Anticipating learning in multi-step ahead predictions of learning-based control

Alexandre Capone * **Armin Lederer** * **Sandra Hirche** *

* *Chair of Information-oriented Control (ITR), Department of Electrical and Computer Engineering, Technical University of Munich, Germany, (e-mail: {alexandre.capone,armin.lederer,hirche}@tum.de)*

Abstract: Although machine learning techniques are increasingly employed in control tasks, few methods exist to predict the behavior of closed-loop learning-based systems. In this paper, we introduce a method for computing confidence regions for closed-loop system trajectories when a learning-based control law is employed. We employ Monte Carlo simulations and exploit system properties to prove that the confidence regions are correct with high probability. In a numerical simulation, we show that the proposed approach accurately predicts the correct confidence regions up to small outliers.

Keywords: Gaussian processes, system identification, nonlinear systems, stochastic systems, Monte Carlo simulation, error estimation

1. INTRODUCTION

Technological advances have led to increasingly challenging control tasks, where first-principles models cannot be obtained either due to prohibitive complexity or lack of model knowledge. In such settings, model uncertainty is generally high, which leads to poor control performance if conventional model-based control techniques are applied. In order to address these issues, machine learning techniques have been increasingly employed (Deisenroth et al., 2015; Capone and Hirche, 2019; Umlauft et al., 2017; Chowdhary et al., 2015; Berkenkamp et al., 2017). These tools effectively deal with a large variety of model uncertainties, and exhibit good control performance in complex settings, provided that enough system data is available. Particular techniques which have recently garnered attention within this context are learning-based model predictive control (Koller et al., 2018; Kamthe and Deisenroth, 2018; Maiworm et al., 2018), as well as gain-varying parametric state feedback control (Beckers et al., 2019; Berkenkamp and Schoellig, 2015) based on learned models. Learning-based model predictive control often employs stochastic, nonparametric models such as Gaussian processes to update prior models in real-time. Since the employed machine learning techniques significantly complicate theoretical analysis, learning-based model predictive control typically either provide no guarantees (Kamthe and Deisenroth, 2018) or very conservative ones (Koller et al., 2018). However, the simulation of stochastic models, which are controlled based on models updated in real-time, is an open research question, such that there are no reliable simulation based evaluation methods for learning-based model predictive control. In contrast, stability is often formally proven for parametric state feedback control with learned models by exploiting robust control approaches. However, control performance is typically not analyzed such that control parameters such as gains (Beckers et al., 2019) or cost function parameters (Berkenkamp and Schoellig, 2015) must be tuned manually. Therefore, a method for simulating learning-based controllers with models updated in real-time is required.

In order to address this open problem, we introduce a multi-step ahead prediction algorithm for learning-based control laws that enables to determine how learning influences the control performance over a long time horizon. We employ unrestrictive assumptions and prove important properties for Gaussian process state space models, which in turn are employed to obtain confidence regions for the closed-loop system trajectory. In a numerical simulation of the cart-pole balancing problem, we show that the computed confidence region holds up to small errors. Moreover, the confidence region illustrates how system uncertainty is expected to decrease over time due to the learning-based nature of the control law.

The remainder of this paper is structured as follows. The problem statement is given in Section 2, after which we introduce Gaussian processes, in Section 3. Section 4 describes how the confidence regions are obtained, and provides corresponding guarantees. A numerical cart-pole experiment is given in Section 5, and Section 6 provides a conclusion and discussion.

2. PROBLEM STATEMENT

Consider a nonlinear system of the form

$$\begin{aligned} \boldsymbol{x}_{t+1} =& \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{g}(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{w}_t := \\ & \boldsymbol{f}(\tilde{\boldsymbol{x}}_t) + \boldsymbol{g}(\tilde{\boldsymbol{x}}_t) + \boldsymbol{w}_t, \end{aligned} \tag{1}$$

where $\boldsymbol{x}_t \in \mathcal{X} = \mathbb{R}^{d_x}$ and $\boldsymbol{u}_t \in \mathcal{U} \subseteq \mathbb{R}^{d_u}$ respectively denote the system's state and control input at time $t \in \mathbb{N}$. The initial condition $\boldsymbol{x}_0$ is fixed and known. The function $\boldsymbol{f} : \mathcal{X} \times \mathcal{U} \mapsto \mathbb{R}^{d_x}$ corresponds to the known component of the system dynamics, whereas $\boldsymbol{g} : \mathcal{X} \times \mathcal{U} \mapsto \mathbb{R}^{d_x}$ is unknown. The system is perturbed by independent and identically distributed (iid) process noise $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$, with $\boldsymbol{Q} = \operatorname{diag}(\sigma_1, \cdots, \sigma_{d_x}) \in \mathbb{R}_+^{d_x}$. Here $\mathbb{R}_+$ denotes

the positive real numbers. The augmented state $\tilde{\boldsymbol{x}}_t = (\boldsymbol{x}_t^{\mathrm{T}}, \boldsymbol{u}_t^{\mathrm{T}})^{\mathrm{T}}$ concatenates the state and input vectors, and is introduced for notational simplicity. We assume that the entries of $\boldsymbol{f}(\cdot)$ are continuously differentiable and exhibit at most polynomial growth, i.e.,

Assumption 1. The entries $f_i(\cdot)$ of $\boldsymbol{f}(\cdot)$ are continuously differentiable and are bounded by a known, positive, isotropic and monotone polynomial function $\pi : \mathbb{R} \mapsto \mathbb{R}$, i.e., $f_i(\tilde{\boldsymbol{x}}) \leq \pi(\|\tilde{\boldsymbol{x}}\|_2) \ \forall \ \tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}$, where $\|\cdot\|_2$ denotes the Euclidean norm.

This is not a very restrictive assumption, as multiple physical systems, such as robotic and electrical systems, are described by functions that satisfy Assumption 1, e.g. polynomial functions. Moreover, the case where no prior model information is available, which is given by $\boldsymbol{f}(\boldsymbol{x}_t) = \boldsymbol{x}_t$, also satisfies this assumption.
We consider a learning-based control law that collects measurements of the system dynamics and employs them to update itself. This is formally expressed as

$$\boldsymbol{u}_t := \boldsymbol{u}_t(\boldsymbol{x}_0, \cdots, \boldsymbol{x}_t), \tag{2}$$

where the functions $\boldsymbol{u}_t : \mathcal{X}^{t+1} \mapsto \mathcal{U}$ depend on the current state $\boldsymbol{x}_t$ and the measurement data collected up to time t. This definition applies to all adaptive control laws and also accomodates more general learning-based control laws. We make the following assumption with respect to the input space:

Assumption 2. The input space $\mathcal{U}$ is bounded, with $\|\boldsymbol{u}\|_2 \leq u_{\max}$ for all $\boldsymbol{u} \in \mathcal{U}$ and some fixed scalar $u_{\max} > 0$.

This is generally the case in practice, as the input $\boldsymbol{u}_t$ is often constrained due to safety or physical limitations.
Let $\boldsymbol{X} := (\boldsymbol{x}_1^{\mathrm{T}}, \cdots, \boldsymbol{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{X}^T$ denote an T-step sample closed-loop trajectory of the true system (1). We wish to obtain a confidence region $\mathcal{S} \subseteq \mathcal{X}^T$ for $\boldsymbol{X}$, such that

$$\mathbb{P}(\boldsymbol{X} \in \mathcal{S}) \geq 1 - \delta, \tag{3}$$

holds for a fixed $\delta \in (0, 1)$.

3. ANTICIPATING LEARNING USING GAUSSIAN PROCESSES

We now introduce Gaussian process (GP) models, and illustrate how they are emplyed to model (1). A GP is a collection of random variables, of which any finite subset is normally distributed (Rasmussen and Williams, 2006). A GP is fully characterized by a *mean* function $m : \tilde{\mathcal{X}} \mapsto \mathbb{R}$ and a symmetric positive definite *kernel* function $k : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \mapsto \mathbb{R}$, and is denoted $\mathcal{GP}(m, k)$. In this paper, we set $m \equiv 0$, which corresponds to a setting where no prior knowledge about $\boldsymbol{g}(\cdot)$ is available, and is applicable without loss of generality (Rasmussen and Williams, 2006). The kernel $k(\cdot, \cdot)$ encodes information about the unknown function $\boldsymbol{g}(\cdot)$, such as differentiability and periodicity. In settings where little is known about the characteristics of $\boldsymbol{g}(\cdot)$, *universal* kernels are often employed, as they uniformly approximate any continuous function in a closed subset of $\tilde{\mathcal{X}}$ (Micchelli et al., 2006).
If the state space is one-dimensional, i.e., $d_x = 1$, given system measurement data $\mathcal{D}_t = \{\tilde{\boldsymbol{X}}_t, \boldsymbol{y}_t\}$, where $\tilde{\boldsymbol{X}}_t := (\tilde{\boldsymbol{x}}_1^{\mathrm{T}}, \cdots, \tilde{\boldsymbol{x}}_t^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{y}_t = (g(\tilde{\boldsymbol{x}}_1) + w_1, \cdots, g(\tilde{\boldsymbol{x}}_t) + w_t)^{\mathrm{T}}$, the posterior mean and variance of the GP are computed as

$$\mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_t) = \boldsymbol{k}^{\mathrm{T}}(\tilde{\boldsymbol{x}}_t) \left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y}_t \tag{4}$$

$$\sigma(\tilde{\boldsymbol{x}}_t|\mathcal{D}_t) = k(\tilde{\boldsymbol{x}}_t, \tilde{\boldsymbol{x}}_t) - \boldsymbol{k}^{\mathrm{T}}(\tilde{\boldsymbol{x}}_t) \left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}(\tilde{\boldsymbol{x}}_t), \tag{5}$$

where $\boldsymbol{k}(\cdot) = [k(\tilde{\boldsymbol{x}}_1, \cdot), \cdots, k(\tilde{\boldsymbol{x}}_n, \cdot)]^{\mathrm{T}}$ and $\boldsymbol{K}$ is the covariance matrix with entries $K_{ij} = k(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j)$.
If $d_x > 1$, we model each entry of $\boldsymbol{g}(\cdot)$ using a separate GP, i.e.,

$$\boldsymbol{g}(\tilde{\boldsymbol{x}}_t) \sim \mathcal{N}\left(\boldsymbol{\mu}(\tilde{\boldsymbol{x}}_t|\mathcal{D}_t), \boldsymbol{\sigma}^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_t)\right), \tag{6}$$

where

$$\boldsymbol{\mu}_t(\tilde{\boldsymbol{x}}_t) := \left(\mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{1,t}) \ \cdots \ \mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{d_x,t})\right), \tag{7}$$

$$\boldsymbol{\sigma}_t^2(\tilde{\boldsymbol{x}}_t) := \operatorname{diag}\left(\sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{1,t}) \ \cdots \ \sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{d_x,t})\right) \tag{8}$$

and the data $\mathcal{D}_{i,t} = \{\tilde{\boldsymbol{X}}_t, \boldsymbol{y}_{i,t}\}$ used to model the i-th entry is chosen as $\tilde{\boldsymbol{X}}_t = (\tilde{\boldsymbol{x}}_1^{\mathrm{T}}, \cdots, \tilde{\boldsymbol{x}}_t^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{y}_{i,t} = (g_i(\tilde{\boldsymbol{x}}_1) + w_{1,i}, \cdots, g_i(\tilde{\boldsymbol{x}}_t) + w_{t,i})^{\mathrm{T}}$. This corresponds to assuming that the entries of $\boldsymbol{g}(\cdot)$ are conditonally independent.
The requirements for modelling (1) are summarized as follows:

Assumption 3. The entries of the unknown function $\boldsymbol{g}(\cdot)$ correspond to samples from a GP with mean $m \equiv 0$ and a known, bounded, and continuously differentiable kernel $k(\cdot, \cdot) \leq k_{\max}$, i.e., $g_i(\cdot) \sim \mathcal{GP}(0, k)$ holds for $i = 1, \cdots, d_x$.

The choice of kernel is typically carried out with some knowledge of the system at hand. Since universal kernels exist that are bounded and continuously differentiable, such as the Gaussian kernel (Micchelli et al., 2006), the required kernel characteristics pose few restrictions on $\boldsymbol{g}(\cdot)$.
For the described setting, the following result applies:

Proposition 1. Let Assumption 3 hold, and choose $k_{\max}$ accordingly. Moreover, let $\sigma_{\min} := \min_{i \in \{1, \cdots, d_x\}} \sigma_i$ be the smallest entry of the process noise covariance matrix $\boldsymbol{Q}$. Then, for any $i \in \{1, \cdots, d_x\}$ and a corresponding data set $\mathcal{D}_{i,t} = \{\tilde{\boldsymbol{X}}_t, \boldsymbol{y}_{i,t}\}$,

$$|\mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})| \leq \sqrt{d_x} \frac{k_{\max}}{\sigma_{\min}} \|\boldsymbol{y}_{i,t}\|_2, \tag{9}$$

$$\sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) \leq k_{\max}. \tag{10}$$

Proof. We begin by proving (9). Let $\lambda_{\min}^{-1}\left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)$ denote the smallest eigenvalue of $\left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)$. From (4), it follows that

$$\begin{aligned} |\mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})| &= \boldsymbol{k}^{\mathrm{T}}(\tilde{\boldsymbol{x}}_t) \left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y}_{i,t} \\ &\leq \|\boldsymbol{k}^{\mathrm{T}}\|_2 \lambda_{\min}^{-1}\left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right) \|\boldsymbol{y}_{i,t}\|_2 \\ &\leq k_{\max} \sqrt{d_x} \sigma_{\min}^{-1} \|\boldsymbol{y}_{i,t}\|_2, \end{aligned} \tag{11}$$

where the last inequality is due to the symmetrict positive semi-definiteness of $\boldsymbol{K}$, i.e., $\boldsymbol{K}$ only has nonnegative eigenvalues. This proves (10).
The inequality (10) also follows straightforwardly from the symmetric positive semi-definiteness of $\boldsymbol{K}$, i.e.,

$$\boldsymbol{k}^{\mathrm{T}}(\tilde{\boldsymbol{x}}_t) \left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}(\tilde{\boldsymbol{x}}_t) \geq 0.$$

The result then follows straightforwardly from (5). □

3.1 Multi-step ahead predictions

Under Assumption 3, the one step dynamics are given by

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\tilde{\boldsymbol{x}}_t) + \boldsymbol{\mu}_t(\tilde{\boldsymbol{x}}_t) + \boldsymbol{\sigma}_t(\tilde{\boldsymbol{x}}_t)\boldsymbol{\zeta}_t, \tag{12}$$

where $\boldsymbol{\zeta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. The mean $\mu(\cdot|\mathcal{D}_{i,t})$ and variance $\sigma^2(\cdot|\mathcal{D}_{i,t})$ for each entry i at time t are obtained by sampling a new data point from the GP and updating the measurement data with the resulting state, i.e.,

$$\boldsymbol{y}_{i,t+1} = \begin{pmatrix} \boldsymbol{y}_{i,t} \\ \mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) + \sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})\zeta_{i,t} \end{pmatrix}. \tag{13}$$

Hence, the closed-loop trajectory $\boldsymbol{X}$ is fully specified by T random samples $\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_{T-1}$, i.e., it is a function of $\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_{T-1}$. In fact, since $\mathcal{X} = \mathbb{R}^{d_x}$, the corresponding mapping is defined for all $\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_{T-1} \in \mathbb{R}^{d_x}$ and is bijective:

Lemma 2. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T) \in \mathcal{X}^T$. Then there exists a unique set of samples $\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_{T-1} \in \mathbb{R}^{d_x}$, such that (12) is satisfied.

Proof. Since $\boldsymbol{\sigma}_t(\tilde{\boldsymbol{x}}_t)$ is a diagonal matrix with positive diagonal entries, it is invertible. Hence,

$$\boldsymbol{\zeta}_t = \boldsymbol{\sigma}_t^{-1}(\tilde{\boldsymbol{x}}_t)\left(\boldsymbol{x}_{t+1} - \boldsymbol{f}(\tilde{\boldsymbol{x}}_t) - \boldsymbol{\mu}_t(\tilde{\boldsymbol{x}}_t)\right) \tag{14}$$

holds for all t, i.e., the samples $\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_{T-1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ that satisfy (12) for a fixed trajectory $\boldsymbol{X}$ are unique. □

We denote the corresponding bijection as $\boldsymbol{\Phi} : \mathcal{X}^T \mapsto \mathcal{X}^T$, $\boldsymbol{\Phi}(\boldsymbol{Z}) = \boldsymbol{X}$, where $\boldsymbol{Z} := (\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_{T-1})$. Moreover, the following holds:

Lemma 3. Let Assumption 3 hold. Then $\boldsymbol{\Phi}(\cdot)$ is differentiable and the corresponding Jacobian $J_{\boldsymbol{\Phi}}(\boldsymbol{Z})$) is continuously differentiable and nonsingular for all $\boldsymbol{Z} \in \mathcal{X}^T$.

Proof. Since the known function $\boldsymbol{f}(\cdot)$ and the kernel $k(\cdot,\cdot)$ are continuously differentiable, we can employ the chain-rule to differentiate the states $\boldsymbol{x}_t$, $t = 1, \cdots, T$ with respect to the samples $\boldsymbol{\zeta}_i$, $i = 0, \cdots, T-1$, and the resulting matrices $d\boldsymbol{x}_t/d\boldsymbol{\zeta}_i$ are continuous. Hence, $\boldsymbol{\Phi}(\cdot)$ is continuously differentiable. Moreover, the components of the Jacobian $J_{\boldsymbol{\Phi}}(\cdot)$ are given by

$$\frac{d\boldsymbol{x}_t}{d\boldsymbol{\zeta}_\tau} = \begin{cases} \mathbf{0}, & \tau \geq t \\ \operatorname{diag}(\boldsymbol{\sigma}_{t-1}^2(\tilde{\boldsymbol{x}}_{t-1})), & \tau = t-1 \\ \sum_{i=\tau+1}^{t-1} \frac{d\boldsymbol{x}_t}{d\tilde{\boldsymbol{x}}_i}\frac{d\tilde{\boldsymbol{x}}_i}{d\boldsymbol{\zeta}_\tau}, & \tau < t-1 \end{cases}. \tag{15}$$

Hence, the Jacobian $J_{\boldsymbol{\Phi}}(\cdot)$ is a lower triangular matrix with determinant

$$\det J_{\boldsymbol{\Phi}}(\boldsymbol{Z}) = \prod_{t=1}^{T}\prod_{i=1}^{d_x} \sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) \neq 0. \tag{16}$$

□

Corollary 4. Let Assumption 3 hold. Then the probability density function $p(\boldsymbol{X})$ of the trajectory $\boldsymbol{X}$ is continuously differentiable.

Proof. Due to Lemma 3, the inverse function theorem is applicable, and the inverse mapping $\boldsymbol{\Phi}^{-1}(\cdot)$ from the sample space to the trajectory space is continuously differentiable. Hence, by the change of variables formula, the probability density function of $\boldsymbol{X}$ is given by $p(\boldsymbol{X}) = p(\boldsymbol{Z})|\det(J_{\boldsymbol{\Phi}^{-1}}(\boldsymbol{Z}))|$, where det denotes the determinant operator and $J_{\boldsymbol{\Phi}^{-1}}$ is the Jacobian of $\boldsymbol{\Phi}^{-1}(\cdot)$. Since the samples $\boldsymbol{\zeta}_t$ are normally distributed, $p(\boldsymbol{Z})$ is continuously differentiable. As $\boldsymbol{\Phi}^{-1}(\cdot)$ is also continuously differentiable, this implies the continuity of $p(\boldsymbol{X})$. □

We introduce the expected system trajectory

$$\mathbb{E}[\boldsymbol{X}] = \left(\mathbb{E}[\boldsymbol{x}_1]^{\mathrm{T}}, \cdots, \mathbb{E}[\boldsymbol{x}_T]^{\mathrm{T}}\right)^{\mathrm{T}}, \tag{17}$$

where the mean of a multi-step ahead prediction $\mathbb{E}[\boldsymbol{x}_t]$ at an arbitrary time step t is given by

$$\mathbb{E}[\boldsymbol{x}_t] = \int_{\tilde{\mathcal{X}}^t} \boldsymbol{x}_t \prod_{\tau=0}^{t-1} p(\boldsymbol{\zeta}_\tau) d\boldsymbol{\zeta}_\tau. \tag{18}$$

Here $p(\cdot)$ is a probability distribution and the state $\boldsymbol{x}_t$ is computed recursively using (12). The variance of the trajectory (19) is then given by

$$\mathbb{V}[\boldsymbol{X}] = \int_{\tilde{\mathcal{X}}^T} (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^{\mathrm{T}} \prod_{t=0}^{T-1} p(\boldsymbol{\zeta}_t) d\boldsymbol{\zeta}_t. \tag{19}$$

We now prove the existence of (18) and (19). For this the following result is essential.

Lemma 5. (Tsagris et al. (2014)). Let $\zeta \sim \mathcal{N}(0,1)$ be a random variable. Then

$$\mathbb{E}[|\zeta|^N] = \int_{\mathbb{R}} |\zeta|^N p(\zeta) d\zeta \tag{20}$$

is finite-valued for all $N \in \mathbb{N}$.

Furthermore, the Euclidean norm of the measurement data vectors $\boldsymbol{y}_{i,t}$ are bounded by the samples $\boldsymbol{\zeta}_t$ as shown in the following lemma:

Lemma 6. Let Assumption 3 hold, choose $i \in \{1, \cdots, d_x\}$, and let $\boldsymbol{y}_{i,t+1}$ be given as in (13). Then

$$\|\boldsymbol{y}_{i,t+1}\|_2 \leq \sum_{j=0}^{t} (1 + \sqrt{d_x}\frac{k_{\max}}{\sigma_{\min}})^{t-j} k_{\max}|\zeta_{i,j}| \tag{21}$$

holds.

Proof. Due to (13), Proposition 1, and the Cauchy-Schwarz inequality,

$$\begin{aligned} \|\boldsymbol{y}_{i,t+1}\|_2 &= \sqrt{\|\boldsymbol{y}_{i,t}\|_2^2 + (\mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) + \sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})\zeta_{i,t})^2} \\ &\leq \|\boldsymbol{y}_{i,t}\|_2 + |\mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})| + |\sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})\zeta_{i,t}| \\ &\leq (1 + \sqrt{d_x}\frac{k_{\max}}{\sigma_{\min}})\|\boldsymbol{y}_{i,t}\|_2 + k_{\max}|\zeta_{i,t}|. \end{aligned} \tag{22}$$

Applying (22) t times yields

$$\begin{aligned} \|\boldsymbol{y}_{i,t+1}\|_2 \leq& (1 + \sqrt{d_x}\frac{k_{\max}}{\sigma_{\min}})^{(t+1)}\|\boldsymbol{y}_{i,0}\|_2 \\ &+ \sum_{j=0}^{t} (1 + \sqrt{d_x}\frac{k_{\max}}{\sigma_{\min}})^{t-j} k_{\max}|\zeta_{i,j}| \\ =& \sum_{j=0}^{t} (1 + \sqrt{d_x}\frac{k_{\max}}{\sigma_{\min}})^{t-j} k_{\max}|\zeta_{i,j}|. \end{aligned} \tag{23}$$

Here the last equality holds because no data is available at the beginning of the simulation. □

Hence, the growth of the data vector depends only on the random samples $\boldsymbol{\zeta}_t$, and not on the trajectory itself. Note that, due to Proposition 1, this directly implies

$$\mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) \leq \sqrt{d_x}\frac{k_{\max}^2}{\sigma_{\min}} \left(\sum_{j=0}^{t} (1 + \sqrt{d_x}\frac{k_{\max}}{\sigma_{\min}})^{t-j} |\zeta_{i,j}| \right) \tag{24}$$

for all $\tilde{\boldsymbol{x}}_t \in \tilde{\mathcal{X}}$. We then obtain the following result:

Lemma 7. Let $x_{t+1,i}$ denote the i-th entry of (12), and let Assumptions 1-3 hold. Then, for every $t \in \{1, \cdots, T\}$, there exists a polynomial function $\pi_t : \mathbb{R}^{d_x \times t} \mapsto \mathbb{R}$, such that

$$|x_{t+1,i}| \leq \pi_t \left(\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_t\right). \tag{25}$$

Proof. Choose the polynomial function $\pi(\cdot)$ as in Assumption 1. Then, due to Lemma 6, there exist positive constants $\tilde{a}_0, \cdots, \tilde{a}_T$, such that for all $t \in \{1, \cdots, T\}$ and all $i \in \{1, \cdots, d_x\}$,

$$\begin{aligned}
|x_{t+1,i}| =& |f_i(\tilde{\boldsymbol{x}}_t) + \mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) + \sigma(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})\zeta_{i,t}| \\
\leq& |\pi(\|\tilde{\boldsymbol{x}}_t\|_2) + \mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) + \sigma(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})\zeta_{i,t}| \\
\leq& \pi(\|\tilde{\boldsymbol{x}}_t\|_2) + \sum_{j=0}^{t} \tilde{a}_j |\zeta_{i,j}| \\
\leq& \pi\left(\sum_{j=1}^{d_x} |x_{t,j}| + \sum_{j=1}^{d_u} |u_{t,j}|\right) + \sum_{j=0}^{t} \tilde{a}_j |\zeta_{i,j}| \\
\leq& \pi\left(\sum_{j=1}^{d_x} |x_{t,j}| + d_u u_{\max}\right) + \sum_{j=0}^{t} \tilde{a}_j |\zeta_{i,j}|
\end{aligned} \tag{26}$$

holds, where $u_{\max}$ is chosen as in Assumption 2. Applying (26) recursively yields the desired result. □

We are now able to prove the existence of (17) and (19).

Lemma 8. The expected value $\mathbb{E}[\boldsymbol{X}]$ and variance $\mathbb{V}[\boldsymbol{X}]$ of the system trajectory as given by (17) and (19) exist and are bounded.

Proof. Due to Lemma 7, the integrands of (18) are bounded by $\pi_t\left(\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_t\right)$, where $\pi_t(\cdot)$ is a polynomial function chosen as in Lemma 7. Moreover, due to Lemma 5 and the fact that $\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_t$ are independent, the integral

$$\int_{\tilde{\mathcal{X}}^T} \pi_t\left(\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_t\right) \prod_{t=0}^{T-1} p(\boldsymbol{\zeta}_t) d\boldsymbol{\zeta}_t \tag{27}$$

is finite-valued. Hence, $\mathbb{E}[\boldsymbol{x}_t]$ is finite-valued. Similarly, the entries of the integrand of (19) satisfy

$$\begin{aligned}
&\left((\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^{\mathrm{T}}\right)_{i,j} \\
\leq& \left(\pi_t\left(\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_t\right) + \|\mathbb{E}[\boldsymbol{x}_t]\|_\infty\right)^2,
\end{aligned} \tag{28}$$

where $i, j \in \{1, \cdots, d_x\}$ denotes the i-th row and j-th column and $\|\cdot\|_\infty$ corresponds to the maximum norm. Since this corresponds to a polynomial function, Lemma 5 implies that the entries of $\mathbb{V}[\boldsymbol{X}]$ are finite-valued. □

Moreover, the variance $\mathbb{V}[\boldsymbol{X}]$ satisfies the following property:

Proposition 1. The trajectory variance $\mathbb{V}[\boldsymbol{X}]$ as given by (19) is symmetric positive definite.

Proof. We prove the result by contradiction. Assume $\mathbb{V}[\boldsymbol{X}]$ is not symmetric-positive definite. Due to Lemma 8, $\mathbb{V}[\boldsymbol{X}]$ is finite valued. It is easy to see from (19) that $\mathbb{V}[\boldsymbol{X}]$ must be symmetric positive-semidefinite. Hence, there exists an $\boldsymbol{\alpha} \in \mathbb{R}^{d_x T}$, $\boldsymbol{\alpha} \neq \boldsymbol{0}$, such that $\boldsymbol{\alpha}^{\mathrm{T}} \mathbb{V}[\boldsymbol{X}] \boldsymbol{\alpha} = 0$. Due to (19) and the continuity of $\mathcal{X}$ with respect to the samples $\boldsymbol{\zeta}_0, \cdots, \boldsymbol{\zeta}_{T-1}$, this implies

$$\boldsymbol{\alpha}^{\mathrm{T}} (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^{\mathrm{T}} \boldsymbol{\alpha} = 0 \tag{29}$$

holds for all $\boldsymbol{X} \in \mathcal{X}^T$. This in turn holds only if $(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^{\mathrm{T}} \boldsymbol{\alpha} = 0$, i.e.,

$$\sum_{t=0}^{T-1} \sum_{i=1}^{d_x} \alpha_{(td_x+i)} \left(f(\tilde{\boldsymbol{x}}_t) + \mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) + \sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})\zeta_{i,t}\right) = 0. \tag{30}$$

Let $\nu d_x + \rho := J = \max_j j$, $\alpha_j \neq 0$ be the highest index corresponding to a nonzero entry of $\boldsymbol{\alpha}$, where ν and ρ are the corresponding time step and dimension, respectively. We rewrite (30) as

$$\sum_{t=0}^{\nu} \sum_{i=1}^{d_x} \alpha_{(td_x+i)} \left(f(\tilde{\boldsymbol{x}}_t) + \mu(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) + \sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t})\zeta_{i,t}\right) = 0. \tag{31}$$

Since varying the value of $\zeta_{\nu,\rho}$ does not affect the states up to time step ν, i.e., $\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_\nu$, (31) implies

$$\sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{\nu,\rho})\zeta_{\nu,\rho} = 0, \tag{32}$$

which is a contradiction due to $\sigma^2(\tilde{\boldsymbol{x}}_t|\mathcal{D}_{i,t}) > 0$ for every i and t. Hence, $\mathbb{V}[\boldsymbol{X}]$ is symmetric positive definite. □

4. MONTE CARLO SIMULATION

Computing (18) and (19) analytically is generally intractable. Hence, we resort to a Monte Carlo simulation to approximate (18) and (19), i.e.,

$$\mathbb{E}[\boldsymbol{x}_t] \approx \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{x}_t^{(m)}, \tag{33}$$

with

$$\boldsymbol{x}_{t+1}^{(m)} = \boldsymbol{f}(\tilde{\boldsymbol{x}}_t^{(m)}) + \boldsymbol{\mu}_t^{(m)}(\tilde{\boldsymbol{x}}_t^{(m)}) + \boldsymbol{\sigma}_t^{(m)}(\tilde{\boldsymbol{x}}_t^{(m)})\boldsymbol{\zeta}_t^{(m)}, \tag{34}$$

where the superscript (i) refers to the i-th Monte Carlo simulation, and M is the number of Monte Carlo simulations. We define a sample trajectory as

$$\boldsymbol{X}^{(m)} := \left(\left(\boldsymbol{x}_1^{(m)}\right)^{\mathrm{T}}, \cdots, \left(\boldsymbol{x}_T^{(m)}\right)^{\mathrm{T}}\right)^{\mathrm{T}}. \tag{35}$$

The estimated mean and unbiased sample variance of a trajectory obtained by the Monte Carlo simulation are given by

$$\bar{\boldsymbol{X}}_M := \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{X}^{(i)}, \tag{36}$$

$$\bar{\boldsymbol{\Sigma}}_M := \frac{1}{M-1} \sum_{i=1}^{M} \left(\boldsymbol{X}^{(i)} - \bar{\boldsymbol{X}}_M\right)\left(\boldsymbol{X}^{(i)} - \bar{\boldsymbol{X}}_M\right)^{\mathrm{T}}, \tag{37}$$

and are employed to approximate the true mean and variance of the trajectory. We now prove the following:

Lemma 9. Let Assumptions 1-3 hold, let $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(M)}$ be M sample trajectories obtained using (12) and (35) and let $\bar{\boldsymbol{X}}_M$ be given as in (36). Assume $M \geq Td_x$ and $Td_x > 1$ and let $\boldsymbol{X}^{(m_1)}, \cdots, \boldsymbol{X}^{(m_{d_x T})}$, $m_1, \cdots, m_{d_x T} \in \{1, \cdots, M\}$ be $d_x T$ arbitrary sample trajectories. Then, with probability 1, $\boldsymbol{X}^{(m_1)}, \cdots, \boldsymbol{X}^{(m_{d_x T})}$ are linearly independent and $\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_M \neq \boldsymbol{0}$ holds for all m.

Proof. Assume the contrary is true and assume, without loss of generality, $m_1 = 1, \cdots, m_{d_x T} = d_x T$. Then there exist scalars $\alpha_1, \cdots, \alpha_{d_x T}$ and an $m \in \{1, \cdots, d_x T\}$, such that

$$\sum_{m=1}^{d_x T - 1} \alpha_i \boldsymbol{X}^{(m)} = \alpha_{T d_x} \boldsymbol{X}^{(T d_x)}, \tag{38}$$

where we assume $\boldsymbol{X}^{(Td_x)} \neq \boldsymbol{0}$ and $\alpha_{Td_x} \neq 0$ without loss of generality. Define

$$\mathcal{H} := \left\{ \boldsymbol{X} \in \mathcal{X} \;\middle|\; \boldsymbol{X} = \sum_{m=1}^{d_x T - 1} \alpha_m \boldsymbol{X}^{(m)},\ \alpha_m \in \mathbb{R} \right\}. \quad (39)$$

Note that $\mathcal{H}$ is a hyperplane in $\mathbb{R}^{d_x T}$, hence has measure zero. Due to Corollary 4, the probability density function $p(\boldsymbol{X})$ is continuous. The probability that (38) holds for some $\alpha_{d_x T} \in \mathbb{R}$ is then given by

$$\mathbb{P}\left(\boldsymbol{X}^{(Td_x)} \in \mathcal{H}\right) = \int_{\mathcal{H}} p(\boldsymbol{X}) d\boldsymbol{X} = 0, \quad (40)$$

where the last equality because $p(\boldsymbol{X})$ is continuous and $\mathcal{H}$ has measure zero. If $\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_M = \boldsymbol{0}$ holds for some m, due to $M > 1$, then this implies

$$\boldsymbol{X}^{(m)} + \frac{M-1}{M^2} \sum_{\substack{j=1 \\ j \neq i}}^{M} \boldsymbol{X}^{(j)} =: \boldsymbol{X}^{(m)} + \bar{\boldsymbol{X}}_{M \setminus m} = \boldsymbol{0}. \quad (41)$$

This holds with probability

$$\mathbb{P}\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_M = \boldsymbol{0}\right) = \int_{\bar{\boldsymbol{X}}_{M \setminus m}} p(\boldsymbol{X}) d\boldsymbol{X} = 0 \quad (42)$$

because $\bar{\boldsymbol{X}}_{M \setminus m}$ is a vector, hence has measure zero. Applying the union bound to both events yields

$$\begin{aligned} &\mathbb{P}\left(\boldsymbol{X}^{(Td_x)} \in \mathcal{H} \bigcup \boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_M = \boldsymbol{0}\right) \\ &\leq \mathbb{P}\left(\boldsymbol{X}^{(Td_x)} \in \mathcal{H}\right) + \mathbb{P}\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_M = \boldsymbol{0}\right) = 0. \end{aligned} \quad (43)$$

□

Corollary 1. Let Assumptions 1-3 hold and let $\bar{\boldsymbol{\Sigma}}_M$ be given as in (37) with $M \geq d_x T$. Then $\bar{\boldsymbol{\Sigma}}_M$ is invertible with probability 1.

Proof. Let $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(M)}$ be the samples used to compute $\bar{\boldsymbol{\Sigma}}_M$ and consider the first $d_x T$ arbitrary samples $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(d_x T)}$. Note that if $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(d_x T)}$ are linearly independent and $\boldsymbol{X}^{(i)} - \bar{\boldsymbol{X}}_M \neq \boldsymbol{0}$ for all i, then the difference vectors $\left(\boldsymbol{X}^{(n_1)} - \bar{\boldsymbol{X}}_M\right), \cdots, \left(\boldsymbol{X}^{(n_{d_x T})} - \bar{\boldsymbol{X}}_M\right)$ are linearly independent. Due to Lemma 9 this holds with probability 1. Hence, for any $\boldsymbol{\alpha} \in \mathbb{R}^{d_x T]}$,

$$\begin{aligned} \boldsymbol{\alpha} \bar{\boldsymbol{\Sigma}}_M \boldsymbol{\alpha}^{\mathrm{T}} &= \frac{1}{M-1} \sum_{m=1}^{M} \left(\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_M \right)^{\mathrm{T}} \boldsymbol{\alpha} \right)^2 \\ &\geq \frac{1}{M-1} \sum_{m=1}^{d_x T} \left(\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_M \right)^{\mathrm{T}} \boldsymbol{\alpha} \right)^2 > 0 \end{aligned} \quad (44)$$

holds with probability 1, i.e., $\bar{\boldsymbol{\Sigma}}_M$ is symmetric positive definite, which implies that $\bar{\boldsymbol{\Sigma}}_M$ is invertible. □

4.1 Choosing sample size

We now give a theoretical analysis of the Monte Carlo method and provide a confidence region for the trajectory of the real system. Since the unbiased sample variance $\bar{\boldsymbol{\Sigma}}_M$ is invertible, it produces a confidence ellipsoid as follows:

Proposition 10. (Stellato et al. (2017)). Let $M > d_x$. Given $M+1$ iid samples $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(M)}, \boldsymbol{X}$, if $\bar{\boldsymbol{\Sigma}}_M$ is nonsingular, then for all $\eta > 0$, it holds that

Algorithm 1 Confidence Regions for Anticipation of Learning in Multi-Step Predictions

Input: $\boldsymbol{x}_0$, δ, η, such that $\eta \geq \sqrt{d_x/\delta}$
  Solve $\tilde{M}^2(\delta\eta^2 - d_x) + 1 - \tilde{M}\eta^2 = 0$ for $\tilde{M}$
  Set $M \geq \tilde{M}$, $M \in \mathbb{N}$
  for $m = 1, \cdots, M$ **do**
    for $t = 1, \cdots, T$ **do**
      Compute $\tilde{\boldsymbol{x}}_t^{(m)} = \left(\left(\boldsymbol{x}_t^{(m)} \right)^{\mathrm{T}} \left(\boldsymbol{u}_t^{(m)} \right)^{\mathrm{T}} \right)^{\mathrm{T}}$
      Sample $\boldsymbol{\zeta}_t^{(m)} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
      Compute $\boldsymbol{\mu}_t^{(m)}(\tilde{\boldsymbol{x}}_{t-1}^{(m)})$, $\boldsymbol{\sigma}_t^{(m)}(\tilde{\boldsymbol{x}}_{t-1}^{(m)})$ as in (7), (8)
      Compute $\boldsymbol{x}_{t+1}^{(m)}$ by solving (12)
    end for
    Set $\boldsymbol{X}^{(m)} = \left(\left(\boldsymbol{x}_1^{(m)} \right)^{\mathrm{T}}, \cdots, \left(\boldsymbol{x}_T^{(m)} \right)^{\mathrm{T}} \right)^{\mathrm{T}}$
  end for
  Compute $\bar{\boldsymbol{\Sigma}}_M, \bar{\boldsymbol{X}}_M$ as in (36) and (37)
Output: $\bar{\boldsymbol{\Sigma}}_M, \bar{\boldsymbol{X}}_M$

$$\begin{aligned} &\mathbb{P}\left(\left(\boldsymbol{X} - \bar{\boldsymbol{X}}_M \right)^{\mathrm{T}} \bar{\boldsymbol{\Sigma}}_M^{-1} \left(\boldsymbol{X} - \bar{\boldsymbol{X}}_M \right) \geq \eta^2 \right) \\ &\leq \min\left\{ 1, \frac{d_x(M^2 - 1 + M\eta^2)}{M^2 \eta^2} \right\} \end{aligned} \quad (45)$$

Note that, because $T \geq 1$, the right-hand side of (45) is greater or equal to $d_x \eta^{-2}$, which effectively imposes restrictions on the radius of the confidence regions associated with high confidence levels. In particular, $\eta \geq \sqrt{d_x}$ must be chosen in order to obtain meaningful confidence regions. These results are summarized in Algorithm 1. We now state our main result:

Theorem 11. Let Assumptions 1-3 hold. Choose δ and η, such that $\eta \geq \sqrt{d_x/\delta}$ and let $\bar{\boldsymbol{\Sigma}}_M, \bar{\boldsymbol{X}}_M$ be given by Algorithm 1. Then $\bar{\boldsymbol{\Sigma}}_M$ is invertible with probability 1 and the ellipsoid

$$\mathcal{S} = \left\{ \boldsymbol{X} \in \mathcal{X}^T \mid (\boldsymbol{X} - \bar{\boldsymbol{X}}_M) \bar{\boldsymbol{\Sigma}}_M^{-1} (\boldsymbol{X} - \bar{\boldsymbol{X}}_M) \right\} \quad (46)$$

corresponds to a $1 - \delta$ a confidence bound for a sample trajectory $\boldsymbol{X}$ of the true system (1).

Proof. This follows straightforwardly from Corollary 1 and Proposition 10.

5. NUMERICAL SIMULATION

We evaluate the performance of the proposed approach in a numerical simulation of a cart-pole. The pendulum dynamics are given by

$$(m_c + m_p)\ddot{x} + m_p l \ddot{\theta} \cos(\theta) - m_p l \dot{\theta}^2 \sin(\theta) = u \quad (47)$$

$$m_p l^2 \ddot{\theta} + m_p g l \sin(\theta) = -m_p l \ddot{x} \cos(\theta), \quad (48)$$

where x denotes the cart's position, θ is the pendulum's angle, and u is the horizontal force applied to the cart. The cart's and pendulum's masses are given by $m_c = 0.5\,\mathrm{kg}$ and $m_p = 0.5\,\mathrm{kg}$, respectively The parameter l denotes the pendulum's length. The discrete-time dynamical system form (1) is obtained by sampling the continuous system (47) every 0.05 seconds. The prior model $\boldsymbol{f}(\cdot)$ corresponds to the linearized dynamics around the origin, where incorrect masses $m_c = m_p = 0.4\,\mathrm{kg}$ are assumed. Additionally, we consider discrete process noise $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$,

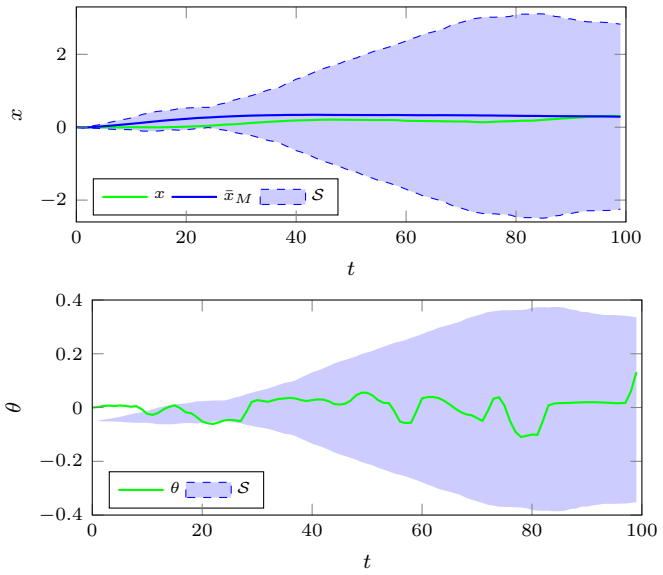

Fig. 1. Expected trajectory, confidence region of cart position x and pendulum angle θ for $M = 100$, $\eta = 10$, and simulation results of the real closed-loop system.

with $\boldsymbol{Q} = 0.01\boldsymbol{I}$. The cart position must stay within the boundaries $-5 \leq x \leq 5$, and we assume that the pendulum hits the floor once it gets to the horizontal position, which corresponds to the contstraints $-\pi/2 \leq \theta \leq \pi/2$. We employ the safe learning-based control law from Koller et al. (2018) to safely move the cart position from $x = 0$ to $x = 0.3$ while simultaneously learning the system dynamics. The control law employs model predictive control and exploits the knowledge about the system dynamics to minimize the 3-step cost function

$$\sum_{t=1}^{3}(x_t - 0.3)^2 + 0.1\dot{x}_t^2 + 0.1\theta_t^2 + 0.1\dot{\theta}_t^2. \tag{49}$$

Simultaneously, it aims to guarantee that safety constraints are not violated with high probability, and that a safe region is always reachable, where a safe backup LQR controller stabilizes the system.

We consider the performance of the closed-loop system over $T = 100$ steps, which corresponds to 5 seconds. We employ $M = 100$ Monte Carlo simulations and choose $\eta = 10$, which yields a confidence level of $1-\delta \approx 0.955$. The expected trajectory and confidence region of the position x and angle θ are shown in Figure 1. Moreover, a simulated trajectory of the true system is shown. As can be seen, the trajectory of the true simulation lies entirely within the confidence region, except for very small deviations at the beginning of the simulation. The confidence region is larger towards the end of the simulation due to the propagation of system uncertainty. However, a slight decline in size in the portion of the confidence region corresponding to the position x takes place after $t = 80$ time steps. This is because the system dynamics are learned during the simulation such that the posterior variance of the Gaussian process decreases in the proximity of the reference. The decline in model uncertainty in turn is expected to lead to a successful stabilization of the system around $x = 0.3$.

6. CONCLUSION

An algorithm for computing confidence regions of multi-step ahead predictions of closed-loop learning-based control systems is presented. We show that the algorithm is applicable almost surely, and that the corresponding confidence region is correct with high probability. In a numerical simulation of a cart-pole system, the confidence region is shown to contain a trajectory of the real system entirely, except for small outliers. Moreover, the effect of learning is shown to lead to an expected decline in system uncertainty over the simulation horizon.

REFERENCES

Beckers, T., Kulic, D., and Hirche, S. (2019). Stable Gaussian process based tracking control of euler-langrange systems. *Automatica*, 103(23), 390–397.

Berkenkamp, F. and Schoellig, A.P. (2015). Safe and robust learning control with Gaussian processes. In *IEEE European Control Conference*, 2496–2501.

Berkenkamp, F., Turchetta, M., Schoellig, A., and Krause, A. (2017). Safe model-based reinforcement learning with stability guarantees. In *Advances in Neural Information Processing Systems*, 908–919.

Capone, A. and Hirche, S. (2019). Backstepping for partially unknown nonlinear systems using Gaussian processes. *IEEE Control Systems Letters*, 3, 416–421.

Chowdhary, G., Kingravi, H.A., How, J.P., Vela, P.A., et al. (2015). Bayesian nonparametric adaptive control using Gaussian processes. *IEEE Trans. Neural Netw. Learning Syst.*, 26(3), 537–550.

Deisenroth, M.P., Fox, D., and Rasmussen, C.E. (2015). Gaussian processes for data-efficient learning in robotics and control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2), 408–423.

Kamthe, S. and Deisenroth, M. (2018). Data-efficient reinforcement learning with probabilistic model predictive control. In *International Conference on Artificial Intelligence and Statistics*, 1701–1710.

Koller, T., Berkenkamp, F., Turchetta, M., and Krause, A. (2018). Learning-based model predictive control for safe exploration. In *IEEE Conference on Decision and Control*, 6059–6066.

Maiworm, M., Limon, D., Manzano, J., and Findeisen, R. (2018). Stability of Gaussian process learning based output feedback model predictive control. In *IFAC-PapersOnLine*, 455–461.

Micchelli, C.A., Xu, Y., and Zhang, H. (2006). Universal kernels. *Journal of Machine Learning Research*, 7(Dec), 2651–2667.

Rasmussen, C.E. and Williams, C.K. (2006). Gaussian processes for machine learning. 2006. *The MIT Press, Cambridge, MA, USA*.

Stellato, B., Van Parys, B.P., and Goulart, P.J. (2017). Multivariate chebyshev inequality with estimated mean and variance. *The American Statistician*, 71(2), 123–127.

Tsagris, M., Beneki, C., and Hassani, H. (2014). On the folded normal distribution. *Mathematics*, 2(1), 12–28.

Umlauft, J., Beckers, T., Kimmel, M., and Hirche, S. (2017). Feedback linearization using Gaussian processes. In *2017 IEEE 56th Annual Conference on Decision and Control*, 5249–5255.