

# Gaussian-Process-Based Control of Underactuated Balance Robots With Guaranteed Performance

Kuo Chen, Jingang Yi , *Senior Member, IEEE*, and Dezhen Song , *Senior Member, IEEE*

**Abstract**—The control of underactuated balance robots is aimed at performing both the external (actuated) subsystem trajectory tracking and internal (unactuated) subsystem balancing tasks. In this article, we propose a learning-based control design for underactuated balance robots. The key idea integrates a model predictive control method to design the desired internal subsystem trajectory and perform the external subsystem tracking task, while an inverse dynamics controller is used to stabilize the internal subsystem to its desired trajectory. The control design is based on Gaussian process (GP) regression models that are learned from experiments without requiring *a priori* knowledge about the robot dynamics or the demonstration of successful stabilization. GP regression models also provide estimates of modeling uncertainties of the robotic systems, and these estimations are used to enhance control robustness to modeling errors. The learning-based control design is analyzed with guaranteed stability and performance. The proposed design is demonstrated by experiments on a Furuta pendulum and an autonomous bikebot.

**Index Terms**—Balance control, Gaussian processes (GPs), model predictive control, nonminimum phase systems, underactuated robots.

## I. INTRODUCTION

UNDERACTUATED robots have fewer control inputs than the degrees of freedom (DOFs) [1]. The control of underactuated balance robots requires trajectory tracking for an actuated subsystem and balance task around unstable equilibria for an unactuated subsystem [2]. Commonly studied underactuated balance robots include the cart–pole system [3], Furuta pendulum [4], [5], [6], autonomous bicycles [7], [8], and bipedal walkers [9], [10], to name a few. The control of these robots is

challenging due to the lack of an analytical causal compensator for the nonminimum phase nature of their dynamics [11]. A well-known approach is to formulate the part of the control task as an inversion problem, that is, to estimate the future state trajectory such that the resultant internal state can be stabilized on the desired profile [2], [3], [12]. In [2], an approximate dynamic inversion approach was proposed to construct the balance equilibrium manifold (BEM) to capture the desired trajectory for the unstable subsystem state. An external/internal convertible (EIC) form of robot dynamics was used to design the controller to perform both the tracking and balancing tasks. Despite the guaranteed stability property, all the above-mentioned nonlinear control designs require accurate dynamics models.

Using machine learning techniques, data-driven model-based controllers have showed promising potential to capture complex systems dynamics and achieve satisfactory control performance [13]. The Gaussian process (GP) is used as a nonparametric learning model and has been widely applied to robot modeling and control [14]. The work in [15] proposed a GP-based inverse dynamics control law, and the feedback gain was adapted to the variance of the predictive distribution. The approaches in [15], [16], and [17] gave theoretically guaranteed stability or safety regions of GP-based inverse dynamics control. The work in [18] achieved robotic impromptu trajectory tracking for a cart–pole system by learning a stable approximate inverse of a nonminimum phase system. The algorithm, however, required a baseline controller to stabilize the system for data collection. In [19], the BEM approach was used to learn model-based control. The learned model was, however, not accurate due to the flexible structure of the GP under the dynamic inversion. The work in [20] presented a GP-based stable tracking control of underactuated balance robots, and the rotary pendulum was used as a demonstration example. However, the design required the knowledge of the physical model structure, which limits the approach to general robot dynamics.

Optimization-based controllers, such as model predictive control (MPC) and reinforcement learning, have been applied to underactuated robot systems. Many GP-based designs take advantage of the predicted Gaussian distribution to achieve robust control performance. For example, in [17], [21], [22], [23], and [24], the objective function was designed to include tracking errors over the prediction horizon with the variance of the predictive distribution. In [25], the predictive variance was used to help reduce the feasible region for the predictive trajectory mean value. Learning-based inverse dynamics control and MPC have been demonstrated in many applications [26], [27], [28],

Manuscript received 21 January 2022; revised 28 May 2022; accepted 6 August 2022. This paper was recommended for publication by Associate Editor S. Oh and Editor E. Yoshida upon evaluation of the reviewers' comments. The work of Jingang Yi was supported in part by the U.S. National Science Foundation under Award CNS-1932370. An earlier version of this paper was presented in part at the 2019 IEEE International Conference on Robotics and Automation, Montreal, Canada, May 20–24, 2019 [DOI: 10.1109/ICRA.2019.8794097]. (Corresponding author: Jingang Yi.)

Kuo Chen was with the Department of Mechanical and Aerospace Engineering, Rutgers University, Piscataway, NJ 08854 USA. He is now with Motional Inc., Boston, MA 02210 USA (e-mail: kc625@rutgers.edu).

Jingang Yi is with the Department of Mechanical and Aerospace Engineering, Rutgers University, Piscataway, NJ 08854 USA (e-mail: jgyi@rutgers.edu).

Dezhen Song is with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: dz-song@cse.tamu.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TRO.2022.3203625>.

Digital Object Identifier 10.1109/TRO.2022.3203625

[29]. In [30], [31], and [32], researchers adopted an inverse dynamics controller with the global and local GP regression models. These inverse dynamics controllers, however, cannot be directly applied to unstable nonminimum phase underactuated robots.

In this article, we take advantage of the EIC form property of the underactuated balance robot dynamics to design GP-based learning controllers. A GP-based inverse dynamics controller is designed and applied to the internal subsystem to ensure rapid stabilization with a specific fast convergence rate. An MPC is used to simultaneously track the output of the external subsystem and obtain the desired balance profiles for the internal subsystem. Both the external and internal subsystem dynamics are learned from experimental data with GP models, and the MPC incorporates predictive model uncertainties into the design to enhance the control robustness with proven stability. We demonstrate the proposed learning-based control design on a Furuta pendulum and an autonomous bikebot.

This article makes several contributions. First, the control design is based on the learning models without needing to obtain the physical dynamics model, and therefore, it is attractive for many complex underactuated balance robots whose dynamics models are difficult to obtain. Our approach differs from previous approaches that relied on either the prior knowledge of robot's physical dynamics model or the demonstration of stable motion to obtain the training data. For example, the controller in [18] explicitly requires a stabilized baseline system, and the learning only improves the tracking accuracy under the baseline controller. The proposed approach does not require such a stabilization design for training data collection, and in this sense, the approach treats the collected training data as open-loop dynamic responses. We demonstrate this "free of stable motion in training data" property in rotary inverted pendulum experiments. This is attractive for many robots with intrinsic unstable balance. Second, the proposed control demonstrates a novel design of explicitly incorporating predictive model uncertainty to enhance control robustness with guaranteed stability and convergence. The guaranteed stability property is attractive for learning-based robot control. The use of the GP-based MPC to design the BEM is novel and avoids the dynamic inversion of nonminimum phase robot dynamics. Compared with the previous conference presentation [33], this article includes significant extensions and new developments in control analysis and new experiments conducted on multiple robotic platforms.

The rest of this article is organized as follows. In Section II, we present the robot dynamics and basics of GP models. Section III presents the control design overview. Section IV presents the GP-based control of underactuated balance robots. We present the control performance analysis in Section V. Experimental results are included in Section VI. Finally, Section VII concludes this article.

## II. ROBOT DYNAMICS AND GPs

### A. Notations

We introduce notations that will be used in this article. Vector  $\alpha$  and matrix  $A$  are denoted with bold lowercase and uppercase variables, respectively. An  $n \times n$  identity matrix is denoted as

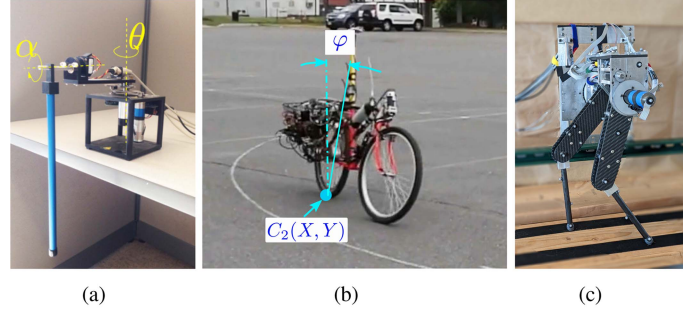


Fig. 1. Three examples of underactuated balance robots. (a) Furuta pendulum. (b) Autonomous bikebot. (c) Robotic biped.

$I_n$ . The smallest and the largest eigenvalues of matrix  $A$  are denoted by  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$ , respectively. For vector  $x \in \mathbb{R}^n$  and real matrix  $A$ ,  $\|x\|$  and  $\|A\|$  represent the (induced) 2-norm, that is,  $\|x\| = \sqrt{x^T x}$  and  $\|A\| = \sqrt{\lambda_{\max}(A^T A)}$ , and the weighted norm is  $\|x\|_P = \sqrt{x^T P x}$  with symmetric positive definite  $P \in \mathbb{R}^{n \times n}$ .  $\text{tr}(A)$  and  $\det(A)$  denote the trace and determinant of matrix  $A$ , respectively. The expression  $x \sim \mathcal{N}(\mu, \Sigma)$  represents that  $x$  is a random variable satisfying Gaussian distribution with mean vector  $\mu$  and covariance matrix  $\Sigma$ . The expectation operator is denoted as  $\mathbb{E}$ , variable  $\Pi$  denotes a probabilistic event, and its probability is written as  $\Pr\{\Pi\}$ .

### B. Robot Dynamics and Problem Statement

Fig. 1 shows three examples of underactuated balance robots. The Furuta pendulum shown in Fig. 1(a) has the actuated base joint angle  $\theta$  and the unstable unactuated pendulum link joint angle  $\alpha$ . The bikebot shown in Fig. 1(b) has three DOFs (i.e., position  $(X, Y)$  at the rear wheel contact point  $C_2$  and unstable platform roll angle  $\varphi$ ) with two inputs (i.e., velocity and steering actuation) [7], [8]. The five-link robotic biped shown in Fig. 1(c) has four actuated joints, and the unstable torso joint is unactuated [9], [34]. We present general dynamics models and property for these underactuated balance robots.

The underactuated robot dynamics model is described as

$$D(q)\ddot{q} + H(q, \dot{q}) = B(q)u \quad (1)$$

where  $q \in \mathcal{Q}$  and  $\dot{q} \in T\mathcal{Q}$  are the generalized coordinate and velocity of the system, respectively,  $\mathcal{Q}, T\mathcal{Q} \subset \mathbb{R}^{m+n}$  are compact sets,  $u \in \mathcal{U}$  is the control input, admissible control set  $\mathcal{U} \subset \mathbb{R}^m$  is compact,  $D(q)$  is the inertia matrix,  $H(q, \dot{q})$  contains the centripetal, Coriolis, and gravitational terms, and  $B(q)$  is the input mapping matrix [35]. The coordinate  $q = [\theta_1^T \alpha_1^T]^T$  is partitioned into  $\theta_1 \in \mathbb{R}^m$  of the actuated subsystem and  $\alpha_1 \in \mathbb{R}^n$  of the unactuated subsystem. Throughout this article, we focus on the subset of the underactuated systems with  $m \geq n$ , that is, the actuated DOFs are not less than the unactuated DOFs. Defining  $\theta_2 = \dot{\theta}_1$  and  $\alpha_2 = \dot{\alpha}_1$ , (1) becomes

$$\begin{bmatrix} \dot{\theta}_2 \\ \dot{\alpha}_2 \end{bmatrix} = D^{-1}(q) \left( \begin{bmatrix} B_1(q)u \\ 0 \end{bmatrix} - H(q, \dot{q}) \right) \quad (2)$$

where  $B_1(\mathbf{q}) \in \mathbb{R}^{m \times m}$  is full rank. A general state-space representation of (2) is formulated as

$$\begin{cases} \Sigma_e : \dot{\theta}_1 = \theta_2, \dot{\theta}_2 = \mathbf{f}_\theta(\theta, \alpha, \mathbf{u}) \\ \Sigma_i : \dot{\alpha}_1 = \alpha_2, \dot{\alpha}_2 = \mathbf{f}_\alpha(\theta, \alpha, \mathbf{u}) \end{cases} \quad (3)$$

where  $\theta = [\theta_1^\top \theta_2^\top]^\top \in \mathbb{R}^{2m}$ ,  $\alpha = [\alpha_1^\top \alpha_2^\top]^\top \in \mathbb{R}^{2n}$ , and  $\mathbf{f}_\theta$  and  $\mathbf{f}_\alpha$  are smooth nonlinear functions for the subsystems  $\Sigma_e$  and  $\Sigma_i$ , respectively. In general, nonlinear dynamics of  $\Sigma_i$  in (3) are unstable due to the balance task of the robotic system.

The key idea of EIC form is that the external and internal subsystems are interchangeable by selecting a different output subspace ( $\theta$  or  $\alpha$ ) [2]. The external subsystem is the subsystem with the chosen system output. For illustration purposes, let us consider a special case of  $m = n$ , and both  $\mathbf{f}_\theta$  and  $\mathbf{f}_\alpha$  are invertible with respect to  $\mathbf{u}$ . Considering that the desired trajectory is specified for  $\theta$  subspace,  $\theta$  is the output, and therefore,  $\theta$  dynamics are considered to be the external subsystem.  $\Sigma_e$  of (3) can be feedback linearized by introducing a virtual input  $\mathbf{z} = \mathbf{f}_\theta(\theta, \alpha, \mathbf{u})$  to obtain the linearized form of (3) as

$$\begin{cases} \dot{\theta}_1 = \theta_2, \dot{\theta}_2 = \mathbf{z} \\ \dot{\alpha}_1 = \alpha_2, \dot{\alpha}_2 = \mathbf{f}_\alpha(\theta, \alpha, \mathbf{f}_\theta^{-1}(\theta, \alpha, \mathbf{z})) \end{cases} \quad (4)$$

Symmetrically, by introducing a virtual input  $\mathbf{v} = \mathbf{f}_\alpha(\theta, \alpha, \mathbf{u})$ , the internal subsystem of (3) is feedback linearized, and we obtain

$$\begin{cases} \dot{\alpha}_1 = \alpha_2, \dot{\alpha}_2 = \mathbf{v} \\ \dot{\theta}_1 = \theta_2, \dot{\theta}_2 = \mathbf{f}_\theta(\theta, \alpha, \mathbf{f}_\alpha^{-1}(\theta, \alpha, \mathbf{v})) \end{cases} \quad (5)$$

Equations (4) and (5) are called EIC dual forms for (3) from different perspectives of output linearization [2]. External/internal subsystem interchangeability is only used to introduce the EIC form as background knowledge to facilitate understanding of the technical development. In the rest of this article, the external subsystem  $\Sigma_e$  refers to the  $\theta$  dynamics and the internal subsystem  $\Sigma_i$  refers to the  $\alpha$  dynamics given in (3).

Given a desired trajectory  $\theta_{1d}$ , the controller aims to perform an asymptotic tracking task of  $\theta_1$  to  $\theta_{1d}$ ; meanwhile, the unstable internal subsystem is stabilized. Defining  $\theta_d = [\theta_{1d}^\top \theta_{2d}^\top]^\top$ ,  $\theta_{2d} = \dot{\theta}_{1d}$ , as the EIC control design [2], a state feedback control, denoted as  $\mathbf{z}_f$ , is first chosen for input  $\mathbf{z}$  in (4) to achieve asymptotic tracking of  $\theta_1$  to  $\theta_{1d}$ . Under control  $\mathbf{z}_f$ , the BEM is defined as element  $\alpha_1$  of the equilibrium in (4), that is,

$$\mathcal{E} = \{\alpha_1^e : \mathbf{f}_\alpha(\theta, \alpha^e, \mathbf{f}_\theta^{-1}(\theta, \alpha^e, \mathbf{z}_f)) = \mathbf{0}, \alpha_2^e = \mathbf{0}\}, \quad (6)$$

where  $\alpha^e = [(\alpha_1^e)^\top (\alpha_2^e)^\top]^\top$ . The BEM specifies the desired trajectory  $\alpha_1$  of the internal subsystem when the external subsystem follows  $\theta_d$ . For example, a bikebot following a straight-line trajectory would result in a BEM of a zero roll angle (i.e., upright position), whereas when following a circular trajectory, the BEM is a constant roll angle whose value depends on the trajectory curvature and the bikebot's velocity.

The BEM is considered the desired trajectory  $\alpha_{1d}$  for  $\alpha_1$ , namely,  $\alpha_{1d} \in \mathcal{E}$ . To obtain  $\alpha_{2d} = \dot{\alpha}_{1d}$  as the desired trajectory

for  $\alpha_2$ , we take the directional derivative of  $\alpha_{1d}$  along the  $\theta$ -dynamics vector field under control  $\mathbf{z}_f$ , i.e.,

$$\alpha_{2d} = \frac{\partial \alpha_{1d}(\theta, \mathbf{z}_f)}{\partial \theta} \dot{\theta} + \frac{\partial \alpha_{1d}(\theta, \mathbf{z}_f)}{\partial t}, \quad \alpha_{1d} \in \mathcal{E}.$$

$\mathbf{z}_f$  is a function of  $\theta$  and  $\theta_d$ , and therefore, the directional derivative does not contain terms with  $\dot{\mathbf{z}}_f$ . With the above construction,  $\alpha_d = [\alpha_{1d}^\top \alpha_{2d}^\top]^\top$ ,  $\alpha_{1d} \in \mathcal{E}$ , is the desired output trajectory for  $\alpha$  under the tracking control  $\mathbf{z}_f$ . The EIC control then updates the above  $\mathbf{z}_f$  design to achieve an exponential convergence of  $\alpha$  and  $\theta$  to the respective neighborhoods of  $\alpha_d$  and  $\theta_d$  simultaneously [2].

One challenge of the above EIC control design lies in an estimation of  $\alpha_d$  that implicitly depends on tracking performance under  $\mathbf{z}_f$ . In [2], a dynamic inversion method was proposed to obtain  $\mathcal{E}$  in (6), and it was involved to invert a nonminimum phase nonlinear dynamics. From (6), it is clear that accurate models (i.e.,  $\mathbf{f}_\alpha$  and  $\mathbf{f}_\theta$ ) are required to guarantee the control performance. With the above background description, we consider the following problem.

**Problem statement:** The control goal for robot dynamics (3) is to track a given desired trajectory  $\theta_d$  of the external subsystem  $\Sigma_e$  and simultaneously obtain and follow the desired trajectory  $\alpha_d$  of the unstable internal subsystem  $\Sigma_i$  by learning robot dynamics models (i.e.,  $\mathbf{f}_\alpha$  and  $\mathbf{f}_\theta$ ) in (3) with guaranteed stability and tracking performance.

### C. Gaussian Processes

A GP is a collection of random variables, any finite number of which has a joint Gaussian distribution. A real-valued random function  $f(\mathbf{x})$  of variable  $\mathbf{x} \in \mathcal{X}$ , where  $\mathcal{X} \subset \mathbb{R}^n$  is compact, is determined by its mean value function  $\mu(\mathbf{x})$  and covariance function  $k(\mathbf{x}, \mathbf{x}')$  as  $\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$  and  $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))]$  for  $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ . The GP with a zero prior mean function is assumed sufficient for the applications in this article and is, therefore, adopted throughout this article. Suppose that the training dataset contains  $N$  data pairs  $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^N$ , where observation  $y_i = f(\mathbf{x}_i) + \varepsilon$  for  $\mathbf{x}_i, \varepsilon \sim \mathcal{N}(0, \sigma^2)$  is a zero-mean Gaussian noise. The observation is denoted as  $\mathbf{y} = [y_1 \cdots y_N]^\top$ , and the input is denoted as  $\mathbf{X} = [\mathbf{x}_1^\top \cdots \mathbf{x}_N^\top]^\top$ . Under the zero prior mean GP assumption, for a testing input  $\mathbf{x}^* \in \mathbb{R}^n$ , the joint distribution of  $f(\mathbf{x}^*)$  and  $\mathbf{y}$  is a zero mean Gaussian distribution, namely

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{x}^*) \end{bmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} \mathbf{K}_X + \sigma^2 \mathbf{I}_N & \mathbf{k}(\mathbf{X}, \mathbf{x}^*) \\ \mathbf{k}(\mathbf{X}, \mathbf{x}^*)^\top & k(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix} \right)$$

where the  $(i, j)$ th element of the Gram matrix  $\mathbf{K}_X$  is  $k(\mathbf{x}_i, \mathbf{x}_j)$  and the  $i$ th element of  $\mathbf{k}(\mathbf{X}, \mathbf{x}^*) \in \mathbb{R}^N$  is  $k(\mathbf{x}_i, \mathbf{x}^*)$ ,  $i, j = 1, \dots, N$ . Throughout this article, the covariance (kernel) function adopts the squared exponential (SE) form. For calculating the Gram matrix  $\mathbf{K}_X$ ,  $k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp[-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{S}(\mathbf{x}_i - \mathbf{x}_j)]$ , where positive-definite matrix  $\mathbf{S}$  and  $\sigma_f > 0$  are hyperparameters determined by the learning process. For calculating vector  $\mathbf{k}(\mathbf{X}, \mathbf{x}^*)$ ,  $k(\mathbf{x}_i, \mathbf{x}^*) = \sigma_f^2 \exp[-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}^*)^\top \mathbf{S}(\mathbf{x}_i - \mathbf{x}^*)]$ .



The prediction of  $f(\mathbf{x}^*)$  is given by the conditional distribution  $f(\mathbf{x}^*)|\mathcal{D} \sim \mathcal{N}(\mu(\mathbf{x}^*|\mathcal{D}), \Sigma(\mathbf{x}^*|\mathcal{D}))$  with mean and covariance functions, respectively, as

$$\mu(\mathbf{x}^*|\mathcal{D}) = \mathbf{k}(\mathbf{X}, \mathbf{x}^*)^\top (\mathbf{K}_X + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{y} \quad (7a)$$

$$\Sigma(\mathbf{x}^*|\mathcal{D}) = \mathbf{k}(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{k}(\mathbf{X}, \mathbf{x}^*)^\top (\mathbf{K}_X + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{k}(\mathbf{X}, \mathbf{x}^*). \quad (7b)$$

For the  $m$ -dimensional function  $\mathbf{f}(\mathbf{x})$ , GPs are adopted to learn each element  $f_i(\mathbf{x})$ ,  $i = 1, \dots, m$ , as  $f_i(\mathbf{x}^*)|\mathcal{D} \sim \mathcal{N}(\mu_i(\mathbf{x}^*|\mathcal{D}), \Sigma_i(\mathbf{x}^*|\mathcal{D}))$  independently. The predictive distribution is written as  $\mathbf{f}(\mathbf{x}^*)|\mathcal{D} \sim \mathcal{N}(\mu(\mathbf{x}^*|\mathcal{D}), \Sigma(\mathbf{x}^*|\mathcal{D}))$ , where  $\mu(\mathbf{x}^*|\mathcal{D})$  with element  $\mu_i(\mathbf{x}^*|\mathcal{D})$  and  $\Sigma(\mathbf{x}^*|\mathcal{D})$  is a diagonal matrix with element  $\Sigma_i(\mathbf{x}^*|\mathcal{D})$ .

The hyperparameters are obtained by the training process, such as maximum *a posteriori*. Once hyperparameters are determined, the values of the above SE covariance function only depend on the distance between two points. If a testing point  $\mathbf{x}^*$  is far away from all training data, i.e.,  $\|\mathbf{x}^* - \mathbf{x}_i\| \rightarrow \infty$  for any  $\mathbf{x}_i \in \mathcal{D}$ , the posterior mean  $\mu(\mathbf{x}^*|\mathcal{D}) \rightarrow 0$  due to  $\mathbf{k}(\mathbf{x}^*, \mathbf{x}_i) \rightarrow 0$ . The posterior covariance is bounded by  $\Sigma(\mathbf{x}^*|\mathcal{D}) \leq \sigma_f^2 + \sigma^2$ . For the  $m$ -dimensional function  $\mathbf{f}$ ,  $\|\Sigma(\mathbf{x}^*|\mathcal{D})\| \leq \max_{1 \leq i \leq m} (\sigma_{f_i}^2 + \sigma_i^2)$ , where  $\sigma_{f_i}$  and  $\sigma_i$  are the hyperparameters for corresponding  $f_i$ . The rest of this article mainly focuses on the posterior GP given training data  $\mathcal{D}$ , and for notation simplicity,  $\mathcal{D}$  will be omitted in  $\mu(\mathbf{x})$  and  $\Sigma(\mathbf{x})$ . The following lemma gives the estimation error bound.

*Lemma 1 (see [36, Th. 6]):* For a function  $f(\mathbf{x})$  with bounded reproducing kernel Hilbert space norm  $\|f(\mathbf{x})\|_k$ , and  $\mathbf{x} \in \mathcal{X}$ ,  $\mathcal{X} \subset \mathbb{R}^n$  is compact, let  $\delta \in (0, 1)$ ; then

$$\Pr\{|\mu(\mathbf{x}) - f(\mathbf{x})| \leq \beta \Sigma^{\frac{1}{2}}(\mathbf{x})\} \geq 1 - \delta$$

where  $\beta = \sqrt{2\|f\|_k^2 + 300\gamma \ln^3(\frac{N+1}{\delta})}$  and  $\gamma = \max_{\mathbf{x}} \frac{1}{2} \log \det(\mathbf{I}_N + \sigma^{-2} \mathbf{K}_X)$  is the maximum information gain for GP.

For the  $m$ -dimensional function  $\mathbf{f}(\mathbf{x})$ , if every dimension is independent, the results in Lemma 1 are extended to  $\Pr\{\|\mu(\mathbf{x}) - \mathbf{f}(\mathbf{x})\| \leq \|\beta^\top \Sigma^{\frac{1}{2}}(\mathbf{x})\|\} \geq (1 - \delta)^m$ , where  $\mu(\cdot)$  and  $\Sigma(\cdot)$  are defined in (7),  $\beta \in \mathbb{R}^m$  with element  $\beta_i = \sqrt{2\|f_i\|_k^2 + 300\gamma_i \ln^3(\frac{N+1}{\delta})}$ , and  $\gamma_i$  is the maximum information gain for  $f_i$ .

### III. ROBOT CONTROL DESIGN OVERVIEW

In this section, we first illustrate the proposed control design by using the physical dynamics model and then give an overview of the GP-based control design.

#### A. Internal Subsystem Control

Instead of using (4) to design tracking control of  $\theta$  as in the EIC-based approach, we first consider driving  $\alpha$  to follow  $\alpha_d$  by (5) assuming that the desired trajectory  $\alpha_d$  is known. Only a subset of  $\mathbf{u}$  is obtained by inverting  $\mathbf{f}_\alpha$  due to  $m \geq n$ , and therefore, letting  $\mathbf{u} = [\mathbf{u}_d^\top \mathbf{u}_f^\top]^\top$ ,  $\mathbf{u}_d \in \mathbb{R}^n$ , and  $\mathbf{u}_f \in \mathbb{R}^{m-n}$ ,  $\mathbf{u}_d$

is obtained by inverting  $\mathbf{f}_\alpha$ , i.e.,

$$\mathbf{u}_d = \mathbf{f}_\alpha^{-1}(\theta, \alpha, \mathbf{v}, \mathbf{u}_f). \quad (8)$$

Robot dynamics (5) become

$$\Sigma_e : \dot{\theta}_1 = \theta_2, \dot{\theta}_2 = \mathbf{f}_\theta(\theta, \alpha, \mathbf{v}, \mathbf{u}_f) \quad (9a)$$

$$\Sigma_i : \dot{\alpha}_1 = \alpha_2, \dot{\alpha}_2 = \mathbf{v}. \quad (9b)$$

In (9a), we slightly abuse notation for function  $\mathbf{f}_\theta$  by directly replacing  $\mathbf{u}$  with  $(\mathbf{v}, \mathbf{u}_f)$  in argument without causing any confusion.

A high-gain feedback control is designed for input  $\mathbf{v}$  in (9b) for rapidly tracking of  $\alpha$  to  $\alpha_d$ , and one commonly used high-gain controller is given as

$$\mathbf{v} = \dot{\alpha}_{2d} - \frac{k_d}{\epsilon} \mathbf{e}_{\alpha 2} - \frac{k_p}{\epsilon^2} \mathbf{e}_{\alpha 1} = \mathbf{w} - \frac{k_d}{\epsilon} \mathbf{e}_{\alpha 2} - \frac{k_p}{\epsilon^2} \mathbf{e}_{\alpha 1} \quad (10)$$

where  $\mathbf{w} = \dot{\alpha}_{2d}$  is the designed input, errors  $\mathbf{e}_{\alpha 1} = \alpha_1 - \alpha_{1d}$ ,  $\mathbf{e}_{\alpha 2} = \alpha_2 - \alpha_{2d}$ ,  $\mathbf{e}_\alpha = \alpha - \alpha_d = [\mathbf{e}_{\alpha 1}^\top \mathbf{e}_{\alpha 2}^\top]^\top$ ,  $\epsilon > 0$  is small, and  $k_p, k_d > 0$  are constant gains. Plugging (10) into (9b), the closed-loop error  $\mathbf{e}_\alpha$  converges to zero exponentially with a rapid rate of  $-1/\epsilon$ . Inspired by singular perturbation theory [37] with detailed analysis provided in later sections, by replacing  $\alpha$  with  $\alpha_d$  under control  $\mathbf{v}$  in (10), an accurate approximation of (9a) is obtained as

$$\dot{\hat{\theta}}_1 = \hat{\theta}_2, \dot{\hat{\theta}}_2 = \mathbf{f}_\theta(\hat{\theta}, \alpha_d, \mathbf{w}, \mathbf{u}_f) \quad (11)$$

where  $\hat{\theta} = [\hat{\theta}_1^\top \hat{\theta}_2^\top]^\top$  is an estimate of  $\theta$ , and it can be shown that  $\|\theta - \hat{\theta}\| = O(\epsilon)$  [37]. In (11),  $\alpha_d$  and  $\mathbf{w}$  are designed by

$$\dot{\alpha}_{1d} = \alpha_{2d}, \dot{\alpha}_{2d} = \mathbf{w}, \alpha_d(t_0) = \alpha_{d0} \quad (12)$$

where input  $\mathbf{w}$  and initial value  $\alpha_{d0}$  determine the trajectory of  $\alpha_d(t)$  at any time  $t \geq t_0$  with initial time  $t_0$ . Function arguments of  $\mathbf{f}_\theta$  in (11) do not contain  $\mathbf{v}$  and  $\alpha$  terms since, under (10),  $\alpha$  converges to  $\alpha_d$  rapidly and  $\mathbf{v}$  is a function of  $\mathbf{w}$  and  $\alpha_d$ .

The rationale of introducing the approximation in (11) is that  $\hat{\theta}$  dynamics are decoupled from closed-loop  $\alpha$  dynamics and the trajectory  $\alpha_d$  is considered as an input to drive  $\hat{\theta}$  close to  $\theta$ . The desired trajectory  $\alpha_d$  exhibits a duplicated dynamics similar to that of  $\alpha$  with decision variables  $\mathbf{w}$  and  $\alpha_{d0}$ , which are designed by an optimization approach. The future value of  $\alpha_d$  can be chosen by design, while  $\alpha$  is the actual state whose value is only available up to the current time. By doing so, we avoid designing a noncausal controller, which is one of the main challenges for the control of nonminimum phase nonlinear dynamic systems.

#### B. External Subsystem Control

We still need to design the tracking control of  $\hat{\theta}$  to follow  $\theta_d$  and discuss how to obtain the desired trajectory  $\alpha_d$ . We formulate an MPC to achieve these two tasks. Since  $\hat{\theta}$  is an accurate approximation of  $\theta$ , one element of the MPC objective function is to minimize the magnitude of error  $\mathbf{e}_{\hat{\theta}} = \hat{\theta} - \theta_d$ , from current time  $t$  to  $t_h = t + t_H$  with a time horizon  $t_H > 0$ . Considering  $\hat{\theta}$ - and  $\alpha_d$ -dynamics in (11) and (12), the MPC decision variables are denoted as  $\mathbf{W}(t) = \{\alpha_d(t) = \alpha_{dt}, \mathbf{w}(\tau), \mathbf{u}_f(\tau), t \leq$

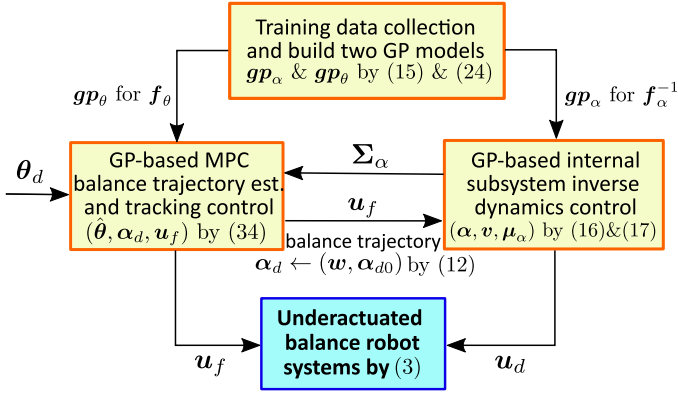


Fig. 2. Overall schematic of the GP-based control algorithms. Two GP models  $gp_\alpha$  and  $gp_\theta$  are trained by offline experiments and then used to estimate  $f_\alpha^{-1}$  and  $f_\theta$ , respectively. A GP-based inverse dynamics control  $u_d$  (through  $v$ ) and an MPC design  $u_f$  are then used for real-time tracking control of the internal and external subsystems, respectively.

$\tau \leq t_h\}$ , and the MPC is given as

$$\min_{\mathbf{W}(\tau)} \|\alpha_{dt}\|_{Q_2}^2 + \int_t^{\tau} (\|e_\theta\|_{Q_1}^2 + \|\mathbf{w}\|_R^2 + \|\mathbf{u}_f\|_R^2) d\tau \quad (13a)$$

$$\text{subject to (11) and (12) with } t \leq \tau \leq t_h \quad (13b)$$

where matrices  $Q_1$ ,  $Q_2$ , and  $R$  are positive definite.

Note that penalizing  $\|\alpha_{dt}\|_{Q_2}^2$  as an initial cost together with the  $\|\mathbf{w}\|_R^2$  stage cost is equivalent to penalizing the weighted norm of  $\alpha_d(\tau)$  over the time horizon  $\tau \in [t, t_h]$  in (13). Besides (11) and (12), no other constraints on system states and inputs are considered in this article. By (13), the desired trajectory  $\alpha_d$  and control inputs  $\mathbf{w}$  and  $\mathbf{u}_f$  are obtained for the internal subsystem control in (10) and state approximation (11).

### C. GP-Based Learned Robot Dynamics

The accurate inverse dynamics model  $f_\alpha^{-1}$  in (8) and the state model  $f_\theta$  in (11) are usually difficult to obtain. We use GP models to estimate them, and a detailed discussion will be presented in Section IV.

Fig. 2 illustrates the overall schematic of the GP-based control design. Two GP models  $gp_\alpha$  and  $gp_\theta$  are trained offline with data that are collected under open-loop control without requiring robot stabilization. A GP-based inverse dynamics control  $u_d$  (through  $v$  by (10) and an extension of (8) with GP model  $gp_\alpha$ ) drives the internal subsystem  $\Sigma_i$  in (3) to follow the desired trajectory  $\alpha_d(t)$ . Meanwhile, an MPC extended from (13) with GP model  $gp_\theta$  is used to design  $\alpha_d(t)$  (through  $\alpha_{d0}$  and  $\mathbf{w}(t)$ ) and control input  $\mathbf{u}_f$  for the external subsystem  $\Sigma_e$  in (3) to follow the given trajectory  $\theta_d$ . The MPC also incorporates the predictive variance  $\Sigma_\alpha$  from the GP distribution  $gp_\alpha$ . The overall tracking convergence and performance of the closed-loop system is analyzed and guaranteed. We are now ready to present the GP-based control design in detail.

## IV. GP-BASED CONTROL DESIGN

### A. GP-Based Inverse Dynamics Control

Instead of directly learning inverse dynamic model  $f_\alpha^{-1}$  in the form (8), we subtract both sides of (8) by  $v$  to yield  $u_d - v = f_\alpha^{-1}(\theta, \alpha, v, u_f) - v$ . We define  $\kappa_\alpha(v) := \kappa_\alpha(\theta, \alpha, v, u_f) = f_\alpha^{-1}(\theta, \alpha, v, u_f) - v$  as the model to learn. For presentation brevity, we drop other three arguments and only retain the third one in  $\kappa_\alpha(v)$ . By doing so and with (8), the robot dynamics model (9b) becomes

$$\dot{\alpha}_1 = \alpha_2, \dot{\alpha}_2 + \kappa_\alpha(\dot{\alpha}_2) = u_d. \quad (14)$$

By (14), the learning model for  $\kappa_\alpha(\dot{\alpha}_2)$  is formulated as

$$\kappa_\alpha(\dot{\alpha}_2) = u_d - \dot{\alpha}_2 \sim gp_\alpha(\theta, \alpha, \dot{\alpha}_2, u_f) \quad (15)$$

where  $gp_\alpha$  is the GP distribution to estimate  $\kappa_\alpha(\dot{\alpha}_2)$ . In (15), we use  $\dot{\alpha}_2$  in the function argument of  $\kappa_\alpha$  because it is part of training data, and when applied in control design,  $v$  is used for inference prediction. The GP model (15) is trained on a dataset that contains the input tuple  $\mathbf{x}_\alpha = \{\theta, \alpha, \dot{\alpha}_2, u_f\}$  and the output  $\mathbf{y}_\alpha = \{u_d - \dot{\alpha}_2\}$ , that is, training dataset  $\mathcal{D}_\alpha = \{\mathbf{x}_\alpha, \mathbf{y}_\alpha\}$ . For each output dimension, an individual GP model is built, and the GPs for all the dimensions are assumed to be independent of each other due to the property of fully actuated  $\alpha$  dynamics of  $\Sigma_i$  in (9b).

The control input  $u_d$  for (15) is predicted and obtained as

$$u_d = v + \mu_\alpha \quad (16)$$

where  $\mu_\alpha := \mu_\alpha(\theta, \alpha, v, u_f)$  (and  $\Sigma_\alpha := \Sigma_\alpha(\theta, \alpha, v, u_f)$ ) is the mean (and variance) of predictive Gaussian distribution  $gp_\alpha(\theta, \alpha, v, u_f) \sim \mathcal{N}(\mu_\alpha, \Sigma_\alpha)$ , which are computed from (7) with inverse dynamics control input  $v$  and MPC input  $u_f$ . Here, we drop dependence on  $(\theta, \alpha, v, u_f)$  for  $\mu_\alpha$  and  $\Sigma_\alpha$  for notation simplicity. One benefit of using the  $\kappa_\alpha(\dot{\alpha}_2)$  representation is that with a zero-mean GP model with SE kernel, when the inference input  $v$  is far away from the training data  $\mathcal{D}_\alpha$ , the inverse dynamics control (16) reduces to  $u_d = v$  due to that  $\mu_\alpha$  is close to zero. Similar to (10),  $v$  is designed as

$$v = \mathbf{w} - \frac{k_d}{\epsilon} e_{\alpha 2} - \frac{k_p}{\epsilon^2} e_{\alpha 1} + \mathbf{r} \quad (17)$$

where  $\mathbf{r}$  is an auxiliary control input that is introduced to guarantee that the controller converges robustly when using learned models  $gp_\alpha$  to approximate unknown  $\kappa_\alpha(v)$ .

Plugging (16) and (17) into (14), the  $\alpha$  error dynamics are

$$\dot{e}_\alpha = \mathbf{A}e_\alpha + \mathbf{B}[\mathbf{r} + \mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)] \quad (18)$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_n \\ -\frac{k_p}{\epsilon^2} \mathbf{I}_n & -\frac{k_d}{\epsilon} \mathbf{I}_n \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_n \end{bmatrix} \quad (19)$$

and  $\mathbf{A}$  is Hurwitz with  $k_p, k_d > 0$ . To show the convergence of  $e_\alpha$ , it is required that disturbance  $\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)$  is bounded. Note that  $\mu_\alpha$  and  $\kappa_\alpha(\dot{\alpha}_2)$  have different arguments, that is, the latter has argument  $\dot{\alpha}_2$  for training, while the former has  $v$  for running inference prediction. We first analyze the difference

$\|\mu_\alpha - \kappa_\alpha(v)\|$  to give a bound for  $\|\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)\|$ . According to Lemma 1, the following probabilistic event

$$\Pi_\alpha = \{\|\mu_\alpha - \kappa_\alpha(v)\| \leq \|\beta_\alpha^\top \Sigma_\alpha^{\frac{1}{2}}\|\} \quad (20)$$

is true with probability  $\Pr\{\Pi_\alpha\} \geq (1 - \delta)^n$ , where  $\beta_\alpha \in \mathbb{R}^n$  is  $\delta$ -dependent and given in Lemma 1. When (20) is true,  $\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)$  is upper-bounded, namely

$$\|\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)\| \leq \rho(e_\alpha, \theta) \quad (21)$$

where  $\rho(e_\alpha, \theta)$  is defined and calculated in Appendix I.A. Because of (21), the auxiliary control  $r$  is designed such that (18) is robustly stable.

**Lemma 2:** If  $k_d^2 > 4k_p > 0$ ,  $A$  in (19) is diagonalizable with  $A = M\Lambda M^{-1}$ , where  $\Lambda$  is a diagonal matrix and  $M$  is a nonsingular matrix. A positive-definite matrix  $P$  exists as the solution of  $A^\top P + PA = -M^\top M^{-1}$ . Furthermore, under (20),  $r$  is designed as

$$r = \begin{cases} -\rho(e_\alpha, \theta) \frac{B^\top P e_\alpha}{\|B^\top P e_\alpha\|}, & \text{if } \|B^\top P e_\alpha\| > \xi \\ -\frac{\rho(e_\alpha, \theta)}{\xi} B^\top P e_\alpha, & \text{if } \|B^\top P e_\alpha\| \leq \xi \end{cases} \quad (22)$$

with constant  $\xi > 0$ . Moreover, errors  $\|e_\alpha(t)\|$  satisfy

$$\|e_\alpha(t)\| \leq d_1 \|e_\alpha(0)\| e^{\frac{\lambda_1}{4\epsilon} t} + d_2 \quad (23)$$

where  $\lambda_1 = \frac{-k_d + \sqrt{k_d^2 - 4k_p}}{2}$ ,  $d_1 = \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}$ ,  $d_2 = \sqrt{-\frac{2\epsilon c_3}{\lambda_1 \lambda_{\min}(P)}}$ , and constant  $c_3 > 0$  is defined in (54).

The proof of Lemma 2 is given in Appendix I.B. Since  $\lambda_1 < 0$  and  $\epsilon > 0$  is small,  $e^{\frac{\lambda_1}{4\epsilon} t}$  converges to zero rapidly. The GP-based inverse dynamics controller (16) only uses the mean value  $\mu_\alpha$  of the predictive distribution, and the covariance  $\Sigma_\alpha$  determines the disturbance error bound  $\rho(e_\alpha, \theta)$  and control performance. We incorporate  $\Sigma_\alpha$  into the MPC design.

### B. MPC-Based Tracking and Balance Control

In this subsection, we discuss how to use MPC to obtain control input  $u_f$  and desired trajectory  $\alpha_d$ . The design can be considered as an extension of (13) with GP models.

A learned GP model is used to predict  $f_\theta$  in (11) with estimated state  $\hat{\theta}$

$$\hat{\theta}_1 = \hat{\theta}_2, \quad \hat{\theta}_2 \sim gp_\theta(\hat{\theta}, \alpha_d, w, u_f) \quad (24)$$

where  $gp_\theta$  is the GP distribution to estimate  $f_\theta$ . The GP model (24) is trained on a dataset that contains the input tuple  $x_\theta = \{\theta, \alpha, \dot{\alpha}_2, u_f\}$  and the output  $y_\theta = \{\hat{\theta}_2\}$ , that is,  $\mathcal{D}_\theta = \{x_\theta, y_\theta\}$ . The strategy is to design the MPC such that  $\hat{\theta}$  converges to  $\theta_d$ , and we then give theoretical guarantee that  $\hat{\theta}$  is an accurate approximation of  $\theta$ . We use the discrete-time representation of (12) and (24) for MPC design as

$$\hat{\theta}(k+i+1|k) \sim F_m \hat{\theta}(k+i|k) + G_m gp_\theta(k+i) \quad (25a)$$

$$\alpha_d(k+i+1) = F_n \alpha_d(k+i) + G_n w(k+i), \quad \alpha_d(k) = \alpha_{dk} \quad (25b)$$

where  $\hat{\theta}(k+i|k) = [\hat{\theta}_1(k+i|k)^\top \hat{\theta}_2(k+i|k)^\top]^\top$ ,  $i = 0, \dots, H+1$ , denotes the  $(k+i)$ th step estimate given the

$k$ th observation and  $\hat{\theta}(k|k) = \theta(k)$ ,  $H$  is the predictive horizon, and  $gp_\theta(k+i) = gp_\theta(\hat{\theta}(k+i|k), \alpha_d(k+i), w(k+i), u_f(k+i))$ .  $\alpha_d(k+i) = [\alpha_{1d}(k+i)^\top \alpha_{2d}(k+i)^\top]^\top$  are obtained from the deterministic model (12) and, therefore, not conditioned on observations.  $F_i$  and  $G_i$ ,  $i = m, n$ , are matrices with dimension  $i$  as

$$F_i = \begin{bmatrix} I_i & \Delta t I_i \\ 0 & I_i \end{bmatrix}, \quad G_i = \begin{bmatrix} 0 \\ \Delta t I_i \end{bmatrix} \quad (26)$$

where  $\Delta t$  is the sampling period.

The evolution of  $\hat{\theta}(k+i+1|k)$  is probabilistic due to the predictive GP model  $gp_\theta(k+i)$  and also the probabilistic nature of its previous state  $\hat{\theta}(k+i|k)$ . Even if  $\hat{\theta}(k+i|k)$  is Gaussian distribution,  $\hat{\theta}(k+i+1|k)$  in (25a) is not exactly Gaussian because the predictive distribution  $gp_\theta(k+i)$  is nonlinear with respect to  $\hat{\theta}(k+i|k)$ . By linearizing  $gp_\theta(k+i)$  in (25a) with respect to  $\hat{\theta}(k+i|k)$ , we approximate  $\hat{\theta}(k+i+1|k)$  with a Gaussian distribution with a closed form for mean and covariance propagation (denoted, respectively, as  $\mu_{\hat{\theta}}(k+i+1|k)$  and  $\Sigma_{\hat{\theta}}(k+i+1|k)$ ) as

$$\mu_{\hat{\theta}}(k+i+1|k) = F_m \mu_{\hat{\theta}}(k+i|k) + G_m \mu_{gp_\theta}(k+i) \quad (27a)$$

$$\Sigma_{\hat{\theta}}(k+i+1|k) = F_m \Sigma_{\hat{\theta}}(k+i|k) F_m^\top + G_m \partial \Sigma_{\hat{\theta}}(k+i) G_m^\top \quad (27b)$$

where  $\partial \Sigma_{\hat{\theta}}(k+i) = \frac{\partial \mu_{gp_\theta}}{\partial \hat{\theta}} \Sigma_{\hat{\theta}}(k+i|k) \frac{\partial \mu_{gp_\theta}^\top}{\partial \hat{\theta}} + \Sigma_{gp_\theta}(k+i)$ , and  $\mu_{gp_\theta}(k+i)$  and  $\Sigma_{gp_\theta}(k+i)$  are the mean and covariance of  $gp_\theta(k+i)$ , respectively. It is straightforward to obtain  $\|\Sigma_{gp_\theta}\| \leq \sigma_{f_{\max}}^2 := \max_{1 \leq j \leq m} (\sigma_{f_{\theta_j}}^2 + \sigma_j^2)$ , where  $j$  is the index of the dimension of  $f_\theta$ . Assuming that  $\mu_{gp_\theta}$  has a bounded gradient, with a small  $\Delta t$ , we obtain a bound of  $\Sigma_{\hat{\theta}}(k+i|k)$  as

$$\|\Sigma_{\hat{\theta}}(k+i|k)\| \leq i(\Delta t)^2 \|\Sigma_{gp_\theta}\| \leq i(\Delta t)^2 \sigma_{f_{\max}}^2. \quad (28)$$

The calculation of (28) is given in Appendix I.C.

We use  $\theta(k+i|k)$ ,  $i \in \mathbb{N}$ , to denote the predicted value of  $\theta(k+i)$  given  $\theta(k)$ . Similar to (25a), the evolution of  $\theta(k+i|k)$  under the GP model (24) with inference by  $\alpha$  (instead of  $\alpha_d$ ) is given as

$$\theta(k+i+1|k) \sim F_m \theta(k+i|k) + G_m gp_\theta(k+i) \quad (29)$$

with the mean value  $\mu_\theta(k+i+1|k)$  and variance  $\Sigma_\theta(k+i+1|k)$  calculated similar to (27), where  $gp_\theta(k+i) = gp_\theta(\theta(k+i|k), \alpha(k+i), v(k+i), u_f(k+i))$ . The main difference between models (29) and (25a) is that the former depends on the actual state  $\alpha(k+i)$ , while the latter uses the estimated  $\alpha_d(k+i)$ . Fig. 3 further illustrates the relationships among three different  $\theta$  and its estimate  $\hat{\theta}$  predictive models (9a), (25a), and (29).

At the  $k$ th step, the MPC input set is

$$W(k) = \{\alpha_{dk}, w(k+i), u_f(k+i), i = 0, \dots, H\}. \quad (30)$$

Unlike  $w(k+i)$  or  $u_f(k+i)$  with predictive values over horizon  $H$ , only the initial value  $\alpha_{dk} = \alpha_d(k)$  at the current step is included in (30), which is the same in (13). For (25), the MPC



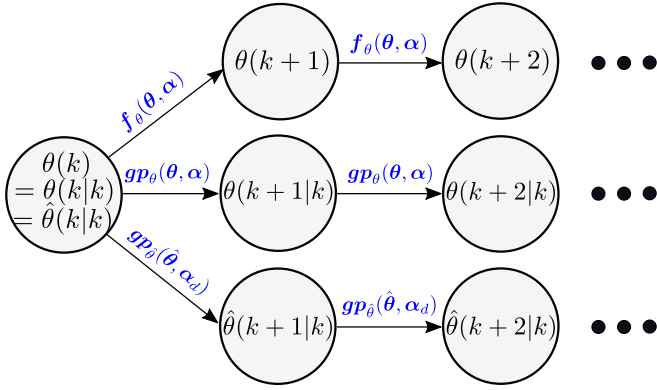


Fig. 3. Flowchart of the state variable  $\theta(k+i)$  and its estimates  $\theta(k+i|k)$  and  $\hat{\theta}(k+i|k)$ ,  $i \in \mathbb{N}$ , by three predictive models. True state  $\theta(k+i)$  is evolved with the external subsystem dynamics  $f_\theta$ ; estimates  $\theta(k+i|k)$  and  $\hat{\theta}(k+i|k)$  are predicted under the  $k$ th step values by using the GP models  $gp_\theta$  and  $gp_{\hat{\theta}}$  with the true and desired trajectories  $\alpha$  and  $\alpha_d$ , respectively. For presentation simplicity, notation  $f_\theta(\theta, \alpha)$  drops the other two arguments  $v$  and  $u_f$ , while  $gp_\theta(\theta, \alpha)$  and  $gp_{\hat{\theta}}(\hat{\theta}, \alpha_d)$  are presented by dropping the last two arguments of  $(v, u_f)$  and  $(w, u_f)$ , respectively.

cost function is designed as

$$\bar{J}_{\hat{\theta}, W}^k = \|\alpha_{dk}\|_{Q_2}^2 + \sum_{i=0}^H l_s(k+i) + l_f(k+H+1) \quad (31)$$

where the stage cost  $l_s(j)$ ,  $j \geq k$ , and the terminal cost  $l_f(k+H+1)$  are

$$l_s(j) = \mathbb{E} \|e_{\hat{\theta}}(j)\|_{Q_1}^2 + \|w(j)\|_R^2 + \|u_f(j)\|_R^2 + \|e_{\mu_{\hat{\theta}}}(j)\|_{Q_1}^2 + \text{tr}(Q_1 \Sigma_{\hat{\theta}}(j|k)) + \|w(j)\|_R^2 + \|u_f(j)\|_R^2 \quad (32a)$$

$$l_f(k+H+1) = \mathbb{E} \|e_{\hat{\theta}}(k+H+1)\|_{Q_3}^2 = \|e_{\mu_{\hat{\theta}}}(k+H+1)\|_{Q_3}^2 + \text{tr}(Q_3 \Sigma_{\hat{\theta}}(k+H+1|k)) \quad (32b)$$

$e_{\hat{\theta}}(j) = \hat{\theta}(j|k) - \theta_d(j)$ ,  $e_{\mu_{\hat{\theta}}}(j) = \mu_{\hat{\theta}}(j|k) - \theta_d(j)$ , and matrix  $Q_3$  is positive definite. We take the expectation operator in (32) because of the probabilistic variable  $\hat{\theta}(j|k)$  from (27) for  $j = k, \dots, k+H+1$ . In (32a), we use the fact that  $\mathbb{E} \|e_{\hat{\theta}}(j)\|_{Q_1}^2 = \mathbb{E} \|\hat{\theta}(j|k) - \mu_{\hat{\theta}}(j|k) + \mu_{\hat{\theta}}(j|k) - \theta_d(j)\|_{Q_1}^2 = \|e_{\mu_{\hat{\theta}}}(j)\|_{Q_1}^2 + \text{tr}(Q_1 \Sigma_{\hat{\theta}}(j|k))$ ; a similar calculation is used in (32b) and others in this article.

The cost (31) is built on the  $\hat{\theta}$ - and  $\alpha_d$ -dynamics in (25), and it does not penalize the  $e_\alpha$  convergence. It is, thus, modified as

$$\bar{J}_{\hat{\theta}, W}^k = \bar{J}_{\hat{\theta}, W}^k + \nu \|\Sigma_\alpha(k)\| \quad (33)$$

where  $\Sigma_\alpha(k)$  is the covariance of the predictive distribution  $gp_\alpha(\theta, \alpha, v, u_f)$  at the  $k$ th step and  $\nu > 0$  is a weighting factor. As expressed previously, variance  $\Sigma_\alpha(k)$  depends on control inputs (i.e.,  $W(k)$ ), and the added term  $\|\Sigma_\alpha(k)\|$  in (33) penalizes to optimize decision variables  $W(k)$  to be close to the training dataset and to reduce the tracking performance and convergence of  $e_\alpha$ . The variance term also measures the GP modeling accuracy, and this can be seen from the fact that  $d_2$  in (23) increases with large  $\|\Sigma_\alpha\|$  value.

Similar to (13), the MPC formulation is given as

$$\min_{W(k)} J_{\hat{\theta}, W}^k \quad (34a)$$

$$\text{subject to (25b) and (27), } i = 0, \dots, H. \quad (34b)$$

The above optimization is solved by a gradient decent method, and the solution is denoted as  $W^*(k) = \{\alpha_{dk}^*, w^*(k+i), u_f^*(k+i), i = 0, \dots, H\}$ . The gradient of the objective function with respect to the decision variables is obtained with the chain rule. In implementation, the states in the MPC prediction horizon are computed and constrained as the function of initial state and the inputs (i.e.,  $W(k)$ ) by the forward propagating models (25b) and (27) iteratively [38, Ch. 2].

### C. MPC Convergence Analysis

Certain terminal cost conditions of (31) need to be satisfied to guarantee the asymptotic stability of the closed-loop system under the finite-horizon optimization in (34). The work in [39] presented the terminal cost condition that guarantees the MPC stability. To apply the results in [39] to (34), the state and input variables in (27) need to be shifted such that the origin of the transformed dynamics is an equilibrium of the system.

Suppose that for given desired trajectory  $\theta_d$ , there exists a set of corresponding nominal inputs  $\{\alpha_{dd}, w_d, u_{fd}\}$  satisfying the mean propagation dynamics similar to (27a), that is,

$$\mu_{\theta_d}(k+i+1) = F_m \mu_{\theta_d}(k+i) + G_m \mu_{gp_{\theta_d}}(k+i)$$

where  $\mu_{gp_{\theta_d}}(k+i)$  is the mean of  $gp_{\theta_d}(k+i) = gp_\theta(\theta_d(k+i), \alpha_{dd}(k+i), w_d(k+i), u_{fd}(k+i))$ , and  $\alpha_{dd}(k+i)$  is obtained from (25b) under  $w_d(k+i)$  and  $\alpha_{dkd}$ . The dynamics of error  $e_{\mu_{\hat{\theta}}}(k+i) = \mu_{\hat{\theta}}(k+i|k) - \theta_d(k+i)$  is obtained by taking the difference between the above equation and (27a)

$$e_{\mu_{\hat{\theta}}}(k+i+1) = F_m e_{\mu_{\hat{\theta}}}(k+i) + G_m \Delta \mu_{gp_\theta}(k+i) \quad (35)$$

where  $\Delta \mu_{gp_\theta}(k+i) := \mu_{gp_\theta}(k+i) - \mu_{gp_{\theta_d}}(k+i)$ . Since the nominal inputs trajectory is fixed for a given  $\theta_d$ , the actual control input for (35) is the difference between the input for (27a) and the nominal input trajectory, i.e.,  $u_e(k) = [\Delta \alpha_{dk}^\top \Delta w^\top(k) \Delta u_f^\top(k)]^\top$ , where  $\Delta \alpha_{dk} = \alpha_{dk} - \alpha_{dkd}$ ,  $\Delta w(k) = w(k) - w_d(k)$  and  $\Delta u_f(k) = u_f(k) - u_{fd}(k)$ . The origin of state  $e_{\mu_{\hat{\theta}}}(k)$  and input  $u_e(k)$  of (35) is an equilibrium point. The following fictitious MPC is used to find input  $u_e$  at the  $k$ th step:

$$\min_{W(k)} \bar{J}_{\hat{\theta}, W}^{k^\circ} \quad (36)$$

where  $\bar{J}_{\hat{\theta}, W}^{k^\circ} = \|\Delta \alpha_{dk}\|_{Q_2}^2 + \sum_{i=0}^H l_s^\circ(k+i) + l_f^\circ(k+H+1)$ , and

$$l_s^\circ(j) = \|e_{\mu_{\hat{\theta}}}(j)\|_{Q_1}^2 + \|\Delta w(j)\|_{R^\circ}^2 + \|\Delta u_f(j)\|_{R^\circ}^2 \quad (37a)$$

$$l_f^\circ(k+H+1) = \|e_{\mu_{\hat{\theta}}}(k+H+1)\|_{Q_3}^2. \quad (37b)$$

Matrix  $R^\circ$  in (37a) is positive definite. The superscript “ $\circ$ ” is used in (36) and (37) to differentiate from (34) and (32), respectively.

By [39], the stability of the fictitious MPC (36) is guaranteed by the particular design of the terminal cost, i.e., matrix  $Q_3$ ,

which is the solution of a Lyapunov equation of the linearization of (35) at its equilibrium. The linearization of (35) around its equilibrium (i.e., the origin) leads to

$$\mathbf{e}_{\mu_{\hat{\theta}}}(k+i+1) = \mathbf{A}_e \mathbf{e}_{\mu_{\hat{\theta}}}(k+i) + \mathbf{B}_e \mathbf{u}_e(k+i) \quad (38)$$

where  $\mathbf{A}_e = \mathbf{F}_m + \mathbf{G}_m \frac{\partial \mu_{gp\theta_d}}{\partial \theta_d}$  and  $\mathbf{B}_e = \mathbf{G}_m \left[ \frac{\partial \mu_{gp\theta_d}^\top}{\partial \alpha_{dk}} \frac{\partial \mu_{gp\theta_d}^\top}{\partial \mathbf{w}_d} \frac{\partial \mu_{gp\theta_d}^\top}{\partial \mathbf{u}_{fd}} \right]^\top$ . The chosen  $\mathbf{Q}_3$  as the Lyapunov equation solution of (38) guarantees that the MPC (36) has a decreasing terminal cost

$$l_f^\circ(k+H+2) \leq l_f^\circ(k+H+1) - l_s^\circ(k+H+1). \quad (39)$$

By comparing (32) with (37), the following terminal cost decreasing property is obtained for the MPC in (34).

**Lemma 3:** For the stage and terminal costs  $l_s(k+i)$  and  $l_f(k+H+1)$  in (32), let  $\mathbf{R}$  satisfy  $\lambda_{\max}(\mathbf{R}) < \lambda_{\min}(\mathbf{R}^\circ)$  and  $\mathbf{Q}_3$  be the solution of the Lyapunov equation of (38). Defining  $\lambda_R = \frac{\lambda_{\min}(\mathbf{R}^\circ)\lambda_{\max}(\mathbf{R})}{\lambda_{\min}(\mathbf{R}^\circ) - \lambda_{\max}(\mathbf{R})}$ , then

$$\begin{aligned} l_f(k+H+2) &\leq l_f(k+H+1) - l_s(k+H+1) \\ &+ \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(k+H+2|k)) + \text{tr}(\mathbf{Q}_1 \Sigma_{\hat{\theta}}(k+H+1|k)) \\ &+ \lambda_R (\|\mathbf{w}_d(k+H+1)\|^2 + \|\mathbf{u}_{fd}(k+H+1)\|^2). \end{aligned} \quad (40)$$

The proof of this lemma is in Appendix I.D. The terminal cost condition (40) is a key step to ensure the following convergence property for the MPC in (34).

**Lemma 4:** Assuming bounded input trajectories for  $\theta_d$ , that is,  $\|\alpha_{dk}\|_{Q_2}^2 \leq \alpha_{d\max}^2$ ,  $\|\mathbf{w}_d(k)\| \leq \mathbf{w}_{d\max}$  and  $\|\mathbf{u}_{fd}(k)\| \leq \mathbf{u}_{fd\max}$ , under the MPC in (34), the tracking error satisfies  $\|\mathbf{e}_{\mu_{\hat{\theta}}}(k+i)\| \leq a_4(i)\|\mathbf{e}_{\theta}(k)\| + a_5(i)$ , where  $\mathbf{e}_{\theta}(k) = \theta(k) - \theta_d(k)$ ,  $a_4(i)$ , and  $a_5(i)$  are given in Appendix I.E.

The proof of Lemma 4 is in Appendix I.E. Lemma 4 gives the bound of tracking error  $\mathbf{e}_{\mu_{\hat{\theta}}}$  under the MPC in (34). However, it does not take into account the error introduced by using the GP model to approximate the unknown robot dynamics. We will show the stability of the overall GP-based control in the next section.

## V. CONTROL PERFORMANCE ANALYSIS

Under the GP-based inverse dynamics control (16) and (17) and the MPC design (34), we consider a Lyapunov function candidate

$$V(k) = V_{\theta}(k) + \zeta V_{\alpha}(k) \quad (41)$$

where constant  $\zeta > 0$ ,  $V_{\alpha}(k) = \mathbf{e}_{\alpha}^\top(k) \mathbf{P} \mathbf{e}_{\alpha}(k)$ ,  $\mathbf{P}$  is defined in Lemma 2,  $V_{\theta}(k) = \bar{J}_{\theta, W^*}^k$  is taken similar to the cost  $\bar{J}_{\theta, W^*}^k$  in (31) under  $\mathbf{W}^*(k)$  but with actual  $\theta(k)$  from (9a), namely

$$\begin{aligned} \bar{J}_{\theta, W^*}^k &= \sum_{i=0}^H \|\mathbf{e}_{\theta}(k+i)\|_{Q_1}^2 + \|\mathbf{w}^*(k+i)\|_R^2 + \|\mathbf{u}_f^*(k+i)\|_R^2 \\ &+ \|\alpha_{dk}^*\|_{Q_2}^2 + \|\mathbf{e}_{\theta}(k+H+1)\|_{Q_3}^2 \end{aligned} \quad (42)$$

and  $\mathbf{e}_{\theta}(k+i) = \theta(k+i) - \theta_d(k+i)$ . It is impossible to directly evaluate above  $\bar{J}_{\theta, W^*}^k$  because of inaccessible states  $\theta(k+i)$ , and its value is instead approximated by  $\bar{J}_{\theta, W^*}^k$ .

Using (33) and (42), the Lyapunov function decrease  $\Delta V(k) = V(k+1) - V(k)$  is calculated as

$$\begin{aligned} \Delta V(k) &= \left( \bar{J}_{\theta, W^*}^{k+1} - \bar{J}_{\theta, W^*}^{k+1} \right) - \left( \bar{J}_{\theta, W^*}^k - \bar{J}_{\theta, W^*}^k \right) \\ &+ \left( J_{\theta, W^*}^{k+1} - J_{\theta, W^*}^k \right) + \zeta [V_{\alpha}(k+1) - V_{\alpha}(k)] \\ &- \nu [\|\Sigma_{\alpha^*}(k+1)\| - \|\Sigma_{\alpha^*}(k)\|] \end{aligned} \quad (43)$$

where  $\Sigma_{\alpha^*}(k)$  is the covariance  $\Sigma_{\alpha}(k)$  under control  $\mathbf{W}^*(k)$ . In (43), term  $\bar{J}_{\theta, W^*}^k - \bar{J}_{\theta, W^*}^k$  quantifies the difference between the approximated and actual cost-to-go at the  $k$ th step.

As shown in Fig. 3, we quantify the difference between  $\hat{\theta}(k+i|k)$  and  $\theta(k+i|k)$ . At  $i=0$ ,  $\hat{\theta}(k|k) = \theta(k|k) = \theta(k)$ . The difference between  $\hat{\theta}(k+i|k)$  and  $\theta(k+i|k)$  comes from the predictions by (27a) and (29). Assuming that the mean value function of the predictive distribution  $\mu_{gp\theta}(\mu_{\theta}, \alpha)^1$  is Lipschitz in  $\mu_{\theta}$  and  $\alpha$ , we have

$$\|\mu_{gp\theta}(\cdot, \alpha) - \mu_{gp\theta}(\cdot, \alpha_d)\| \leq L_2 \|\mathbf{e}_{\alpha}\| \quad (44a)$$

$$\|\mu_{gp\theta}(\mu_{\theta}, \cdot) - \mu_{gp\theta}(\mu_{\hat{\theta}}, \cdot)\| \leq L_3 \|\mu_{\theta} - \mu_{\hat{\theta}}\| \quad (44b)$$

with constants  $L_2, L_3 > 0$ ; error  $\tilde{\mu}_{\theta}(k+i) = \mu_{\theta}(k+i|k) - \mu_{\hat{\theta}}(k+i|k)$  satisfies  $\|\tilde{\mu}_{\theta}(k+i)\| \leq \varrho_{\hat{\theta}}(i)\|\mathbf{e}_{\alpha}(k)\| + \varrho_2(i)$ , where  $\varrho_{\hat{\theta}}(i)$  and  $\varrho_2(i)$  are given by Lemma 5 in Appendix I.F. We then inspect the difference between  $\theta(k+i|k)$  and  $\theta(k+i)$ , which comes from the predictions between (9a) and (29), as shown in Fig. 3. Similar to (20) for  $\kappa_{\alpha}$ , if event

$$\Pi_{\theta} = \{\|\mu_{gp\theta}(\theta, \alpha) - \mathbf{f}_{\theta}(\theta, \alpha)\| \leq \|\beta_{\theta}^\top \Sigma_{gp\theta}^{\frac{1}{2}}\|\} \quad (45)$$

for the GP estimate  $\mu_{gp\theta}(\theta, \alpha)$  for  $\mathbf{f}_{\theta}(\theta, \alpha)$  is true with high probability, that is,  $\Pr\{\Pi_{\theta}\} \geq (1-\delta)^m$ ,  $0 < \delta < 1$ , the upper bound of  $\theta_{\mu}(k+i) = \mu_{\theta}(k+i|k) - \theta(k+i)$  is given by Lemma 6 in Appendix I.F as  $\|\theta_{\mu}(k+i)\| \leq \varrho_{\mu_{\theta}}(i) = \Delta t \sum_{j=0}^{i-1} \|\beta_{\theta}^\top \Sigma_{gp\theta}^{\frac{1}{2}}(k+j|k)\|$ . In (45),  $\beta_{\theta} \in \mathbb{R}^m$  is  $\delta$ -dependent and given in Lemma 1. Combining these results, we have the upper bound of  $\tilde{\theta}_{\mu}(k+i) = \mu_{\hat{\theta}}(k+i|k) - \theta(k+i)$  as

$$\begin{aligned} \|\tilde{\theta}_{\mu}(k+i)\| &\leq \|\tilde{\mu}_{\theta}(k+i)\| + \|\theta_{\mu}(k+i)\| \\ &\leq \varrho_{\hat{\theta}}(i)\|\mathbf{e}_{\alpha}(k)\| + a_2(i) \end{aligned}$$

where  $a_2(i) = \varrho_2(i) + \varrho_{\mu_{\theta}} > 0$ . We then obtain the bound for  $|\bar{J}_{\theta, W^*}^k - \bar{J}_{\theta, W^*}^k|$ , as shown in Lemma 7 in Appendix I.F.

We denote  $\mathbf{e}(k) = [\mathbf{e}_{\theta}^\top(k) \mathbf{e}_{\alpha}^\top(k)]^\top$  as the systems error at the  $k$ th step,  $\underline{\lambda} = \min(\lambda_{\min}(\mathbf{Q}_1), \zeta \lambda_{\min}(\mathbf{Q}))$  and  $\bar{\lambda} = \max(\lambda_{\max}(\mathbf{Q}_1), \zeta \lambda_{\max}(\mathbf{Q}))$ , where matrices  $\mathbf{Q}$  and  $\mathbf{Q}_1$  are defined in Lemma 2 and (31), respectively. We have the following main result with proof given in Appendix I.G.

**Theorem 1:** For parameters  $\bar{\xi}_j(i)$ ,  $i=0, 1, \dots, H+2$ ,  $j=1, \dots, 5$ , given in Lemma 7, defining  $\xi_j = \bar{\lambda}[\bar{\xi}_j(0) + 2 \sum_{i=1}^{H+1} \bar{\xi}_j(i) + \bar{\xi}_j(H+2)]$ ,  $\gamma_1 = \sqrt{\eta}$ ,  $\gamma_2 = \frac{\xi_3}{2\gamma_1}$ ,

<sup>1</sup>For presentation simplicity, we drop the third and four arguments and use notation  $\mu_{gp\theta}(\mu_{\theta}, \alpha)$  to represent  $\mu_{gp\theta}(\mu_{\theta}, \alpha, \mathbf{v}, \mathbf{u}_f)$  and  $\mu_{gp\theta}(\mu_{\theta}, \alpha_d)$  for  $\mu_{gp\theta}(\mu_{\theta}, \alpha_d, \mathbf{w}, \mathbf{u}_f)$ . The same notation is applied to both  $\mu_{gp\theta}$  and  $\mathbf{f}_{\theta}$  in (45) with dropped  $(\mathbf{v}, \mathbf{u}_f)$  as the third and fourth arguments.



$$\begin{aligned} \gamma_3 &= \sqrt{\lambda_{\min}(\mathbf{Q}_1)}, \quad \gamma_4 = \frac{\xi_4}{\gamma_3}, \quad \text{and} \quad \gamma_5 = \frac{\xi_4^2}{\gamma_3} + \frac{\xi_3^2}{4\gamma_1^2} + \xi_5 + \\ &\hat{\alpha}_{\max}^2 + \nu\sigma_{\kappa_{\max}}^2 + \zeta c_3 \Delta t + m\lambda_m(H+2)(\Delta t)^2\sigma_f^2, \\ \lambda_m &= \lambda_{\max}(\mathbf{Q}_1) + \lambda_{\max}(\mathbf{Q}_3), \text{ and} \\ \eta &= \frac{1}{4}\zeta\lambda_{\min}(\mathbf{Q})\Delta t - \xi_1 - \frac{\xi_2^2}{2\lambda_{\min}(\mathbf{Q}_1)} - \frac{\lambda_{\min}(\mathbf{Q}_1)}{4} > 0 \end{aligned} \quad (46)$$

the following property is then held:

$$V(k+1) \leq \gamma_\lambda V(k) + \gamma_5 \quad (47)$$

where  $0 < \gamma_\lambda = 1 - \frac{\gamma_3^2}{4\lambda} < 1$ .

Applying (47) iteratively for  $i$  consecutive steps, we have

$$V(k+i) \leq \gamma_\lambda^i V(k) + \frac{4\gamma_5\lambda(1-\gamma_\lambda^i)}{\gamma_3^2}. \quad (48)$$

Introducing the steady-state values  $V_{ss} = \lim_{i \rightarrow \infty} V(k+i)$  and  $e_{ss} = \lim_{i \rightarrow \infty} \|e(k+i)\|$  for any fixed  $k$ , taking the limit on both sides of (48) as  $i \rightarrow \infty$  results  $V_{ss} \leq \frac{4\lambda}{\gamma_3^2}\gamma_5$ . From (60) in Appendix I.G,  $\underline{\lambda}e_{ss}^2 \leq V_{ss}$  and, therefore, combining with the above result, we have  $e_{ss} \leq \sqrt{\frac{4\lambda}{\gamma_3^2\lambda}}\gamma_5$ . Theorem 1 implies that the error magnitude decreases exponentially until reaching the region that satisfies  $\|e(t)\| \leq e_{ss}$ . Condition (46) is satisfied by choosing a small value for  $\epsilon$ . As the value of  $\epsilon$  is small,  $\lambda_{\min}(\mathbf{Q})$  becomes large according to Lemma 2 and  $\varrho_{\hat{\theta}}$  is small by Lemma 5, and henceforth, both  $\xi_1$  and  $\xi_2$  values are small. GP learning model errors are important factors for control performance since the results in Theorem 1 are obtained under conditions (20) and (45). As the error bound  $\|\beta_\alpha^T \Sigma_\alpha^{\frac{1}{2}}\|$  for  $\kappa_\alpha$  estimation increases, values of  $d_2$  and  $\varrho_2(i)$  increase,  $a_2(i)$  increases,  $\bar{\xi}_3, \bar{\xi}_4$ , and  $\bar{\xi}_5$  increase,  $\gamma_5$  increases, and, finally, the bound of  $e_{ss}$  increases. Similarly, as the error bound  $\|\beta_\theta^T \Sigma_{gp\theta}^{\frac{1}{2}}\|$  for  $f_\theta$  estimation increases, the values of  $\varrho_{\mu_\theta}(i)$  and  $a_2(i)$  increase; therefore, both  $\gamma_5$  and the bound of  $e_{ss}$  increase.

## VI. EXPERIMENTS

The proposed robot control is implemented and demonstrated independently using a Furuta pendulum and a bikebot.

### A. Experimental Testbeds

The Furuta pendulum shown in Fig. 1(a) is a commercial robotic platform provided by Quanser, Inc. The actuated joint is the base angle  $\theta$  that is driven by a motor. The unactuated joint is the pendulum angle  $\alpha$ , and its value is defined to be zero when the pendulum arm is vertically upright. The applied motor voltage  $V_m$  is the control input. The control goal is to balance the pendulum around the upright position, while the rotary base tracks a desired trajectory  $\theta_{1d}$ . The motion of the external subsystem is captured by  $\theta_1 = \theta$  and  $\theta_2 = \dot{\theta}$ , and for the internal subsystem,  $\alpha_1 = \alpha$  and  $\alpha_2 = \dot{\alpha}$ . Defining  $\theta = [\theta_1 \ \theta_2]^T$ ,  $\alpha = [\alpha_1 \ \alpha_2]^T$ , and  $u_d = V_m$ , the dynamics model is

$$\begin{cases} \dot{\theta}_1 = \theta_2, \quad \dot{\theta}_2 = f_\theta(\theta, \alpha, u_d) \\ \dot{\alpha}_1 = \alpha_2, \quad \dot{\alpha}_2 + \kappa_\alpha(\theta, \alpha, \dot{\alpha}_2) = u_d \end{cases} \quad (49)$$

where functions  $f_\theta$  and  $\kappa_\alpha$  are given in Appendix II.A.

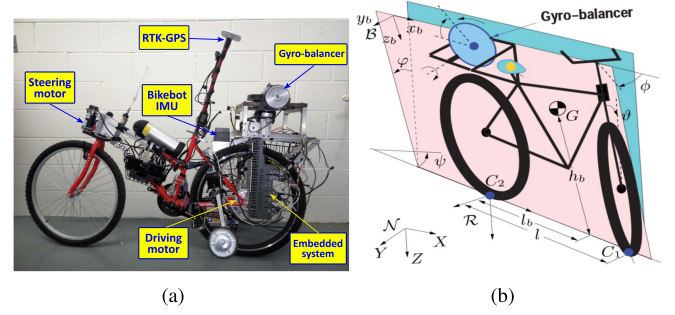


Fig. 4. (a) Autonomous bikebot. (b) Schematic of the bikebot model.

The bikebot shown in Figs. 1(b) and 4(a) is a single-track mobile robot [8], [40]. The detailed description of the system can be found in [8]. Fig. 4(b) illustrates the modeling configuration of the bikebot. The bikebot platform consists of a main body structure (with the rear wheel) and a front wheel. The rear wheel contact point is denoted as  $C_2$ , and its planar position is denoted as  $\theta_1 = [X \ Y]^T$  in the inertial frame  $\mathcal{N}$ . The yaw and roll angles of the bikebot are denoted as  $\psi$  and  $\varphi$ , respectively. The steering angle is denoted as  $\phi$ , and the magnitude of linear velocity at  $C_2$  is denoted as  $v_c$ . Due to the nonholonomic constraint of  $C_2$ , its velocity is obtained as  $\theta_2 = [\dot{X} \ \dot{Y}]^T = [v_c \cos \psi \ v_c \sin \psi]^T$ . The external subsystem motion is captured by  $\theta = [\theta_1^T \ \theta_2^T]^T$  and the internal subsystem motion  $\alpha = [\alpha_1 \ \alpha_2]^T$ ,  $\alpha_1 = \varphi$ , and  $\alpha_2 = \dot{\varphi}$ . The control inputs are  $u = [u_d \ u_f]^T$  with  $u_d = \phi$  and  $u_f = \dot{v}_c$ . The bikebot dynamics model is written in the form of (14) as

$$\begin{cases} \dot{\theta}_1 = \theta_2, \quad \dot{\theta}_2 = f_\theta(\theta, \alpha, u_d, u_f) \\ \dot{\alpha}_1 = \alpha_2, \quad \dot{\alpha}_2 + \kappa_\alpha(\theta, \alpha, \dot{\alpha}_2, u_f) = u_d \end{cases} \quad (50)$$

where  $f_\theta$  and  $\kappa_\alpha$  are given in Appendix II.A.

### B. Experimental Results

1) *Furuta Pendulum Experiments*: To obtain the learned model, we perturbed the system to collect training data. An open-loop control was implemented as

$$V_m = \begin{cases} a_1 \sin(\omega_1 t) + a_2 \sin(\omega_2 t), & |\alpha| \leq \frac{\pi}{3} \\ 0, & |\alpha| > \frac{\pi}{3} \end{cases} \quad (51)$$

where  $a_1$  and  $a_2$  were chosen to satisfy the voltage bound  $|V_m| \leq 5$  V, and  $\omega_1$  and  $\omega_2$  were designed to excite the system by both low and high frequencies. In the experiment, we chose  $a_1 = 3$ ,  $a_2 = 1.5$ ,  $\omega_1 = 8$  rad/s, and  $\omega_2 = 40$  rad/s as an example. Under this input, we swung up the pendulum manually by giving an initial velocity when angle  $|\alpha| \geq \frac{\pi}{3}$ . The input (51) cannot stabilize the pendulum around the upright position. Fig. 5 shows an example of  $\theta$  and  $\alpha$  angles under (51). For each swing, angle  $\alpha$  often stayed in the range of  $|\alpha| \leq \frac{\pi}{3}$  for less than 1 s and then fell.

Joint angles  $\theta$  and  $\alpha$  were measured with encoders, and their velocities and accelerations were obtained by numerical differentiation. The open-loop controller and data collection were implemented at a frequency of 100 Hz. The closed-loop controller was implemented through MATLAB Real-Time Workshop with

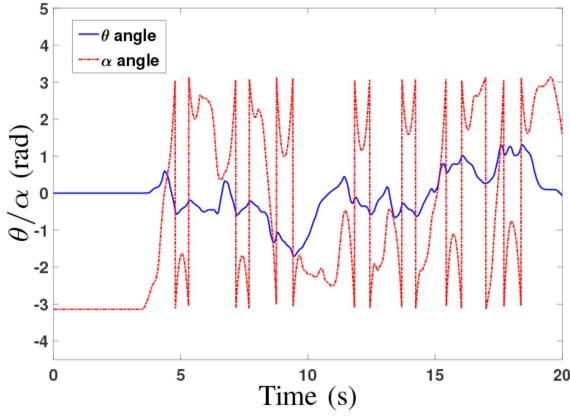


Fig. 5. Example profiles of the collected training data for Furuta pendulum experiments (under open-loop control).

a frequency of 50 Hz, that is,  $\Delta t = 20$  ms, and a preview horizon of  $H = 27$ . The weight matrices in (31) were chosen as  $\mathbf{Q}_1 = \mathbf{Q}_3 = \text{diag}\{1000, 100\}$ ,  $\mathbf{Q}_2 = \text{diag}\{100, 100\}$ ,  $\mathbf{R} = 10\mathbf{I}_2$ , and  $\nu = 1$ . The chosen  $\mathbf{Q}_3$  value satisfied the stability condition in [39] to obtain the property in (39). The inverse dynamics control parameters were chosen as  $k_p = k_d = 5$  and  $\epsilon = 0.1$ .

A set of 800 points was collected and used as the training data. In testing experiments, the desired trajectory was first designed as a composite sine  $\theta_{1d} = 0.6 \sin(t) + 0.4 \sin(4t)$  rad and  $\theta_d = [\theta_{1d} \dot{\theta}_{1d}]^T$ . We chose this smooth curve as a representative profile to demonstrate the performance. Fig. 6(a) shows the tracking results of the base angle  $\theta$  and Fig. 6(b) for the roll angle  $\alpha$ . For comparison purposes, the physical model-based EIC control in [2] was implemented and included in the plots. The EIC control was used as the benchmark, and other physical model-based controls (e.g., sliding-mode control [5], orbital stabilization [4], [6], etc.) produced similar results. The parameter values of the physical model were obtained from the vendor's manual and also validated by experiments. In Fig. 6(b), the BEMs  $\alpha_{1d}$  for the pendulum link angle show similar profiles under the learning control and the EIC control. The  $\alpha_d$  profile under the learning control was obtained from the MPC design. Fig. 6(c) and (d) compares the tracking errors  $e_\theta$  and  $e_\alpha$  [with their corresponding  $\alpha_{1d}$  shown in Fig. 6(b)] under these two controllers. Fig. 6(e) and (f) shows the control input  $u_d$  and the MPC solution  $w$  for the learning control, respectively. Fig. 7(a) and (b) further shows the error mean and standard deviation profiles over multiple experimental runs. We also implemented tracking a square wave  $\theta_{1d} = 0.6 \text{sgn}(\sin(t))$  rad to show the performance under a sharp change trajectory, where function  $\text{sgn}(x) = 1$  for  $x \geq 0$  and  $-1$  for  $x < 0$ . Fig. 7(c) and (d) similarly shows the error mean and standard deviation profiles. Table I lists the root-mean-square errors (RMSEs) for both the trajectory profiles under the two controllers. It is clear that the learning-based design outperforms the physical model-based controller with more than a 50% reduction in both mean values and variance of errors for both the trajectories.

TABLE I  
RMSES AND THEIR STANDARD DEVIATIONS OF THE BASE AND ROLL ANGLES UNDER TWO CONTROLLERS FOR THE FURUTA PENDULUM

| Trajectories   | EIC control       |                   | Learning control  |                   |
|----------------|-------------------|-------------------|-------------------|-------------------|
|                | $\theta$ (degree) | $\alpha$ (degree) | $\theta$ (degree) | $\alpha$ (degree) |
| Composite sine | $11.4 \pm 8.1$    | $1.6 \pm 1.2$     | $4.2 \pm 2.9$     | $1.5 \pm 1.1$     |
| Square wave    | $30.4 \pm 10.6$   | $3.3 \pm 2.1$     | $11.4 \pm 5.7$    | $1.5 \pm 1.3$     |

TABLE II  
RMSES (STANDARD DEVIATIONS) OF THE TRACKING POSITION AND ROLL ANGLE COMPARISON UNDER TWO CONTROLLERS FOR THE BIKEBOT

| Trajectories  | EIC control   |               | Learning control |               |
|---------------|---------------|---------------|------------------|---------------|
|               | Position (m)  | Roll (degree) | Position (m)     | Roll (degree) |
| Straight-line | $0.6 \pm 0.2$ | $2.3 \pm 1.2$ | $0.3 \pm 0.1$    | $0.9 \pm 0.5$ |
| Sinusoidal    | $0.9 \pm 0.5$ | $4.1 \pm 2.3$ | $0.4 \pm 0.3$    | $2.0 \pm 1.1$ |
| Circular      | $0.9 \pm 0.3$ | $2.5 \pm 1.5$ | $0.7 \pm 0.2$    | $1.7 \pm 1.1$ |

2) *Bikebot Experiments*: Because the falling experiments would severely damage the hardware platform, for training data collection, the EIC controller was used to track sinusoidal-shape trajectories [8]. Different sinusoidal-shape trajectories were used as  $X_d = v_d t$ ,  $Y_d = A_y \sin(\frac{2\pi}{T_y} t)$ , where  $v_d = 2$  m/s is the  $x$ -direction desired velocity,  $A_y$  is the magnitude around the  $y$ -direction, and period  $T_y = 3.5$  s. The training data were collected from seven different experiment trials, each lasting about 7 s. In these experiments,  $A_y$  values were chosen from 0.2 to 0.5 m, and the use of these different trajectories aimed to perturb the bikebot dynamics. The use of the EIC control for training data collection was mainly for experiment convenience. Other open-loop control strategies might work as well provided that a fall-protection mechanism is used in the experiment without changing system dynamics.

Using the trained model, we conducted the experiments to track various desired trajectories, such as straight lines, sinusoidal (peak-to-peak amplitude around 0.8 m), and circular (radius of around 3.8 m) shapes. For comparison purposes, we also included the results under the EIC controller. Fig. 8 shows the comparison results under the learning-based and EIC controllers for a sample experimental run. It is clear that the trajectory tracking results under the learning-based control outperform those under the EIC controller [see Fig. 8(a)–(c)]. Similarly, the results shown in Fig. 8(d)–(f) demonstrate that the roll angles under the learning-based control oscillated less significantly than those under the EIC controller. The learning-based controller also demonstrates faster reaction in circular tracking than the EIC controller.

Fig. 9 further shows the tracking errors and roll angle errors under the two controllers over the path arc length. In the figure, we plot the mean and standard deviation of the trajectory and roll angle tracking errors by using five experimental trials. Fig. 9(a)–(c) shows the error and deviation profiles for the straight-line, sinusoidal, and circular trajectories, respectively. The RMSEs are listed in Table II under both the controllers. These results confirm again that the learning-based control outperforms the EIC control. Fig. 9(d)–(f) shows the roll angle tracking errors for

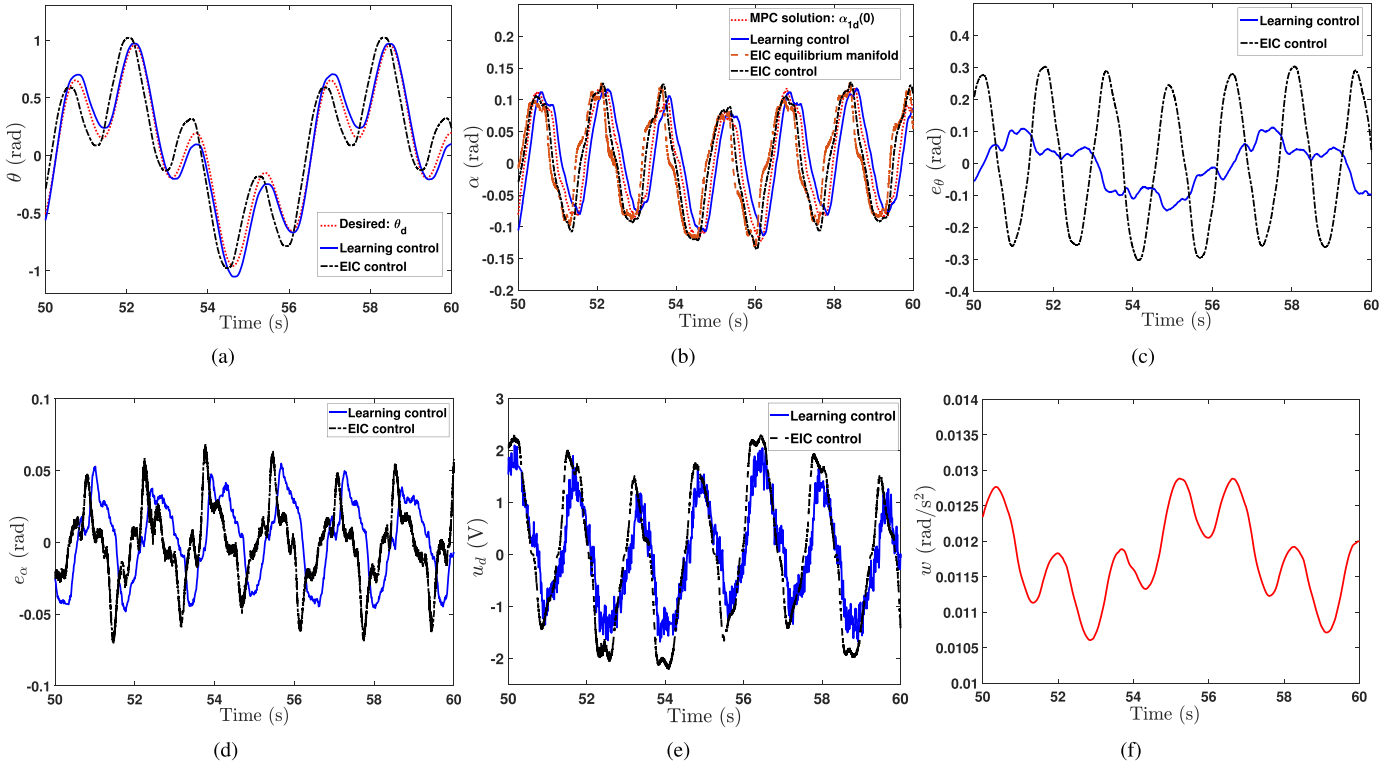


Fig. 6. One experimental performance under the learning-based and EIC controllers for the Furuta pendulum. (a) Rotary link angle  $\theta$  tracking profiles. (b) Pendulum link angle  $\alpha$  tracking profiles. The balance equilibrium angles are calculated differently for the physical model and the GP-based learning model.  $\alpha_{1d}(0)$  is the first element of  $\alpha_{d0}$ . (c) Tracking error  $e_\theta = \theta - \theta_{1d}$ . (d) Tracking error  $e_\alpha = \alpha - \alpha_{1d}$ , where  $\alpha_{1d}$  (the first element of  $\alpha_d$ ) is the BEMs that are obtained by the physical model (for EIC control) or the GP-based model (for learning control). (e) Control inputs  $u_d$  (motor voltage). (f) MPC solution  $w$  in the learning control design.

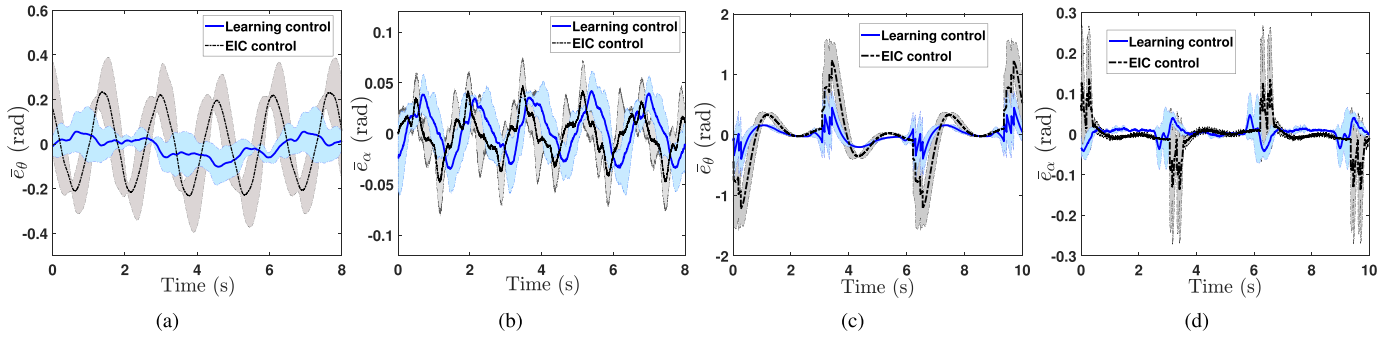


Fig. 7. Mean tracking errors  $\bar{e}_\theta$  and  $\bar{e}_\alpha$  with standard deviation profiles by multiple experimental runs under two controllers for the Furuta pendulum. (a) and (b) For tracking the composite sine. (c) and (d) For the square wave trajectories.

three types of trajectories in multiple runs. The roll angle error magnitudes and variances under the learning-based controller are much lower than those under the EIC controller. In Table II, we also list the RMSEs for the roll angles during these runs, and the calculations confirm small variations under the learning control, as shown in the figures.

To understand the influence of training data on control performance, the sizes of training datasets varied from 200 to 800 points to obtain different learned models in pendulum experiments. These models were used to track the same previously used

composite sine trajectory. Fig. 10(a) shows the error distribution under different sizes of training datasets. These contours in the figure are plotted as the smallest convex coverage of the corresponding error data points. For each learned model, the plot includes tracking errors of a 90-s duration. The results clearly imply that with 200 data points, the controller barely achieved the balancing and tracking tasks with large errors. With increased training data points, the magnitudes of both the balancing and tracking errors decreased. With a set of 800 training data points, the learned model-based controller achieved



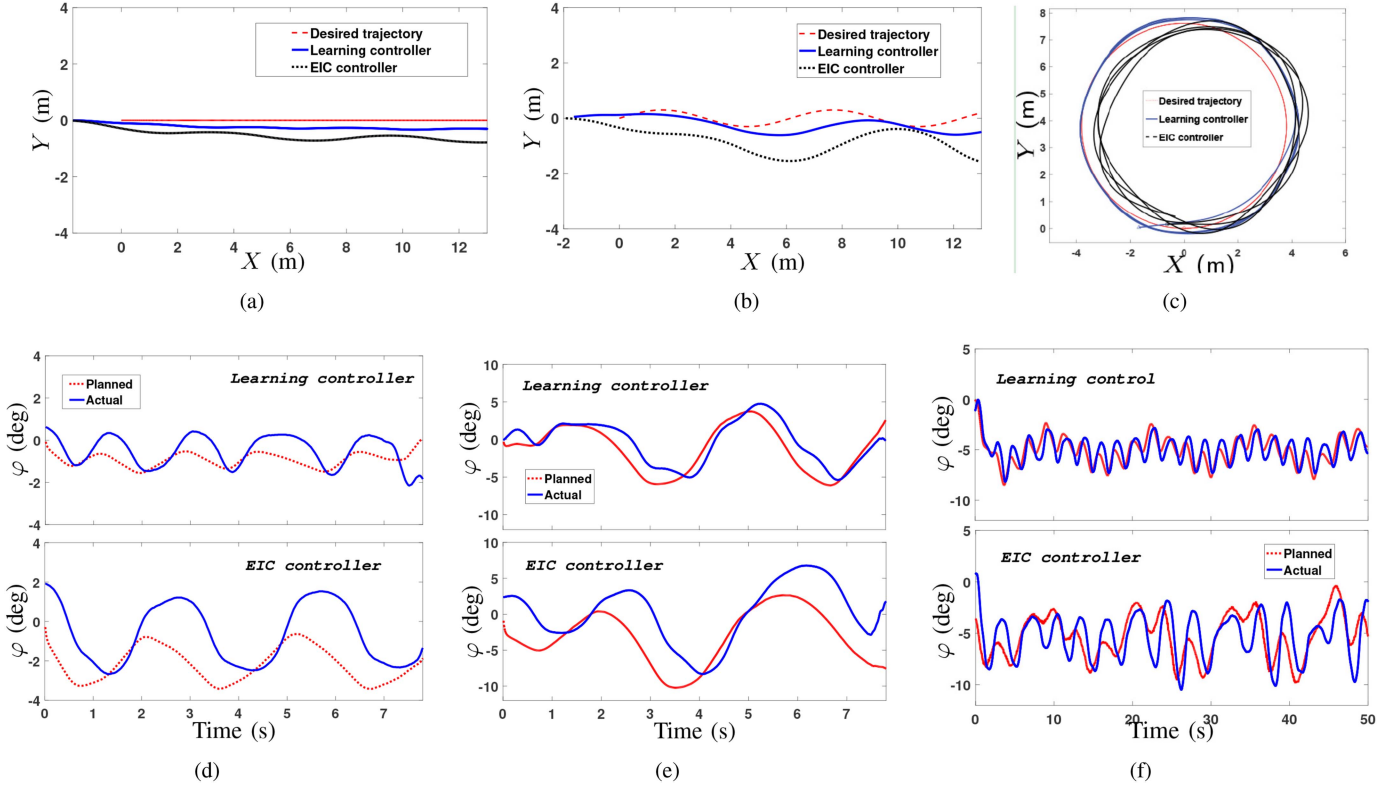


Fig. 8. Performance comparison of the bikebot tracking under the learning-based control and the EIC control designs for one experimental run. (a)–(c) For  $XY$  position tracking profiles and (d)–(f) for roll angle profiles for straight-line, sinusoidal, and circular trajectories, respectively.

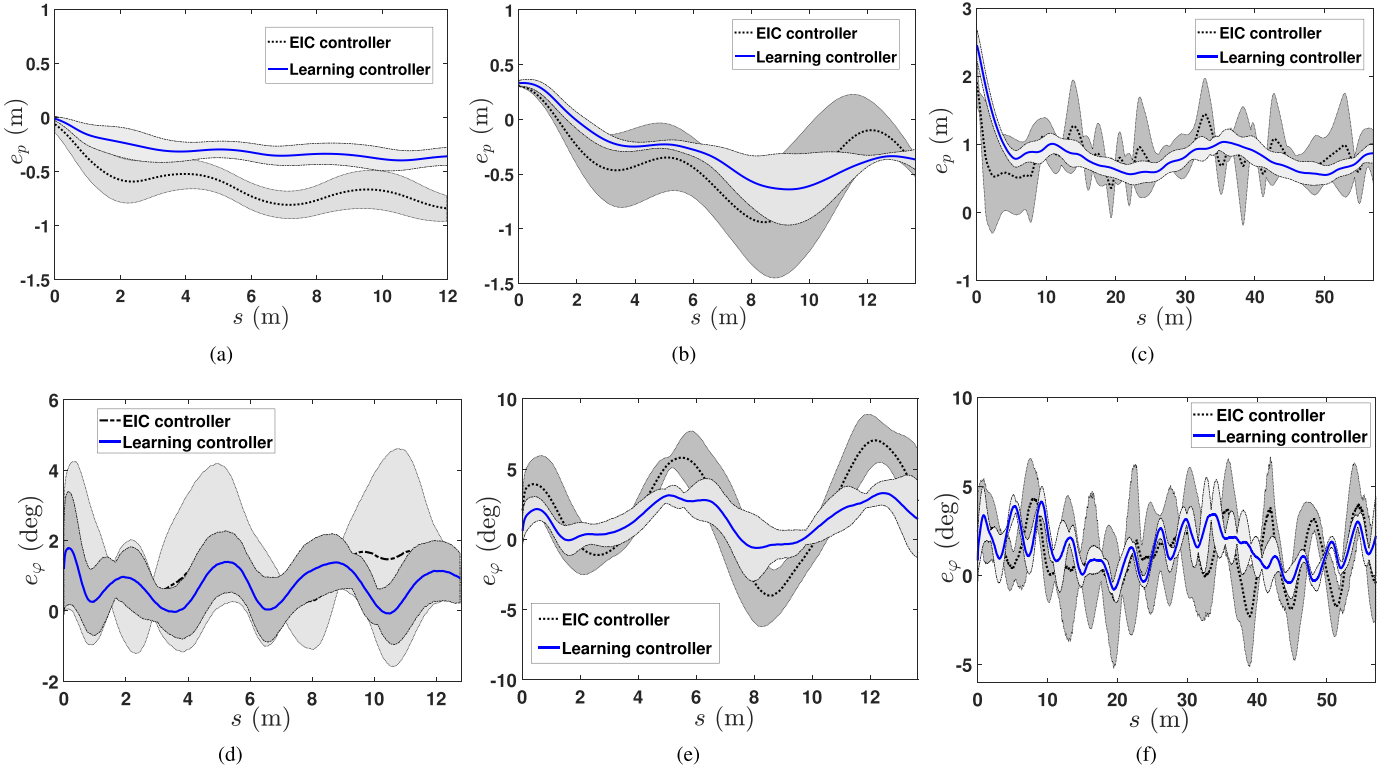


Fig. 9. Bikebot position and roll angle tracking error profiles with multiple experimental runs under the learning-based and the EIC controllers. (a)–(c) For the position tracking error  $e_p$  profiles and (d)–(f) for roll angle error  $e_\phi$  profiles for straight-line, sinusoidal, and circular trajectories, respectively. All the results are plotted over the path arc length  $s$ .

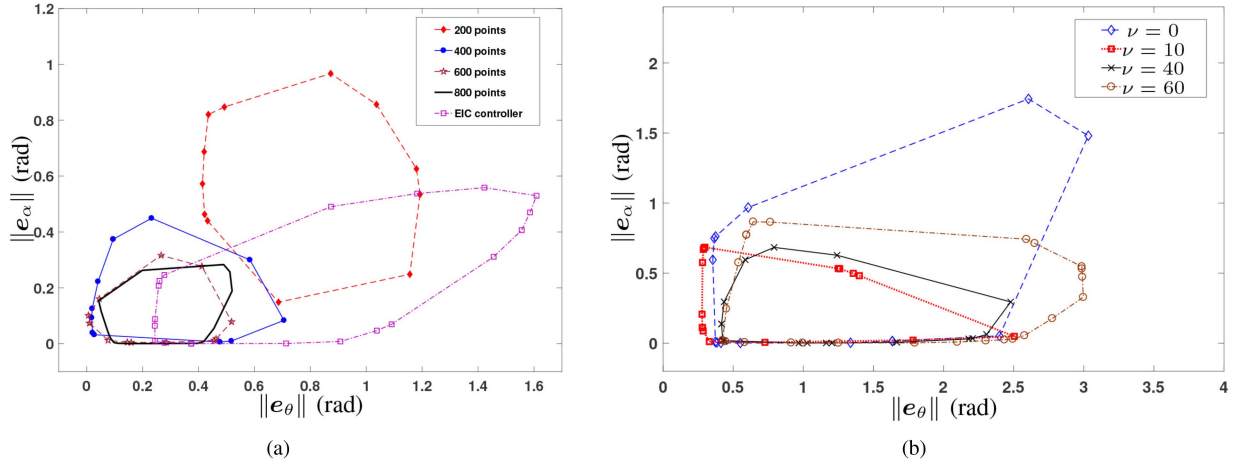


Fig. 10. (a) Comparison results of the errors  $\|e_\alpha\|$  and  $\|e_\theta\|$  under the learning-based control by various training data points and the EIC control for the Furuta pendulum. (b) Comparison of the balance and tracking errors  $\|e_\theta\|$  and  $\|e_\alpha\|$  under various values of the weight factor  $\nu$ .

superior performance relative to the EIC controller. The plots in Fig. 10(a) also confirm the error analysis and trend given by Theorem 1.

The tradeoff between tracking and balancing performance is tuned by the choice of  $\nu$  value in (33). Experiments were conducted on the Furuta pendulum to demonstrate the performance under different values of  $\nu$ . We chose a slightly inaccurate learned model obtained by using 200 training data points, and the value of  $\|\Sigma_\alpha\|$  in (33) was relatively large. Fig. 10(b) shows the contours of tracking and balancing errors with different  $\nu$  values. When  $\nu = 0$ , the system shows large error distributions due to the poor inverse dynamics model. With  $\nu = 10$ , the system achieves a good tradeoff between balancing and tracking tasks. With a further increased  $\nu$  value (i.e.,  $\nu = 40, 60$ ), the tracking performance becomes similar or slightly worse than those with  $\nu = 10$ , and when  $\nu > 80$ , the controller even fails to balance the pendulum. The average variances of the inverse dynamics model for 60-s trials are  $0.255^\circ$ ,  $0.174^\circ$ ,  $0.108^\circ$ , and  $0.108^\circ$  for  $\nu = 0, 10, 40, 60$ , respectively. The results clearly show that with increased  $\nu$  values, the magnitude of  $\Sigma_\alpha$  decreases. This confirms that the integration of  $\Sigma_\alpha$  in the cost function helps improve control performance.

## VII. CONCLUSION

In this article, we proposed a GP-based learning control design for underactuated balance robots. One characteristic of underactuated balance robots is that the equilibria of the internal subsystem depend on the tracking of the external subsystem trajectory. The control design consisted of an integrated trajectory tracking of the external subsystem and the stabilization of the internal subsystem. The trajectory tracking of the external subsystem was designed through a GP-based MPC method, while a GP-based inverse dynamics controller was used to simultaneously predict and stabilize the internal subsystem trajectory. The GP models were used to estimate system dynamics and to provide predictive distribution of model uncertainties. The learned GP models were obtained without prior knowledge about the robot dynamics

or successful balance demonstration. Moreover, the stability and closed-loop performance were guaranteed through control design and analysis. We demonstrated the controller design independently using a Furuta pendulum and a bikebot testbed.

## APPENDIX I TECHNICAL RESULTS AND PROOFS

### A. Calculation of (21)

We calculate  $\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)$  by Taylor expansion as

$$\begin{aligned} \mu_\alpha - \kappa_\alpha(\dot{\alpha}_2) &= \mu_\alpha - \kappa_\alpha(v) + \kappa_\alpha(v) - \kappa_\alpha(\dot{\alpha}_2) \\ &= \mu_\alpha - \kappa_\alpha(v) - \frac{\partial \kappa_\alpha}{\partial v} [\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)] + O(\|\dot{\alpha}_2 - v\|^2). \end{aligned}$$

The second equality results from (14) and (16). The positive-definite matrix  $A_\kappa = I_n + \frac{\partial \kappa_\alpha}{\partial v}$  is the linearization of the left-hand side of (14). From the above equation, we have

$$\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2) = A_\kappa^{-1} [O(\|\dot{\alpha}_2 - v\|^2) + \mu_\alpha - \kappa_\alpha(v)].$$

Taking the norm of both the sides, if event  $\Pi_\alpha$  is true and considering that  $O(\|\dot{\alpha}_2 - v\|^2) \leq c_2 \|e_\alpha\|^2 + c_1 \|e_\alpha\| + c_0$ , with constants  $c_i > 0$ ,  $i = 0, 1, 2$ , we obtain

$$\|\mu_\alpha - \kappa_\alpha(\dot{\alpha}_2)\| \leq \rho(e_\alpha, \theta)$$

$$\text{where } \rho(e_\alpha, \theta) = \lambda_{\min}^{-1}(A_\kappa) \left( \sum_{i=0}^2 c_i \|e_\alpha\|^i + \|\beta_\alpha^\top \Sigma_\alpha^{\frac{1}{2}}\| \right).$$

### B. Proof of Lemma 2

To show that  $A$  is diagonalizable, introducing

$$M = \begin{bmatrix} \epsilon I_n & \epsilon I_n \\ \lambda_1 I_n & \lambda_2 I_n \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \frac{\lambda_1}{\epsilon} I_n & 0 \\ 0 & \frac{\lambda_2}{\epsilon} I_n \end{bmatrix} \quad (52)$$

where  $\lambda_{1,2} = \frac{-k_d \pm \sqrt{k_d^2 - 4k_p}}{2} < 0$ , it is clear that  $A = M\Lambda M^{-1}$ . Since  $\Lambda$  is Hurwitz and diagonal, it is straightforward to find a positive-definite matrix  $P_\alpha$  such that  $\Lambda^\top P_\alpha + P_\alpha \Lambda = -I_{2n}$ .  $P_\alpha$  has  $n$  eigenvalues at  $-\frac{\epsilon}{2\lambda_1} > 0$  and the

other  $n$  eigenvalues at  $-\frac{\epsilon}{2\lambda_2} > 0$ . Let  $\mathbf{P} = \mathbf{M}^{-\top} \mathbf{P}_\alpha \mathbf{M}^{-1}$  and  $\mathbf{A}^\top \mathbf{P} + \mathbf{P} \mathbf{A} = -\mathbf{M}^{-\top} \mathbf{M}^{-1}$ .

Introducing  $\mathbf{e}_\alpha = \mathbf{M} \mathbf{e}_{\alpha'}$ , error dynamics (18) becomes

$$\dot{\mathbf{e}}_{\alpha'} = \mathbf{\Lambda} \mathbf{e}_{\alpha'} + \mathbf{M}^{-1} \mathbf{B} [\mathbf{r} + \boldsymbol{\mu}_\alpha - \boldsymbol{\kappa}_\alpha(\dot{\boldsymbol{\alpha}}_2)]. \quad (53)$$

We choose the Lyapunov function candidate  $V_{\alpha'} = \mathbf{e}_{\alpha'}^\top \mathbf{P}_\alpha \mathbf{e}_{\alpha'}$ . It is clear that  $-\frac{\epsilon}{2\lambda_2} \|\mathbf{e}_{\alpha'}\|^2 \leq V_{\alpha'} \leq -\frac{\epsilon}{2\lambda_1} \|\mathbf{e}_{\alpha'}\|^2$  and

$$\dot{V}_{\alpha'} = -\mathbf{e}_{\alpha'}^\top \mathbf{e}_{\alpha'} + 2(\mathbf{B}^\top \mathbf{M}^{-\top} \mathbf{P}_\alpha \mathbf{e}_{\alpha'})^\top [\mathbf{r} + \boldsymbol{\mu}_\alpha - \boldsymbol{\kappa}_\alpha(\dot{\boldsymbol{\alpha}}_2)].$$

Defining  $V_\alpha = \mathbf{e}_\alpha^\top \mathbf{P} \mathbf{e}_\alpha$ , by the choice of  $\mathbf{r}$  in (22) with  $\xi = \frac{\lambda_{\min}(\mathbf{A}_\alpha)}{c_2 \|\mathbf{M}\|^2}$ , we obtain

$$\begin{aligned} \dot{V}_{\alpha'} \leq & -\|\mathbf{e}_{\alpha'}\|^2 + \frac{\xi \rho(\mathbf{e}_\alpha, \boldsymbol{\theta})}{2} = -\frac{1}{2} \|\mathbf{e}_{\alpha'}\|^2 + \frac{c_1}{2c_2 \|\mathbf{M}\|} \|\mathbf{e}_{\alpha'}\| \\ & + \frac{c_0}{2c_2 \|\mathbf{M}\|^2} + \frac{\|\beta_\alpha^\top \Sigma_\alpha^{1/2}\|}{2c_2 \|\mathbf{M}\|^2} = -\frac{1}{4} \|\mathbf{e}_{\alpha'}\|^2 - \frac{1}{4} \left( \|\mathbf{e}_{\alpha'}\| \right. \\ & \left. - \frac{c_1}{c_2 \|\mathbf{M}\|} \right)^2 + c_3 \leq -\frac{1}{4} \|\mathbf{e}_{\alpha'}\|^2 + c_3 \end{aligned} \quad (54)$$

where  $c_3 = \frac{1}{4} \frac{c_1^2}{c_2^2 \|\mathbf{M}\|^2} + \frac{1}{2} \frac{c_0}{c_2 \|\mathbf{M}\|^2} + \frac{1}{2} \frac{\|\beta_\alpha^\top \Sigma_\alpha^{1/2}\|}{2c_2 \|\mathbf{M}\|^2} > 0$ .

From (54), we obtain  $\dot{V}_{\alpha'} \leq \frac{\lambda_1}{2\epsilon} V_{\alpha'} + c_3$ , and therefore

$$V_{\alpha'}(t) \leq V_{\alpha'}(0) e^{\frac{\lambda_1}{2\epsilon} t} - \frac{2\epsilon}{\lambda_1} c_3. \quad (55)$$

Noting that  $V_\alpha = V_{\alpha'}$  and  $\lambda_{\min}(\mathbf{P}) \|\mathbf{e}_\alpha\|^2 \leq V_\alpha \leq \lambda_{\max}(\mathbf{P}) \|\mathbf{e}_\alpha\|^2$ , from (55), we have

$$\begin{aligned} \|\mathbf{e}_\alpha(t)\| & \leq \sqrt{\frac{\lambda_{\max}(\mathbf{P})}{\lambda_{\min}(\mathbf{P})}} \|\mathbf{e}_\alpha(0)\| e^{\frac{\lambda_1}{4\epsilon} t} + \sqrt{-\frac{2\epsilon c_3}{\lambda_1 \lambda_{\min}(\mathbf{P})}} \\ & = d_1 \|\mathbf{e}_\alpha(0)\| e^{\frac{\lambda_1}{4\epsilon} t} + d_2. \end{aligned}$$

This proves the lemma.

### C. Calculation of (28)

Taking norm on both sides of (27b) and applying the upper bound of the gradient  $\|\frac{\partial \mu_{gp\hat{\theta}}}{\partial \hat{\theta}}\| \leq L_1$ , we obtain  $\|\Sigma_{\hat{\theta}}(k+i+1|k)\| \leq (\|\mathbf{F}_m\|^2 + \|\mathbf{G}_m\|^2 L_1^2) \|\Sigma_{\hat{\theta}}(k+i|k)\| + \|\mathbf{G}_m\|^2 \sigma_{\mathbf{f}}^2$ . Applying the process iteratively with  $\Sigma_{\hat{\theta}}(k|k) = \mathbf{0}$ , we have

$$\|\Sigma_{\hat{\theta}}(k+i|k)\| \leq \frac{1 - (\|\mathbf{F}_m\|^2 + \|\mathbf{G}_m\|^2 L_1^2)^i}{1 - (\|\mathbf{F}_m\|^2 + \|\mathbf{G}_m\|^2 L_1^2)} \|\mathbf{G}_m\|^2 \sigma_{\mathbf{f}}^2.$$

From (26),  $\|\mathbf{F}_m\| = \sqrt{1 + \frac{\Delta t}{2} (\Delta t + \sqrt{(\Delta t)^2 + 4})}$ ,  $\|\mathbf{G}_m\| = \Delta t$  and for  $\Delta t \ll 1$ , taking  $\|\mathbf{F}_m\| \approx 1$  and that  $(1+x)^n \approx 1 + nx$  for  $|x| \ll 1$ , we obtain (28).

### D. Proof of Lemma 3

The terminal and stage costs in (32) and (37) are related as  $l_f^\circ(j) = l_f(j) - \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(j|k))$  and  $l_s^\circ(j) = l_s(j) - \text{tr}(\mathbf{Q}_1 \Sigma_{\hat{\theta}}(j|k)) - \|\mathbf{w}(j)\|_R^2 - \|\mathbf{u}_f(j)\|_R^2 + \|\Delta \mathbf{w}(j)\|_{R^\circ}^2 + \|\Delta \mathbf{u}_f(j)\|_{R^\circ}^2$ . Plugging the above results into (39), we obtain

$$l_f(j+1) \leq l_f(j) + \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(j+1|k)) - \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(j|k))$$

$$\begin{aligned} & -l_s(j) + \text{tr}(\mathbf{Q}_1 \Sigma_{\hat{\theta}}(j|k)) + \|\mathbf{w}_d(j) + \Delta \mathbf{w}(j)\|_R^2 \\ & + \|\mathbf{u}_{fd}(j) + \Delta \mathbf{u}_f(j)\|_R^2 - \|\Delta \mathbf{w}(j)\|_{R^\circ}^2 - \|\Delta \mathbf{u}_f(j)\|_{R^\circ}^2 \end{aligned}$$

where  $j = k + H + 1$ . With  $\lambda_{\max}(\mathbf{R}) < \lambda_{\min}(\mathbf{R}^\circ)$ , we obtain

$$\begin{aligned} & \|\mathbf{w}_d(j) + \Delta \mathbf{w}(j)\|_R^2 - \|\Delta \mathbf{w}(j)\|_{R^\circ}^2 \\ & \leq \lambda_{\max}(\mathbf{R}) \|\mathbf{w}_d(j) + \Delta \mathbf{w}(j)\|^2 - \lambda_{\min}(\mathbf{R}^\circ) \|\Delta \mathbf{w}(j)\|^2 \\ & \leq \lambda_R \|\mathbf{w}_d(j)\|^2. \end{aligned}$$

Similarly,  $\|\mathbf{u}_{fd}(j) + \Delta \mathbf{u}_f(j)\|_R^2 - \|\Delta \mathbf{u}_f(j)\|_{R^\circ}^2 \leq \lambda_R \|\mathbf{u}_{fd}(j)\|^2$ . Combining the above two inequalities with the fact that  $\text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(j|k)) \geq 0$ , we obtain (40), and this proves the lemma.

### E. Proof of Lemma 4

We choose  $J_{\hat{\theta}, W^*}^k$  under  $\mathbf{W}^*(k)$  in (34) as the Lyapunov function candidate and first show the decreasing property of  $J_{\hat{\theta}, W^*}^k$ . Inspired by the design in [38, Ch. 2.4], we take the technique to construct a following intermediary policy  $\mathbf{W}^e(k+1)$  extended from computed  $\mathbf{W}^*(k)$  as

$$\begin{aligned} \mathbf{W}^e(k+1) & = \{\boldsymbol{\alpha}_{d(k+1)}^e, \mathbf{w}^*(k+i), \mathbf{u}_f^*(k+i), \\ & \quad \mathbf{w}^e(k+H+1), \mathbf{u}_f^e(k+H+1), i = 1, \dots, H\} \end{aligned}$$

where  $\boldsymbol{\alpha}_{d(k+1)}^e = \mathbf{F}_1 \boldsymbol{\alpha}_{dk}^* + \mathbf{G}_1 \mathbf{w}^*(k)$ , and  $\mathbf{w}^e(k+H+1)$  and  $\mathbf{u}_f^e(k+H+1)$  are chosen from admissible input set. The choice of the above design guarantees that  $\mathbf{W}^e(k+1)$  contains the same terms of  $\mathbf{W}^*(k)$  for  $i = 1, \dots, H$ . Consequently, the predicted states at the  $(k+1)$ th control cycle  $\boldsymbol{\mu}_{\hat{\theta}}^e(k+i|k+1)$ ,  $\Sigma_{\hat{\theta}}^e(k+i|k+1)$  by (27) under  $\mathbf{W}^e(k+1)$  are the same as those at the  $k$ th control cycle under control  $\mathbf{W}^*(k)$  i.e.  $\boldsymbol{\mu}_{\hat{\theta}}^e(k+i|k+1) = \boldsymbol{\mu}_{\hat{\theta}}^*(k+i|k)$ ,  $\Sigma_{\hat{\theta}}^e(k+i|k+1) = \Sigma_{\hat{\theta}}^*(k+i|k)$  for  $i = 1, \dots, H+1$ .

Let  $l_s^e(k+i)$  ( $l_f^e(k+i)$ ) and  $l_s^*(k+i)$  ( $l_f^*(k+i)$ ) denote the stage (terminal) costs under  $\mathbf{W}^e(k+1)$  and  $\mathbf{W}^*(k)$ , respectively. It is straightforward to obtain that  $l_s^e(k+i) = l_s^*(k+i)$  for  $i = 1, \dots, H$  and  $l_f^e(k+H+1) = l_f^*(k+H+1)$

$$\begin{aligned} J_{\hat{\theta}, W^e}^{k+1} - J_{\hat{\theta}, W^*}^k & = l_s^e(k+H+1) + l_f^e(k+H+2) - l_s^*(k) \\ & \quad - l_f^*(k+H+1) + \nu \Delta \Sigma_\alpha(k) + \Delta \alpha_{dk} \end{aligned}$$

where  $\Delta \Sigma_\alpha(k) = \|\Sigma_\alpha^e(k+1)\| - \|\Sigma_\alpha^*(k)\|$  and  $\Delta \alpha_{dk} = \|\boldsymbol{\alpha}_{d(k+1)}^e\|_{Q_2}^2 - \|\boldsymbol{\alpha}_{dk}^*\|_{Q_2}^2$ . By Lemma 3, we have

$$\begin{aligned} J_{\hat{\theta}, W^e}^{k+1} - J_{\hat{\theta}, W^*}^k & \leq -l_s^*(k) + \nu \Delta \Sigma_\alpha(k) + \Delta \alpha_{dk} + \\ & \quad \text{tr}(\mathbf{Q}_1 \Sigma_{\hat{\theta}}(k+H+1|k)) + \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(k+H+2|k)) \\ & \quad + \lambda_R (\|\mathbf{w}_d(k+H+1)\|^2 + \|\mathbf{u}_{fd}(k+H+1)\|^2). \end{aligned}$$

Because of  $J_{\hat{\theta}, W^*}^{k+1} \leq J_{\hat{\theta}, W^e}^{k+1}$  by optimality, from the above result, we have

$$\begin{aligned} J_{\hat{\theta}, W^*}^{k+1} - J_{\hat{\theta}, W^*}^k & \leq -\lambda_{\min}(\mathbf{Q}_1) \|\mathbf{e}_{\mu_{\hat{\theta}}}(k)\|^2 + \nu \Delta \Sigma_\alpha(k) \\ & \quad + \Delta \alpha_{dk} + \text{tr}(\mathbf{Q}_1 \Sigma_{\hat{\theta}}(k+H+1|k)) \\ & \quad + \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(k+H+2|k)) + \lambda_R (\|\mathbf{w}_d(k+H+1)\|^2 \end{aligned}$$



$$+ \|\mathbf{u}_{fd}(k+H+1)\|^2. \quad (56)$$

From (28),  $\Sigma_{\hat{\theta}}(k+H+1|k) \leq (H+1)(\Delta t)^2 \sigma_{f \max}^2$ . Also,  $\|\Sigma_{\alpha}^e(k+1)\| \leq \max_{1 \leq i \leq n} (\sigma_{\alpha_i}^2 + \sigma_i^2) := \sigma_{\kappa \max}^2$ . Applying the above inequalities and the nominal input bounds, we have

$$\begin{aligned} J_{\hat{\theta}, W^*}^{k+1} - J_{\hat{\theta}, W^*}^k &\leq -\lambda_{\min}(\mathbf{Q}_1) \|\mathbf{e}_{\mu_{\hat{\theta}}}(k)\|^2 + \nu \sigma_{\kappa \max}^2 + \alpha_{d \max}^2 \\ &+ m[\lambda_{\max}(\mathbf{Q}_1) + \lambda_{\max}(\mathbf{Q}_3)](H+2)(\Delta t)^2 \sigma_{f \max}^2 \\ &+ \lambda_R(w_{d \max}^2 + u_{fd \max}^2). \end{aligned} \quad (57)$$

Furthermore, we have  $J_{\hat{\theta}, W^*}^{k+1} \geq \lambda_{\min}(\mathbf{Q}_1) \|\mathbf{e}_{\mu_{\hat{\theta}}}(k+1)\|^2$ . By the monotonicity of the value function (see [38, Lemma 2.15]), we have

$$\begin{aligned} J_{\hat{\theta}, W^*}^k &\leq l_f(k) + \|\alpha_{dk}^*\|_{\mathbf{Q}_2}^2 + \nu \|\Sigma_{\alpha}(k)\| \\ &\leq \lambda_{\max}(\mathbf{Q}_3) \|\mathbf{e}_{\mu_{\hat{\theta}}}(k)\|^2 + \alpha_{d \max}^2 + \nu \sigma_{\kappa \max}^2. \end{aligned}$$

Substituting the above inequalities into (57), we obtain  $J_{\hat{\theta}, W^*}^{k+1} \leq d_3 J_{\hat{\theta}, W^*}^k + d_4$ , where  $d_3 = 1 - \frac{\lambda_{\min}(\mathbf{Q}_1)}{\lambda_{\max}(\mathbf{Q}_3)}$ ,  $d_4 = m\lambda_m(H+2)(\Delta t)^2 \sigma_{f \max}^2 + (2-d_3)(\nu \sigma_{\kappa \max}^2 + \alpha_{d \max}^2) + \lambda_R(w_{d \max}^2 + u_{fd \max}^2)$ . Therefore,  $J_{\hat{\theta}, W^*}^{k+i} \leq d_3^i J_{\hat{\theta}, W^*}^k + d_4 \frac{1-d_3^i}{1-d_3}$ , and consequently,  $\|\mathbf{e}_{\mu_{\hat{\theta}}}(k+i)\| \leq a_4(i) \|\mathbf{e}_{\theta}(k)\| + a_5(i)$ , where  $a_4(i) = d_3^{\frac{i}{2}} \sqrt{\frac{\lambda_{\max}(\mathbf{Q}_3)}{\lambda_{\min}(\mathbf{Q}_1)}}$ ,  $a_5(i) = \sqrt{\frac{d_3^i (\alpha_{d \max}^2 + \nu \sigma_{\kappa \max}^2) + d_4 \frac{1-d_3^i}{1-d_3}}{\lambda_{\min}(\mathbf{Q}_1)}}$ , and  $\lambda_m = \lambda_{\max}(\mathbf{Q}_1) + \lambda_{\max}(\mathbf{Q}_3)$ . This proves the lemma.

#### F. Other Technical Lemmas

**Lemma 5:** Given (44),  $\|\tilde{\mu}_{\theta}(k+i)\| \leq \varrho_{\hat{\theta}}(i) \|\mathbf{e}_{\alpha}(k)\| + \varrho_2(i)$ , where

$$\varrho_{\hat{\theta}}(i) = d_1 L_2 \Delta t \left[ \left( \frac{1-a_1^i}{1-a_1} - i \right) \left( 1 - \frac{L_3 \Delta t}{1-a_1} \right) + i \right]$$

$a_1 = e^{\frac{\lambda_1}{4\epsilon} \Delta t}$  and  $\varrho_2(i) = d_2 L_2 \Delta t [i + \frac{1}{2} L_3 \Delta t (i-1)]$ .  $d_1$ ,  $d_2$ , and  $\lambda_1$  are defined in Lemma 2.

*Proof:* Denoting  $j = k+i$  and plugging (27a) for  $\mu_{\hat{\theta}}(j|k)$  and counterpart from (29) for  $\mu_{\theta}(j|k)$  into  $\tilde{\mu}_{\theta}(j) = \mu_{\theta}(j|k) -$

$\mu_{\hat{\theta}}(j|k)$ , we obtain

$$\begin{aligned} \|\tilde{\mu}_{\theta}(j)\| &= \|\mathbf{F}_m \tilde{\mu}_{\theta}(j-1) + \mathbf{G}_m (\mu_{gp_{\hat{\theta}}}(j) - \mu_{gp_{\theta}}(j))\| \\ &\leq \|\mathbf{F}_m\| \|\tilde{\mu}_{\theta}(j-1)\| + \|\mathbf{G}_m\| \|\mu_{gp_{\hat{\theta}}}(\mu_{\theta}(j-1), \alpha(j-1)) \\ &\quad - \mu_{gp_{\theta}}(\mu_{\hat{\theta}}(j-1), \alpha_d(j-1))\| \\ &= \|\mathbf{F}_m\| \|\tilde{\mu}_{\theta}(j-1)\| + \|\mathbf{G}_m\| \|\mu_{gp_{\theta}}(\mu_{\theta}(j-1), \alpha(j-1)) \\ &\quad - \mu_{gp_{\theta}}(\mu_{\theta}(j-1), \alpha_d(j-1)) + \mu_{gp_{\theta}}(\mu_{\theta}(j-1), \alpha_d(j-1)) \\ &\quad - \mu_{gp_{\theta}}(\mu_{\hat{\theta}}(j-1), \alpha_d(j-1))\| \\ &\leq \|\mathbf{F}_m\| \|\tilde{\mu}_{\theta}(j-1)\| + \|\mathbf{G}_m\| \times (L_2 \|\mathbf{e}_{\alpha}(j-1)\| \\ &\quad + L_3 \|\tilde{\mu}_{\theta}(j-1)\|) \\ &= (\|\mathbf{F}_m\| + L_3 \|\mathbf{G}_m\|) \|\tilde{\mu}_{\theta}(j-1)\| + L_2 \|\mathbf{G}_m\| \|\mathbf{e}_{\alpha}(j-1)\|. \end{aligned}$$

We use the Lipschitz condition above. When  $i = 1$ , with the fact that  $\mu_{\theta}(k|k) = \mu_{\hat{\theta}}(k|k)$ , we have  $\tilde{\mu}_{\theta}(k+1) \leq L_2 \Delta t \|\mathbf{e}_{\alpha}(k)\|$ . For  $i \geq 2$ , applying the above process iteratively, we obtain

$$\begin{aligned} \tilde{\mu}_{\theta}(k+i) &\leq \sum_{j=0}^{i-1} (\|\mathbf{F}_m\| + L_3 \|\mathbf{G}_m\|)^{i-j-1} L_2 \|\mathbf{G}_m\| \\ &\quad \times \|\mathbf{e}_{\alpha}(k+j)\|. \end{aligned}$$

From Lemma 2, we have  $\|\mathbf{e}_{\alpha}(k+j)\| \leq d_1 e^{\frac{\lambda_1}{4\epsilon} j \Delta t} \|\mathbf{e}_{\alpha}(k)\| + d_2$ . For  $\Delta t \ll 1$ ,  $\|\mathbf{F}_m\| \approx 1$  and  $\|\mathbf{G}_m\| = \Delta t$ , using  $(1 + L_3 \Delta t)^{i-j-1} \approx 1 + (i-j-1)L_3 \Delta t$  for small  $L_3 \Delta t \ll 1$ , we obtain the upper bound, as shown in the lemma. ■

**Lemma 6:** Under condition (45),  $\|\theta_{\mu}(k+i)\| \leq \varrho_{\mu_{\theta}}(i) := \Delta t \sum_{j=0}^{i-1} \|\beta_{\theta}^T \Sigma_{gp_{\theta}}^{\frac{1}{2}}(k+j|k)\|$ .

*Proof:* Substituting the iterative model similar to (27) for both  $\mu_{\theta}(k+i|k)$  and  $\theta(k+i)$ , the error is reduced to

$$\begin{aligned} \theta_{\mu}(k+i) &= \mathbf{F}_m \theta_{\mu}(k+i-1) + \mathbf{G}_m [\mu_{gp_{\theta}}(k+i-1|k) \\ &\quad - \mathbf{f}_{\theta}(k+i-1)]. \end{aligned}$$

Taking the norm on both sides of the above equation and using approximation  $\|\mathbf{F}_m\| \approx 1$ ,  $\|\mathbf{G}_m\| = \Delta t$ , and event in (45) is true with high probability, we have

$$\|\theta_{\mu}(k+i)\| \leq \|\theta_{\mu}(k+i-1)\| + \Delta t \|\beta_{\theta}^T \Sigma_{gp_{\theta}}^{\frac{1}{2}}(k+i-1|k)\|.$$

---


$$\begin{aligned} \bar{J}_{\hat{\theta}, W^*}^k - \bar{J}_{\hat{\theta}, W^*}^k &= \sum_{i=0}^H \mathbb{E} \|\mathbf{e}_{\hat{\theta}}(k+i|k)\|_{\mathbf{Q}_1}^2 - \|\mathbf{e}_{\theta}(k+i)\|_{\mathbf{Q}_1}^2 + \mathbb{E} \|\mathbf{e}_{\hat{\theta}}(k+H+1)\|_{\mathbf{Q}_3}^2 - \|\mathbf{e}_{\theta}(k+H+1)\|_{\mathbf{Q}_3}^2 \\ &= \sum_{i=0}^H \|\mathbf{e}_{\mu_{\hat{\theta}}}(k+i)\|_{\mathbf{Q}_1}^2 + \text{tr}(\mathbf{Q}_1 \Sigma_{\hat{\theta}}(k+i)) - \|\mathbf{e}_{\theta}(k+i)\|_{\mathbf{Q}_1}^2 + \|\mathbf{e}_{\mu_{\hat{\theta}}}(k+H+1|k)\|_{\mathbf{Q}_1}^2 + \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(k+H+1)) \\ &\quad - \|\mathbf{e}_{\theta}(k+H+1)\|_{\mathbf{Q}_3}^2 \\ &= \sum_{i=0}^H -\|\tilde{\theta}_{\mu}(k+i)\|_{\mathbf{Q}_1}^2 + \text{tr}(\mathbf{Q}_1 \Sigma_{\hat{\theta}}(k+i|k)) + 2\tilde{\theta}_{\mu}^T(k+i) \mathbf{Q}_1 \mathbf{e}_{\mu_{\hat{\theta}}}(k+i) - \|\tilde{\theta}_{\mu}(k+H+1)\|_{\mathbf{Q}_3}^2 + \text{tr}(\mathbf{Q}_3 \Sigma_{\hat{\theta}}(k+H+1|k)) \\ &\quad + 2\tilde{\theta}_{\mu}^T(k+H+1) \mathbf{Q}_3 \mathbf{e}_{\mu_{\hat{\theta}}}(k+H+1). \end{aligned} \quad (58)$$

With  $\mu_\theta(k|k) = \theta(k)$ , recursively using the above inequality, we obtain  $\|\theta_\mu(k+i)\| \leq \varrho_{\mu_\theta}(i)$ . ■

**Lemma 7:** Under (20) and (45),  $|\bar{J}_{\hat{\theta}, W^*}^k - \bar{J}_{\theta, W^*}^k| \leq \rho_J(e_\alpha, e_\theta)$ , where

$$\rho_J(e_\alpha, e_\theta) = \lambda_{\max}(\mathbf{Q}_3) \sum_{i=0}^{H+1} \left\{ \bar{\xi}_1(i) \|e_\alpha(k)\|^2 + \bar{\xi}_3(i) \|e_\alpha(k)\| \right. \\ \left. + \bar{\xi}_2(i) \|e_\alpha(k)\| \|e_\theta(k)\| + \bar{\xi}_4(i) \|e_\theta(k)\| + \bar{\xi}_5(i) \right\}$$

$\bar{\xi}_1(i) = \varrho_{\hat{\theta}}^2(i)$ ,  $\bar{\xi}_2(i) = 2\varrho_{\hat{\theta}}(i)a_4(i)$ ,  $\bar{\xi}_3(i) = 2\varrho_{\hat{\theta}}(i)[a_2(i) + a_5(i)]$ ,  $\bar{\xi}_4(i) = 2a_2(i)a_4(i)$ , and  $\bar{\xi}_5(i) = a_2(i)(a_2(i) + 2a_5(i)) + m(\Delta t)^2\sigma_{f_{\max}}^2$ .  $\varrho_{\hat{\theta}}(i)$  is defined in Lemmas 5,  $a_4(i)$  and  $a_5(i)$  are defined in Lemma 4.

*Proof:* Using (31) and (42), we obtain (58) shown at the bottom of the previous page. The last equality comes from the observation

$$\|e_{\mu_{\hat{\theta}}}\|_{Q_1}^2 - \|e_\theta\|_{Q_1}^2 = \|e_{\mu_{\hat{\theta}}}\|_{Q_1}^2 - \|e_{\mu_{\hat{\theta}}} - \tilde{\theta}_\mu\|_{Q_1}^2 \\ = \|e_{\mu_{\hat{\theta}}}\|_{Q_1}^2 - \|e_{\mu_{\hat{\theta}}}\|_{Q_1}^2 - \|\tilde{\theta}_\mu\|_{Q_1}^2 + 2\tilde{\theta}_\mu^\top Q_1 e_{\mu_{\hat{\theta}}} \\ = -\|\tilde{\theta}_\mu\|_{Q_1}^2 + 2\tilde{\theta}_\mu^\top Q_1 e_{\mu_{\hat{\theta}}}$$

with  $e_{\mu_{\hat{\theta}}} = \mu_{\hat{\theta}} - \theta_d$ ,  $\tilde{\theta}_\mu = \mu_{\hat{\theta}} - \theta$ , and  $e_\theta = \theta - \theta_d$ . For presentation simplicity, we drop the discrete-time index for each error variable above. The rationale to use the above formulation is to obtain a bound on  $\bar{J}_{\hat{\theta}, W^*}^k - J_{\theta, W^*}^k$  by terms  $\|\tilde{\theta}_\mu\|$  and  $\|e_{\mu_{\hat{\theta}}}\|$ .

Since  $\lambda_{\max}(\mathbf{Q}_1) < \lambda_{\max}(\mathbf{Q}_3)$ , from (58), we have

$$|\bar{J}_{\hat{\theta}, W^*}^k - \bar{J}_{\theta, W^*}^k| \leq \lambda_{\max}(\mathbf{Q}_3) \sum_{i=0}^{H+1} \left\{ \|\tilde{\theta}_\mu(k+i)\|^2 + \right. \\ \left. \text{tr}(\Sigma_{\hat{\theta}}(k+i|k)) + 2\|e_{\mu_{\hat{\theta}}}(k+i)\| \|\tilde{\theta}_\mu(k+i)\| \right\}. \quad (59)$$

From Lemma 4,  $\|e_{\mu_{\hat{\theta}}}(k+i)\| \leq a_4(i)\|e_\theta(k)\| + a_5(i)$ . From (28),  $\|\Sigma_{\hat{\theta}}(k+i)\| \leq i(\Delta t)^2\sigma_{f_{\max}}^2$ . For diagonal  $\Sigma_{\hat{\theta}}(k+i|k)$ , we have  $\text{tr}(\Sigma_{\hat{\theta}}(k+i|k)) \leq m\|\Sigma_{\hat{\theta}}(k+i)\| \leq im(\Delta t)^2\sigma_{f_{\max}}^2$ . Adding the above upper bounds for each term in (59), we obtain that  $|\bar{J}_{\hat{\theta}, W^*}^k - \bar{J}_{\theta, W^*}^k| \leq \rho_J(e_\alpha, e_\theta)$ . ■

### G. Proof of Theorem 1

It is straightforward to obtain that  $V(k) \geq \lambda_{\min}(\mathbf{Q}_1)\|e_\theta(k)\|^2 + \zeta\lambda_{\min}(\mathbf{Q})\|e_\alpha(k)\|^2 \geq \underline{\lambda}\|e(k)\|^2$  and similarly  $V(k) \leq \bar{\lambda}\|e(k)\|^2$ . Therefore, we have

$$\underline{\lambda}\|e(k)\|^2 \leq V(k) \leq \bar{\lambda}\|e(k)\|^2. \quad (60)$$

From (43), we obtain

$$\Delta V(k) \leq |\bar{J}_{\hat{\theta}, W^*}^{k+1} - \bar{J}_{\theta, W^*}^{k+1}| + |\bar{J}_{\hat{\theta}, W^*}^k - \bar{J}_{\theta, W^*}^k| + \zeta[V_\alpha(k+1) \\ - V_\alpha(k)] + \nu[\|\Sigma_\alpha^*(k)\| - \|\Sigma_\alpha^*(k+1)\|] \\ + (J_{\hat{\theta}, W^*}^{k+1} - J_{\theta, W^*}^k). \quad (61)$$

We apply the results in Lemma 7 to the first two terms in (61). For the third difference term, from Lemma 2, we have

$$V_\alpha(k+1) - V_\alpha(k) \leq -\frac{1}{4}\Delta t\lambda_{\min}(\mathbf{Q})\|e_\alpha(k)\|^2 + c_3\Delta t.$$

For the last two difference terms in (61), by (56), we have

$$J_{\hat{\theta}, W^*}^{k+1} - J_{\theta, W^*}^k + \nu[\|\Sigma_\alpha^*(k)\| - \|\Sigma_\alpha^*(k+1)\|] \\ \leq -\lambda_{\min}(\mathbf{Q}_1)\|e_\theta(k)\|^2 + \nu[\|\Sigma_\alpha^e(k+1)\| - \|\Sigma_\alpha^*(k+1)\|] \\ + \Delta\alpha_{dk} + \text{tr}(\mathbf{Q}_1\Sigma_{\hat{\theta}}(k+H+1|k)) \\ + \text{tr}(\mathbf{Q}_3\Sigma_{\hat{\theta}}(k+H+2|k)) \\ \leq -\lambda_{\min}(\mathbf{Q}_1)\|e_\theta(k)\|^2 + \alpha_{d_{\max}}^2 + \nu\sigma_{\kappa_{\max}}^2 \\ + m\lambda_m(H+2)(\Delta t)^2\sigma_{f_{\max}}^2.$$

In the above last inequality, we use the facts that  $\|\alpha_{d(k+1)}^*\|_{Q_2}^2 \leq \alpha_{d_{\max}}^2$  and  $\|\Sigma_\alpha^e(k+1)\| \leq \sigma_{\kappa_{\max}}^2$ .

Substituting the above derivations into (61), we obtain (62) shown at the bottom of this page if (46) is held. Considering the above result and (60), we have  $V(k+1) \leq \gamma_\lambda V(k) + \gamma_5$ , and this proves the theorem.

## APPENDIX II

### PHYSICAL MODELS FOR TWO PLATFORMS

#### A. Furuta Pendulum

The dynamics model is given in the form of (1) with  $\mathbf{q} = [\theta_1 \ \alpha_1]^\top$ ,  $\dot{\mathbf{q}} = [\theta_2 \ \alpha_2]^\top$ , and  $u = u_d = V_m$ ,

$$\mathbf{D}(\mathbf{q}) = \begin{bmatrix} D_1 & D_3 \\ D_3 & D_2 \end{bmatrix}, \mathbf{H}(\mathbf{q}, \dot{\mathbf{q}}) = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}, \mathbf{B}(\mathbf{q}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

where  $D_1 = m_p l_r^2 + \frac{1}{4}m_p l_p^2 \sin^2 \alpha_1 + J_r$ ,  $D_2 = J_p + \frac{1}{4}m_p l_p^2$ ,

$$D_3 = -\frac{1}{2}m_p l_p l_r \cos \alpha_1, H_1 = \frac{1}{2}m_p l_p^2 \theta_2 \alpha_2 \sin \alpha_1 \cos \alpha_1 \\ + \frac{1}{2}m_p l_p l_r \alpha_2^2 \sin \alpha_1 + d_r \theta_2 + k_g^2 k_t k_m / R_m \theta_2$$

$$\Delta V(k) \leq \xi_1\|e_\alpha(k)\|^2 + \xi_2\|e_\alpha(k)\| \|e_\theta(k)\| + \xi_3\|e_\alpha(k)\| + \xi_4\|e_\theta(k)\| + \xi_5 - \frac{\zeta}{4}\Delta t\lambda_{\min}(\mathbf{Q})\|e_\alpha(k)\|^2 + \zeta c_3\Delta t - \\ \lambda_{\min}(\mathbf{Q}_1)\|e_\theta(k)\|^2 + \alpha_{d_{\max}}^2 + \nu\sigma_{\kappa_{\max}}^2 + m\lambda_m(H+2)(\Delta t)^2\sigma_{f_{\max}}^2 \\ = -\frac{1}{2}\left(\gamma_3\|e_\theta(k)\| - \frac{\xi_2}{\gamma_3}\|e_\alpha(k)\|\right)^2 - \frac{\gamma_3^2}{4}\|e(k)\|^2 - (\gamma_1\|e_\alpha(k)\| - \gamma_2)^2 - \left(\frac{\gamma_3}{2}\|e_\theta(k)\| - \gamma_4\right)^2 + \gamma_5 \\ \leq -\frac{\gamma_3^2}{4}\|e(k)\|^2 + \gamma_5. \quad (62)$$

$$H_2 = -\frac{1}{4}m_p l_p^2 \cos \alpha_1 \sin \alpha_1 \theta_2^2 + d_p \alpha_2 - \frac{1}{2}m_p l_p g \sin \alpha_1.$$

$l_r$  and  $J_r$  denote the base arm's length and moment of inertia, respectively. Parameters  $l_p$ ,  $m_p$ , and  $J_p$  denote the pendulum link's length, mass, and moment of inertia, respectively. Parameters  $d_r$  and  $d_p$  are the viscous damping coefficients of the base arm and pendulum joints, respectively.  $k_g$ ,  $k_t$ ,  $k_m$ , and  $R_m$  are dc motor's electromechanical parameters [41]. It is straightforward to write the dynamics in the form of (14) with

$$f_\theta(\theta, \alpha, u_d) = \frac{D_2 u_d - H_1 D_2 + D_3 H_2}{D_1 D_2 - D_3^2}$$

$$\text{and } \kappa_\alpha(\theta, \alpha, \dot{\alpha}_2) = H_1 - \frac{1}{D_3}[(D_3 + D_1 D_2 - D_3^2)\dot{\alpha}_2 + H_2 D_1].$$

### B. Autonomous Bikebot

As shown in Fig. 4(b), the kinematic model at  $C_2$  gives

$$\ddot{X} = \dot{v}_c \cos \psi - v_c \sin \psi \dot{\psi}, \ddot{Y} = \dot{v}_c \sin \psi + v_c \cos \psi \dot{\psi}. \quad (63)$$

The yaw rate is calculated from the geometric relationship between the steering and rear frames as [42]

$$\dot{\psi} = \frac{v_c \cos \vartheta}{l \cos \varphi} \tan \phi \quad (64)$$

and the equation of roll motion is obtained as [8]

$$J_t \ddot{\varphi} = -m_b h_b \frac{v_c^2}{l} \cos \vartheta \tan \phi + m_b h_b g \sin \varphi \quad (65)$$

where  $m_b$  is the total mass of the bikebot,  $l$  is the wheelbase,  $\vartheta$  is the steering casting angle,  $h_b$  is the height of mass center, and  $J_t = m_b h_b^2 + J_b$  and  $J_b$  are the mass moments of inertia of the bikebot around the  $x$ -axis direction of the moving frame  $\mathcal{R}$  and the body frame  $\mathcal{B}$ , respectively [see Fig. 4(b)]. Plugging (64) into (63) and combining with (65), we obtain the dynamics in (50) with

$$f_\theta(\theta, \alpha, u_d, u_f) = \begin{bmatrix} u_f \cos \psi - \frac{v_c^2 \sin \psi \cos \vartheta}{l \cos \varphi} \tan u_d \\ u_f \sin \psi + \frac{v_c^2 \cos \psi \cos \vartheta}{l \cos \varphi} \tan u_d \end{bmatrix}$$

$$\kappa_\alpha(\theta, \alpha, \dot{\alpha}_2, u_f) = \tan^{-1} \left( \frac{m_b h_b g l \sin \varphi - J_t l \ddot{\varphi}}{m_b h_b v_c^2 \cos \vartheta} \right) - \dot{\varphi}, \text{ and yaw angle } \psi = \text{atan2}(\dot{Y}, \dot{X}).$$

### ACKNOWLEDGMENT

The authors would like to thank three anonymous reviewers and the Associate Editor for their constructive and helpful comments and suggestions to improve the quality of this article.

### REFERENCES

- [1] A. Choukchou-Braham, B. Cherki, M. Djemaï, and K. Busawon, *Analysis and Control of Underactuated Mechanical Systems*. New York, NY, USA: Springer, 2014.
- [2] N. Getz, "Dynamic inversion of nonlinear maps with applications to nonlinear control and robotics," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, CA, USA, 1995.
- [3] J. Lee, R. Mukherjee, and H. K. Khalil, "Output feedback stabilization of inverted pendulum on a cart in the presence of uncertainties," *Automatica*, vol. 54, pp. 146–157, 2015.
- [4] A. S. Shiriaev, L. B. Freidovich, A. Robertsson, R. Johansson, and A. Sandberg, "Virtual-holonomic-constraints-based design of stable oscillations of Furuta pendulum: Theory and experiments," *IEEE Trans. Robot.*, vol. 23, no. 4, pp. 827–832, Aug. 2007.
- [5] M.-S. Park and D. Chwa, "Orbital stabilization of inverted-pendulum systems via coupled sliding-mode control method," *IEEE Trans. Ind. Electron.*, vol. 56, no. 9, pp. 3556–3570, Sep. 2009.
- [6] L. B. Freidovich, A. S. Shiriaev, F. Gordillo, F. Gómez-Estern, and J. Aracil, "Partial-energy-shaping control for orbital stabilization of high-frequency oscillations of the Furuta pendulum," *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 4, pp. 853–858, Jul. 2009.
- [7] J. Yi, D. Song, A. Levandowski, and S. Jayasuriya, "Trajectory tracking and balance stabilization control of autonomous motorcycles," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2006, pp. 2583–2589.
- [8] P. Wang, J. Yi, T. Liu, and Y. Zhang, "Trajectory tracking and balance control of an autonomous bikebot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 2414–2419.
- [9] E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, and B. Morris, *Feedback Control of Dynamic Bipedal Robot Locomotion*. Boca Raton, FL, USA: CRC Press, 2007.
- [10] K. Chen, M. Trkov, and J. Yi, "Hybrid zero dynamics of human walking with foot slip," in *Proc. Amer. Control Conf.*, 2017, pp. 2124–2129.
- [11] J. Grizzle, M. Di Benedetto, and F. Lamnabhi-Lagarigue, "Necessary conditions for asymptotic tracking in nonlinear systems," *IEEE Trans. Autom. Control*, vol. 39, no. 9, pp. 1782–1794, Sep. 1994.
- [12] S. S. Sastry, *Nonlinear Systems: Analysis, Stability, and Control*. New York, NY, USA: Springer, 1999.
- [13] D. Nguyen-Tuong and J. Peters, "Model learning for robot control: A survey," *Cogn. Process.*, vol. 12, pp. 319–340, 2011.
- [14] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [15] T. Beckers, J. Umlauf, D. Kulic, and S. Hirche, "Stable Gaussian process based tracking control of Lagrangian systems," in *Proc. IEEE Conf. Decis. Control*, 2017, pp. 5180–5185.
- [16] M. K. Helwa, A. Heins, and A. P. Schoellig, "Provably robust learning-based approach for high-accuracy tracking control of Lagrangian systems," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1587–1594, Apr. 2019.
- [17] T. Koller, F. Berkenkamp, M. Turchetta, and A. Krause, "Learning-based model predictive control for safe exploration," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 6059–6066.
- [18] S. Zhou, M. K. Helwa, and A. P. Schoellig, "An inversion-based learning approach for improving impromptu trajectory tracking of robots with non-minimum phase dynamics," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1663–1670, Sep. 2018.
- [19] K. Chen, J. Yi, and T. Liu, "Learning-based modeling and control of underactuated balance robotic systems," in *Proc. IEEE Conf. Autom. Sci. Eng.*, 2017, pp. 1118–1123.
- [20] F. Han and J. Yi, "Stable learning-based tracking control of underactuated balance robots," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 1543–1550, Apr. 2021.
- [21] G. Cao, E. M.-K. Lai, and F. Alam, "Gaussian process model predictive control of an unmanned quadrotor," *J. Intell. Robot. Syst.*, vol. 88, no. 1, pp. 147–162, 2017.
- [22] M. P. Deisenroth, D. Fox, and C. E. Rasmussen, "Gaussian processes for data-efficient learning in robotics and control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 2, pp. 408–423, Feb. 2015.
- [23] R. Murray-Smith, D. Sbarbaro, C. E. Rasmussen, and A. Girard, "Adaptive, cautious, predictive control with Gaussian process priors," *IFAC Proc.*, vol. 36, no. 16, pp. 1155–1160, 2003.
- [24] C. D. McKinnon and A. P. Schoellig, "Learning probabilistic models for safe predictive control in unknown environments," in *Proc. Eur. Control Conf.*, 2019, pp. 2472–2479.
- [25] C. J. Ostafew, A. P. Schoellig, and T. D. Barfoot, "Robust constrained learning-based NMPC enabling reliable mobile robot path tracking," *Int. J. Robot. Res.*, vol. 35, no. 13, pp. 1547–1563, 2016.
- [26] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1238–1274, 2013.
- [27] I. Lenz, R. Knepper, and A. Saxena, "DeepMPC: Learning deep latent features for model predictive control," in *Proc. Robot.: Sci. Syst. Conf.*, 2015.
- [28] A. Tamar, G. Thomas, T. Zhang, S. Levine, and P. Abbeel, "Learning from the hindsight plan—Episodic MPC improvement," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 336–343.
- [29] G. Williams et al., "Information theoretic MPC for model-based reinforcement learning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 1714–1721.

- [30] S. Schaal and C. G. Atkeson, "Learning control in robotics," *IEEE Robot. Autom. Mag.*, vol. 17, no. 2, pp. 20–29, Jun. 2010.
- [31] G. Chowdhary, H. A. Kingravi, J. P. How, and P. A. Vela, "Bayesian nonparametric adaptive control using Gaussian processes," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 3, pp. 537–550, Mar. 2015.
- [32] J. Umlauft and L. P. S. Hirche, "An uncertainly-based control Lyapunov approach for control-affine systems modeled by Gaussian process," *IEEE Control Syst. Lett.*, vol. 2, no. 3, pp. 483–488, Jul. 2018.
- [33] K. Chen, J. Yi, and D. Song, "Gaussian processes model-based control of underactuated balance robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2019, pp. 4458–4464.
- [34] M. Mihalec and J. Yi, "Control of a bipedal walker under foot slipping condition using whole-body operational space framework," in *Proc. Model. Estimation Control Conf.*, 2021, pp. 278–283.
- [35] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Modeling and Control*. New York, NY, USA: Wiley, 2006.
- [36] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3250–3265, May 2012.
- [37] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
- [38] J. B. Rawlings, D. Q. Mayne, and M. M. Diehl, *Model Predictive Control: Theory, Computation, and Design*, 2nd ed. Santa Barbara, CA, USA: Nob Hill Publishing, 2020.
- [39] H. Chen and F. Allgower, "A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability," *Automatica*, vol. 34, pp. 1205–1217, 1998.
- [40] K. Chen, Y. Zhang, J. Yi, and T. Liu, "An integrated physical-learning model of physical human-robot interactions with application to pose estimation in bikebot riding," *Int. J. Robot. Res.*, vol. 35, no. 12, pp. 1459–1476, 2016.
- [41] J. Apkarian, P. Karam, and M. Levis, *Instructor Workbook: Inverted Pendulum Experiment for Matlab/Simulink Users*. Markham, ON, Canada: Quanser Inc., 2011.
- [42] P. Wang, J. Yi, and T. Liu, "Stability and control of a rider-bicycle system: Analysis and experiments," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 1, pp. 348–360, Jan. 2020.



**Kuo Chen** received the B.S. degree in information science from Nankai University, Tianjin, China, in 2011, and the Ph.D. degree in mechanical and aerospace engineering from Rutgers University, New Brunswick, NJ, USA, in 2019.

He is currently a Senior Software Engineer with Motional Inc., Boston, MA, USA. His research interests include underactuated system control, human–robot interactions, and autonomous driving.

Dr. Chen received the 2014 ASME Dynamic Systems and Control Division's Best Student Paper

Award on Mechatronics and the Best Student Paper Award at 2015 IEEE/ASME International Conference on Advanced Intelligent Mechatronics.

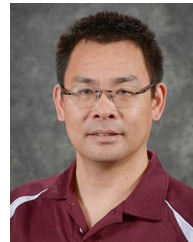


**Jingang Yi** (Senior Member, IEEE) received the B.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 1993, the M.Eng. degree in precision instruments from Tsinghua University, Beijing, China, in 1996, and the M.A. degree in mathematics and the Ph.D. degree in mechanical engineering from the University of California, Berkeley, CA, USA, in 2001 and 2002, respectively.

He is currently a Professor of Mechanical Engineering with Rutgers University, New Brunswick, NJ, USA. His research interests include autonomous

robotic systems, dynamic systems and control, mechatronics, and automation science and engineering, with applications to biomedical systems, civil infrastructure, and transportation systems.

Dr. Yi is a Fellow of the American Society of Mechanical Engineers (ASME). He was a recipient of the 2010 U.S. National Science Foundation Faculty Early Career Development Award. He is a Senior Editor for IEEE ROBOTICS AND AUTOMATION LETTERS and IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING. He is an Associate Editor for *International Journal of Intelligent Robotics and Applications*. He was an Associate Editor for IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, IEEE/ASME TRANSACTIONS ON MECHATRONICS, IEEE ROBOTICS AND AUTOMATION LETTERS, *IFAC Journal Mechatronics*, *Control Engineering Practice*, and *ASME Journal of Dynamic Systems, Measurement and Control*.



**Dezhen Song** (Senior Member, IEEE) received the Ph.D. degree in industrial engineering from the University of California, Berkeley, CA, USA, in 2004.

He is currently a Professor with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA. His research interests include networked robotics, distributed sensing, computer vision, surveillance, and stochastic modeling.

Dr. Song received the Kayamori Best Paper Award at 2005 IEEE International Conference on Robotics and Automation (with J. Yi and S. Ding) and the National Science Foundation Faculty Early Career Development Award in 2007. From 2008 to 2012, he was an Associate Editor for IEEE TRANSACTIONS ON ROBOTICS. From 2010 to 2014, he was an Associate Editor for IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING. From 2017 to 2021, he was a Senior Editor for IEEE ROBOTICS AND AUTOMATION LETTERS.