EVENT-TRIGGERED STRUCTURED ACTIVE EXPLORATION WITH GAUSSIAN PROCESS SAFETY FILTERS

Anonymous authors
Paper under double-blind review

ABSTRACT

For robotic systems operating under uncertain environmental conditions, accurate system models are often unavailable. To ensure their safety despite this uncertainty, the integration of Gaussian process (GP) models into safety filters based on control barrier function (CBF) has been shown to be promising. Since these GP-CBF safety filters become infeasible when the model uncertainty is too large, they are typically updated online, but existing strategies for data generation insufficiently cover the control space, introduce excessive conservatism, and temporarily compromise safety. We address this problem by proposing an event-triggered structured exploration approach for safe and sample-efficient learning of CBF-based safety filters using GPs. Before infeasibility issues can occur, our approach activates a safe exploratory control sequence, which is constructed via orthogonal perturbations of a safe control direction. Under standard regularity assumptions and sufficient CBF robustness, we prove high-probability recursive feasibility and safety, as well as finite-trigger and thereby sample complexity guarantees. Experiments on two safe robot control benchmarks show improved safety rates with lower conservatism and lower update frequency than state-of-the-art baselines. Manipulator simulations demonstrate that the GP model underlying the safety filter can be trained solely using distance-derived labels for certain system classes.

1 INTRODUCTION

Robotic systems such as drones, underwater vehicles and legged robots need to continuously satisfy safety constraints as required, e.g., for obstacle avoidance, but increasingly face uncertainty about their dynamic behavior due to a deployment in unknown environments. To overcome this key challenge for real-world deployment of autonomous systems (Dulac-Arnold et al., 2021), learning-based approaches for designing safe control policies have found growing interest in recent years (Brunke et al., 2022a; Dawson et al., 2023). In particular, Gaussian processes (GPs) (Rasmussen & Williams, 2006) are frequently employed in safety-critical applications as their probabilistic nature provides an intuitive representation of uncertainty. Due to their strong theoretical foundations, they allow the rigorous quantification of learning errors (Srinivas et al., 2012; Chowdhury & Gopalan, 2017; Lederer et al., 2019) along with their asymptotic dependency on data (Vakili et al., 2021; Lederer et al., 2026c). Therefore, they can be used to systematically tighten safety constraints, enabling the data-efficient design of provably safe control policies (Sui et al., 2015; Capone et al., 2025).

Related Work. Early ideas for the design of safe control policies base on results from black box optimization (Berkenkamp et al., 2016). By considering constraints in Bayesian optimization (Sui et al., 2015), parameters of control policies can be safely optimized through interactions with the unknown system. This has proven highly effective for simple parameterizations (Berkenkamp et al., 2023), while recent variations also indicate promising paths towards high-dimensional parameter spaces (Wei et al., 2025). By considering an underlying MDP structure, these ideas can be extended to model-based reinforcement learning (RL) with GPs (Turchetta et al., 2016; Curi et al., 2022; Wachi et al., 2023). However, these approaches are fundamentally based on an episodic learning paradigm, such that their safety guarantees are valid merely for a finite time window.

To overcome this limitation of finite-time safety certificates, so-called safety filters (Wabersich et al., 2023; Hsu et al., 2024) have become popular, which adapt nominal control inputs to maintain safety guarantees are maintained indefinitely. Among these techniques, control barrier functions (CBFs)

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

are particularly favored for safe RL (Cheng et al., 2019; Emam et al., 2025), learning-based control (Taylor et al., 2020; Qin et al., 2021), and imitation learning (Xiao et al., 2023b; Huang et al., 2026) due to their computational efficiency and ease of use. Given an accurate, control-affine dynamics model of a system, CBFs induce a quadratic program, whose solution is a safe control input (Ames et al., 2017). When model uncertainty is only state-dependent, GPs can be integrated into this approach by tightening CBF constraints using learning error bounds (Cheng et al., 2019; Jagtap et al., 2020). The effect of this tightening is frequently mitigated by updating the GP online, e.g., via regularly scheduled updates (time-triggered) (Brunke et al., 2022b; Zhang et al., 2025b) or initiated depending on performance specifications (event-triggered) (Dhiman et al., 2023; Wu et al., 2022). In particular, event-triggered GP learning has found wider adoption in control, with many works initiating model updates when Lyapunov-stability or tracking-performance conditions approach violation (Umlauft & Hirche, 2019; Jiao et al., 2022; Dai et al., 2023; Zhang et al., 2024). These works are restricted to state-dependent uncertainty, which significantly simplifies the online learning problem since model errors are independent of the applied control inputs. This limitation also holds for the majority of CBF-based event-triggered learning approaches relying on GPs (Wu et al., 2022; Xiao et al., 2023a; Zhang et al., 2025a; Wang et al., 2018).

While the integration of GPs into CBF-based safety filters is also possible for control-dependent model uncertainty (Long et al., 2022), the feasibility of the resulting optimization problem and thus all safety guarantees depend crucially on a sufficient model accuracy (Mestres & Cortés, 2024). This insight has inspired different methods for actively exploring the space of control inputs when the model accuracy falls below a threshold, so that feasibility is provably retained (Capone et al., 2025; Castaneda et al., 2025; Lederer et al., 2026a). As illustrated in Fig. 1, single point methods (Castaneda et al., 2025; Lederer et al., 2026a) only generate one exploratory control input along the most probable safe direction. This control over the GP variance in the selected control direction is exploited to establish recursive feasibility. However, uncertainty remains high in large sections of the input space with negative impact on conservatism and required frequency for triggering updates. Bandit exploration (Capone et al., 2025) generates a provably sample-efficient exploratory control sequence by executing a GP-UCB algorithm to identify control directions that maximize the CBF constraint. This approach can be shown to eventually recover feasibility after it is lost, but it ignores the CBF constraints during the execution of the exploratory control sequence, which can compromise safety.

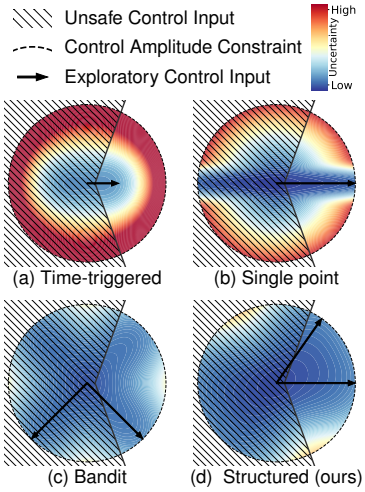


Figure 1: Only our structured exploration ensures a sufficient coverage of the control space to uniformly reduce uncertainty of the GP model of the CBF derivative (color) while avoiding unsafe control inputs (shaded area).

Contributions. We propose an event-triggered structured exploration approach for event-triggered learning of CBF-based safety filters using GPs that maintains CBF safety guarantees during exploration, ensures recursive feasibility, and is provably sample-efficient. The specific contributions are:

1. We propose a new control-sequence design for event-triggered learning that collects multidirectional information through structured exploration while maintaining safety guarantees.
2. We prove the recursive-feasibility ensured by this exploratory control sequence. The key for this proof is a new GP variance decomposition that uses observations along finitely many directions to obtain a uniform posterior-variance bound over the entire bounded control-input domain.
3. To the best of our knowledge, we provide the first sample-complexity bound for event-triggered learning with safe exploratory control signals.
4. We provide an implementable event-triggered learning algorithm and demonstrate, on two robotic-control benchmarks, its advantages over state-of-the-art event-triggered learning methods in terms of safety, conservatism, and model-update frequency. In a torque-controlled manipulator simulation, we demonstrate the capability of our approach to solely learn from training labels based on measured distances, but without global closed-form CBF knowledge.

The remainder of the paper is organized as follows. Section 2 formally introduces the problem of safely learning CBF-based safety filters. Our structured exploration approach for safe event-triggered learning is presented in Section 3. In Section 4, we prove its safety and sample efficiency, before we demonstrate its practical benefits in a numerical comparison with baselines in Section 5.

2 PROBLEM FORMULATION

We consider an unknown nonlinear control-affine system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{G}(\boldsymbol{x})\boldsymbol{u} \tag{1}$$

with system state $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n$ and control input $\boldsymbol{u} \in \mathcal{U} \subset \mathbb{R}^m$. Note that control-affine dynamics comprise many practically relevant robotic systems, e.g., quadcopters (Mahony et al., 2012), robot manipulators (Murray et al., 1994), and underwater vehicles (Fossen, 2011). We assume a unit norm bounded set of control inputs in the following, i.e., $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^m : \|\boldsymbol{u}\| \leq 1\}$. Moreover, we assume access to a nominal model $\dot{\boldsymbol{x}} = \widetilde{\boldsymbol{f}}(\boldsymbol{x}) + \widetilde{\boldsymbol{G}}(\boldsymbol{x})\boldsymbol{u}$ with locally Lipschitz functions $\widetilde{\boldsymbol{f}}$ and $\widetilde{\boldsymbol{G}}$ on the considered compact operating domain. This model may come from simplified physics, nominal parameters, or offline identification and need not coincide with the true dynamics.

Our primary objective is to ensure the safety of the system, which we define as forward invariance of a constraint set $\mathcal{C} = \{\boldsymbol{x} \in \mathcal{X} : h(\boldsymbol{x}) \geq 0\}$, where $h : \mathcal{X} \to \mathbb{R}$ is a continuously differentiable function. Forward invariance guarantees that the state trajectory $\boldsymbol{x}(t)$ will remain in $\mathcal{C}$ for all $t \geq 0$ if initialized within $\mathcal{C}$. If $h(\cdot)$ is a CBF, this form of safety can be straightforwardly ensured.

Lemma 1 (Control Barrier Function (Cohen et al., 2024)). *Consider system* (1) *and the set* $\mathcal{C}$*. Assume that there exists an extended class* $\mathcal{K}_\infty$ *function*[1] $\alpha(\cdot)$ *such that* $\max_{\|\boldsymbol{u}\|\leq 1} \phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})\boldsymbol{u} > 0,\ \forall \boldsymbol{x} \in \mathcal{C}$*, where* $\phi(\boldsymbol{x}) = \nabla h(\boldsymbol{x})^\top \boldsymbol{f}(\boldsymbol{x}) + \alpha(h(\boldsymbol{x}))$ *and* $\boldsymbol{\psi}^\top(\boldsymbol{x}) = \nabla h(\boldsymbol{x})^\top \boldsymbol{G}(\boldsymbol{x})$*. Then* $h(\cdot)$ *is a CBF, and any Lipschitz continuous control law satisfying* $\phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})\boldsymbol{u} \geq 0$ *renders* $\mathcal{C}$ *forward invariant.*

Given a CBF, a safety filter for a nominal policy $\pi_{\text{nom}} : \mathcal{X} \to \mathcal{U}$ is obtained as (Ames et al., 2017)

$$\boldsymbol{\pi}^*(\boldsymbol{x}) = \arg\min_{\boldsymbol{u}:\|\boldsymbol{u}\|\leq 1} \|\boldsymbol{\pi}_{\text{nom}}(\boldsymbol{x}) - \boldsymbol{u}\|^2 \qquad \text{s.t. } \phi(\boldsymbol{x}) + \boldsymbol{\psi}^T(\boldsymbol{x})\boldsymbol{u} \geq 0. \tag{2}$$

However, this safety filter requires precise knowledge of the system dynamics (1). While we could define the constraint in (2) using the nominal CBF derivative estimates $\tilde{\phi}(\boldsymbol{x}) = \nabla h(\boldsymbol{x})^\top \tilde{\boldsymbol{f}}(\boldsymbol{x}) + \alpha(h(\boldsymbol{x}))$ and $\tilde{\boldsymbol{\psi}}^\top(\boldsymbol{x}) = \nabla h(\boldsymbol{x})^\top \tilde{\boldsymbol{G}}(\boldsymbol{x})$, this can pose significant risks for safety (Taylor et al., 2020). Therefore, we require the ability to collect data for learning to compensate for mismatches.

Assumption 1 (CBF Data Collection). A CBF for system (1) is known, which induces a compact constraint set $\mathcal{C}$. Moreover, we can collect state-input pairs $\boldsymbol{z}_t = [\boldsymbol{x}^\top(t), \boldsymbol{u}^\top(t)]^\top$ and noisy estimates of the residual CBF derivative $y_t = \Delta\phi(\boldsymbol{x}(t)) + \Delta\boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t) + \omega_t$, where ω_t is i.i.d. R-sub-Gaussian noise, $\Delta\phi(\boldsymbol{x}) = \nabla h(\boldsymbol{x})^\top(\boldsymbol{f}(\boldsymbol{x}) - \tilde{\boldsymbol{f}}(\boldsymbol{x}))$ and $\Delta\boldsymbol{\psi}^\top(\boldsymbol{x}) = \nabla h(\boldsymbol{x})^\top(\boldsymbol{G}(\boldsymbol{x}) - \tilde{\boldsymbol{G}}(\boldsymbol{x}))$.

Noise-free state measurements and sub-Gaussian target noise are primarily imposed to streamline the theoretical analysis. In practice, the targets y_t can be obtained by numerically differentiating sequential CBF evaluations, while small, bounded state perturbations can be accommodated through robustness margins in the CBF conditions and error bounds (Cosner et al., 2021; Lederer, 2023). Moreover, Assumption 1 only requires point-wise CBF evaluations rather than a globally known closed-form CBF, thereby admitting, e.g., distance-based CBF representations. A detailed discussion of these practical implications and possible theoretical extensions is provided in Section A.

When learning the model mismatches $\Delta\phi(\cdot)$ and $\Delta\psi_i(\cdot)$ from data, regression errors are unavoidable in general. Therefore, we impose the following kernel-based regularity requirement on these functions to allow the quantification of these learning errors.

Assumption 2. Let $k_\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and $k_{\psi_i} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, be stationary kernels with bounded Hessian traces, i.e., $\operatorname{trace}(\nabla_{\boldsymbol{x}} \nabla_{\boldsymbol{x}'}^\top k_\phi(\boldsymbol{x}, \boldsymbol{x}')\big|_{\boldsymbol{x}=\boldsymbol{x}'}) \leq L_{k_\phi}$ and $\operatorname{trace}(\nabla_{\boldsymbol{x}} \nabla_{\boldsymbol{x}'}^\top k_{\psi_i}(\boldsymbol{x}, \boldsymbol{x}')\big|_{\boldsymbol{x}=\boldsymbol{x}'}) \leq L_{k_{\psi_i}}$. The unknown functions $\Delta\phi(\cdot)$ and $\Delta\psi_i(\cdot)$ have a bounded norm in the reproducing kernel Hilbert spaces (RKHSs) defined through their respective kernels $k_\phi(\cdot,\cdot)$ and $k_{\psi_i}(\cdot,\cdot)$, i.e., $\|\Delta\phi(\cdot)\|_{k_\phi} \leq B_\phi$ and $\|\Delta\psi_i(\cdot)\|_{k_{\psi_i}} \leq B_{\psi_i}$, $i = 1, \ldots, m$.

The smoothness requirement in Assumption 2 is satisfied by commonly used kernels, including squared exponential kernels, while bounded RKHS norms are standard assumptions for controlling function complexity in kernel-based error bounds (Srinivas et al., 2012; Chowdhury & Gopalan, 2017). However, evaluating these bounds requires numerical values for B_ϕ and B_{ψ_i}, which can be

[1] An extended class-$\mathcal{K}_\infty$ function is a continuous and strictly increasing function $\alpha : \mathbb{R} \to \mathbb{R}$ satisfying $\alpha(0) = 0$, $\lim_{r\to\infty} \alpha(r) = \infty$, and $\lim_{r\to-\infty} \alpha(r) = -\infty$.

difficult to specify, leading to controversy about this assumption Fiedler et al. (2024). A detailed discussion of this requirement and possible mitigation strategies is provided in Section A.2.

To ensure that the safety constraint in (2) remains feasible despite learning errors, we require the CBF to admit a safety margin as specified in the following.

Assumption 3 (Robust Feasibility)**.** There exist known constants $\theta > 0$ and $\epsilon \in (0, 1)$ such that for all $\boldsymbol{x} \in \mathcal{C}$, there exists a control input $\boldsymbol{u}$ with $\|\boldsymbol{u}\| \leq 1 - \epsilon$ satisfying $\phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})\boldsymbol{u} \geq \theta$.

This assumption can be frequently found in robust CBF literature (Capone et al., 2025; Taylor et al., 2020; Xu et al., 2015). It merely requires that the system possesses sufficient control authority, such that the CBF constraint in (2) can be tightened to robustly handle uncertainties. Notably, since the robustness margin θ can be arbitrarily small, this requirement is generally not restrictive.

Problem Setting: Based on these assumptions, we seek a safe and sample-efficient event-triggered exploration strategy that activates before model uncertainty makes the learned safety filter infeasible and collects informative data across the control-input space. The collected data should maintain recursive feasibility of computationally efficient min-norm safety filters akin to (2), while reducing their conservatism and limiting the number of exploration events.

3 Structured Exploration for Gaussian Process Safety Filters

We address this problem by combining GP learning of the CBF constraints (Castaneda et al., 2025) with an event-triggered structured exploration sequence. Exploration is activated before excessive model uncertainty makes the safety filter infeasible, and the resulting sequence safely samples multiple control directions to reduce uncertainty across the input space. In Section 3.1, we introduce GP regression and explain how the CBF constraints are learned. The event-triggered structured exploration algorithm is presented in Section 3.2.

3.1 Safety Filters with Gaussian Process Control Barrier Constraints

We employ GP regression to estimate the unknown residual $\Delta H(\boldsymbol{x}, \boldsymbol{u}) = \Delta\phi(\boldsymbol{x}(t)) + \Delta\boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t)$ from a dataset $\mathbb{D}_N = \{(\boldsymbol{z}^{(i)}, y^{(i)})\}_{i=1}^N$. A GP is a collection of random variables, any finite number of which follow a joint Gaussian distribution. Similar to the Gaussian distribution, a GP is fully specified by a prior mean function, which we subsequently set to 0 without loss of generality, and a covariance function, also referred to as a kernel. To exploit the control-affine structure of $\Delta H(\boldsymbol{x}, \boldsymbol{u})$, we place independent GP priors on the components, i.e., $\Delta\phi(\cdot) \sim \mathcal{GP}(0, k_\phi)$ and $\Delta\psi_i(\cdot) \sim \mathcal{GP}(0, k_{\psi_i})$. This induces a composite kernel $k(\boldsymbol{z}, \boldsymbol{z}') = k_\phi(\boldsymbol{x}, \boldsymbol{x}') + \sum_{i=1}^m u_i k_{\psi_i}(\boldsymbol{x}, \boldsymbol{x}')u_i'$, where $\boldsymbol{z} = [\boldsymbol{x}^\top, \boldsymbol{u}^\top]^\top$. The posterior mean $\mu(\boldsymbol{z})$ and variance $\sigma^2(\boldsymbol{z})$ for this composite GP take the form (Castaneda et al., 2025; Umlauft & Hirche, 2019)

$$\mu(\boldsymbol{x}, \boldsymbol{u}) = \Delta\tilde{\phi}(\boldsymbol{x}) + \Delta\tilde{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{u}, \quad \sigma^2(\boldsymbol{x}, \boldsymbol{u}) = \begin{bmatrix}1 & \boldsymbol{u}^\top\end{bmatrix} \boldsymbol{\Sigma}(\boldsymbol{x}) \begin{bmatrix}1 & \boldsymbol{u}^\top\end{bmatrix}^\top, \tag{3}$$

where the nominal mean components are computed as $\Delta\tilde{\phi}(\boldsymbol{x}) = \boldsymbol{k}_\phi^\top(\boldsymbol{x})(\boldsymbol{K} + \sigma_{\text{on}}^2\boldsymbol{I})^{-1}\boldsymbol{y}$, $\Delta\tilde{\psi}_i(\boldsymbol{x}) = \boldsymbol{k}_{\psi_i}^\top(\boldsymbol{x})\boldsymbol{U}_i(\boldsymbol{K} + \sigma_{\text{on}}^2\boldsymbol{I})^{-1}\boldsymbol{y}$, and the positive definite covariance matrix is given by

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = \begin{bmatrix} k_\phi(\boldsymbol{x}, \boldsymbol{x}) & 0 & \cdots & 0 \\ 0 & k_{\psi_1}(\boldsymbol{x}, \boldsymbol{x}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_{\psi_m}(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} - \begin{bmatrix} \boldsymbol{k}_\phi^\top(\boldsymbol{x}) \\ \boldsymbol{k}_{\psi_1}^\top(\boldsymbol{x})\boldsymbol{U}_1 \\ \vdots \\ \boldsymbol{k}_{\psi_m}^\top(\boldsymbol{x})\boldsymbol{U}_m \end{bmatrix} (\boldsymbol{K} + \sigma_{\text{on}}^2\boldsymbol{I})^{-1} \begin{bmatrix} \boldsymbol{k}_\phi^\top(\boldsymbol{x}) \\ \boldsymbol{k}_{\psi_1}^\top(\boldsymbol{x})\boldsymbol{U}_1 \\ \vdots \\ \boldsymbol{k}_{\psi_m}^\top(\boldsymbol{x})\boldsymbol{U}_m \end{bmatrix}^\top.$$

Here, $\boldsymbol{U}_i = \text{diag}(u_i^{(1)}, \ldots, u_i^{(N)})$ are diagonal matrices composed of the i-th element of the control samples. Moreover, $\boldsymbol{k}_\phi(\boldsymbol{x})$ is defined as $[\boldsymbol{k}_\phi(\boldsymbol{x})]_i = k_\phi(\boldsymbol{x}, \boldsymbol{x}^{(i)})$ with an analogous definition of $\boldsymbol{k}_{\psi_i}(\boldsymbol{x})$, $\boldsymbol{K}$ is defined via $[\boldsymbol{K}]_{ij} = k(\boldsymbol{z}^{(i)}, \boldsymbol{z}^{(j)})$, and the aggregated target vector is $[\boldsymbol{y}]_i = y^{(i)}$.

Since the GP mean $\mu(\boldsymbol{x}, \boldsymbol{u})$ is merely a learned estimate of the true CBF constraint in (2), we tighten it using a confidence bound to account for model uncertainty. This leads to the robust safety filter

$$\boldsymbol{\pi}^*(\boldsymbol{x}) = \arg\min_{\|\boldsymbol{u}\| \leq 1} \|\boldsymbol{u} - \boldsymbol{\pi}_{\text{nom}}(\boldsymbol{x})\|^2 \tag{4a}$$

$$\text{s.t. } \hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{u} - \beta\left\| \begin{bmatrix}1 & \boldsymbol{u}^\top\end{bmatrix} \boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{x})\right\|_2 - \eta_1 \geq 0 \tag{4b}$$

with $\hat{\phi}(\boldsymbol{x}) = \tilde{\phi}(\boldsymbol{x}) + \Delta\tilde{\phi}(\boldsymbol{x})$, $\hat{\boldsymbol{\psi}}(\boldsymbol{x}) = \tilde{\boldsymbol{\psi}}(\boldsymbol{x}) + \Delta\tilde{\boldsymbol{\psi}}(\boldsymbol{x})$, scaling factor $\beta \in \mathbb{R}_{>0}$ and constant constraint tightening $\eta_1 \in \mathbb{R}_{\geq 0}$. Castaneda et al. (2025) show that this optimization problem can be reformulated in to a SOCP, which is convex and can be efficiently solved numerically (Boyd & Vandenberghe, 2004). However, unlike the QP (2) for known dynamics, GP-CBF-SOCPs (4) generally lack feasibility guarantees (Mestres & Cortés, 2024). Specifically, if the model uncertainty encoded in $\boldsymbol{\Sigma}(\boldsymbol{x})$ is large relative to the system's control authority, the feasible control space becomes empty.

3.2 Event-Triggered Learning with Structured Exploration

To avoid potentially infeasible GP-CBF-SOCPs (4), we augment them using event-triggered online learning with structured exploration as outlined in Algorithm 1. For triggering model updates, we first compute a nominally safe control input under the GP model via[2]

$$\hat{\boldsymbol{u}}^* = \arg\max_{\|\boldsymbol{u}\|=1-\epsilon} \hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{u}. \tag{5}$$

such that we can assess the robustness margin of the CBF constraint. Based on this control input, we define the event trigger as

$$t_{\kappa+1} = \arg\inf_{t \geq t_\kappa + mT_s} t \tag{6a}$$

$$\text{such that } 3\beta \max_{\|\boldsymbol{u}\|\leq 1} \sigma(\boldsymbol{x}(t), \boldsymbol{u}) \geq \hat{\phi}(\boldsymbol{x}(t)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t))\hat{\boldsymbol{u}}^* - \eta_2, \tag{6b}$$

i.e., we start an exploratory control sequence when the robustness margin of the nominal CBF constraint $\hat{\phi}(\boldsymbol{x}(t)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t))\boldsymbol{u}^*$ is small compared to the scaled maximum of the GP standard deviation $3\beta \max_{\|\boldsymbol{u}\|\leq 1} \sigma(\boldsymbol{x}_t, \boldsymbol{u})$. The trigger condition (6b) serves two purposes. On the one hand, it can be straightforwardly seen that the feasibility of the GP-CBF-SOCP (4) is guaranteed if the trigger condition (6b) is not satisfied. Indeed, we require additional conservatism via a factor 3 to ensure feasibility without a GP update at the next sampling time in our theoretical analysis. Note that $\eta_2 \in \mathbb{R}_{>0}$ ensures a robustness margin for feasibility, which can be exploited by the exploratory sequence. On the other hand, (6b) ensures that model uncertainty cannot grow excessively in any control direction due to the consideration of the maximum GP standard deviation $\max_{\|\boldsymbol{u}\|\leq 1} \sigma(\boldsymbol{x}_t, \boldsymbol{u})$. Thus, the exploratory sequence is also triggered to reduce potential conservatism of the GP-CBF constraint (4b).

Algorithm 1: Safe Event-Triggered Online Learning with Structured Exploration

Input: Sampling time T_s, tightening constants η_1, η_2, exploration magnitude ϵ, scaling factor β
Initialize: $t = 0$, $i = m + 1$
while true do
  $\boldsymbol{x}_t \leftarrow \boldsymbol{x}(t)$
  Compute $\hat{\boldsymbol{u}}^*$ using (5)
  if $i = m + 1$ **and** (6b) **then**
    Construct orthonormal basis: $\{\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon}, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$
    Define exploratory control sequence: $\boldsymbol{u}_1 = \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon}, \boldsymbol{u}_2 = \hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_1, \ldots, \boldsymbol{u}_m = \hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_{m-1}$
    Set $i \leftarrow 1$
  if $i \leq m$ **then**
    Apply $\boldsymbol{u}_t = \boldsymbol{u}_i$
    Record $y_t = \Delta\phi(\boldsymbol{x}_t) + \Delta\boldsymbol{\psi}^\top(\boldsymbol{x}_t)\boldsymbol{u}_t + \omega_t$
    $\mathbb{D}_{N+1} \leftarrow \mathbb{D}_N \cup \{(\boldsymbol{z}_t, y_t)\}$, $N \leftarrow N+1$, $i \leftarrow i+1$
    Update GP posterior $\mu(\cdot)$ and $\sigma(\cdot)$ using (3)
  else
    Compute and apply $\boldsymbol{\pi}(\boldsymbol{x}_t)$ by solving (4)
  $t \leftarrow t + T_s$

When an update event is triggered, an exploratory control sequence is constructed by perturbing the control input (5) in all possible orthogonal directions. We obtain these perturbations by constructing an orthonormal basis $\{\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon}, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$ with $\boldsymbol{d}_i^\top \hat{\boldsymbol{u}}^* = 0$, $\boldsymbol{d}_i^\top \boldsymbol{d}_j = 0$ for all $i, j = 1 \ldots, m-1$, $i \neq j$ and $\|\boldsymbol{d}_i\| = 1$. Subsequently, we add the scaled orthogonal vectors $\boldsymbol{d}_i$ to the anchor control direction $\hat{\boldsymbol{u}}^*$, which results in the exploratory control sequence $\boldsymbol{u}_1 = \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon}, \boldsymbol{u}_2 = \hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_1, \ldots, \boldsymbol{u}_m = \hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_{m-1}$. Finally, the computed control sequence is sequentially applied, each input generating one training sample used for updating the GP model. In combination with the control input obtained from the GP-CBF-SOCP (4), this results in the time-varying control signal

$$\boldsymbol{u}(t) = \begin{cases} \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} & \text{if } t \in [t_\kappa, t_\kappa + T_s) \\ \hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i & \text{if } t \in [t_\kappa + iT_s, t_\kappa + (i+1)T_s) \\ \boldsymbol{\pi}^*(\boldsymbol{x}(t_k)) & \text{if } t_k \neq t_\kappa + iT_s \text{ for } i = 0, \ldots, m-1 \wedge t \in [t_k, t_k + T_s). \end{cases} \tag{7}$$

[2] If no unique maximum exists ($\hat{\boldsymbol{\psi}}(\boldsymbol{x}) = \boldsymbol{0}$), we assume the $\arg\inf$ operator to return one of the local optima.

These m exploratory inputs span the entire control-input space, which we utilize to transform the variance reduction in the exploration directions into an uncertainty bound on any admissible control direction. A larger ϵ makes this sequence closer to orthogonal exploration, while a smaller ϵ makes each input closer to the safe anchor direction, thereby preserving a larger safety margin. Practical guidance for selecting the parameters β, ϵ, η_1, η_2, and T_s is provided in Section B, while the use of scalable GP implementations to reduce computational complexity are discussed in Section C.

4 Theoretical Guarantees

In this section, we establish the theoretical guarantees of the proposed event-triggered active exploration framework. In Section 4.1, we show that our algorithm ensures safety with high probability. We derive an upper bound on the number of triggered events and consequently the sample complexity in Section 4.2. Formal proofs for all results are deferred to Section D.

4.1 Closed-Loop Safety and Recursive Feasibility

Since the GP-CBF-SOCP safety filter (4) guarantees safety if it is feasible (Castaneda et al., 2025), proving the safety of Algorithm 1 mainly requires an analysis of the proposed event-triggered structured exploration approach. In particular, we show that the orthogonal perturbations inherit safety from their anchor direction $\hat{\boldsymbol{u}}^*$ under Assumption 3. Moreover, we derive a novel bound of the posterior GP variance $\sigma^2(\boldsymbol{x}, \boldsymbol{u})$ for arbitrary control inputs $\boldsymbol{u}$ based on GP variances in the directions of the orthogonal perturbations (c.f. Lemma 6). Since the variance after GP updates can be bounded by a constant $\underline{\sigma}^2 \in \mathbb{R}_{\geq 0}$ depending on the GP hyperparameters (Lederer et al., 2026c), this establishes recursive feasibility using the updates and consequently the safety of Algorithm 1.

Theorem 1 (Probabilistic Safety)**.** *Consider the continuous-time system* (1) *and a continuously differentiable CBF* $h(\boldsymbol{x})$*. Suppose Assumptions 1, 2 and 3 are satisfied so that*

$$\Gamma = \theta - 2\beta\sqrt{(m+1)^3\sigma_{\phi,0}^2 + \tfrac{2(m+1)^3}{\epsilon^2}\underline{\sigma}^2} > 0. \tag{8}$$

Furthermore, assume that the GP mean and variance defined in (3) *are based on differentiable stationary kernels* $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$*. If at most* $\bar{N} \in \mathbb{N}$ *training samples are generated and* (6b) *does not hold at* $t = 0$*, there exist thresholds* $\Xi_1(\bar{N}, \delta), \Xi_2(\Gamma, \beta, \epsilon), \Xi_3(T_s), \Xi_4(T_s, \beta) \in \mathbb{R}_{\geq 0}$*, such that the control signal* (7) *and the event trigger* (6) *resulting from Algorithm 1 with parameters*

$$\beta \geq \Xi_1(\bar{N}, \delta), \qquad \sqrt{T_s} \leq \Xi_2(\Gamma, \beta, \epsilon), \qquad \eta_1 \geq \Xi_3(T_s) \qquad \eta_2 \geq \Xi_4(T_s, \beta) \tag{9}$$

guarantees the safety of the closed-loop system, i.e., $\boldsymbol{x}(t) \in \mathcal{C}$ *for all* $t \geq 0$*, with probability* $1 - \delta$*.*

Despite its seeming complexity, this result allows simple interpretations. The lower bound on β in (9) is necessary for the GP error bound to hold with probability $1 - \delta$, while η_1 and η_2 are lower bounded to provide sufficient robustness for the discrete-time implementation of the safety filter derived using continuous-time principles. Condition (8) ensures a feasibility margin if all exploratory control inputs could be applied at the same time. This margin subsequently imposes an upper bound on the sampling time in (9) since we can only apply one input at a time in practice.

4.2 Sample Complexity of Event-Triggered Learning

To bound the number of triggered exploration sequences in Algorithm 1, we exploit sample complexity bounds for GP regression (Srinivas et al., 2012). By adapting them to control-affine kernels and lower bounding the right side of the triggering condition (6b), we obtain the following result.

Theorem 2 (Sample Complexity)**.** *Consider the continuous-time system* (1) *and a continuously differentiable CBF* $h(\boldsymbol{x})$*. Suppose Assumptions 1, 2 and 3 are satisfied so that* (8) *holds. Furthermore, assume that the GP mean and variance defined in* (3) *are based on differentiable stationary kernels* $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$*. If Algorithm 1 is initialized with* $\mathbb{D}_0 = \emptyset$*, there exist constants* $\Xi_1(\bar{\kappa}m, \delta), \bar{\Xi}_2(\Gamma, \beta, \epsilon, \eta_2, \theta), \Xi_3(T_s), \Xi_4(T_s, \beta), \tilde{\Xi}_4(\beta, \epsilon, \theta) \in \mathbb{R}_{\geq 0}$*, such that the control signal* (7) *and the event trigger* (6) *resulting from Algorithm 1 with parameters satisfying*

$$\beta \geq \Xi_1(\bar{\kappa}m, \delta), \quad \sqrt{T_s} \leq \bar{\Xi}_2(\Gamma, \beta, \epsilon, \eta_2, \theta), \quad \eta_1 \geq \Xi_3(T_s), \quad \Xi_4(T_s, \beta) \leq \eta_2 \leq \tilde{\Xi}_4(\beta, \epsilon, \theta) \tag{10}$$

guarantee an upper bound on the number of trigger events

$$\bar{\kappa} = \min_{\kappa \in \mathbb{N}} \kappa \qquad \textit{such that } \kappa \geq 2c(\theta - \eta_2, \epsilon)\beta^2 \gamma_{\kappa m}, \tag{11}$$

where $\gamma_{\kappa m}$ *is the maximum information gain* $\gamma_{\kappa m} = \max_{\mathbb{D}:|\mathbb{D}|\leq \kappa m} \log\det\left(\boldsymbol{I} + \sigma_{\text{on}}^{-2}\boldsymbol{K}\right)$ *and* $c(\theta - \eta_2, \epsilon) \in \mathbb{R}_{\geq 0}$ *is a constant.*

The bound in (11) is finite because its left-hand side grows linearly in κ, whereas the maximum information gain $\gamma_{\kappa m}$ typically grows sublinearly. For squared exponential kernels, this observation also gives an explicit dependence of the trigger bound on the exploration parameters. The formal result and its proof are provided in Corollary 1.

5 Numerical Experiments

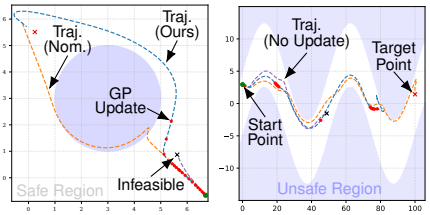
To illustrate the practical effectiveness of our approach for safely learning GP safety filters using event-triggered structured exploration, we compare Algorithm 1 to state-of-the-art baselines in two practical robot control benchmarks. Section 5.1 demonstrates the reduced conservatism despite fewer model updates and strict safety guarantees achieved by our approach in comparison to the baselines. In Section 5.2, we illustrate that our approach robustly achieves this performance for a range of parameter choices. Finally, we demonstrate that our approach can be straightforwardly extended to scenarios without explicit CBF expressions using distance measurements in Section 5.3.

5.1 Comparison with State-of-the-Art Baselines

We evaluate our approach on two robot control benchmarks: a quadrotor navigating an open three-dimensional space adapted from Capone et al. (2025) and an autonomous underwater vehicle (AUV) traversing a narrow tunnel based on standard marine-craft dynamics (Fossen, 2011). Fig. 2 illustrates these benchmarks along with exemplary trajectories that highlight the need for online learning to ensure the safety of the robots. Detailed settings are reported in Appendix E.

We compare against three state-of-the-art baselines in GP-CBF-based safe online learning:

- The on-the-fly **Bandit** exploration approach (Capone et al., 2025) as an instance for event-triggered learning with an exploratory control sequence. This approach formulates exploration as an event-triggered unconstrained bandit optimization problem.
- The **Single Point** exploration method proposed in Lederer et al. (2026a) as an example for an event-triggered learning baseline with a single exploratory control input. The control inputs used by this method maximize the probability of safety under the current model knowledge.
- A **Time-Triggered** learning strategy as example for online learning without active exploration (Zhang et al., 2025b). The GP model is updated at regular times using the GP-SOCP control inputs.



(a) Quadrotor. (b) AUV.

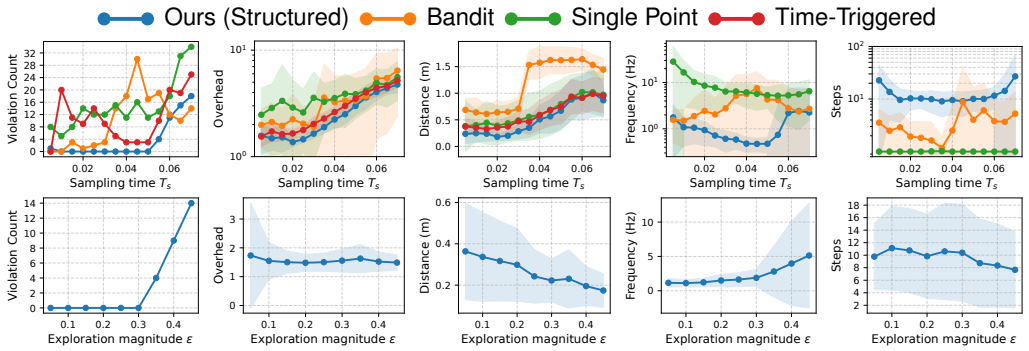
Figure 2: Safe robot control benchmarks with example trajectories. Using only the GP mean leads to constraint violations, while the absence of online learning causes infeasible optimization problems, illustrating the importance of safe online learning.

All methods use the same nominal controller, CBF formulation, and initial data to ensure a fair comparison. We run 100 randomized trials and evaluate safety (number of trajectories with CBF violations and infeasible SOCPs), conservatism (path overhead ratio and minimum distance to the obstacle), and sample efficiency (trigger frequency and minimum interval between triggered updates).

Table 1 shows that our structured exploration approach for safe event-triggered learning consistently achieves zero CBF violations and ensures the feasibility of the GP-CBF-SOCP on both benchmarks, matching the reliability of Bandit exploration and outperforming the Single Point Strategy and Time-Triggered learning. The failures of the Time-Triggered approach indicate that fixed-rate updates alone cannot maintain feasibility under high uncertainty, while the violations of the Single Point approach demonstrate that exploring only along $\hat{\boldsymbol{u}}^*$ can require impractically high sampling rates to achieve theoretical safety guarantees. Among the methods with zero CBF violations, our approach is significantly less conservative. It achieves lower path overhead than Bandit exploration

Table 1: Performance on the quadrotor and AUV benchmarks. Lower values indicate better performance for all metrics apart from the minimum interval between triggered updates. Our method reduces conservatism while improving sample efficiency and maintaining safety. Single Point exploration achieves lower distances in the AUV benchmark only at the cost of safety violations. See Section E.3 for a discussion of the minimum trigger intervals of our approach in the AUV benchmark.

Metric	Quadrotor				AUV			
	Ours	**Bandit**	**Single Point**	**Time-Triggered**	**Ours**	**Bandit**	**Single Point**	**Time-Triggered**
CBF Violations	**0**	**0**	3	17	**0**	**0**	32	5
Infeasibilities	**0**	**0**	**0**	26	**0**	**0**	**0**	28
Path Overhead Ratio	$\mathbf{1.52 \pm 0.61}$	1.78 ± 0.92	2.70 ± 2.35	1.95 ± 0.32	$\mathbf{1.25 \pm 0.24}$	2.40 ± 1.32	2.11 ± 0.91	1.34 ± 0.13
Min Dist (m)	$\mathbf{0.26 \pm 0.16}$	0.62 ± 0.21	0.36 ± 0.31	0.32 ± 0.14	1.32 ± 0.42	3.31 ± 0.14	$\mathbf{0.632 \pm 0.23}$	1.65 ± 0.33
Trigger Freq. (Hz)	$\mathbf{1.22 \pm 0.70}$	1.27 ± 0.83	13.81 ± 10.73	1.00 ± 0.00	$\mathbf{0.043 \pm 0.035}$	0.265 ± 0.21	15.40 ± 7.40	0.050 ± 0.00
Min Trigger Interval (steps)	$\mathbf{9.3 \pm 9.1}$	3.1 ± 2.7	1.0 ± 0.0	—	$\mathbf{501.4 \pm 972.4}$	1.2 ± 1.5	1.0 ± 0.0	—



(a) CBF Viol. (b) Overhead (c) Min Dist. (d) Trigger Freq. (e) Min Interval

Figure 3: Ablation study on the sampling time T_s and exploration magnitude ϵ. Our approach is relatively insensitive to parameter changes, except when theoretical requirements become violated.

on both the quadrotor and AUV benchmarks, and also allows smaller minimum distances to the unsafe regions without losing safety. This is consistent with our structured exploration design: Bandit exploration terminates once feasibility is recovered in any direction, whereas our fixed orthogonal exploration sequence reduces uncertainty in all control directions, allowing the safety filter to use less conservative actions. This effect is also impacting the triggering behavior: Compared with Bandit exploration, our approach increases minimum trigger intervals on the quadrotor and substantially lowers the average triggering frequency on the confined AUV task. Thus, these results clearly demonstrate the benefits of the proposed event-triggered structured exploration on sample efficiency while preserving safety and reducing conservatism.

5.2 Robustness Against Parameter Choices

We further evaluate the sensitivity of our approach to different parameter choices in the quadrotor benchmark. We exemplarily present the results for the sampling time T_s and the exploration magnitude ϵ here, while results for other parameters can be found in Appendix F. When possible, we compare these dependencies against the baselines from the previous subsection.

The first row of Fig. 3 shows the effect of the sampling time T_s. For small and moderate values of T_s, our method maintains zero CBF violations while triggering less frequently than Single Point exploration and exhibiting a considerably weaker conservatism increase than Bandit exploration. For large T_s, violations eventually also appear for our method, which can be attributed to a violation of condition (9) in Theorem 1. Note that the occurrence of safety violations happens at considerably larger sampling times than for Bandit exploration. The exploration magnitude ϵ has only a minor effect on most metrics over a broad range, as illustrated in the second row of Fig. 3. The main visible trend is a gradual reduction of the minimum distance to obstacles, indicating slightly less conservative behavior due to the improved exploration. This improved exploration however comes at the cost of applying control inputs that are closer to and eventually violate the CBF conditions, which explains the increasing triggering frequency as well as CBF violations for large ϵ. Note that this effect can be easily explained using our theory via a violation of Assumption 3 that occurs when ϵ is chosen too large. Therefore, these results highlight that our approach for event-triggered learning

GP-CBF safety filters using structured exploration is relatively insensitive to parameter choices, with considerable deterioration only with apparent violations of assumptions in the theoretical analysis.

5.3 Learning with Point-Wise CBF Evaluations

We further consider randomized obstacle avoidance with a seven-degree-of-freedom Panda manipulator under joint-torque control. To enable safe learning with unknown obstacle geometry, the safety filter relies solely on local distance measurements without the need for a global closed-form CBF expression. Following existing distance-based CBF constructions (Long et al., 2021; Singletary et al., 2022; De Sa et al., 2024; Begzadić et al., 2025), numerical differentiation via point-wise CBF evaluations yields $\ddot{h}$ labels paired with state–input data for GP training. The GP thereby learns the derivative of a second-order control barrier function (Xiao & Belta, 2022). Figure 4 illustrates an executed end-effector trajectory around the obstacle and the locations of online GP updates. Implementation details are provided in Appendix G.

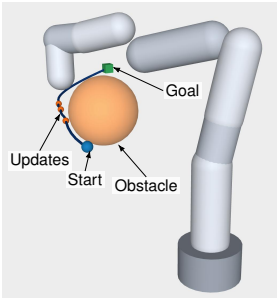


Figure 4: Panda obstacle avoidance. The curve shows the end-effector trajectory; markers indicate GP update locations.

Table 2 depicts the comparison of this variant requiring only distance measurements to the base algorithm evaluated in Section 5.1, which exploits closed-form CBF knowledge for generating training targets. Both approaches complete the same 100 randomized tasks without safety violations or GP-CBF-SOCP infeasibility. Compared with residual learning using closed-form CBF expressions, learning the complete derivative from point-wise CBF evaluations via distance measurements is only slightly more conservative: mean path overhead increases by approximately 4.0%, and mean minimum clearance increases by 1.0 cm. Meanwhile, mean trigger frequency increases by a factor of 2.13. These observations are consistent with the greater learning burden of identifying the complete CBF derivative rather than a residual error. Crucially, they illustrate that our event-triggered structured exploration mechanism natively supports safe online learning in scenarios without prior information about the obstacle geometry by replacing closed-form CBF expressions with point-wise CBF evaluations determined from distance measurements.

6 Conclusion and Open Problems

We proposed an event-triggered structured exploration framework that enables the safe learning of CBF-based safety filters using GPs. Our method safely collects informative measurements along a structured exploration sequence, such that the infeasibility of the learned safety filter is prevented. We establish probabilistic safety and finite-trigger guarantees, formally showing that the proposed exploration mechanism in combination with GP-CBF-SOCPs ensures safety with high probability while requiring a finite number of updates. Experiments on quadrotor and AUV benchmarks demonstrate that the method improves safety, increases data efficiency, and achieves less conservative behavior than existing exploration strategies; an additional distance-based experiment further indicates the applicability even when a globally valid closed-form CBF expression is not available.

Table 2: Comparison of our approach with global closed-form CBF knowledge and point-wise CBF evaluations via distance measurements on the Panda obstacle avoidance problem. Both settings complete all tasks without safety violations, SOCP infeasibility.

Metric	Closed-Form CBF	Point-Wise CBF Eval.
Path Overhead	1.127 ± 0.028	1.172 ± 0.055
Min Dist. (cm)	2.6 ± 0.5	3.6 ± 1.0
Trigger Freq.	0.186 ± 0.103	0.396 ± 0.175

Limitations and Open Problems. While an explicit CBF expression is not required for our method, the presented variant is restricted to certain system classes and constraints. A combination of our approach with an iterative application of techniques for learning CBFs (Dawson et al., 2023; Qin et al., 2021) appears to be a promising research direction to overcome this limitation. A particularly relevant problem when dealing with learned CBFs is the consideration and quantification of CBF uncertainties in the analysis. Requiring additional robustness of the CBF constraint could be a useful way to address such challenges (Mestres & Cortés, 2024). To reduce the cost for exploring all control input directions, a prioritized execution of directions that minimize the upper bound on the variance combined with a stopping condition based on the satisfaction of the desired variance bound can open a promising avenue. We believe that exploring these research directions can extend the applicability of GP-CBF-safety filters to robotic systems beyond the types considered in our experiments.

AI USE STATEMENT

We used generative AI tools solely for translation, language editing, and LaTeX formatting of the manuscript. The authors reviewed all AI-assisted revisions to verify that they preserved the intended meaning, technical content, and mathematical notation. The authors take full responsibility for the final content of the paper.

ETHICS STATEMENT

This work investigates safety filters for autonomous systems under model uncertainty. All reported experiments are conducted in simulation. Although the method aims to improve operational safety, its theoretical guarantees depend on the stated assumptions and parameter conditions; the empirical results do not establish safety for arbitrary physical deployments. Violations of these conditions could lead to unsafe behavior and physical harm. Deployment on real systems would require application-specific validation, assessment of sensing, actuation, and computational limitations, and independent protective measures. The study does not involve human participants or personal data.

REPRODUCIBILITY STATEMENT

The problem formulation and assumptions are stated in Section 2, and the method is described in Section 3 and Algorithm 1. Proofs of the theoretical results are provided in Appendix D, with practical parameter-selection guidance in Appendix B. The quadrotor and AUV simulation settings, implementation details, hyperparameters, evaluation metrics, and computing environment are documented in Appendix E. The Panda simulation model, control and learning settings, and randomized-task generation and evaluation protocol are specified in Appendix G. We plan to release the source code under an open-source license upon acceptance of the paper.

REFERENCES

WolframAlpha. https://www.wolframalpha.com/input?i=solve+b%5Clog%28x%29%3Dx%5E%281%2Fa%29+for+x. Accessed: 2026-04-16.

Aaron D. Ames, Xiangru Xu, Jessy W. Grizzle, and Paulo Tabuada. Control Barrier Function Based Quadratic Programs for Safety Critical Systems. *IEEE Transactions on Automatic Control*, 62(8): 3861–3876, 2017.

Yarden As, Bhavya, Lenart Treven, Carmelo Sferrazza, Stelian Coros, and Andreas Krause. Actsafe: Active exploration with safety constraints for reinforcement learning. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 29752–29776, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/495b0bb32959448b8b5b21ae9c0e9e34-Paper-Conference.pdf.

Azra Begzadić, Armin Lederer, Jorge Cortés, and Sylvia Herbert. Learning high-order CBFs using Gaussian processes for systems in Brunovskỳ canonical form. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pp. 2945–2950. IEEE, 2025.

Felix Berkenkamp, Angela P. Schoellig, and Andreas Krause. Safe controller optimization for quadrotors with Gaussian processes. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 491–496, 2016. doi: 10.1109/ICRA.2016.7487170.

Felix Berkenkamp, Matteo Turchetta, Angela P. Schoellig, and Andreas Krause. Safe Model-based Reinforcement Learning with Stability Guarantees. In *Advances in Neural Information Processing Systems*, pp. 908–918, 2017.

Felix Berkenkamp, Andreas Krause, and Angela P Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *Machine Learning*, 112(10):3713–3747, 2023.

Franco Blanchini and Stefano Miani. *Set-Theoretic Methods in Control*, volume 78. Springer, 2008.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Lukas Brunke, Melissa Greeff, Adam W Hall, Zhaocong Yuan, Siqi Zhou, Jacopo Panerati, and Angela P Schoellig. Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(1):411–444, 2022a.

Lukas Brunke, Siqi Zhou, and Angela P Schoellig. Barrier bayesian linear regression: Online learning of control barrier conditions for safety-critical control of uncertain systems. In *Learning for Dynamics and Control Conference*, pp. 881–892. PMLR, 2022b.

Alexandre Capone, Ryan Kazuo Cosner, Aaron Ames, and Sandra Hirche. Learning safe control via on-the-fly bandit exploration. In *International Conference on Machine Learning*, pp. 6726–6745. PMLR, 2025.

Fernando Castaneda, Jason J Choi, Wonsuhk Jung, Bike Zhang, Claire J Tomlin, and Koushil Sreenath. Recursively feasible probabilistic safe online learning with control barrier functions. *IEEE Open Journal of Control Systems*, 2025.

Ioannis Chatzigeorgiou. Bounds on the lambert function and their application to the outage analysis of user cooperation. *IEEE Communications Letters*, 17(8):1505–1508, 2013.

Richard Cheng, Gábor Orosz, Richard M. Murray, and Joel W. Burdick. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 3387–3395, 2019.

Sayak Ray Chowdhury and Aditya Gopalan. On Kernelized Multi-armed Bandits. In *Proceedings of the International Conference on Machine Learning*, pp. 844–853, 2017.

Earl A. Coddington and Norman Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

Max H. Cohen, Tamas G. Molnar, and Aaron D. Ames. Safety-critical control for autonomous systems: Control barrier functions via reduced-order models. *Annual Reviews in Control*, 57: 100947, 2024.

Ryan K Cosner, Andrew W Singletary, Andrew J Taylor, Tamas G Molnar, Katherine L Bouman, and Aaron D Ames. Measurement-robust control barrier functions: Certainty in safety with uncertainty in state. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6286–6291, 2021.

Sebastian Curi, Armin Lederer, Sandra Hirche, and Andreas Krause. Safe Reinforcement Learning via Confidence-Based Filters. In *Proceedings of the IEEE Conference on Decision and Control*, 2022.

Xiaobing Dai, Armin Lederer, Zewen Yang, and Sandra Hirche. Can learning deteriorate control? analyzing computational delays in gaussian process-based event-triggered online learning. In *Learning for Dynamics and Control Conference*, pp. 445–457. PMLR, 2023.

Charles Dawson, Sicun Gao, and Chuchu Fan. Safe control with learned certificates: A survey of neural lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics*, 39(3):1749–1767, 2023.

Massimiliano De Sa, Prasanth Kotaru, and Koushil Sreenath. Point cloud-based control barrier function regression for safe and efficient vision-based control. In *IEEE International Conference on Robotics and Automation*, pp. 366–372, 2024.

Vikas Dhiman, Mohammad Javad Khojasteh, Massimo Franceschetti, and Nikolay Atanasov. Control barriers in bayesian learning of system dynamics. *IEEE Transactions on Automatic Control*, 68(1):214–229, 2023. doi: 10.1109/TAC.2021.3137059.

Steven Diamond and Stephen Boyd. Cvxpy: A python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

Gabriel Dulac-Arnold, Nir Levine, Daniel J Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, 110(9):2419–2468, 2021.

Yousef Emam, Gennaro Notomista, Paul Glotfelter, Zsolt Kira, and Magnus Egerstedt. Safe reinforcement learning using robust control barrier functions. *IEEE Robotics and Automation Letters*, 10(3):2886–2893, 2025. doi: 10.1109/LRA.2022.3216996.

Christian Fiedler, Johanna Menn, Lukas Kreisköther, and Sebastian Trimpe. On Safety in Safe Bayesian Optimization. *Transactions on Machine Learning Research*, 2024.

Thor I Fossen. *Handbook of marine craft hydrodynamics and motion control*. John Wiley & Sons, 2011.

Franka Robotics. Panda joint limits: Official franka_ros robot description. `https://github.com/frankarobotics/franka_ros/blob/develop/franka_description/robots/panda/joint_limits.yaml`. Accessed: 2026-09-09.

Jacob R. Gardner, Geoff Pleiss, David Bindel, Kilian Q. Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In *Advances in Neural Information Processing Systems*, 2018.

Paul J. Goulart and Yuwen Chen. Clarabel: An interior-point solver for conic programs with quadratic objectives, 2024.

Lukas Hewing, Kim P. Wabersich, Marcel Menner, and Melanie N. Zeilinger. Learning-Based Model Predictive Control: Toward Safe Learning in Control. *Annual Review of Control, Robotics, and Autonomous Systems*, 3:269–296, May 2020.

Kai-Chieh Hsu, Haimin Hu, and Jaime F. Fisac. The Safety Filter: A Unified View of Safety-Critical Control in Autonomous Systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 7(1):47–72, July 2024.

Tzu-Yuan Huang, Armin Lederer, Dai-Jie Wu, Xiaobing Dai, Sihua Zhang, Hsiu-Chin Lin, Shao-Hua Sun, Stefan Georg Sosnowski, and Sandra Hirche. SAD-flower: Flow matching for safe, admissible, and dynamically consistent planning. In *Forty-third International Conference on Machine Learning*, 2026. URL `https://openreview.net/forum?id=wb97oR78OX`.

Pushpak Jagtap, George J. Pappas, and Majid Zamani. Control Barrier Functions for Unknown Nonlinear Systems using Gaussian Processes. In *Proceedings of the IEEE Conference on Decision and Control*, pp. 3699–3704, 2020.

Junjie Jiao, Alexandre Capone, and Sandra Hirche. Backstepping tracking control using gaussian processes with event-triggered online learning. *IEEE Control Systems Letters*, 6:3176–3181, 2022.

Andreas Krause and Cheng Ong. Contextual Gaussian process bandit optimization. *Advances in Neural Information Processing Systems*, 24, 2011.

Amon Lahr, Johannes Köhler, Anna Scampicchio, and Melanie Zeilinger. Optimal kernel regression bounds under energy-bounded noise. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 100411–100444. Curran Associates, Inc., 2025. doi: 10.52202/085713-3357. URL `https://proceedings.neurips.cc/paper_files/paper/2025/file/90e73f3cf1a6c84c723a2e8b7fb2b2c1-Paper-Conference.pdf`.

Armin Lederer. *Gaussian Processes in Control: Performance Guarantees through Efficient Learning*. PhD thesis, Technische Universität München, 2023.

Armin Lederer, Jonas Umlauft, and Sandra Hirche. Uniform error bounds for Gaussian process regression with application to safe control. *Advances in Neural Information Processing Systems*, 32, 2019.

Armin Lederer, Alejandro J Ordónez Conejo, Korbinian A Maier, Wenxin Xiao, Jonas Umlauft, and Sandra Hirche. Gaussian process-based real-time learning for safety critical applications. In *International Conference on Machine Learning*, pp. 6055–6064. PMLR, 2021.

Armin Lederer, Azra Begzadić, Sandra Hirche, Jorge Cortés, and Sylvia Herbert. Safe event-triggered gaussian process learning for barrier-constrained control. *IEEE Transactions on Automatic Control*, 71(9):6179–6186, 2026a. doi: 10.1109/TAC.2026.3680122.

Armin Lederer, Anuj Srivastava, Marco Bagatella, and Andreas Krause. Policy Search via Bayesian Optimization with Temporal Difference Gaussian Processes. In *Proceedings of the International Conference on Machine Learning*, 2026b.

Armin Lederer, Jonas Umlauft, and Sandra Hirche. Episodic Gaussian process-based learning control with vanishing tracking errors. *IEEE Transactions on Automatic Control*, 71(2):962–977, 2026c.

Jialin Li, Marta Zagorowska, Giulia De Pasquale, Alisa Rupenyan, and John Lygeros. Safe Time-Varying Optimization based on Gaussian Processes with Spatio-Temporal Kernel. In *Advances in Neural Information Processing Systems*, 2024.

Haitao Liu, Yew-Soon Ong, Xiaobo Shen, and Jianfei Cai. When gaussian process meets big data: A review of scalable gps. *IEEE transactions on neural networks and learning systems*, 31(11): 4405–4423, 2020.

Kehan Long, Cheng Qian, Jorge Cortés, and Nikolay Atanasov. Learning barrier functions with memory for robust safe navigation. *IEEE Robotics and Automation Letters*, 6(3):4931–4938, 2021.

Kehan Long, Vikas Dhiman, Melvin Leok, Jorge Cortes, and Nikolay Atanasov. Safe Control Synthesis With Uncertain Dynamics and Constraints. *IEEE Robotics and Automation Letters*, 7: 7295–7302, 2022.

Robert Mahony, Vijay Kumar, and Peter Corke. Multirotor aerial vehicles: Modeling, estimation, and control of quadrotor. *IEEE Robotics & Automation Magazine*, 19(3):20–32, 2012.

Rowan McAllister and Carl Edward Rasmussen. Data-efficient reinforcement learning in continuous state-action Gaussian-POMDPs. *Advances in Neural Information Processing Systems*, 30, 2017.

Pol Mestres and Jorge Cortés. Feasibility and regularity analysis of safe stabilizing controllers under uncertainty. *Automatica*, 167:111800, 2024.

Richard M. Murray, Zexiang Li, and S. Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

Vern I Paulsen and Mrinal Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. Cambridge University Press, 2016.

Zengyi Qin, Kaiqing Zhang, Yuxiao Chen, Jingkai Chen, and Chuchu Fan. Learning safe multi-agent control with decentralized neural barrier certificates. In *International Conference on Learning Representations*, 2021.

Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, 2006.

Robert Reed, Luca Laurenti, and Morteza Lahijanian. Error bounds for Gaussian process regression under bounded support noise with applications to safety certification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20157–20164, 2025.

Andrew Singletary, William Guffey, Tamas Molnar, Ryan Sinnet, and Aaron Ames. Safety-critical manipulation for collision-free food preparation. *IEEE Robotics and Automation Letters*, 7(4): 10954–10961, 2022.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias W Seeger. Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, 2012.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with Gaussian processes. In *International Conference on Machine Learning*, pp. 997–1005. PMLR, 2015.

Andrew Taylor, Andrew Singletary, Yisong Yue, and Aaron Ames. Learning for safety-critical control with control barrier functions. In *Learning for dynamics and control*, pp. 708–717. PMLR, 2020.

Johannes Teutsch, Oleksii Molodchyk, Marion Leibold, Timm Faulwasser, and Armin Lederer. On uniform error bounds for kernel regression under non-gaussian noise. *arXiv preprint arXiv:2605.09757*, 2026.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012. doi: 10.1109/IROS.2012.6386109.

Abdullah Tokmak, Thomas B Schön, and Dominik Baumann. Pacsbo: Probably approximately correct safe bayesian optimization. In *Symposium on Systems Theory in Data and Optimization*, pp. 3–18. Springer, 2024.

Abdullah Tokmak, Kiran G. Krishnan, Thomas B. Schön, and Dominik Baumann. Safe exploration in reproducing kernel Hilbert spaces. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, March 2025.

Matteo Turchetta, Felix Berkenkamp, and Andreas Krause. Safe exploration in finite Markov decision processes with Gaussian processes. *Advances in Neural Information Processing Systems*, 29, 2016.

Jonas Umlauft and Sandra Hirche. Feedback Linearization based on Gaussian Processes with Event-Triggered Online Learning. *IEEE Transactions on Automatic Control*, 65(10):4154–4169, 2019.

Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in Gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 82–90. PMLR, 2021.

Kim P. Wabersich, Andrew J. Taylor, Jason J. Choi, Koushil Sreenath, Claire J. Tomlin, Aaron D. Ames, and Melanie N. Zeilinger. Data-Driven Safety Filters: Hamilton-Jacobi Reachability, Control Barrier Functions, and Predictive Methods for Uncertain Systems. *IEEE Control Systems Magazine*, 43(5):137–177, October 2023.

Akifumi Wachi, Wataru Hashimoto, Xun Shen, and Kazumune Hashimoto. Safe exploration in reinforcement learning: A generalized formulation and algorithms. In *Neural Information Processing Systems*, 2023.

Li Wang, Evangelos A Theodorou, and Magnus Egerstedt. Safe learning of quadrotor dynamics using barrier certificates. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2460–2465. IEEE, 2018.

Yunyue Wei, Zeji Yi, Hongda Li, Saraswati Soedarmadji, and Yanan Sui. Safe bayesian optimization for the control of high-dimensional embodied systems. In *Proceedings of the Conference on Robot Learning*, pp. 4771–4792, 2025.

Tizian Wenzel, Abdullah Tokmak, and Christian Fiedler. Reliable sampling-based rkhs norm estimation via superconvergence. *arXiv preprint arXiv:2605.20091*, 2026.

Zhixuan Wu, Rui Yang, Lei Zheng, and Hui Cheng. Safe learning-based feedback linearization tracking control for nonlinear system with event-triggered model update. *IEEE Robotics and Automation Letters*, 7(2):3286–3293, 2022.

Wei Xiao and Calin Belta. High-order control barrier functions. *IEEE Transactions on Automatic Control*, 67(7):3655–3662, 2022.

Wei Xiao, Calin Belta, and Christos G. Cassandras. Event-triggered control for safety-critical systems with unknown dynamics. *IEEE Transactions on Automatic Control*, 68(7):4143–4158, 2023a. doi: 10.1109/TAC.2022.3202088.

Wei Xiao, Tsun-Hsuan Wang, Chuang Gan, Ramin Hasani, Mathias Lechner, and Daniela Rus. Safediffuser: Safe planning with diffusion probabilistic models. In *Proceedings of the International Conference on Learning Representations*, 2023b.

Xiangru Xu, Paulo Tabuada, Jessy W Grizzle, and Aaron D Ames. Robustness of control barrier functions for safety critical control. *IFAC-PapersOnLine*, 48(27):54–61, 2015.

Zewen Yang, Dongfa Zhang, Xiaobing Dai, Fengyi Yu, Chi Zhang, Bingkun Huang, Hamid Sadeghian, and Sami Haddadin. Streaming generated gaussian process experts for online learning and control. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 27719–27727, 2026.

Chenxu Zhang, Xiaohan Li, Xinghu Wang, and Haibo Ji. Event-triggered learning-based control of quadrotors for accurate agile trajectory tracking. *IEEE Robotics and Automation Letters*, 9(6): 5855–5862, 2024.

Sihua Zhang, Di-Hua Zhai, Xiaobing Dai, Tzu-yuan Huang, and Yuanqing Xia. Gaussian process-based prescribed-time input-to-state safety for unknown system control. *IEEE Transactions on Industrial Electronics*, 72(12):14271–14280, 2025a.

Yu Zhang, Long Wen, Zhenshan Bing, Xiangtong Yao, Linghuan Kong, Wei He, and Alois Knoll. Adaptive safety-critical control for high-order systems: A real-time Gaussian process approach. *IEEE Transactions on Automation Science and Engineering*, 22:21296–21310, 2025b.

Ding-Xuan Zhou. Derivative reproducing properties for kernel methods in learning theory. *Journal of Computational and Applied Mathematics*, 220(1-2):456–463, 2008.

A Discussion of the CBF Data Collection Assumption

A.1 CBF Data Collection Assumption

Noise-free state measurements are a standard requirement in learning-based control design (McAllister & Rasmussen, 2017; Berkenkamp et al., 2017; Hewing et al., 2020), used to streamline theoretical analyses. It is a clear simplification of reality due to measurement noise and the frequent need for state observers. However, it is a reasonable simplification when estimation errors and noise are small and bounded.

To see this, let $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{\epsilon}_x$ be the estimated, perturbed state, where $\|\boldsymbol{\epsilon}_x\| \leq \bar{e}$ for some constant $\bar{e} \in \mathbb{R}_{\geq 0}$. If $\tilde{\phi}(\cdot)$, $\tilde{\boldsymbol{\psi}}(\cdot)$, $\mu(\cdot,\cdot)$ and $\sigma^2(\cdot,\cdot)$ are Lipschitz continuous with constants $L_{\tilde{\phi}}$, $L_{\tilde{\psi}}$, L_{μ_x} and $L_{\sigma_x}\bar{e}$, we have

$$\hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{u} - \beta \left\| \begin{bmatrix} 1 & \boldsymbol{u}^\top \end{bmatrix} \boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{x}) \right\|_2 \geq \hat{\phi}(\hat{\boldsymbol{x}}) + \hat{\boldsymbol{\psi}}^\top(\hat{\boldsymbol{x}})\boldsymbol{u} - \beta \left\| \begin{bmatrix} 1 & \boldsymbol{u}^\top \end{bmatrix} \boldsymbol{\Sigma}^{\frac{1}{2}}(\hat{\boldsymbol{x}}) \right\|_2 - (L_{\tilde{\phi}} + L_{\tilde{\psi}} + L_{\mu_x})\bar{e} + \sqrt{L_{\sigma_x}\bar{e}}.$$

Thus, we simply have to choose a larger constraint tightening constant η_1 in (4) to account for state perturbations in the safety filter, which does not introduce significant conservatism if $\bar{e}$ is small. Note that GP means and variances are Lipschitz continuous for Lipschitz continuous kernels, such that this requirement is not very restrictive in general. Additionally, we can bound the learning errors for GP regression using similar Lipschitz continuity requirements on the ground-truth functions $\phi(\cdot)$ and $\psi(\cdot)$ because continuity allows us to obtain $\left|\phi(\boldsymbol{x}) + \psi^\top(\boldsymbol{x})\boldsymbol{u} - \left(\phi(\hat{\boldsymbol{x}}) + \psi^\top(\hat{\boldsymbol{x}})\boldsymbol{u}\right)\right| \leq (L_\phi + L_\psi)\bar{e}$,i.e., we can treat state perturbations as bounded errors in the training targets. As shown in Lederer (2023, Lemma 6.1), this approach merely introduces a need for a larger value of β in (4) when dealing with perturbed states.

Noise on training targets y_t allows for addressing perturbations, such that y_t can be obtained via numerical differentiation based on sequential evaluations of the control barrier function. While the assumption of zero-mean sub-Gaussian noise is a simplification to streamline the derivation of error bounds, it is not crucial. Extensions for other mean-free noise distributions exist (Reed et al., 2025; Teutsch et al., 2026). Moreover, bias terms with a known uniform bound can be treated

analogously to bounded target perturbations caused by noisy or estimated states discussed in the previous paragraph.

Finally, Assumption 1 requires local CBF knowledge via point-wise evaluations. While finding a CBF is generally a difficult problem, structural knowledge, e.g., about full actuation, and approximate models significantly simplify it for many practical systems, which we also exploit in our numerical simulations. Moreover, Assumption 1 does not require a globally valid closed-form expression of the CBF, such that it is applicable to settings in which we do not know about the safe set in advance, e.g., because we do not know the location of all obstacles. Thus, this assumption admits settings with data-driven CBF representations based on distance functions, which are attracting increasing attention in robotics (Long et al., 2021; Singletary et al., 2022; De Sa et al., 2024; Begzadić et al., 2025).

A.2 Discussion of the RKHS Regularity Assumption

Many commonly used kernels are sufficiently smooth to satisfy the bounded-Hessian-trace requirement in Assumption 2. Examples include squared exponential kernels and Matérn kernels with $\nu \geq 3/2$ (Rasmussen & Williams, 2006). The differentiability of these kernels transfers to functions in their RKHSs (Zhou, 2008), and therefore gives the required differentiability of $\Delta\phi(\cdot)$ and $\Delta\psi_i(\cdot)$ under the assumed bounded RKHS norms.

Bounded RKHS norms are commonly used to restrict the complexity of unknown functions when deriving kernel regression error bounds (Srinivas et al., 2012; Chowdhury & Gopalan, 2017; Lahr et al., 2025). To evaluate any such bound, valid numerical upper bounds on the RKHS norms must be available. This requirement is not specific to our method and common when addressing safety using kernel methods (Berkenkamp et al., 2017; Li et al., 2024; Capone et al., 2025; As et al., 2025). Despite the frequent use of RKHS norm bounds, this assumption is not undisputed in the literature, see e.g., Fiedler et al. (2024). In particular, in the context of safety, it has been criticized that knowledge of a RKHS norm bound is an impractical assumption: Even if we know that a RKHS would be the suitable space for representing an unknown function, we still may not have any knowledge of the norm of the function in the RKHS.

However, if we substantially overestimate the RKHS norm bound, β increases a lot, which can render pessimistic approaches prohibitively conservative. In our approach, it enlarges the uncertainty term subtracted in (4b), making the safety filter more conservative and potentially reducing its feasible control set. It also makes (6b) easier to satisfy and can therefore cause earlier or more frequent exploration. In contrast, underestimating the norm bounds can invalidate the regression error bound and hence subsequent safety guarantees.

While these issues do exist, it is important to note that they are not specific to our approach, but they are shared among all related work that addresses safety using kernel methods (Berkenkamp et al., 2017; Li et al., 2024; Capone et al., 2025; As et al., 2025), which has multiple implications. Firstly, this large number of existing works demonstrates empirically that methods relying on the RKHS norm assumption can perform well when treating β as a design parameter that is manually fixed to a constant value (often fixed to 2). Thus, there exists substantial numerical evidence suggesting that precise knowledge of the RKHS norm bound may not be necessary for good empirical performance. Secondly, the wide use of this assumption has led to the development of approaches to reliably estimate the RKHS norm from data (Tokmak et al., 2024; 2025; Wenzel et al., 2026). By accounting for the uncertainty in these estimates, we can immediately integrate them into our approach to retain safety guarantees without requiring an a priori known RKHS norm bound. Note that our approach will similarly benefit from future improvements in estimating RKHS norms, which seem reasonable due to the field's increasing attention in recent years.

B Practical Parameter Selection

While the conditions in our theorems provide sufficient conditions and qualitative guidance for choosing all constants, we retain the problem of tuning these parameters in practice. When directly deploying our approach on a real-world system, one can start with conservative choices—an approach that we also exemplarily execute using a small number of preliminary runs in our numerical experiments. For this, one starts with a sufficiently large value of β, which we empirically found

to be in the range between 5–10, which is considerably larger than the commonly used value of $\beta = 2$ in many papers. Note that this is a range which we also frequently obtain when evaluating Bayesian regression error bounds that rely on the GP posterior (Lederer et al., 2019). This value can be slowly reduced while monitoring the validity of the regression error bound.

Since large $\epsilon \approx 1$ can cause violations of the CBF conditions during the exploratory control sequence, one starts with a small value, e.g., $\epsilon = 0.1$, and gradually increases it to improve structured exploration while maintaining a clear positive CBF margin. As the parameter η_1 tightens the CBF constraint to account for the digital implementation, one starts again at large values and gradually reduces them until the discretization errors come close to the choice of η_1. However, we want to note that the required magnitude of η_1 depends on the system behavior: While a smaller value may suffice for slower systems, larger values are generally required for fast systems.

Moreover, η_1 effectively reduces the size of the safe set, such that its value needs to be chosen in relation to the maximum value of the CBF: the larger the max of the CBF, the larger the value of η_1 that one should start with. The value of η_2 is lower bounded by η_1, such that one begins at $\eta_2 = \eta_1$ and gradually increases the value of η_2. Finally, the sampling time T_s is usually determined by the update period of the low-level controller, such that it does not need to be tuned.

C Computational Scalability

For N training samples, recomputing an exact GP posterior from scratch has cubic time complexity and quadratic storage complexity. The event-triggered design reduces the frequency of model updates, and Theorem 2 bounds the number of triggered exploration sequences under its stated conditions. However, it does not remove the dependence of each GP update on the accumulated number of samples. Exact GP regression can therefore remain computationally expensive during long deployments or in control loops with short sampling times.

Scalable GP approximations can be used to address this issue; an overview is provided by (Liu et al., 2020). In particular, online GP methods designed for control loops (Lederer et al., 2021; Yang et al., 2026) can reduce the computational cost of updating the model after an exploration event.

When such an approximation is used in our framework, it must provide a valid error or uncertainty bound that can replace $\beta\sigma(\boldsymbol{x}, \boldsymbol{u})$ in the robust CBF constraint (4b) and the triggering condition (6b). If the approximation introduces an additional error, this error must instead be included through an additional tightening of the CBF constraint. Safety guarantees for event-triggered learning with scalable GP updates have, for example, been established in Lederer et al. (2021).

D Proofs for Theoretical Guarantees

In these appendices, we provide the rigorous problem definitions, auxiliary lemmas, and formal proofs for the theoretical results stated in the main text.

D.1 Learning Error Bounds for GP Regression

Lemma 2 (Uniform Error Bound (Lederer et al., 2019)). *Suppose Assumptions 1 and 2 hold. Given a compact domain $\Omega \subset \mathbb{R}^n \times \mathbb{R}^m$, a probability $\delta \in (0, 1)$, and a maximum number of data points $\bar{N} \in \mathbb{N}, \bar{N} \geq |\mathbb{D}_N|$, let $\beta \in \mathbb{R}_{\geq 0}$ be such that*

$$\beta \geq B + R\sqrt{2\gamma_{\bar{N}} + 1 + \log(\tfrac{1}{\delta})}, \tag{12}$$

where $\gamma_{\bar{N}}$ is the maximum information gain

$$\gamma_{\bar{N}} = \max_{\mathbb{D}:|\mathbb{D}|\leq\bar{N}} \log\det\left(\boldsymbol{I} + \sigma_{\text{on}}^{-2}\boldsymbol{K}\right) \tag{13}$$

with $[\boldsymbol{K}]_{ij} = k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ for all $\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \in \mathbb{D}$. Then, it holds that

$$P\left(|\hat{\phi}_i(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}_i^\top(\boldsymbol{x})\boldsymbol{u} - \phi(\boldsymbol{x}) - \boldsymbol{\psi}^\top(\boldsymbol{x})\boldsymbol{u}| \leq \beta\sigma_i(\boldsymbol{x}, \boldsymbol{u}),\ \forall(\boldsymbol{x}, \boldsymbol{u}) \in \Omega, i = 1, \ldots, \bar{N}\right) \geq 1-\delta, \tag{14}$$

where the index i denotes the number of data points the GP is conditioned on.

Proof. Define $c(\boldsymbol{x}(t), \boldsymbol{u}(t)) = \phi(\boldsymbol{x}(t)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t)$. Then, it follows from Paulsen & Raghupathi (2016) that

$$\|c(\cdot,\cdot)\|_k^2 \leq 2\|\phi(\cdot)\|_{k_\phi}^2 + \sum_{i=1}^{m} \|\psi_i(\cdot)\|_{k_{\psi_i}}^2, \tag{15}$$

such that $\|c(\cdot,\cdot)\|_k \leq B_\phi + \sum_{i=1}^m B_{\psi_i}$. This allows us the application of Chowdhury & Gopalan (2017, Theorem 2), which results in

$$|\hat{\phi}_i(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}_i^\top(\boldsymbol{x})\boldsymbol{u} - c(\boldsymbol{x},\boldsymbol{u})| \leq \left(B + R\sqrt{2\gamma_i + 1 + \log(\tfrac{1}{\delta})}\right)\sigma_i(\boldsymbol{x},\boldsymbol{u}) \tag{16}$$

with probability $1-\delta$ jointly for all $(\boldsymbol{x},\boldsymbol{u}) \in \Omega, i \in \mathbb{N}$. Since γ_i is non-decreasing, it can be upper bounded by $\gamma_{\bar{N}}$, which immediately proves the result. □

D.2 Safety During Active Exploration

To prove the safety of the open-loop exploration sequence, we establish two supporting lemmas. The first lemma guarantees that the optimal control direction for the GP mean function satisfies a robust safety margin, while the second lemma extends this safety guarantee to the subsequent orthogonal perturbations.

Lemma 3 (Safety of Nominal Maximizing Direction)**.** *Consider the continuous-time system* (1) *and a continuously differentiable CBF $h(\boldsymbol{x})$. Suppose Assumptions 1 and 2 are satisfied and let $\hat{\boldsymbol{u}}^*$ be defined as the maximizer of the nominal learned CBF condition*

$$\hat{\boldsymbol{u}}^* = \arg\max_{\|\boldsymbol{u}\| \leq 1-\epsilon} \hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{u}. \tag{17}$$

If the GP predictive uncertainty strictly satisfies the condition

$$\beta \max_{\|\boldsymbol{u}\| \leq 1} \sigma(\boldsymbol{x},\boldsymbol{u}) \leq \max_{\|\boldsymbol{u}\| \leq 1-\epsilon} \hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{u} - \tilde{\theta} \tag{18}$$

for some $\tilde{\theta} \in \mathbb{R}_{\geq 0}$ and β satisfying (12)*, then, the true unmodeled CBF derivative satisfies*

$$\phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} \geq \tilde{\theta} \tag{19}$$

with probability at least $1-\delta$.

Proof. We can apply the upper bound (14) of the GP prediction error to obtain

$$\hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} \leq \phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} + \beta(\mathbb{D}_N,\delta)\sigma(\boldsymbol{x}, \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon}) \tag{20}$$

with probability $1-\delta$. Since $\sigma(\boldsymbol{x}, \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon})$ can be bounded by its maximum, we obtain

$$\hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} \leq \phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} + \beta(\mathbb{D}_N,\delta)\max_{\|\boldsymbol{u}\| \leq 1}\sigma(\boldsymbol{x},\boldsymbol{u}) \tag{21}$$

with probability $1-\delta$. Due to (18) and $\hat{\psi}^\top(\boldsymbol{x})\hat{\boldsymbol{u}}^* \leq \hat{\psi}^\top(\boldsymbol{x})\frac{\hat{\boldsymbol{u}}^*}{1-\epsilon}$, (19) immediately follows. □

Building upon the safety of the anchor direction $\hat{\boldsymbol{u}}^*$, we extend the analysis to the perturbed input directions. Since the anchor input satisfies $\|\hat{\boldsymbol{u}}^*\| \leq 1-\epsilon$, unit norm perturbations can be easily defined using orthonormal basis vectors $\boldsymbol{d}_i$ via $\|\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i\|^2 = \|\hat{\boldsymbol{u}}^*\|^2 + \epsilon^2\|\boldsymbol{d}_i\|^2 \leq 1$.

Lemma 4 (Safety of Basis Completion)**.** *Consider the continuous-time system* (1) *and a continuously differentiable CBF $h(\boldsymbol{x})$. Suppose Assumptions 1 and 2 are satisfied, let $\hat{\boldsymbol{u}}^*$ be defined as in* (17)*, and let $\boldsymbol{d}_i \in \mathbb{R}^m$ be an arbitrary unit-norm vector that is orthogonal to $\hat{\boldsymbol{u}}^*$. If* (18) *holds for some $\tilde{\theta} \in \mathbb{R}_{\geq 0}$ and β satisfying* (12)*, the true CBF derivative satisfies*

$$\phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})(\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i) \geq \tilde{\theta} \tag{22}$$

with probability at least $1-\delta$ for every $\epsilon \in (0,1)$.

Proof. Due to the GP regression error bound from Lemma 2, we have

$$\phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})(\hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) \geq \hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})(\hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) - \beta\sigma(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) \tag{23}$$

with probability $1-\delta$. Recall that the anchor direction $\hat{\boldsymbol{u}}^*$ is defined as the maximizer of the affine function $\hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{u}$ over a symmetric ball. Therefore, it is straightforward to see that $\hat{\boldsymbol{u}}^*$ is parallel to $\hat{\boldsymbol{\psi}}(\boldsymbol{x})$ due to linearity of objective in $\boldsymbol{u}$. Because the vector $\boldsymbol{d}_i$ is orthogonal to the anchor direction, i.e., $\boldsymbol{d}_i^\top \hat{\boldsymbol{u}}^* = 0$, it must also be orthogonal to the nominal gradient vector $\hat{\boldsymbol{\psi}}(\boldsymbol{x})$, i.e., $\hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\boldsymbol{d}_i = 0$. Therefore, we can simplify (23) to

$$\phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})(\hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) \geq \hat{\phi}(\boldsymbol{x}) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x})\hat{\boldsymbol{u}}^* - \beta \max_{\|\boldsymbol{u}\| \leq 1} \sigma(\boldsymbol{x}, \boldsymbol{u}), \tag{24}$$

with probability $1-\delta$, where we additionally bound the GP standard deviation $\sigma(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i)$ by its maximum. Due to (18), this immediately yields (22) with probability $1-\delta$. □

With these two foundational lemmas established, we now formalize the safety proof for the entire active exploration sequence across the continuous time window.

Proposition 1 (Safety of Active Exploration Sequence)**.** *Consider the continuous-time system* (1) *and a continuously differentiable CBF $h(\boldsymbol{x})$. Suppose Assumptions 1, 3 and 2 are satisfied, let $\hat{\boldsymbol{u}}^*$ be defined as in* (17)*, and let $\{\hat{\boldsymbol{u}}^*/(1-\epsilon), \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$ be an orthonormal basis of $\mathbb{R}^m$. Define Δ such that*

$$\|\boldsymbol{x}(t_1) - \boldsymbol{x}(t_2)\| \leq \Delta |t_1 - t_2| \tag{25}$$

for all $\boldsymbol{u} \in \mathcal{U}$, $t_1, t_2 \in \mathbb{R}_{\geq 0}$. Assume that (18) *holds at time t^* with β satisfying* (12) *and*

$$\tilde{\theta} \geq m\Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s, \tag{26}$$

where L_ϕ and $\boldsymbol{L}_{\boldsymbol{\psi}}$ are Lipschitz constants of $\phi(\cdot)$ and $\boldsymbol{\psi}(\cdot)$, respectively. Then, the open-loop control signal

$$\boldsymbol{u}(t) = \begin{cases} \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} & \text{if } t \in [t^*, t^* + T_s) \\ \hat{\boldsymbol{u}}^* + \epsilon d_i & \text{if } t \in [t^* + iT_s, t^* + (i+1)T_s) \end{cases} \tag{27}$$

guarantees $\phi(\boldsymbol{x}(t)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t) \geq 0$ jointly for all $t \in [t^*, t^* + mT_s)$ with probability $1-\delta$.

Proof. Due to the restriction to a compact safe set in Assumption 1, the Lipschitz continuity of $\boldsymbol{f}(\cdot)$ and $\boldsymbol{G}(\cdot)$ in combination with the boundedness of $\boldsymbol{u}$ imply a boundedness of the right-hand side of (1), which immediately guarantees the existence of a constant Δ satisfying (25). Moreover, the Lipschitz constants L_ϕ and $\boldsymbol{L}_{\boldsymbol{\psi}}$ are guaranteed to exist due to the bounded RKHS norm of $\phi(\cdot)$ and $\boldsymbol{\psi}(\cdot)$ (Lederer et al., 2026b). Exploiting these smoothness properties, we can bound

$$\phi(\boldsymbol{x}(t)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t) \geq \phi(\boldsymbol{x}(t^*)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t^*))\boldsymbol{u}(t) - (L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\| \max_{\boldsymbol{u} \in \mathcal{U}} \|\boldsymbol{u}\|)\Delta(t - t^*). \tag{28}$$

Since $t - t^* \leq mT_s$ for $t \in [t^*, t^* + mT_s)$ and $\max_{\boldsymbol{u} \in \mathcal{U}} \|\boldsymbol{u}\| = 1$ due to the definition of $\mathcal{U}$, this bound further simplifies to

$$\phi(\boldsymbol{x}(t)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t) \geq \phi(\boldsymbol{x}(t^*)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t^*))\boldsymbol{u}(t) - (L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)\Delta m T_s. \tag{29}$$

We finally apply Lemmas 3 and 4 to obtain

$$\phi(\boldsymbol{x}(t)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t) \geq \tilde{\theta} - (L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)\Delta m T_s, \tag{30}$$

which is guaranteed to be lower bounded by 0 due to (26). □

D.3 Variance Bound via Basis Decomposition

To ensure the safety filter recovers feasibility, the executed basis sequence must ensure an upper bound for the maximum GP variance over the control input space $\mathcal{U}$. We analyze the structural properties of the GP covariance matrix to explicitly derive this bound.

Lemma 5 (Variance Reduction in Identical Direction)**.** *Let $\mathbb{D}_{N+1} = \mathbb{D}_N \cup \{([\boldsymbol{x}^\top, \boldsymbol{u}^\top]^\top, y)\}$ be the dataset augmented with a single measurement. If the GP posterior variance* (3) *for the data set $\mathbb{D}_{N+1}$ is based on stationary kernels $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$, its evaluation at the new data point $(\boldsymbol{x}, \boldsymbol{u})$ satisfies*

$$\sigma^2(\boldsymbol{x}, \boldsymbol{u}) \leq \frac{\sigma^2_{\text{on}} \left(\sigma^2_{\phi,0} + \sum_{j=1}^m u_j^2 \sigma^2_{\psi_j,0}\right)}{\sigma^2_{\text{on}} + \sigma^2_{\phi,0} + \sum_{j=1}^m u_j^2 \sigma^2_{\psi_j,0}}, \tag{31}$$

where σ^2_{on} is the observation noise variance, and $\sigma^2_{\phi,0} = k_\phi(\boldsymbol{x}, \boldsymbol{x})$, $\sigma^2_{\psi_j,0} = k_{\psi_j}(\boldsymbol{x}, \boldsymbol{x})$ are the prior variances.

Proof. The proof is an immediate consequence of Lederer et al. (2026c, Corollary 2). □

Lemma 6 (Global Variance Bound via Basis)**.** *Let $\{\hat{\boldsymbol{u}}^*/(1-\epsilon), \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$ with $\hat{\boldsymbol{u}}^*$ defined in* (17) *form an orthonormal basis for $\mathbb{R}^m$. If the posterior variance* (3) *is based on kernels $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$, its evaluation at an arbitrary control input $\boldsymbol{u} \in \mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^m : \|\boldsymbol{u}\| \leq 1\}$ is upper-bounded by a linear combination of the variances evaluated at the basis vectors, i.e.,*

$$\begin{aligned}\sigma^2(\boldsymbol{x}, \boldsymbol{u}) \leq & \frac{(m+1)}{\epsilon^2} \sum_{i=1}^{m-1} \sigma^2(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) + \frac{(m+1)^3}{\epsilon^2} \sigma^2\left(\boldsymbol{x}, \frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) \\ & + (m+1)^3 \sigma^2(\boldsymbol{x}, \boldsymbol{0}).\end{aligned} \tag{32}$$

Proof. To establish this upper bound, we express $\boldsymbol{u}$ as a linear combination of the basis sequence. Define the augmented basis matrix $\boldsymbol{B}$ as

$$\boldsymbol{B} = [\boldsymbol{b}_1 \quad \cdots \quad \boldsymbol{b}_{m-1} \quad \boldsymbol{b}_m \quad \boldsymbol{b}_{m+1}] = \begin{bmatrix} 1 & \cdots & 1 & 1 & 1 \\ \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_1 & \cdots & \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_{m-1} & \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon} & \boldsymbol{0} \end{bmatrix}, \tag{33}$$

and correspondingly define the projection coefficient vector $\boldsymbol{\eta}$ for any arbitrary control input $\boldsymbol{u}$ as

$$\boldsymbol{\eta} = \boldsymbol{B}^{-1} \begin{bmatrix} 1 \\ \boldsymbol{u} \end{bmatrix}. \tag{34}$$

Substituting the basis expansion of $\boldsymbol{u}$ into the definition of the GP variance in (3), we obtain

$$\sigma^2(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{\eta}^\top \boldsymbol{B}^\top \boldsymbol{\Sigma}(\boldsymbol{x}) \boldsymbol{B} \boldsymbol{\eta} = \|\boldsymbol{B}\boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}(\boldsymbol{x})} = \left\| \sum_{j=1}^{m+1} \eta_j \boldsymbol{b}_j \right\|^2_{\boldsymbol{\Sigma}(\boldsymbol{x})}. \tag{35}$$

Applying Young's inequality, we can further simplify this expression to

$$\sigma^2(\boldsymbol{x}, \boldsymbol{u}) \leq (m+1) \sum_{j=1}^{m+1} |\eta_j|^2 \|\boldsymbol{b}_j\|^2_{\boldsymbol{\Sigma}(\boldsymbol{x})}. \tag{36}$$

Due to the definition of the basis vectors $\boldsymbol{b}_i$, the terms $\|\boldsymbol{b}_j\|^2_{\boldsymbol{\Sigma}(\boldsymbol{x})}$ correspond to GP variances, which yields

$$\begin{aligned}\sigma^2(\boldsymbol{x}, \boldsymbol{u}) \leq & (m+1) \sum_{i=1}^{m-1} |\eta_i|^2 \sigma^2(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) + (m+1)|\eta_m|^2 \sigma^2\left(\boldsymbol{x}, \frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) \\ & + (m+1)|\eta_{m+1}|^2 \sigma^2(\boldsymbol{x}, \boldsymbol{0}).\end{aligned} \tag{37}$$

It remains to rigorously bound the coefficients η_j. To achieve this, we employ an intermediate projection onto an orthonormal basis based on $\{\hat{\boldsymbol{u}}^*/(1-\epsilon), \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$. The coefficients for this

projection are given by

$$\begin{aligned}\tilde{\eta}_1 &= \boldsymbol{u}^\top \boldsymbol{d}_1 \\ &\vdots \\ \tilde{\eta}_{m-1} &= \boldsymbol{u}^\top \boldsymbol{d}_{m-1} \\ \tilde{\eta}_m &= \frac{1}{1-\epsilon}\boldsymbol{u}^\top \hat{\boldsymbol{u}}^*(\boldsymbol{x}) \\ \tilde{\eta}_{m+1} &= 1.\end{aligned} \tag{38}$$

Because the basis vectors are orthonormal and the arbitrary input satisfies the norm bound $\|\boldsymbol{u}\| \leq 1$, it inherently follows that $|\tilde{\eta}_j| \leq 1$. Moreover, it is straightforward to see that the coefficients η_j of the non-orthonormal basis can be easily found based on the intermediate coefficients $\tilde{\eta}_j$ via

$$\begin{aligned}\eta_1 &= \frac{1}{\epsilon}\tilde{\eta}_1 \\ &\vdots \\ \eta_{m-1} &= \frac{1}{\epsilon}\tilde{\eta}_{m-1} \\ \eta_m &= \tilde{\eta}_m - \frac{1-\epsilon}{\epsilon}\sum_{i=1}^{m-1}\tilde{\eta}_i \\ \eta_{m+1} &= \tilde{\eta}_{m+1} - \left(\tilde{\eta}_m - \frac{1-\epsilon}{\epsilon}\sum_{i=1}^{m-1}\tilde{\eta}_i\right) - \frac{1}{\epsilon}\sum_{i=1}^{m-1}\tilde{\eta}_i = \tilde{\eta}_{m+1} - \sum_{i=1}^{m}\tilde{\eta}_i.\end{aligned} \tag{39}$$

Therefore, it follows that

$$\begin{aligned}|\eta_1| &\leq \frac{1}{\epsilon} \\ &\vdots \\ |\eta_{m-1}| &\leq \frac{1}{\epsilon} \\ |\eta_m| &\leq m + \frac{m-1}{\epsilon} \leq \frac{2m-1}{\epsilon} \\ \eta_{m+1} &\leq m+1.\end{aligned} \tag{40}$$

Finally, substituting these absolute coefficient bounds back into the foundational expansion (37), and relaxing the term $(m+1)\left(\frac{2m-1}{\epsilon}\right)^2 \leq \frac{(m+1)^3}{\epsilon^2}$ we yield the complete global geometric inequality

$$\begin{aligned}\sigma^2(\boldsymbol{x},\boldsymbol{u}) \leq &\frac{(m+1)}{\epsilon^2}\sum_{i=1}^{m-1}\sigma^2(\boldsymbol{x},\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i) + \frac{(m+1)^3}{\epsilon^2}\sigma^2\left(\boldsymbol{x}, \frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) \\ &+ (m+1)^3\sigma^2\left(\boldsymbol{x},\boldsymbol{0}\right).\end{aligned} \tag{41}$$

□

Through a combination of the local variance reduction ensured by Lemma 5 and the basis decomposition of the variance provided by Lemma 6, we obtain the following result.

Proposition 2 (Variance Bound via Basis)**.** *Define the data set*

$$\mathbb{D}_{\text{new}} = \{([\boldsymbol{x}(t_1); \tfrac{\hat{\boldsymbol{u}}^*}{1-\epsilon}], y_1), ([\boldsymbol{x}(t_2); \hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_1], y_2), \ldots, ([\boldsymbol{x}(t_m); \hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_{m-1}], y_m)\} \tag{42}$$

with $t_i = t^* + (i-1)T_s$, $\hat{\boldsymbol{u}}$ defined in (17)*, and $\{\hat{\boldsymbol{u}}^*/(1-\epsilon), \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$ and orthonormal basis for $\mathbb{R}^m$. Let $\mathbb{D}_{N+m} = \mathbb{D}_N \cup \mathbb{D}_{\text{new}}$ and assume that* (25) *holds. If the posterior variance* (3) *for the data set $\mathbb{D}_{N+m}$ is based on differentiable stationary kernels $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$, its evaluation at $\boldsymbol{x}(t^*)$ and arbitrary $\boldsymbol{u} \in \mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^m : \|\boldsymbol{u}\| \leq 1\}$ satisfies*

$$\sigma^2(\boldsymbol{x}(t^*), \boldsymbol{u}) \leq \Upsilon \tag{43}$$

where

$$\Upsilon = (m+1)^3\sigma_{\phi,0}^2 + \frac{2(m+1)^3}{\epsilon^2}\underline{\sigma}^2 + \frac{(m+1)^3}{\epsilon^2}L_{\sigma^2}\Delta T_s \tag{44}$$

$$\underline{\sigma}^2 = \frac{\sigma_{\text{on}}^2\left(\sigma_{\phi,0}^2 + \max_{j=1,\dots,m}\sigma_{\psi_j,0}^2\right)}{\sigma_{\text{on}}^2 + \sigma_{\phi,0}^2 + \max_{j=1,\dots,m}\sigma_{\psi_j,0}^2} \tag{45}$$

and L_{σ^2} is the Lipschitz constant of the GP variance $\sigma^2(\cdot,\boldsymbol{u})$.

Proof. Due to Lemma 6, we have

$$\sigma^2(\boldsymbol{x}(t^*),\boldsymbol{u}) \leq \frac{(m+1)}{\epsilon^2}\sum_{i=1}^{m-1}\sigma^2(\boldsymbol{x}(t^*),\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i) + \frac{(m+1)^3}{\epsilon^2}\sigma^2\left(\boldsymbol{x}(t^*),\frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) + (m+1)^3\sigma^2\left(\boldsymbol{x}(t^*),\boldsymbol{0}\right). \tag{46}$$

Note that differentiability of the kernel immediately implies differentiability of the GP variance $\sigma^2(\cdot,\cdot)$. Thus, (25) allows us to bound $\sigma^2(\boldsymbol{x}(t^*),\boldsymbol{u})$ via $\sigma^2(\boldsymbol{x}(t_i),\boldsymbol{u})$ yielding

$$\sigma^2(\boldsymbol{x}(t^*),\boldsymbol{u}) \leq \frac{(m+1)}{\epsilon^2}\sum_{i=1}^{m-1}\sigma^2(\boldsymbol{x}(t_{i+1}),\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i) + iL_{\sigma^2}\Delta T_s + \frac{(m+1)^3}{\epsilon^2}\sigma^2\left(\boldsymbol{x}(t_1),\frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) + (m+1)^3\sigma^2\left(\boldsymbol{x}(t^*),\boldsymbol{0}\right) \tag{47}$$

The variance $\sigma^2(\boldsymbol{x}(t^*),\boldsymbol{0})$ can be bounded through the prior $\sigma_{\phi,0}^2$. For other variance terms, we apply Lemma 5, which yields

$$\sigma^2(\boldsymbol{x}(t^*),\boldsymbol{u}) \leq \frac{(m+1)}{\epsilon^2}\sum_{i=1}^{m-1}\frac{\sigma_{\text{on}}^2(\sigma_{\phi,0}^2 + (\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i)^\top\boldsymbol{\Sigma}_0(\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i))}{\sigma_{\text{on}}^2 + \sigma_{\phi,0}^2 + (\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i)^\top\boldsymbol{\Sigma}_0(\hat{\boldsymbol{u}}^* + \epsilon\boldsymbol{d}_i)} + iL_{\sigma^2}\Delta T_s + \frac{(m+1)^3}{\epsilon^2}\frac{\sigma_{\text{on}}^2\left(\sigma_{\phi,0}^2 + \frac{1}{(1-\epsilon)^2}(\hat{\boldsymbol{u}}^*)^\top\boldsymbol{\Sigma}_0\hat{\boldsymbol{u}}^*\right)}{\sigma_{\text{on}}^2 + \sigma_{\phi,0}^2 + \frac{1}{(1-\epsilon)^2}(\hat{\boldsymbol{u}}^*)^\top\boldsymbol{\Sigma}_0\hat{\boldsymbol{u}}^*} + (m+1)^3\sigma_{\phi,0}^2 \tag{48}$$

where we use $\sum_{j=1}^m u_j^2\sigma_{\psi_j,0}^2 = \boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u}$ for $\boldsymbol{\Sigma}_0 = \text{diag}(\sigma_{\psi_1,0}^2,\dots,\sigma_{\psi_m,0}^2)$. By bounding

$$\frac{\sigma_{\text{on}}^2(\sigma_{\phi,0}^2 + \boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u})}{\sigma_{\text{on}}^2 + \sigma_{\phi,0}^2 + \boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u}} \leq \max_{\boldsymbol{u}\in\mathcal{U}}\frac{\sigma_{\text{on}}^2(\sigma_{\phi,0}^2 + \boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u})}{\sigma_{\text{on}}^2 + \sigma_{\phi,0}^2 + \boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u}} \tag{49}$$

for all $\boldsymbol{u}\in\mathcal{U}$, we obtain a bound that is independent of the data points, i.e.,

$$\sigma^2(\boldsymbol{x}(t^*),\boldsymbol{u}) \leq (m+1)^3\sigma_{\phi,0}^2 + \left(\frac{(m+1)^2 + (m+1)^3}{\epsilon^2}\right)\max_{\boldsymbol{u}\in\mathcal{U}}\frac{\sigma_{\text{on}}^2(\sigma_{\phi,0}^2 + \boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u})}{\sigma_{\text{on}}^2 + \sigma_{\phi,0}^2 + \boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u}} + \frac{m+1}{\epsilon^2}\sum_{i=1}^{m-1}iL_{\sigma^2}\Delta T_s. \tag{50}$$

The last summand can be straightforwardly shown to satisfy

$$\frac{m+1}{\epsilon^2}\sum_{i=1}^{m-1}iL_{\sigma^2}\Delta T_s = \frac{m(m+1)(m-1)}{2\epsilon^2}L_{\sigma^2}\Delta T_s. \tag{51}$$

Moreover, since $\frac{a}{1+a}$ is a monotonically growing function for positive a, the maximizer of the optimization problem over $\boldsymbol{u}$ in (50) is given by the control input acting only in the direction of the largest eigenvalue of $\boldsymbol{\Sigma}_0$. Because $\boldsymbol{\Sigma}_0$ is a diagonal matrix, we immediately obtain the solution of

the optimization problem as

$$\max_{\boldsymbol{u}\in\mathcal{U}} \frac{\sigma_{\text{on}}^2(\sigma_{\phi,0}^2+\boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u})}{\sigma_{\text{on}}^2+\sigma_{\phi,0}^2+\boldsymbol{u}^\top\boldsymbol{\Sigma}_0\boldsymbol{u}} = \underline{\sigma}^2 \tag{52}$$

for $\underline{\sigma}^2$ defined in (45). Substituting (51) and (52) into (50) and noting that $m(m+1)(m-1) \leq (m+1)^3$, (43) follows for Υ defined in (44). □

D.4 Recursive Feasibility and Probabilistic Safety

To formally establish the global probabilistic safety of the closed-loop system (Theorem 1), we must prove its foundational prerequisite: the recursive feasibility of the control framework. Specifically, we must demonstrate that the optimization problem systematically recovers its feasibility after every active exploration event.

Since the system state continuously drifts during the physical execution of the open-loop sequence, we first introduce an auxiliary lemma that bounds the spatial variation of the GP posterior variance. This step is critical to ensure that the predictive uncertainty remains bounded despite state drift.

Lemma 7 (Temporal Variation of GP Posterior Variance)**.** *Consider the system dynamics* (1) *and assume that* (25) *holds. Let $L_{\sigma^2} > 0$ be the Lipschitz constant of the GP variance function $\sigma^2(\cdot, \boldsymbol{u})$ with respect to the state vector. Moreover, assume that the posterior variance* (3) *is based on differentiable stationary kernels $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$. Then, it satisfies*

$$\sigma^2(\boldsymbol{x}(t'), \boldsymbol{u}) \leq \sigma^2(\boldsymbol{x}(t), \boldsymbol{u}) + L_{\sigma^2}\Delta(t-t'), \quad \forall \boldsymbol{u} \in \mathcal{U}, \tag{53}$$

for all $t \geq t'$, $t, t' \in \mathbb{R}_{\geq 0}$.

Proof. Differentiability of the kernels immediately implies differentiability of the GP variance $\sigma^2(\cdot,\cdot)$ and consequently the existence of L_{σ^2}. Thus, it follows from (25) that

$$\sigma^2(\boldsymbol{x}(t'), \boldsymbol{u}) - \sigma^2(\boldsymbol{x}(t), \boldsymbol{u}) \leq L_{\sigma^2}\|\boldsymbol{x}(t') - \boldsymbol{x}(t)\| \leq L_{\sigma^2}\Delta(t-t'), \tag{54}$$

which concludes the proof. □

Together with Propositions 1 and 2, this result allows us to prove the safety of Algorithm 1 as shown in the following.

Theorem 1 (Probabilistic Safety) *Consider the continuous-time system* (1) *and a continuously differentiable CBF $h(\boldsymbol{x})$. Suppose Assumptions 1, 3 and 2 are satisfied so that*

$$\Gamma = \theta - 2\beta\sqrt{(m+1)^3\sigma_{\phi,0}^2 + \frac{2(m+1)^3}{\epsilon^2}\frac{\sigma_{\text{on}}^2\left(\sigma_{\phi,0}^2 + \max_{j=1,\ldots,m}\sigma_{\psi_j,0}^2\right)}{\sigma_{\text{on}}^2 + \sigma_{\phi,0}^2 + \max_{j=1,\ldots,m}\sigma_{\psi_j,0}^2}} > 0, \tag{55}$$

where $\sigma_{\phi,0}^2 = k_\phi(\boldsymbol{x},\boldsymbol{x})$, $\sigma_{\psi_j,0}^2 = k_{\psi_j}(\boldsymbol{x},\boldsymbol{x})$ are prior variances. Furthermore, assume that the GP mean and variance defined in (3) *are based on differentiable stationary kernels $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$. If at most $\bar{N} \in \mathbb{N}$ training samples are generated and* (6b) *does not hold at $t = 0$, there exist thresholds $\Xi_1(\bar{N}, \delta)$ defined in* (12) *and*

$$\Xi_2(\Gamma,\beta,\epsilon) = \frac{\sqrt{\beta^2\left(\frac{(m+1)^3}{m}+\epsilon^2\right)L_{\sigma^2} + (L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)\Gamma\epsilon^2} - \beta\sqrt{\left(\frac{(m+1)^3}{m}+\epsilon^2\right)L_{\sigma^2}}}{\epsilon\sqrt{m\Delta}(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)} \tag{56}$$

$$\Xi_3(T_s) = \Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s \tag{57}$$

$$\Xi_4(T_s,\beta) = (m+1)\Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s + 2\beta\sqrt{L_{\sigma^2}\Delta T_s}, \tag{58}$$

such that the control signal (7) *and the event trigger* (6) *resulting from Algorithm 1 with parameters*

$$\beta \geq \Xi_1(\bar{N},\delta), \qquad \sqrt{T_s} \leq \Xi_2(\Gamma,\beta,\epsilon), \qquad \eta_1 \geq \Xi_3(T_s) \qquad \eta_2 \geq \Xi_4(T_s,\beta) \tag{59}$$

guarantees the safety of the closed-loop system, i.e., $\boldsymbol{x}(t) \in \mathcal{C}$ for all $t \geq 0$, with probability $1-\delta$.

Proof. The event-triggered active exploration framework introduces discrete switching at the triggering instances, rendering the closed-loop control law discontinuous in time t. However, because the nominal dynamics $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{G}(\boldsymbol{x})$ are locally Lipschitz continuous, and the applied control inputs are constrained within the compact set $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^m : \|\boldsymbol{u}\| \leq 1\}$, the right-hand side of the closed-loop system (1) remains a measurable function of time t and is Lipschitz continuous with respect to the state $\boldsymbol{x}$ within the compact safe set $\mathcal{C}$. Under these classical conditions, standard nonsmooth differential equation theory guarantees the existence and uniqueness of an extended solution for the closed-loop system in the Carathéodory sense (Coddington & Levinson, 1955, Chapter 2).

To formally prove the forward invariance of the safe set $\mathcal{C}$, we employ Nagumo's Theorem (Blanchini & Miani, 2008, Theorem 4.7), which requires that the true CBF derivative satisfies $c(\boldsymbol{x}(t), \boldsymbol{u}(t)) = \phi(\boldsymbol{x}(t)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(t) \geq 0$ for all $t \geq 0$. In the following, we will prove that $c(\boldsymbol{x}(t), \boldsymbol{u}(t)) \geq 0$ holds for each time interval $[t_k, t_k + T_s)$. Note that all following statements jointly hold with probability $1 - \delta$ due to their reliance on Lemma 2, but we do not explicitly state this for ease of reading.

For $t = 0$, the control input $\boldsymbol{u}(0)$ is defined by (4) since the triggering condition is assumed not to be satisfied. Hence, we have

$$\hat{\phi}(\boldsymbol{x}(0)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(0))\boldsymbol{u}(t) - \beta\sigma(\boldsymbol{x}(0), \boldsymbol{u}(0)) - \eta_1 \geq 0 \tag{60}$$

by construction. Since $\beta \geq \Xi_1(\bar{N}, \delta)$ for $\Xi_1(\bar{N}, \delta)$ in (12), it follows from Lemma 2 that $c(\boldsymbol{x}(0), \boldsymbol{u}(0)) \geq \eta_1$. Due to the restriction to a compact safe set in Assumption 1, the Lipschitz continuity of $\boldsymbol{f}(\cdot)$ and $\boldsymbol{G}(\cdot)$ in combination with the boundedness of $\boldsymbol{u}$ imply a boundedness of the right-hand side of (1), which immediately guarantees the existence of a constant Δ satisfying (25). Moreover, the Lipschitz constants L_ϕ and $\boldsymbol{L}_{\boldsymbol{\psi}}$ are guaranteed to exist due to the bounded RKHS norm of L_ϕ and $\boldsymbol{L}_{\boldsymbol{\psi}}$ (Lederer et al., 2026b). Thus, we have

$$\begin{aligned}\phi(\boldsymbol{x}(t)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t))\boldsymbol{u}(0) &\geq \phi(\boldsymbol{x}(0)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(0))\boldsymbol{u}(0) - \Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s &(61)\\ &\geq \eta_1 - \Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s, &(62)\end{aligned}$$

for all $t \in [0, T_s)$. Note that the right-hand side is lower bounded by 0 since $\eta_1 \geq \Xi_3(T_s)$ with $\Xi_3(T_s)$ in (57), such that $c(\boldsymbol{x}(t), \boldsymbol{u}(0)) \geq 0$ for all $t \in [0, T_s)$.

For all subsequent time intervals, we consider three different cases depending on the recent activation of the event trigger:

1. If there is no trigger at t_k and no trigger event occurred during the $m-1$ time steps before, the satisfaction of $c(\boldsymbol{x}(t), \boldsymbol{u}(t_k)) \geq 0$ for all $t \in [t_k, t_k + T_s)$ follows analogously to the time interval $[0, T_s)$.

2. If a trigger occurs at t_k, we can distinguish two sub-cases:

   (a) In case the last trigger activation occurred more than m time steps ago, it holds that

   $$3\beta \max_{\|\boldsymbol{u}\| \leq 1} \sigma(\boldsymbol{x}(t_{k-1}), \boldsymbol{u}) < \hat{\phi}(\boldsymbol{x}(t_{k-1})) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_{k-1}))\boldsymbol{u}^* - \eta_2 \tag{63}$$

   since the trigger condition was not satisfied at t_{k-1}. Using Lemma 2, we have

   $$\begin{aligned}\hat{\phi}(\boldsymbol{x}(t_{k-1})) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_{k-1}))\boldsymbol{u}^* &\leq \phi(\boldsymbol{x}(t_{k-1})) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t_{k-1}))\boldsymbol{u}^*\\ &\quad + \beta \max_{\|\boldsymbol{u}\| \leq 1} \sigma(\boldsymbol{x}(t_{k-1}), \boldsymbol{u}) &(64)\end{aligned}$$

   due to the lower bound $\Xi_1(\bar{N}, \delta)$ defined in (12). Using the same continuity properties as before, we map it to time step k, i.e.,

   $$\begin{aligned}\hat{\phi}(\boldsymbol{x}(t_{k-1})) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_{k-1}))\boldsymbol{u}^* &\leq \phi(\boldsymbol{x}(t_k)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t_k))\boldsymbol{u}^* + \beta \max_{\|\boldsymbol{u}\| \leq 1} \sigma(\boldsymbol{x}(t_k), \boldsymbol{u})\\ &\quad + \Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s + \beta\sqrt{L_{\sigma^2}\Delta T_s}. &(65)\end{aligned}$$

We apply Lemma 2 again, which yields

$$\hat{\phi}(\boldsymbol{x}(t_{k-1})) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_{k-1}))\boldsymbol{u}^* \leq \hat{\phi}(\boldsymbol{x}(t_k)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_k))\boldsymbol{u}^* + 2\beta \max_{\|\boldsymbol{u}\|\leq 1} \sigma(\boldsymbol{x}(t_k), \boldsymbol{u}) + \Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s + \beta\sqrt{L_{\sigma^2}\Delta T_s}. \tag{66}$$

Using Lipschitz continuity of the GP variance, this finally results in

$$\beta \max_{\|\boldsymbol{u}\|\leq 1} \sigma(\boldsymbol{x}(t_k), \boldsymbol{u}) < \hat{\phi}(\boldsymbol{x}(t_k)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_k))\boldsymbol{u}^* - \eta_2 + \Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s + 2\beta\sqrt{L_{\sigma^2}\Delta T_s}. \tag{67}$$

As $\eta_2 \geq \Xi_4(T_s, \beta)$ for $\Xi_4(T_s, \beta)$ specified in (58), this implies

$$\beta \max_{\|\boldsymbol{u}\|\leq 1} \sigma(\boldsymbol{x}(t_k), \boldsymbol{u}) \leq \max_{\|\boldsymbol{u}\|\leq 1-\epsilon} \hat{\phi}(\boldsymbol{x}(t_k)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_k))\boldsymbol{u} - \tilde{\theta} \tag{68}$$

with $\tilde{\theta} \geq m\Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s$, i.e., (18), (26) are satisfied. Thus, $c(\boldsymbol{x}(t), \boldsymbol{u}(t_k)) \geq 0$ for $t \in [t_k, t_k + T_s)$ follows from Proposition 1.

(b) In case the last trigger activation happened exactly m time steps ago, it follows from Proposition 2 and the same continuity arguments as before that

$$\max_{\|\boldsymbol{u}\|\leq 1} \sigma^2(\boldsymbol{x}(t_k), \boldsymbol{u}) \leq \Upsilon + mL_{\sigma^2}\Delta T_s. \tag{69}$$

Moreover, observe that

$$\max_{\|\boldsymbol{u}\|\leq 1-\epsilon} \hat{\phi}(\boldsymbol{x}(t_k)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_k))\boldsymbol{u} \geq \hat{\phi}(\boldsymbol{x}(t_k)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_k))\boldsymbol{u}^* \tag{70}$$

$$\geq \phi(\boldsymbol{x}(t_k)) + \boldsymbol{\psi}^\top(\boldsymbol{x}(t_k))\boldsymbol{u}^* - \beta\sigma(\boldsymbol{x}(t_k), \boldsymbol{u}^*) \tag{71}$$

$$\geq \theta - \beta\sigma(\boldsymbol{x}(t_k), \boldsymbol{u}^*) \tag{72}$$

$$\geq \theta - \beta\sqrt{\Upsilon + mL_{\sigma^2}\Delta T_s} \tag{73}$$

with probability $1 - \delta$ for $\boldsymbol{u}^*$ defined as

$$\boldsymbol{u}^* = \arg\max_{\|\boldsymbol{u}\|\leq 1-\epsilon} \phi(\boldsymbol{x}) + \boldsymbol{\psi}^\top(\boldsymbol{x})\boldsymbol{u}, \tag{74}$$

in (74), where the first line follows from optimality, the second line follows from Lemma 2, and the third line is due to Assumption 3. Therefore, we have

$$\max_{\|\boldsymbol{u}\|\leq 1-\epsilon} \hat{\phi}(\boldsymbol{x}(t_k)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t_k))\boldsymbol{u} - \beta \max_{\|\boldsymbol{u}\|\leq 1-\epsilon} \sigma(\boldsymbol{x}(t_k), \boldsymbol{u}) \geq \theta - 2\beta\sqrt{\Upsilon + mL_{\sigma^2}\Delta T_s}. \tag{75}$$

Using (55), we can further refine this expression to obtain

$$\theta - 2\beta\sqrt{\Upsilon + mL_{\sigma^2}\Delta T_s} \geq \Gamma - 2\beta\sqrt{\frac{(m+1)^3}{\epsilon^2}L_{\sigma^2}\Delta T_s + mL_{\sigma^2}\Delta T_s}. \tag{76}$$

It immediately follows from $\sqrt{T_s} \leq \Xi_2(\Gamma, \beta, \epsilon)$ with $\Xi_2(\Gamma, \beta, \epsilon)$ defined in (56) that

$$\Gamma - 2\beta\sqrt{\frac{(m+1)^3}{\epsilon^2}L_{\sigma^2}\Delta T_s + mL_{\sigma^2}\Delta T_s} \geq m\Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s \tag{77}$$

holds since $\Xi_2(\Gamma, \beta, \epsilon)$ corresponds to the positive root of this quadratic function. Note that this root is guaranteed to be positive since $\Gamma > 0$ by assumption. As a consequence, (18), (26) are satisfied, such that $c(\boldsymbol{x}(t), \boldsymbol{u}(t_k)) \geq 0$ for $t \in [t_k, t_k + T_s)$ follows from Proposition 1.

3. If the trigger event occurred less than $m - 1$ time steps ago, $c(\boldsymbol{x}(t), \boldsymbol{u}(t_k)) \geq 0$ for $t \in [t_k, t_k + T_s)$ follows analogous to case 2. from Proposition 1.

As these cases cover all possible time intervals, it follows that $c(\boldsymbol{x}(t), \boldsymbol{u}(t)) \geq 0$ holds for all $t \geq 0$, such that Nagumo's theorem (Blanchini & Miani, 2008, Theorem 4.7) guarantees forward invariance of the set $\mathcal{C}$ with probability $1 - \delta$, such that safety is guaranteed with probability $1 - \delta$. □

D.5 Sample Complexity of Event-Triggered Learning

To bound the total number of triggers, we quantify the minimum joint information gain achieved during each execution of the exploration sequence. We begin by bounding the sum of the predictive variances over the exploration directions.

Lemma 8 (Lower Bound on Basis Variance Sum)**.** *Let* $\{\hat{\boldsymbol{u}}^*/(1-\epsilon), \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$ *with* $\hat{\boldsymbol{u}}^*$ *defined in* (17) *form an orthonormal basis for* $\mathbb{R}^m$*. If the posterior variance* (3) *is based on stationary kernels* $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$*, it holds that*

$$\sigma^2_{\phi,0} + \sigma^2\left(\boldsymbol{x}, \frac{\hat{\boldsymbol{u}}^*}{1-\epsilon}\right) + \sum_{j=1}^{m-1} \sigma^2(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_j) \geq C_{\text{basis}} \max_{\|\boldsymbol{u}\|\leq 1} \sigma^2(\boldsymbol{x}, \boldsymbol{u}), \tag{78}$$

where

$$C_{\text{basis}} = \frac{\epsilon^2}{(m+1)^3} \tag{79}$$

Proof. Due to Lemma 6, it holds that

$$\begin{aligned}\max_{\|\boldsymbol{u}\|\leq 1} \sigma^2(\boldsymbol{x}, \boldsymbol{u}) \leq & \frac{m+1}{\epsilon^2} \sum_{i=1}^{m-1} \sigma^2(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) + \frac{(m+1)^3}{\epsilon^2} \sigma^2\left(\boldsymbol{x}, \frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) \\ & + (m+1)^3 \sigma^2(\boldsymbol{x}, \boldsymbol{0}).\end{aligned}$$

The variance $\sigma(\boldsymbol{x}, 0)$ is bounded by its prior $\sigma^2_{\phi,0}$. Therefore, we obtain

$$\begin{aligned}\max_{\|\boldsymbol{u}\|\leq 1} \sigma^2(\boldsymbol{x}, \boldsymbol{u}) \leq & \frac{m+1}{\epsilon^2} \sum_{i=1}^{m-1} \sigma^2(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) + \frac{(m+1)^3}{\epsilon^2} \sigma^2\left(\boldsymbol{x}, \frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) \\ & + (m+1)^3 \sigma^2_{\phi,0}.\end{aligned}$$

We simplify the bound by extracting the maximum constant multiplier among the terms, which yields

$$\max_{\|\boldsymbol{u}\|\leq 1} \sigma^2_{N_k}(\boldsymbol{x}, \boldsymbol{u}) \leq \frac{(m+1)^3}{\epsilon^2} \left(\sum_{i=1}^{m-1} \sigma^2(\boldsymbol{x}, \hat{\boldsymbol{u}}^* + \epsilon \boldsymbol{d}_i) + \sigma^2\left(\boldsymbol{x}, \frac{1}{1-\epsilon}\hat{\boldsymbol{u}}^*\right) + \sigma^2_{\phi,0} \right). \tag{80}$$

since $\epsilon \in (0,1)$. Multiplying this expression by C_{basis} as defined in (79) yields the lemma. □

Next, we establish the relationship between this variance sum and the information gain, explicitly taking into account the continuous-time state drift over the time window necessary for data collection.

Lemma 9 (Batch Information Gain with State Drift Penalty)**.** *Suppose Assumptions 1, 3 and 2 are satisfied. Let* $\hat{\boldsymbol{u}}$ *be defined as in* (17)*, and* $\{\hat{\boldsymbol{u}}^*/(1-\epsilon), \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{m-1}\}$ *an orthonormal basis for* $\mathbb{R}^m$*. Define the batch dataset*

$$\mathcal{B} = \{(\boldsymbol{x}^{(i)}, \boldsymbol{u}^{(i)})_{i=1}^m\} \quad \textit{with} \quad \boldsymbol{x}^{(i)} = \boldsymbol{x}(t^* + (i-1)T_s), \boldsymbol{u}^{(1)} = \tfrac{\boldsymbol{u}^*}{1-\epsilon}, \boldsymbol{u}^{(i)} = \boldsymbol{u}^* + \epsilon \boldsymbol{d}_{i-1}. \tag{81}$$

Assume that (25) *holds and let the posterior variance* (3) *for a data set* $\mathbb{D}_N$ *be based on differentiable stationary kernels* $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$*. If* (6b) *holds at* t^* *for* β *satisfying* (12)*, the batch information gain* $I_k(\mathcal{B})$ *satisfies*

$$I_k(\mathcal{B}) \geq \frac{1}{2} \log\left(1 + \sigma_{\text{on}}^{-2}\left(C_{\text{basis}}\left(\frac{\theta - \eta_2}{4\beta}\right)^2 - \frac{(m-1)m}{2} L_{\sigma^2} \Delta T_s - \sigma^2_{\phi,0}\right)\right), \tag{82}$$

with probability $1-\delta$*, where*

$$I_k(\mathcal{B}) = \frac{1}{2}\log\det\left(\boldsymbol{I}_m + \sigma_{\text{on}}^{-2}\boldsymbol{K}_N\right) \tag{83}$$

$$\boldsymbol{K}_{N,ij} = \begin{bmatrix}1 & (\boldsymbol{u}^{(i)})^\top\end{bmatrix}\boldsymbol{\Sigma}(\boldsymbol{x}^{(i)},\boldsymbol{x}^{(j)})\begin{bmatrix}1\\ \boldsymbol{u}^{(j)}\end{bmatrix} \tag{84}$$

and, with slight abuse of notation,

$$\boldsymbol{\Sigma}(\boldsymbol{x},\boldsymbol{x}') = \begin{bmatrix} k_\phi(\boldsymbol{x},\boldsymbol{x}') & 0 & \cdots & 0\\ 0 & k_{\psi_1}(\boldsymbol{x},\boldsymbol{x}') & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & k_{\psi_m}(\boldsymbol{x},\boldsymbol{x}')\end{bmatrix} \tag{85}$$

$$- \begin{bmatrix}\boldsymbol{k}_\phi^\top(\boldsymbol{x})\\ \boldsymbol{k}_{\psi_1}^\top(\boldsymbol{x})\boldsymbol{U}_1\\ \vdots\\ \boldsymbol{k}_{\psi_m}^\top(\boldsymbol{x})\boldsymbol{U}_m\end{bmatrix}(\boldsymbol{K}+\sigma_{\text{on}}^2\boldsymbol{I})^{-1}\begin{bmatrix}\boldsymbol{k}_\phi^\top(\boldsymbol{x}')\\ \boldsymbol{k}_{\psi_1}^\top(\boldsymbol{x}')\boldsymbol{U}_1\\ \vdots\\ \boldsymbol{k}_{\psi_m}^\top(\boldsymbol{x}')\boldsymbol{U}_m\end{bmatrix}^\top. \tag{86}$$

Proof. To establish a lower bound for the information gain $I_k(\mathcal{B})$ in (83), we apply the determinant-trace inequality resulting in

$$\det\left(\boldsymbol{I}_m + \sigma_{\text{on}}^{-2}\boldsymbol{K}_N\right) \geq 1 + \sigma_{\text{on}}^{-2}\text{Tr}(\boldsymbol{K}_N). \tag{87}$$

Because the trace of a matrix is strictly defined as the sum of its diagonal elements, we explicitly substitute the trace with the summation of the GP variances

$$\text{Tr}(\boldsymbol{K}_N) = \sum_{i=1}^{m}\sigma^2(\boldsymbol{x}^{(i)},\boldsymbol{u}^{(i)}). \tag{88}$$

Note that differentiability of the kernel immediately implies differentiability of the GP variance $\sigma^2(\cdot,\cdot)$. Thus, (25) allows us to bound $\sigma^2(\boldsymbol{x}^{(i)},\boldsymbol{u}^{(i)})$ via $\sigma^2(\boldsymbol{x}^{(1)},\boldsymbol{u}^{(i)})$ yielding

$$\sum_{i=1}^{m}\sigma^2(\boldsymbol{x}^{(i)},\boldsymbol{u}^{(i)}) \geq \sum_{i=1}^{m}\sigma^2(\boldsymbol{x}^{(1)},\boldsymbol{u}^{(i)}) - \frac{(m-1)m}{2}L_{\sigma^2}\Delta T_s. \tag{89}$$

By applying Lemma 8, the sum of the variances can be bounded in terms of the maximum variance

$$\sum_{i=1}^{m}\sigma^2(\boldsymbol{x}^{(1)},\boldsymbol{u}^{(i)}) \geq C_{\text{basis}}\max_{\|\boldsymbol{u}\|\leq 1}\sigma^2(\boldsymbol{x}^{(1)},\boldsymbol{u}) - \sigma_{\phi,0}^2. \tag{90}$$

Since (6b) holds at t^*, we have

$$\max_{\|\boldsymbol{u}\|\leq 1}\sigma(\boldsymbol{x}^{(1)},\boldsymbol{u}) \geq \frac{\hat{\phi}(\boldsymbol{x}(t)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t))\boldsymbol{u}^* - \eta_2}{3\beta}. \tag{91}$$

Moreover, due to Lemma 2 and Assumption 3, we have

$$\hat{\phi}(\boldsymbol{x}(t)) + \hat{\boldsymbol{\psi}}^\top(\boldsymbol{x}(t))\boldsymbol{u}^* \geq \theta - \beta\max_{\|\boldsymbol{u}\|\leq 1}\sigma(\boldsymbol{x}^{(1)},\boldsymbol{u}) \tag{92}$$

with probability $1-\delta$ analogously to (73), such that

$$\frac{4}{3}\max_{\|\boldsymbol{u}\|\leq 1}\sigma(\boldsymbol{x}^{(1)},\boldsymbol{u}) \geq \frac{\theta-\eta_2}{3\beta} \tag{93}$$

holds with probability $1-\delta$. Finally, substituting (88), (89), (90), and (93) into (87), the lemma is proven. □

With a lower bound on the informational contribution of each distinct exploration phase, we conclude the theoretical analysis by proving the sample complexity of Algorithm 1.

Theorem 2 (Sample Complexity) *Consider the continuous-time system* (1) *and a continuously differentiable CBF* $h(\boldsymbol{x})$*. Suppose Assumptions 1, 3 and 2 are satisfied so that* (55) *holds. Furthermore, assume that the GP mean and variance defined in* (3) *are based on differentiable stationary kernels* $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \dots, k_{\psi_m}(\cdot,\cdot)$*. Assume that control signal* (7) *and the event trigger* (6) *result from Algorithm 1 with parameters satisfying* (59) *and*

$$T_s \leq \tilde{\Xi}_2(\beta, \eta_2, \epsilon, \theta), \qquad \sqrt{T_s} \leq \hat{\Xi}_2^2(\beta, \epsilon, \theta), \qquad \eta_2 \leq \tilde{\Xi}_4(\beta, \epsilon, \theta) \tag{94}$$

for

$$\tilde{\Xi}_2(\beta, \eta_2, \epsilon, \theta) = \frac{C_{\text{basis}}(\tilde{\Xi}_4(\beta, \epsilon, \theta) - \eta_2)^2}{16\beta^2 m(m-1)L_{\sigma^2}\Delta} \tag{95}$$

$$\hat{\Xi}_2(\beta, \epsilon, \theta) = \frac{\sqrt{((m+1)\Delta(L_\phi + \|\boldsymbol{L_\psi}\|))^2 + 8\beta\sqrt{L_{\sigma^2}\Delta}\tilde{\Xi}_4(\beta, \epsilon, \theta)} - (m+1)\Delta(L_\phi + \|\boldsymbol{L_\psi}\|)}{4\beta\sqrt{L_{\sigma^2}\Delta}} \tag{96}$$

$$\tilde{\Xi}_4(\beta, \epsilon, \theta) = \theta - \frac{2\beta\sigma_{\phi,0}}{\sqrt{C_{\text{basis}}}}. \tag{97}$$

If Algorithm 1 is initialized with $\mathbb{D}_0 = \emptyset$*, the number of trigger events* κ *is bounded by*

$$\bar{\kappa} = \min_{\kappa \in \mathbb{N}} \kappa \qquad \text{such that } \kappa \geq 2\beta^2 \gamma_{\kappa m} \left(\frac{32\sigma_{\text{on}}^2}{C_{\text{basis}} \left(\tilde{\Xi}_4(\beta, \epsilon, \theta) - \eta_2 \right)^2} + 1 \right) = c(\theta - \eta_2, \epsilon)\beta^2 \gamma_{\kappa m} \tag{98}$$

with probability $1 - \delta$*, where* $\gamma_{\kappa m}$ *is the maximum information gain defined in* (13)*.*

Proof. Due to Theorem 1, safety is probabilistically guaranteed, i.e., $\boldsymbol{x}(t) \in \mathcal{C}$ for all $t \geq 0$ with probability $1 - \delta$. Therefore, all generated state training samples lie in the compact space $\mathcal{C}$ with probability $1 - \delta$. Since this implies that all subsequent results hold jointly with probability $1 - \delta$, we streamline the presentation and subsequently do not explicitly mention the probability anymore.

For bounding the sample complexity, we exploit the fundamental property of Gaussian process regression that the sum of the isolated mutual information gains acquired during each individual exploration batch is strictly upper-bounded by this maximum joint capacity (Srinivas et al., 2012)

$$\sum_{i=1}^{\kappa} I_{(i-1)m}(\mathcal{B}_i) \leq \gamma_{\kappa m}, \tag{99}$$

where $\gamma_{\kappa m}$ is the maximum information gain defined in (13) and $\mathcal{B}_i$ are batch datasets (81). Thus, the number of update triggers κ is bounded by

$$\bar{\kappa} \geq \inf_{\kappa \in \mathbb{R}_{\geq 0}} \kappa \qquad \text{such that } \sum_{i=1}^{\kappa} I_{(i-1)m}(\mathcal{B}_i) \geq \gamma_{\kappa m}. \tag{100}$$

Due to Lemma 9, we can tighten the constraint to

$$\frac{\kappa}{2} \log\left(1 + \sigma_{\text{on}}^{-2} \left(C_{\text{basis}} \left(\frac{\theta - \eta_2}{4\beta} \right)^2 - \frac{(m-1)m}{2} L_{\sigma^2} \Delta T_s - \sigma_{\phi,0}^2 \right)\right) \geq \gamma_{\kappa m}. \tag{101}$$

In order to ensure that the infimum in (100) is attainable, we need to show that the argument of the logarithm is larger or equal to 1. Using (97), we can lower bound this argument by

$$\begin{aligned} C_{\text{basis}} \left(\frac{\theta - \eta_2}{4\beta} \right)^2 - \frac{(m-1)m}{2} L_{\sigma^2} \Delta T_s - \sigma_{\phi,0}^2 \\ \geq C_{\text{basis}} \left(\frac{\tilde{\Xi}_4(\beta, \epsilon, \theta) - \eta_2}{4\beta} \right)^2 - \frac{(m-1)m}{2} L_{\sigma^2} \Delta T_s. \end{aligned} \tag{102}$$

Note that (97) is compatible with (58) since

$$(m+1)\Delta(L_\phi + \|\boldsymbol{L}_{\boldsymbol{\psi}}\|)T_s + 2\beta\sqrt{L_{\sigma^2}\Delta T_s} \leq \tilde{\Xi}_4(\beta,\epsilon,\theta) \tag{103}$$

is ensured to hold for $\sqrt{T_s} \leq \hat{\Xi}_2(\beta,\epsilon,\theta)$ with $\hat{\Xi}_2(\beta,\epsilon,\theta)$ defined in (96). Moreover, $T_s \leq \tilde{\Xi}_2(\beta,\eta_2,\epsilon,\theta)$ for $\tilde{\Xi}_2(\beta,\eta_2,\epsilon,\theta)$ in (95) simplifies this bound to

$$\frac{\kappa}{2}\log\left(1 + \frac{C_{\text{basis}}}{2\sigma_{\text{on}}^2}\left(\frac{\tilde{\Xi}_4(\beta,\epsilon,\theta) - \eta_2}{4\beta}\right)^2\right) \geq \gamma_{\kappa m} \tag{104}$$

Since $\log(1+c) \geq 1 - \frac{1}{1+c} = \frac{c}{1+c}$, we can further tighten the bound to obtain the constraint

$$\kappa \geq 2\gamma_{\kappa m}\left(\frac{32\beta^2\sigma_{\text{on}}^2}{C_{\text{basis}}\left(\tilde{\Xi}_4(\beta,\epsilon,\theta) - \eta_2\right)^2} + 1\right), \tag{105}$$

which can be simplified to obtain (98). □

D.6 Sample Complexity with SE Base Kernels

By restricting our analysis to a specific covariance function, e.g., the squared exponential kernel, we can upper bound the solution to (98). The following lemma upper bounds the maximum information gain of our control-affine kernel when it is based on a squared exponential kernel to enable this more specific analysis.

Lemma 10. *If the posterior variance* (3) *for the data set $\mathbb{D}_N$ is based on differentiable stationary kernels $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$, the maximum information gain γ_N satisfies*

$$\gamma_N \in \tilde{\mathcal{O}}\left((m+1)\log(N)^{n+1}\right). \tag{106}$$

Proof. The kernel k is an additive combination of m kernels of the form $u_i k_{\psi_i}(\boldsymbol{x},\boldsymbol{x}')u_i'$ and one squared exponential kernel $k_\phi(\boldsymbol{x},\boldsymbol{x}')$. Hence, it follows from Krause & Ong (2011, Thm. 3) that

$$\gamma_N \leq 2m\log(N) + \gamma_{k_\phi,N} + \sum_{i=1}^{m}\gamma_{u_i k_{\psi_i}(\boldsymbol{x},\boldsymbol{x}')u_i',N}.$$

where $\gamma_{k_\phi,N}$ and $\gamma_{u_i k_{\psi_i}(\boldsymbol{x},\boldsymbol{x}')u_i',N}$ represent the maximum information gain for kernels $k_\phi(\cdot,\cdot)$ and $u_i k_{\psi_i}(\boldsymbol{x},\boldsymbol{x}')u_i'$. Since $u_i k_{\psi_i}(\boldsymbol{x},\boldsymbol{x}')u_i'$ is a product kernel between a scalar linear and a squared exponential covariance function, it follows from Krause & Ong (2011, Thm. 2) that

$$\gamma_{u_i k_{\psi_i}(\boldsymbol{x},\boldsymbol{x}')u_i',N} \leq \gamma_{k_{\psi_i},N} + \log(N).$$

We combine both bounds to obtain

$$\gamma_N \leq 3m\log(N) + \gamma_{k_\phi,N} + \sum_{i=1}^{m}\gamma_{k_{\psi_i},N}.$$

Since the information gain satisfies $\gamma_{k_\phi,N} \in \tilde{\mathcal{O}}(\log(N)^{n+1})$, $\gamma_{k_{\psi_i},N} \in \mathcal{O}(\log(N)^{n+1})$ for the squared exponential kernels (Srinivas et al., 2012), it immediately follows that (106) holds. □

Given this bound for the maximum information gain, we obtain the asymptotics for the SE base kernel as follows.

Corollary 1 (Bound for Squared Exponential Kernels)**.** *Assume that $k_\phi(\cdot,\cdot), k_{\psi_1}(\cdot,\cdot), \ldots, k_{\psi_m}(\cdot,\cdot)$ are squared exponential kernels and let*

$$\vartheta = \tilde{\Xi}_4(\beta,\epsilon,\theta) - \eta_2, \qquad \beta \in \mathcal{O}(\Xi_1(\bar{\kappa}m,\delta)). \tag{107}$$

For $\vartheta > 0$, the solution of (98) *satisfies*

$$\bar{\kappa} \in \tilde{\mathcal{O}}\left(\frac{1}{\vartheta^2\epsilon^2}\right). \tag{108}$$

Proof. By definition $\Xi_1(\bar{\kappa}m, \delta) \in \tilde{\mathcal{O}}(\sqrt{\gamma_{\bar{N}}})$. Thus, we have $\beta^2 \in \tilde{\mathcal{O}}(\gamma_{\bar{\kappa}m})$ due to (107) and consequently

$$c(\theta - \eta_2, \epsilon)\beta^2\gamma_{\kappa m} \in \mathcal{O}(c(\theta - \eta_2, \epsilon)\gamma_{\kappa m}^2) \subset \mathcal{O}\left(\frac{\gamma_{\kappa m}^2}{\vartheta^2\epsilon^2}\right), \tag{109}$$

where the set inclusion follows from the definition of $c(\theta - \eta_2, \epsilon)$ in (98) and (107). Using Lemma 10, we can resolve the dependency on the maximum information gain, yielding

$$c(\theta - \eta_2, \epsilon)\beta^2\gamma_{\kappa m} \in \mathcal{O}\left(\frac{\log(\kappa m)^{2n+2}}{\vartheta^2\epsilon^2}\right), \tag{110}$$

where we drop dependencies on m where possible. Due to the definition of $\mathcal{O}$ notation, this implies the existence of a constant $C \in \mathbb{R}_{>0}$ such that

$$c(\theta - \eta_2, \epsilon)\beta^2\gamma_{\kappa m} \leq \frac{C\log(\kappa m)^{2n+2}}{\vartheta^2\epsilon^2}. \tag{111}$$

Since $b\log(x) = x^{1/a}$ has a solution $x = (-abW(-\frac{1}{ab}))^a$, $a = 2n+2$, $b = \left(\frac{m}{\vartheta^2\epsilon^2}\right)^{1/(2n+2)}$ (Wol), where $W(\cdot)$ is the Lambert W function, we have

$$\kappa m \leq \frac{Cm(m+1)^2}{\vartheta^2\epsilon^2}\left(-2(n+1)W_{-1}\left(-\frac{1}{2(n+1)\left(\frac{Cm(m+1)^2}{\vartheta^2\epsilon^2}\right)^{1/(2n+2)}}\right)\right)^{2n+2} \tag{112}$$

where $W_{-1}(\cdot)$ is the lower branch of the Lambert W function. Thus, it immediately follows that

$$\kappa \in \mathcal{O}\left(-\tfrac{1}{\vartheta^2\epsilon^2}W_{-1}\left(-\frac{1}{2(n+1)\left(\frac{Cm}{\vartheta^2\epsilon^2}\right)^{1/(2n+2)}}\right)^{2n+2}\right). \tag{113}$$

Since the lower branch of the Lambert W function can be bounded by $W(x) \geq \log(-x) - \sqrt{2(1-\log(-x))} \geq 3\log(-x)$ for sufficiently small x (Chatzigeorgiou, 2013), we have

$$\kappa \in \mathcal{O}\left(-\tfrac{1}{\vartheta^2\epsilon}\log\left(\frac{1}{2(n+1)\left(\frac{Cm}{\vartheta^2\epsilon}\right)^{1/(2n+2)}}\right)^{2n+2}\right). \tag{114}$$

Dropping poly-log terms and noting $\log(1/x) = -\log(x)$, we directly obtain (108) □

E Experimental Setup

E.1 Quadrotor

To evaluate the proposed event-triggered active exploration framework, we construct a continuous-time quadrotor navigation environment subject to unmodeled physical disturbances. The system state is $\boldsymbol{x} = [\boldsymbol{p}^\top, \boldsymbol{v}^\top, \boldsymbol{q}^\top]^\top \in \mathbb{R}^{10}$, comprising the global position $\boldsymbol{p} \in \mathbb{R}^3$, linear velocity $\boldsymbol{v} \in \mathbb{R}^3$, and attitude parameterized by a unit quaternion $\boldsymbol{q} \in \mathbb{R}^4$. The full control input is nominally $\boldsymbol{u} = [T, \boldsymbol{\omega}_{cmd}^\top]^\top \in \mathbb{R}^4$, where T is the thrust acceleration and $\boldsymbol{\omega}_{cmd} = [\omega_x, \omega_y, \omega_z]^\top$ is the commanded body-frame angular velocity.

The true evolution of the quadrotor follows the control-affine dynamics $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{G}(\boldsymbol{x})\boldsymbol{u}$, formulated as (Mahony et al., 2012)

$$\dot{\boldsymbol{p}} = \boldsymbol{v} \tag{115}$$

$$\dot{\boldsymbol{v}} = -g\boldsymbol{e}_z + \boldsymbol{a}_{wind} + \eta(\boldsymbol{v})\boldsymbol{R}(\boldsymbol{q})\boldsymbol{e}_z T \tag{116}$$

$$\dot{\boldsymbol{q}} = \frac{1}{2}\boldsymbol{Q}(\boldsymbol{q})\boldsymbol{M}\boldsymbol{\omega}_{cmd} \tag{117}$$

where g is the gravitational acceleration, $\boldsymbol{e}_z = [0, 0, 1]^\top$, $\boldsymbol{R}(\boldsymbol{q}) \in SO(3)$ is the rotation matrix, and $\boldsymbol{Q}(\boldsymbol{q}) \in \mathbb{R}^{4\times3}$ is the quaternion kinematics matrix.

The environment introduces three physical disturbances: 1. A steady wind disturbance modeled as a constant additive acceleration $\boldsymbol{a}_{wind} \in \mathbb{R}^3$. 2. An aerodynamic efficiency decay reducing effective thrust at high speeds, modeled as $\eta(\boldsymbol{v}) = 1 - \min\left(\eta_{max}, (\|\boldsymbol{v}\|/v_{scale})^2\right)$. 3. A motor mixing cross-coupling matrix $\boldsymbol{M} \in \mathbb{R}^{3\times3}$, which introduces efficiency loss on the diagonal elements and unmodeled cross-axis coupling on the off-diagonal elements.

The nominal dynamics $\tilde{\boldsymbol{f}}(\boldsymbol{x})$ and $\tilde{\boldsymbol{G}}(\boldsymbol{x})$ used by the probabilistic safety filter assume $\boldsymbol{a}_{wind} = \boldsymbol{0}$, $\eta(\boldsymbol{v}) = 1$, and $\boldsymbol{M} = \boldsymbol{I}_{3\times3}$. The structural differences between the true functions and these nominal approximations constitute the residual dynamics $\Delta\phi(\boldsymbol{x})$ and $\Delta\boldsymbol{\psi}(\boldsymbol{x})$ that the GP evaluates online.

To enforce spatial safety, the quadrotor must avoid an infinitely tall vertical cylindrical obstacle. Let the obstacle be defined by its horizontal centroid $\boldsymbol{p}_{obs,xy} \in \mathbb{R}^2$ and radius R. Defining the relative horizontal position vector $\Delta\boldsymbol{p}_{xy} = [p_x - p_{obs,x}, p_y - p_{obs,y}, 0]^\top$, the safe set $\mathcal{C}$ is synthesized using the following extended CBF

$$h(\boldsymbol{x}) = 2\Delta\boldsymbol{p}_{xy}^\top\boldsymbol{v} + \alpha_{\text{qr}}(\|\Delta\boldsymbol{p}_{xy}\|^2 - R^2) - \lambda_{\text{qr}}\left(1 - \frac{\Delta\boldsymbol{p}_{xy}^\top\boldsymbol{z}_B}{R}\right) \tag{118}$$

where $\boldsymbol{z}_B = \boldsymbol{R}(\boldsymbol{q})\boldsymbol{e}_z$ is the body vertical axis, and $\alpha_{\text{qr}}, \lambda_{\text{qr}} > 0$ are design parameters. The nominal CBF constraint functions $\tilde{\phi}(\boldsymbol{x})$ and $\tilde{\boldsymbol{\psi}}(\boldsymbol{x})$ are derived directly by taking the Lie derivatives of $h(\boldsymbol{x})$ along the nominal dynamics.

Because the obstacle is a vertical cylinder, a rotation purely around the vertical body axis (yaw) does not instantaneously change the horizontal distance or the attitude penalty. Analytically, the coefficient corresponding to the yaw rate ω_z in the control mapping vector $\boldsymbol{\psi}(\boldsymbol{x})$ evaluates identically to zero. Due to this zero-coefficient property, the ω_z input cannot contribute to satisfying the CBF constraint. Consequently, ω_z is explicitly excluded from the GP residual learning and the active exploration basis. The effective control input space monitored and regulated by the learning framework is strictly reduced to $\boldsymbol{u}_{eff} = [T, \omega_x, \omega_y]^\top \in \mathbb{R}^3$.

E.2 Autonomous Underwater Vehicle (AUV)

To evaluate the proposed event-triggered active exploration framework in tightly confined environments, we construct a continuous-time autonomous underwater vehicle (AUV) navigation simulation subject to unmodeled physical disturbances. The system state is defined as $\boldsymbol{x} = [\boldsymbol{\eta}^\top, \boldsymbol{\nu}^\top]^\top \in \mathbb{R}^{12}$, comprising the global position and Euler angles $\boldsymbol{\eta} = [x, y, z, \phi, \theta, \psi]^\top \in \mathbb{R}^6$, and the body-frame linear and angular velocities $\boldsymbol{\nu} = [u, v, w, p, q, r]^\top \in \mathbb{R}^6$. The full control input is nominally $\boldsymbol{u} = [\tau_X, \tau_Y, \tau_Z, \tau_K, \tau_M, \tau_N]^\top \in \mathbb{R}^6$, representing the three-dimensional forces and moments applied in the body frame.

The true evolution of the AUV follows the control-affine dynamics $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{G}(\boldsymbol{x})\boldsymbol{u}$, formulated as (Fossen, 2011)

$$\dot{\boldsymbol{\eta}} = \boldsymbol{J}(\boldsymbol{\eta})\boldsymbol{\nu} \tag{119}$$

$$\boldsymbol{M}\dot{\boldsymbol{\nu}} + \boldsymbol{C}(\boldsymbol{\nu})\boldsymbol{\nu} + \boldsymbol{D}(\boldsymbol{\nu})\boldsymbol{\nu} + \boldsymbol{g}(\boldsymbol{\eta}) = \boldsymbol{\tau}_{act} \tag{120}$$

where $\boldsymbol{J}(\boldsymbol{\eta}) \in \mathbb{R}^{6\times6}$ is the Jacobian transformation matrix, $\boldsymbol{M} \in \mathbb{R}^{6\times6}$ is the inertia matrix, $\boldsymbol{C}(\boldsymbol{\nu}) \in \mathbb{R}^{6\times6}$ is the Coriolis and centripetal matrix, $\boldsymbol{D}(\boldsymbol{\nu}) \in \mathbb{R}^{6\times6}$ is the nonlinear hydrodynamic damping matrix, and $\boldsymbol{g}(\boldsymbol{\eta}) \in \mathbb{R}^6$ represents the restoring forces from gravity and buoyancy. The term $\boldsymbol{\tau}_{act}$ denotes the actual generalized forces produced by the actuators.

The environment introduces three unmodeled physical disturbances to the actual actuator output ($\boldsymbol{\tau}_{act} = \boldsymbol{\Gamma}_{bio}\boldsymbol{\Gamma}_{spd}(\boldsymbol{\nu})\boldsymbol{M}_{err}\boldsymbol{u}$): a constant biofouling efficiency decay modeled as a diagonal matrix $\boldsymbol{\Gamma}_{bio} \in \mathbb{R}^{6\times6}$; a speed-dependent thrust decay capturing the nonlinear loss of effective thrust at high velocities, formulated as a state-dependent diagonal matrix $\boldsymbol{\Gamma}_{spd}(\boldsymbol{\nu}) \in \mathbb{R}^{6\times6}$ parameterized by a hyperbolic tangent function; and a motor mixing cross-coupling matrix $\boldsymbol{M}_{err} \in \mathbb{R}^{6\times6}$, which introduces unmodeled cross-axis rotational coupling.

Table 3: Quantiles of the minimum trigger interval for our method on the AUV benchmark. The distribution is right-skewed, so these statistics complement the mean and standard deviation reported in Table 1.

Metric	10% Quantile	Median	90% Quantile
Min Trigger Interval (steps)	11	43	1649

The nominal dynamics $\tilde{\boldsymbol{f}}(\boldsymbol{x})$ and $\tilde{\boldsymbol{G}}(\boldsymbol{x})$ used by the probabilistic safety filter assume an idealized environment devoid of disturbances, obtained by substituting $\boldsymbol{\Gamma}_{bio} = \boldsymbol{I}_{6\times6}$, $\boldsymbol{\Gamma}_{spd}(\boldsymbol{\nu}) = \boldsymbol{I}_{6\times6}$, and $\boldsymbol{M}_{err} = \boldsymbol{I}_{6\times6}$. The structural differences between the true functions and these nominal approximations constitute the residual dynamics $\Delta\phi(\boldsymbol{x})$ and $\Delta\boldsymbol{\psi}(\boldsymbol{x})$ that the GP evaluates online.

To enforce spatial safety, the AUV must navigate through a sinusoidal tunnel without colliding with the walls. Instead of relying on a priori global analytical models, the CBF is constructed dynamically using simulated local sensor measurements. We assume the AUV is equipped with sensors capable of measuring the minimum distance to the tunnel's central axis, denoted by $\boldsymbol{p}_c \in \mathbb{R}^3$, and estimating the local tangent vector of the tunnel, denoted by $\boldsymbol{t}_d \in \mathbb{R}^3$. Defining the relative position vector $\Delta\boldsymbol{p} = \boldsymbol{\eta}_{1:3} - \boldsymbol{p}_c$ and the expanded tunnel radius R, the safe set $\mathcal{C}$ is synthesized using the following High-Order CBF (HOCBF) that integrates both spatial boundaries and attitude alignment

$$h(\boldsymbol{x}) = (R^2 - \|\Delta\boldsymbol{p}\|^2) - \lambda_{\text{auv}}(1 - \boldsymbol{x}_B^\top \boldsymbol{t}_d) \tag{121}$$

where $\boldsymbol{x}_B \in \mathbb{R}^3$ is the longitudinal axis of the vehicle expressed in the world frame, and $\lambda_{\text{auv}} > 0$ is the attitude penalty parameter. Because the control inputs (forces and moments) appear in the second derivative of this function, we apply a differentiation-based cascade $\ddot{h}(\boldsymbol{x}, \boldsymbol{u}) + (\alpha_{\text{auv1}} + \alpha_{\text{auv2}})\dot{h}(\boldsymbol{x}) + \alpha_{\text{auv1}}\alpha_{\text{auv2}}h(\boldsymbol{x}) \geq 0$. The nominal CBF constraint parameters $\tilde{\phi}(\boldsymbol{x})$ and $\tilde{\boldsymbol{\psi}}(\boldsymbol{x})$ are derived directly from this cascade along the nominal dynamics.

Because the attitude penalty strictly aligns the vehicle's longitudinal axis with the tunnel's tangent, a rotation purely around this longitudinal body axis (roll) does not instantaneously change the spatial distance or the attitude alignment. Analytically, the coefficient corresponding to the roll torque τ_K in the control mapping vector $\boldsymbol{\psi}(\boldsymbol{x})$ evaluates identically to zero. Due to this zero-coefficient property, the τ_K input cannot contribute to satisfying the CBF constraint. Consequently, τ_K is explicitly excluded from the GP residual learning and the active exploration basis. The effective control input space monitored and regulated by the learning framework is strictly reduced to $\boldsymbol{u}_{eff} = [\tau_X, \tau_Y, \tau_Z, \tau_M, \tau_N]^\top \in \mathbb{R}^5$.

E.3 Additional Discussion on AUV Trigger Intervals

Table 1 reports the minimum trigger interval on the AUV benchmark as 501.4 ± 972.4 simulation steps for our method. In the confined AUV environment, the triggering behavior is highly non-uniform across trajectories: some runs require relatively frequent early updates when the vehicle enters narrow regions, whereas other runs exhibit very long trigger-free segments after the GP uncertainty has been sufficiently reduced. As a result, the empirical distribution is strongly right-skewed, making the mean and standard deviation less representative of the typical run.

To clarify this behavior, we additionally report distributional statistics for the AUV minimum trigger interval in Table 3. The median value is 43 steps, while the 10% and 90% quantiles are 11 and 1649 steps, respectively. Thus, although the average is inflated by runs with very long trigger-free segments, the typical minimum interval remains substantially larger than immediate one-step re-triggering. This indicates that the proposed structured exploration does not rely on repeatedly triggering updates at every step; instead, after an exploratory sequence reduces uncertainty in multiple control directions, the GP-CBF-SOCP can often remain feasible for extended periods. Together with the zero CBF violations and zero infeasibilities reported in Table 1, these results suggest that the large standard deviation reflects heterogeneous AUV trajectory difficulty rather than unstable or pathological triggering behavior.

E.4 Implementation Details and Hyperparameters

This section provides the numerical and software details used in the experiments of Section 5.1. All reported trajectories are generated by simulation; we do not use external datasets, pretrained models, or newly introduced assets. Panda-specific models, control and learning parameters, data collection, and task generation are described in Appendix G.

Gaussian process residual model. We implement the control-affine GP model using GPyTorch (Gardner et al., 2018). All GP computations use double precision. The scalar residual CBF derivative is modeled as

$$\Delta H(\boldsymbol{x}, \boldsymbol{u}) = \Delta\phi(\boldsymbol{x}) + \sum_{i=1}^{m} \Delta\psi_i(\boldsymbol{x}) u_i, \tag{122}$$

where $\Delta\phi(\cdot)$ and $\Delta\psi_i(\cdot)$ are independent zero-mean exact GPs with ARD squared-exponential kernels. This induces the composite control-affine kernel used in Section 3.1. The state and target values are normalized using the means and standard deviations in Tables 5 and 6. Control inputs are first mapped to the normalized action coordinates used by the safety filter; no additional control scaling is applied inside the GP handler. GP hyperparameters are fixed to the values in the configuration tables. The Gaussian likelihood noise is set according to the noise standard deviation σ_{noise} reported in Tables 5 and 6. For numerical stability, a Cholesky jitter of 10^{-6} is added when forming the posterior cache. During each triggered exploration episode, all new samples from the structured sequence are added and the GP posterior is recomputed as a batch update. Note that this has no impact on the theoretical guarantees and the performance of Algorithm 1 since the GP is not used during exploration, i.e., when exploratory control inputs $\boldsymbol{u}_t$ in line 8 of Algorithm 1 do not change with GP model updates.

Normalized action coordinates. The theoretical development uses a unit Euclidean input ball for the construction of the structured exploration sequence, whereas the simulated robotic systems are subject to componentwise physical actuator limits. These two descriptions are reconciled by applying an affine offset-scaling map before the safety filter and the GP model. For a physical control input $\boldsymbol{u}^{\text{phys}}$ with componentwise bounds $\boldsymbol{u}_{\min}^{\text{phys}} \le \boldsymbol{u}^{\text{phys}} \le \boldsymbol{u}_{\max}^{\text{phys}}$, we define

$$\boldsymbol{o} = \frac{\boldsymbol{u}_{\max}^{\text{phys}} + \boldsymbol{u}_{\min}^{\text{phys}}}{2}, \qquad \boldsymbol{s} = \frac{\boldsymbol{u}_{\max}^{\text{phys}} - \boldsymbol{u}_{\min}^{\text{phys}}}{2},$$

and use the normalized input

$$\bar{\boldsymbol{u}} = \operatorname{diag}(\boldsymbol{s})^{-1} \left(\boldsymbol{u}^{\text{phys}} - \boldsymbol{o}\right), \qquad \boldsymbol{u}^{\text{phys}} = \boldsymbol{o} + \operatorname{diag}(\boldsymbol{s})\bar{\boldsymbol{u}}.$$

Thus, the physical actuator box is mapped to the normalized box

$$\bar{\mathcal{U}}_\infty = \{\bar{\boldsymbol{u}} \in \mathbb{R}^m : \|\bar{\boldsymbol{u}}\|_\infty \le 1\}.$$

The structured exploration sequence is then constructed inside the normalized Euclidean ball

$$\bar{\mathcal{U}}_2 = \{\bar{\boldsymbol{u}} \in \mathbb{R}^m : \|\bar{\boldsymbol{u}}\|_2 \le 1\}.$$

Since $\|\bar{\boldsymbol{u}}\|_2 \le 1$ implies $\|\bar{\boldsymbol{u}}\|_\infty \le 1$, we have $\bar{\mathcal{U}}_2 \subseteq \bar{\mathcal{U}}_\infty$. Consequently, every exploratory control input satisfying the theoretical ℓ_2 constraint also satisfies the normalized box constraint and, after denormalization, lies inside the physical actuator limits.

The same affine change of coordinates is applied to the CBF constraint. If the physical-coordinate CBF derivative is written as $\phi^{\text{phys}}(\boldsymbol{x}) + (\boldsymbol{\psi}^{\text{phys}}(\boldsymbol{x}))^\top \boldsymbol{u}^{\text{phys}}$, then in normalized coordinates it becomes

$$\bar{\phi}(\boldsymbol{x}) + \bar{\boldsymbol{\psi}}^\top(\boldsymbol{x})\bar{\boldsymbol{u}}, \qquad \bar{\phi}(\boldsymbol{x}) = \phi^{\text{phys}}(\boldsymbol{x}) + (\boldsymbol{\psi}^{\text{phys}}(\boldsymbol{x}))^\top \boldsymbol{o}, \qquad \bar{\boldsymbol{\psi}}(\boldsymbol{x}) = \operatorname{diag}(\boldsymbol{s})\boldsymbol{\psi}^{\text{phys}}(\boldsymbol{x}).$$

This is the coordinate system used by the GP safety filter and by the exploration basis. The nonzero offset is particularly relevant for the quadrotor thrust channel, whose physical bound is not symmetric, while symmetric bounds such as angular-rate or AUV force limits correspond to the special case $\boldsymbol{o} = \boldsymbol{0}$.

Safety-filter optimization. The GP-CBF safety filter is implemented in CVXPY (Diamond & Boyd, 2016) and solved as a second-order cone program using Clarabel (Goulart & Chen, 2024).

The standard SOCP minimizes the squared deviation from the nominal controller subject to the tightened GP-CBF constraint and box constraints on the normalized control input. If the standard problem fails to return a feasible control, we solve a relaxed fallback problem with a nonnegative slack variable δ and penalty $10^6\delta^2$. The slack is used only as a numerical fallback and is counted as an infeasibility event in the reported metrics.

Nominal controllers. For the quadrotor benchmark, the nominal controller is a cascaded position–attitude PID controller. Given the goal position $\boldsymbol{p}_{\text{goal}}$, it first computes the desired acceleration

$$\boldsymbol{a}_d = \boldsymbol{K}_p^p(\boldsymbol{p}_{\text{goal}} - \boldsymbol{p}) + \boldsymbol{K}_i^p\boldsymbol{\xi}_p - \boldsymbol{K}_d^p\boldsymbol{v} + g\boldsymbol{e}_z,$$

where $\boldsymbol{\xi}_p$ is the componentwise saturated integral of the position error. The commanded thrust is

$$T_{\text{nom}} = \text{sat}_{[0,20]}\left(\boldsymbol{a}_d^\top \boldsymbol{z}_B\right), \qquad \boldsymbol{z}_B = \boldsymbol{R}(\boldsymbol{q})\boldsymbol{e}_z.$$

The desired attitude $\boldsymbol{R}_d$ is chosen such that its body z-axis aligns with $\boldsymbol{a}_d/\|\boldsymbol{a}_d\|$, using a fixed reference heading. The angular-rate command is

$$\boldsymbol{\omega}_{\text{nom}} = \text{sat}_{[-5,5]}\left(\boldsymbol{K}_p^R \,\text{Log}\big(\boldsymbol{R}(\boldsymbol{q})^\top \boldsymbol{R}_d\big)\right).$$

We use $\boldsymbol{K}_p^p = (1,1,2)$, $\boldsymbol{K}_i^p = (0.1,0.1,0.5)$, $\boldsymbol{K}_d^p = (3,3,5)$, and $\boldsymbol{K}_p^R = (5,5,2)$.

For the AUV benchmark, we use a cascaded 6-DOF PID controller. Let $\boldsymbol{p}_{\text{goal}} = (100,0,0)^\top$ and $\psi_{\text{goal}} = 0$. The desired inertial velocity is

$$\boldsymbol{v}_d^I = \text{sat}_{[-2,2]}\left(\boldsymbol{K}_p^p(\boldsymbol{p}_{\text{goal}} - \boldsymbol{p})\right),$$

which is transformed into the body frame as $\boldsymbol{v}_d^B = \boldsymbol{R}(\boldsymbol{\eta})^\top \boldsymbol{v}_d^I$. The desired angular velocity is

$$\boldsymbol{\omega}_d = \text{sat}_{[-0.5,0.5]}\left(\boldsymbol{K}_p^\eta \begin{bmatrix} \text{ssa}(-\phi) \\ \text{ssa}(-\theta) \\ \text{ssa}(\psi_{\text{goal}} - \psi) \end{bmatrix}\right),$$

where $\text{ssa}(\cdot)$ wraps angles to the principal interval. With $\boldsymbol{\nu}_d = [(\boldsymbol{v}_d^B)^\top, \boldsymbol{\omega}_d^\top]^\top$, the nominal generalized force command is

$$\boldsymbol{\tau}_{\text{nom}} = \text{sat}\left(\boldsymbol{K}_p^\nu(\boldsymbol{\nu}_d - \boldsymbol{\nu}) + \boldsymbol{K}_i^\nu\boldsymbol{\xi}_\nu\right),$$

where $\boldsymbol{\xi}_\nu$ is the componentwise saturated integral of the velocity error. We use $\boldsymbol{K}_p^p = (0.5,0.5,0.5)$, $\boldsymbol{K}_p^\eta = (1,1,1)$, $\boldsymbol{K}_p^\nu = (30,30,30,20,20,20)$, and $\boldsymbol{K}_i^\nu = (5,5,5,2,2,2)$.

Evaluation metrics. For each trajectory i that completes the task without CBF violations, we compute the actual path length as

$$L_{\text{actual}}^{(i)} = \sum_{k=0}^{T_i-1} \left\|\boldsymbol{p}_{k+1}^{(i)} - \boldsymbol{p}_k^{(i)}\right\|_2,$$

where $\boldsymbol{p}_k^{(i)}$ denotes the position at the k-th simulation step. The nominal reference length is the straight-line start-to-end distance

$$L_{\text{nominal}}^{(i)} = \left\|\boldsymbol{p}_{\text{end}}^{(i)} - \boldsymbol{p}_0^{(i)}\right\|_2.$$

The reported mean path overhead is

$$\text{Overhead}_{\text{mean}} = \frac{1}{N_{\text{safe}}} \sum_{i=1}^{N_{\text{safe}}} \frac{L_{\text{actual}}^{(i)}}{L_{\text{nominal}}^{(i)}},$$

where N_{safe} is the number of safe trajectories included in the statistic. Thus, $L_{\text{nominal}}^{(i)}$ is a geometric straight-line reference rather than the length of the trajectory generated by the nominal controller.

Table 4: Implementation-level parameters shared by the main benchmark experiments. Note that the GP prediction is not employed by Algorithm 1 while the exploratory control sequence is applied, such that batch updates do not affect our theoretical results and empirical results.

Item	**Value**
Programming language	Python 3.10.20
Floating-point precision	64-bit floating point
GP library	GPyTorch 1.15.2
Optimization interface	CVXPY 1.7.5
SOCP solver	Clarabel via CVXPY
GP posterior factorization	Cholesky factorization with jitter 10^{-6}
Maximum Cholesky size	4000
Exploration basis construction	Gram–Schmidt orthogonalization
GP update mode	Batch update after triggered exploration sequence
Random seeds	Generated as `np.random.randint(0,1000000,size=total_tests)`

Table 5: Quadrotor benchmark parameters used in Section 5.1.

Category	**Parameter**	**Value**
Simulation	sampling/integration step	$0.01\,\mathrm{s}$
Simulation	horizon	$20\,\mathrm{s}$
State/control	state dimension	10
State/control	full control input	$[T, \omega_x, \omega_y, \omega_z]^\top$
State/control	effective learned input	$[T, \omega_x, \omega_y]^\top$, dimension 3
State/control	thrust bound	$T \in [0, 20]\,\mathrm{m/s^2}$
State/control	angular-rate bound	$\omega_i \in [-5, 5]\,\mathrm{rad/s}$
Obstacle/CBF	obstacle radius	1.5
Obstacle/CBF	CBF parameters	$\lambda_{\mathrm{qr}} = 0.25$, $\alpha_{\mathrm{qr}} = 1.5$, $\gamma_{\mathrm{qr}} = 2.5$
True-model disturbance	steady wind acceleration	$(-0.5, 0.5, 0)$
True-model disturbance	motor mixing parameters	efficiency 0.9, coupling factor 0.1
True-model disturbance	thrust-efficiency decay rate	0.1
Nominal model	disturbances	wind, thrust-efficiency decay, and motor mixing omitted
Initial data	initial sampling steps / size	30 steps / 10 samples
GP scaling	state mean	$\mathbf{0}_{10}$
GP scaling	state std.	$(10, 10, 10, 10, 10, 10, 1, 1, 1, 1)$
GP scaling	target mean/std.	0 / 10
GP kernel	offset lengthscale/std.	$\ell_0 = 1$, $s_0 = 0.2$
GP kernel	coefficient lengthscales	$\ell_c = (1, 2, 2)$
GP kernel	coefficient prior stds.	$s_c = (1.25, 0.02, 0.02)$
GP noise/confidence	noise std.	$\sigma_{\mathrm{noise}} = 10^{-2}$
GP noise/confidence	failure probability	$\delta = 0.01$
GP noise/confidence	confidence multipliers	$\beta_0 = 5$, $\beta_c = (5, 5, 5)$
Structured exploration	exploration magnitude	$\epsilon = 0.2$
Structured exploration	tightening constants	$\eta_1 = 0.5$, $\eta_2 = 0.5$
Randomized trials	start/goal sampling	start outside obstacle with margin $d_{\mathrm{margin}} = 3$
Randomized trials	radial sampling scale	$\mathtt{mul_r} = 1.5$

For the AUV benchmark, the motor-mixing matrix used in the true actuator model is

$$\boldsymbol{M}_{err} = \begin{bmatrix} 1.0 & 0.1 & 0.2 & 0.2 & 0.1 & 0.3 \\ 0.1 & 1.0 & 0.1 & 0.2 & 0.2 & 0.4 \\ 0.2 & 0.1 & 1.0 & 0.1 & 0.3 & 0.1 \\ 0.1 & 0.2 & 0.2 & 1.0 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.3 & 0.1 & 1.0 & 0.1 \\ 0.2 & 0.3 & 0.1 & 0.2 & 0.1 & 1.0 \end{bmatrix}. \tag{123}$$

Table 8 reports elapsed wall-clock time for each batch of 100 randomized trials. The quadrotor and AUV batches use the compute environment in Table 7. The listed batches total $7685.922\,\mathrm{s}$ ($2.135\,\mathrm{h}$). These are batch execution times rather than per-control-step latencies.

Table 6: AUV benchmark parameters used in Section 5.1.

Category	Parameter	Value
Simulation	sampling/integration step	$0.02\,\mathrm{s}$
Simulation	horizon	$100\,\mathrm{s}$
State/control	state dimension	12
State/control	full control input	$[\tau_X, \tau_Y, \tau_Z, \tau_K, \tau_M, \tau_N]^\top$
State/control	effective learned input	$[\tau_X, \tau_Y, \tau_Z, \tau_M, \tau_N]^\top$, dimension 5
State/control	force bounds	$\tau_X, \tau_Y, \tau_Z \in [-100, 100]\,\mathrm{N}$
State/control	moment bounds	$\tau_K, \tau_M, \tau_N \in [-50, 50]\,\mathrm{Nm}$
Task	terminal x coordinate	100
Task	target yaw	0
Tunnel	radius	$R = 5$
Tunnel	randomized amplitude	$A = R(1 + 0.5\rho)$, $\rho \sim \mathcal{U}(0, 1)$
Tunnel	randomized period	$100(0.5 + 0.5\rho)$
Tunnel	center height	$z_c = 0$
CBF	HOCBF parameters	$\alpha_{\text{auv1}} = 1.5$, $\alpha_{\text{auv2}} = 1.5$, $\lambda_{\text{auv}} = 0.5$
Rigid-body model	mass	$30\,\mathrm{kg}$
Rigid-body model	weight and buoyancy	$W = B = 30 \cdot 9.81$
Rigid-body model	center-of-gravity offset	$\boldsymbol{r}_{bg} = (0, 0, 0.02)$
Rigid-body model	rigid-body inertia	$\mathrm{diag}(30, 30, 30, 1.5, 3, 3)$
Rigid-body model	added mass	$\mathrm{diag}(3, 15, 15, 0.5, 2, 2)$
Rigid-body model	linear damping	$\mathrm{diag}(5, 15, 15, 2, 5, 5)$
True-model disturbance	additive ocean-current force	$(5, 15, 20, 5, -5, 5)$
True-model disturbance	biofouling efficiency	$(0.8, 0.75, 0.85, 0.8, 0.9, 0.75)$
True-model disturbance	maximum speed-dependent decay	$(0.05, 0.1, 0.15, 0.15, 0.3, 0.15)$
True-model disturbance	velocity scales	$(1.5, 1.5, 1.5, 0.5, 0.5, 0.5)$
Nominal model	disturbances	actuator decay and motor mixing omitted
Initial data	initial sampling steps / size	75 steps / 25 samples
GP scaling	state mean	$\mathbf{0}_{12}$
GP scaling	state std.	$(100, 10, 5, 1, 1, 3.14, 3, 3, 3, 1, 1, 1)$
GP scaling	target mean/std.	0 / 2.75
GP kernel	offset lengthscale/std.	$\ell_0 = 1$, $s_0 = 0.85$
GP kernel	coefficient lengthscales	$\ell_c = (2, 2, 2, 2, 2)$
GP kernel	coefficient prior stds.	$s_c = (1, 0.9, 0.9, 0.02, 0.3)$
GP noise/confidence	noise std.	$\sigma_{\text{noise}} = 10^{-2}$
GP noise/confidence	failure probability	$\delta = 0.01$
GP noise/confidence	confidence multipliers	$\beta_0 = 5$, $\beta_c = (5, 5, 5, 5, 5)$
Structured exploration	exploration magnitude	$\epsilon = 0.3$
Structured exploration	tightening constants	$\eta_1 = 5$, $\eta_2 = 5$
Randomized trials	initial radial offset	$r_0 \sim \mathcal{U}(0.05R, 0.75R)$
Randomized trials	initial roll/pitch	independently sampled from $\pm 2.5^\circ$
Randomized trials	initial yaw perturbation	tangent yaw plus $\pm 10^\circ$

F Additional Ablation on Action Scaling

In this appendix, we provide additional ablations on the action scale used in the quadrotor experiments. These parameters are not algorithmic parameters in the theoretical analysis. Instead, they change the physical actuation limits before the inputs are normalized in the safety filter, and should therefore be interpreted as practical stress tests on the available control authority.

The first row of Fig. 5 varies the maximum thrust-acceleration scale $a_T^{\max}$. This ablation should not be interpreted as a monotone increase of symmetric control authority, since the quadrotor thrust channel is nonnegative and affects the translational dynamics differently from the angular-rate channels. Increasing $a_T^{\max}$ can provide more vertical acceleration authority, but it can also make the closed-loop motion more aggressive and increase the state displacement between consecutive safety-filter updates. As a result, the CBF-violation curves exhibit method-dependent regimes rather than a single monotone trend. For example, Time Trigger can show larger violation counts at intermediate thrust scales while improving again at the two ends of the tested range, because its fixed update schedule interacts with the induced closed-loop trajectory. Bandit is comparatively insensitive to

Table 7: Software dependencies and compute resources. The GPU was available on the machine but was not used for the reported experiments.

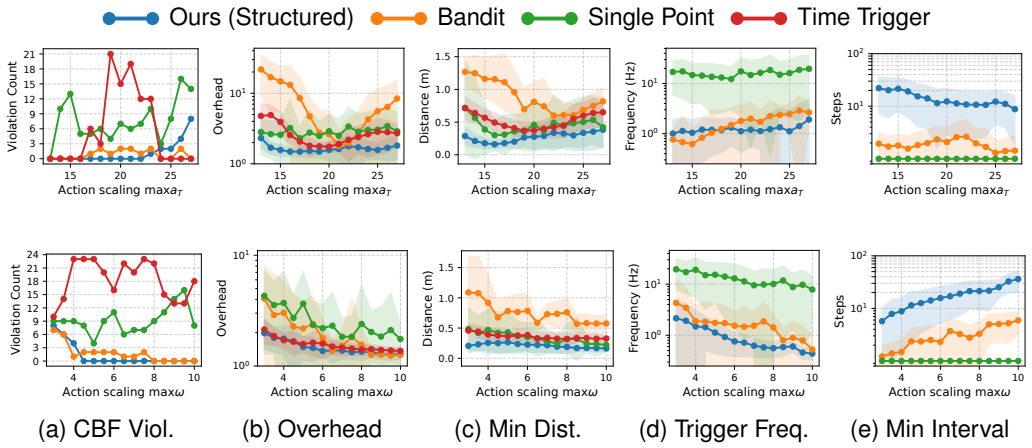
Item	Version / specification	Role
PyTorch	2.11.0	tensor computations
GPyTorch	1.15.2, MIT license	GP residual model
CVXPY	1.7.5, Apache-2.0 license	convex optimization interface
Clarabel	Apache-2.0 license	SOCP solver
NumPy	2.4.4	numerical arrays
SciPy	1.17.1	rotations and numerical utilities
PythonVehicleSimulator	1.0.1, MIT license	AUV GNC utilities
Pandas	3.0.2	result aggregation
Matplotlib	3.10.9	plotting
Operating system	Ubuntu 22.04	runtime environment
CPU	AMD Ryzen 9 7945HX, 2.50 GHz	reported experiments
Memory	32 GB RAM	reported experiments
GPU	NVIDIA RTX 4060 Laptop GPU, 8 GB	available but not used
Machine	Lenovo Legion R9000P laptop 2023	local compute worker

Table 8: Wall-clock runtimes for the reported randomized evaluations. The GPU was available but not used; all runs were executed on the local CPU worker reported in Table 7.

Experiment	Method	100-trial runtime (s)	Mean per trial (s)
Quadrotor	Ours	457.3	4.57
Quadrotor	Bandit	476.9	4.77
Quadrotor	Single Point	739.4	7.39
Quadrotor	Time-Triggered	323.7	3.24
AUV	Ours	1157.5	11.58
AUV	Bandit	1298.6	12.99
AUV	Single Point	1939.2	19.39
AUV	Time-Triggered	882.1	8.82
Panda	Closed-Form CBF	233.1	2.33
Panda	Point-Wise CBF Eval.	178.2	1.78

$a_T^{\max}$ in this experiment, since its exploration rule is mainly driven by recovering SOCP feasibility rather than by systematically covering all scaled input directions. Single Point remains more variable because it reduces uncertainty only along a single locally safe direction. Our structured method stays close to zero violations over the tested range, with only isolated increases under more aggressive thrust scaling. Thus, the apparent non-monotonicity in Fig. 5 is not a contradiction, but reflects the interaction between physical action scaling, closed-loop motion, and the different exploration mechanisms. Finally, the ablation curves are computed from finite randomized rollouts and use parameter values different from the default setting in Table 1; small discrepancies in low-count CBF-violation statistics are therefore expected.

The second row varies the maximum angular-rate scale $\omega_{\max}$, which more directly affects the quadrotor's ability to reorient and compensate for model uncertainty. When $\omega_{\max}$ is small, all methods have limited corrective authority, and CBF violations can occur. Once sufficient angular-rate authority is available, our structured exploration achieves zero CBF violations and keeps the path overhead low. In contrast, Single Point and Time Trigger continue to show violations over most values of $\omega_{\max}$, indicating that increased actuation alone is not sufficient without an exploration mechanism that reduces uncertainty in the relevant input directions. Moreover, our trigger frequency decreases and the minimum trigger interval increases as $\omega_{\max}$ grows, suggesting that improved angular authority makes it easier to maintain feasible safety-filter operation over longer intervals.



(a) CBF Viol. (b) Overhead (c) Min Dist. (d) Trigger Freq. (e) Min Interval

Figure 5: Additional action-scaling ablation on the maximum thrust-acceleration scale $a_T^{\max}$ and the maximum angular-rate scale $\omega_{\max}$. The columns report CBF violations, path overhead, minimum distance to obstacles, trigger frequency, and minimum trigger interval.

Table 9: Fixed transforms in the Panda simulation model.

Joint	Translation t_i (m)	Rotation α_i (rad)
1	$(0, 0, 0.333)$	0
2	$(0, 0, 0)$	$-\pi/2$
3	$(0, -0.316, 0)$	$\pi/2$
4	$(0.0825, 0, 0)$	$\pi/2$
5	$(-0.0825, 0.384, 0)$	$-\pi/2$
6	$(0, 0, 0)$	$\pi/2$
7	$(0.088, 0, 0)$	$\pi/2$

G Experimental Setup for Panda Obstacle Avoidance

G.1 Manipulator and Task Model

We simulate a fixed-base, seven-degree-of-freedom Franka Panda manipulator in MuJoCo (Todorov et al., 2012). The state is $x = [q^\top, \dot{q}^\top]^\top \in \mathbb{R}^{14}$ and the physical control input is the joint torque $u \in \mathbb{R}^7$. Joint positions, velocities, and torques are expressed in rad, rad/s, and N·m, respectively. All seven joints participate in safety filtering and structured exploration.

Without contact, the nominal and true simulation models satisfy

$$M_s(q)\ddot{q} + C_s(q, \dot{q})\dot{q} + g_s(q) + D\dot{q} = u, \qquad s \in \{\mathrm{n}, \mathrm{t}\}. \tag{124}$$

Here, M_s is the mass matrix, $C_s\dot{q}$ contains Coriolis and centrifugal terms, g_s is gravity, and $D = 0.1I_7$ is viscous damping in N·m·s/rad. Each joint has an additional armature inertia of 0.1 kg·m^2, so $M_s = M_{\mathrm{rigid},s} + 0.1I_7$. Gravity is $(0, 0, -9.81)^\top$ m/s^2. The two models share geometry, damping, and armature parameters.

Model mismatch is introduced by a payload present only in the true model: a uniform cube of mass 1 kg and side length 0.08 m, whose center is at $(0, 0, 0.05)$ m in the flange frame. Its inertia about its center is $1.0666667 \times 10^{-3} I_3$ kg·m^2. This load changes inertial, velocity-dependent, and gravitational terms.

All joints rotate around their local z axes. Successive link transforms are ${}^{i-1}T_i(q_i) = T(t_i)R_x(\alpha_i)R_z(q_i)$, with the fixed transforms in Table 9. The translation from link 7 to the flange is $(0, 0, 0.107)$ m. The constrained end-effector point is at $(0, 0, 0.05)$ m relative to the flange, or $(0, 0, 0.157)$ m relative to link 7. No gripper degree of freedom is included. Writing its position as $p(q)$ and its translational Jacobian as $J(q) = \partial p/\partial q$ gives $\dot{p} = J\dot{q}$ and $\ddot{p} = J\ddot{q} + \dot{J}\dot{q}$.

Table 10: Link masses and centers of mass.

Link	Mass (kg)	Center of mass (m)
1	4.970684	$(0.003875, 0.002081, -0.047620)$
2	0.646926	$(-0.003141, -0.028720, 0.003495)$
3	3.228604	$(0.027518, 0.039252, -0.066502)$
4	3.587895	$(-0.053170, 0.104419, 0.027454)$
5	1.225946	$(-0.011953, 0.041065, -0.038437)$
6	1.666555	$(0.060149, -0.014117, -0.010517)$
7	0.735522	$(0.010517, -0.004252, 0.061597)$

Table 11: Link inertias in kg·m^2.

Link	I_{xx}	I_{yy}	I_{zz}	I_{xy}	I_{xz}	I_{yz}
1	0.703370	0.706610	0.009117	-0.000139	0.006772	0.019169
2	0.007962	0.028110	0.025995	-0.003925	0.010254	0.000704
3	0.037242	0.036155	0.010830	-0.004761	-0.011396	-0.012805
4	0.025853	0.019552	0.028323	0.007796	-0.001332	0.008641
5	0.035549	0.029474	0.008627	-0.002117	-0.004037	0.000229
6	0.001964	0.004354	0.005433	0.000109	-0.001158	0.000341
7	0.012516	0.010027	0.004815	-0.000428	-0.001196	-0.000741

Link masses, centers of mass, and inertias are listed in Tables 10 and 11. Centers of mass use the corresponding link frame. Inertias are taken about the center of mass and expressed along the link-frame axes; off-diagonal entries are the actual elements of the inertia matrix.

Joint-position limits follow the official Panda description (Franka Robotics). We use torque limits of 60 N·m for joints 1–4 and 10 N·m for joints 5–7, below the corresponding hardware reference limits of 87 and 12 N·m. Table 12 specifies these ranges. The position coordinates of x use these joint limits, with initial and target states further restricted by the task-generation conditions in Appendix G.4.

Let $S_u = \mathrm{diag}(60, 60, 60, 60, 10, 10, 10)$ and $\bar{u} = S_u^{-1} u$. Both the learning safety filter and structured exploration use $\|\bar{u}\|_2 \le 1$, corresponding to the physical torque domain

$$\mathcal{U}_{\mathrm{Panda}} = \{u \in \mathbb{R}^7 : u^\top S_u^{-2} u \le 1\}. \tag{125}$$

Thus, the unit input ball in the main text refers to normalized torques and maps to an ellipsoid inside the physical torque box. The obstacle constraint covers only the end-effector point. Other links, payload volume, and self-collision are not included; joint ranges are monitored by a separate protection mechanism.

G.2 Obstacle Constraint and Learning Settings

The simulated obstacle has center $c \in \mathbb{R}^3$ and radius $r = 0.10$ m. Let $d(q) = p(q) - c$. We use the barrier function and surface clearance

$$h(x) = d(q)^\top d(q) - r^2, \qquad \delta(x) = \|d(q)\|_2 - r. \tag{126}$$

Their units are m^2 and m, respectively. The barrier derivatives are

$$\begin{aligned} \dot{h} &= 2d^\top J\dot{q}, \\ \ddot{h} &= 2\|J\dot{q}\|_2^2 + 2d^\top(J\ddot{q} + \dot{J}\dot{q}). \end{aligned} \tag{127}$$

Using a relative-degree-two HOCBF (Xiao & Belta, 2022), we define

$$b_1 = \dot{h} + \lambda h, \qquad b_2 = \ddot{h} + 2\lambda\dot{h} + \lambda^2 h, \qquad \lambda = 80\,\mathrm{s}^{-1}. \tag{128}$$

Table 12: Panda joint-position limits and experimental torque limits.

Joint	$q_{\min}$ (rad)	$q_{\max}$ (rad)	Torque range (N·m)
1	-2.8973	2.8973	$[-60, 60]$
2	-1.7628	1.7628	$[-60, 60]$
3	-2.8973	2.8973	$[-60, 60]$
4	-3.0718	-0.0698	$[-60, 60]$
5	-2.8973	2.8973	$[-10, 10]$
6	-0.0175	3.7525	$[-10, 10]$
7	-2.8973	2.8973	$[-10, 10]$

The safety condition is $b_2 > 0$. With $a_s = -M_s^{-1}(C_s\dot{q} + g_s + D\dot{q})$, the constraint has the control-affine form

$$\begin{aligned} b_{2,s}(x,u) &= \phi_s(x) + \psi_s(x)^\top u, \\ \phi_s &= 2\|J\dot{q}\|_2^2 + 2d^\top(Ja_s + \dot{J}\dot{q}) + 160\dot{h} + 6400h, \\ \psi_s^\top &= 2d^\top J M_s^{-1}, \qquad \bar{\psi}_s^\top = \psi_s^\top S_u. \end{aligned} \tag{129}$$

Global closed-form CBF knowledge: residual learning. Using global closed-form CBF knowledge, the filter evaluates the nominal constraint and learns $\Delta H = b_{2,\mathrm{t}} - b_{2,\mathrm{n}}$. Training labels are the difference between the true and nominal simulated constraints with added zero-mean Gaussian label noise. With posterior mean $\mu_{\Delta H}$ and standard deviation $\sigma_{\Delta H}$, the robust constraint is

$$\phi_{\mathrm{n}} + \bar{\psi}_{\mathrm{n}}^\top \bar{u} + \mu_{\Delta H}(x,\bar{u}) - \beta\sigma_{\Delta H}(x,\bar{u}) \geq \eta_1. \tag{130}$$

Point-wise CBF evaluations via distance measurements: learning the complete second derivative. Numerical differentiation of point-wise CBF evaluations obtained via distance measurements yields $\ddot{h}$ labels paired with state–input data for GP training. The GP thereby learns the complete control-dependent $\ddot{h}(x,u)$ without global closed-form CBF knowledge or nominal CBF dynamics. Joint states and local relative-position features remain available. The deterministic first derivative is computed kinematically as $\dot{h} = 2(p-c)^\top J\dot{q}$, without a separate GP. The robust constraint is

$$\mu_{\ddot{h}}(x,\bar{u}) + 160\dot{h} + 6400h - \beta\sigma_{\ddot{h}}(x,\bar{u}) \geq \eta_1. \tag{131}$$

In the notation of the main text, the known terms are $\widetilde{\phi} = 160\dot{h} + 6400h$ and $\widetilde{\psi} = 0$, while the GP supplies the unknown control-affine second derivative. Both settings use the same event trigger and seven-dimensional structured exploration mechanism.

G.3 Control and Learning Parameters

Common nominal controller. Both settings use the same computed-torque PD controller to generate the requested torque. During online control, $q_r = q_{\mathrm{g}}$ and the reference velocity and acceleration are zero. The requested acceleration and torque are

$$\begin{aligned} \ddot{q}_{\mathrm{cmd}} &= 40(q_r - q) - 2\sqrt{40}\dot{q}, \\ u_{\mathrm{PD}} &= M_{\mathrm{n}}\ddot{q}_{\mathrm{cmd}} + C_{\mathrm{n}}\dot{q} + g_{\mathrm{n}} + D\dot{q}. \end{aligned} \tag{132}$$

The requested torque is clipped to the experimental torque ranges and mapped to normalized coordinates. The safety filter minimizes its squared deviation from this request in normalized coordinates, subject to the input domain (125) and the corresponding robust constraint. During initial data collection, the reference is the initial joint position.

Control and initialization. The control frequency is 200 Hz, with $T_c = 0.005$ s and zero-order-held torques. Online simulation lasts 5 s, excluding initialization, and uses MuJoCo's implicit-fast integrator. Each task collects its own initial data using the holding action and paired positive and negative excitations along the seven control axes. Each action is held for two control periods; 90 collection steps give an initialization duration of 0.45 s. We retain 15 initial GP samples.

Sample selection first ensures full rank of the action design matrix including its constant column, then covers the control axes and excitation signs, and finally fills the remaining samples by a max–min criterion in normalized state–action space. Collection checks clearance, joint protection, state

Table 13: Panda control, initialization, and learning parameters.

Parameter	Value
Control frequency / period	200 Hz / 0.005 s
Online horizon	5 s
PD gains K_p, K_d	$40I_7$, $2\sqrt{40}I_7$
Initial collection steps / action hold steps	90 / 2
Initial GP samples	15
Required action-design rank / minimum samples	8 / 8
Initial excitation / minimum realized amplitude	0.10 / 0.05
Initial clearance guard	0.15 m
Maximum action-design condition number	99
Maximum normalized state drift	0.65
Confidence multiplier β	5
Exploration magnitude ϵ	0.15
Filter / trigger margins η_1, η_2	5, 5
Quadratic slack penalty	10^6
HOCBF gain λ	$80\,\mathrm{s}^{-1}$

drift, and action clipping. Table 13 summarizes the parameters. Online data are not shared between test tasks.

Distance-difference labels. Given distance observations ρ_k, we evaluate $h_k = \rho_k^2 - r^2$. For three consecutive equally spaced observations with interval Δt, once the observation at time k is available, the second-derivative label assigned to the center time $k-1$ is

$$\hat{\ddot{h}}_{k-1} = \frac{h_k - 2h_{k-1} + h_{k-2}}{\Delta t^2}. \tag{133}$$

The label is therefore constructed only from observations available by time k and is paired with the state and executed action at time $k-1$. We accept it only when both adjacent observation windows share the same constant executed torque; no second-derivative label crosses an action switch. Initialization and online exploration follow this rule, and GP posteriors are updated when valid exploration labels become available. Trigger events, GP updates, and training samples are recorded separately.

Input normalization. Both main GPs use the 17-dimensional feature vector $\chi = [q^\top, \dot{q}^\top, (p(q) - c)^\top]^\top$ and normalized torque $\bar{u}$. We apply $\bar{\chi} = (\chi - \mu_\chi) \oslash s_\chi$, where $\oslash$ denotes componentwise division. The official Panda joint-position ranges and velocity specifications provide physical reference scales for the $[q, \dot{q}]$ coordinates (Franka Robotics). The offsets and scales below are fixed experimental preprocessing constants, rather than manufacturer-provided distribution statistics or quantities fitted to test trajectories.

$$\begin{aligned} \mu_q &= (0, 0, 0, -1.6, 0, 2.5, 0)^\top, \\ \mu_{\dot{q}} &= 0_7, \qquad \mu_d = 0_3, \\ s_q &= (2.9, 1.9, 2.9, 1.5, 2.9, 2.1, 3.1)^\top, \\ s_{\dot{q}} &= (2.7, 2.7, 2.7, 2.7, 5.3, 4.2, 5.3)^\top, \\ s_d &= (0.5, 0.5, 0.5)^\top. \end{aligned} \tag{134}$$

We concatenate these blocks to obtain μ_χ and s_χ. Position and velocity offsets and scales use rad and rad/s, respectively, while the relative-position block uses m. These constants remain fixed across tasks and are not hard bounds on the state.

The software environment for the Panda experiments is summarized in Table 14.

G.4 Randomized Tasks and Evaluation

At 200 Hz, the two settings share 100 randomized tasks generated with master seed 42, yielding 200 learning-control trajectories. Task hashes match for all 100 pairs. Tasks are fixed by geometric checks and nominal-trajectory screening before learned-controller evaluation; acceptance does not depend on learned-controller outcomes.

Table 14: Software environment for the Panda experiments. PyTorch uses the CPU-only build.

Software	**Version**	**Role**
Python	3.11.15	Programming language
PyTorch	`2.13.0+cpu`	Tensor computations
GPyTorch	1.15.2	Gaussian process models
`linear_operator`	0.6.1	GP linear algebra
CVXPY	1.9.2	Convex optimization interface
Clarabel	0.11.1	SOCP solver
MuJoCo	3.11.0	Robot dynamics simulation
NumPy	2.4.6	Numerical arrays
SciPy	1.17.1	Scientific computing utilities
Pandas	3.0.5	Result aggregation
Matplotlib	3.11.1	Visualization

Table 15: GP parameters for the Panda settings.

Parameter	Closed-Form CBF	Point-Wise CBF Eval.
Output scale	0.50	0.75
Label-noise standard deviation	0.01	0.05
Drift length scale ℓ_0	2.0	2.0
Coefficient length scales ℓ_j	1.5 for all j	1.5 for all j
Drift amplitude s_0	0.75	2.5
Coefficient amplitudes $(s_1, \ldots, s_7)$	$(6.5, 2.9, 3.2,$ $6.1, 3.8, 2.1, 0.1)$	$(14.5, 8.5, 8.5,$ $11.0, 8.0, 5.5, 0.3)$
Initial diagonal regularization	10^{-12}	10^{-12}
Confidence multiplier β	5	5

Start and goal sampling. Let $w = q_{\max} - q_{\min}$ and $q_{\text{ref}} = (0, -\pi/4, 0, -3\pi/4, 0, \pi/2, \pi/4)^\top$. We sample a task center and joint displacement as

$$q_{\text{c}} = q_{\text{ref}} + 0.06w \odot z_{\text{c}}, \qquad \Delta q = \rho w \odot \frac{z_{\text{d}}}{\|z_{\text{d}}\|_2}, \tag{135}$$

where $z_{\text{c}}, z_{\text{d}} \sim \mathcal{N}(0, I_7)$, $\rho \sim \mathcal{U}[0.14, 0.28]$, and $\odot$ is componentwise multiplication. The endpoints are $q_{\text{c}} \mp \Delta q/2$ and are swapped with probability $1/2$. Initial velocity is zero. These geometric rejection checks enforce joint-margin, kinematic-conditioning, and workspace-geometry requirements rather than selecting tasks based on the performance of any learned method. Rejection criteria are listed in Table 16; the resulting distribution is conditional on these checks and the trajectory screening below. A normalized joint margin is the distance to the nearest position limit divided by the joint's full range.

Obstacle placement. The center lies in $[0.28, 0.62] \times [-0.28, 0.28] \times [0.10, 0.50]$ m. Let t be the unit tangent of the start-to-goal end-effector chord and e_z the world vertical unit vector. Define

$$n_{\text{up}} = \frac{e_z - (e_z^\top t)t}{\|e_z - (e_z^\top t)t\|_2}, \qquad n_{\text{lat}} = t \times n_{\text{up}}. \tag{136}$$

For $p_\xi = p(q_{\text{s}}) + \xi[p(q_{\text{g}}) - p(q_{\text{s}})]$, we place the center at

$$c = p_\xi - r(a_{\text{up}} n_{\text{up}} + a_{\text{lat}} n_{\text{lat}}), \tag{137}$$

with $\xi \sim \mathcal{U}[0.38, 0.62]$, $a_{\text{up}} \sim \mathcal{U}[0.45, 0.70]$, and $a_{\text{lat}} \sim \mathcal{U}[-0.18, 0.18]$. Positive a_{up} places the center below the chord. Initial and target clearances must lie in $[0.20, 0.40]$ m and $[0.05, 0.30]$ m, respectively. The obstacle must block the chord, while the required upward lift $c_z + r - (p_\xi)_z$ is positive and at most 0.10 m. Directional checks require nonnegative upward control acceleration at both endpoints and nonnegative outward control acceleration at the start.

Candidate ranking and nominal proxies. We organize candidate generation into screening batches. Within each batch, task candidates are sampled and subjected to the geometric rejection checks above until 16 geometrically admissible candidates are retained. These 16 candidates are then ranked and evaluated by the nominal proxy trajectories described below. If no candidate in

Table 16: Geometric checks for Panda start and goal configurations. Jacobian values use meter–radian coordinates.

Quantity	Requirement
Minimum normalized joint margin	0.20
Joint-displacement norm	$[0.30, 1.60]$ rad
Maximum absolute joint displacements	$(0.85, 0.70, 0.85, 0.65,$ $0.85, 0.70, 0.85)$ rad
End-effector chord length	$[0.42, 0.65]$ m
Absolute vertical displacement / chord length	≤ 0.35
Maximum downward end-effector displacement	0.10 m
Minimum translational Jacobian singular value	0.025
Maximum initial translational Jacobian singular value	0.65

Table 17: Nominal-trajectory screening parameters.

Quantity	Requirement
Duration of each proxy	5 s
Minimum predicted joint margin	0.30 rad
Maximum PD-proxy penetration	$0.1r = 0.01$ m
Minimum nominal-CBF-proxy clearance	0 m
Additional nominal-CBF-proxy checks	Constraint satisfaction; no slack, action clipping, or simulation failure
Constraint tolerance	10^{-7}
Maximum geometric attempts per batch	50,000
Maximum obstacle samples per endpoint pair	32
Maximum screening batches	8

the batch passes the proxy screening, a new batch is generated. Pareto ranking favors a smaller maximum initial translational Jacobian singular value, smaller downward displacement, and greater nominal CBF control authority near the goal. Ordering within a Pareto front is randomized. Candidates are evaluated in the nominal simulator with two full proxy trajectories: computed-torque PD alone and the same controller with a nominal CBF filter. Both reference controllers use nominal dynamics. We accept the first candidate passing both proxies under Table 17. The screening removes degenerate tasks dominated by joint-limit proximity, poor kinematic conditioning, or unsuitable obstacle geometry, so that the evaluation focuses on the intended obstacle-avoidance setting rather than unrelated failure modes. Candidate acceptance is determined before evaluating the learned methods and does not depend on GP uncertainty or structured exploration.

Task completion and joint protection. A task is completed when $\|q_{\text{final}} - q_{\text{g}}\|_2 \leq 0.05$ rad. For each joint, the protection mechanism uses

$$m_i^{\text{pred}} = \min(q_i - q_{i,\min}, q_{i,\max} - q_i) - |\dot{q}_i| T_c. \tag{138}$$

The run terminates when any predicted margin falls below 0.05 rad. This protection is separate from the obstacle HOCBF.

Evaluation. Safety uses the HOCBF condition $b_2 > 0$. SOCP infeasibility follows the slack-fallback convention in Appendix E.4. Path overhead follows the definition in that appendix. Minimum clearance is $\delta_{\min} = \min_k(\|p(q_k) - c\|_2 - r)$, and trigger frequency is the number of exploration events divided by the actual online duration, excluding initialization. Both settings complete all 100 tasks with no safety violations, infeasible SOCPs, or joint-protection terminations; all continuous metrics therefore include 100 trajectories per setting and are reported as means $\pm$ standard deviations. Batch runtimes are reported in Table 8.