

Active Sensing for Data Quality Improvement in Model Learning

Olga Napolitano^{ID}, *Student Member, IEEE*, Marco Cognetti^{ID}, *Member, IEEE*,
Lucia Pallottino^{ID}, *Senior Member, IEEE*, Dimitrios Kanoulas^{ID}, Paolo Salaris^{ID}, *Member, IEEE*,
and Valerio Modugno^{ID}

Abstract—In machine learning for robotics, training data quality assumes a crucial role. Many methods use exploration algorithms to select the most informative data points for the model, often ignoring the impact of measurement noise on data. This letter introduces a method to enhance dataset quality for model learning, optimizing a combination of exploration and active sensing metrics. We introduce a novel Exploration Gramian metric based on a Gaussian Process predicted covariance matrix, optimized to explore the state space regions where the knowledge about the unknown model is maximum. These are integrated with an active sensing metric (Constructibility Gramian) to mitigate measurement noise effects. The effectiveness of this approach is demonstrated through simulations on a unicycle and a quadruped robot, confirming that combining active sensing and exploration significantly enhances performance in model learning.

Index Terms—Optimization, robotics, information theory and control.

I. INTRODUCTION

IN THE field of machine learning (ML), many researchers focus on developing powerful models for processing complex datasets to enhance predictions. However, the importance of the quality and informativeness of these datasets, particularly in real-world applications, often receives less attention,

Manuscript received 7 March 2024; revised 7 May 2024; accepted 22 May 2024. Date of publication 5 June 2024; date of current version 27 June 2024. This work was supported in part by the UKRI Future Leaders Fellowship (RoboHike) under Grant MR/V025333/1; in part by the CDT for Foundational Artificial Intelligence; in part by the European Union's Horizon 2020 Research and Innovation Programme (DARKO) under Agreement 101017274; and in part by the European Union by the Next Generation EU—Ecosistema dell'Innovazione Tuscany Health Ecosystem (THE, PNRR, Spoke 9: Robotics and Automation for Health) and the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab and ForeLab Project (Departments of Excellence) under Project ECS00000017. The work of Marco Cognetti's was supported by the Chaire de Professeur Junior under Grant ANR-22-CPJ1-0064-01. Recommended by Senior Editor S. Olaru. (Corresponding author: Olga Napolitano.)

Olga Napolitano, Lucia Pallottino, and Paolo Salaris are with the Research Center "E. Piaggio" and the Diploma of Information Engineering, University of Pisa, 56122 Pisa, Italy (e-mail: olga.napolitano@phd.unipi.it).

Marco Cognetti is with LAAS-CNRS, Université de Toulouse, CNRS, UPS, 31000 Toulouse, France.

Dimitrios Kanoulas and Valerio Modugno are with the RPL Lab, University College London, WC1E 6BT London, U.K.

Digital Object Identifier 10.1109/LCSYS.2024.3410152

despite its significant impact on ML performance [1], [2]. The quality of the datasets in terms of informativeness plays a crucial role in ML applications in robotics and automation, especially when used for unknown model learning. Because of the time constraint of a real-time implementation, only a few data points can be processed to provide a timely prediction. Therefore, to best predict the model, during the learning process, it is of paramount importance to identify the few data points that contain the largest amount of information [3]. To this purpose, strategies for constructing training sets that minimize the *epistemic error* (systematic errors due to limitations on model knowledge) are becoming popular. These include the use of information-theoretic optimal experimental design for selecting optimal training dataset [4], and exploration algorithms that maximize information gain, quantified by the Fisher Information Matrix (FIM) [5], [6].

In learning unknown dynamics, a significant challenge is to extend the concept of information gain to a trajectory optimization problem. To tackle this issue, in [7] a static active learning strategy is adopted to dynamic settings, aligning the sampling strategy with trajectory-dependent states.

Another factor affecting data quality is *measurement error*, primarily stemming from noisy sensor readings. While filters can passively increase the robustness of model learning by limiting the effects of noise, actively optimizing measurement acquisition through active sensing/perception approaches further mitigates these effects. Although widely adopted in robotics to enhance estimation [8], [9], active sensing has been minimally explored for improving filter estimates in training set construction. In particular, in this letter, we use the Constructibility Gramian (CG) as active sensing metric [10]

Our proposed method integrates active sensing with exploration to identify informative trajectories for sample collection. We introduce the Exploration Gramian (EG), a new metric that uses the state transition matrix to capture how the system evolves along a trajectory, thus improving sample collection in trajectory optimization problems. The EG is obtained from the predicted covariance matrix of a Gaussian Process, which we utilize as a learning method in our study. To the best of our knowledge, this represents the first implementation of EG in an exploration algorithm. Here, exploration and active sensing metrics are combined, providing new cost functions maximized in a Model Predictive Control (MPC) framework.

The effectiveness of our methodology is demonstrated through comparative analysis with a standard exploration algorithm based on the GP covariance matrix. Differently from the EG, this metric does not contain the state dynamics and hence could not properly steer the system to unexplored states. Finally, we test our EG-based methodology on simulated unicycle and quadrupedal robots, proving significant improvements in model learning.

II. PRELIMINARIES

The primary objective of this letter is to maximize data-derived information to reduce both epistemic and measurement errors. Consequently, we employ two metrics: the *active sensing* measure (CG) enhances observer estimate performance and minimizes *measurement error*, while the *active exploration* measure (EG) boosts the efficiency of a Gaussian Process used in learning unknown dynamics, thus diminishing *epistemic error*.

Let us consider a generic nonlinear system

$$\dot{\mathbf{q}}(t) = \mathbf{f}_n(\mathbf{q}(t), \mathbf{u}(t)) + \mathbf{f}_u(\mathbf{q}(t), \mathbf{u}(t)), \quad \mathbf{q}(t_0) = \mathbf{q}_0 \quad (1)$$

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{q}(t)) + \mathbf{v}(t) \quad (2)$$

where $\mathbf{q}(t) \in \mathbb{R}^n$ is the state of the system, $\mathbf{u}(t) \in \mathbb{R}^m$ is the control inputs, $\mathbf{z}(t) \in \mathbb{R}^p$ is the sensor outputs (i.e., the measurements available through onboard sensors at time t), $\mathbf{f}_n(\cdot)$ is the known nominal system dynamics while $\mathbf{f}_u(\cdot)$ is the unknown one. Finally, $\mathbf{v}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}(t))$ is a white, normally-distributed Gaussian noise with zero mean and covariance matrix $\mathbf{R}(t)$. We assume that the nonlinear system is affected by negligible process noise.

Let us consider the linear time-varying (LTV) system obtained by linearizing (1)-(2) around a given trajectory, with $\mathbf{q}(t_0) = \mathbf{q}_0$:

$$\dot{\mathbf{q}}(t) = \mathbf{A}(t)\mathbf{q}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{F}_u(\mathbf{q}(t), \mathbf{u}(t)), \quad (3)$$

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{q}(t) + \mathbf{v}(t), \quad (4)$$

where $\mathbf{A}(t) = \frac{\partial \mathbf{f}_n(\mathbf{q}(t), \mathbf{u}(t))}{\partial \mathbf{q}(t)}$, $\mathbf{B}(t) = \frac{\partial \mathbf{f}_n(\mathbf{q}(t), \mathbf{u}(t))}{\partial \mathbf{u}(t)}$, and $\mathbf{C}(t) = \frac{\partial \mathbf{h}(\mathbf{q}(t))}{\partial \mathbf{q}(t)}$. Moreover, for the LTV system, the unknown dynamics is $\mathbf{F}_u(\mathbf{q}(t), \mathbf{u}(t)) = \mathbf{f}_u(\mathbf{q}(t), \mathbf{u}(t)) + O(\mathbf{q}^2)$ and consists of the linearized unknown dynamics $\mathbf{f}_u(\mathbf{q}(t), \mathbf{u}(t))$ and nonlinear higher-order terms $O(\mathbf{q}^2)$ resulting from the Taylor series, which are neglected during the linearization process. The chosen formulation is general enough to cover a broad class of practical cases, ranging from simplified linear models of complex systems with highly nonlinear dynamics to nonlinear dynamics with unmodeled dynamics, disturbances, and neglected higher-order terms.

A. Active Sensing Measures

In real scenarios, the real state $\mathbf{q}(t)$ of a system is typically unknown, with only an available estimate $\hat{\mathbf{q}}(t)$ provided by an observer. The observer's performance, in terms of uncertainty and estimation error, heavily relies on the information coming from noisy sensor readings. Addressing this, active sensing/perception control aims to maximize sensory information for optimal state estimation. A crucial key step is the selection of an appropriate measure of information to optimize [11], [12], [13]. The measure chosen in this letter is the CG (see Section I) which is defined as

$$\mathcal{G}_c(t_0, t_f) \triangleq \int_{t_0}^{t_f} \Phi^\top(\tau, t) \mathbf{C}(\tau)^\top (\tau) \mathbf{W}_c(\tau) \mathbf{C}(\tau) \Phi(\tau, t) d\tau,$$

where $t_f > t_0$ is the final integration time, and $\mathbf{W}_c(\tau) \in \mathbb{R}^{p \times p}$ is a symmetric positive definite weight matrix. $\Phi(t, t_0)$ is the *state transition matrix* of the LTV system in (3)-(4) (with $\mathbf{F}_u(\mathbf{q}(t), \mathbf{u}(t)) = 0$) and solution of the following matrix differential equation

$$\dot{\Phi}(t, t_0) = \mathbf{A}(t)\Phi(t, t_0), \quad \Phi(t_0, t_0) = \mathbf{I}. \quad (5)$$

In [10], the authors showed that the CG is also strictly related to the inverse of the estimation error covariance matrix \mathbf{P} provided by an Extended Kalman Filter (EKF) built on (3)-(4) (with $\mathbf{F}_u(\mathbf{q}(t), \mathbf{u}(t)) = 0$ and negligible process noise). Indeed, the following equivalence holds: $\mathbf{P}^{-1}(t) = \mathcal{G}_c(-\infty, t)$, where $\mathcal{G}_c(-\infty, t)$ represents the amount of the collected information in the time window $[-\infty, t]$. Thus, maximizing the CG is equivalent to minimizing the state estimation uncertainty.

B. Exploration Measures

A Gaussian Processes regressor is exploited to reconstruct the unknown dynamics \mathbf{F}_u . GPs can be seen as an infinite collection of random Gaussian variables, any finite number of which has a joint Gaussian distribution [14]. Therefore, \mathbf{F}_u is learned from N samples $\mathbf{x}_i := (\hat{\mathbf{q}}_i, \mathbf{u}_i)$, $i = 1, \dots, N$, which consists of estimated state and inputs and represents the input data of the GP. The corresponding outputs

$$\mathbf{y}_i = \hat{\mathbf{q}}_i - \mathbf{f}_n(\hat{\mathbf{q}}_i, \mathbf{u}_i) + \mathbf{w} \quad (6)$$

represent the deviation of the estimated nominal system dynamics from $\hat{\mathbf{q}}_i$, which is an estimate of $\hat{\mathbf{q}}_i$. Moreover, $\mathbf{w} \sim \mathcal{N}(0, \Lambda)$ is a white, normally-distributed Gaussian noise with zero mean and covariance matrix Λ that affects each element of $\mathbf{y} \in \mathbb{R}^n$. The computation of $\hat{\mathbf{q}}(t)$ in (6), requires an estimator. An observer built on (1)-(2) allows to only retrieve $\hat{\mathbf{q}}(t)$ but not its time derivative. Hence we introduce an augmented state $\mathbf{q}_{\text{aug}}(t) = (\mathbf{q}(t), \dot{\mathbf{q}}(t), \mathbf{u}(t))^\top$ with $\mathbf{u}_{\text{aug}}(t) = \dot{\mathbf{u}}(t)$ the new input. Starting from (1) with $\mathbf{f}_u(\mathbf{q}(t), \mathbf{u}(t)) = 0$, the dynamics of $\mathbf{q}_{\text{aug}}(t)$ is

$$\dot{\mathbf{q}}_{\text{aug}}(t) = \begin{pmatrix} \dot{\mathbf{q}}(t) \\ \mathbf{A}(t)\mathbf{f}_n(\mathbf{q}(t), \mathbf{u}(t)) \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{n \times m} \\ \mathbf{B}(t) \\ \mathbf{I}_{m \times m} \end{pmatrix} \mathbf{u}_{\text{aug}}(t). \quad (7)$$

The training dataset of the GP is hence given by $\mathcal{D} := \{\mathbf{X} := [\mathbf{x}_1, \dots, \mathbf{x}_N]^\top, \mathbf{Y} := [\mathbf{y}_1, \dots, \mathbf{y}_N]^\top\}$, with $\mathbf{X} \in \mathbb{R}^{N \times (n+m)}$, $\mathbf{Y} \in \mathbb{R}^{N \times n}$, and N the number of training set samples.

The accuracy of $\dot{\hat{\mathbf{q}}}$ and $\hat{\mathbf{q}}$ directly influences the training set quality and thus the effectiveness of the GP in learning the unknown model. Consequently, the introduction of an active sensing control strategy is crucial for enhancing training set quality, underlining the importance of the active perception-exploration approaches tackled in this letter.

Since \mathbf{F}_u is approximated by a GP, providing $\hat{\mathbf{F}}_u$, in each output dimension $a = 1, \dots, n$ it is fully characterized by its mean $\mu^a(\cdot)$ and variance $Q^a(\cdot)$, which at an unobserved point \mathbf{x}_* are

$$\begin{aligned} \mu^a(\mathbf{x}_*) &= \mathbf{K}_{\mathbf{x}_* \mathbf{x}_*}^a (\mathbf{K}_{\mathbf{X} \mathbf{X}}^a + \Lambda)^{-1} [\mathbf{y}]_{..a}, \\ Q^a(\mathbf{x}_*) &= \mathbf{K}_{\mathbf{x}_* \mathbf{x}_*}^a - \mathbf{K}_{\mathbf{x}_* \mathbf{X}}^a (\mathbf{K}_{\mathbf{X} \mathbf{X}}^a + \Lambda)^{-1} \mathbf{K}_{\mathbf{X} \mathbf{x}_*}^a. \end{aligned}$$

where K_{Xx}^a , K_{XX}^a , $K_{x_*x_*}^a$ and $K_{x_*X}^a$ depends on the kernel $k^a(\cdot, \cdot)$ (see [14]). The prior mean is assumed zero. Each dimension of the output is learned independently, resulting in a multivariate GP approximation of the unknown dynamics F_u , given by stacking the individual output dimensions $\hat{F}_u(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \mathbf{Q}(\mathbf{x}))$, where $\boldsymbol{\mu} = [\mu^1, \dots, \mu^n]^T$ and $\mathbf{Q} = \text{diag}(\{Q^1, \dots, Q^n\})$ are the predicted mean vector and covariance matrix provided by the GP about the unknown dynamics at a generic input \mathbf{x} and conditioned by the training set \mathcal{D} .

To enhance the reconstruction of unknown dynamics $\hat{F}_u(\mathbf{x})$, the GP training set will be extended with novel samples \mathbf{x} that are expected to increase the knowledge about $F_u(\mathbf{x})$. In the context of online learning, where the training set is continuously updated, the informativeness of new sampling points is assessed using mutual information between these points and their respective observations [15]. Therefore, let us consider new samples $\mathbf{X}_{\text{new}} = [\mathbf{x}_1, \dots, \mathbf{x}_\kappa]$ and the a -th component of the outputs $\mathbf{y}_{\text{new}}^a = [y_1^a, \dots, y_\kappa^a]$, with $a = 1, \dots, n$. Since the observations are affected by an additive Gaussian white noise $\mathcal{N}(0, \Lambda)$ with $\Lambda = \lambda^2 \mathbf{I}$, the *Information Gain* (IG) for each component of the output can be expressed as (see [16, Lemma 5.3] for details)

$$\begin{aligned} I^a(\mathbf{y}_{\text{new}}^a, \hat{F}_{u_{\text{new}}}^a) &= \mathbf{H}(\mathbf{y}_{\text{new}}^a) - \mathbf{H}(\mathbf{y}_{\text{new}}^a | \hat{F}_{u_{\text{new}}}^a) \\ &= \frac{1}{2} \sum_{\alpha=1}^{\kappa} \log(2\pi e (\lambda^2 + \mathbf{Q}^a(\mathbf{x}_\alpha))) \\ &\quad - \frac{1}{2} \sum_{\alpha=1}^{\kappa} \log(2\pi e \lambda^2) \\ &= \frac{1}{2} \sum_{\alpha=1}^{\kappa} \log(1 + \lambda^{-2} \mathbf{Q}^a(\mathbf{x}_\alpha)), \end{aligned}$$

where $\hat{F}_{u_{\text{new}}}^a = [\hat{F}_u^a(\mathbf{x})]_{\mathbf{x} \in \mathbf{X}_{\text{new}}}$, and $\alpha = 1, \dots, \kappa$ is an index used to iterate over the new samples of \mathbf{X}_{new} . One can conclude that for the future points to sample, IG depends on the predictive variances $\mathbf{Q}(\mathbf{x})$ of the GP [16]. Moreover, in parameter identification problems, \mathbf{Q} can be an alternative to the FIM, as the two are inversely proportional [17].

Remark 1: The eigenvalues of \mathbf{Q} measure the information gain associated with the new samples taken along the directions denoted by their eigenvectors. Samples that maximize \mathbf{Q} , also maximize IG, steering the system toward unexplored state-input space.

Unfortunately, $\mathbf{Q}(\mathbf{x})$ does not consider the information about the state evolution along the planned trajectory that, instead, cannot be neglected since we are learning the unknown part of a dynamic system. Hence, let us start from the explicit solution of (3), i.e.,

$$\begin{aligned} \mathbf{q}(t) &= \boldsymbol{\Phi}(t, t_0) \mathbf{q}_0 \\ &\quad + \int_{t_0}^t \boldsymbol{\Phi}(t, \tau) (\mathbf{B}(\tau) \mathbf{u}(\tau) + \mathbf{F}_u(\mathbf{q}(\tau), \mathbf{u}(\tau))) d\tau. \end{aligned} \quad (8)$$

Following [18], the first and the second moments of (8) are

$$\begin{aligned} \boldsymbol{\mu}_q(t) &= \boldsymbol{\Phi}(t, t_0) \mathbf{q}_0 + \int_{t_0}^t \boldsymbol{\Phi}(t, \tau) (\mathbf{B}(\tau) \mathbf{u}(\tau) + \boldsymbol{\mu}(\mathbf{x}(\tau))) d\tau, \\ \mathbf{Q}_q(t) &= \int_{t_0}^t \boldsymbol{\Phi}(t, \tau) \mathbf{Q}(\mathbf{x}(\tau)) \boldsymbol{\Phi}^\top(t, \tau) d\tau. \end{aligned}$$

Unlike $\mathbf{Q}(t)$, $\mathbf{Q}_q(t)$ describes the evolution of the uncertainty on $\mathbf{F}_u(\mathbf{q}(t), \mathbf{u}(t))$ in the time window $[t_0, t]$ taking into account the information about the $\mathbf{f}_n(\mathbf{q}(t), \mathbf{u}(t))$ around the nominal trajectory. The maximization (of some norm) of $\mathbf{Q}_q(t)$ is expected to steer the system toward unexplored state space regions, where the expected IG about $\mathbf{F}_u(\mathbf{q}(t), \mathbf{u}(t))$ is maximized. For this reason, hereafter, we will rename $\mathbf{Q}_q(t)$ as $\mathcal{G}_{\text{exp}}(t_0, t)$ and we will refer to it as *Exploration Gramian*.

We conclude by showing an important link between the EG and the solution $\mathbf{X}(t)$ of the Continuous Riccati Differential Equation (CRDE) associated with (3) in the absence of output equations (i.e., sensor readings). In this case, the CRDE becomes a Differential Lyapunov Equation (DLE), i.e.,

$$\dot{\mathbf{X}}(t) = \mathbf{A}(t) \mathbf{X}(t) + \mathbf{X}(t) \mathbf{A}^\top(t) + \mathbf{Q}(t), \quad (9)$$

with initial condition $\mathbf{X}(t_0) = \mathbf{X}_0$. The solution of (9) is (see [19])

$$\begin{aligned} \mathbf{X}(t) &= \boldsymbol{\Phi}(t, t_0) \mathbf{X}_0 \boldsymbol{\Phi}^\top(t, t_0) + \int_{t_0}^t \boldsymbol{\Phi}(t, \tau) \mathbf{Q}(\tau) \boldsymbol{\Phi}^\top(t, \tau) d\tau \\ &= \boldsymbol{\Phi}(t, t_0) \mathbf{X}_0 \boldsymbol{\Phi}^\top(t, t_0) + \mathcal{G}_{\text{exp}}(t_0, t). \end{aligned} \quad (10)$$

Analogously to what has been done in [10] for the CG, also in this case equation (10) can be expressed as a function of the sole EG. Let $\mathcal{G}_{\text{exp}}(-\infty, t)$ represent the EG computed over the interval $(-\infty, t]$,

$$\begin{aligned} \mathcal{G}_{\text{exp}}(-\infty, t) &= \int_{-\infty}^{t_0} \boldsymbol{\Phi}(t, \tau) \mathbf{Q}(\tau) \boldsymbol{\Phi}^\top(t, \tau) d\tau + \\ &\quad + \int_{t_0}^t \boldsymbol{\Phi}(t, \tau) \mathbf{Q}(\tau) \boldsymbol{\Phi}^\top(t, \tau) d\tau. \end{aligned}$$

Since $\boldsymbol{\Phi}(t, \tau) = \boldsymbol{\Phi}(t, t_0) \boldsymbol{\Phi}(t_0, \tau)$, it follows that

$$\mathcal{G}_{\text{exp}}(-\infty, t) = \boldsymbol{\Phi}(t, t_0) \mathcal{G}_{\text{exp}}(-\infty, t_0) \boldsymbol{\Phi}^\top(t, t_0) + \mathcal{G}_{\text{exp}}(t_0, t). \quad (11)$$

By comparing (10) with (11), it follows that $\mathbf{X}(t) = \mathcal{G}_{\text{exp}}(-\infty, t)$. Differently from $\mathbf{Q}(t)$, the EG takes into account the integral informativeness of both the state-input sample collected along the path and the state evolution of the system encoded in the state transition matrix $\boldsymbol{\Phi}(t_0, t)$. However, the state transition matrix is often not available in explicit form, especially for nonlinear systems or LTV systems, and needs to be computed by solving (5). Therefore, a simplified version of the EG, which is easier to compute, is

$$\mathcal{G}_{\text{exp}}^{\text{DF}}(t_0, t) = \int_{t_0}^t \mathbf{Q}(\tau) d\tau.$$

In the following, we will refer to $\mathcal{G}_{\text{exp}}^{\text{DF}}$ as *Dynamic-Free Exploration Gramian* (DF-EG). Of course, the DF-EG shares somehow the same disadvantage as the Empirical Observability Gramian (EOG) [20] related to neglecting the system state transition matrix. The EOG cannot approximate the local observability Gramian for the states that do not appear in the sensor model and hence its optimization cannot

steer the system along those directions. For the same reason, the DF-EG cannot model the IG for the state that does not appear in $F_u(\mathbf{q}(t), \mathbf{u}(t))$, and hence its optimization cannot steer the system along those directions. It is important to note that, DF-EG is not a new metric for active exploration, as maximizing Q has previously been employed in other works. For instance, in [21], where the GP covariance is maximized within an active tactile object exploration algorithm.

III. ACTIVE PERCEPTION-EXPLORATION COUPLING

This section first introduces the objective function chosen as a baseline in our simulations. Subsequently, effective new cost functions that integrate active sensing and exploration measures are defined and an optimal control problem that maximizes these functions is stated.

A. Perception-Exploration Performance Metrics

Let $t \in [-\infty, t_f]$, the first metric that we introduce in this letter is the trace of DF-EG, i.e.,

$$J_1(\mathbf{q}(t), \mathbf{u}(t)) = \text{tr}(\mathcal{G}_{\text{exp}}^{\text{DF}}(-\infty, t_f)),$$

chosen as the state-of-the-art baseline to prove the effectiveness of our methodology. Furthermore, the DF-EG represents the continuous-time version of the cost function proposed in [7], wherein the IG is replaced by its equivalent definition, Q , as described in [16]. The maximization of both IG and Q yields the same optimal solution, as the logarithmic operation does not alter the problem optimality. By maximizing J_1 , the system will visit unexplored state-input spaces collecting training set samples that maximize the information needed for good model learning.

The second metric is the weighted combination of $J_1(\mathbf{q}(t), \mathbf{u}(t))$ and the trace of the CG, i.e.,

$$J_2(\mathbf{q}(t), \mathbf{u}(t)) = \sigma_{\text{EG}} \text{tr}(\mathcal{G}_{\text{exp}}^{\text{DF}}(-\infty, t_f)) + \sigma_{\text{CG}} \text{tr}(\mathcal{G}_c(-\infty, t_f)),$$

where $\sigma_{\text{EG}}, \sigma_{\text{CG}} > 0$ are used to assign different levels of importance to each component. The maximization of J_2 will result in finding the trajectory that simultaneously leverages the sensor information for improving state estimation and explores the regions where the expected knowledge about the unknown model is maximized.

However, the DF-EG does not consider the system state evolution along it. As a consequence, a better combination is the following

$$J_3(\mathbf{q}(t), \mathbf{u}(t)) = \sigma_{\text{EG}} \text{tr}(\mathcal{G}_{\text{exp}}(-\infty, t_f)) + \sigma_{\text{CG}} \text{tr}(\mathcal{G}_c(-\infty, t_f)).$$

Unlike J_2 , in J_3 both $\mathcal{G}_{\text{exp}}(-\infty, t_f)$ and $\mathcal{G}_c(-\infty, t_f)$ quantify the sensory information and the evolution of the unknown model uncertainty along the planned trajectory taking into account the state evolution of the system encoded in the transition matrix.

Remark 2: The trace operator (aka A -optimality criterion) measures both the average IG about the unknown model and the amount of measurement information. It also satisfies the Bellman principle of optimality, i.e., $\text{tr}(M+N) = \text{tr}(M) + \text{tr}(N)$ implying that subpaths of an optimal path are optimal as well. Other optimality criteria, not satisfying the Bellman principle, could be used as, e.g., the D -optimality (matrix determinant) or E -optimality (matrix eigenvalue) [22].

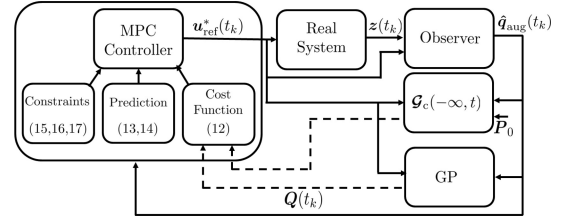


Fig. 1. Control system block diagram of the active perception-exploration coupling methodology proposed in this letter.

Remark 3: It is crucial to find the correct trade-off between exploration and active sensing by adjusting the weights in the proposed objective functions when precise estimation and thorough exploration are required. Indeed, the combination of active sensing and active exploration may exhibit contrasting behaviors. For instance, let us consider a vehicle equipped with a sensor providing intermittent measurements w.r.t. fixed markers. The active sensing control strategy steers the robot around the markers, maximizing the amount of sensory information. In contrast, active exploration guides the vehicle to unexplored state space regions for retrieving the missing dynamics where, however, the availability of the measurement is not guaranteed, possibly reducing (because of Q , which is related to the unknown dynamics) the amount of information and hence the quality of estimation. As a consequence, an effective control strategy needs to balance the benefits of sensory information maximization with the advantages of state space exploration for improving the system's state estimation and unknown dynamics reconstruction.

B. Online Optimal Perception-Exploration Problem

Let us consider a generic observer built on the augmented system $\mathbf{q}_{\text{aug}}(t) = (\mathbf{q}(t), \dot{\mathbf{q}}(t), \mathbf{u}(t))^T$ for estimating $\hat{\mathbf{q}}_{\text{aug}}(t)$, needed for constructing the training set. The goal is to develop an online exploration-perception control strategy (see Fig. 1) by solving, at each time t , the following

Problem 1: [Online Optimal Perception-Exploration Control] Given the prediction horizon consisting of L samples, the control input $\mathbf{u}(t) \in \mathcal{S}(\Delta)$, where $\mathcal{S}(\Delta)$ is the family of piece-wise constant functions with sampling period ΔT , the predicted trajectory of the nominal system $\tilde{\mathbf{q}}_{\text{aug}}$ obtained by applying $\mathbf{u}(t)$ starting from the initial estimated state $\hat{\mathbf{q}}_{\text{aug}}(t_k)$ at time t_k , find $\forall t_- \in [t_k, t_{k+L}]$, the optimal control sequence

$$\mathbf{u}_{\text{ref}}^*(t_-) = \arg \max_{\mathbf{u} \in \mathcal{S}(\Delta)} J_l(\tilde{\mathbf{q}}_{\text{aug}}(t_-), \mathbf{u}(t_-)) \quad l = 1, 2, 3 \quad (12)$$

s.t.

$$\dot{\tilde{\mathbf{q}}}_{\text{aug}}(t_-) = \mathbf{f}_n(\tilde{\mathbf{q}}_{\text{aug}}(t_-), \mathbf{u}(t_-)) \quad (13)$$

$$\tilde{\mathbf{q}}_{\text{aug}}(t_k) = \hat{\mathbf{q}}_{\text{aug}}(t_k) \quad (14)$$

$$\underline{\mathbf{q}}_{\text{aug}} \leq \tilde{\mathbf{q}}_{\text{aug}}(t_-) \leq \bar{\mathbf{q}}_{\text{aug}} \quad (15)$$

$$\underline{\mathbf{u}} \leq \mathbf{u}(t_-) \leq \bar{\mathbf{u}} \quad (16)$$

$$\mathbf{c}(\mathbf{u}(t_-), \tilde{\mathbf{q}}_{\text{aug}}(t_-)) \leq 0 \quad (17)$$

where $J_l(\tilde{\mathbf{q}}_{\text{aug}}(t_-), \mathbf{u}(t_-))$ with $l = 1, \dots, 3$ represents one of the cost functions introduced in Section III-A, integrated along the time interval $[t_k, t_{k+L}]$. eq (13) is the nominal system dynamics used to predict the state evolution starting from the initial state (14) provided at runtime by the observer. eq (15) are the state constraints, (16) are the control constraints, while (17) are other possible constraints such as the total

TABLE I

RMS OF THE MISMATCH ERROR OF RANDOM, J_1 , J_2 , AND J_3 FOR THE UNICYCLE VEHICLE AND QUADRUPELAL ROBOT CASE STUDIES

	Unicycle vehicle			Quadrupedal robot		
	RMS(e)	Improvement w.r.t. random	Improvement w.r.t. J_1	RMS(e)	Improvement w.r.t. random	Improvement w.r.t. J_1
random	[2.789, 3.697, 0.100]	–	–	[1.689, 2.475, 5.197]	–	–
J_1	[0.303, 0.567, 0.341]	[-89%, -85%, +60%]	–	[0.591, 0.550, 1.486]	[-65%, -78%, -71%]	–
J_2	[0.534, 0.141, 0.045]	[-81%, -96%, -46%]	[43%, -75%, -87%]	[0.399, 0.405, 0.607]	[-76%, -84%, -88%]	[-32%, -26%, -59%]
J_3	[0.148, 0.298, 0.141]	[-95%, -92%, +29%]	[-51%, -47%, -57%]	[0.316, 0.381, 0.783]	[-81%, -85%, -85%]	[-47%, -31%, -47%]

energy consumption for the execution of the task [10] or a Lyapunov constraint to better ensure stability [23].

Problem 1 is solved using the CasADi tool [24] by rewriting it as a Nonlinear Programming (NLP) problem. $\mathbf{u}_{\text{ref}}^*(t)$, is exploited to compute the current system state estimate, $\hat{\mathbf{q}}_{\text{aug}}(t)$. The state estimation-input pair is then added to the GP training set. The GP is updated at intervals of ΔT_{update} , where $\Delta T_{\text{update}} > \Delta T$. By setting the update time distinct from the sampling time, the samples are spatially more separated, improving the GP's ability to capture the characteristics of the unknown model more effectively.

IV. RESULTS

To validate our methodology, we conduct tests on simulated unicycle and quadrupedal robots. An Extended Kalman Filter is employed as an observer, and training set samples are collected online by optimizing each objective function detailed in Section III-A. In addition, we also compare our method with a random approach generated by applying random input. Each learned GP, yielding estimates of the unknown dynamics \hat{f}_u , is validated on a testing trajectory of 200 samples for both case studies. The testing trajectory consists of $(\mathbf{q}^{\text{test}}, \dot{\mathbf{q}}^{\text{test}})^\top$, along with the corresponding control inputs \mathbf{u}^{test} . For each sample of the testing trajectory, we compute the mismatch error

$$\mathbf{e} = \dot{\mathbf{q}}^{\text{test}} - \left(\mathbf{f}_n(\mathbf{q}^{\text{test}}, \mathbf{u}^{\text{test}}) + \hat{\mathbf{f}}_u(\mathbf{q}^{\text{test}}, \mathbf{u}^{\text{test}}) \right) \quad (18)$$

where the second term in brackets is the reconstructed dynamics. To show the effectiveness of our approach, we compare the Root Mean Square (RMS) of the mismatch errors. Moreover, we consider $\sigma_e = \sigma_a = 1$, for both case studies.

A. Unicycle Vehicle

Let us consider a unicycle vehicle affected by an unknown external disturbance $\mathbf{f}_u(\mathbf{q}, \mathbf{u})$, whose dynamics is described by (1) with $\mathbf{q} = [p_x, p_y, \theta]^\top$, $\mathbf{u} = [v, \omega]^\top$ and $\mathbf{f}_n(\mathbf{q}, \mathbf{u}) = [v \cos \theta, v \sin \theta, \omega]^\top$. According to (7), the augmented unicycle vehicle has $\mathbf{q}_{\text{aug}} = [p_x, p_y, \theta, \dot{p}_x, \dot{p}_y, \omega, v]^\top$ as state and $\mathbf{u}_{\text{aug}} = [\dot{v}, \dot{\omega}]^\top$ as input. The vehicle starts from $\mathbf{q}_0 = \mathbf{0}_6$ with zero estimation error and initial uncertainty $\mathbf{P}_0 = 0.4 \mathbf{I}_{6 \times 6}$. The onboard sensors provide noisy distances w.r.t. four markers located at (0, -5) m, (0, 5) m, (10, 6) m, (-10, -10) m. The measurement noise covariance matrix is $\mathbf{R} = 0.25 \mathbf{I}_{4 \times 4}$. Moreover, $-3 \text{ m/s} \leq v \leq 3 \text{ m/s}$ and $-3 \text{ rad/s} \leq \omega \leq 3 \text{ rad/s}$, while, $\Delta T = 0.1 \text{ s}$, and $L = 10$ samples. Finally, the unknown disturbance is defined as $\mathbf{f}_u(\mathbf{q}, \mathbf{u}) = (-0.3 \sin \theta, 0.3 \cos \theta, 0.3 \cos \theta \sin \theta)^\top$. The training set is built online by solving Problem 1 and updating the GP every $\Delta T_{\text{update}} = 0.5 \text{ s}$, continuing until the training set reaches 200 samples. Each sample of the training set consists of the input $\mathbf{x}^j = [\hat{p}_x^j, \hat{p}_y^j, \hat{\theta}^j, v^j, \omega^j]^\top$ and the output $\mathbf{y}^j = [\hat{p}_x^j, \hat{p}_y^j, \hat{\theta}^j]^\top - \mathbf{f}_n(\hat{\mathbf{q}}^j, \mathbf{u}^j)$. Once a testing trajectory $(\mathbf{q}^{\text{test}}, \dot{\mathbf{q}}^{\text{test}})$

and corresponding input \mathbf{u}^{test} have been chosen, we compute the mismatch error defined in (18) along such a trajectory, with $\mathbf{q}^{\text{test}} = [p_x^{\text{test}}, p_y^{\text{test}}, \theta^{\text{test}}]^\top$, $\dot{\mathbf{q}}^{\text{test}}$, and $\mathbf{u}^{\text{test}} = [v^{\text{test}}, \omega^{\text{test}}]^\top$. For this case study, we have $\dot{\mathbf{q}}^{\text{test}} = \mathbf{f}_n(\mathbf{q}^{\text{test}}, \mathbf{u}^{\text{test}}) + \mathbf{f}_u(\mathbf{q}^{\text{test}}, \mathbf{u}^{\text{test}})$, and hence (18) reduces to $\mathbf{e} = \mathbf{f}_u(\mathbf{q}^{\text{test}}, \mathbf{u}^{\text{test}}) - \hat{\mathbf{f}}_u(\mathbf{q}^{\text{test}}, \mathbf{u}^{\text{test}})$. Finally, the RMS of the mismatch errors are compared in Table I. Overall, all three performance indices improve the quality of the training set over the random approach. Comparing J_2 and J_3 to J_1 indicates that the active sensing part further enhances the quality of the training dataset. In general, the RMSE values for J_2 and J_3 are smaller than those for J_1 . Moreover, J_3 allows to achieve the best results in terms of mismatch errors. Finally, we have evaluated the time needed for the GP update and for solving Problem 1, both w.r.t. the training set size. The update time increases with the growth of the training set dimensions, confirming the computational expense of updating the GP scales as $\mathcal{O}(nN^2)$ [25]. Furthermore, J_3 consistently exhibits long times for finding the optimal solution of Problem 1, mainly due to a not-optimized code w.r.t. the one of the quadrupedal robot.

B. Quadrupedal Robot

We now test the proposed metrics on a more complex dynamic system such as a quadrupedal robot. According to [27], its approximated linear Single Rigid Body Dynamics (SRBD) in the world frame is

$$\dot{\mathbf{q}} = \begin{pmatrix} \dot{\mathbf{p}} \\ \dot{\Theta} \\ \mathbf{f}_n(\mathbf{q}, \mathbf{u}) \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \end{pmatrix}, \quad (19)$$

with $\mathbf{f}_n(\mathbf{q}, \mathbf{u}) = [\sum_{i=1}^4 \mathbf{f}_i + \mathbf{g}, \sum_{i=1}^4 \mathbf{I}(\psi) \mathbf{r}_i \times \mathbf{f}_i]^\top$, $\mathbf{q} = [p, \Theta, \dot{p}, \dot{\Theta}]^\top$ and $\mathbf{u} = [f_1, \dots, f_4]^\top$. Moreover, $\mathbf{p} \in \mathbb{R}^3$ is the robot's position, m is the robot's mass, $\mathbf{g} \in \mathbb{R}^3$ is the gravity vector, and $\mathbf{I} \in \mathbb{R}^3$ is the robot's inertia tensor, $\Theta = [\phi, \theta, \psi]^\top$ is the robot's orientation with ϕ , θ , and ψ are the roll, pitch, and yaw angles, respectively. Notice that, we assumed that the roll and pitch angles do not significantly vary during the robot's motion. Additionally, $\mathbf{r}_i \in \mathbb{R}^3$ is the vector connecting the center of mass (CoM) to the point where the force $\mathbf{f}_i \in \mathbb{R}^3$ is applied with $i = 1 \dots 4$. The system dynamics expressed by (19) represents a simplified version of the more complex nonlinear one. Therefore, $\mathbf{f}_u(\mathbf{q}(t), \mathbf{u}(t))$ ($\mathcal{O}(q^2)$ is zero in this case) represents the legs dynamics, the robot-ground interaction, and the small angles approximation on pitch and roll in the SRBD. According to (7), the new augmented state is $\mathbf{q}_{\text{aug}} = [p, \Theta, \dot{p}, \dot{\Theta}, \ddot{p}, \ddot{\Theta}, f_1, \dots, f_4]^\top$ with input $\mathbf{u}_{\text{aug}} = \dot{f}_i$. The robot is simulated in PyBullet, as shown in Fig. 2. We consider $\mathbf{q}_0 = \mathbf{0}_{18}$ with zero estimation error and initial uncertainty $\mathbf{P}_0 = 0.3^2 \mathbf{I}_{18 \times 18}$. An onboard IMU provides the linear acceleration and the angular velocity of the robot CoM. Additionally, a laser scan provides noisy distances of the floating base CoM w.r.t. fixed landmarks. Moreover,

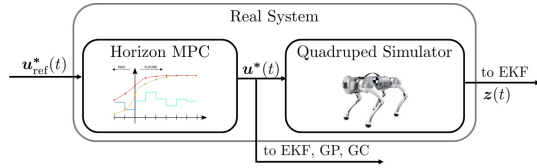


Fig. 2. Real system block in Fig. 1 for the quadrupedal robot. u_{ref}^* are the robot CoM reference. Horizon [26] converts it to the optimal input u^* used to update the state estimate and hence the training set.

$R = 0.1^2 I_{9 \times 9}$, $\Delta T = 0.04\text{s}$ and $L = 20$ samples. The training set is built with GP sample collected every $\Delta T_{\text{update}} = 0.12\text{s}$, until the training set reaches 200 samples. Each sample of the training set consists of the input $x^j = [\hat{p}^j, \hat{\Theta}^j, \hat{p}^j, \hat{\Theta}^j, f_i^j, r_i^j]^\top$ and the output $y^j = [\hat{p}^j, \hat{\Theta}^j]^\top - f_n(\hat{q}^j, u^j)$ with $j = 1, \dots, 200$. The mismatch error (18) is computed on a testing trajectory with $q^{\text{test}} = [\hat{p}^{\text{test}}, \hat{\Theta}^{\text{test}}]^\top$, \dot{q}^{test} , $u^{\text{test}} = [f_i^{\text{test}}, r_i^{\text{test}}]$, and $i = 1, \dots, 4$, provided by the PyBullet simulator. Since we noticed small variations in the height, roll, and pitch angles during motions, we have excluded these variables from the discussion of the results. Table I shows the comparison between the RMS of mismatch errors. The RMSE values for J_2 and J_3 are smaller than those for the random approach and J_1 , confirming that the active sensing component contributes to enhancing the quality of the training dataset. The case study involving the quadrupedal robot demonstrates a rising trend in the GP update time as the size of the training set increases, following an $\mathcal{O}(nN^2)$ as discussed for the unicycle vehicle. As stated in the previous section, the time required to find the optimal solution of Problem 1 should exhibit a growing trend with the size of the training set. This trend is evident for the unicycle vehicle but not for the quadrupedal robot. This is due to a more optimized code version used in the latter case (based on the open-source framework Horizon [26]).

V. CONCLUSION AND FUTURE WORKS

In this letter, we have proposed an online optimal perception-exploration control approach. Results have clearly shown the data quality improvement obtained by including active sensing. We have also observed an increasing trend in GP update times as a function of the training set size for both case studies. Future works will aim at optimizing the GP model size by adding a sub-sampling level, e.g., based on Nystrom-type methods [28] or a subset data technique [25], to reduce the growth rate of the training set size, which is of paramount importance for real-time experiments.

REFERENCES

- [1] L. Budach et al., “The effects of data quality on machine learning performance,” 2022, *arXiv:2207.14529*.
- [2] P. Oliveira, F. Rodrigues, P. Henriques, and H. Galhardas, “A taxonomy of data quality problems,” in *Proc. 2nd Int. Workshop Data Inf. Qual.*, 2005, pp. 219–233.
- [3] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, “Robots that can adapt like animals,” *Nature*, vol. 521, no. 7553, pp. 503–507, 2015.
- [4] A. Jain, T. Nghiem, M. Morari, and R. Mangharam, “Learning and control using gaussian processes,” in *Proc. ACM/IEEE 9th Int. Conf. Cyber-Phys. Syst. (ICCCPS)*, 2018, pp. 140–149.
- [5] M. Blanke and M. Lelarge, “FLEX: An adaptive exploration algorithm for nonlinear systems,” 2023, *arXiv:2304.13426*.
- [6] M. Blanke and M. Lelarge, “Online greedy identification of linear dynamical systems,” in *Proc. IEEE 61st Conf. Decis. Control (CDC)*, 2022, pp. 5363–5368.
- [7] M. Buisson-Fenet, F. Solowjow, and S. Trimpe, “Actively learning Gaussian process dynamics,” in *Proc. Learn. Dyn. Control*, 2020, pp. 5–15.
- [8] O. Napolitano, D. Fontanelli, L. Pallottino, and P. Salaris, “Gramian-based optimal active sensing control under intermittent measurements,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 9680–9686.
- [9] K.-C. Ma, L. Liu, and G. S. Sukhatme, “An information-driven and disturbance-aware planning method for long-term ocean monitoring,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2016, pp. 2102–2108.
- [10] P. Salaris, M. Cognetti, R. Spica, and P. Robuffo Giordano, “Online optimal perception-aware trajectory generation,” *IEEE Trans. Robot.*, vol. 35, no. 6, pp. 1307–1322, Dec. 2019.
- [11] J. Shan and Q. Sun, “Observability analysis and optimal information gathering of mobile robot navigation system,” in *Proc. IEEE Int. Conf. Inf. Autom.*, 2015, pp. 731–736.
- [12] J. Ott, E. Balaban, and M. J. Kochenderfer, “Sequential Bayesian optimization for adaptive informative path planning with multimodal sensing,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2023, pp. 7894–7901.
- [13] P. Tokekar, J. Vander Hook, and V. Isler, “Active target localization for bearing based robotic telemetry,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2011, pp. 488–493.
- [14] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. (Adaptive computation and machine learning). Cambridge, MA, USA: MIT Press, 2006.
- [15] T. M. Cover and J. A. Thomas, *Elements of Information Theory Second Edition Solutions to problems* (Internet Access). Hoboken, NJ, USA: Wiley, 2006, pp. 19–20.
- [16] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, “Gaussian process optimization in the bandit setting: No regret and experimental design,” 2009, *arXiv:0912.3995*.
- [17] J. M. B. Dias and J. M. N. Leitão, “A method for computing the information matrix of stationary gaussian processes,” in *Proc. 8th Eur. Signal Process. Conf. (EUSIPCO)*, 1996, pp. 1–4.
- [18] M. Mukadam, J. Dong, X. Yan, F. Dellaert, and B. Boots, “Continuous-time gaussian process motion planning via probabilistic inference,” *Int. J. Robot. Res.*, vol. 37, no. 11, pp. 1319–1340, 2018.
- [19] M. Behr, P. Benner, and J. Heiland, “Solution formulas for differential Sylvester and Lyapunov equations,” *Calcolo*, vol. 56, no. 4, p. 51, 2019.
- [20] K. Hausman, J. Preiss, G. S. Sukhatme, and S. Weiss, “Observability-aware trajectory optimization for self-calibration with application to UAVs,” *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1770–1777, Jul. 2017.
- [21] Z. Yi et al., “Active tactile object exploration with gaussian processes,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2016, pp. 4925–4930.
- [22] F. Pukelsheim, *Optimal Design of Experiments*. Philadelphia, PA, USA: SIAM, 2006.
- [23] O. Napolitano, D. Fontanelli, L. Pallottino, and P. Salaris, “Information-aware Lyapunov-based MPC in a feedback-feedforward control strategy for autonomous robots,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4765–4772, Apr. 2022.
- [24] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, “CasADi—A software framework for nonlinear optimization and optimal control,” *Math. Program. Comput.*, vol. 11, no. 1, pp. 1–36, 2019.
- [25] F. Amadio, A. Dalla Libera, R. Antonello, D. Nikovski, R. Carli, and D. Romeres, “Model-based policy search using monte carlo gradient estimation with real systems application,” *IEEE Trans. Robot.*, vol. 38, no. 6, pp. 3879–3898, Dec. 2022.
- [26] F. Rucelli, A. Laurenzi, N. G. Tsarakakis, and E. Mingo Hoffman, “Horizon: A trajectory optimization framework for robotic systems,” *Front. Robot. AI*, vol. 9, Jul. 2022, Art. no. 899025.
- [27] J. Di Carlo, P. M. Wensing, B. Katz, G. Bledt, and S. Kim, “Dynamic locomotion in the mit cheetah 3 through convex model-predictive control,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2018, pp. 1–9.
- [28] A. Rudi, R. Camoriano, and L. Rosasco, “Less is more: Nystrom computational regularization,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–31.