Gaussian Process-Based Representation Learning via Timeseries Symmetries

Petar Bevanda [*1] **Max Beier** [*1] **Armin Lederer** [2] **Alexandre Capone** [3] **Stefan Sosnowski** [1] **Sandra Hirche** [1]

Abstract

Credible forecasting and representation learning of dynamical systems are of ever-increasing importance for reliable decision-making. To that end, we propose a family of Gaussian processes for dynamical systems with linear time-invariant responses, which are nonlinear only in initial conditions. This linearity allows us to tractably quantify forecasting and representational uncertainty simultaneously—alleviating the challenge of multistep uncertainty propagation in GP models and enabling a new probabilistic treatment of learning representations. Using a novel data-based symmetrization, we improve the generalization ability of GPs and obtain tractable, continuous-time posteriors without the need for multiple models or approximate uncertainty propagation.

1 Introduction

Learning predictive models for forecasting dynamical systems is of paramount importance due to complex and often unknown interactions between quantities of interest (Brunton & Kutz, 2019). The great utility of such models helps advance various different fields such as fluid mechanics (Kundu et al., 2015), molecular biology (Lindorff-Larsen et al., 2011), robotics (Billard et al., 2022) or safety-constrained decision making (Hewing et al., 2020b; Brunke et al., 2022). Dynamical system descriptions commonly require simulation for forecasting and uncertainty propagation, which can be difficult for non-parametric data-driven models (Hewing et al., 2020a; Beckers & Hirche, 2022). From real-world measurements of complex and unknown dynamical systems, data often comes in the form of sequential data that can be arbitrarily and non-uniformly sampled. Despite such discrete observations, there is often a certain regularity in the evolution of quantities interest (Biloš et al., 2023) in various different domains, from medical, industrial to financial (Sezer et al., 2020; Deb et al., 2017; Lim & Zohren, 2021)—making it important to impose structure that discourages temporal fluctuations. To account for these challenges in modeling dynamical systems, the choice of *representations* when learning from data becomes a deciding factor in the difficulty of forecasting as well as inference, especially when modeling complex phenomena (Mezić & Banaszuk, 2004) or long time-series (Gu et al., 2022). To address the above challenges we consider non-parametric learning paradigms along with their appeals and challenges, emphasising *uncertainty quantification* and *forecasting simplicity*.

The linearity of Koopman operators and the forecast simplicity of *linear time-invariant* (LTI) models coming from their eigendecompositions, lead to their increasing popularity in learning dynamical systems (Bevanda et al., 2021; Otto & Rowley, 2021; Brunton et al., 2022). Nevertheless, existing LTI predictor models based on operator regression are limited to dissecting long-term components of stationary dynamics (Korda & Mezić, 2018; Klus et al., 2020; Kostic et al., 2022; 2023). While this approach is well-versed for stationary data and reversible dynamics, most real-world dynamical systems are irreversible and often even nonstationary (Wu & Noé, 2020). Thus, an increasing amount of methods considers *dynamics-informed* kernels (Zhao & Giannakis, 2016; Berry & Sauer, 2016; Banisch & Koltai, 2017; Alexander & Giannakis, 2020; Burov et al., 2021; Dufée et al., 2024; Bevanda et al., 2023). By incorporating samples of the dynamics from sequential data into the kernel, eigenfunctions of Koopman operators can be accessed for both ergodic (Dufée et al., 2024) and transient settings (Bevanda et al., 2023). However, approaches of this type offer no forecast confidence bounds, lack principled tools for model selection and handling observation noise, cf. Table 1

Approach	LTI	End-to-end	Bayesian
GP-based	✗	✓	✓
Koopman-based	✓	✗	✗
this paper	✓	✓	✓

Table 1: Dynamical systems modeling approaches

[*]Equal contribution [1]Chair of Information-oriented Control, School of Computation, Information and Technology, TU Munich, Germany [2]Learning & Adaptive Systems Group, Department of Computer Science, ETH Zurich, Switzerland [3]Robotics Institute, CMU, Pittsburgh PA, United States of America. Correspondence to: Petar Bevanda <petar.bevanda@tum.de>, Max Beier <max.beier@tum.de>.

Presented at the ELLIS Workshop on Geometry-grounded Representation Learning and Generative Modeling at 41st International Conference on Machine Learning, Vienna, Austria.

In this work we present **K**oopman-**E**quivariant **G**aussian **P**rocesses (KEGPs), the first universal GP models with fully tractable and closed-form confidence bounds for multi-step prediction. This is enabled by latent dynamics based on simple LTI responses from a nonlinear function of the initial condition. Our GP model provides enhanced generalization ability through intrinsic symmetries (Koopman-equivariants) and delivers continuous-time posteriors without requiring time-derivative data. KEGPs allow for tractable quantification of both forecasting and representational uncertainty simultaneously – alleviating a traditional challenge of GPs and enabling a novel probabilistic treatment of learning representations of dynamical systems.

2 Preliminaries and problem statement

The following covers the necessary prerequisites for setting up the subsequent interplay between GPs, linear operators, and intrinsic dynamical system symmetries.

2.1 Preliminaries

System class We consider state-space models with Lipschitz dynamics and measurement functions

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \in \mathbb{X} \subset \mathbb{R}^n, \quad y = h(\boldsymbol{x}) \in \mathbb{R} \tag{1}$$

that have a well-defined flow $\boldsymbol{F}_t(\boldsymbol{x}_0) := \int_0^t \boldsymbol{f}(\boldsymbol{x}(\tau))d\tau$. The required local Lipschitz continuity is natural to physical systems that often evolve "smoothly". The canonical forecasting model for (1) is $y(t, \boldsymbol{x}_0) := h_t(\boldsymbol{x}_0) \equiv h \circ \boldsymbol{F}_t(\boldsymbol{x}_0)$, see Figure 1 (left). In practice, however, a numerical integration scheme is used to approximately solve the integral $\tilde{\boldsymbol{F}}_{t_\Delta} \approx \boldsymbol{F}_{t_\Delta}$ for a shorter time-interval t_Δ so $H = t/t_\Delta \in \mathbb{N}$, so the actual forecast becomes

$$y(t, \boldsymbol{x}_0) \approx h \circ \underbrace{\tilde{\boldsymbol{F}}_{t_\Delta} \circ \cdots \circ \tilde{\boldsymbol{F}}_{t_\Delta}}_{H\times}(\boldsymbol{x}_0) \tag{2}$$

Koopman operator-based modeling We can obtain a simpler (linear) forecasting model, by utilizing the fact that the composition of a function h with the flow $\boldsymbol{F}_t$ can be replaced by the linear Koopman operator $\mathcal{A}_t : \mathcal{C}(\mathbb{X}) \to \mathcal{C}(\mathbb{X})$ with $\mathcal{A}_t h(\boldsymbol{x}_0) := h(\boldsymbol{x}_t)$ (Koopman, 1931; Cvitanović et al., 2016). This representation is depicted in Figure 1 (middle). For a forward complete system (Krstic, 2009) on a compact non-recurrent domain $\mathbb{X}_{[0,T]}$, we can forecast using an *LTI predictor* (Bevanda et al., 2023)

$$y(t, \boldsymbol{x}_0) \approx [A_t \tilde{h}](\boldsymbol{x}_0), \quad A_t : \mathcal{H} \to \mathcal{H} \tag{3}$$

where $\mathcal{H}$ is a suitable finite-dimensional hypothesis space and $\tilde{h}$ is the restriction of h to $\mathcal{H}$. Note that $A_t \equiv P_{\mathcal{H}} \mathcal{A}_{t|_{\mathcal{H}}}$, where $P_{\mathcal{H}}$ is the (orthogonal) projector onto $\mathcal{H}$. Constructing a suitable $\mathcal{H}$ is nontrivial as the domain $\mathcal{C}(\mathbb{X})$ of the true operator $\mathcal{A}_t$ is infinite-dimensional. For practicable models, it is critical to obtain an accurate finite-dimensional restriction $\mathcal{A}_{t|_{\mathcal{H}}}: \mathcal{H} \to \mathcal{C}(\mathbb{X})$, which is a challenging task (Kostic et al., 2024). As such, it is at the core of our representation learning problem.

2.2 Problem statement

In this paper, we consider the problem of determining a representation of $\mathcal{A}_t$ acting on an observable h satisfying the following properties.

1. **Closure:** The hypothesis $\mathcal{H}$ satisfies $\|A_t - \mathcal{A}_{t|_{\mathcal{H}}}\| = 0$.
2. **Richness:** $\mathcal{H}$ is "big enough" to span h, i.e. $\mathcal{H}$ is universal – dense in $\mathcal{C}(\mathbb{X})$.
3. **Generalization:** Fast estimation error decay due to a, non-local, non-stationary kernel that does not introduce bias.

Closure and richness are crucial to control learning bias. If the *closure* property is not upheld, the dynamics of the learned system drift away and lose their consistency with the actual dynamics. Insufficient *richness* leads to the model being incapable of making accurate predictions, as the observable of interest can not be represented. Finally, generalization is paramount in scenarios with limited data, as non-asymptotic effects dominate the quality of the model.

By exploiting the linearity of the model (3) together with the symmetry contained in trajectories, we learn a suitable hypothesis space $\mathcal{H}$ using Gaussian process regression. This allows us to tractably quantify both forecasting and representational uncertainty simultaneously – alleviating the traditional challenge of GPs and enabling a novel probabilistic treatment of learning representations in the form of LTI predictors.

3 Closure via latent LTI symmetries of trajectory data

Our work builds upon the extensive literature on GPs (Rasmussen & Williams, 2006; van der Vaart & van Zanten, 2008; Kanagawa et al., 2018; Ridderbusch et al.), their interplay with linear operators (Särkkä, 2011; Matsumoto & Sullivan, 2024) and our newly introduced concept of *Koopman-equivariance*. As depicted in Figure 1, Koopman operators allow us to relate the action of the dynamics on the input of a measurement function (state-space) to its action on the output of the measurement functional (functional-space). Equivariance – the equivalence – of input and output action provides a constructive way to extract LTI factors of dynamical systems based on trajectory data.

LTI dynamical factors When a Koopman operator is spectral, we can consider dynamical systems at the level of flows

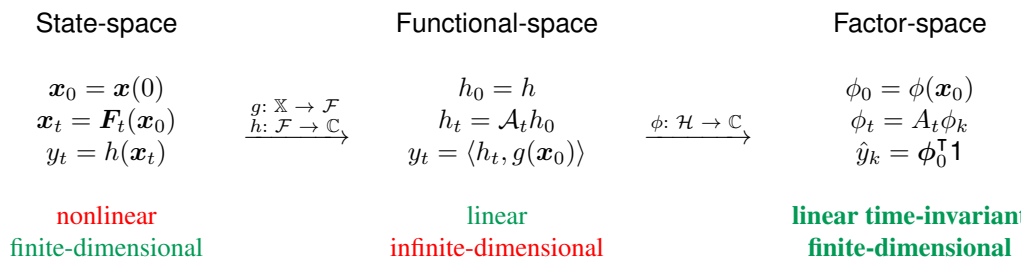

Figure 1: Using operator-theoretic ideas to trade-in a state-space model for a functional- or factor-space model.

through the *semi-conjugacy* $\phi \circ \boldsymbol{F}_t = \mathrm{e}^{\lambda t} \circ \phi$, that defines *dynamical factors/modes* (Mezić & Banaszuk, 2004)

$$\phi_t(\boldsymbol{x}) := \phi\left(\boldsymbol{F}_t(\boldsymbol{x})\right) = \mathrm{e}^{\lambda t}\, \phi(\boldsymbol{x}). \tag{4}$$

where $\lambda \in \mathsf{spec}(\mathcal{L})$ is an eigenvalue of the Lie derivative (Koopman generator) (Lasota & Mackey, 1994). The map $\mathrm{e}^{\lambda t}$ is often called a (topological) factor of $\boldsymbol{F}_t$, and, conversely, the flow $\boldsymbol{F}_t$ is called an extension of $\mathrm{e}^{\lambda t}$ (Zeng, 2018). This notion of semi-conjugacy provides a rigorous description of (simple) linear time-invariant dynamics $\mathrm{e}^{\lambda t}$ that are an *LTI factor* of (1). Those factors can be combined to yield an LTI basis for system representation, as depicted in Figure 1 (right). Nevertheless, it may be unclear how to obtain such dynamical factors in a generic fashion for unknown systems without an additional feature regression stage. Next, we will introduce the key concept to enable that, which we refer to as Koopman-equivariance.

Subgroup equivariance for dynamical systems Let $T_t : X \to X$ be a set of transformations on X for the abstract semigroup with parameter $t \in \mathcal{T}$. We say a function $\phi : X \to Y$ is equivariant to t if there exists an equivalent transformation on its output space $S_t : Y \to Y$ such that $\phi\left(T_t(\mathbf{x})\right) = S_t(\phi(\mathbf{x}))$ (Satorras et al., 2021). This effectively means that T_t and S_t essentially describe the same transformation but in different spaces.

Based on the above idea of equivariance, we formalize an intrinsic symmetry of a dynamical system through the interplay of the flow semigroup $\{\boldsymbol{F}_t\}_{t\in\mathbb{T}}$ that produces dynamical system trajectories and its corresponding *LTI dynamical factors* (4).

Definition 3.1 (Koopman-equivariance)**.** Let $\mathbb{T}$ be a compact subset of the time axis and $\mathcal{M}$ a manifold. A map $\phi_\lambda : \mathcal{M} \mapsto \mathcal{M}$ is called *Koopman-equivariant* if $\phi_\lambda \circ \boldsymbol{F}_t = \mathrm{e}^{\lambda t}\, \phi_\lambda$ on $\mathcal{M}$ for any $t \in \mathbb{T}$.

Considering trajectories as the unit of data, the defined Koopman-equivariance allows us to obtain LTI dynamical factors for our dynamical system before the inference itself even starts and satisfes the *closure* property.

3.1 Koopman-equivariant GPs: from past to future

Following Definition 3.1, we would require access to future trajectories for Koopman-equivariance. However, it suffices to symmetrize for the evaluation of a dynamical factor at the current state $\phi_\lambda\,|_{\boldsymbol{x}}$ in order to benefit from improved performance at test time. Hence, we form a causal model that utilizes the past for symmetrization to forecast the future model : $\boldsymbol{x}_{[T_p,t_0]} \mapsto y_{[t_0,T_f]}$ in a reliable fashion. To do so, we employ a symmetrization approach that guarantees that any function is immediately an LTI dynamical factor of the underlying dynamical system given access to trajectory data.

Since this approach allows us to use a standard GP regression formulation, hyperparameter optimization is straightforward, and GP approximations can be readily employed.

To implement the idea of Koopman-equivariance, we follow the agnostic symmetrization approach of (Kim et al., 2023; Nguyen et al., 2023) to obtain a constructive expression.

$$\phi_\lambda\,|_{\boldsymbol{x}} = \mathcal{E}_\lambda[\psi_{\boldsymbol{\theta}}] = \mathbb{E}_{t\sim\mathcal{U}([T_p,t_0])}\left[\mathrm{e}^{-\lambda t}\,\psi\left(\boldsymbol{x}(t)\right)\right] \tag{5a}$$

$$\psi(\boldsymbol{x}) \sim \mathcal{GP}(\hat{\psi}(\boldsymbol{x}), k_\psi(\boldsymbol{x}, \boldsymbol{x}')) \tag{5b}$$

By the closedness of Gaussian processes under linear operators (Matsumoto & Sullivan, 2024), ϕ_λ follows a Koopman-equivariant Gaussian process distribution, i.e., $\phi_\lambda \sim \mathcal{GP}(\hat{\phi}_\lambda, k_{\phi_\lambda})$, equipped with a covariance in the space of equivariant objects based on the past trajectory data, fulfilling Definition 3.1.

4 Rich representations using LTI factors

While (5) addresses the *closure* property, it is not *rich* enough to represent generic observables h over a time interval $[t_0, T_f]$. Fortunately, given a sufficiently rich set of eigenvalues, we get a *rich representation* (3) by setting the hypothesis to $\mathcal{H} := \mathrm{span}\{\phi_{\lambda_1}, \cdots, \phi_{\lambda_D}\}$ and the operator to $A_t := \mathrm{diag}([\mathrm{e}^{\lambda_1 t} \cdots \mathrm{e}^{\lambda_D t}])$ in (3). Then, dynamics (1) are universally representable (Korda & Mezić, 2020) in the

form of a *modal decomposition*

$$\tilde{y}_t(\cdot) := \sum_{j\in[D]} \mathrm{e}^{\lambda_j t}\, \phi_{\lambda_j}(\cdot). \tag{6}$$

so $\forall \varepsilon \in \mathbb{R}^+ \; \exists \tilde{y}_t(\cdot)$ such that $\|y_t - \tilde{y}_t\|_\infty \leq \varepsilon$ with $\phi_{\lambda_j} \in \mathcal{C}(\mathbb{X}), \lambda_j \in \mathbb{C}$ and $t \in [0, T]$. Such a perspective is shown to be a fruitful theoretical framework for building predictive models for dynamical systems with the mild requirement of non-recurrence (Bollt, 2021; Korda & Mezić, 2020; Bevanda et al., 2023; 2024).

4.1 Mode decomposition-induced GP structure

Trajectory-based GP With the predictor structure fully defined in (6), we place a prior on constituent factors via the eigenvalue distribution $\lambda \sim \hat{\rho}_\theta \approx \rho(\mathsf{spec}(\mathcal{A}_t))$ of the Koopman operator. As our goal is to obtain a high-likelihood representation – enabling automated model selection – we integrate the parameters of the distribution into Bayesian model selection. Finally, in order to use the mode decomposition (6) structure we employ a finite sample approximation of $\hat{\rho}_\theta$ using D spectral components. Each of the sampled spectral components defines a prior

$$\{\phi_{\lambda_j}(\boldsymbol{x}_{\mathsf{traj}}) \sim \mathcal{GP}(\hat{\phi}_{\lambda_j}(\boldsymbol{x}_{\mathsf{traj}}), k_{\phi_{\lambda_j}}(\boldsymbol{x}_{\mathsf{traj}}, \boldsymbol{x}'_{\mathsf{traj}}))\}_{j\in[D]}. \tag{7}$$

Adding LTI temporal dynamics Assuming no correlation between different factors $\{\phi_{\lambda_j}\}_{j\in[D]}$ immediately induces a structured prior from (5) given by $\tilde{y}_t(\cdot) \sim \mathcal{GP}(\hat{y}_t(\cdot), k_y(\cdot,\cdot))$, where the corresponding kernel factorizes into a linear temporal and nonlinear spatial component.

$$\hat{y}(t,\cdot) = \sum_{j\in[D]} \mathrm{e}^{\lambda_j t}\, \hat{\phi}_{\lambda_j}(\cdot), \tag{8a}$$

$$k_y((t,\cdot),(t',\cdot')) := \sum_{j\in[D]} k_{\mathsf{T}j}(t,t') k_{\phi_{\lambda_j}}(\cdot,\cdot'), \tag{8b}$$

where $k_{\mathsf{T}j}(t,t') = \mathrm{e}^{\lambda_j t}\, \mathrm{e}^{\lambda_j^* t'}$ and $\hat{\phi}_{\lambda_j}(\cdot)$, $k_{\phi_{\lambda_j}}(\cdot,\cdot)$ are defined via (5). Conceptually, this kernel is akin to a simulation-induced kernel (Chen, 2018). However, it is universal for nonlinear system responses and symmetrized using Koopman-equivariance.

Intuitively, the LTI featurization $\left\{k_{\mathsf{T}_j}(t,t')\right\}_{j\in[D]}$ can be considered as a prior on the temporal properties of the predictor. Temporal covariance with decay $|\lambda|$ close to zero will result in models with uniform uncertainty over time, while taking negative or positive decays will result in models with contracting or expanding uncertainty, respectively. Using this featurization, the model is guaranteed to be LTI.

5 Numerical experiments

We implement KEPGs using `GPJax` (Pinder & Dodd, 2022) and compare them to a simple continuous multi-task GP (MTGP): $\mathcal{GP}(\hat{\psi}(x,t), k_\psi((x,t),(x',t'))$ as a baseline model (Bonilla et al., 2007). All covariances are modeled with Gaussian radial basis functions with automatic relevance determination (ARD). The spectral distribution of KEGP is modeled as a parametrized uniform distribution and jointly optimized with the generative parameters. To compute (8), we sample $D = 128$ eigenvalues

		predator-prey	halfcheetah	weather
	$\#N$	256	4048	1024
	$\#n$ /$\#H$	2/16	24 / 16	8/18
KEGP	MSLL	**-1.19**±0.05	**-0.70**±0.02	**-0.74**±0.03
	RMSE	**0.31**±0.01	**0.50**±0.01	**0.51**±0.01
MTGP	MSLL	-0.61±0.10	0.00±0.00	0.00±0.00
	RMSE	0.42±0.05	1.00±0.00	1.08±0.00

Table 2: KEGP vs MTGP in MSLL and RMSE over 5 iid. simulations. $\#N$ denotes the number of trajectories, $\#H$ of forecast timesteps, and $\#n$ the state-space dimension.

5.1 The benefits of equivariance

We illustrate the benefits of our modeling approach compared to MTGP. To this end, we use the predator-prey model:

$$\dot{x}_1 = r_1 x_1 + c_i \gamma_1 x_1 x_1, \quad \dot{x}_2 = r_2 x_2 + c_i \gamma_2 x_1 x_2. \tag{9}$$

First, we display how KEGP modeling manifests in the covariance functions in Figure 2. We compare initial and marginal log-likelihood optimized covariances for KEGP and MTGP. The spatial covariance corresponds to that of a trajectory $\boldsymbol{x}_{\mathsf{traj}}$. The temporal covariance $[t_0, T_f]$ is separately displayed for the same trajectory. Notably, while the learned covariances for KEGP and MTGP are of similar shape, the initial spatial covariance of KEGP is less local than that of MTGP, already encoding the trajectory structure before optimization. Further, the KEGP temporal covariance is spectrally rich, as it is the superposition of multiple one-dimensional LTI factors, as opposed to a single nonlinear covariance. Notably, since the spectral distribution is a parametrized uniform distribution, KEGP has only seven parameters, whereas MTGP has 100. To illustrate how the non-local covariance affects the model's predictive distribution, Figure 3 shows a predicted (test) trajectory with KEGP and MTGP. It can be observed that the mean and confidence intervals are "smoother" for KEGP, resembling the smoothness in trajectories of the predator-prey model.

5.2 Quantitative study

To investigate the advantages of KEGPs on realistic data, we perform quantitative study on a range of benchmark examples of varying complexity. From the robotics domain, we consider expert demonstrations from D4RL (Fu et al., 2020) from the halfcheetah environment and forecast the first state and action. We take temperature data from the Monash TSF

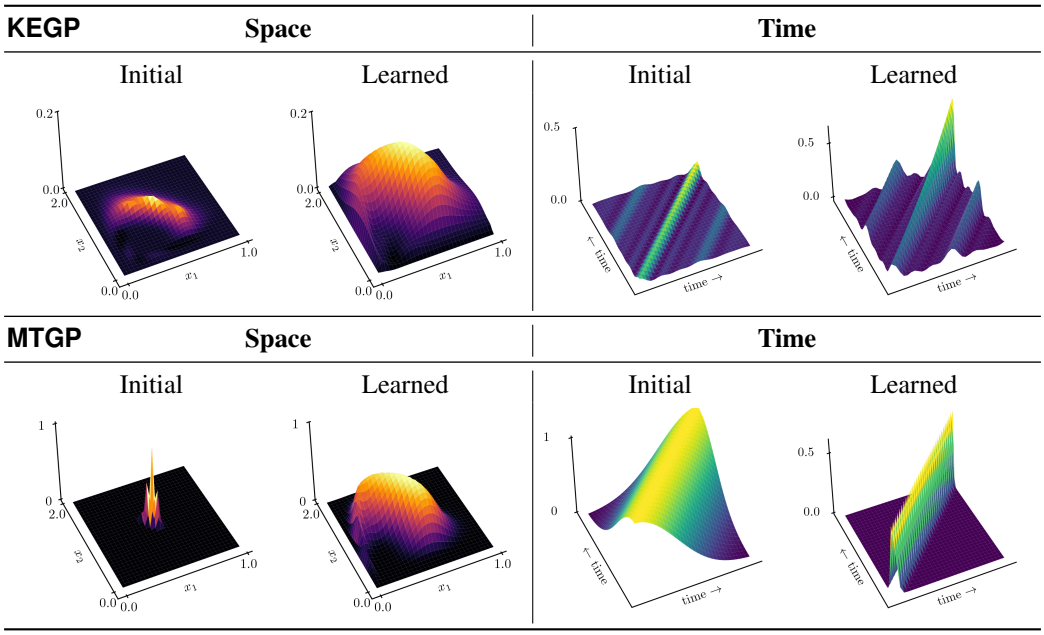

Figure 2: Visualization of the GP covariances in space and time. The spatial, KEGP prior already strongly indicates the shape of the optimized covariance.

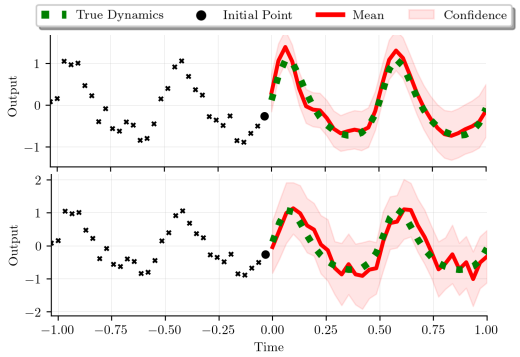

Figure 3: Multi-step mean and variance from noisy data of the predator population (9). **Top:** Koopman-equivariant GPs. **Bottom:** Multi-task GPs.

benchmark (Godahewa et al., 2021) as a sample for highly complex weather dynamics. Since these datasets provide a single long trajectory, we split off the last chunk as test data and partition the trajectory into $\#N$ input-task pairs to comply with our model : $[T_p, t_0] \to [t_0, T_f]$. Since exact inference in GPs is costly for larger datasets, we use an approximate model based on stochastic sparse variational inference (Hensman et al., 2015). We evaluate the performance in RMSE and MSLL (Rasmussen & Williams, 2006) as displayed in Table 2. We observe that KEGP can perform on all tasks, while MTGP frequently has zero MSLL, indicating convergence to the trivial model – resembling the mean and variance of the data. We attribute this difference to the dynamics-informed prior of our KEGPs facilitating the optimization procedure, which is crucial for complex problems.

6 Conclusion

We have presented a novel approach to incorporate a generic dynamical system structure as a prior through the holistic use of LTI dynamical factors and Koopman-equivariance. We realize this by defining a model using a GP with linear latent dynamics we call Koopman-equivariant Gaussian process (KEGP). Though a scalar-valued model, it allows multi-step error minimization and compression of trajectory information through a Koopman-equivariant hypothesis space. We demonstrate the versatility of regression with KEGP both qualitatively and quantitatively, demonstrating superior performance to vanilla multi-task GPs. Nevertheless, the main caveat remains the computational complexity of the proposed covariance function. Future work should investigate efficient computations w.r.t. input-sequence length, possibly drawing from recent ideas in neural sequence-to-sequence models (Guo et al., 2023). Also, work towards a sharp analysis of the proposed method is of great interest.

Acknowledgements

This work is supported by the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research.

References

Alexander, R. and Giannakis, D. Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. *Physica D: Nonlinear Phenomena*, 409:132520, 2020. ISSN 0167-2789. doi: https://doi.org/10.1016/j.physd.2020.132520.

Banisch, R. and Koltai, P. Understanding the geometry of transport: Diffusion maps for Lagrangian trajectory data unravel coherent sets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(3):035804, 02 2017.

Beckers, T. and Hirche, S. Prediction with Approximated Gaussian Process Dynamical Models. *IEEE Transactions on Automatic Control*, 67(12):6460–6473, 2022. doi: 10.1109/TAC.2021.3131988.

Berry, T. and Sauer, T. Local kernels and the geometric structure of data. *Applied and Computational Harmonic Analysis*, 40(3):439–469, 2016. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2015.03.002.

Bevanda, P., Sosnowski, S., and Hirche, S. Koopman operator dynamical models: Learning, analysis and control. *Annual Reviews in Control*, 52:197–212, 2021.

Bevanda, P., Beier, M., Lederer, A., Sosnowski, S., Hüllermeier, E., and Hirche, S. Koopman Kernel Regression. In *Advances in Neural Information Processing Systems*, volume 37, 2023.

Bevanda, P., Driessen, B., Iacob, L. C., Toth, R., Sosnowski, S., and Hirche, S. Nonparametric control-koopman operator learning: Flexible and scalable models for prediction and control. 5 2024. URL http://arxiv.org/abs/2405.07312.

Billard, A., Mirrazavi, S., and Figueroa, N. *Learning for Adaptive and Reactive Robot Control: A Dynamical Systems Approach*. MIT Press, 2022.

Biloš, M., Rasul, K., Schneider, A., Nevmyvaka, Y., and Günnemann, S. Modeling temporal data as continuous functions with stochastic process diffusion. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 2452–2470. PMLR, 23–29 Jul 2023.

Bollt, E. M. Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation. *Communications in Nonlinear Science and Numerical Simulation*, 100, 9 2021. ISSN 10075704. doi: 10.1016/j.cnsns.2021.105833.

Bonilla, E. V., Agakov, F. V., and Williams, C. K. I. Kernel multi-task learning using task-specific features. In Meila, M. and Shen, X. (eds.), *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, volume 2 of *Proceedings of Machine Learning Research*, pp. 43–50, San Juan, Puerto Rico, 21–24 Mar 2007. PMLR. URL https://proceedings.mlr.press/v2/bonilla07a.html.

Brunke, L., Greeff, M., Hall, A. W., Yuan, Z., Zhou, S., Panerati, J., and Schoellig, A. P. Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(1):411–444, 2022. doi: 10.1146/annurev-control-042920-020211.

Brunton, S. L. and Kutz, J. N. *Data-Driven Science and Engineering*. Cambridge University Press, 1 2019. ISBN 9781108380690. doi: 10.1017/9781108380690.

Brunton, S. L., Budišić, M., Kaiser, E., and Kutz, J. N. Modern Koopman Theory for Dynamical Systems. *SIAM Review*, 64(2):229–340, 2022.

Burov, D., Giannakis, D., Manohar, K., and Stuart, A. Kernel analog forecasting: Multiscale test problems. *Multiscale Modeling & Simulation*, 19(2):1011–1040, 2021. doi: 10.1137/20M1338289.

Chen, T. On kernel design for regularized LTI system identification. *Automatica*, 90:109–122, 2018. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2017.12.039.

Cvitanović, P., Artuso, R., Mainieri, R., Tanner, G., and Vattay, G. *Chaos: Classical and Quantum*. Niels Bohr Inst., Copenhagen, 2016. URL http://ChaosBook.org/.

Deb, C., Zhang, F., Yang, J., Lee, S. E., and Shah, K. W. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017. ISSN 1364-0321.

Dufée, B., Hug, B., Étienne Mémin, and Tissot, G. Ensemble forecasts in reproducing kernel hilbert space family. *Physica D: Nonlinear Phenomena*, 459:134044, 2024. ISSN 0167-2789. doi: https://doi.org/10.1016/j.physd.2023.134044.

Fu, J., Kumar, A., Nachum, O., Tucker, G., and Levine, S. D4rl: Datasets for deep data-driven reinforcement learning, 2020.

Godahewa, R., Bergmeir, C., Webb, G. I., Hyndman, R. J., and Montero-Manso, P. Monash time series forecasting archive. In *Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.

Gu, A., Goel, K., and Ré, C. Efficiently modeling long sequences with structured state spaces. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*, 2022.

Guo, Z. C., Dümbgen, F., Forbes, J. R., and Barfoot, T. D. Data-driven batch localization and slam using koopman linearization. 9 2023. URL http://arxiv.org/abs/2309.04375.

Hensman, J., Matthews, A., and Ghahramani, Z. Scalable Variational Gaussian Process Classification. In Lebanon, G. and Vishwanathan, S. V. N. (eds.), *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 351–360, San Diego, California, USA, 09–12 May 2015. PMLR. URL https://proceedings.mlr.press/v38/hensman15.html.

Hewing, L., Arcari, E., Fröhlich, L. P., and Zeilinger, M. N. On simulation and trajectory prediction with Gaussian process dynamics. In *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, pp. 424–434. PMLR, 10–11 Jun 2020a.

Hewing, L., Wabersich, K. P., Menner, M., and Zeilinger, M. N. Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):269–296, 2020b. doi: 10.1146/annurev-control-090419-075625.

Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. K. Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences. 7 2018. URL http://arxiv.org/abs/1807.02582.

Kim, J., Nguyen, D., Suleymanzade, A., An, H., and Hong, S. Learning probabilistic symmetrization for architecture agnostic equivariance. 36:18582–18612, 2023.

Klus, S., Schuster, I., and Muandet, K. Eigendecompositions of Transfer Operators in Reproducing Kernel Hilbert Spaces. *Journal of Nonlinear Science*, 30 (1):283–315, 2020. ISSN 1432-1467. doi: 10.1007/s00332-019-09574-z.

Koopman, B. O. Hamiltonian Systems and Transformation in Hilbert Space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 5 1931. ISSN 0027-8424. doi: 10.1073/pnas.17.5.315.

Korda, M. and Mezić, I. On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator. *Journal of Nonlinear Science*, 28(2):687–710, 4 2018. ISSN 14321467. doi: 10.1007/s00332-017-9423-0.

Korda, M. and Mezić, I. Optimal Construction of Koopman Eigenfunctions for Prediction and Control. *IEEE Transactions on Automatic Control*, 65(12):5114–5129, 12 2020. ISSN 15582523. doi: 10.1109/TAC.2020.2978039.

Kostic, V., Novelli, P., Maurer, A., Ciliberto, C., Rosasco, L., and Pontil, M. Learning Dynamical Systems via Koopman Operator Regression in Reproducing Kernel Hilbert Spaces. In *Advances in Neural Information Processing Systems*, pp. 4017–4031, 2022.

Kostic, V., Lounici, K., Novelli, P., and Pontil, M. Sharp spectral rates for koopman operator learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 32328–32339, 2023.

Kostic, V. R., Novelli, P., Grazzi, R., Lounici, K., and massimiliano pontil. Learning invariant representations of time-homogeneous stochastic dynamical systems. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=twSnZwiOIm.

Krstic, M. Forward-Complete Systems. In *Delay Compensation for Nonlinear, Adaptive, and PDE Systems*, pp. 171–190. Birkhäuser, Boston, 2009.

Kundu, P. K., Cohen, I. M., and Dowling, D. R. *Fluid mechanics*. Academic press, 2015.

Lasota, A. and Mackey, M. C. *Chaos, Fractals, and Noise*, volume 97. Springer New York, 1994. ISBN 978-1-4612-8723-0. doi: 10.1007/978-1-4612-4286-4.

Lim, B. and Zohren, S. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379:20200209, 4 2021. ISSN 1364-503X.

Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011. doi: 10.1126/science.1208351.

Matsumoto, T. and Sullivan, T. J. Images of gaussian and other stochastic processes under closed, densely-defined, unbounded linear operators. *Analysis and Applications*, 22:619–633, 4 2024. ISSN 0219-5305. doi: 10.1142/S0219530524400025.

Mezić, I. and Banaszuk, A. Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1):101–133, 2004.

Nguyen, T. D., Kim, J., Yang, H., and Hong, S. Learning symmetrization for equivariance with orbit distance minimization. 11 2023. URL http://arxiv.org/abs/2311.07143.

Otto, S. E. and Rowley, C. W. Koopman Operators for Estimation and Control of Dynamical Systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 4: 2021, 2021.

Pinder, T. and Dodd, D. Gpjax: A gaussian process framework in jax. *Journal of Open Source Software*, 7(75):4455, 2022. doi: 10.21105/joss.04455. URL https://doi.org/10.21105/joss.04455.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, 2006.

Ridderbusch, S., Ober-Blöbaum, S., and Goulart, P. Approximate uncertainty propagation for continuous gaussian process dynamical systems.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 18–24 Jul 2021.

Sezer, O. B., Gudelek, M. U., and Ozbayoglu, A. M. Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, 90:106181, 2020. ISSN 1568-4946.

Särkkä, S. Linear operators and stochastic partial differential equations in gaussian process regression. pp. 151–158. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21738-8.

van der Vaart, A. W. and van Zanten, J. H. Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, 36, 6 2008. ISSN 0090-5364. doi: 10.1214/009053607000000613.

Wu, H. and Noé, F. Variational approach for learning markov processes from time series data. *Journal of Nonlinear Science*, 30:23–66, 2 2020. ISSN 0938-8974. doi: 10.1007/s00332-019-09567-y.

Zeng, S. On systems theoretic aspects of koopman operator theoretic frameworks. In *2018 IEEE Conference on Decision and Control (CDC)*, pp. 6422–6427, 2018. doi: 10.1109/CDC.2018.8619090.

Zhao, Z. and Giannakis, D. Analog forecasting with dynamics-adapted kernels. *Nonlinearity*, 29(9):2888, aug 2016. doi: 10.1088/0951-7715/29/9/2888.