Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Confidence Regions for Predictions of Online 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 of closed-loop system trajectories under an online learning-based control law. We employ a sampling-based approximation and exploit system properties to prove that the computed confidence regions are correct with high probability. In a numerical simulation, we show that the proposed approach accurately predicts correct confidence regions.

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 systems cannot be modeled using first-principles. In such settings, model uncertainty is generally high, and conventional control techniques yield poor performance. A promising technique to address these issues is online learning-based control, i.e., control laws that learn with measurements collected online. This type of control law operates most often by either learning a model (Umlauft and Hirche, 2019; Chowdhary et al., 2015; Kamthe and Deisenroth, 2018), or by directly learning a control policy (Bakker et al., 2006; Berkenkamp et al., 2017). Even though online learning-based tools fall under the broader category of adaptive control, they are often too intricate to be analyzed with classical control techniques. As such, most online learning-based methods yield no theoretical guarantees (Kamthe and Deisenroth, 2018; Bakker et al., 2006), with few exceptions (Umlauft and Hirche, 2019; Berkenkamp et al., 2017). However, most of the theoretical results obtained so far pertain to stability of the closed-loop system. This does not give insight into how the control performance evolves due to learning.

Quantifying how learning affects long-term performance is closely related to dual control (Bar-Shalom and Tse, 1974). The goal of dual control is to simultaneously address system identification and a control objective. However, this often only be accomplished approximately, yielding no theoretical guarantees (Mesbah, 2018).

In spite of the intricacy of online learning-based control, it is generally still possible to predict performance if the underlying model uncertainty is well understood. Gaussian processes (GP) have been increasingly employed to quantify model uncertainty, enabling both rigorous analyses and good practical results (Umlauft and Hirche, 2019; Capone and Hirche, 2019; Chowdhary et al., 2015; Deisenroth et al., 2015; Beckers et al., 2019). Despite their widespread use, GPs have seldom been employed to predict how online learning affects closed-loop performance.

In this paper, we introduce a multi-step ahead prediction algorithm for learning-based control laws. Our approach computes confidence regions for the closed-loop system trajectory, determining how learning influences the control performance over a long time horizon. We employ nonrestrictive assumptions and prove useful properties of Gaussian process models, which in turn are employed to obtain confidence regions. In a numerical Monte Carlo simulation of the cart-pole balancing problem, we show that the computed confidence region holds with high probability.

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 brief conclusion.

Notation Let $\mathbb{N}$ denote the natural numbers, $\mathbb{R}$ the real numbers, and $\mathbb{R}_+$ the nonnegative real numbers. We use $\mathbb{P}(\cdot)$ to denote the probability of an event, $\mathbb{E}[\cdot]$ to denote the expected value operator, and $\mathbb{V}[\cdot]$ to denote the variance of an event. For $m, n, d \in \mathbb{N}$ and a matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$, $[\boldsymbol{A}]_{m,n}$ denotes the entry in the m-th row and n-th column of $\boldsymbol{A}$. The determinant of $\boldsymbol{A}$ is denoted by $\det(\boldsymbol{A})$. The Euclidean norm of a vector $\boldsymbol{a} \in \mathbb{R}^d$ is denoted by $\|\boldsymbol{a}\|_2$, its maximum norm by $\|\boldsymbol{a}\|_\infty$. For a continuously differentiable function $\boldsymbol{h} : \mathbb{R}^m \to \mathbb{R}^n$, $J_{\boldsymbol{h}}$ denotes the Jacobian of $\boldsymbol{h}(\cdot)$.

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 $\boldsymbol{f} : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^{d_x}$ corresponds to the known component of the system dynamics, whereas $\boldsymbol{g} : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^{d_x}$ is unknown. The system is perturbed by independent and identically distributed (iid) process noise $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$, with $\boldsymbol{Q} = \operatorname{diag}(\sigma_1, \ldots, \sigma_{d_x}) \in \mathbb{R}_+^{d_x}$. The augmented state vector $\tilde{\boldsymbol{x}}_t = (\boldsymbol{x}_t^{\mathrm{T}}, \boldsymbol{u}_t^{\mathrm{T}})^{\mathrm{T}}$ concatenates the state and input, and is introduced for simplicity of exposition.

We consider a control law that learns online using measurements of the system dynamics. This is formally expressed as

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

where the functions $\boldsymbol{u}_t : \mathcal{X}^{t+1} \to \mathcal{U}$ depend on the current state $\boldsymbol{x}_t$ and the measurement data collected up to time t. This definition applies to most online learning-based control laws.

Our goal is to determine a confidence region $\mathcal{S} \subseteq \mathcal{X}^T$ for a T step closed-loop trajectory $\boldsymbol{X}_T := (\boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{X}^T$ of the true system (1), such that

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

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

In this paper, we restrict ourselves to the case where $\boldsymbol{f}(\cdot)$ is continuously differentiable and exhibits at most polynomial growth. This is detailed in the following.

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 $\pi : \mathbb{R} \to \mathbb{R}$, i.e., $f_i(\tilde{\boldsymbol{x}}) \leq \pi(\|\tilde{\boldsymbol{x}}\|_2) \; \forall \; \tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}$.

This is not a very restrictive assumption, as it holds for many physical systems. For example, robotic and electrical systems are often described by polynomial functions.

In addition, we consider a control laws that satisfy the following assumption.

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, e.g., due to safety or physical limitations.

3. PROBABILISTIC MODEL

We now introduce Gaussian process (GP) models, and illustrate how they are employed 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 $m : \tilde{\mathcal{X}} \to \mathbb{R}$ and a symmetric positive definite *kernel* function $k : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \to \mathbb{R}$. 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 $\boldsymbol{g}(\cdot)$ is continuous, *universal* kernels are often employed, as they uniformly approximate any continuous function in a closed subset of $\tilde{\mathcal{X}}$ (Micchelli et al., 2006).

In the following, we only show how to recursively draw samples from a GP, as opposed to training a GP. However, the corresponding equations are identical up to a noise term (Rasmussen and Williams, 2006). We begin by introducing GPs for the one-dimensional case, i.e., $d_x = 1$. Consider a set of sample augmented trajectory and GP evaluation pairs $\mathcal{T}_t^s = \{\tilde{\boldsymbol{X}}_t^s, \boldsymbol{y}_t^s\}$, where $t \geq 0$, $\tilde{\boldsymbol{X}}_t^s := ((\tilde{\boldsymbol{x}}_1^s)^{\mathrm{T}}, \ldots, (\tilde{\boldsymbol{x}}_t^s)^{\mathrm{T}})^{\mathrm{T}}$ is a sample trajectory and $\boldsymbol{y}_t^s := (g^s(\tilde{\boldsymbol{x}}_1^s), \ldots, g^s(\tilde{\boldsymbol{x}}_t^s))^{\mathrm{T}}$ concatenates sample evaluations from the GP distribution over $g(\cdot)$. With a slight abuse of notation, we use the superscript s here to distinguish sampled trajectories from trajectories of the true system (1). The posterior mean and variance of a GP sample at time step $t+1$ are computed as

$$\mu(\tilde{\boldsymbol{x}}_t^s | \mathcal{T}_t^s) = \boldsymbol{k}^{\mathrm{T}}(\tilde{\boldsymbol{x}}_t^s) \boldsymbol{K}^{-1} \boldsymbol{y}_t^s \tag{4}$$

$$\sigma(\tilde{\boldsymbol{x}}_t^s | \mathcal{T}_t^s) = k(\tilde{\boldsymbol{x}}_t^s, \tilde{\boldsymbol{x}}_t^s) - \boldsymbol{k}^{\mathrm{T}}(\tilde{\boldsymbol{x}}_t^s) \boldsymbol{K}^{-1} \boldsymbol{k}(\tilde{\boldsymbol{x}}_t^s), \tag{5}$$

where $\boldsymbol{k}(\cdot) = [k(\tilde{\boldsymbol{x}}_1^s, \cdot), \ldots, k(\tilde{\boldsymbol{x}}_n^s, \cdot)]^{\mathrm{T}}$ and $\boldsymbol{K}$ is the covariance matrix with entries $[\boldsymbol{K}]_{ij} = k(\tilde{\boldsymbol{x}}_i^s, \tilde{\boldsymbol{x}}_j^s)$.

If $d_x > 1$, we model each entry of $\boldsymbol{g}(\cdot)$ using a separate GP, i.e.,

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

where

$$\boldsymbol{\mu}_t(\tilde{\boldsymbol{x}}_t^s) := \left(\mu(\tilde{\boldsymbol{x}}_t^s | \mathcal{T}_{1,t}^s) \; \ldots \; \mu(\tilde{\boldsymbol{x}}_t^s | \mathcal{T}_{d_x,t}^s)\right), \tag{7}$$

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

and the sample trajectory $\mathcal{T}_{i,t}^s = \{\tilde{\boldsymbol{X}}_t^s, \boldsymbol{y}_{i,t}^s\}$ used to model the i-th entry of the unknown function $\boldsymbol{g}(\cdot)$ is given by $\tilde{\boldsymbol{X}}_t^s = (\tilde{\boldsymbol{x}}_1^s, \ldots, \tilde{\boldsymbol{x}}_t^s)^{\mathrm{T}}$ and $\boldsymbol{y}_{i,t}^s = (g_i^s(\tilde{\boldsymbol{x}}_1^s), \ldots, g_i^s(\tilde{\boldsymbol{x}}_t^s))^{\mathrm{T}}$. This corresponds to assuming that the entries of $\boldsymbol{g}(\cdot)$ are conditionally independent.

We assume that the function $\boldsymbol{g}(\cdot)$ is drawn from a GP with known kernel $k(\cdot, \cdot)$. Furthermore, we assume that $k(\cdot, \cdot)$ exhibits some commonly found kernel properties. This is specified in the following.

Assumption 3. The entries of the unknown function $\boldsymbol{g}(\cdot)$ correspond to samples from a GP with mean $m \equiv 0$ and known, bounded, nondegenerate and continuously differentiable kernel $k(\cdot, \cdot) \leq k_{\max}$.

The squared-exponential, which is often employed in practice, is an example that satisfies the requirements of Assumption 3 (Wahba, 1990).

Using this assumption, we can compute useful bounds for the GP mean and variance as follows.

Lemma 4. Let Assumption 3 hold, and choose $k_{\max}$ accordingly. Moreover, let $\sigma_{\min} := \min_{i \in \{1, \ldots, d_x\}} \sigma_i$ be the smallest entry of the process noise covariance matrix $\boldsymbol{Q}$. Then, for any $i \in \{1, \ldots, d_x\}$ and a corresponding sample trajectory $\mathcal{T}_{i,t}^s = \{\tilde{\boldsymbol{X}}_t^s, \boldsymbol{y}_{i,t}^s\}$,

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

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

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

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

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 (9). The inequality (10) also follows straightforwardly from the symmetric positive semi-definiteness of $\boldsymbol{K}$, i.e.,

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

□

3.1 Multi-step ahead predictions

Under Assumption 3, the predicted one step dynamics are given by

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

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

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

Hence, the predicted closed-loop trajectory $\boldsymbol{X}_T^s$ is fully specified by the samples $\boldsymbol{Z}_T := (\boldsymbol{\zeta}_0,\dots,\boldsymbol{\zeta}_{T-1}) \in \mathcal{X}^T$ and $\boldsymbol{W}_T^s := (\boldsymbol{w}_0^s,\dots,\boldsymbol{w}_{T-1}^s)\in\mathcal{X}^T$. In other words, it is a function of $\boldsymbol{Z}_T$ and $\boldsymbol{W}_T^s$.

In order to derive probabilistic bounds for the closed-loop system, we need to show that the probability distribution of $\boldsymbol{X}_T^s$ is well-behaved. To this end, we require the following preliminary results.

Lemma 5. Let $\tilde{\boldsymbol{x}}_t^s$, $\boldsymbol{w}_t^s$ be fixed, and let $\boldsymbol{x}_{t+1}^s$ be given as in (6). Then $\det(\boldsymbol{\sigma}_{t+1}(\tilde{\boldsymbol{x}}_{t+1}^s))\neq 0$ holds for almost every $\boldsymbol{\zeta}_t$.

Proof. Due to Assumption 3, the kernel $k(\cdot,\cdot)$ is non-degenerate. Hence, the matrix $\boldsymbol{\sigma}_{t+1}([(\boldsymbol{x}_{t+1}^s)^{\mathrm{T}}\boldsymbol{u}_{t+1}^{\mathrm{T}}]^{\mathrm{T}})$ is invertible for almost every $\boldsymbol{x}_{t+1}^s$ (Rasmussen and Williams, 2006). This implies that the mapping from $\boldsymbol{x}_{t+1}^s$ to $\boldsymbol{\zeta}_t$ is bijective up to subsets of measure zero, and is given by

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

This implies the desired result. □

Corollary 6. Let Assumption 3 hold. For a fixed $\boldsymbol{W}_T^s$, define the function $\boldsymbol{\Phi}_{\boldsymbol{W}}:\mathcal{X}^T\to\mathcal{X}^T$, $\boldsymbol{\Phi}_{\boldsymbol{W}}(\boldsymbol{Z}_T)=\boldsymbol{X}_T^s$, where the trajectory $\boldsymbol{X}_T^s$ is computed using (6). Then $\boldsymbol{\Phi}(\cdot)$ is differentiable, and the corresponding Jacobian $J_{\boldsymbol{\Phi}}(\boldsymbol{Z}_T)$ is continuously differentiable and nonsingular for almost every $\boldsymbol{Z}_T\in\mathcal{X}^T$.

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 $\boldsymbol{x}_t^s$, $t=1,\dots,T$ with respect to the samples $\boldsymbol{\zeta}_i$, $i=0,\dots,T-1$, and the resulting matrices $d\boldsymbol{x}_t^s/d\boldsymbol{\zeta}_i$ are continuous. Hence, $\boldsymbol{\Phi}(\cdot)$ is continuously differentiable. Moreover, the components of the Jacobian $J_{\boldsymbol{\Phi}}(\cdot)$ are given by

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

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

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

which is nonzero for almost every $\boldsymbol{Z}_t$ due to Lemma 5. □

Using this result, we can show that $\boldsymbol{X}_t^s$ is within a predefined set of measure zero with probability zero.

Corollary 7. Let Assumption 3 hold, and let $\tilde{\mathcal{X}}_0^T\subset\tilde{\mathcal{X}}^T$ be a subset of measure zero. Then $\mathbb{P}(\boldsymbol{X}_T^s\in\tilde{\mathcal{X}}_0^T)=0$.

Proof. Due to Corollary 6, the change of variables formula applies (Spivak, 2018), and the probability density function of $\boldsymbol{X}_T^s$ is given by $p(\boldsymbol{X}_T^s)=p(\boldsymbol{Z}_T)|\det(J_{\boldsymbol{\Phi}^{-1}}(\boldsymbol{Z}_T))|$ up to a set of measure zero. Since $p(\boldsymbol{Z}_T)$ is continuously differentiable, the result follows. □

Furthermore, the norm of the sample vectors $\boldsymbol{y}_{i,t}^s$ is bounded, as shown in the following lemma.

Lemma 8. Let Assumption 3 hold, choose $i\in\{1,\dots,d_x\}$, and let $\boldsymbol{y}_{i,t+1}^s$ be given as in (13). Then

$$\|\boldsymbol{y}_{i,t+1}^s\|_2\leq\sum_{j=0}^{t}\left(1+\frac{k_{\max}\sqrt{t}}{\sigma_{\min}}\right)^{t-j}k_{\max}|\zeta_{i,j}| \tag{17}$$

holds.

Proof. Due to (13), Lemma 4, and the Cauchy-Schwarz inequality,

$$\begin{aligned}\|\boldsymbol{y}_{i,t+1}^s\|_2 &= \sqrt{\|\boldsymbol{y}_{i,t}^s\|_2^2+(\mu(\tilde{\boldsymbol{x}}_t^s|\mathcal{T}_{i,t}^s)+\sigma^2(\tilde{\boldsymbol{x}}_t^s|\mathcal{T}_{i,t}^s)\zeta_{i,t})^2} \\ &\leq \|\boldsymbol{y}_{i,t}^s\|_2+|\mu(\tilde{\boldsymbol{x}}_t^s|\mathcal{T}_{i,t}^s)|+|\sigma^2(\tilde{\boldsymbol{x}}_t^s|\mathcal{T}_{i,t}^s)\zeta_{i,t}| \\ &\leq \left(1+\frac{k_{\max}\sqrt{t}}{\sigma_{\min}}\right)\|\boldsymbol{y}_{i,t}^s\|_2+k_{\max}|\zeta_{i,t}|.\end{aligned} \tag{18}$$

Applying (18) t times yields

$$\begin{aligned}\|\boldsymbol{y}_{i,t+1}^s\|_2 &\leq \left(1+\frac{k_{\max}\sqrt{t}}{\sigma_{\min}}\right)^{(t+1)}\|\boldsymbol{y}_{i,0}^s\|_2 \\ &\quad+\sum_{j=0}^{t}\left(1+\frac{k_{\max}\sqrt{t}}{\sigma_{\min}}\right)^{t-j}k_{\max}|\zeta_{i,j}| \\ &=\sum_{j=0}^{t}\left(1+\frac{k_{\max}\sqrt{t}}{\sigma_{\min}}\right)^{t-j}k_{\max}|\zeta_{i,j}|.\end{aligned} \tag{19}$$

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

Hence, the growth of the sample vector depends only on the random samples $\boldsymbol{\zeta}_t$, and not on the trajectory itself. Note that, due to Proposition 4, this directly implies

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

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

Lemma 9. Let $x^s_{t+1,i}$ denote the i-th entry of (12), and let Assumptions 1-3 hold. Then, for every $t \in \{1,\dots,T\}$, there exists a polynomial function $\pi_t : \mathbb{R}^{d_x \times t} \to \mathbb{R}$, such that

$$|x^s_{t+1,i}| \leq \pi_t\left(\boldsymbol{\zeta}_0,\dots,\boldsymbol{\zeta}_t\right). \tag{20}$$

Proof. Choose the polynomial function $\pi(\cdot)$ as in Assumption 1. Then, due to Lemma 8, there exist positive constants $\tilde{a}_0,\dots,\tilde{a}_T$, such that for all $t \in \{1,\dots,T\}$ and all $i \in \{1,\dots,d_x\}$,

$$\begin{aligned}
|x^s_{t+1,i}| &= \left|f_i(\tilde{\boldsymbol{x}}^s_t) + \mu(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t}) + \sigma(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t})\zeta_{i,t}\right| \\
&\leq \left|\pi(\|\tilde{\boldsymbol{x}}^s_t\|_2) + \mu(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t}) + \sigma(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t})\zeta_{i,t}\right| \\
&\leq \pi(\|\tilde{\boldsymbol{x}}^s_t\|_2) + \sum_{j=0}^{t}\tilde{a}_j|\zeta_{i,j}| \\
&\leq \pi\left(\sum_{j=1}^{d_x}|x^s_{t,j}| + d_u u_{\max}\right) + \sum_{j=0}^{t}\tilde{a}_j|\zeta_{i,j}|
\end{aligned} \tag{21}$$

holds, where $u_{\max}$ is chosen as in Assumption 2. Applying (21) recursively yields the desired result. □

4. PREDICTING PERFORMANCE OF ONLINE LEARNING-BASED CONTROL

Ideally, we would like to use (6)-(8) to compute the expected trajectory and its variance. These are given by

$$\mathbb{E}[\boldsymbol{X}_T] = \int_{\tilde{\mathcal{X}}^t} \boldsymbol{X}^s_T \prod_{t=0}^{T-1} p(\boldsymbol{\zeta}_t) d\boldsymbol{\zeta}_t. \tag{22}$$

$$\mathbb{V}[\boldsymbol{X}_T] = \int_{\tilde{\mathcal{X}}^T} (\boldsymbol{X}^s_T - \mathbb{E}[\boldsymbol{X}^s_T])(\boldsymbol{X}^s_T - \mathbb{E}[\boldsymbol{X}^s_T])^{\mathrm{T}} \prod_{t=0}^{T-1} p(\boldsymbol{\zeta}_t) d\boldsymbol{\zeta}_t. \tag{23}$$

respectively. Using (22) and (23), we would be able to directly obtain confidence regions using Chebyshev-type inequalities. However, computing (22) and (23) directly is generally intractable. Hence, we employ a finite number of GP samples to approximate (22) and (23). To this end, we define a sample trajectory as

$$\boldsymbol{X}^{(m)}_T := \left(\left(\boldsymbol{x}^{(m)}_1\right)^{\mathrm{T}},\dots,\left(\boldsymbol{x}^{(m)}_T\right)^{\mathrm{T}}\right)^{\mathrm{T}}, \tag{24}$$

with

$$\boldsymbol{x}^{(m)}_{t+1} = \boldsymbol{f}(\tilde{\boldsymbol{x}}^{(m)}_t) + \boldsymbol{\mu}^{(m)}_t(\tilde{\boldsymbol{x}}^{(m)}_t) + \boldsymbol{\sigma}^{(m)}_t(\tilde{\boldsymbol{x}}^{(m)}_t)\boldsymbol{\zeta}^{(m)}_t. \tag{25}$$

Here the superscript (m) refers to the m-th sample simulation, and M is the total number of sample simulations. We then approximate (22) and (23) with the estimated mean and unbiased sample variance

$$\bar{\boldsymbol{X}}_{T,M} := \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{X}^{(i)}, \tag{26}$$

$$\bar{\boldsymbol{\Sigma}}_{T,M} := \frac{1}{M-1}\sum_{i=1}^{M}\left(\boldsymbol{X}^{(i)} - \bar{\boldsymbol{X}}_{T,M}\right)\left(\boldsymbol{X}^{(i)} - \bar{\boldsymbol{X}}_{T,M}\right)^{\mathrm{T}}. \tag{27}$$

The corresponding confidence region is then given by

$$\mathcal{S} = \left\{\boldsymbol{X} \in \mathcal{X}^T \middle| (\boldsymbol{X} - \bar{\boldsymbol{X}}_{T,M})\bar{\boldsymbol{\Sigma}}^{-1}_{T,M}(\boldsymbol{X} - \bar{\boldsymbol{X}}_{T,M}) \leq \eta^2\right\}, \tag{28}$$

where $\eta \in \mathbb{R}$ is a design parameter that scales the confidence regions' radius.

4.1 Choosing sample size

We now aim to derive guarantees for the confidence region (28) based on the sample size M. To achieve this, we require some preliminary results, described in the following.

Lemma 10. (Tsagris et al. (2014)). Let $\zeta \sim \mathcal{N}(0,1)$ be a random variable. Then

$$\mathbb{E}[|\zeta|^N] = \int_{\mathbb{R}} |\zeta|^N p(\zeta) d\zeta \tag{29}$$

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

We are now able show (22) and (23) are bounded.

Lemma 11. The expected value $\mathbb{E}[\boldsymbol{X}_T]$ and variance $\mathbb{V}[\boldsymbol{X}_T]$ of the system trajectory as given by (22) and (23) are bounded.

Proof. Due to Lemma 9, the integrands of (22) are bounded by $\pi_t\left(\boldsymbol{\zeta}_0,\dots,\boldsymbol{\zeta}_t\right)$, where $\pi_t(\cdot)$ is a polynomial function chosen as in Lemma 9. Moreover, due to Lemma 10 and the independence of $\boldsymbol{\zeta}_0,\dots,\boldsymbol{\zeta}_t$, the integral

$$\int_{\tilde{\mathcal{X}}^T} \pi_t\left(\boldsymbol{\zeta}_0,\dots,\boldsymbol{\zeta}_t\right) \prod_{t=0}^{T-1} p(\boldsymbol{\zeta}_t) d\boldsymbol{\zeta}_t \tag{30}$$

is finite-valued. Hence, $\mathbb{E}[\boldsymbol{X}_t]$ is finite-valued. Similarly, the entries of the integrand of (23) satisfy

$$\begin{aligned}
&\left[(\boldsymbol{X}^s_T - \mathbb{E}[\boldsymbol{X}_T])(\boldsymbol{X}^s_T - \mathbb{E}[\boldsymbol{X}_T])^{\mathrm{T}}\right]_{i,j} \\
&\leq \left(\pi_t\left(\boldsymbol{\zeta}_0,\dots,\boldsymbol{\zeta}_t\right) + \|\mathbb{E}[\boldsymbol{x}^s_t]\|_\infty\right)^2.
\end{aligned} \tag{31}$$

Since this corresponds to a polynomial function, Lemma 10 implies that the entries of $\mathbb{V}[\boldsymbol{X}_T]$ are finite-valued. □

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

Lemma 12. The trajectory variance $\mathbb{V}[\boldsymbol{X}_T]$ as given by (23) is symmetric positive definite.

Proof. We prove the result by contradiction. Assume the variance matrix $\mathbb{V}[\boldsymbol{X}_T]$ is not symmetric-positive definite. Due to Lemma 11, $\mathbb{V}[\boldsymbol{X}_T]$ is finite valued. It is easy to see from (23) that $\mathbb{V}[\boldsymbol{X}_T]$ must be symmetric positive-semidefinite. Hence, there exists an $\boldsymbol{\alpha} \in \mathbb{R}^{d_x T}$, $\boldsymbol{\alpha} \neq \boldsymbol{0}$, such that $\boldsymbol{\alpha}^{\mathrm{T}}\mathbb{V}[\boldsymbol{X}_T]\boldsymbol{\alpha} = 0$. Due to (23) and the continuity of $\boldsymbol{X}^s_T$ with respect to $\boldsymbol{Z}_T$, this implies

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

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

$$\sum_{t=0}^{T-1}\sum_{i=1}^{d_x}\alpha_{(td_x+i)}\left(f(\tilde{\boldsymbol{x}}^s_t) + \mu(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t}) + \sigma^2(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t})\zeta_{i,t}\right) = 0, \tag{32}$$

holds for almost every sample trajectory $\boldsymbol{X}^s_T \in \mathcal{X}^T$. Choose a time step $\nu \in \mathbb{N}$, $\nu \leq T$ and state dimension $\rho \in \mathbb{N}$, $\rho \leq d_x$, such that $\nu d_x + \rho := \max_j j$, $\alpha_j \neq 0$ is the highest index corresponding to a nonzero entry of $\boldsymbol{\alpha}$. We rewrite (32) as

$$\sum_{t=0}^{\nu}\sum_{i=1}^{d_x}\alpha_{(td_x+i)}\left(f(\tilde{\boldsymbol{x}}^s_t) + \mu(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t}) + \sigma^2(\tilde{\boldsymbol{x}}^s_t|\mathcal{T}^s_{i,t})\zeta_{i,t}\right) = 0. \tag{33}$$

Since varying the value of $\zeta_{\nu,\rho}$ does not affect the states up to time step ν, $\tilde{\boldsymbol{x}}_1^s, \ldots, \tilde{\boldsymbol{x}}_\nu^s$, (33) implies

$$\sigma^2(\tilde{\boldsymbol{x}}_t^s | \mathcal{T}_{\nu,\rho}^s)\zeta_{\nu,\rho} = 0, \tag{34}$$

for almost every $\boldsymbol{X}_T^s \in \mathcal{X}^T$. This is a contradiction due to Lemma 5. Hence, $\mathbb{V}[\boldsymbol{X}_T]$ is symmetric positive definite. □

We are now able to prove that randomly sample trajectories are linearly independent, as shown in the sequel.

Lemma 13. Let Assumptions 1-3 hold, let $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(M)}$ be M sample trajectories obtained using (12) and (24), and let $\bar{\boldsymbol{X}}_{T,M}$ be given as in (26). Furthermore, assume $M \geq Td_x$ and $Td_x > 1$. Let $\boldsymbol{X}^{(m_1)}, \ldots, \boldsymbol{X}^{(m_{d_xT})}$, where $m_1, \ldots, m_{d_xT} \in \{1, \ldots, M\}$, be d_xT arbitrary sample trajectories. Then $\boldsymbol{X}^{(m_1)}, \ldots, \boldsymbol{X}^{(m_{d_xT})}$ are linearly independent and $\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_{T,M} \neq \boldsymbol{0}$ holds for all $m \leq M$ with probability 1,.

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

$$\sum_{m=1}^{d_xT-1} \alpha_i \boldsymbol{X}^{(m)} = \alpha_{Td_x} \boldsymbol{X}^{(Td_x)}, \tag{35}$$

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

$$\mathcal{H} := \left\{ \boldsymbol{X} \in \mathcal{X}^T \;\middle|\; \boldsymbol{X} = \sum_{m=1}^{d_xT-1} \alpha_m \boldsymbol{X}^{(m)},\ \alpha_m \in \mathbb{R} \right\}. \tag{36}$$

Note that $\mathcal{H}$ is a hyperplane in $\mathbb{R}^{d_xT}$, hence has measure zero. Due to Corollary 7, $\mathbb{P}(\boldsymbol{X}^{(Td_x)} \in \mathcal{H}) = 0$ holds. As $\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_{T,M} = \boldsymbol{0}$ holds for some m, due to $M > 1$, then this implies

$$\boldsymbol{X}^{(m)} + \frac{M-1}{M^2} \sum_{\substack{j=1 \\ j \neq i}}^{M} \boldsymbol{X}^{(j)} =: \boldsymbol{X}^{(m)} + \bar{\boldsymbol{X}}_{M \setminus m} = \boldsymbol{0}. \tag{37}$$

This holds with probability

$$\mathbb{P}\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_{T,M} = \boldsymbol{0}\right) = \int_{\bar{\boldsymbol{X}}_{M \setminus m}} p(\boldsymbol{X}) d\boldsymbol{X} = 0 \tag{38}$$

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

$$\begin{aligned} &\mathbb{P}\left(\boldsymbol{X}^{(Td_x)} \in \mathcal{H} \bigcup \boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_{T,M} = \boldsymbol{0}\right) \\ &\leq \mathbb{P}\left(\boldsymbol{X}^{(Td_x)} \in \mathcal{H}\right) + \mathbb{P}\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_{T,M} = \boldsymbol{0}\right) = 0, \end{aligned} \tag{39}$$

which is a contradiction. □

We are now able to prove that the approximate variance (27) is invertible with probability 1.

Corollary 14. Let Assumptions 1-3 hold and let $\bar{\boldsymbol{\Sigma}}_{T,M}$ be given as in (27) with $M \geq d_xT$. Then $\bar{\boldsymbol{\Sigma}}_{T,M}$ is invertible with probability 1.

Proof. Let $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(M)}$ be the sample trajectories used to compute $\bar{\boldsymbol{\Sigma}}_{T,M}$ and consider the first d_xT samples $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(d_xT)}$. Note that if $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(d_xT)}$ are linearly independent and $\boldsymbol{X}^{(i)} - \bar{\boldsymbol{X}}_{T,M} \neq \boldsymbol{0}$ for all i, then the deviations $\left(\boldsymbol{X}^{(n_1)} - \bar{\boldsymbol{X}}_{T,M}\right), \ldots, \left(\boldsymbol{X}^{(n_{d_xT})} - \bar{\boldsymbol{X}}_{T,M}\right)$ are linearly independent. Due to Lemma 13 this holds with probability 1. Hence, for any $\boldsymbol{\alpha} \in \mathbb{R}^{d_xT}$,

$$\begin{aligned} \boldsymbol{\alpha} \bar{\boldsymbol{\Sigma}}_{T,M} \boldsymbol{\alpha}^{\mathrm{T}} &= \frac{1}{M-1} \sum_{m=1}^{M} \left(\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_{T,M} \right)^{\mathrm{T}} \boldsymbol{\alpha} \right)^2 \\ &\geq \frac{1}{M-1} \sum_{m=1}^{d_xT} \left(\left(\boldsymbol{X}^{(m)} - \bar{\boldsymbol{X}}_{T,M} \right)^{\mathrm{T}} \boldsymbol{\alpha} \right)^2 > 0 \end{aligned} \tag{40}$$

holds with probability 1. Hence, $\bar{\boldsymbol{\Sigma}}_{T,M}$ is symmetric positive definite, i.e., it is invertible. □

Using this result, we can obtain probabilistic guarantees for the confidence region (28). This is our main result, and is detailed in the following.

Theorem 15. Let Assumptions 1-3 hold. Choose $0 < \delta < 1$ and $\eta > \sqrt{Td_x/\delta}$. Furthermore, choose $M \in \mathbb{N}$, such that

$$\frac{Td_x(M^2 - 1 + M\eta^2)}{M^2\eta^2} \geq \delta$$

holds, and let $\bar{\boldsymbol{\Sigma}}_{T,M}, \bar{\boldsymbol{X}}_{T,M}$, and $\mathcal{S}$ be given by (26), (27), and (28), respectively. Then $\bar{\boldsymbol{\Sigma}}_{T,M}$ is invertible with probability 1 and $\mathcal{S}$ corresponds to a $1-\delta$ a confidence region for a trajectory $\boldsymbol{X}$ of the true system (1).

To prove Theorem 15, we employ the following result.

Lemma 16. (Stellato et al. (2017)). Let $M > d_x$ be a positive scalar. Given $M+1$ iid samples $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(M)}, \boldsymbol{X}$, if $\bar{\boldsymbol{\Sigma}}_{T,M}$ is nonsingular, then for all $\eta > 0$, it holds that

$$\begin{aligned} &\mathbb{P}\left(\left(\boldsymbol{X} - \bar{\boldsymbol{X}}_{T,M} \right)^{\mathrm{T}} \bar{\boldsymbol{\Sigma}}_{T,M}^{-1} \left(\boldsymbol{X} - \bar{\boldsymbol{X}}_{T,M} \right) \geq \eta^2 \right) \\ &\leq \min\left\{ 1, \frac{Td_x(M^2 - 1 + M\eta^2)}{M^2\eta^2} \right\}. \end{aligned} \tag{41}$$

Proof of Theorem 15. This follows directly from Corollary 14 and Lemma 16. □

Note that, for large M, the right-hand side of (41) is approximately equal to $Td_x\eta^{-2}$. This effectively imposes restrictions on the radius of the confidence regions associated with high confidence levels. In particular, $\eta \geq \sqrt{Td_x}$ must be chosen in order to obtain meaningful confidence regions.

5. NUMERICAL SIMULATION

We evaluate the performance of the proposed approach in a numerical simulation of a cart-pole system, given by

$$(m_c + m_p)\ddot{\xi} + m_p l \ddot{\theta} \cos(\theta) - m_p l \dot{\theta}^2 \sin(\theta) = u \tag{42}$$

$$m_p l^2 \ddot{\theta} + m_p g l \sin(\theta) = -m_p l \ddot{\xi} \cos(\theta). \tag{43}$$

Here ξ is the cart's position, θ is the pole's angle, and u is the horizontal force applied to the cart. The cart and pole masses are given by $m_c = 0.5\,\mathrm{kg}$ and $m_p = 0.5\,\mathrm{kg}$, respectively The parameter l denotes the pole's length. The discrete-time dynamical system form (1) is obtained by sampling the continuous system (42) every 0.05 seconds. The prior model $\boldsymbol{f}(\cdot)$ corresponds to the linearized dynamics around the origin, with incorrectly assumed masses $m_c = m_p = 0.4\,\mathrm{kg}$. Additionally, we consider discrete process noise $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$, with $\boldsymbol{Q} = 0.01\boldsymbol{I}$. We employ the safe learning-based control law from Koller et al.

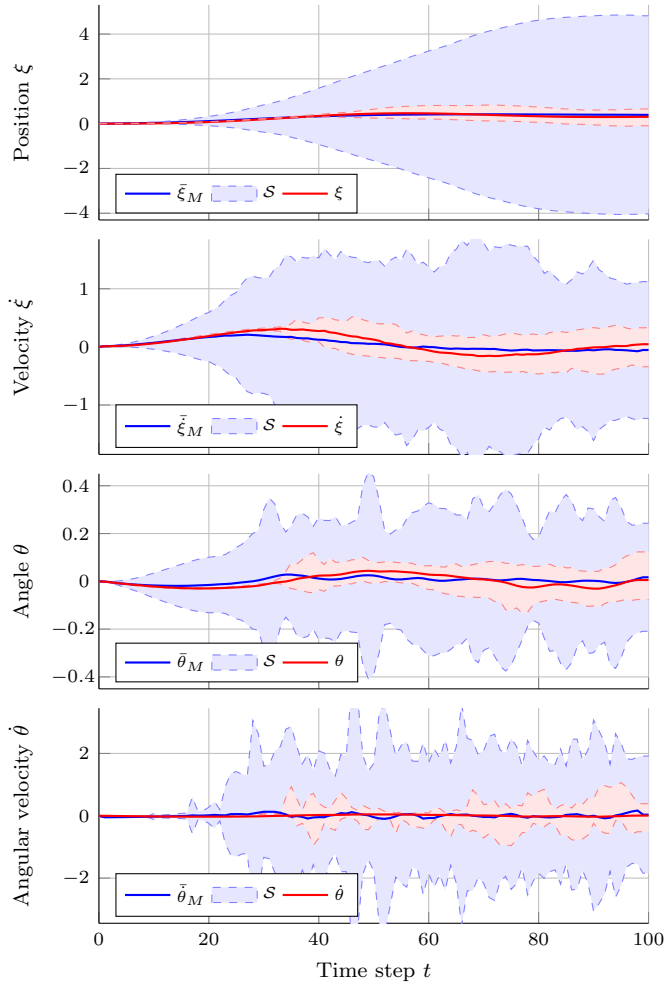

Fig. 1. Mean and confidence region of predicted cart pole trajectory (blue and light blue). Median, lower and upper deciles of Monte Carlo simulation of true system (red and light red).

(2018) to move the cart position from $x = 0$ to $x = 0.3$ while simultaneously learning the system dynamics.

We predict the performance of the closed-loop system over $T = 100$ steps, which corresponds to 5 seconds. We choose $\eta = 10$ and sample $M = 100$ trajectories to compute the confidence region $\mathcal{S}$. This yields a confidence level of $1-\delta \approx 0.955$. Moreover, we perform a Monte Carlo simulation of the true system, consisting of 100 runs. The results are shown in Figure 1. As can be seen, the median, lower and upper quartiles corresponding to the true system lie entirely within the confidence region. Towards the end of the simulation, the confidence region is larger than at the beginning. This is due to the propagation of model uncertainty. However, the confidence region shrinks slightly at $t = 80$ time steps. This is because the system dynamics are learned during the simulation, and the posterior variance of the Gaussian process decreases in the proximity of the reference.

6. CONCLUSION

We have presented a technique for computing confidence regions of online learning-based control systems. The proposed method yields a correct confidence region 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.

In future work, we aim to apply the presented approach to support the design of online learning-based control laws.

REFERENCES

Bakker, B., Zhumatiy, V., Gruener, G., and Schmidhuber, J. (2006). Quasi-online reinforcement learning for robots. In *2006 IEEE International Conference on Robotics and Automation*, 2997–3002.

Bar-Shalom, Y. and Tse, E. (1974). Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, 19(5), 494–500.

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., 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.

Mesbah, A. (2018). Stochastic model predictive control with active uncertainty learning: a survey on dual control. *Annual Reviews in Control*, 45, 107–117.

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*.

Spivak, M. (2018). *Calculus on manifolds: a modern approach to classical theorems of advanced calculus*. CRC press.

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. and Hirche, S. (2019). Feedback linearization based on Gaussian processes with event-triggered online learning. *IEEE Transactions on Automatic Control*.

Wahba, G. (1990). *Spline models for observational data*, volume 59. Siam.