IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

A Multi-Fidelity Bayesian Approach to Safe Controller Design

Ethan Lau Vaibhav Srivastava Shaunak D. Bopardikar

***Abstract*— Safely controlling unknown dynamical systems is one of the biggest challenges in the field of control systems. Oftentimes, an approximate model of a system's dynamics exists which provides beneficial information for control design. However, differences between the approximate and true systems present challenges as well as safety concerns. We propose an algorithm called SAFESLOPE to safely evaluate points from a Gaussian process model of a function when its Lipschitz constant is unknown. We establish theoretical guarantees for the performance of SAFESLOPE and quantify how multi-fidelity modeling improves the algorithm's performance. Finally, we present a case where SAFESLOPE achieves lower cumulative regret than a naive sampling method by applying it to find the control gains of a linear time-invariant system.**

I. INTRODUCTION

In the realm of control systems, there exist many instances in which the dynamics are not fully modeled. While an approximation of the dynamics may exist, variations in the system's components or environment may cause the system to deviate from the design model. For example, consider off-the-shelf robotics kits. Though identically designed, each robot possesses variations that cause its performance to vary from the design model. In this case, we can consider each robot to be a *black-box system*, possessing accessible input-output data but inaccessible exact dynamics. We study how the true system output can be used with a design or simulated model to create an improved model of the true dynamical system.

Gaussian process (GP) regression is a popular non-parametric technique for optimizing unknown or difficult-to-evaluate cost functions. The upper confidence bound (UCB) algorithm [1] guarantees asymptotic zero regret when iteratively sampling a GP. Multi-fidelity Gaussian processes (MF-GPs) predict a distribution from multiple correlated inputs. The linear auto-regressive (AR-1) model is an MF-GP that uses a cheaper model to assist in evaluating a more complex model [2]. The AR-1 model's recursive structure allows it to effectively model correlated processes while its decoupled form enables computationally efficient parameter learning. Analytical guarantees have also been established when applying Bayesian optimization to MF-GPs [3], [4].

Recently, GPs have been explored for control design. GPs and MF-GPs have been applied to finding ideal control gains for linear time-invariant (LTI) systems [5], [6]. MF-GPs have also been applied to falsification frameworks for testing system safety [7]. However, these papers primarily contain experimental results, without any mathematical guarantees for the approach.

Other data-driven methods have been proposed to control LTI systems. Model-based approaches reconstruct a model of the system dynamics from trajectories of similar systems [8], [9] and have been studied for robustness [10]. When data is abundant, model predictive control may be used to find an ideal control strategy [11]. Model-free approaches aim to directly control a system without learning the system dynamics [12]–[14].

Whether model-based or model-free, a critical aspect of controller design is safety. A recent review of safe learning in control classifies approaches based on the strength of the safety guarantee and the required knowledge of the system's dynamics [15]. An ideal approach ensures strict constraints are met for a system with unknown dynamics. Despite proposed solutions, there is a gap in work involving using GPs for safe control design.

We consider a data-driven Bayesian optimization approach to find optimal controllers of black-box systems. The following are our main contributions:

1) We establish SAFESLOPE, a safe exploration algorithm with analytical bounds when the Lipschitz constant of a black-box cost function is unknown. Unlike SAFEOPT [16], which relies on a known Lipschitz constant, we upper bound the slope using the posterior distribution of the GP.

2) We formalize how an AR-1 model can improve the choice of inputs. In particular, we show how its conditional covariance matrix can be used to reduce the upper bound on the information gain. We also numerically compare the performance of an AR-1 model to a single-fidelity GP.

II. PROBLEM OVERVIEW

A. Motivating Scenario

For this problem, we model a true system with LTI dynamics, $\boldsymbol{z}_{j+1} = A\boldsymbol{z}_j + B\boldsymbol{u}_j$, where $\boldsymbol{z} \in \mathbb{R}^n$ is the state, $\boldsymbol{u} \in \mathbb{R}^p$ is the input, and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$ are the system matrices. Under feedback control, the system input is $\boldsymbol{u}_j = -K\boldsymbol{z}_j$, where $K \in \mathbb{R}^{p\times n}$ is the control gain. Given an initial state $\boldsymbol{z}_0$ and weighting matrices Q and R, the system's infinite-horizon LQR cost for a set of gains K is

$$J(K)=\sum_{j=0}^{\infty} \boldsymbol{z}_0^T(A-BK)^{Tj}[Q+K^TRK](A-BK)^j\boldsymbol{z}_0. \quad (1)$$

Our goal is to minimize (1) by finding the ideal gain K^*.[1]

This work was supported in part by ARO grant W911NF-18-1-0325 and in part by NSF Award CNS-2134076.

The authors are with the Electrical and Computer Engineering Department at Michigan State University.

[1]We demonstrate the algorithm on an LTI system with a quadratic cost for simplicity's sake. However, our algorithm may also be applied to any system possessing a parameterized controller with a measurable performance metric.

When A and B are *unknown*, determining an ideal K^* becomes more challenging. We consider a situation in which a design model of the system has the evolution $\boldsymbol{z}_{j+1} = \hat{A}\boldsymbol{z}_j + \hat{B}\boldsymbol{u}_j$ and associated cost $\hat{J}$, with $\hat{A} \in \mathbb{R}^{n\times n}$, $\hat{B} \in \mathbb{R}^{n\times p}$. The design model has the same dimension as the true system, but its entries differ from those in the true system. We aim to leverage the design model to quickly find an ideal K^* while avoiding gains that cause instability.

We propose using an MF-GP framework that *only requires the input-output data* from the auxiliary and the true systems. Here, the input is the choice of gain K, and the output is $J(K)$. We apply an AR-1 model by treating $(\hat{A}, \hat{B})$ and (A, B) as the low- and high-fidelity models, respectively. By using a search algorithm that guarantees safety, we seek to avoid sampling unstable controller gains.

B. Multi-Fidelity Gaussian Processes (MF-GPs)

A Gaussian process is a collection of random variables such that every finite set of random variables has a multivariate Gaussian distribution [17]. A GP is defined over a space $\mathcal{X} \subset \mathbb{R}^n$ by its mean function $\mu : \mathcal{X} \to \mathbb{R}$ and its covariance (kernel) function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

Given a set of points $\boldsymbol{X}_t = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t\}$, we create a covariance matrix $\boldsymbol{k}(\boldsymbol{X}_t, \boldsymbol{X}_t) = [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1}^{t,t}$, which is always positive definite. The covariance between a point and a set of points yields a covariance vector $\boldsymbol{k}(\boldsymbol{x}) := \boldsymbol{k}(\boldsymbol{X}_t, \boldsymbol{x}) = [k(\boldsymbol{x}_1, \boldsymbol{x}) \ldots k(\boldsymbol{x}_t, \boldsymbol{x})]^T$.

Let f be a sample from a GP with mean μ and kernel k. Suppose we have prior data $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t = \{y_1, \ldots, y_t\}$, where $y_i = f(\boldsymbol{x}_i) + \eta$ has measurement noise $\eta \sim N(0, \xi^2)$. Then the posterior distribution of f at $\boldsymbol{x}$ is a normally distributed random variable with mean $\mu_{f,t+1}$, covariance $k_{f,t+1}$, and standard deviation $\sigma_{f,t+1}$ given by

$$\mu_{f,t+1}(\boldsymbol{x}) = \boldsymbol{k}^T(\boldsymbol{x})[\boldsymbol{k}(\boldsymbol{X}_t, \boldsymbol{X}_t) + \xi^2 I]^{-1}\boldsymbol{Y}_t \tag{2}$$

$$k_{f,t+1}(\boldsymbol{x}, \boldsymbol{x}') = k_{f,t}(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{k}^T(\boldsymbol{x})[\boldsymbol{k}(\boldsymbol{X}_t, \boldsymbol{X}_t) + \xi^2 I]^{-1}\boldsymbol{k}(\boldsymbol{x}')$$

$$\sigma_{f,t+1}(\boldsymbol{x}) = \sqrt{k_{f,t+1}(\boldsymbol{x}, \boldsymbol{x})}. \tag{3}$$

To incorporate data from multiple sources, we use an AR-1 model, which models f as a linear combination of a low-fidelity GP $f_L(\boldsymbol{x})$ and an error GP $\delta(\boldsymbol{x})$ according to

$$f(\boldsymbol{x}) = \rho f_L(\boldsymbol{x}) + \delta(\boldsymbol{x}), \tag{4}$$

where ρ is a scaling constant [2]. In general, an AR-1 model is beneficial when the low-fidelity observations $\boldsymbol{X}_L$ are more abundant than the high-fidelity observations $\boldsymbol{X}_H$.

Let $\boldsymbol{k}^{(L)}$ denote the kernel of $f_L(\boldsymbol{x})$ and $\boldsymbol{k}^{(\delta)}$ denote the kernel of $\delta(\boldsymbol{x})$. Then, letting $\boldsymbol{X} = [\boldsymbol{X}_L, \boldsymbol{X}_H]$, the covariance matrix of the AR-1 model has the form

$$\boldsymbol{k}^{(MF)}(\boldsymbol{X}, \boldsymbol{X}) = \begin{bmatrix} \boldsymbol{k}^{(L)}_{L,L} & \rho\boldsymbol{k}^{(L)}_{L,H} \\ \rho\boldsymbol{k}^{(L)}_{H,L} & \rho^2\boldsymbol{k}^{(L)}_{H,H} + \boldsymbol{k}^{(\delta)}_{H,H} \end{bmatrix}, \tag{5}$$

where $\boldsymbol{k}^{(L)}_{L,H}$ is shorthand notation for the single-fidelity covariance matrix $\boldsymbol{k}^{(L)}(\boldsymbol{X}_L, \boldsymbol{X}_H)$.

C. Problem Statement

Consider a finite domain $\mathcal{X} \subset \mathbb{R}^n$, with $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}$. Let $f : \mathcal{X} \to \mathbb{R}$ be an unknown realization of a GP and let $\boldsymbol{x}^*$ be a minimizer of f. Given a safety barrier $h \in \mathbb{R}$ and precision $\epsilon > 0$, our goal is to design a sequence $\{\boldsymbol{x}_t\}_{t\in\mathbb{N}}$ such that for some sufficiently large t^*,

$$f(\boldsymbol{x}_t) < f(\boldsymbol{x}^*) + \epsilon,\ \forall t > t^*;\ \text{and } f(\boldsymbol{x}_t) \le h\ \forall t \in \mathbb{N}.$$

We develop an iterative algorithm to design such a sequence $\{\boldsymbol{x}_t\}_{t\in\mathbb{N}}$. We apply this framework to the multi-fidelity case when an approximation of $f(\boldsymbol{x})$ is available.

III. Algorithms and Main Results

In this section, we first review the SafeOpt algorithm, which forms the framework of SafeSlope. Next, we introduce SafeSlope and describe how it deviates from SafeOpt. We then discuss how SafeSlope applies to MF-GPs, then discuss the theoretical properties of this algorithm.

A. The SafeOpt Algorithm [16]

SafeOpt is an exploration algorithm that uses the Lipschitz constant L of a function f to avoid searching in an unsafe domain. To accomplish this, SafeOpt uses the predictive confidence interval

$$Q_{f,t}(\boldsymbol{x}) := \left[Q^-_{f,t}(\boldsymbol{x}), Q^+_{f,t}(\boldsymbol{x})\right], \tag{6}$$

where $Q^\pm_{f,t}(\boldsymbol{x}) := \mu_{f,t-1}(\boldsymbol{x}) \pm \beta^{1/2}_{f,t}\sigma_{f,t-1}(\boldsymbol{x})$ and $\beta_{f,t}$ is a parameter which controls exploration.

Step 1: Given an initial safe set S_0, we define $C_{f,0}(\boldsymbol{x}) := [h, \infty)$, $\forall \boldsymbol{x} \in S_0$ and $\mathbb{R}$ otherwise. Then, the nested confidence interval $C_{f,t}(\boldsymbol{x}) = C_{f,t-1}(\boldsymbol{x}) \cap Q_{f,t}(\boldsymbol{x})$ is used to define the upper and lower confidence bounds of f as

$$u_{f,t}(\boldsymbol{x}) := \max C_{f,t}(\boldsymbol{x}) \text{ and } \ell_{f,t}(\boldsymbol{x}) := \min C_{f,t}(\boldsymbol{x}). \tag{7}$$

Step 2: These confidence bounds are used to establish the subsequent safe sets S_t according to

$$S_t = \bigcup_{\boldsymbol{x}\in S_{t-1}} \{\boldsymbol{x}' \in \mathcal{X} \mid u_{f,t}(\boldsymbol{x}) + Ld(\boldsymbol{x}, \boldsymbol{x}') \le h\},$$

where $d(\boldsymbol{x}, \boldsymbol{x}')$ is the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$.

Step 3: Two subsets of S_t guide the search process. The set of points that potentially minimize f is given by

$$M_t = \left\{ x \in S_t \middle| \ell_{f,t}(\boldsymbol{x}) \le \min_{\boldsymbol{x}'\in S_t} u_{f,t}(\boldsymbol{x}') \right\}.$$

Step 4: Meanwhile, the set of points that potentially increase the size of S_t is given by

$$G_t = \{x \in S_t | g_t(\boldsymbol{x}) > 0\},$$

where $g_t(\boldsymbol{x})$ is the cardinality of the set of points that sampling at $\boldsymbol{x}$ could add to S_t, defined by

$$g_t(\boldsymbol{x}) := \left| \{\boldsymbol{x}' \in \mathcal{X}\backslash S_t \mid \ell_{f,t}(\boldsymbol{x}) + Ld(\boldsymbol{x}, \boldsymbol{x}') \le h\} \right|.$$

Step 5: From the union of M_t and G_t, SafeOpt selects points using the width of the confidence interval $w_t(\boldsymbol{x}) := u_{f,t}(\boldsymbol{x}) - \ell_{f,t}(\boldsymbol{x})$ according to the function

$$\boldsymbol{x}_t \in \operatorname*{arg\,max}_{\boldsymbol{x}\in M_t\cup G_t} w_t(\boldsymbol{x}). \tag{8}$$

B. The SAFESLOPE Algorithm

The SAFESLOPE algorithm is an adaptation of SAFEOPT with the following modification: *we assume the global Lipschitz constant is unknown* and instead use local slope predictions to avoid searching beyond the safety limit.

To do so, we model the slopes of f as GPs. For ease of presentation, we organize $\mathcal{X}$ into a hypercube with r^n points. Along each axis $i \in \{1, \dots, n\}$, we create an incidence matrix W_i with size $(r-1)r^{n-1} \times r^n$. Each W_i corresponds to the union of directed line graphs along the i-th axis. Then, at iteration t, we represent the slopes between adjacent points along the i-th axis using $m_i \in \mathbb{R}^{(r-1)r^{n-1}}$. Each m_i is a realization of a GP with mean and covariance

$$\boldsymbol{\mu}_{m_i} = W_i \cdot \boldsymbol{\mu}_{f,t}(\mathcal{X}), \quad \boldsymbol{k}_{m_i} = W_i \cdot \boldsymbol{k}_{f,t}(\mathcal{X}) \cdot W_i^T.$$

Essentially, the elements of m_i consist of evaluations of

$$m_i(\boldsymbol{x}, \boldsymbol{x}') = [\mu_f(\boldsymbol{x}') - \mu_f(\boldsymbol{x})]/d(\boldsymbol{x}', \boldsymbol{x}),$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are adjacent points along the i-th axis, $x_i' > x_i$, and $d(\boldsymbol{x}', \boldsymbol{x})$ is the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$.

Step 1: We preserve the format of SAFEOPT's safety condition by using the magnitude of the slope. Here, we use the greatest magnitude of the confidence bounds, defined by

$$q_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}') := \max\left\{\text{abs}(Q^-_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}')), \text{abs}(Q^+_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}'))\right\}, \tag{9}$$

where

$$Q^{\pm}_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}') := \mu_{m_i,t-1}(\boldsymbol{x}, \boldsymbol{x}') \pm \beta^{1/2}_{m,t}\sigma_{m_i,t-1}(\boldsymbol{x}, \boldsymbol{x}').$$

Then, we replace L with the nested upper bound on the slope

$$\hat{u}_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}') := \min\{q_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}'), \hat{u}_{m_i,t-1}(\boldsymbol{x}, \boldsymbol{x}')\}, \tag{10}$$

where $\hat{u}_{m_i,0} = \infty$.

Step 2: We now redefine the safe set as

$$S_t = \bigcup_{\boldsymbol{x} \in S_{t-1}} \bigcup_{i=1,\dots,n} \left\{\boldsymbol{x}' \in V_i(\boldsymbol{x}) \mid s_t(\boldsymbol{x}, \boldsymbol{x}') \le h\right\}, \tag{11}$$

where

$$s_t(\boldsymbol{x}, \boldsymbol{x}') = u_{f,t}(\boldsymbol{x}) + \hat{u}_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}') \cdot d(\boldsymbol{x}, \boldsymbol{x}')$$

and the vicinity V_i of $\boldsymbol{x}$ is given by

$$V_i(\boldsymbol{x}) = \{\boldsymbol{x}' \in \mathcal{X} | \boldsymbol{x}', \boldsymbol{x} \text{ are adjacent and } x_i' = x_i\}.$$

Steps 3 and 4: The definitions of M_t and G_t are the same as those in SAFE-OPT, but the growth criterion becomes

$$g_t(\boldsymbol{x}) = \left| \{\boldsymbol{x}' \in V_i(\boldsymbol{x}) \backslash S_t | \ell_{f,t}(\boldsymbol{x}) + \hat{u}_{m_i,t} d(\boldsymbol{x}, \boldsymbol{x}') \le h\} \right|.$$

Step 5: Similar to SAFEOPT, points are sampled using the redefined M_t and G_t according to (8).

C. Multi-fidelity Extension of SAFESLOPE

We can use SAFESLOPE to sample points from the highest fidelity of an MF-GP. Consider an AR-1 GP with fidelities, f_L and f. We evaluate f_L at every $\boldsymbol{x} \in \mathcal{X}$ to construct a data set $(\boldsymbol{Y}_L, \boldsymbol{X}_L)$. We also evaluate f at a starting point $\boldsymbol{x}_0 = \arg\min_{\boldsymbol{x} \in \mathcal{X}} f_L(\boldsymbol{x})$. Then, with $\boldsymbol{x}_0$ as S_0, SAFESLOPE is used to explore the AR-1 GP and find $\boldsymbol{x}^*$.

D. Reachability

Similar to SAFEOPT, the theoretical guarantees of SAFESLOPE rely on the reachability operator. Define $\hat{u}_t := [\hat{u}_{m_1,t}, \dots, \hat{u}_{m_n,t}]^T$. Then the reachability operator at time t is the set of points given by

$$R_{\epsilon,\hat{u}_t}(S) := S \cup \left\{ \boldsymbol{x}' \in \mathcal{X} \,\middle|\, \begin{array}{c} \exists \boldsymbol{x} \in S, \exists i \in \{1, \dots, n\}, \boldsymbol{x}' \in V_i(\boldsymbol{x}), \\ f(\boldsymbol{x}) + \hat{u}_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}') \cdot d(\boldsymbol{x}, \boldsymbol{x}') + \epsilon \le h \end{array} \right\},$$

where $\hat{u}_{m_i,t}(\boldsymbol{x}, \boldsymbol{x}')$ is the upper bound on the slope between $\boldsymbol{x}$ and $\boldsymbol{x}'$ at time t. Given the current set of safe points, the reachability operator provides the total collection of points that could be sampled as f is learned within S.

The T-step reachability operator is defined by

$$R^T_\epsilon(S) := R_{\epsilon,\hat{u}_T}(R_{\epsilon,\hat{u}_{T-1}} \dots (R_{\epsilon,\hat{u}_0}(S))). \tag{12}$$

By taking the limit, we obtain the closure set $\bar{R}_\epsilon(S) := \lim_{T\to\infty} R^T_\epsilon(S)$. Because SAFESLOPE never explores outside $\bar{R}_\epsilon(S_0)$ with probability 1, we modify our optimization goal from Section II-C to take the equivalent form,

$$f^*_\epsilon = \min_{\boldsymbol{x} \in \bar{R}_\epsilon(S_0)} f(\boldsymbol{x}).$$

E. Theoretical Results

For Bayesian approaches, we measure the information gain after sampling a set of points $A \subseteq \mathcal{X}$ as $I(\boldsymbol{y}_A; \boldsymbol{f}_A) = H(\boldsymbol{y}_A) - H(\boldsymbol{y}_A|f)$, where $\boldsymbol{y}_A$ is a random vector of noisy observations of f evaluated at every point in A, $\boldsymbol{f}_A$ is the vector of true values of f at every point in A, and H is the entropy of the vector. The maximum information gain after T evaluations of f is given by

$$\gamma_T = \max_{A \subset \mathcal{X}, |A| = T} I(\boldsymbol{y}_A; \boldsymbol{f}_A). \tag{13}$$

A bound on the γ_T can be found in [1, Eq. (8)]. With the information gain defined, we now move to the main theorem.

Theorem 3.1 (Single-Fidelity SAFESLOPE *Guarantees):* Define $\hat{\boldsymbol{x}}_t := \arg\min_{x \in S_t} u_{f,t}(\boldsymbol{x})$. Select $\delta_f, \delta_m \in (0, 1)$. Set $\beta_{f,t} = 2\log(|\mathcal{X}|\pi_t/\delta_f)$ and $\beta_{m,t} = 2\log(|\mathcal{X}|n\pi_t/\delta_m)$, where $\sum_{t \ge 1} \pi_t^{-1} = 1$ with $\pi_t > 0$. Given an initial safe set $S_0 \neq \varnothing$, with $f(\boldsymbol{x}) \le h$ for each $\boldsymbol{x} \in S_0$, let t^* be the smallest positive integer satisfying

$$\frac{t^*}{\gamma_{t^*}\beta_{f,t^*}} \ge \frac{C_1\left(|\bar{R}_0(S_0)| + 1\right)}{\epsilon^2},$$

where $C_1 = 8v^2/\log(1 + v^2\xi^{-2})$, v^2 is the kernel variance, and $|\bullet|$ denotes cardinality. Then, for any $\epsilon > 0$, using SAFESLOPE with $\beta_{f,t}$ and $\beta_{m,t}$ results in the following.

- With probability at least $1 - \delta_f - \delta_m$,

$$\forall t \ge 1, f(\boldsymbol{x}_t) \le h.$$

- With probability at least $1 - \delta_f$,

$$\forall t \ge t^*, f(\hat{\boldsymbol{x}}_t) < f^*_\epsilon + \epsilon. \qquad \blacksquare$$

The first point of Theorem 3.1 states that with high probability, SAFESLOPE will sample points under a threshold h. This probability is directly tied to β_f and β_m, parameters

that quantify the algorithm's tendency to explore points in unexplored regions. The second point states that with high probability, after time t^*, the minimum yielded by SAFESLOPE will fall within an ϵ-neighborhood of f_ϵ^*. This value of t^* scales intuitively with the information gain γ_{t^*}, since more information to learn requires a greater search iteration count. Because γ_{t^*} lacks a closed-form solution, a bound on γ_{t^*} is typically used instead.

Our second main result is an extension of Theorem 3.1 to an AR-1 model. But first, we establish an upper bound on the information gain γ_T for an AR-1 model.

Theorem 3.2 (Information Gain Bound for an AR-1 GP): Consider the information gain γ_T from (13). For a linear auto-regressive GP with noise-free ($\xi_L^2 = 0$) low-fidelity observations at $\boldsymbol{X}_L$ and high-fidelity observations at $\boldsymbol{X}_H \subseteq \boldsymbol{X}_L$, the information gain γ_T is upper bounded by

$$\tilde{\gamma}_T := \frac{1/2}{1-e^{-1}} \max_{m_1,\dots,m_T} \sum_{t=1}^{T} \log\left(1+\xi^{-2} m_t \lambda_t^{(\delta)}\right), \quad (14)$$

where $\sum_{i=1}^{T} m_i = T$ and $\lambda_t^{(\delta)}$ are the eigenvalues of the error covariance matrix $\boldsymbol{k}_{H,H}^{(\delta)}$.

Proof: Suppose we have the high- and low-fidelity input points $\boldsymbol{X}_H$ and $\boldsymbol{X}_L$, where $\boldsymbol{X}_H \subseteq \boldsymbol{X}_L$, $\boldsymbol{X}_{H'} = \boldsymbol{X}_L \backslash \boldsymbol{X}_H$, and each entry of $\boldsymbol{X}_L$ is unique. Then, $\boldsymbol{X}_L = \boldsymbol{X}_H \cup \boldsymbol{X}_{H'}$. Since the covariance matrix is always positive definite, $\boldsymbol{k}_{L,L}^{(L)}$ is invertible, and the covariance of the high-fidelity data conditioned on the low-fidelity data is given by

$$\begin{aligned}
&\boldsymbol{k}(f_H(\boldsymbol{X}_H), f_H(\boldsymbol{X}_H) | f_L(\boldsymbol{X}_L) = \boldsymbol{y}_L, f_H(\boldsymbol{X}_H) = \boldsymbol{y}_H) \\
&= \rho^2 \boldsymbol{k}_{H,H}^{(L)} + \boldsymbol{k}_{H,H}^{(\delta)} - \rho^2 \boldsymbol{k}_{H,L}^{(L)} [\boldsymbol{k}_{L,L}^{(L)}]^{-1} \boldsymbol{k}_{L,H}^{(L)} \\
&= \rho^2 \boldsymbol{k}_{H,H}^{(L)} + \boldsymbol{k}_{H,H}^{(\delta)} - \rho^2 \begin{bmatrix} \boldsymbol{k}_{H,H'}^{(L)} & \boldsymbol{k}_{H,H}^{(L)} \end{bmatrix} \\
&\quad \times \begin{bmatrix} \boldsymbol{k}_{H',H'}^{(L)} & \boldsymbol{k}_{H',H}^{(L)} \\ \boldsymbol{k}_{H,H'}^{(L)} & \boldsymbol{k}_{H,H}^{(L)} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{k}_{H',H}^{(L)} \\ \boldsymbol{k}_{H,H}^{(L)} \end{bmatrix} = \boldsymbol{k}_{H,H}^{(\delta)},
\end{aligned}$$

where the last line is obtained using properties of block matrix inversion. In words, the conditional covariance is simply the covariance of the error GP $\delta(\boldsymbol{x})$. By applying the above result to [1, Eq. (8)], we complete the proof. ■

Remark 3.1: As the quality of a low-fidelity model improves, the variance of the error GP approaches 0. Since the eigenvalues of a covariance matrix are directly proportional to the kernel's variance, Theorem 3.2 shows that improving the low-fidelity quality decreases the eigenvalues of $\boldsymbol{k}_{H,H}^{(\delta)}$, thereby decreasing the information gain.

Theorem 3.3 (Multi-Fidelity SAFESLOPE *Guarantees):* Assume f is an AR-1 GP with the structure given in (4). Consider $\hat{\boldsymbol{x}}_t$, δ_f, δ_m, $\beta_{f,t}$, $\beta_{m,t}$, π_t, and S_0 as defined in Theorem 3.1. Let t_{MF}^* denote the smallest positive integer satisfying

$$\frac{t_{MF}^*}{\tilde{\gamma}_{t_{MF}^*} \beta_{f,t_{MF}^*}} \geq \frac{C_1\left(|\bar{R}_0(S_0)|+1\right)}{\epsilon^2},$$

where $\tilde{\gamma}_{t_{MF}^*}$ is defined by (14), $C_1 = 8v_{MF}^2/\log(1+v_{MF}^2\xi^{-2})$, and v_{MF}^2 is the variance of the AR-1 GP, given by $v_{MF}^2 = \rho v_L^2 + v_\delta^2$. Then, for any $\epsilon > 0$, using SAFESLOPE with $\beta_{f,t}$ and $\beta_{m,t}$, with probability at least $1-\delta_f$,

$$\forall t \geq t_{MF}^*, f(\hat{\boldsymbol{x}}_t) < f_\epsilon^* + \epsilon. \quad \blacksquare$$

This theorem indicates that the quality of a multi-fidelity model impacts the time t_{MF}^* to identify an optimal $\hat{\boldsymbol{x}}$. In particular, improving the quality of the low-fidelity model lowers the information gain bound $\tilde{\gamma}_{t_{MF}^*}$, thereby decreasing the time to find an optimal $\hat{\boldsymbol{x}}$.

IV. NUMERICAL RESULTS

We now apply SAFESLOPE to our motivating scenario, in which we try to find the best controller for a system when an approximate model of the system exists.

For the motivating scenario from Section II-A, consider a 2×2 LTI system. For the true system, we let

$$A = \begin{bmatrix} 0.785 & -0.260 \\ -0.260 & 0.315 \end{bmatrix}, \; B = \begin{bmatrix} 1.475 \\ 0.607 \end{bmatrix}. \quad (15)$$

By applying system identification [18] to (15) with $N_s = 12$ snapshots, we obtain the approximate model,

$$\hat{A} = \begin{bmatrix} 0.700 & -0.306 \\ -0.306 & 0.342 \end{bmatrix}, \; \hat{B} = \begin{bmatrix} 1.543 \\ 0.524 \end{bmatrix}. \quad (16)$$

Since unstable controllers result in extremely large costs, we modify the cost functions to be

$$f(\boldsymbol{x}) = \log(J(\boldsymbol{x})), \;\; f_L(\boldsymbol{x}) = \log(\hat{J}(\boldsymbol{x})), \quad (17)$$

where J and $\hat{J}$ are approximated by a 20-step horizon quadratic cost with $Q = I$, $R = 1$ and $\boldsymbol{x}$ now represents the choice of controller gains. Gaussian noise with variance $\xi^2 = 10^{-4}$ and $\xi_L^2 = 10^{-8}$ is added to evaluations of f and f_L to ensure kernel matrices are well-conditioned.

Our goal is to find the controller gains $\boldsymbol{x}^* = [x_1^* \;\; x_2^*]$ such that (17) is minimized. First, we set a search domain $\mathcal{X}$ and select an initial safe set S_0. In practice, input constraints and low-fidelity data could guide the choice of $\mathcal{X}$ and S_0. Here, we set $x_1 \in [-0.5, 4.5]$, $x_2 \in [-3.5, 1.5]$, and resolution $r = 26$. Matérn kernels are used to correlate points for each fidelity [17]. For 10 different S_0's of three points each, we observe the safety and regret of SAFESLOPE with parameters $h = 0$, $\delta_f = 0.1$, $\delta_m = 0.1$, and $\pi_t = t^2\pi^2/6$. We compare SAFESLOPE to SAFEUCB, a naive approach that solely relies on $u_{f,t}(\boldsymbol{x})$ for safety and selects points according to

$$\boldsymbol{x}_t = \underset{\boldsymbol{x} \in S_t}{\arg\max} \; w(\boldsymbol{x}_t), \text{ where } S_t = \{\boldsymbol{x} \in \mathcal{X} | u_{f,t}(x) \leq h\}.$$

We use SAFEUCB with $h = 0$, $\delta_f = 0.1$, and $\pi_t = t^2\pi^2/6$.

To compare SAFESLOPE to SAFEUCB, we use the cumulative regret up to time T, given by $R_T = \sum_{t=0}^{T} (f(\boldsymbol{x}_t) - f^*)$. Fig. 1 plots the cumulative regret and cumulative number of unsafe samples over 150 iterations. We see that in this example the multi-fidelity SAFESLOPE algorithm performs the best, with a plateau in regret after 25 iterations. In general, SAFESLOPE obtains better cumulative regret than SAFEUCB at higher iteration counts. By limiting evaluations to growth or minimizer points, SAFESLOPE eliminates non-ideal points in fewer trials. This differs from

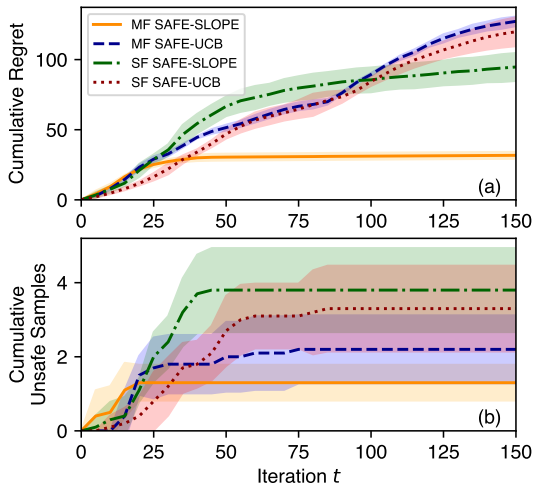

Fig. 1. (a) Cumulative regret and (b) the cumulative number of unsafe samples using SAFESLOPE and SAFEUCB, averaged across 10 trials. Error bars indicate one standard deviation.

SAFEUCB, which seeks to limit uncertainty across all safe points, rather than growth and minimizer points only. We also see both algorithms sample fewer unsafe points on MF models, with MF SAFESLOPE sampling the fewest unsafe points on average.

V. CONCLUSION

We propose SAFESLOPE, a safe exploration algorithm that leverages a function's posterior mean to predict its slopes. We preserve the safety result from SAFEOPT with a reduction in probability. By applying SAFESLOPE to an AR-1 GP, we show the search time for an optimal point corresponds to the quality of the low-fidelity approximation. Finally, we examine SAFESLOPE's performance by comparing it to a naive approach applied to single- and multi-fidelity models. We observe that applying SAFESLOPE to an MF-GP achieves lower cumulative regret while sampling fewer unsafe points.

Future research includes applying SAFESLOPE to nonlinear systems, LTI systems with disturbances, or experimental robotic applications. Another direction is designing a search algorithm which can select either fidelity for evaluation.

REFERENCES

[1] 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. on Inf. Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.

[2] M. C. Kennedy and A. O'Hagan, "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, vol. 87, no. 1, pp. 1–13, 2000.

[3] K. Kandasamy, G. Dasarathy, J. Oliva, J. Schneider, and B. Poczos, "Multi-fidelity Gaussian process bandit optimisation," *Journal of Artificial Intell. Res.*, vol. 66, pp. 151–196, 2019.

[4] J. Song, Y. Chen, and Y. Yue, "A general framework for multi-fidelity Bayesian optimization with Gaussian processes," in *Int. Conf. Artif. Intell. & Stats.*, 2019, pp. 3158–3167.

[5] A. Marco, P. Hennig, S. Schaal, and S. Trimpe, "On the design of LQR kernels for efficient controller learning," in *56th Annual Conf. Decision Control (CDC)*. IEEE, 2017, pp. 5193–5200.

[6] A. Marco *et al.*, "Virtual vs. real: Trading off simulations and physical experiments in reinforcement learning with Bayesian optimization," in *2017 Int. Conf. Robot. Autom. (ICRA)*. IEEE, 2017, pp. 1557–1563.

[7] Z. Shahrooei, M. J. Kochenderfer, and A. Baheri, "Falsification of learning-based controllers through multi-fidelity Bayesian optimization," *arXiv preprint arXiv:2212.14118*, 2022.

[8] S. Oymak and N. Ozay, "Non-asymptotic identification of LTI systems from a single trajectory," in *2019 Am. Control Conf. (ACC)*. IEEE, 2019, pp. 5655–5661.

[9] L. Xin, L. Ye, G. Chiu, and S. Sundaram, "Learning dynamical systems by leveraging data from similar systems," *arXiv preprint arXiv:2302.04344*, 2023.

[10] Y. Zheng, L. Furieri, M. Kamgarpour, and N. Li, "Sample complexity of linear quadratic Gaussian (LQG) control for output feedback systems," in *Learning for Dynamics and Control*, 2021, pp. 559–570.

[11] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, "Learning-based model predictive control: Toward safe learning in control," *Annual Rev. of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 269–296, 2020.

[12] G. Baggio, V. Katewa, and F. Pasqualetti, "Data-driven minimum-energy controls for linear systems," *IEEE Control Systems Letters*, vol. 3, no. 3, pp. 589–594, 2019.

[13] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Trans. on Autom. Control*, vol. 65, no. 3, pp. 909–924, 2019.

[14] Y. Sun and M. Fazel, "Learning optimal controllers by policy gradient: Global optimality via convex parameterization," in *60th IEEE Conf. Decision Control (CDC)*. IEEE, 2021, pp. 4576–4581.

[15] L. Brunke *et al.*, "Safe learning in robotics: From learning-based control to safe reinforcement learning," *Annual Rev. of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 411–444, 2022.

[16] Y. Sui, A. Gotovos, J. Burdick, and A. Krause, "Safe exploration for optimization with Gaussian processes," in *Int. Conf. Mach. Learn.* PMLR, 2015, pp. 997–1005.

[17] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. Springer, 2006, vol. 1.

[18] B. Ho and R. E. Kálmán, "Effective construction of linear state-variable models from input/output functions," *at-Automatisierungstechnik*, vol. 14, no. 1-12, pp. 545–548, 1966.

[19] E. Lau, V. Srivastava, and S. D. Bopardikar, "A multi-fidelity Bayesian approach to safe controller design," *arXiv preprint arXiv:2304.11023*, 2023.

APPENDIX

The following steps compose the proof of Theorem 3.1. We start by restating the upper confidence bound from Lemma 5.1 in [1].

Lemma 1.1 (UCB Bound): Let f be a function sampled from a GP. For all $t \geq 1$ and $\beta_{f,t} = 2\log(|\mathcal{X}|\pi_t/\delta_f)$ with probability $1-\delta_f$,

$$\text{abs}[f(\boldsymbol{x}) - \mu_{f,t}(\boldsymbol{x})] \leq \beta_{f,t}^{1/2}\sigma_{f,t}(\boldsymbol{x}), \ \ \forall \boldsymbol{x} \in \mathcal{X}. \qquad \blacksquare$$

Next, we show that even though multiple GPs are used to model the slopes, the UCB bound still applies.

Lemma 1.2: Suppose we have n GPs m_i over $\mathcal{X}$. For all $t \geq 1$ and $\beta_{m,t} = 2\log(|\mathcal{X}|n\pi_t/\delta_m)$ with probability at least $1-\delta_m$, the following holds for all $i = 1, \ldots, n$:

$$\text{abs}[m_i(\boldsymbol{x}) - \mu_{m_i,t}(\boldsymbol{x})] \leq \beta_{m,t}^{1/2}\sigma_{m_i,t}(\boldsymbol{x}), \ \ \forall \boldsymbol{x} \in \mathcal{X}.$$

Proof: Let A_i be the event

$$A_i = \{\text{abs}[m_i(\boldsymbol{x}) - \mu_{m_i,t}(\boldsymbol{x})] \leq \beta_{m,t}^{1/2}\sigma_{m_i,t}(\boldsymbol{x}) \forall \boldsymbol{x} \in \mathcal{X}_i\}.$$

Then, $P[A_i^c] \leq |\mathcal{X}| \cdot e^{-\beta_{m,t}/2}$. By applying DeMorgan's laws and the union bound, we obtain $P[\cap_i A_i] \geq 1 - |\mathcal{X}|ne^{-\beta_{m,t}/2}$. The remainder of the proof is identical to the proof of Lemma 5.1 in [1]. $\blacksquare$

We now establish properties of sets used in SAFESLOPE.

Lemma 1.3: The following properties hold for all $t \geq 1$.

(i) $S_{t+1} \supseteq S_t \supseteq S_0$.

(ii) $S \subseteq D \implies R_{\epsilon,\hat{u}_t}(S) \subseteq R_{\epsilon,\hat{u}_t}(D)$.
(iii) $S \subseteq D \implies \bar{R}_\epsilon(S) \subseteq \bar{R}_\epsilon(D)$.

Proof: (i) From Lemma 2 of [16], we know that (i) holds when the Lipschitz constant L of $f(\cdot)$ is known. By replacing L with $\hat{u}_{m_i,t}(x,x')$, it follows that for every $t \geq 1$ and given any $\boldsymbol{x}$, $\boldsymbol{x}'$,

$$u_{f,t+1}(\boldsymbol{x}) + \hat{u}_{m_i,t+1}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') \\ \leq u_{f,t}(\boldsymbol{x}) + \hat{u}_{m_i,t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') \leq h.$$

From the definition of $u_{f,t}$ and $\hat{u}_{m_i,t}$, it follows that these bounds are non-increasing over time, for all $\boldsymbol{x}$. The second inequality follows from (11). Therefore, $S_{t+1} \supseteq S_t \supseteq S_0$.

(ii) Let $\boldsymbol{x} \in R_{\epsilon,\hat{u}_t}(S)$. By definition of the reachability set, $\exists \boldsymbol{x}' \in S$ such that $f(\boldsymbol{x}') + \hat{u}_{m_i,t} \cdot d(\boldsymbol{x},\boldsymbol{x}') + \epsilon \leq h$. As $S \subseteq D$, this implies $\boldsymbol{x}' \in D$, which implies $\boldsymbol{x} \in R_{\epsilon,\hat{u}_t}$.

(iii) This directly follows from repeatedly applying part (ii). Each reachability step is a union of two subsets of $\mathcal{X}$, so the union is bounded by $\mathcal{X}$ and the limit exists. ■

Next, we show that the width $w(x)$ is bounded by some $\epsilon > 0$ using upper confidence bounds. Unlike [1], [16], we consider a non-unit variance for the kernel function k.

Lemma 1.4: Given a kernel with variance v^2 and measurement noise ξ^2, for each $t \geq 1$, define T_t as the smallest positive integer satisfying $\frac{T_t}{\beta_{f,t+T_t}\gamma_{t+T_t}} \geq \frac{C_1}{\epsilon^2}$, where $C_1 = 8v^2/\log(1+v^2\xi^{-2})$. If $S_{t+T_t} = S_t$, then for any $\boldsymbol{x} \in G_{t+T_t} \cup M_{t+T_t}$, it holds that $w_{t+T_t}(\boldsymbol{x}) \leq \epsilon$. ■

The proof follows the same steps as Lemma 5 in [16] and Lemma 5.4 in [1] with the difference of a non-unit kernel variance. The complete proof is provided in [19].

In the following lemmas, we assume C_1 and T_t are defined as in Lemma 1.4. We next establish guarantees on how S_t evolves with time using the reachability operator.

Lemma 1.5: For any $t \geq 1$, if $\bar{R}_\epsilon(S_0)\backslash S_t \neq \varnothing$, then with probability at least $1-\delta_f$,

$$S_{t+T_t} \supsetneq S_t. \tag{18}$$

Proof: We prove this by contradiction. First, for any $t \geq 1$, if $\bar{R}_\epsilon(S_t)\backslash S_t \neq \varnothing$, then $R_{\epsilon,\hat{u}_t}(S_t)\backslash S_t \neq \varnothing$ (by following steps identical to those in the proof of Lemma 6 in [16]). By the definition of $R_{\epsilon,\hat{u}_t}(S_t)$, we know that (a) $\exists \boldsymbol{x}' \in R_{\epsilon,\hat{u}_t}(S_t)\backslash S_t$ and (b) $\exists \boldsymbol{x} \in S_t$ so that

$$f(\boldsymbol{x}) + \epsilon + \hat{u}_{m_i,t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}) \leq h. \tag{19}$$

Now, assume that contrary to (18), $S_{t+T_t} = S_t$. This implies that $\boldsymbol{x}' \in V(S_{t+T_t})\backslash S_{t+T_t}$ and $\boldsymbol{x} \in S_{t+T_t}$. As a result, with probability at least $1-\delta_f$,

$$\begin{aligned} &\ell_{f,t+T_t}(\boldsymbol{x}) + \hat{u}_{m_i,t+T_t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') \\ &\leq f(\boldsymbol{x}) + \hat{u}_{m_i,t+T_t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') && \text{by Lemma 1.1} \\ &\leq f(\boldsymbol{x}) + \hat{u}_{m_i,t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') && \text{by (10)} \\ &\leq f(\boldsymbol{x}) + \epsilon + \hat{u}_{m_i,t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') \leq h && \text{by (19).} \end{aligned}$$

Therefore, $g_{t+T_t}(\boldsymbol{x}) > 0$ and $\boldsymbol{x} \in G_{t+T_t}$. Since we assumed that $S_{t+T_t} = S_t$ with $\boldsymbol{x} \in G_{t+T_t}$, we have

$$\begin{aligned} &u_{f,t+T_t}(\boldsymbol{x}) + \hat{u}_{m_i,t+T_t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') \\ &\leq u_{f,t+T_t}(\boldsymbol{x}) + \hat{u}_{m_i,t}(\boldsymbol{x},\boldsymbol{x}') \cdot d(\boldsymbol{x},\boldsymbol{x}') && \text{by (10)} \\ &\leq u_{f,t+T_t}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon + h && \text{by (19)} \\ &\leq w_{t+T_t}(\boldsymbol{x}) - \epsilon + h \leq h && \text{by Lemmas 1.1, 1.4.} \end{aligned}$$

Eq. (11) implies $\boldsymbol{x}' \in S_{t+T_t}$. This contradicts our assumption that $\boldsymbol{x}' \in V(S)\backslash S_{t+T_t}$. Therefore, $S_{t+T_t} \supsetneq S_t$. ■

Lemma 1.6: For any $t \geq 1$, if $S_{t+T_t} = S_t$, then with probability at least $1-\delta_f$,

$$f(\hat{\boldsymbol{x}}_{t+T_t}) \leq \min_{\boldsymbol{x} \in \bar{R}_\epsilon(S_0)} f(\boldsymbol{x}) + \epsilon.$$

Proof: By solving a minimization rather than a maximization, the first part of the proof of Lemma 8 in [16] shows that $f(\hat{\boldsymbol{x}}_{t+T_t}) \leq f(\boldsymbol{x}^*) + \epsilon$, where $\boldsymbol{x}^* := \arg\max_{\boldsymbol{x} \in S_{t+T_t}} f(\boldsymbol{x})$. Then, since $S_{t+T_t} = S_t$, Lemma 1.5 implies that $\bar{R}_\epsilon(S_0) \subseteq S_t = S_{t+T_t}$. Therefore,

$$\begin{aligned} \min_{\boldsymbol{x} \in \bar{R}_\epsilon(S_0)} f(\boldsymbol{x}) + \epsilon &\geq \min_{\boldsymbol{x} \in S_{t+T_t}} f(\boldsymbol{x}) + \epsilon \\ &= f(\boldsymbol{x}^*) + \epsilon \geq f(\hat{\boldsymbol{x}}_{t+T_t}). \end{aligned}$$

■

Corollary 1.1: For any $t \geq 1$, if $S_{t+T_t} = S_t$, then with probability at least $1-\delta_f$,

$$\forall t' \geq 0, f(\hat{\boldsymbol{x}}_{t+T_t+t'}) \leq \min_{\boldsymbol{x} \in \bar{R}_\epsilon(S_0)} f(\boldsymbol{x}) + \epsilon.$$ ■

Similar to the proof of Corollary 3 in [16], this directly follows from Lemma 1.6.

Having analyzed the evolution of the S_t, we now bound the time it takes to achieve the optimization goal.

Lemma 1.7: Let t^* be the smallest integer resulting in $t^* \geq |\bar{R}_0(S_0)|T_{t^*}$. Then, there exists a $t_0 \leq t^*$ such that $S_{t_0+T_{t_0}} = S_{t_0}$. ■

The proof of this lemma is similar to the proofs of Lemma 9 and 10 in [16], with the key difference of R depending on the upper bound of $\hat{u}_t$ instead of a global constant L. Complete proof provided in [19].

Corollary 1.2: Let t^* be the smallest integer resulting in $\frac{t^*}{\beta_{f,t^*}\gamma_{t^*}} \geq \frac{C_1(|\bar{R}_0(S_0)|+1)}{\epsilon^2}$. Then, there exists a $t_0 \leq t^*$ so that $S_{t_0+T_{t_0}} = S_{t_0}$. ■

The proof results directly from Lemmas 1.4 and 1.7.

Proof of Theorem 3.1: For the first point of Theorem 3.1, the steps are similar to the proof of Lemma 11 in [16]. For the induction step, assume $f(\boldsymbol{x}) \leq h$ for some $t \geq 1$ and any $\boldsymbol{x} \in S_{t-1}$. Then, for any $\boldsymbol{x} \in S_t$, $\exists \boldsymbol{x}' \in S_{t-1}$ along some axis i so that $h \geq u_{f,t}(\boldsymbol{x}') + \hat{u}_{m_i,t}(\boldsymbol{x}',\boldsymbol{x}) \cdot d(\boldsymbol{x}',\boldsymbol{x})$.

With probability at least $1 - e^{-\frac{1}{2}\beta_{f,t}}$,

$$h \geq f(\boldsymbol{x}') + \hat{u}_{m_i,t}(\boldsymbol{x}',\boldsymbol{x}) \cdot d(\boldsymbol{x}',\boldsymbol{x}) \qquad \text{by Lemma 1.1.}$$

With probability at least $1 - e^{-\frac{1}{2}\beta_{m,t}}$,

$$\begin{aligned} &\geq f(\boldsymbol{x}') + |m(\boldsymbol{x}',\boldsymbol{x})| \cdot d(\boldsymbol{x}',\boldsymbol{x}), && \text{by Lemma 1.2} \\ &\geq f(\boldsymbol{x}), && \text{by the definition of } m. \end{aligned}$$

By applying the union bound across $|\mathcal{X}|$ realizations of $\boldsymbol{x}$, the resulting inequality holds with probability $1-\delta_f-\delta_m$.

The second point results from Corollaries 1.1 and 1.2. ■