

# Estimation of Dynamic Gaussian Processes

Jilles van Hulst, Roy van Zuijlen, Duarte Antunes, W.P.M.H. (Maurice) Heemels

**Abstract**—Gaussian processes provide a compact representation for modeling and estimating an unknown function, that can be updated as new measurements of the function are obtained. This paper extends this powerful framework to the case where the unknown function dynamically changes over time. Specifically, we assume that the function evolves according to an integro-difference equation and that the measurements are obtained locally in a spatial sense. In this setting, we will provide the expressions for the conditional mean and covariance of the process given the measurements, which results in a generalized estimation framework, for which we coined the term Dynamic Gaussian Process (DGP) estimation. This new framework generalizes both Gaussian process regression and Kalman filtering. For a broad class of kernels, described by a set of basis functions, fast implementations are provided. We illustrate the results on a numerical example, demonstrating that the method can accurately estimate an evolving continuous function, even in the presence of noisy measurements and disturbances.

## I. INTRODUCTION

Gaussian Processes (GPs) have seen increasing interest in recent years due to their universal applicability in the estimation of unknown functions. This widely used framework is non-parametric and generally provides accurate estimates with relatively little data. In this paper, we extend this powerful technique to the case when the unknown function *dynamically* changes with *time*.

We consider the estimation of a class of spatio-temporal models, which have a time dimension, as well as a space dimension. The model class includes an evolving function that is described by a generalized version of an integro-difference equation (IDE) [1], [2]. The class also includes an observation equation, which specifies that we take local, noisy measurements of the evolving function at every discrete time step. We call this model class the Dynamic Gaussian Process (DGP).

Our main results pertain to the exact Bayesian estimation of the DGP. We provide analytical expressions for the conditional mean and covariance (with respect to available measurement data and priors). These computations include the calculation of an analytical integral at every time step where new measurements are received. As a result, in the general case, the DGP estimate expression and therefore the

computation may become very complex as the number of time steps increases. Therefore, we show that for a restricted class of DGPs, the required computations can be simplified considerably, leading to a rather efficient numerical implementation. The restricted class relies on basis functions and yields computations with complexity  $O(M^3)$ , with  $M$  the number of bases. Importantly, the computation no longer scales with the number of time steps. This is highly beneficial computationally, both when the DGP fits this simplified framework, and when it is approximated to do so. Interestingly, we also show that our Bayesian estimation framework generalizes both Gaussian process regression and Kalman filtering, two of the most important estimation frameworks in the literature.

This is not the first work to examine a connection between the Kalman filter (KF) [3] and the Gaussian process (GP) [4]. In fact, such connections have been studied extensively [5], [6], [7]. The computational complexity of a GP estimator can be drastically reduced by incorporating KF logic, given temporal data [8], [9]. Alternatively, the computation can be simplified by making use of sparse GP methods [10], [11], recursive methods [12] or by introducing basis functions [13], [14]. GPs have also been used alongside the Kalman filter to estimate state-dependent disturbances [15]. Moreover, in [16], GPs are used to estimate the state dynamics and observation model for different types of the Bayes filter, including the extended Kalman filter.

Turning to the estimation of evolving functions and spatio-temporal models, we find the KF and GP working in tandem. For example, [17] proposes an evolving GP estimator in continuous time using a vector of functions with spatial input, effectively creating an infinite-dimensional KF in continuous time. Moreover, [18] developed a framework for the estimation of evolving functions described by IDEs in the case that they are restricted to a set of basis functions. In the current work, however, we also consider the exact Bayesian estimation of the general unrestricted class. Additionally, the current work relaxes specific structure assumptions such as stationary measurement locations which are often not met in practically relevant problems. The Kriged Kalman filter is developed in [19], which presents a Kalman filter projected onto basis functions, making it suitable for the estimation of evolving functions. However, in this work, we establish a rigorous connection to a spatio-temporal model class, which allows our method to estimate IDEs. To the best of the authors' knowledge, there are currently no estimation frameworks that simultaneously generalize the KF and GP-based regression.

The remainder of this paper is organized as follows. In Section II, the main estimation problem of this paper is introduced. Moreover, it is explained how the problem

The research is carried out as part of the ITEA4 20216 ASIMOV project. The ASIMOV activities are supported by the Netherlands Organisation for Applied Scientific Research (TNO) and the Dutch Ministry of Economic Affairs and Climate (project number: AI211006). The research leading to these results is partially funded by the German Federal Ministry of Education and Research (BMBF) within the project ASIMOV-D under grant agreement No. 01IS21022G [DLR], based on a decision of the German Bundestag.

The authors are with the Control Systems Technology Section, Department of Mechanical Engineering, Eindhoven University of Technology, the Netherlands. Emails: {j.s.v.hulst, r.a.c.zuijlen, d.antunes, m.heemels}@tue.nl

generalizes the estimation problems in GP-based regression and the KF. In Section III, the exact solution to the estimation problem is detailed and analyzed. Section IV introduces the special class of DGPs that results in efficient computation. Section V presents the results of applying the proposed method in a numerical case study. A code example is available at [https://github.com/JvHulst/Dynamic\\_Gaussian\\_Processes](https://github.com/JvHulst/Dynamic_Gaussian_Processes). Lastly, Section VI gives conclusions and recommendations for future research.

**Notation.** Let  $\mathbb{R} = (-\infty, \infty)$ , and  $\mathbb{R}_{\geq 0} = [0, \infty)$ . Let  $\mathbb{N} = \{0, 1, \dots\}$  denote the natural numbers. For  $a, b \in \mathbb{N}$ , let  $\mathbb{N}_{[a,b]} = \{a, a+1, \dots, b\}$ . The identity matrix of size  $n$  is denoted by  $I_n$ . Let  $\mathbb{S}_+^n := \{A \in \mathbb{R}^{n \times n} \mid A \succeq 0\}$  denote the set of symmetric positive semidefinite matrices of size  $n \times n$ . A normal distribution with mean vector  $\bar{m} \in \mathbb{R}^n$  and covariance matrix  $\Sigma \in \mathbb{S}_+^n$  is denoted  $\mathcal{N}(\bar{m}, \Sigma)$ . The expected value of a random variable  $x \in \mathbb{R}^n$  is denoted  $\mathbb{E}[x]$ . The covariance of two random variables is denoted  $\text{cov}(x_1, x_2) := \mathbb{E}[(x_1 - \mathbb{E}[x_1])(x_2 - \mathbb{E}[x_2])^\top]$ . A Gaussian process with the mean function  $f : \mathcal{X} \rightarrow \mathbb{R}$  and the covariance function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is denoted  $\mathcal{GP}(\bar{f}(x), k(x, x'))$ .

Given a function  $f : \mathcal{X} \rightarrow \mathbb{R}$ , and a vector  $X = [x_1, x_2, \dots, x_n]^\top \in \mathcal{X}^n$ , let  $\mathbf{f}(X) = [f(x_1), f(x_2), \dots, f(x_n)]^\top \in \mathbb{R}^n$ . Let  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  be a kernel function, and let  $X = [x_1, x_2, \dots, x_n]^\top \in \mathcal{X}^n$  and  $X' = [x'_1, x'_2, \dots, x'_m]^\top \in \mathcal{X}^m$  be vectors. We define the kernel matrix  $\mathbf{k}(X, X') \in \mathbb{R}^{n \times m}$  corresponding to  $k$  to be the matrix whose  $(i, j)$ -th entry is given by  $k(x_i, x'_j)$ , i.e.,  $\mathbf{k}(X, X')_{ij} := k(x_i, x'_j)$ , for  $i \in \mathbb{N}_{[1,n]}$  and for  $j \in \mathbb{N}_{[1,m]}$ .

## II. MODEL CLASS

### A. Model Class and Problem Statement

The DGP consists of a dynamically evolving function and an observation model. The function at time  $t \in \mathbb{N}$  is denoted by  $f_t : \mathcal{X} \rightarrow \mathbb{R}$ , for  $x \in \mathcal{X} \subseteq \mathbb{R}$ , and its evolution dynamics are described by

$$f_{t+1}(x) = \int_{\mathcal{X}} f_t(s) \mu(x, ds) + w_t(x), \quad (1)$$

where  $\mu(x, \cdot)$  is a set of measures parametrized by  $x$  and such that the integral in (1) is given by the integral with respect to the measure  $\mu$  is defined as

$$\int_{\mathcal{X}} f(s) \mu(x, ds) = \int_{\mathcal{X}} f(s) k_f(x, s) ds + \sum_{i=1}^m f(s_i(x)) b_i(x). \quad (2)$$

Here,  $k_f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a continuous kernel function, and  $s_i : \mathcal{X} \rightarrow \mathcal{X}, i \in \mathbb{N}_{[1,m]}$ , continuous point mass functions with corresponding continuous weight functions  $b_i : \mathcal{X} \rightarrow \mathbb{R}, i \in \mathbb{N}_{[1,m]}$ . We assume that  $f_0(x) \sim \mathcal{GP}(\bar{f}_0(x), Q_f(x, x'))$  in which  $\bar{f}_0 : \mathcal{X} \rightarrow \mathbb{R}$  and  $Q_f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  are, respectively, the mean function and positive semidefinite covariance function of the initial condition function  $f_0$ , which are both continuous. The disturbances acting on the evolving function are denoted  $w_t(x) \sim \mathcal{GP}(0, Q_w(x, x'))$ ,  $t \in \mathbb{N}$ , in which  $Q_w : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a continuous positive semidefinite covariance function.

Because the integral transform is a linear transformation, it directly follows that  $f_t$  is a GP for any  $t \in \mathbb{N}$  [4], [20].

For many of the results presented in this paper, the discrete part of (2), i.e., the sum  $\sum_{i=1}^m f(s_i(x)) b_i(x)$ , is not considered. This leaves the evolving function dynamics (1) as an integro-difference equation (IDE) with a stochastic disturbance. The main purpose of the discrete part is to accompany both GPs and the KF in the same framework.

While our main results consider first general kernel functions  $k_f, Q_f, Q_w$  for Theorems 1 and 2 in Section III below, they will lead to computationally efficient implementations when the kernels take a special so-called separable form. We say that a kernel  $k(x, x')$  is separable, if

$$k(x, x') = U^\top(x) \Lambda U(x'), \quad (3)$$

for a vector of functions  $U(x) := [u_1(x), \dots, u_M(x)]^\top$  and a matrix  $\Lambda \in \mathbb{R}^{M \times M}$ . This special form is further examined in Section IV.

The observation model in the DGP is given by

$$Y_t = \mathbf{f}_t(X_t) + \mathbf{v}_t(X_t), \quad (4)$$

where  $Y_t \in \mathbb{R}^p$  is a vector of observations of the dynamic function  $f_t$  in (1) at each time step  $t \in \mathbb{N}$ , following a vector of corresponding locations  $X_t \in \mathcal{X}^p$ .  $\mathbf{v}_t$  follows from  $v_t(x) \sim \mathcal{GP}(0, Q_v(x, x'))$ ,  $t \in \mathbb{N}$ , which is noise on the observation, in which the positive semidefinite kernel  $Q_v : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  gives the covariance. It is assumed that  $f_0(x)$ ,  $w_t(x)$ , and  $v_t(x)$  are uncorrelated for any  $t, x$ .

The problem considered in this paper is the estimation of the function  $f_N$  using the data set  $\mathcal{D}_N := \{X^{(N)}, Y^{(N)}\}$  in which  $X^{(N)} := [X_0^\top, X_1^\top, \dots, X_N^\top]^\top \in \mathcal{X}^{Np}$  and corresponding  $Y^{(N)} := [Y_0^\top, Y_1^\top, \dots, Y_N^\top]^\top \in \mathbb{R}^{Np}$  as in (4), and  $N \in \mathbb{N}$  is arbitrary. This problem can be motivated from two different perspectives:

- 1) as an extension of the GP-based estimation problem to the case where the Gaussian process evolves in time;
- 2) as an extension of the Kalman filter problem to infinite-dimensional systems.

The connection between GP and DGP is shown in Section II-B below, while Section II-C addresses the connection between the KF and the DGP.

### B. Generalization of Gaussian Process Regression

A Gaussian process (GP) is in general an infinite-dimensional set of Gaussian variables, of which any finite subset constitutes a multivariate Gaussian distribution [4]. The GP framework is often used to estimate unknown functions based on observations that are subject to Gaussian noise, i.e.,

$$y_t = f(x_t) + v_t, \quad (5)$$

with  $x_t \in \mathcal{X}$  the function's argument,  $y_t \in \mathbb{R}$  the observation,  $f : \mathcal{X} \rightarrow \mathbb{R}$  the unknown function to be estimated, and  $v_t \sim \mathcal{N}(0, \sigma^2)$ ,  $t \in \mathbb{N}$  the Gaussian noise on the observation. An estimate of  $f$  from (5) for  $t \in \{0, 1, \dots, N\}$  is completely characterized by the measurements  $Y = \{y_0, y_1, \dots, y_N\} \in \mathbb{R}^N$ , the corresponding inputs  $X = \{x_0, x_1, \dots, x_N\} \in \mathcal{X}^N$ , and the prior beliefs on its mean and covariance, denoted respectively  $\bar{f} : \mathcal{X} \rightarrow \mathbb{R}$ , and  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ . In particular,

given a data set  $\mathcal{D} := \{X, Y\}$  of inputs and corresponding outputs according to (5), we obtain the following joint distribution

$$\begin{bmatrix} Y \\ f(x) \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \bar{f}(X) \\ \bar{f}(x) \end{bmatrix}, \begin{bmatrix} \mathbf{k}(X, X) + \sigma^2 I_N & \mathbf{k}(X, x) \\ \mathbf{k}(x, X) & k(x, x) \end{bmatrix} \right). \quad (6)$$

We can perform Gaussian process regression by conditioning the joint distribution in (6) with the data  $\mathcal{D} = \{X, Y\}$  using Bayes' rule. The resulting posterior distribution is Gaussian and can be expressed in terms of a mean estimate  $\hat{f} : \mathcal{X} \rightarrow \mathbb{R}$ , and a covariance estimate  $\hat{c} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ :

$$\hat{f}(x) := \mathbb{E}[f(x) \mid \mathcal{D}] = \bar{f}(x) + \mathbf{L}(x, X)(Y - \bar{f}(X)), \quad (7)$$

$$\begin{aligned} \hat{c}(x, x') &:= \text{cov}(f(x), f(x') \mid \mathcal{D}) \\ &= k(x, x') - \mathbf{L}(x, X)\mathbf{k}(X, x'), \end{aligned} \quad (8)$$

where  $\mathbf{L}(x, X) := \mathbf{k}(x, X) [\mathbf{k}(X, X) + \sigma^2 I_N]^{-1}$ .

The standard GP regression framework can be considered a special case of the problem posed in the previous subsection when (1) is restricted to

$$f_{t+1}(x) = f_t(x) \quad (9)$$

with  $f_0(x) = f(x)$ . The function dynamics (9) are obtained by choosing  $k_f = Q_w = 0$ ,  $m = 1$ ,  $s_1 = x$ , and  $b_1 = 1$ . In this sense, we can see the stated problem in Section II-A as an extension of the GP framework to the case where the unknown function  $f$  evolves according to (1). Note that if we choose  $Q_w$  arbitrarily instead, the DGP behaves as a GP-based estimator with a forgetting factor.

### C. Generalization of the Kalman Filter

The Kalman filter is a well-known algorithm to estimate the state of a dynamical system described by

$$x_{t+1} = Ax_t + w_t, \quad (10)$$

with the observation model

$$y_t = C_t x_t + v_t, \quad (11)$$

where  $x_t \in \mathbb{R}^n$  the state to be estimated,  $y_t \in \mathbb{R}^p$  the system output,  $w_t \sim \mathcal{N}(0, W)$ , the disturbance on the state with covariance matrix  $W \in \mathbb{S}_+^n$ , and  $v_t \sim \mathcal{N}(0, V)$  the measurement noise with covariance matrix  $V \in \mathbb{S}_+^p$  at time  $t \in \mathbb{N}$ . Lastly,  $x_0 \sim \mathcal{N}(\bar{x}_0, \bar{S})$  with mean  $\bar{x}_0 \in \mathbb{R}^n$  and covariance matrix  $\bar{S} \in \mathbb{S}_+^n$ .

The problem tackled by the Kalman filter is to obtain the estimate of the system's state that minimizes the covariance of the estimation error, given a history of observations  $\{y_i\}_{i=0}^N$  generated by (11). We express the state estimate in terms of a mean  $\hat{x}_{t|l} := \mathbb{E}[x_t \mid \{y_i\}_{i=0}^l]$  and covariance  $S_{t|l} := \text{cov}(x_t, x_t \mid \{y_i\}_{i=0}^l)$ . The famous KF result yields equations with which to optimally update these estimates given observations, as well as predict the evolution of the system using the dynamical model [3]. The update and prediction steps can be combined into the expressions

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + AL_t(y_t - C_t\hat{x}_{t|t-1}), \quad (12)$$

and

$$S_{t+1|t} = AS_{t|t-1}A^\top - AL_tC_tS_{t|t-1}A^\top + W, \quad (13)$$

with  $L_t = S_{t|t-1}C_t^\top(C_tS_{t|t-1}C_t^\top + V)^{-1}$  the Kalman gain. Starting with an initial condition  $\hat{x}_{0|-1} = \bar{x}_0$  and  $S_{0|-1} = \bar{S}$ , we can use the recursive relationships (12) and (13) to obtain an estimate for any  $N \in \mathbb{N}$ . Importantly, the KF has low memory requirements, due to the compact representation of the estimate. Additionally, the computation scales independently of the number of time steps  $N$ .

The standard Kalman filter can be seen as the solution to a special case of the estimation problem considered in this paper when we consider the estimation of  $f_t$  for a finite set of points  $\Xi = \{\xi_1, \xi_2, \dots, \xi_n\} \in \mathcal{X}^n$ . The KF state is then given by  $\mathbf{f}_t(\Xi)$ . We obtain linear state dynamics by choosing  $k_f(x, s) = 0$   $s_i = \xi_i$ , and  $b_i(\xi_j) = A_{ij}$  in (1), with  $A_{ij}$  the elements of  $A$  in (10) for  $i, j \in \mathbb{N}_{[1, n]}$ . Lastly, the observation model (11) is a result of the sampling locations  $X_t$  in (4).

The following section details the solution to the DGP estimation problem.

### III. ESTIMATION OF DYNAMIC GAUSSIAN PROCESSES

Here, we consider the estimation problem for the system (1) with observation model (4). To solve the DGP estimation problem, we will first state two theorems, which together constitute the main contribution of this paper. The theorems conceptually generalize Theorem 4.1 in [21, Chapter 7], related to the Kalman filter.

We aim to estimate the true evolving function  $f_N(x)$  for arbitrary  $N \in \mathbb{N}$  using the data set  $\mathcal{D}_N$ , and the model (1) with the continuous part only, i.e.,

$$f_{t+1}(x) = \int_{\mathcal{X}} f_t(s)k_f(x, s)ds + w_t(x). \quad (14)$$

The estimate is characterized by a mean function

$$\hat{f}_{t|l}(x) := \mathbb{E}[f_t(x) \mid \mathcal{D}_l], \quad (15)$$

and a covariance function

$$\hat{c}_{t|l}(x, x') := \text{cov}(f_t(x), f_t(x') \mid \mathcal{D}_l). \quad (16)$$

The first theorem is concerned with the conditioning of the mean and variance estimate based on newly obtained data.

**Theorem 1 (DGP update).** Consider the evolving function dynamics (14) with a prior belief of  $f_t$  based on  $\mathcal{D}_{t-1}$ , expressed in terms of a mean  $\hat{f}_{t|t-1}$  and a covariance  $\hat{c}_{t|t-1}$  as defined in (15) and (16). Consider the updating of this belief based on new data  $X_t, Y_t$ , obtained according to (4). The posterior mean and covariance, conditioned on  $X_t$  and  $Y_t$ , are then given by

$$\hat{f}_{t|t}(x) = \hat{f}_{t|t-1}(x) + \mathbf{L}_t(x, X_t)(Y_t - \hat{\mathbf{f}}_{t|t-1}(X_t)), \quad (17)$$

$$\hat{c}_{t|t}(x, x') = \hat{c}_{t|t-1}(x, x') - \mathbf{L}_t(x, X_t)\hat{\mathbf{c}}_{t|t-1}(X_t, x'), \quad (18)$$

in which

$$\mathbf{L}_t(x, X_t) = \hat{\mathbf{c}}_{t|t-1}(x, X_t) [\hat{\mathbf{c}}_{t|t-1}(X_t, X_t) + \mathbf{Q}_v(X_t, X_t)]^{-1}.$$

*Proof.* The first step in the proof pertains to the separation of the new data  $X_t, Y_t$  from the previously obtained data  $\mathcal{D}_{t-1}$  in terms of their contribution to the belief. To do so, we define  $\tilde{Y}_t := Y_t - \mathbb{E}[Y_t \mid \mathcal{D}_{t-1}]$ , which is independent from  $\mathcal{D}_{t-1}$ , following Theorem 3.2 in [21, Chapter 7]. Note that since we only consider finite-dimensional data here, the

result from [21] can directly be applied, even though we are considering an infinite-dimensional system. We obtain

$$\begin{aligned}\mathbb{E}[f_t(x) \mid \mathcal{D}_{t-1}, X_t, Y_t] &= \mathbb{E}[f_t(x) \mid \mathcal{D}_{t-1}, X_t, \tilde{Y}_t] \\ &= \hat{f}_{t|t-1}(x) + \mathbb{E}[f_t(x) \mid X_t, \tilde{Y}_t] - \mathbb{E}[f_t(x)].\end{aligned}$$

The first equality holds as it is a transformation of variables. The second holds due to independence of  $X_t$  and  $\tilde{Y}_t$  with  $\mathcal{D}_{t-1}$ , following Theorem 3.3 in [21, Chapter 7]. The expression shows that the incorporation of newly obtained data  $X_t, Y_t$  can be performed as an update on  $\hat{f}_{t|t-1}$ .

Next, we can express the second term in the last expression by conditioning  $f_t$  on the data  $X_t$  and  $\tilde{Y}_t$  using Bayes' rule. We have

$$\mathbb{E}[f_t(x) \mid X_t, \tilde{Y}_t] = \mathbb{E}[f_t(x)] + \mathbf{L}_t(x, X_t)(\tilde{Y}_t),$$

in which

$$\mathbf{L}_t(x, X_t) = \hat{\mathbf{c}}_{t|t-1}(x, X_t) [\hat{\mathbf{c}}_{t|t-1}(X_t, X_t) + \mathbf{Q}_v(X_t, X_t)]^{-1}.$$

To get the posterior covariance, we use the definition of the covariance and simply multiply the estimation error  $e_{t|t} := f_t - \hat{f}_{t|t}$  with its transpose to obtain

$$\hat{\mathbf{c}}_{t|t}(x, x') = \hat{\mathbf{c}}_{t|t-1}(x, x') - \mathbf{L}_t(x, X_t)\hat{\mathbf{c}}_{t|t-1}(X_t, x'),$$

which completes the proof.  $\square$

We now proceed with the second theorem, which details the propagation of the mean and covariance estimates through the function dynamics (14).

**Theorem 2 (DGP prediction).** Consider the evolving function dynamics given by (14) with a prior belief about the mean and variance of  $f_t$  based on the information set  $\mathcal{D}_t$ , denoted  $\hat{f}_{t|t}$ , and  $\hat{\mathbf{c}}_{t|t}$ , respectively, as in (15) and (16). Then, the belief evolves due to the functional dynamics in (14) to obtain

$$\hat{f}_{t+1|t}(x) = \int_{\mathcal{X}} k_f(x, s) \hat{f}_{t|t}(s) ds, \quad (19)$$

and

$$\hat{\mathbf{c}}_{t+1|t}(x, x') = \int_{\mathcal{X}} \int_{\mathcal{X}} k_f(x, s) \hat{\mathbf{c}}_{t|t}(s, s') k_f(s', x') ds ds' + Q_w(x, x'). \quad (20)$$

*Proof.* Starting with the mean estimate, we have

$$\begin{aligned}\mathbb{E}[f_{t+1}(x) \mid \mathcal{D}_t] &= \mathbb{E}\left[\int_{\mathcal{X}} k_f(x, s) f_t(s) ds + w_t(x) \mid \mathcal{D}_t\right], \\ &= \int_{\mathcal{X}} k_f(x, s) \mathbb{E}[f_t(s) \mid \mathcal{D}_t] ds,\end{aligned}$$

where the first equality follows from the dynamics (1), and the second follows from the zero-mean assumption on  $w_t$  and from the linearity of the expectation, since the integral transform is a linear map.

Next, the covariance estimate. We first define the deviation from the expected mean as  $e_{t+1|t} := f_{t+1} - \hat{f}_{t+1|t}$ . We have

$$\begin{aligned}\text{cov}(f_{t+1}(x), f_{t+1}(x') \mid \mathcal{D}_t) &= \mathbb{E}[e_{t+1|t}(x) e_{t+1|t}(x') \mid \mathcal{D}_t] \\ &= \mathbb{E}\left[w_t(x) w_t(x') + \left(\int_{\mathcal{X}} k_f(x, s) (f_t(s) - \hat{f}_{t|t}(s)) ds\right) \right. \\ &\quad \times \left. \left(\int_{\mathcal{X}} k_f(x', s') (f_t(s') - \hat{f}_{t|t}(s')) ds'\right) \mid \mathcal{D}_t\right] \\ &= \int_{\mathcal{X}} \int_{\mathcal{X}} k_f(x, s) \hat{\mathbf{c}}_{t|t}(s, s') k_f(s', x') ds ds' + Q_w(x, x').\end{aligned}$$

The first equality holds by definition of the covariance. The second holds because  $w_t$  is independent of both  $f_t$  and  $\hat{f}_{t|t}$ . The last equality holds because the integral transform is linear, because of the independence between  $w_t$  and  $\mathcal{D}_t$ , and by definition of  $\hat{\mathbf{c}}_{t|t}$  and  $Q_w$ . This completes the proof.  $\square$

We can combine the theorems to estimate the DGP at arbitrary time steps  $N \in \mathbb{N}$ . To do so, we initialize the estimator's mean and covariance, ideally, but not necessarily, as  $\hat{f}_{0|-1} = \bar{f}_0$ , and  $\hat{\mathbf{c}}_{0|-1} = Q_f$ . Then we perform the following algorithm at each time step  $t$ :

- 1) Update the DGP estimator using Theorem 1 according to newly obtained data;
- 2) Predict the DGP evolution using Theorem 2.

These steps can be performed recursively until we obtain  $\hat{f}_{N|N}$  and  $\hat{\mathbf{c}}_{N|N}$ , which presents the solution to the problem posed in Section II.

The implementation of Theorems 1 and 2 applies to the estimation of DGPs with a general structure. However, it may get computationally very expensive to perform as  $N$  increases due to the continuous integral, which in general produces complicated expressions for the mean and covariance estimates.

The next section details how the computation of the estimator is reduced when the kernels of the DGP are separable.

#### IV. SEPARABLE KERNELS

In this section, we show that when  $k_f$ ,  $Q_f$ , and  $Q_w$  are separable kernels we can perform the required computations of the previous section efficiently. We start with exact DGP when the kernels are truly separable, and detail afterward that separable kernels can be used to approximate the problem from Section II-A. To show the reductions in computational steps in case of separability of kernels  $k_f$ ,  $Q_f$ , and  $Q_w$ , consider the following lemmas.

**Lemma 3.** Suppose that

$$\begin{aligned}k_f(x, x') &= U^\top(x) \Lambda U(x'), \\ Q_f(x, x') &= U^\top(x) \Lambda_f U(x'), \\ Q_w(x, x') &= U^\top(x) \Lambda_w U(x')\end{aligned} \quad (21)$$

with  $\Lambda \in \mathbb{R}^{M \times M}$ ,  $\Lambda_f, \Lambda_w \in \mathbb{S}_+^M$  and  $U(x) := [u_1(x), u_2(x), \dots, u_M(x)]^\top$  a vector of basis functions in which  $u_i : \mathcal{X} \rightarrow \mathbb{R}, i \in \{1, 2, \dots, M\}$  continuous, then the DGP covariance update step in (18) reduces to

$$\hat{\mathbf{c}}_{t|t}(x, x') = U^\top(x) \Psi_{t|t} U(x') \quad (22)$$

with  $\Psi_{t|t} := \Psi_{t|t-1} + \Gamma_t \mathbf{U}^\top(X_t) \Psi_{t|t-1}$ , in which  $\Psi_{0|-1} = \Lambda_f$ , and

$$\Gamma_t := \Psi_{t|t-1} \mathbf{U}(X_t) [\mathbf{U}^\top(X_t) \Psi_{t|t-1} \mathbf{U}(X_t) + \mathbf{Q}_v(X_t, X_t)]^{-1}$$

with  $\mathbf{U}(X_t) := [\mathbf{u}_1(X_t) \quad \mathbf{u}_2(X_t) \quad \dots \quad \mathbf{u}_1(X_t)]^\top \in \mathbb{R}^{M \times p}$ . Additionally, the covariance prediction step in (20) is reduced to

$$\hat{\mathbf{c}}_{t+1|t}(x, x') = U^\top(x) \Psi_{t+1|t} U(x') \quad (23)$$

with  $\Psi_{t+1|t} := \Lambda \Lambda_U \Psi_{t|t} \Lambda_U^\top + \Lambda_w$ , in which  $\Lambda_U := \int_{\mathcal{X}} U(x) U^\top(x) dx$ .

*Proof.* Initialize  $\hat{c}_{0|-1}(x, x') = U^\top(x)\Psi_{0|-1}U(x')$  with  $\Psi_{0|-1} = \Lambda_f$ , then substitute (21) into (18) and (20).  $\square$

*Remark.* Note that  $Q_v$  does not need to be separable, as the measurement noise function  $v_t$  is always evaluated for a finite set of spatial locations  $X_t$  when taking observations according to (4), and  $Q_v(X_t, X_t)$  can be simply computed and used in the calculations above.

To state *Lemma 3* in other words, the assumption in (21) causes the covariance estimate  $\hat{c}_{t|l}$  to remain in the basis under operation of both the update and the prediction steps. This enables a computationally efficient implementation of the estimator since we now only need to keep track of the matrix  $\Psi_{t|l}$ . However, the mean estimate computation still scales poorly as the time index  $N$  increases. Efficient implementation of the mean estimate is enabled by the application of the following lemma.

**Lemma 4.** If (21) holds and additionally

$$\bar{f}_0(x) = U^\top(x)\bar{z}, \quad (24)$$

with  $\bar{z} \in \mathbb{R}^M$  and  $U$  the same vector of functions as in (21), then the update of the mean estimate (17) reduces to

$$\hat{f}_{t|t}(x) = U^\top(x)z_{t|t}, \quad (25)$$

with  $z_{t|t} := z_{t|t-1} + \Gamma_t(Y_t - \mathbf{U}^\top(X_t)z_{t|t-1})$ . Additionally, the prediction step in (19) is reduced to

$$\hat{f}_{t+1|t}(x) = U^\top(x)z_{t+1|t}, \quad (26)$$

with  $z_{t+1|t} = \Lambda U z_{t|t}$ .

*Proof.* Initialize  $\hat{f}_{0|-1}(x) = U^\top(x)z_{0|-1}$  with  $z_{0|-1} = \bar{z}$ , then substitute (21) into (17) and (19).  $\square$

Note the similarity of the expressions in the lemmas to the standard KF. With *Lemma 3* and *Lemma 4*, the DGP can be interpreted as a Kalman filter where the state estimates are projected onto the functions in  $U$ , similar to [19]. Additionally, it can be evaluated at every spatial location and the measurement data can occur anywhere in the set  $\mathcal{X}$ .

If both lemmas apply, the problem presented in Section II can be exactly estimated using the efficient equations presented in this section. For the more general case, where (1) does not satisfy the lemmas, the problem can be approximated using basis functions. In order to do this, we approximate the function  $\bar{f}_0$  in the basis vector  $U$ , and the kernel functions  $k_f, Q_f, Q_w$  in the basis vector  $U(x) \otimes U(x')$ , with  $\otimes$  the Kronecker product. We can find  $\Lambda, \Lambda_f, \Lambda_w$  and  $\bar{z}$  by performing standard parametric function approximation methods such as least squares using the  $L^2$ -norm of the approximation error. Note that this least squares problem can be numerically approximated using Riemann sums. Two important choices for the basis structure in  $U$  are highlighted. The first is discrete bins, while the second is the Fourier basis. Note that the class of separable kernels is dense in the space of  $L^2$  integrable kernels for both of these choices.

The key benefit of the lemmas lies in the reduced computational complexity. When the DGP structure admits the basis functions, either through approximation or exactly, the computation scales with  $O(M^3)$ , similar to the standard KF.

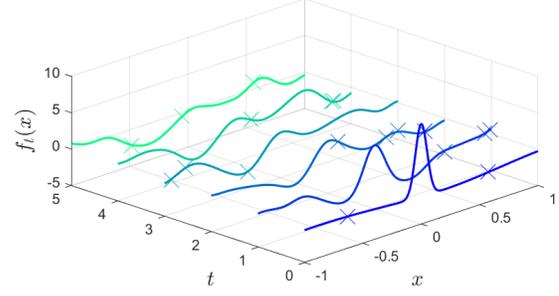


Fig. 1. Function  $f_t$  over time steps  $t$  using a smoothing evolution kernel  $k_f$ , a Gaussian initial mean function  $\bar{f}$ , a smooth initial condition covariance  $Q_f$ , and smooth zero-mean disturbances  $w_t$ .

Additionally, this implies that the approximation framework provides a choice to the user in regard to computational burden.

## V. RESULTS

In this section, an application of DGPs and of the proposed approximation method presented in Section IV to a numerical example is presented. To assess the estimation accuracy of the approximate methods, a ground truth model is considered, which uses 625 basis functions in the form of discrete bins compared to up to 91 Fourier bases in the approximation. We select  $\mathcal{X} = [-1, 1]$ . We model

$$k_f(x, x') = \phi(5.13, 0.07, (x - x')),$$

in which  $\phi(a, \sigma, d) := ae^{-\frac{|d|^2}{2\sigma^2}}$  is a parametrized squared exponential kernel. This choice results in local smoothing and decay of the function  $f_t$  every time step  $t$ . Furthermore,

$$Q_f(x, x') = \phi(1, 0.7, (x - x')),$$

which ensures that the initial condition  $f_0$  smoothly deviates from its mean. The disturbances are modeled to have smooth covariance using

$$Q_w(x, x') = \phi(0.35, 0.15, (x - x')).$$

Next,

$$Q_v(x, x') = \sigma_v^2 \delta(x - x')$$

with  $\delta$  the Dirac delta function and  $\sigma_v = 0.1$  simulates white noise on the measurements since this choice implies there is no spatial correlation in  $v_t$ . Lastly,

$$\bar{f}_0(x) = \phi(10, 0.05, x)$$

simulates a Gaussian in the center of the function domain at the initial time step  $t = 0$ . The spatial locations  $X_t$  are selected randomly every  $t$  using a uniform distribution over  $\mathcal{X}$  with  $p = 3$ . Note that now,  $X_t$  is a random variable. This still allows us to use the theorems and lemmas, however, as it is independent of  $\mathcal{D}_{t-1}$ . The results are shown in Fig. 1, Fig. 2, and Fig. 3.

Observe Fig. 1, which shows a simulation of the ground truth model. Note the smoothing of the function as the time index  $t$  increases as a result of the choice of kernel  $k_f$ . Note furthermore that disturbances affect the function during the entire simulation, for instance at the location  $x = -0.3$  at time step  $t = 1$ .

Fig. 2 displays the approximation of the DGP in Fig. 1 using 31 Fourier bases. Note that wherever an observation

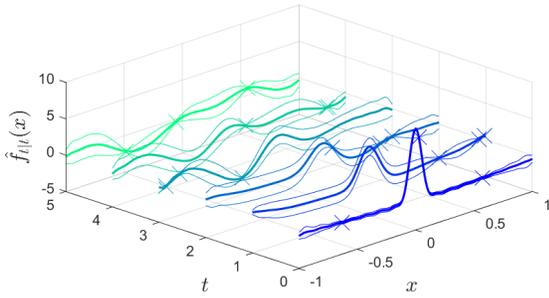


Fig. 2. Estimate  $\hat{f}_{t|x}$  of function  $f_t$  using 31 Fourier bases over time steps  $t$ . A confidence interval of 95% is plotted alongside the mean estimate.

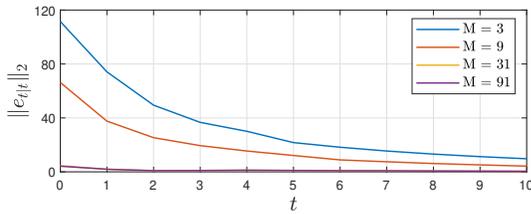


Fig. 3. 2-norm of the estimation error  $e_{t|x}$  over time steps  $t$  without disturbances. A comparison is made between 3, 9, 31, and 91 Fourier bases in  $U$ .

occurs, indicated by a cross, the local confidence is improved. Note furthermore that even in the absence of measurements, the estimate changes due to the prediction of the dynamics. Finally, observe that the local disturbance in the true function at  $t = 1$  only appears in the estimate at  $t = 3$ , because the local observation  $X_t$  did not occur in that local area immediately when the disturbance happened.

Lastly, consider Fig. 3. More bases yield a lower error 2-norm at  $t = 0$ , since the initial mean function is approximated better as the number of bases increases, with diminishing returns. Note that even when the number of bases is high, the error at  $t = 0$  is nonzero due to the uncertainty function's initial condition. After some measurements, this uncertainty is resolved and the error decreases. As the function evolves, it smoothens out, which reduces the amount of high-frequency content, making it easier to approximate even when the number of bases is lower.

## VI. CONCLUSIONS

This paper presented an exact method for the Bayesian estimation of dynamic Gaussian processes (DGPs) describing dynamically evolving uncertain functions. An explicit connection to the powerful techniques of Kalman filtering and Gaussian process regression is made. It is also shown that the computation of the DGP estimation is reduced when a separability structure is present in the kernel and mean functions describing the problem. Next to this exact method, we exploited the separable structure to derive an approximation method using basis functions with reduced memory and computational burden during operation. The proposed framework is demonstrated in a numerical case study.

Future research avenues include the analysis of the error resulting from the approximation, the extension of the problem to include control inputs, and connecting the framework to the estimation of partial differential equations.

## REFERENCES

- [1] N. Cressie and C. K. Wikle, *Statistics for Spatio-Temporal Data*, 1st ed. Hoboken: John Wiley & Sons Inc, 2011.
- [2] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-Temporal Data Mining," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–41, jul 2019.
- [3] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, mar 1960.
- [4] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, T. Dietterich, Ed. MIT Press, 2006.
- [5] M. P. Deisenroth and H. Ohlsson, "A general perspective on Gaussian filtering and smoothing: Explaining current and deriving new algorithms," in *Proceedings of the 2011 American Control Conference*. IEEE, jun 2011, pp. 1807–1812.
- [6] S. Särkkä, J. Hartikainen, L. Svensson, and F. Sandblom, "On the relation between Gaussian process quadratures and sigma-point methods," *Journal of Advances in Information Fusion*, vol. 11, no. 1, pp. 31–46, apr 2016.
- [7] S. Reece and S. Roberts, "An introduction to Gaussian processes for the Kalman filter expert," in *2010 13th International Conference on Information Fusion*. IEEE, jul 2010, pp. 1–9.
- [8] J. Hartikainen and S. Särkkä, "Kalman filtering and smoothing solutions to temporal Gaussian process regression models," *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2010*, pp. 379–384, 2010.
- [9] M. Todescato, A. Carron, R. Carli, G. Pillonetto, and L. Schenato, "Efficient spatio-temporal Gaussian regression via Kalman filtering," *Automatica*, vol. 118, p. 109032, aug 2020.
- [10] J. Quiñero, Q. Quiñero-Candela, C. E. Rasmussen, and C. M. De, "A Unifying View of Sparse Approximate Gaussian Process Regression," *The Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, dec 2005.
- [11] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When Gaussian Process Meets Big Data: A Review of Scalable GPs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4405–4423, nov 2020.
- [12] M. F. Huber, "Recursive Gaussian process: On-line regression and learning," *Pattern Recognition Letters*, vol. 45, no. 1, pp. 85–91, aug 2014.
- [13] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 209–226, jan 2008.
- [14] J. Hensman, N. Durrande, and A. Solin, "Variational Fourier Features for Gaussian Processes," *Journal of Machine Learning Research*, vol. 18, pp. 1–52, 2018.
- [15] T. Lee, "Adaptive Learning Kalman Filter with Gaussian Process," in *2020 American Control Conference (ACC)*, vol. 2020-July. IEEE, jul 2020, pp. 4442–4447.
- [16] J. Ko and D. Fox, "GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, sep 2008, pp. 3471–3476.
- [17] S. Sarkka, A. Solin, and J. Hartikainen, "Spatiotemporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing: A Look at Gaussian Process Regression Through Kalman Filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, jul 2013.
- [18] C. K. Wikle and N. Cressie, "A Dimension-Reduced Approach to Space-Time Kalman Filtering," *Biometrika*, vol. 86, no. 4, pp. 815–829, 1999.
- [19] K. V. Mardia, C. Goodall, E. J. Redfern, and F. J. Alonso, "The Kriged Kalman filter," *Test*, vol. 7, no. 2, pp. 217–282, dec 1998.
- [20] A. Papoulis and U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. McGraw-Hill Europe, 2002.
- [21] K. J. Astrom, *introduction to stochastic control theory*. New York: Academic Press, 1970.