# LEARNING HAMILTONIAN DYNAMICS UNDER UNCERTAINTY VIA SYMPLECTIC GAUSSIAN PROCESSES

**Yusuke Tanaka** *
NTT Communication Science Laboratories
Kyoto, Japan
ysk.tanaka@ntt.com

**ABSTRACT**

This paper focuses on a research question: *How do we infer Hamiltonian vector fields under uncertainty from data?* One promising approach is a Gaussian process (GP) that can consider the uncertainty involved in data. Several GP models have recently incorporated the symplectic structure, allowing the accurate prediction of Hamiltonian vector fields; however, the predictive variance has not been adequately analyzed. This paper experimentally confirms that GP models encoding the symplectic structure can appropriately estimate predictive variances for Hamiltonian vector fields. Also, the experiments demonstrate the robustness of the GP models in simulating Hamiltonian systems in out-of-distribution settings.

***Keywords*** Gaussian processes · uncertainty · Hamiltonian systems · symplectic structure

## 1 Introduction

Modeling, simulating, and forecasting physical phenomena are fundamental in engineering and physical sciences. There is great interest in a machine learning approach for inferring physical dynamics from data, which is advantageous in that it allows for obtaining physical dynamics without explicitly designing mathematical models. Many real-world systems can be described by ordinary differential equations (ODEs), the dynamics of which are defined by *vector fields* in a state space (also called a *phase space*). Many previous works have estimated ODE dynamics by modeling the vector fields using machine learning methods, such as neural networks [3]. The Hamiltonian neural network (HNN) [4] has been designed based on Hamilton's equations, allowing the learning of vector fields that satisfy energy conservation laws. This idea has also been introduced in probabilistic modeling, in which the Hamiltonian is modeled by Gaussian processes (GPs) [1, 6, 8]. The GP models allow one to learn Hamiltonian dynamics from noisy and sparse data.

One advantage of GP models is the ability to evaluate predictive variances [5]; however, the prior works have not thoroughly analyzed them in the experiments on Hamiltonian systems [1, 6, 8]. This paper provides a numerical analysis of predictive variances when inferring Hamiltonian vector fields via the GP models based on Hamiltonian mechanics. The experiments on two simple physical systems show that the GP models can 1) accurately predict Hamiltonian vector fields in low-data settings than HNN and 2) evaluate predictive variances, which appropriately supports prediction errors. In addition, we observe that the GP models improved the predictive performance, especially in out-of-distribution settings: Regions on the phase space where the training samples are extremely small. Thus, we consider the problem of simulating state trajectories in the out-of-distribution settings. These experiments demonstrate that GP models are more robust than HNN in such situations.

## 2 Symplectic Gaussian Processes

**Problem setting**. The continuous-time evolution of Hamiltonian systems is described in phase space, that is, the product space of generalized coordinates $\boldsymbol{x}^{\mathrm{q}} = (x_1^{\mathrm{q}}, \ldots, x_M^{\mathrm{q}})$ and generalized momenta $\boldsymbol{x}^{\mathrm{p}} = (x_1^{\mathrm{p}}, \ldots, x_M^{\mathrm{p}})$, where

*https://sites.google.com/view/yusuketanaka/english

$M$ is the degree of freedom. Let $\boldsymbol{x} = (\boldsymbol{x}^{\mathrm{q}}, \boldsymbol{x}^{\mathrm{p}}) \in \mathbb{R}^D$ be a state of the system, where $D = 2M$. Let $\dot{\boldsymbol{x}} \in \mathbb{R}^D$ denote a time-derivative of the state $\boldsymbol{x}$. We assume that the noisy observation $\dot{\boldsymbol{y}} \in \mathbb{R}^D$ of $\dot{\boldsymbol{x}}$ is available. Suppose that we have $N$ samples $\{(\boldsymbol{x}_n, \dot{\boldsymbol{y}}_n) \mid n = 1, \ldots, N\}$. Our aim is to learn a vector field $\boldsymbol{f}(\boldsymbol{x}) : \mathbb{R}^D \to \mathbb{R}^D$ that can predict the time-derivative for arbitrary state $\boldsymbol{x}$, as $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$.

**Model**. We briefly describe a Gaussian process (GP) model that is designed based on the theory of Hamiltonian mechanics [6, 1, 8], which we call a *symplectic Gaussian process (SGP)*. In the Hamiltonian formalism, the dynamics is totally defined by a *Hamiltonian* $H(\boldsymbol{x}) : \mathbb{R}^D \to \mathbb{R}$, which represents a total energy of the system. In SGP, the *unknown* Hamiltonian is assumed to be a single-output GP with zero mean. According to Hamilton's equations, the vector field $\boldsymbol{f}(\boldsymbol{x})$ is given by

$$\boldsymbol{f}(\boldsymbol{x}) = \mathcal{L}\, H(\boldsymbol{x}), \quad \text{where} \quad H(\boldsymbol{x}) \sim \mathcal{GP}(0, \gamma(\boldsymbol{x}, \boldsymbol{x}')). \tag{1}$$

Here, $\mathcal{L} := \mathbf{S}\nabla$ is a differential operator from Hamilton's equations, and $\gamma(\boldsymbol{x}, \boldsymbol{x}') : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is a covariance function. We can use any differentiable kernel such as the ARD (Automatic Relevance Determination) Gaussian kernel [5] as the covariance function. $\mathbf{S} \in \mathbb{R}^{D \times D}$ is the skew-symmetric matrix, represented by

$$\mathbf{S} = \begin{pmatrix} \mathbf{O} & \mathbf{I} \\ -\mathbf{I} & \mathbf{O} \end{pmatrix}, \tag{2}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{O}$ is the zero matrix. Since $\mathcal{L}$ is a linear operator, the derivative of a GP is again a GP [7]; thus, $\boldsymbol{f}(\boldsymbol{x})$ (1) can be written by a multi-output GP,

$$\boldsymbol{f}(\boldsymbol{x}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}(\boldsymbol{x}, \boldsymbol{x}')), \tag{3}$$

where $\mathbf{0}$ is a column vector of 0's, and $\mathbf{K}(\boldsymbol{x}, \boldsymbol{x}') : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^{D \times D}$ is the matrix-valued covariance function, represented by

$$\mathbf{K}(\boldsymbol{x}, \boldsymbol{x}') = \mathcal{L}\mathcal{L}^\top \gamma(\boldsymbol{x}, \boldsymbol{x}') = \mathbf{S}\nabla^2 \mathbf{S}^\top \gamma(\boldsymbol{x}, \boldsymbol{x}'). \tag{4}$$

Here, $\nabla^2$ is the Hessian operator. A realization of $\boldsymbol{f}(\boldsymbol{x})$ from (3) satisfies the symplectic structure important for modeling energy-conserving dynamics. We can use the SGP as the prior distribution to learn the Hamiltonian vector fields from data.

**Inference**. The posterior of SGP can be derived in a similar way to the vanilla GP. If the observation model of $\dot{\boldsymbol{y}}$ is a Gaussian distribution with a variance of $\sigma^2$, then the posterior $\boldsymbol{f}^*(\boldsymbol{x})$ is given by

$$\boldsymbol{f}^*(\boldsymbol{x}) \sim \mathcal{GP}\left(\boldsymbol{m}^*(\boldsymbol{x}), \mathbf{K}^*(\boldsymbol{x}, \boldsymbol{x}')\right), \tag{5}$$

where $\boldsymbol{m}^*(\boldsymbol{x}) : \mathbb{R}^D \to \mathbb{R}^D$ and $\mathbf{K}^*(\boldsymbol{x}, \boldsymbol{x}') : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^{D \times D}$ are the mean function and the covariance function for $\boldsymbol{f}^*(\boldsymbol{x})$, represented by

$$\boldsymbol{m}^*(\boldsymbol{x}) = \mathbf{K}(\boldsymbol{x}, \mathbf{X}) \left[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}\right]^{-1} \dot{\mathbf{Y}}, \tag{6}$$

$$\mathbf{K}^*(\boldsymbol{x}, \boldsymbol{x}') = \mathbf{K}(\boldsymbol{x}, \boldsymbol{x}') - \mathbf{K}(\boldsymbol{x}, \mathbf{X}) \left[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}\right]^{-1} \mathbf{K}(\mathbf{X}, \boldsymbol{x}'), \tag{7}$$

respectively. Here, $\dot{\mathbf{Y}} = (\dot{\boldsymbol{y}}_1, \ldots, \dot{\boldsymbol{y}}_N)^\top \in \mathbb{R}^{DN}$, and $\mathbf{K}(\boldsymbol{x}, \mathbf{X}) \in \mathbb{R}^{D \times DN}$ denotes the covariance matrix evaluated at all pairs of training points and an arbitrary test point $\boldsymbol{x}$, and similarly for the other matrices $\mathbf{K}(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{DN \times DN}$ and $\mathbf{K}(\mathbf{X}, \boldsymbol{x}') \in \mathbb{R}^{DN \times D}$, where $\mathbf{X} = \{\boldsymbol{x}_n \mid n = 1, \ldots, N\}$. One can predict the time-derivative for $\boldsymbol{x}$ using the posterior mean (6) and evaluate the predictive variance (7). The noise variance $\sigma^2$ and the hyperparameters of the kernel $\gamma(\boldsymbol{x}, \boldsymbol{x}')$ can be estimated by maximizing the marginal likelihood of $\dot{\mathbf{Y}}$, represented by $p(\dot{\mathbf{Y}}) = \mathcal{N}(\mathbf{0}, \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})$.

## 3 Experiments

**Data**. We compared the SGP with the HNN using two physical systems: *Spring* and *pendulum*. The Hamiltonian of the spring is $H(\boldsymbol{x}) = \frac{1}{2}k(x^{\mathrm{q}})^2 + \frac{(x^{\mathrm{p}})^2}{2m}$, where $k$ is the spring constant and $m$ is the mass constant. The Hamiltonian of the pendulum is $H(\boldsymbol{x}) = 2mgl(1 - \cos x^{\mathrm{q}}) + \frac{l^2 (x^{\mathrm{p}})^2}{2m}$, where we denote the gravitational constant by $g$ and the length of the pendulum by $l$. In the experiments, we set $g = 3$ and $k = m = l = 1$. We obtained $N = \{40, 60, 100\}$ samples of $(\boldsymbol{x}_n, \dot{\boldsymbol{y}}_n)$, where the states $\{\boldsymbol{x}_n\}$ were generated with total energies uniformly distributed across a predefined range $[1.3, 2.0]$. The observations $\{\dot{\boldsymbol{y}}_n\}$ were obtained by adding a Gaussian noise with variance $\sigma^2 = 0.1$ to the time-derivatives from the Hamilton's equation: $\dot{\boldsymbol{x}}_n = \mathcal{L}H(\boldsymbol{x}_n)$. We randomly split the samples and used 80% for training and 20% for validation. We conducted the experiments five times by resampling the training and validation sets.

**Task 1: Predicting vector fields.** In the test phase, we prepared $15 \times 15$ regular grid points $\{\boldsymbol{x}_i^* \mid i = 1, \ldots, I\}$ on a two-dimensional phase space defined by $[-3.2, 3.2]^2$, where $I$ is the number of test points.

Table 1: MSE and standard deviation for the predicted vector fields. Bold font indicates statistically significant differences between SGP and HNN (a paired t-test) at level of $P < 0.05$.

| | Spring | | | Pendulum | | |
|---|---|---|---|---|---|---|
| Sample size | 40 | 60 | 100 | 40 | 60 | 100 |
| HNN | 0.600±0.098 | 0.476±0.048 | 0.368±0.034 | 2.253±0.107 | 1.996±0.099 | 1.863±0.089 |
| SGP | **0.041±0.053** | **0.040±0.021** | **0.011±0.001** | **0.323±0.161** | **0.237±0.056** | **0.200±0.030** |

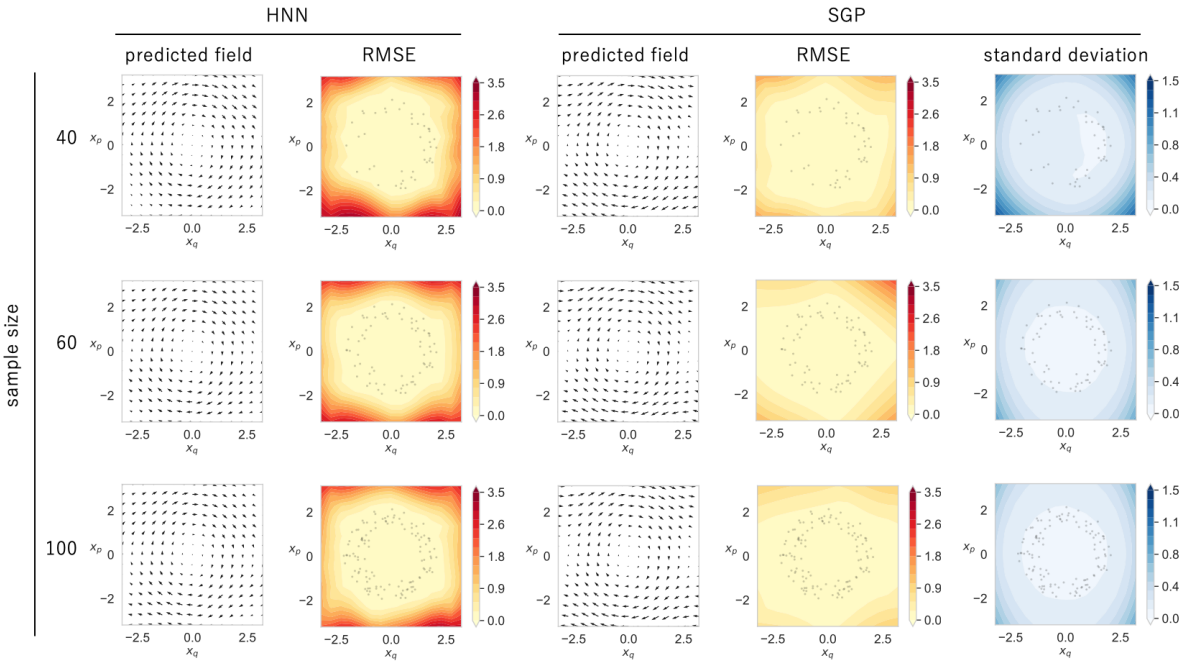



Figure 2: Prediction results of HNN and SGP. The gray dots represent training data points.

We evaluated performance at the test points by comparing the predicted vector fields $\{\dot{\boldsymbol{x}}_i^{\text{pred}}\}$ with the ground truth $\{\dot{\boldsymbol{x}}_i^{\text{true}}\}$. Figure 1 shows the ground truth for the pendulum. The evaluation metric is the mean squared error (MSE), $\frac{1}{I}\sum_{i=1}^{I}(\| \dot{\boldsymbol{x}}_i^{\text{pred}} - \dot{\boldsymbol{x}}_i^{\text{true}} \|^2)$, where $\| \cdot \|^2$ is the Euclidean norm. Table 1 shows the MSE and standard deviation of the vector fields predicted by HNN and SGP. This result shows that SGP yields a better prediction of vector fields than HNN in all the cases. As expected, SGP proves to be particularly effective when the sample size is small. Figure 2 shows the predicted vector fields by HNN and SGP, the respective root MSEs (RMSEs), and the standard deviation to the SGP prediction for the pendulum data. Comparing the true vector field (Figure 1) to its prediction (Figure 2) shows that SGP is more accurate than HNN, especially when the sample size is small. The standard deviation for the SGP prediction reflects the number of samples well; by comparing it to the RMSE, we can see that the standard deviation provides a reasonable estimate of the confidence in the prediction. Furthermore, and importantly, as shown in the RMSE in Figure 2, one observes that SGP has smaller errors than HNN even in regions where there are no training samples.

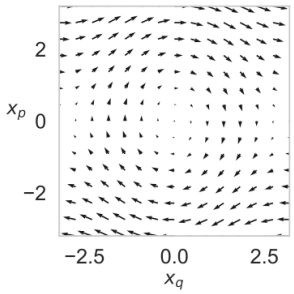


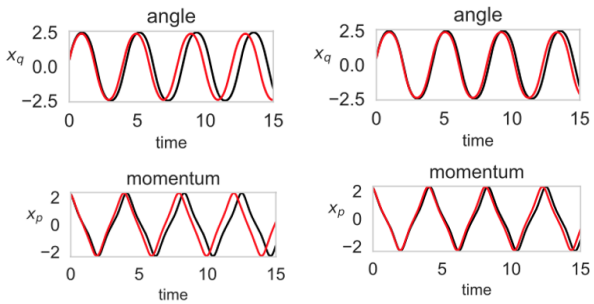
Figure 1: Pendulum dynamics.

**Task 2: Simulating trajectories in out-of-distribution settings.** Experimental results from Task 1 show that SGP can adequately predict dynamics even in regions on the phase space where the training samples are extremely small. For further discussion, we consider the problem of simulating trajectories in out-of-distribution settings. In the test phase, we sampled 25 initial conditions with total energies uniformly distributed across a range [2.0,2.3] different from training data. We compared true trajectories and their prediction based on the vector fields estimated by HNN and SGP, where we generated the trajectories by using a numerical integrator, i.e., the Dormand–Prince method with adaptive time-stepping, implemented in torchdiffeq [2, 3]. Each test trajectory was sampled at a frequency of 100 Hz for 15 seconds.

Table 2: MSE and standard deviation for the predicted trajectories. Bold font indicates statistically significant differences between SGP and HNN (a paired t-test) at level of $P < 0.05$.

| | Spring | | | Pendulum | | |
|---|---|---|---|---|---|---|
| Sample size | 40 | 60 | 100 | 40 | 60 | 100 |
| HNN | 2.183±1.119 | 1.151±1.047 | 0.399±0.072 | 1.473±0.422 | 1.331±0.276 | 0.334±0.074 |
| SGP | **0.302±0.414** | 0.165±0.043 | **0.127±0.072** | **0.651±0.384** | **0.269±0.062** | 0.216±0.024 |

Table 2 shows the MSE and standard deviation of the trajectories predicted by HNN and SGP. We can see that SGP dramatically improves performance than HNN when the number of training samples is small. Figure 3 shows the simulation results of the angle and momentum for the pendulum. The results show that the prediction accuracy of HNN degrades significantly over time, while SGP tracks the true trajectory well. Therefore, SGP is a more robust predictor in out-of-distribution settings than HNN.




(a) HNN (b) SGP

Figure 3: Simulation results for pendulum. The black and red lines are the true and predicted trajectories, respectively.

## 4 Discussion

We have reviewed symplectic Gaussian process (SGP) models that incorporate the theory of Hamiltonian mechanics and have evaluated the effectiveness of SGP experimentally. As also shown in the previous work [8], GP-based models can make accurate predictions even with noisy and sparse data. In addition, this paper has shown that the SGP can adequately estimate the uncertainty in the prediction. It is also interesting to note that the SGP is robust in out-of-distribution settings. There is still little research on modeling and predicting dynamical systems in out-of-distribution settings, which needs to be addressed in the future. This is also significant in practical situations because it is often more difficult to collect experimental data on phenomena with large energy than those with small energy. Gaussian processes could be a promising approach to solving such problems.

## Acknowledgments

This work was supported by JST, ACT-X Grant Number JPMJAX210D, Japan.

## References

[1] Nicholas M. Boffi, Stephen Tu, and Jean-Jacques E. Slotine. Nonparametric adaptive control and prediction: Theory and randomized algorithms. *J. Mach. Learn. Res.*, 23(1), jan 2022.

[2] Ricky T. Q. Chen, Brandon Amos, and Maximilian Nickel. Learning neural event functions for ordinary differential equations. In *International Conference on Learning Representations*, 2021.

[3] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

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

[5] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[6] Katharina Rath, Christopher G. Albert, Bernd Bischl, and Udo von Toussaint. Symplectic Gaussian process regression of maps in Hamiltonian systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(5):053121, 2021.

[7] E. Solak, R. Murray-smith, W. Leithead, D. Leith, and Carl Rasmussen. Derivative observations in Gaussian process models of dynamic systems. In *Advances in Neural Information Processing Systems*, volume 15, 2003.

[8] Yusuke Tanaka, Tomoharu Iwata, and Naonori Ueda. Symplectic spectrum gaussian processes: Learning hamiltonians from noisy and sparse data. In *Advances in Neural Information Processing Systems*, 2022.