}$ such that*

$$u = -J^*v \iff u^b := -\sum_{\#\alpha=0}^{N} (-1)^{\#\alpha} P_{a,b}^{\alpha} D^{\alpha} v^{\alpha}. \tag{2}$$

Definition 3 *Let $\delta_{\boldsymbol{x}}$ denotes the functional derivative operator such that*

$$\int \delta_x F(\boldsymbol{z}) \phi(\boldsymbol{z}) \, d\boldsymbol{z} = \lim_{\varepsilon \to 0} \frac{F[\boldsymbol{x} + \varepsilon\phi] - F[\boldsymbol{x}]}{\varepsilon} = \left[\frac{d}{d\varepsilon} F[\boldsymbol{x} + \varepsilon\phi] \right]_{\varepsilon=0}, \tag{3}$$

where ϕ is an arbitrary function.

Definition 4 *Let $\mathcal{X}$ be the space of vector-valued smooth functions on $[0, \infty] \times \mathcal{Z}$ (energy configurations), and $\mathcal{F}_d$ be the space of vector value smooth functions on $\mathcal{Z}$ (distributed flows). Assume that $\mathcal{E}_d \equiv \mathcal{F}_d$ is its dual (distributed efforts). Moreover, denote by J a skew-adjoint constant differential operator, and by G_d a constant differential operator.*

Let $\mathcal{H} \colon \mathcal{X} \to \mathbb{R}$ be the Hamiltonian functional such that

$$\mathcal{H}(\boldsymbol{x}) = \int_{\mathcal{Z}} H(z, x) dV, \tag{4}$$

where $H \colon \mathcal{Z} \times \mathcal{X} \to \mathbb{R}$ is the energy density. Denote by $\mathcal{W}$ the space of vector-valued smooth functions on $\partial\mathcal{Z}$ representing the boundary terms $\mathcal{W} := \{w | w = B_{\mathcal{Z}}(\delta_{\boldsymbol{x}}\mathcal{H}, \boldsymbol{u})\}$ defined by the boundary operator $B_{\mathcal{Z}}$. Then, the general formulation of a multivariable dPHS Σ with constant

Stokes–Dirac structure is fully described by[1]

$$\Sigma(J, R, \mathcal{H}, G) = \begin{cases} \frac{\partial \boldsymbol{x}}{\partial t} = (J - R)\delta_{\boldsymbol{x}}\mathcal{H} + G_d\boldsymbol{u} \\ \boldsymbol{y} = G_d^*\delta_{\boldsymbol{x}}\mathcal{H} \\ w = B_{\mathcal{Z}}(\delta_{\boldsymbol{x}}\mathcal{H}, \boldsymbol{u}), \end{cases} \tag{5}$$

where R is a constant differential operator taking into account energy dissipation and $(\boldsymbol{u}, \boldsymbol{y}) \in \mathcal{F}_d \times \mathcal{E}_d$. With the state $\boldsymbol{x}(t, \boldsymbol{z}) \in \mathbb{R}^n$ (also called energy variable) at time $t \in \mathbb{R}_{\geq 0}$ location $\boldsymbol{z} \in \mathcal{Z}$ and the I/O ports $\boldsymbol{u}, \boldsymbol{y} \in \mathbb{R}^m$.

Remark 1 *We focus here on the multivariate distributed Port-Hamiltonian system with constant Stokes–Dirac structure.*

Generally speaking, the interconnection of the elements in the dPHS is defined by the J matrix, whereas the Hamiltonian H characterizes their dynamical behavior. In PDE systems, the constitution of the J matrix predominantly involves partial differential operators. The port variables $\boldsymbol{u}$ and $\boldsymbol{y}$ are conjugate variables in the sense that their duality product defines the energy flows exchanged with the environment of the system, for instance, currents and voltages in electrical circuits or forces and velocities in mechanical systems, see Van der Schaft (2000) for more information.

2.2. Gaussian Process Regression

The Gaussian Process model is a nonparametric model widely used in machine learning to perform regression and prediction tasks, see Williams and Rasmussen (1995). Compared to other learning methods in regression tasks, GPs enable a natural quantification of uncertainty instead of just point estimates. The observations of the system are usually limited in practice. However, GPs can provide good predictions with relatively small data sets by leveraging the structure of the data through the kernel.

A GP can be fully defined by giving a mean function m and a covariance function k. For each pair of observation $\boldsymbol{x}, \boldsymbol{x}'$, the kernel function $k(\boldsymbol{x}, \boldsymbol{x})$ describes the covariance between these points. Before observing any data, we define a prior distribution over functions. For a set of input points, the prior is a multivariate (zero-mean) Gaussian process: $f \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$. Let $\{X = \boldsymbol{x}^{\{1\}}, \boldsymbol{x}^{\{2\}}, ..., \boldsymbol{x}^{\{N\}}\}$ denotes the input data set, together with $\{Y = \boldsymbol{y}^{\{1\}}, \boldsymbol{y}^{\{2\}}, ..., \boldsymbol{y}^{\{N\}}\}$, the output data set, to compose the training data $\mathcal{D}$ with N data points. The output data might be corrupted by zero-mean Gaussian noise with variance σ_n^2. After obtaining the data $\mathcal{D}$, we update our belief about the function f using the Bayes' Theorem to get the posterior distribution given a test input x^*. The posterior is again a Gaussian process with updated mean and covariance

$$\mu(y^*|\boldsymbol{x}^*) = k(\boldsymbol{x}^*, X)[K(X, X) + \sigma_n^2 I]^{-1}Y \tag{6}$$

$$\Sigma(y^*|\boldsymbol{x}^*) = k(\boldsymbol{x}^*, \boldsymbol{x}^*) - K(\boldsymbol{x}^*, X)[K(X, X) + \sigma_n^2 I]^{-1}K(\boldsymbol{x}^*, X). \tag{7}$$

---

1. Vectors $\boldsymbol{a}$ and vector-valued functions $\boldsymbol{f}(\cdot)$ are denoted with bold characters. Matrices are described with capital letters. I_n is the n-dimensional identity matrix and 0_n the zero matrix. The expression $A_{:,i}$ denotes the i-th column of A. For a positive semidefinite matrix Λ, $\|x - y\|_\Lambda^2 = (x - y)^\top \Lambda(x - y)$. $\mathbb{R}_{>0}$ denotes the set of positive real number whereas $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers. $\mathcal{C}^i$ denotes the class of i-th times differentiable functions. The operator $\nabla_{\boldsymbol{x}}$ with $\boldsymbol{x} \in \mathbb{R}^n$ denotes $[\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}]^\top$

In the formulation, $K(\boldsymbol{x}^*, X)$ is the covariance between the new point and the training points, and $k(\boldsymbol{x}^*, \boldsymbol{x}^*)$ is the prior variance at the new point. The predictive distribution $p(y^* | \boldsymbol{x}^*, X, Y)$ provides both a mean estimate and a measure of uncertainty for the predictions at a new point $\boldsymbol{x}^*$, which is the essence of uncertainty quantification in the GP regression learning model.

2.3. Problem Setting

We are considering building a physical-constrained learning system for a class of PDE systems. For many physical systems, their dynamics can be written in distributed Port-Hamiltonian form, see (5). In our case, the evolution of the states of a PDE system $\boldsymbol{x}(t, z) \in \mathbb{R}^n$ over time $t \in \mathbb{R}^n_{\geq 0}$ and spatial domain $z \in \mathcal{Z}$ should be defined by the equation with initial state $\boldsymbol{x}(0, z)$ as described in Section 2.1. The Hamiltonian functional $\mathcal{H} \in \mathcal{C}^\infty$ is assumed to be *completely unknown* due to unstructured uncertainties in the system imposed by, e.g., nonlinear stress/strain curve in soft materials, physical coupling effects, or highly nonlinear electrical and magnetic fields. Thus, we aim to learn a dPHS model

$$\frac{\partial \boldsymbol{x}}{\partial t} = (J - R)\delta_{\boldsymbol{x}}\hat{\mathcal{H}} + G_d \boldsymbol{u} \tag{8}$$

with an estimated Hamiltonian functional $\hat{\mathcal{H}}$ based on observations of the system (5). To address the problem, we need to introduce the following assumptions.

Assumption 1 *The structure of the system matrices J, R, and G are assumed to be known, with the exception of a finite set of parameters $\Theta \subset \mathbb{R}^{n_\Theta}$, where n_Θ is the total number of unknown parameters inside of these matrices.*

Assumption 2 *We can observe the state of the PDE system (5) at some temporal t_i and spatial points $\boldsymbol{z}_j$ to generate a set of observations $\{\boldsymbol{x}(t_1, \boldsymbol{z}_1), \ldots, x(t_1, \boldsymbol{z}_{N_z}) \ldots, \boldsymbol{x}(t_{N_t}, \boldsymbol{z}_1), \ldots, x(t_{N_t}, \boldsymbol{z}_{N_z})\}$.*

Assumption 3 *There exists a unique and smooth solution $\boldsymbol{x}(t, z)$ for the PDE system under the known boundary conditions w (5).*

Assumption 1 is only mildly restrictive since J and G are typically matrix differential operators, given by prior knowledge of the PDE system to be learned. Even though we imply knowing the dissipation matrix R, we allow unknown parameters such that only the general structure needs to be known, e.g., the friction model but not the parameters. Assumption 2 ensures that we can collect data from the PDE system, requiring observability of the state. If that is not the case, an observer needs to be implemented. Finally, Assumption 3 guarantees the existence of a solution so that the derivation of a model is meaningful.

3. Gaussian Process dPHS

In this section, we propose the Gaussian Process Distributed Port-Hamiltonian system (GP-dPHS). The objective of the proposed method is to develop a data-driven model for the learning of low-dimensional but nonlinear PDE systems, which is characterized as a distributed Port-Hamiltonian system with a structure as delineated in (5). The main challenge of extending this approach to high-dimensional problems is that the later represented kernel function leads to a Gram matrix $\boldsymbol{K}$, whose

size grows with an increase of the dimension of system state $\boldsymbol{x}$. Thus, calculating the inverse of the covariance matrix would be relatively expensive. First, we create a dataset that consists of state and state derivatives of the PDE dynamics. On the basis of this data set, we can train a GP model with a physics-constrained kernel function to learn the PDE dynamics. Finally, we use the model to predict future states.

3.1. GP-dPHS data generation

As stated in the problem setting, we consider the observations as follows

$$\mathcal{D} = \{t_i, \boldsymbol{z}_j, \boldsymbol{x}(t_i, \boldsymbol{z}_j), \boldsymbol{u}(t_i)\}_{i=1,j=1}^{i=N_t,j=N_z}, \tag{9}$$

where $\boldsymbol{x}(t_i, \boldsymbol{z}_j)$ encapsulates the observed state at time t_i and spatial point $\boldsymbol{z}_j$ for N_t time steps and N_z spatial points, corresponding to the input sequence $\{\boldsymbol{u}(t_1), \boldsymbol{u}(t_2), ..., \boldsymbol{u}(t_{N_t})\}$. With this time-sequenced observational data $\mathcal{D}$ at our disposal, the first step in modeling the PDE system through the GP-dPHS is the procurement of the system's state derivatives $\frac{\partial \boldsymbol{x}}{\partial t}$. Furthermore, we typically have only a small set of spatial observation points N_z due to technical limitations. Referencing the equation detailed in (3), and employing a smooth kernel k such as the squared exponential kernel, our approach is to construct n distinct GPs for the training sets $\mathcal{D}$. To expound for each l-th dimension of the state vector, we have a training set $\mathcal{D}_l = \{t_i, \boldsymbol{z}_j, x_l(t_i, \boldsymbol{z}_j), \boldsymbol{u}(t_i)\}_{i=1,j=1}^{i=N_t,j=N_z}$, where $l = 1, \ldots, n$. Based on these data sets, we train n separate GP models with $\{t_i, \boldsymbol{z}_j\}_{i=1,j=1}^{i=N_t,j=N_e}$ as input data and $\{x_l(t_i, \boldsymbol{z}_j)\}_{i=1,j=1}^{i=N_t,j=N_z}$ as output data.

By differentiating the kernel function k, we can then acquire the desired derivative $\frac{\partial x_l}{\partial t}$ of the corresponding states x_l. The properties of $\frac{\partial x_l}{\partial t}$ can be fully described by the GP mean function $\mu(\frac{\partial x_l}{\partial t}|t, \mathcal{D}_l)$ and the variance function $\Sigma(\frac{\partial x_l}{\partial t}|t, \mathcal{D}_l)$. Thus, we can obtain a new dataset $\mathcal{E}$, which contains the states $X = [\tilde{\boldsymbol{x}}(t_1), \ldots, \tilde{\boldsymbol{x}}(t_{N_t})]$ and the state derivative $\dot{X} = [\frac{\partial \tilde{\boldsymbol{x}}(t_1)}{\partial t}, \ldots, \frac{\partial \tilde{\boldsymbol{x}}(t_{N_t})}{\partial t}]$, where $\tilde{\boldsymbol{x}}(t_i) = [\boldsymbol{x}(t_i, \boldsymbol{z}_1)^\top, \ldots, \boldsymbol{x}(t_i, \boldsymbol{z}_{N_e})^\top]^\top$ denotes the spatially stacked state at time t_i. Note that we upsampled the number of spatial points from N_z to $N_e >> N_z$ by evaluating the GP models on new spatial points. We visualize this step in Fig. 2.
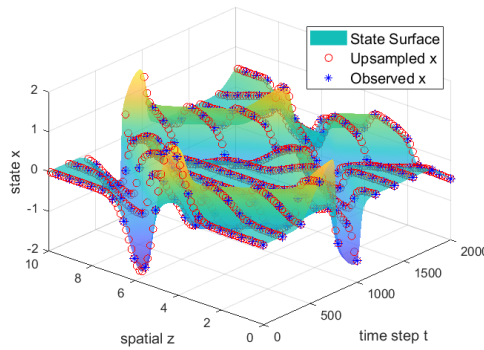

Figure 2: Example of states observations and the upsampling process.

3.2. Modeling

After the generation of the data, we now introduce the GP-dPHS model. The central concept underpinning this approach involves the employment of a Gaussian process to approximate the un-

known Hamiltonian functional while simultaneously treating the parametric indeterminacies within the matrices J, R, and G as hyperparameters. This requirement is accommodated by exploiting the characteristic that GPs are invariant under linear transformations (Jidling et al., 2017), which allows inclusion of the Hamiltonian functional derivative within the GP formulation. The dynamical behavior of the system is thus expressed as a linear transformation of the Hamiltonian within the Port-Hamiltonian (PH) model. This integration of the PHS dynamics into the GP is mathematically represented as

$$\frac{\partial \boldsymbol{x}}{\partial t} \sim \mathcal{GP}(\hat{G}_\Theta \boldsymbol{u}, k_{dphs}(\boldsymbol{x}, \boldsymbol{x}')), \tag{10}$$

where the system dynamics are encapsulated within the probabilistic model. Specifically, the new kernel function k_{dphs} can be formally described as

$$k_{dphs}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2(\hat{J}_\Theta - \hat{R}_\Theta)\delta_{\boldsymbol{x}} \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{x}'||^2}{2\varphi_l^2}\right) \delta_{\boldsymbol{x}'}^\top(\hat{J}_\Theta - \hat{R}_\Theta)^\top. \tag{11}$$

This kernel function is based on the squared exponential function as the Hamiltonian functional is a smooth function. The model can then be trained as a standard GP with the dataset $\mathcal{E}$ (10) from the previous section. The matrices J, R and G of the dPHS system are inferred through the estimate $\hat{J}_\Theta, \hat{R}_\Theta, \hat{G}_\Theta$, where the unknown parameters given by Θ are treated as hyperparameters as introduced in Section 2.3. The unknown hyperparameters $\Theta, \varphi_l, \sigma_f$ are minimized by means of the negative log marginal likelihood. In this way, we complete the training of the proposed GP-dPHS. By exploiting again the invariance under linear transformation, we can rewrite (10) as

$$\frac{\partial}{\partial t}\boldsymbol{x}(t, \boldsymbol{z}) = (\hat{J}_\Theta - \hat{R}_\Theta)\delta_{\boldsymbol{x}}\hat{\mathcal{H}} + \hat{G}_\Theta \boldsymbol{u}, \tag{12}$$

where the Hamiltonian $\hat{\mathcal{H}}$ follows a GP. It can be seen that the dynamics describes a distribution over distributed Port-Hamiltonian systems. Analogously to the approach delineated in (5), our strategy uses data-driven techniques to infer the characteristics of the distributed physical system while rigorously maintaining compliance with intrinsic physical constraints. This multifaceted learning is integrated to construct our Gaussian Process distributed Port-Hamiltonian systems, which embodies a comprehensive model encapsulating the dynamical interplay within diverse domains.

3.3. Prediction

In pursuit of generating samples from the posterior distribution, we employ the joint distribution with mean-adjusted output data at test states $\boldsymbol{x}^* \in \mathbb{R}^n$

$$\begin{bmatrix} X \\ \boldsymbol{f}(x^*) \end{bmatrix} = \mathcal{N}\left(\hat{G}_\Theta \boldsymbol{u}, \begin{bmatrix} K_{dphs}(X, X) & K_{dphs}(X, \boldsymbol{x}^*) \\ K_{dphs}(X, \boldsymbol{x}^*)^T & K_{dphs}(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix}\right), \tag{13}$$

which facilitates the procurement of a vector field $\boldsymbol{f}$ that represents the right-hand side of (12). From this distribution, we could sample a $\boldsymbol{f}$, i.e., a sample of the learned dPHS system, and solve the resulting PDE with a numerical solver. However, the demands of numerical integration necessitate the ability to access this vector field at various points, which is computationally expensive (Beckers and Hirche, 2022). Instead, the approach advocates to sample the estimated Hamiltonian functional

variable named $\hat{\mathcal{H}}$, which leads to (12). Thus, we sample a deterministic Hamiltonian $\hat{\mathcal{H}}$ from the GP distribution and use a numerical PDE solver to compute a solution for (12). Since a sample from GP is the deterministic function and the dPHS is designed to follow 5, it is fully compliant with the Stokes-Dirac structure. This procedure guarantees that the solution of the learned dynamics is physically correct in terms of the energy evolution, since it has to satisfy the dPHS form.

4. Numerical evaluation

For evaluating the proposed physics-constrained learning methodology, we aim to learn the dynamics of the lateral vibration of a string with a nonlinear stress-strain curve. This system is succinctly represented by the ensuing general equation

$$\frac{\partial^2 x}{\partial t^2} - s\left(\frac{\partial x}{\partial z}\right)\frac{\partial^2 x}{\partial z^2} = 0, \tag{14}$$

where s denotes the stress/strain curve that is, for soft materials, a highly nonlinear function of the stress $\frac{\partial x}{\partial z}$. In our example, we assume that s follows a sigmoid function. To use our proposed GP-dPHS model, we need to translate the system into dPHS form, where the functional Hamiltonian $\mathcal{H}$ is assumed to be unknown. Recalling equation (5), it is postulated that the constant matrix differential operators $\mathcal{J}$, $\mathcal{R}$, and $\mathcal{G}$ are structurally known differential operators. In our setting, we are examining a string system with both ends fixed and the spatial domain $z \in [0, 10]$. For the sake of simplicity, we assume no input and imply a nontrivial initial condition of the system to generate a meaningful dataset. We set the damping factor to 0.01. Then, the desired system can be modeled via a dPHS in the following form

$$\frac{\partial}{\partial t}\begin{bmatrix} p(t,z) \\ q(t,z) \end{bmatrix} = \underbrace{\begin{bmatrix} -0.01 & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial z} & 0 \end{bmatrix}}_{J-R} \delta_{\boldsymbol{x}}\mathcal{H}, \tag{15}$$

where $p = \frac{\partial x}{\partial t}$ and $q = \frac{\partial x}{\partial z}$. This transformation is necessary to ensure that J is an adjoint operator as defined in Definition 2. The Hamiltonian is given by $\mathcal{H} = \int_{\mathcal{Z}}\int s(q)q(t,z)dq + p(t,z)^2 dz$. We refer to (15) as the ground truth.

Following the problem statement, we assume that the structure of (15) is known, but the Hamiltonian is unknown to us. This example demonstrates that this is a reasonable assumption as the nonlinearity of the system is encapsulated in the Hamiltonian. At the time point t_0, we set the initial condition to $x_0 = \exp(-(z-5)^2)$. We observe the ground truth system for 20 seconds with a time step of $\Delta t = 0.01$ seconds. The spatial domain is observed at $N_z = 8$ points. We upsample it to $N_e = 400$ points by evaluating the GP models on new spatial points. To accelerate the training, we flatten all the observations, apply a downsample rate = 50 to all the data, and train the proposed GP-dPHS model using the steps mentioned above. Then, we compare the GP-dPHS model with the ground truth for a new initial condition, i.e., $x_0 = \sin(\frac{10z}{\pi})$. The movements of the string are visualized in Fig. 3. It can be seen that the mean prediction of our proposed GP-dPHS model accurately mimics the ground truth. Figure 4 shows the nonlinear ground truth and the learned derivative of the Hamiltonian $\mathcal{H}$. As shown in the figure, our proposed method learned the desired Hamiltonian surface, which is very close to the ground truth. Also, we visualize 95% confidence interval guarantees of uncertainty in the same figure to show that the proposed method offers confidence intervals for uncertainty quantification instead of giving a deterministic prediction that can be inaccurate,
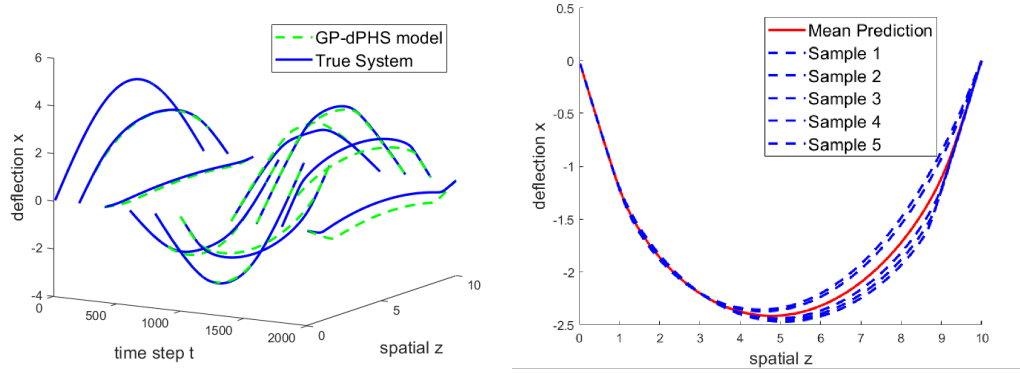
Figure 3: Graphical illustration of string PDE system. Observe on the left figure of each time slice; the deflection learned by GP-dPHS (cyan dotted) is close to the ground-truth trajectory on each time step t. The mean prediction of time slice $t = 600$ and five samples of the GP-dPHS model are shown on the right.

As a physics-constrained method, we expect that the model generalizes well even for small datasets. The following figure on the right side shows that, using very limited observations, our proposed method is capable of learning a relatively accurate physics-constrained model. Note, that independently of the number of data points, the GP-dPHS model always leads to physically correct solutions.
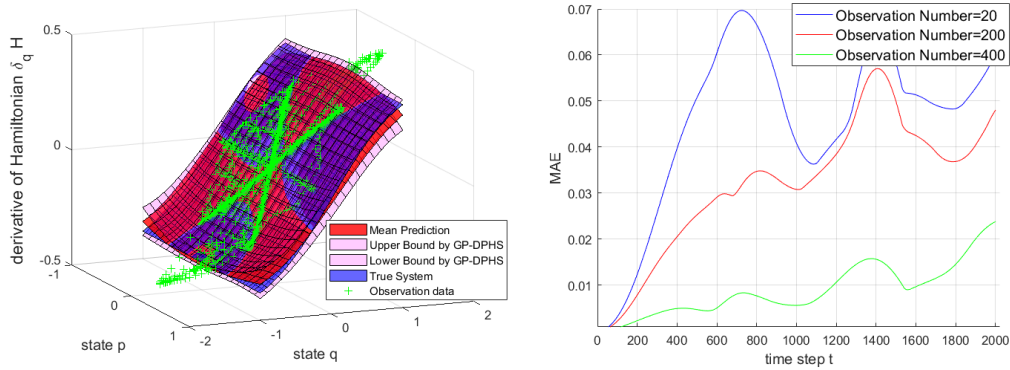

Figure 4: The left figure visualizes the nonlinear Hamiltonian derivative $\delta_q H$. The learned surface (red) is very close to the true surface (blue). The pink surfaces visualize the uncertainty of the Hamiltonian derivative under a 95% prediction uncertainty guarantee. The figure on the right shows the model error for different numbers of training points

5. Conclusion

In this paper, we propose GP-dPHS, a physics-constrained data-driven method for learning a class of PDE systems based on data. Our approach guarantees the physical correctness of the model, generalizes well for small datasets, and provides uncertainty quantification. In a numerical example, we demonstrate the proposed approach for a simple nonlinear PDE problem. In the future, we will take advantage of the dPHS framework to design a physics-based controller for PDE systems.

References

Yohai Bar-Sinai, Stephan Hoyer, Jason Hickey, and Michael P Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, 116(31):15344–15349, 2019.

Thomas Beckers. Data-driven bayesian control of port-Hamiltonian systems. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 8708–8713. IEEE, 2023.

Thomas Beckers and Sandra Hirche. Prediction with approximated Gaussian process dynamical models. *IEEE Transactions on Automatic Control*, 67(12):6460–6473, 2022. doi: 10.1109/TAC. 2021.3131988.

Tom Bertalan, Felix Dietrich, Igor Mezić, and Ioannis G Kevrekidis. On learning Hamiltonian systems from data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12):121107, 2019.

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.

Sławomir Cieślik. Mathematical modeling of the dynamics of linear electrical systems with parallel calculations. *Energies*, 14(10):2930, 2021.

Shaan A Desai, Marios Mattheakis, David Sondak, Pavlos Protopapas, and Stephen J Roberts. Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems. *Physical Review E*, 104(3):034312, 2021.

Han Feng and Javad Lavaei. Learning of dynamical systems under adversarial attacks. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 3010–3017. IEEE, 2021.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Zhong-Sheng Hou and Zhuo Wang. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235:3–35, 2013.

Carl Jidling, Niklas Wahlström, Adrian Wills, and Thomas B Schön. Linearly constrained Gaussian processes. *Advances in Neural Information Processing Systems*, 30, 2017.

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, 2021.

Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. PDE-net: Learning PDEs from data. In *International conference on machine learning*, pages 3208–3216. PMLR, 2018.

Zichao Long, Yiping Lu, and Bin Dong. Pde-net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399:108925, 2019.

Alessandro Macchelli, Arjan J Van Der Schaft, and Claudio Melchiorri. Port Hamiltonian formulation of infinite dimensional systems i. modeling. In *2004 43rd IEEE Conference on Decision and Control (CDC)*, volume 4, pages 3762–3767. IEEE, 2004.

Subramanya P Nageshrao, Gabriel AD Lopes, Dimitri Jeltsema, and Robert Babuška. Port-Hamiltonian systems in adaptive and learning control: A survey. *IEEE Transactions on Automatic Control*, 61(5):1223–1238, 2015.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Robert Stephany and Christopher Earls. PDE-read: Human-readable partial differential equation discovery using deep learning. *Neural Networks*, 154:360–382, 2022.

Arjan Van der Schaft. *L2-gain and passivity techniques in nonlinear control*. Springer, 2000.

Christopher Williams and Carl Rasmussen. Gaussian processes for regression. *Advances in neural information processing systems*, 8, 1995.

Kailiang Wu and Dongbin Xiu. Data-driven deep learning of partial differential equations in modal space. *Journal of Computational Physics*, 408:109307, 2020.

Tongjia Zheng and Hai Lin. PDE-based dynamic control and estimation of soft robotic arms. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 2702–2707. IEEE, 2022.

Navid Zobeiry and Keith D Humfeld. A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. *Engineering Applications of Artificial Intelligence*, 101:104232, 2021.