# Safe Active Learning with Future Viability Guarantees in Time-Series Models

**Hyeon-jun Park**[1] * **Whiyoung Jung**[2] **Deunsol Yoon**[2]
**Sunghoon Hong**[2] **Kanghoon Lee**[2] **Kyungjae Lee**[3†]
[1]Department of Artificial Intelligence, Chung-Ang University
[2]LG AI Research
[3]Department of Statistics, Korea University

## Abstract

In time-series systems, safe active learning seeks to collect informative data while maintaining safety during data collection. A key challenge is that one-step safety does not guarantee future viability: an input can be immediately safe yet lead the system into a state from which no safe continuation exists. For example, in pressure regulation, an input may keep the current pressure below its limit but still drive the system toward an unsafe pressure spike over the next few steps. We study this problem in nonlinear time-series systems with unknown dynamics and safety constraints. We introduce the notion of $m$-step dead-end inputs and propose a future-aware safe active learning framework (FA-SAL) that enforces finite-horizon safety via margin-certified rollout constraints. Our approach constructs a conservative approximation of the true future-safe set by combining Gaussian process uncertainty with a recursive error bound over predicted trajectories. We show that FA-SAL provides high-probability safety guarantees, recovers interior future-safe inputs as uncertainty decreases, and provably excludes separated dead-end inputs. Experiments on synthetic and real-world benchmarks demonstrate that FA-SAL reduces safety violations and dead-end selections while maintaining competitive model learning performance.

## 1 Introduction

Learning an accurate model typically requires informative data, yet collecting such data can be expensive, time-consuming, or operationally difficult in many real-world systems. Active learning addresses this issue by selecting queries that are expected to be most informative for model estimation [22]. In many real-world systems, however, informativeness alone is not enough: each query must also satisfy safety constraints. This requirement is particularly important in scientific experimentation, robotics, industrial process control, and medical decision support, where an unsafe query may damage equipment, violate operational constraints, or drive the system into an unrecoverable state [1, 9–11]. These challenges have led to the development of safe active learning, a framework for learning unknown systems while maintaining safety throughout the data collection process [28].

A key limitation of existing safe active learning methods is that safety is usually certified only for the next query. The query itself must be safe, but its longer-term consequences are not taken into account [15, 23, 26, 35]. This limitation is especially problematic in sequential systems with temporal dependence [3], where an action that is safe at the current step may nevertheless move the system into a region from which no safe continuation is possible. In other words, immediate safety does not necessarily imply future viability. A simple example arises in the active learning of physical dynamical systems [9]. Suppose a learner probes a thermal or chemical process by selecting control

*Contact:hjp202@cau.ac.kr, kyungjae_lee@korea.ac.kr
†Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

inputs to identify its dynamics. A candidate input may keep the current temperature or pressure within the allowable range and therefore appear safe under a one-step criterion. However, it may still steer the system toward an irreversible operating regime in which every admissible future input leads to a safety violation within a few subsequent steps.

This loss of future feasibility is closely related to the notion of *dead-ends*, which has been studied in stochastic shortest path (SSP) Markov decision processes [2] and goal-oriented planning [8, 12–14]. In that literature, a dead-end state is one from which the goal can no longer be reached, and its presence fundamentally changes the structure of the decision problem [13, 14]. This viewpoint is valuable because it captures a broader class of real-world settings in which certain decisions may irreversibly eliminate the possibility of safe or successful future evolution. However, despite its conceptual and practical value, this viewpoint has not yet been incorporated into safe active learning for time-series models [15, 28, 35].

In this paper, we address this gap by introducing a future-aware formulation of safe active learning for time-series systems. We consider a nonlinear autoregressive exogenous (NARX) setting [3] in which the learner sequentially selects inputs to probe an unknown system with an unknown safety function. Our key requirement is that a candidate query should not only be safe at the next step, but should also preserve safe continuation over a finite future horizon. To formalize this idea, we define $m$-step future viability and introduce $m$-step dead-end inputs, namely, inputs that are immediately safe but inevitably lead to a safety violation within the next $m$ steps. Building on this formulation, we propose a Gaussian-process-based [21] safe active learning algorithm with a margin-based rollout certificate. We summarize our contributions as follows:

- We formulate safe active learning with finite-horizon future viability in NARX time-series models. This formulation distinguishes pointwise safety from future safety and introduces $m$-step dead-end inputs, which are immediately safe but future-unsafe. To the best of our knowledge, this is the first work to introduce a dead-end notion in safe active learning.
- We propose a future-aware safe active learning (FA-SAL) algorithm (Algorithm 1), which learns the unknown transition model by selecting informative queries that are certified to remain viable over a finite future horizon.
- We establish high-probability safety guarantees (Theorem 1) and interior recovery guarantees for the future-safe set (Theorem 2), showing that FA-SAL safely approximates the true future-safe set up to a certification-error boundary layer. To the best of our knowledge, this is the first theoretical analysis of future-safe set recovery in safe active learning for time-series systems.
- The empirical results support our theoretical findings. FA-SAL better preserves finite-horizon safety and reduces dead-end selections while maintaining informative transition-model learning.

## 2 Related Work

Gaussian-process-based safe active learning selects informative queries while maintaining safety constraints [23, 28, 34, 35]. Recent work has considered time-series [35] and time-varying safe active learning [15], while related lines of work study active learning, safe exploration, and safe Bayesian optimization for dynamical systems and black-box objectives under safety constraints [1, 4, 5, 9–11, 16–18, 22, 25–27, 29–31]. These methods typically focus on pointwise safe evaluations, trajectory-level safety probabilities, or control-oriented exploration. In contrast, our setting targets active learning of an unknown transition model in NARX time-series systems, where a query changes the future regressor state and must preserve safe continuation over a finite horizon. Notably, prior work introduced future-aware safe active learning for time-varying systems [15]. This approach reduces prediction uncertainty over current and future states, while enforcing pointwise safety at the selected query. We instead use future safety to determine whether a current query is admissible, requiring its NARX rollout to satisfy the safety margin at every step. Our formulation is also related to dead-end reasoning in planning, stochastic shortest path problems, medical decision making, safe reinforcement learning, and dialogue systems [7, 8, 12–14, 32, 33]. We bring this perspective into safe active learning by defining $m$-step dead-end inputs and a margin-certified rollout condition that excludes immediately safe inputs with no $m$-step safe continuation. We provide a more detailed discussion of related work in Appendix A.

## 3 Preliminaries

In this section, we introduce the problem formulation, notation, and assumptions required to present and analyze our proposed algorithm. We first formulate safe active learning in time-series models, then define finite-horizon future safety and dead-end inputs, and review the Gaussian process and assumptions used in our theoretical analysis.

**Safe active learning in time-series models.** At each round $t$, the learner selects a manipulated input $u_t \in \mathcal{U} \subseteq \mathbb{R}$, and the system returns an output $y_t \in \mathbb{R}$. For time-series modeling, we adopt a nonlinear autoregressive exogenous (NARX) representation [3]. The regression state is defined by a finite history of past outputs and inputs $x_t = (y_{t-1}, \dots, y_{t-d_y}, u_t, u_{t-1}, \dots, u_{t-d_u+1}) \in \mathcal{X}$, where $d_y$ and $d_u$ denote the output and input window lengths, respectively. The system output is generated according to $y_t = f(x_t) + \varepsilon_t$, where $f : \mathcal{X} \to \mathbb{R}$ is an unknown transition map in regression form and $\varepsilon_t$ is observation noise. In addition to the transition map, we assume an unknown safety function $g : \mathcal{X} \to \mathbb{R}$. At each evaluated state, the learner also receives a noisy safety measurement $y_t^g = g(x_t) + \varepsilon_t^g$, where $\varepsilon_t^g$ is safety observation noise. Let $h_t = \{(u_s, y_s, y_s^g)\}_{s=1}^t$ denote the history available at round $t$. A regression state is safe if $g(x) \geq 0$, and the true safe set is $\mathcal{S} := \{x \in \mathcal{X} : g(x) \geq 0\}$. The learner's objective is to actively collect informative data for estimating the unknown transition map $f$, while ensuring that the system remains safe throughout the data collection process. The proposed formulation and algorithm naturally support multi-dimensional inputs and outputs.

**Gaussian process models and certified safety.** We use Gaussian process (GP) as nonparametric estimators for both the unknown transition map $f$ and the unknown safety function $g$. In particular, we model the unknown transition map $f$ and safety function $g$ using independent GP priors $f \sim \mathcal{GP}(0, k_f)$ and $g \sim \mathcal{GP}(0, k_g)$, where $k_f$ and $k_g$ are positive definite kernels on $\mathcal{X}$ [21].

We introduce the regularity and confidence assumptions used in the method and analysis. These assumptions are standard in safe learning and Bayesian optimization with Gaussian processes [23, 25, 26, 28, 35].

**Assumption 1** (Sub-Gaussian observation noise)**.** *The transition noises $\{\varepsilon_t\}_{t\geq 1}$ and safety measurement noises $\{\varepsilon_t^g\}_{t\geq 1}$ are conditionally mean-zero and $\sigma_\varepsilon$-sub-Gaussian given $h_{t-1}$ and $u_t$. The GP posterior and information-gain formulas use conditionally independent Gaussian observation likelihoods with variance $\sigma_\varepsilon^2$ for the two measurements.*

**Assumption 2** (RKHS-bounded unknown functions)**.** *The unknown functions $f$ and $g$ satisfy $f \in \mathcal{H}_{k_f}$ and $g \in \mathcal{H}_{k_g}$ where $\mathcal{H}_{k_f}$ and $\mathcal{H}_{k_g}$ are the RKHSs induced by $k_f$ and $k_g$, respectively.*

Given observations up to round $t$, the posterior model for $f$ provides a mean $\mu_t^f(x)$ and standard deviation $\sigma_t^f(x)$. Analogously, the posterior model for $g$ provides $\mu_t^g(x)$ and $\sigma_t^g(x)$. The posterior mean and variance of the transition GP are given by $\mu_t^f(x) = k_t^f(x)^\top (K_t^f + \sigma_\varepsilon^2 I)^{-1} y_{1:t}$ and $(\sigma_t^f(x))^2 = k_f(x,x) - k_t^f(x)^\top (K_t^f + \sigma_\varepsilon^2 I)^{-1} k_t^f(x)$ where $k_t^f(x) := (k_f(x_1, x), \dots, k_f(x_t, x))^\top$ and $K_t^f := [k_f(x_i, x_j)]_{i,j=1}^t$. The posterior quantities for $g$ are defined analogously using the safety observations $y_{1:t}^g$, the kernel $k_g$, and the safety observation likelihood. Our safety and recovery guarantees use the confidence events defined below.

**Lemma 1** (GP confidence on a finite domain; Lemma 5.1 in [24])**.** *Let $q \in \{f, g\}$, and let $D \subset \mathcal{X}$ be a finite domain. Pick $\delta_q \in (0,1)$ and set $\beta_t^q = 2\log(|D|\pi_t/\delta_q)$, where $\pi_t > 0$ satisfies $\sum_{t\geq 1} \pi_t^{-1} = 1$. Then, with probability at least $1 - \delta_q$, the event $\mathcal{E}_t^q := \{|q(z) - \mu_t^q(z)| \leq \sqrt{\beta_t^q}\, \sigma_t^q(z),\ \forall z \in D\}$ holds for all $t \geq 1$.*

**Rollouts and finite-horizon future safety.** Let $U^{\mathrm{cand}} \subseteq \mathcal{U}$ denote a fixed finite candidate input set. Given a candidate current input $u \in U^{\mathrm{cand}}$, the induced next regressor is $\hat{x}_{t+1}(u) = (y_t, \dots, y_{t-d_y+1}, u, u_t, \dots, u_{t-d_u+2})$. To evaluate the future consequence of $u$, we consider a planned continuation sequence $\Pi_t(u) = \big(u_{t+2}^{\mathrm{plan}}(u), \dots, u_{t+m}^{\mathrm{plan}}(u)\big)$, over a planning horizon $m$. Under this continuation, for $j = 1, \dots, m-1$, the hallucinated rollout is defined recursively by

$$\hat{y}_{t+j}(u) = \mu_t^f(\hat{x}_{t+j}(u)) \quad \text{and} \quad \hat{x}_{t+j+1}(u) = \Phi\big(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\big), \tag{1}$$

where $\Phi$ denotes the NARX shift operator. Specifically, for $x = (y_1, \ldots, y_{d_y}, u_1, \ldots, u_{d_u})$, we define $\Phi(x, y, u) := (y, y_1, \ldots, y_{d_y-1}, u, u_1, \ldots, u_{d_u-1})$, so that applying $\Phi$ inserts the new output $y$ and input $u$, and drops the oldest output $y_{d_y}$ and input $u_{d_u}$ from the history.

The corresponding noiseless true rollout is denoted by $\bar{x}_{t+1}(u), \bar{x}_{t+2}(u), \ldots, \bar{x}_{t+m}(u)$, where the same planned continuation is applied but future outputs are generated by the true transition map $f$. In particular, $\bar{x}_{t+1}(u) = \hat{x}_{t+1}(u)$, and for $j = 1, \ldots, m-1$,

$$\bar{y}_{t+j}(u) = f(\bar{x}_{t+j}(u)) \quad \text{and} \quad \bar{x}_{t+j+1}(u) = \Phi\big(\bar{x}_{t+j}(u), \bar{y}_{t+j}(u), u^{\text{plan}}_{t+j+1}(u)\big). \tag{2}$$

Throughout the theoretical analysis, $\hat{x}_{t+j}(u)$ denotes the hallucinated rollout state, whereas $\bar{x}_{t+j}(u)$ denotes the corresponding noiseless true rollout state. We define the true $m$-step future-safety score of a candidate input $u$ by

$$G^{(m)}_t(u) := \min_{1 \le j \le m} g(\bar{x}_{t+j}(u)). \tag{3}$$

Thus, $G^{(m)}_t(u) \ge 0$ if and only if the noiseless rollout induced by $u$ remains safe for the next $m$ steps. The corresponding true future-safe candidate set is $U^{\text{true}}_t := \{u \in U^{\text{cand}} : G^{(m)}_t(u) \ge 0\}$. For $\epsilon > 0$, we also define the $\epsilon$-interior of the true future-safe set as $U^{\text{int}}_{t,\epsilon} := \{u \in U^{\text{cand}} : G^{(m)}_t(u) \ge \epsilon\}$.

**Lipschitz regularity.** The rollout certificate propagates GP uncertainty through the NARX recursion. To control how this uncertainty affects future states and safety scores, we use the following regularity bounds. For the same input $u$, $L_x$ bounds the effect of changing the current regressor, and $L_y$ bounds the effect of changing the new output:

$$\|\Phi(x, y, u) - \Phi(x', y', u)\| \le L_x \|x - x'\| + L_y |y - y'|. \tag{4}$$

Under the Euclidean norm, the NARX shift operator satisfies this bound with $L_x = L_y = 1$. We also assume Lipschitz continuity of the transition function $f$, the safety function $g$, and the lower confidence bound (LCB) safety score $\ell_t(x) := \mu^g_t(x) - \sqrt{\beta^g_t}\sigma^g_t(x)$.

**Assumption 3** (Lipschitz regularity)**.** *Assume that, for each $q \in \{f, g, \ell_t\}$, the function $q : \mathcal{X} \to \mathbb{R}$ is $L_q$-Lipschitz, i.e., $|q(x) - q(x')| \le L_q \|x - x'\|, \quad \forall x, x' \in \mathcal{X}, \forall q \in \{f, g, \ell_t\}$.*

**$m$-step dead-end inputs.** The distinction between one-step safety and future viability is captured by $m$-step dead-end inputs. A candidate input $u \in U^{\text{cand}}$ is called an $m$-step dead-end input if it is safe at the next step but its induced noiseless rollout violates safety within the next $m$ steps: $g(\bar{x}_{t+1}(u)) \ge 0$ and $G^{(m)}_t(u) < 0$. The following definition adds a margin to this dead-end condition.

**Definition 1** ($\epsilon$-separated $m$-step dead-end input)**.** *For $\epsilon \ge 0$, a candidate input $u \in U^{\text{cand}}$ is called an* $\epsilon$-separated $m$-step dead-end input *if $g\big(\bar{x}_{t+1}(u)\big) \ge \epsilon$ and $G^{(m)}_t(u) \le -\epsilon$.*

## 4 Future-Aware Safe Active Learning

In this section, we introduce a future-aware safe active learning algorithm (Algorithm 1) for time-series systems. The main idea is to ensure that a query is safe not only at the next step, but also along its $m$-step future rollout.

**Algorithm.** At each round $t$, the algorithm first updates the GP posteriors of the transition map $f$ and the safety function $g$, and constructs the lower confidence bound (LCB) safety score $\ell_t(x) := \mu^g_t(x) - \sqrt{\beta^g_t}\sigma^g_t(x)$, which induces the certified safe set $S^{\text{LCB}}_t := \{x \in \mathcal{X} : \ell_t(x) \ge 0\}$. The algorithm then considers each candidate input $u \in U^{\text{cand}}$ and constructs a hallucinated rollout $\hat{x}_{t+1}(u), \ldots, \hat{x}_{t+m}(u)$ using the GP posterior mean predictor $\mu^f_t$ together with a planned continuation sequence $\Pi_t(u) = \big(u^{\text{plan}}_{t+2}(u), \ldots, u^{\text{plan}}_{t+m}(u)\big)$. At each rollout step, we consider continuation inputs whose next predicted state satisfies the safety margin. We then choose the input with the largest transition-GP uncertainty at that state.

As the rollout is propagated forward, the algorithm maintains a recursive error envelope $\bar{e}_{t,j}(u)$, which upper bounds the mismatch between the noiseless true rollout state $\bar{x}_{t+j}(u)$ and its hallucinated

counterpart $\hat{x}_{t+j}(u)$. A rollout state is certified only if $\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell \bar{e}_{t,j}(u)$ for $j = 1, \ldots, m$. The largest certified rollout depth is denoted by $\hat{m}_t(u)$. If $\hat{m}_t(u) = m$, then $u$ is inserted into the certified feasible set $U^{\text{feas}}_{t,m}$. For each feasible candidate, the algorithm constructs the planned rollout

$$X^{\text{plan}}_t(u) = \{\hat{x}_{t+1}(u), \ldots, \hat{x}_{t+m}(u)\}, \tag{5}$$

and evaluates both the transition-model and safety-model rollout information gains:

$$I^f_t(u) := I\big(f(X^{\text{plan}}_t(u)); y(X^{\text{plan}}_t(u)) \mid h_t\big), \;\; I^g_t(u) := I\big(g(X^{\text{plan}}_t(u)); y^g(X^{\text{plan}}_t(u)) \mid h_t\big). \tag{6}$$

The final query is selected by maximizing the uncertainty-reduction objective induced by the theoretical certification bound. The full procedure is summarized in Algorithm 1.

**Recursive error.** Since the true transition map $f$ is unknown, future states cannot be rolled out exactly. FA-SAL therefore constructs hallucinated rollouts using the transition surrogate $\hat{f}_t = \mu^f_t$. For a candidate input $u$, let $\hat{x}_{t+1}(u), \ldots, \hat{x}_{t+m}(u)$ denote the hallucinated rollout generated by the GP posterior mean, and let $\bar{x}_{t+1}(u), \ldots, \bar{x}_{t+m}(u)$ denote the corresponding noiseless true rollout generated by the true transition map $f$ under the same planned continuation. The two rollouts share the same initial next regressor, $\bar{x}_{t+1}(u) = \hat{x}_{t+1}(u)$, but may deviate at later steps because $f$ is unknown. To control this mismatch, FA-SAL maintains a computable error envelope $\bar{e}_{t,j}(u)$, initialized by $\bar{e}_{t,1}(u) = 0$ and updated as

$$\bar{e}_{t,j+1}(u) = (L_x + L_y L_f)\bar{e}_{t,j}(u) + L_y \sqrt{\beta^f_t}\, \sigma^f_t(\hat{x}_{t+j}(u)), \qquad j = 1, \ldots, m-1. \tag{7}$$

The first term propagates accumulated mismatch through the NARX regressor, while the second term accounts for local uncertainty of the transition GP. On the transition confidence event (Lemma 1), Lemma 4 gives

$$\|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| \leq \bar{e}_{t,j}(u), \qquad j = 1, \ldots, m. \tag{8}$$

The envelope provides a simple geometric interpretation of the safety test. The hallucinated trajectory must not only lie in the LCB safe region, but must lie inside it with enough margin to absorb the possible rollout error:

$$\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell \bar{e}_{t,j}(u), \qquad j = 1, \ldots, m. \tag{9}$$

Indeed, by Lipschitz continuity of $\ell_t$, this condition implies

$$\ell_t(\bar{x}_{t+j}(u)) \geq \ell_t(\hat{x}_{t+j}(u)) - L_\ell \|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| \tag{10}$$

$$\geq \ell_t(\hat{x}_{t+j}(u)) - L_\ell \bar{e}_{t,j}(u) \geq 0. \tag{11}$$

Thus the noiseless true rollout remains in $S^{\text{LCB}}_t$, and hence in $\mathcal{S}$ on the safety confidence event, for all $j = 1, \ldots, m$ (Lemma 2).

**Feasible input set.** The recursive margin condition (9) induces the certified $m$-step feasible set

$$U^{\text{feas}}_{t,m} := \left\{u \in U^{\text{cand}} : \ell_t(\hat{x}_{t+j}(u)) \geq L_\ell \bar{e}_{t,j}(u), \quad j = 1, \ldots, m\right\}. \tag{12}$$

Thus, a candidate $u$ is feasible only if its hallucinated rollout remains inside the LCB safe region with sufficient margin at every step of the planning horizon. Candidates that fail this test are rejected because their hallucinated rollouts are either unsafe or too close to the safety boundary to be certified under transition-model uncertainty.

**Query selection rule.** After constructing $U^{\text{feas}}_{t,m}$, FA-SAL selects the next query by maximizing the uncertainty-reduction objective induced by the finite-time certification bound, with $C_\eta := 2(\sigma^2_\varepsilon + 1)$:

$$u_{t+1} \in \operatorname*{arg\,max}_{u \in U^{\text{feas}}_{t,m}} \left[ 2\sqrt{\beta^g_t}\sqrt{m C_\eta I^g_t(u)} + (L_g + L_\ell) L_y \sqrt{\beta^f_t}\sqrt{m C_\eta I^f_t(u)} \sum_{r=0}^{m-2} (L_x + L_y L_f)^r \right]. \tag{13}$$

**Remark 1.** *We first present FA-SAL in a single-query online form. In this version, the learner applies one input at each round, while the $m$-step hallucinated rollout is used internally to certify future viability and to evaluate the acquisition score. This formulation isolates the core safety certificate: a candidate input is accepted only if its induced hallucinated rollout remains inside the LCB safe region with enough margin to absorb transition-model error. We therefore present the trajectory-query variant as a direct extension, where the certification procedure is unchanged and the final query object is lifted from the first input to the full hallucinated trajectory (Algorithm 2).*

## 5 Theoretical Results

This section shows that the certificates (9) used inside Algorithm 1 are not merely heuristic screening rules, but provide high-probability guarantees for finite-horizon safety.

### 5.1 Finite-Horizon Safety Certification and Future-Safe Set Recovery

We first show that the safety confidence event in Lemma 1 implies that every state in the LCB-certified safe set is truly safe.

**Lemma 2** (LCB-certified safety implies true safety)**.** *Let $D \subset \mathcal{X}$ be a finite domain on which the GP confidence event is defined, and suppose that all states evaluated by the safety certificate lie in $D$. On the safety confidence event $\mathcal{E}_t^g$ from Lemma 1, we have $S_t^{\mathrm{LCB}} \cap D \subseteq \mathcal{S} \cap D$ for all $t \geq 1$.*

The proof is given in Appendix D. This lemma establishes safety for states certified by the LCB. For future rollouts, the LCB score is evaluated on hallucinated states $\hat{x}_{t+j}(u)$, while safety must be guaranteed for the corresponding noiseless true states $\bar{x}_{t+j}(u)$. This additional gap is handled by the confidence event $\mathcal{E}_t^f$ for the transition function, which controls the rollout mismatch through the recursive error envelope. Throughout the analysis, all candidate rollout states evaluated by the algorithm are assumed to lie in a finite discretization $D \subset \mathcal{X}$. Combining the safety confidence event with the rollout-error bound yields the finite-horizon certificate in Theorem 1.

**Theorem 1** (Explicit certified depth)**.** *Suppose that Assumptions 1–3 hold. Let $D \subset \mathcal{X}$ be a finite domain containing all rollout states evaluated by the certificate and assume $\mathbb{P}(\bigcap_{s\geq 1}(\mathcal{E}_s^f \cap \mathcal{E}_s^g)) \geq 1-\delta$, where $\delta \in (0,1)$. Fix a round $t$ and a candidate input $u \in U^{\mathrm{cand}}$. If $\ell_t(\hat{x}_{t+1}(u)) \geq 0$ and, for every $j = 2, \ldots, m$,*

$$\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell L_y \left( \max_{1\leq s\leq m-1} \sqrt{\beta_t^f}\, \sigma_t^f(\hat{x}_{t+s}(u)) \right) \sum_{q=0}^{m-2} (L_x + L_y L_f)^q, \tag{14}$$

*then $\hat{m}_t(u) = m$. Moreover, with probability at least $1 - \delta_f - \delta_g$, the noiseless true rollout induced by $u$ satisfies $\bar{x}_{t+1:t+m}(u) \in (S_t^{\mathrm{LCB}} \cap D)^m \subseteq (\mathcal{S} \cap D)^m$.*

The proof is given in Appendix E. Theorem 1 justifies the rollout-margin test in Algorithm 1. For each candidate input $u$, the algorithm constructs a hallucinated rollout $\hat{x}_{t+1}(u), \ldots, \hat{x}_{t+m}(u)$ using the GP transition surrogate $\hat{f}_t = \mu_t^f$. Because the true transition map $f$ is unknown, this rollout may differ from the noiseless true rollout. The recursive envelope $\bar{e}_{t,j}(u)$ upper bounds this mismatch on the transition confidence event $\mathcal{E}_t^f$. The theorem gives an explicit sufficient margin under which the algorithm certifies the candidate to full depth $m$. The uncertainty term $\max_{1\leq s\leq m-1} \sqrt{\beta_t^f}\, \sigma_t^f(\hat{x}_{t+s}(u))$ is the largest transition-model uncertainty encountered along the hallucinated rollout, while the geometric factor $\sum_{q=0}^{m-2}(L_x + L_y L_f)^q$ bounds how this uncertainty can accumulate through the NARX recursion. The factor $L_y$ converts output prediction error into state mismatch, and $L_\ell$ converts state mismatch into the required LCB safety-score margin. Thus, any candidate satisfying (14) passes the algorithmic feasibility test and is inserted into $U_{t,m}^{\mathrm{feas}}$. On the confidence event, the noiseless true rollout corresponding to the hallucinated rollout remains inside the certified safe set $S_t^{\mathrm{LCB}}$ for all $m$ steps, and therefore inside the true safe set $\mathcal{S}$ by Lemma 2.

**Remark 2** (Pointwise safety versus future viability)**.** *Pointwise safe learning methods certify a candidate by requiring the next query to be safe, which in our notation corresponds to a condition such as $\ell_t(\hat{x}_{t+1}(u)) \geq 0$. This prevents immediate unsafe evaluations but does not exclude inputs that destroy future feasibility. FA-SAL instead screens candidates by the stronger margin condition*

$$\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell \bar{e}_{t,j}(u), \qquad j = 1, \ldots, m. \tag{15}$$

*Thus, feasibility is defined over the entire certified rollout, not only the next state. The error envelope $\bar{e}_{t,j}(u)$ makes this constraint robust to transition-model uncertainty, and Theorem 1 shows that satisfying this algorithmic constraint implies true $m$-step safety on the confidence event.*

**Theorem 2** ($\epsilon$-interior recovery)**.** *Suppose that Assumptions 1–3 hold and $k_f(x,x), k_g(x,x) \le 1$ for all $x \in \mathcal{X}$. Let $D \subset \mathcal{X}$ be a finite domain containing all rollout states evaluated by the certificate. Assume certified rollouts regularly allow observations that help certify interior candidates. Fix $\delta \in (0,1)$, and assume $\mathbb{P}(\bigcap_{s\ge 1}(\mathcal{E}_s^f \cap \mathcal{E}_s^g)) \ge 1-\delta$. The certification-error radius of Algorithm 2 with fixed $m \ge 2$ satisfies $\epsilon_t(\delta) = O(1/\sqrt{\log t})$. For any $\epsilon > 0$, choose a finite $t^*(\epsilon,\delta)$ such that $\epsilon_s(\delta) \le \epsilon$ for all $s \ge t^*(\epsilon,\delta)$. Then, with probability at least $1-\delta$, for all $t \ge t^*(\epsilon,\delta)$, $U_{t,\epsilon}^{\mathrm{int}} \subseteq U_{t,m}^{\mathrm{feas}} \subseteq U_t^{\mathrm{true}}$. In particular,*

$$U_t^{\mathrm{true}} \setminus U_{t,m}^{\mathrm{feas}} \subseteq \{u \in U^{\mathrm{cand}} : 0 \le G_t^{(m)}(u) < \epsilon\}. \tag{16}$$

The precise observation condition and proof are given in Appendix F. The quantity $\gamma_t$ denotes the maximal information gain of Gaussian processes, which is a standard complexity measure in GP learning and Bayesian optimization [24]. In particular, for the squared exponential kernel in fixed dimension, the maximal information gain satisfies $\gamma_t = \mathcal{O}((\log t)^{d+1})$, where $d$ is the input dimension [24]. Theorem 2 characterizes the feasible set constructed by Algorithm 2. The second inclusion, $U_{t,m}^{\mathrm{feas}} \subseteq U_t^{\mathrm{true}}$, is a soundness guarantee: every candidate certified by the algorithm is truly $m$-step future-safe on the confidence event. Thus, the rollout-margin test does not certify inputs whose noiseless true future trajectories violate safety. The first inclusion $U_{t,\epsilon}^{\mathrm{int}} \subseteq U_{t,m}^{\mathrm{feas}}$ describes how the conservativeness of the certificate decreases with data. At early rounds, large posterior uncertainty in the transition and safety models makes the recursive error envelope and the LCB certificate conservative, so some truly future-safe inputs may be rejected. Once $\epsilon_t(\delta) \le \epsilon$, every candidate whose true $m$-step safety margin is at least $\epsilon$ is included in the algorithmic feasible set. The boundary-layer statement (16) gives the same conclusion from the opposite direction: after $t^*(\epsilon,\delta)$, any truly future-safe candidate that remains uncertified must lie within an $\epsilon$-neighborhood of the $m$-step safety boundary. Hence, the algorithm may remain conservative near the boundary, but it recovers all sufficiently interior future-safe candidates. When $L_x + L_y L_f > 1$, the worst-case data budget sufficient to achieve $\epsilon_t(\delta) \le \epsilon$ can grow exponentially with the horizon $m$, due to the geometric factor $\sum_{r=0}^{m-2}(L_x + L_y L_f)^r$. Therefore, the statement that additional data reduces conservativeness should be understood for each fixed finite horizon $m$.

Theorem 2 concerns recovery of the future-safe feasible set. We next translate this set-level guarantee into the query-level dead-end avoidance property. Since Algorithm 2 selects only from $U_{t,m}^{\mathrm{feas}}$, any candidate excluded by the future-safety screen cannot be selected, regardless of its information-gain value. The proof of the following corollary is given in Appendix G.

**Corollary 1** (Exclusion of dead-end inputs)**.** *Suppose that the conditions of Theorem 2 hold. Fix $\epsilon > 0$, and let $t^*(\epsilon,\delta)$ be defined as in Theorem 2. Then, with probability at least $1-\delta$, for every $t \ge t^*(\epsilon,\delta)$, every $\epsilon$-separated $m$-step dead-end input $u \in U^{\mathrm{cand}}$ is excluded from the certified feasible set, that is, $u \notin U_{t,m}^{\mathrm{feas}}$. Therefore, for every $t \ge t^*(\epsilon,\delta)$, Algorithm 2 does not select an $\epsilon$-separated $m$-step dead-end input.*

## 6 Experiments

We evaluate FA-SAL on three benchmarks: a synthetic NARX environment, a rail-pressure fluid-system, and an NGSIM car-following benchmark. Our evaluation focuses on certified-set soundness (Theorem 1), recovery of future-safe candidates as uncertainty decreases (Theorem 2), and avoidance of one-step-safe but future-unsafe dead-end inputs (Corollary 1).

**Baselines.** We compare FA-SAL against two baselines on the same candidate sets. Since our experiments use trajectory-level queries (Algorithm 2), we consider baselines with trajectory-based sampling rules. SAL-NX [35] is a NARX adaptation of trajectory-based safe active learning, using trajectory-level safety screening and uncertainty-based exploration. Tebbe-ABM [28] is a trajectory-based safe active learning method that certifies candidate trajectories using adaptive Monte Carlo bounds for Gaussian-process safety models.

**Experimental setup.** The main experiments use unit-variance squared-exponential kernels with length scale 1 and fixed GP hyperparameters. We report mean $\pm$ standard deviation over 10 random seeds after 150 trajectory-selection rounds, each executing up to $m$ transitions. Appendix H gives the noise levels, confidence schedule, candidate grids, Lipschitz settings, and NGSIM simulator and split

Table 1: Base synthetic benchmark results. Values are reported as mean ± standard deviation over 10 trials. Lower is better for RMSE, SVR, D-SVR, and false certification rate; higher is better for IoU metrics. The corresponding curves are provided in Section I.1.

| Method | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | Dead-end IoU ↑ | False Cert. ↓ |
|---|---|---|---|---|---|---|
| FA-SAL | $0.468 \pm 0.100$ | $\mathbf{0.059 \pm 0.006}$ | $\mathbf{0.069 \pm 0.012}$ | $0.991 \pm 0.018$ | $0.638 \pm 0.072$ | $\mathbf{0.000 \pm 0.000}$ |
| SAL-NX | $\mathbf{0.415 \pm 0.120}$ | $0.129 \pm 0.007$ | $0.152 \pm 0.007$ | $\mathbf{0.998 \pm 0.002}$ | $0.646 \pm 0.077$ | $\mathbf{0.000 \pm 0.000}$ |
| Tebbe-ABM | $0.423 \pm 0.173$ | $0.122 \pm 0.005$ | $0.149 \pm 0.016$ | $0.996 \pm 0.005$ | $\mathbf{0.656 \pm 0.076}$ | $\mathbf{0.000 \pm 0.000}$ |

Table 2: Penalty-stress synthetic benchmark results.

| Method | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | Dead-end IoU ↑ | False Cert. ↓ |
|---|---|---|---|---|---|---|
| FA-SAL | $\mathbf{2.428 \pm 0.563}$ | $\mathbf{0.018 \pm 0.006}$ | $\mathbf{0.033 \pm 0.013}$ | $0.929 \pm 0.095$ | $\mathbf{0.383 \pm 0.298}$ | $\mathbf{0.000 \pm 0.000}$ |
| SAL-NX | $4.168 \pm 0.979$ | $0.227 \pm 0.036$ | $0.312 \pm 0.051$ | $\mathbf{0.998 \pm 0.002}$ | $0.135 \pm 0.022$ | $0.127 \pm 0.258$ |
| Tebbe-ABM | $3.971 \pm 0.884$ | $0.167 \pm 0.028$ | $0.224 \pm 0.035$ | $0.996 \pm 0.006$ | $0.200 \pm 0.133$ | $0.100 \pm 0.300$ |

details. Appendix J reports horizon and continuation ablations, Lipschitz sensitivity, GP adaptation, a comparison with nominal active-learning MPC, NARX window ablations, runtime, and paired statistical tests.

**Metrics.** We report both model-learning and safety-recovery metrics. The root-mean-squared error (RMSE) measures prediction error of the learned transition model on held-out evaluation regressors. The safety violation rate (SVR) measures the fraction of selected queries whose induced trajectory violates the true safety constraint within the planning horizon. These metrics are commonly used in prior safe active learning to assess both model accuracy and safety during data collection [28, 35]. The dead-end safety violation rate (D-SVR) measures how often a selected query is initially safe but leads to a future unsafe or unrecoverable state. The false certification rate measures whether the algorithmic feasible set contains candidates that are not truly future-safe. Finally, safety-set intersection over union (Safety-Set IoU) measures recovery of the true pointwise safe set, while dead-end-free safety-set IoU measures recovery of the subset of safe candidates that also admit an $m$-step safe continuation.

**Synthetic environment.** The synthetic benchmark is designed to highlight the gap between pointwise safety and finite-horizon future viability. The system follows a scalar-output NARX model with regressor $x_t = (y_t, y_{t-1}, u_t)$, where $y_t$ is the current output, $y_{t-1}$ is the previous output, and $u_t$ is the control input. The transition is inspired by a damped double-integrator model [20]:

$$y_{t+1} = y_t + a(y_t - y_{t-1}) + b_{\text{eff}}(x_t)u_t - c_d(y_t - y_{t-1})|y_t - y_{t-1}| + r(x_t) + \varepsilon_t. \tag{17}$$

Here, $a$ controls inertial drift, $b_{\text{eff}}(x_t)$ is a state-dependent control gain, $c_d$ is a nonlinear drag coefficient, and $r(x_t)$ introduces runaway drift that creates finite-horizon dead-end regions. The pointwise safety function is $g(x_t) = \min\{y_{\max} - y_t, y_t + y_{\max}\}$. Thus, $g(x_t) \geq 0$ only encodes instantaneous safety. Dead-ends are identified by a finite-horizon oracle that checks whether there exists an admissible continuation that remains safe for the next $m$ steps.

Tables 1 and 2 summarize the final-round results on the two synthetic settings. On the base benchmark, all methods achieve zero false certification, indicating that their certified feasible sets remain conservative with respect to the true future-safe set (Table 1). FA-SAL achieves the lowest SVR and D-SVR, while SAL-NX has a slightly lower RMSE and Tebbe-ABM has a slightly higher IoU. This suggests that, in the milder setting, all methods can recover the pointwise safe region reasonably well, but FA-SAL is more effective at avoiding unsafe and dead-end selections.

The difference becomes more pronounced in the penalty-stress benchmark, where entering future-unsafe regions produces more difficult observations and stronger dead-end effects (Table 2). FA-SAL achieves the lowest RMSE, SVR, D-SVR, and false certification rate, and also obtains the highest dead-end IoU. In contrast, SAL-NX and Tebbe-ABM maintain high pointwise Safety IoU but suffer substantially higher safety violation and false certification rates. This gap shows that recovering the pointwise safe set is not sufficient in this environment. Methods must also identify which safe-looking candidates remain viable over the finite horizon. Overall, the synthetic results support the main role of the proposed future-aware certificate, namely, to reject candidates that appear immediately safe but do not admit an $m$-step safe continuation.

**Rail-pressure fluid-system benchmark [28, 35].** We next evaluate FA-SAL on a rail-pressure benchmark. This setting represents a physical control problem in which the learner must

Table 3: Rail-pressure benchmark results. The corresponding curves are provided in Section I.2.

| Method | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | Dead-end IoU ↑ | False Cert. ↓ |
|---|---|---|---|---|---|---|
| FA-SAL | $\mathbf{1.798 \pm 0.068}$ | $\mathbf{0.061 \pm 0.030}$ | $\mathbf{0.058 \pm 0.032}$ | $\mathbf{0.859 \pm 0.033}$ | $\mathbf{0.962 \pm 0.065}$ | $0.010 \pm 0.028$ |
| SAL-NX | $1.939 \pm 0.074$ | $0.087 \pm 0.041$ | $0.085 \pm 0.041$ | $0.834 \pm 0.061$ | $0.884 \pm 0.090$ | $\mathbf{0.000 \pm 0.000}$ |
| Tebbe-ABM | $1.959 \pm 0.064$ | $0.083 \pm 0.038$ | $0.081 \pm 0.037$ | $0.835 \pm 0.071$ | $0.934 \pm 0.056$ | $0.011 \pm 0.033$ |

Table 4: NGSIM collision-absorbing benchmark results. The corresponding curves are provided in Section I.3.

| Method | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE-Free IoU ↑ | False Cert. ↓ |
|---|---|---|---|---|---|---|
| FA-SAL | $\mathbf{2.234 \pm 0.395}$ | $\mathbf{0.290 \pm 0.046}$ | $\mathbf{0.290 \pm 0.046}$ | $\mathbf{0.868 \pm 0.021}$ | $0.915 \pm 0.064$ | $\mathbf{0.000 \pm 0.000}$ |
| SAL-NX | $2.245 \pm 0.363$ | $0.308 \pm 0.030$ | $0.308 \pm 0.030$ | $0.853 \pm 0.033$ | $\mathbf{0.954 \pm 0.046}$ | $\mathbf{0.000 \pm 0.000}$ |
| Tebbe-ABM | $2.337 \pm 0.356$ | $0.314 \pm 0.028$ | $0.314 \pm 0.028$ | $0.857 \pm 0.027$ | $0.941 \pm 0.058$ | $\mathbf{0.000 \pm 0.000}$ |

sample input trajectories while avoiding unsafe pressure responses. The regressor is $x_t = (n_{t-3}, n_{t-2}, n_{t-1}, n_t, u_{t-3}, u_{t-2}, u_{t-1}, u_t)$, where $n_t$ denotes rail speed and $u_t$ denotes the actuation input. The dynamics target is the rail pressure $f(x_t) = p_t$. We define the safety score using a pressure limit $p_{\max}$: $g(x_t) = 1 - \exp\left((p_t - p_{\max})/\lambda_p\right)$ so that $g(x_t) \geq 0$ if and only if $p_t \leq p_{\max}$. We use $p_{\max} = 18.0$ and $\lambda_p = 10.0$.

At each round, the learner selects a short ramp trajectory rather than an isolated point. A candidate query is parameterized by an endpoint in the rail-speed-actuation space; the environment then constructs an $m$-step ramp from the current operating point to the selected endpoint. Safety is evaluated over the full trajectory, which reflects the physical fact that changing the system input requires executing a short transition path. All GP training, planning, and acquisition computations are performed on normalized regressors, while true pressure and safety labels are computed in the original physical units.

Table 3 summarizes the final-round performance on the rail-pressure benchmark. FA-SAL achieves the lowest RMSE, SVR, and D-SVR, indicating that incorporating future-aware safety reduces both prediction error and safety violations when dead-end selections are costly. Its highest Safety IoU and Dead-End IoU further show that the certified feasible set provides a closer approximation to both the pointwise safe region and the subset of safe candidates that remain viable over the rollout horizon.

FA-SAL shows a small but nonzero false certification rate. By contrast, SAL-NX attains zero false certification but lower Safety IoU and Dead-End IoU, suggesting that its feasible set is more conservative and excludes a larger fraction of future-safe candidates. Overall, the rail-pressure results support the intended behavior of FA-SAL: it better preserves finite-horizon safety while improving future-safe trajectory recovery and pressure-dynamics learning.

**NGSIM car-following benchmark [19].** We evaluate FA-SAL on a real-world vehicle trajectory benchmark derived from NGSIM car-following data. For each ego vehicle, we identify its preceding vehicle and construct state-action regressors $x_t = (d_t, \Delta v_t, v_t, a_t)$, where $d_t$ is the headway distance, $\Delta v_t$ is the relative velocity, $v_t$ is the ego velocity, and $a_t$ is the ego acceleration. The dynamics target is the next headway distance, $f(x_t) = d_{t+1}$, and safety is defined by a minimum headway constraint $g(x_t) = d_t - d_{\min}$. We set $d_{\min} = 8$ meters. Since NGSIM is an offline dataset, we construct an empirical counterfactual simulator. The nominal update follows the kinematic model $d_{t+1}^{\text{nom}} = d_t + \Delta t(\Delta v_t - \Delta t\, a_t)$, and a data-driven residual is added using nearest observed transitions. This preserves real trajectory statistics while allowing the learner to evaluate counterfactual acceleration sequences. At each round, the learner selects an acceleration sequence from the finite action set $\mathcal{A} = \{-4, -2, 0, 1, 2\}$ $\mathrm{m/s^2}$, over horizon $m = 4$. The resulting trajectory is then evaluated for finite-horizon safety. The initial GP training set and held-out evaluation set are disjoint by ego-vehicle identity and trajectory segment. A headway violation sends the simulator to an absorbing collision state; Appendix H specifies the residual interpolation, split checks, and collision rule.

Table 4 reports the final-round results on the NGSIM collision-absorbing benchmark. FA-SAL achieves the lowest RMSE, SVR, and D-SVR, together with the highest safety IoU. Since all methods have zero false certification, the key difference is not set soundness but how effectively the certified set guides safe query selection. The lower SVR and D-SVR of FA-SAL indicate that its rollout margin certificate better avoids future-unsafe selections under collision-absorbing dynamics. Although SAL-NX achieves a higher DE-Free IoU, it also incurs higher SVR and D-SVR. This suggests a trade-off between broader recovery of the dead-end-free safe region and conservative finite-horizon

safety preservation. FA-SAL explores boundary regions less aggressively because it only selects candidates certified over the rollout horizon, which can slightly reduce DE-Free IoU but yields safer queries.

## 7 Conclusion

We proposed FA-SAL, a future-aware safe active learning framework that certifies finite-horizon safety in time-series systems using margin-certified GP rollouts and recursive uncertainty propagation. FA-SAL distinguishes pointwise safety from future viability and rejects queries that are immediately safe but may lead to finite-horizon dead-ends. Theoretically, we establish high-probability future-safety guarantees, interior recovery of the future-safe set, and exclusion of separated dead-end inputs. Empirically, FA-SAL reduces safety violations and dead-end selections on synthetic and real-world benchmarks while maintaining competitive transition-model learning.

**Limitations.** FA-SAL requires valid GP confidence bounds and Lipschitz upper bounds; empirical estimates alone do not establish these conditions. Longer horizons can increase conservativeness and computation because transition uncertainty is propagated over more rollout steps. The current analysis uses a finite rollout domain and does not directly provide guarantees for continuous action spaces. The theoretical guarantees assume fixed GP hyperparameters and do not automatically apply to the empirical adaptation procedure in Appendix J.

## Acknowledgements

This work was supported by LG AI Research; and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2022-NR068754) and (No.RS-2023-00211357, Smart Assembler: Robot Active Learning for Unseen Parts Assembly).

## References

[1] Yarden As, Lenart Treven, Carmelo Sferrazza, Stelian Coros, and Andreas Krause. Actsafe: Active exploration with safety constraints for reinforcement learning. In *International Conference on Learning Representations*, volume 2025, pages 29752–29776, 2025.

[2] Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 4. Athena scientific, 2012.

[3] Stephen A Billings. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[4] Alessandro Bottero, Carlos Luis, Julia Vinogradska, Felix Berkenkamp, and Jan R Peters. Information-theoretic safe exploration with gaussian processes. *Advances in Neural Information Processing Systems*, 35:30707–30719, 2022.

[5] Mona Buisson-Fenet, Friedrich Solowjow, and Sebastian Trimpe. Actively learning gaussian process dynamics. In *Learning for dynamics and control*, pages 5–15. PMLR, 2020.

[6] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208, 1995.

[7] Mehdi Fatemi, Taylor W Killian, Jayakumar Subramanian, and Marzyeh Ghassemi. Medical dead-ends and learning to identify high-risk states and treatments. *Advances in Neural Information Processing Systems*, 34:4856–4870, 2021.

[8] Mehdi Fatemi, Shikhar Sharma, Harm Van Seijen, and Samira Ebrahimi Kahou. Dead-ends and secure exploration in reinforcement learning. In *International Conference on Machine Learning*, pages 1873–1881. PMLR, 2019.

[9] Marcello Fiducioso, Sebastian Curi, Benedikt Schumacher, Markus Gwerder, and Andreas Krause. Safe contextual bayesian optimization for sustainable room temperature pid control tuning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5850–5856. International Joint Conferences on Artificial Intelligence Organization, 7 2019.

[10] Steve Heim, Alexander von Rohr, Sebastian Trimpe, and Alexander Badri-Spröwitz. A learnable safety measure. In *Conference on Robot Learning*, pages 627–639. PMLR, 2020.

[11] Achin Jain, Truong Nghiem, Manfred Morari, and Rahul Mangharam. Learning and control using gaussian processes. In *2018 ACM/IEEE 9th international conference on cyber-physical systems (ICCPS)*, pages 140–149. IEEE, 2018.

[12] Emil Keyder and Hector Geffner. The hmdpp planner for planning with probabilities. *Sixth International Planning Competition at ICAPS*, 8:37, 2008.

[13] Andrey Kolobov, Daniel Weld, et al. Sixthsense: Fast and reliable recognition of dead ends in mdps. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pages 1108–1114, 2010.

[14] Andrey Kolobov, Daniel Weld, et al. A theory of goal-oriented mdps with dead ends. *arXiv preprint arXiv:1210.4875*, 2012.

[15] Markus Lange-Hegermann and Christoph Zimmer. Future-aware safe active learning of time varying systems using gaussian processes. *Transactions on Machine Learning Research*, 2025.

[16] Congwen Lu and Joel A Paulson. No-regret bayesian optimization with unknown equality and inequality constraints using exact penalty functions. *IFAC-PapersOnLine*, 55(7):895–902, 2022.

[17] Edwin Lughofer. On-line active learning: A new paradigm to improve practical useability of data stream modeling methods. *Information Sciences*, 415:356–376, 2017.

[18] Quoc Phong Nguyen, Wan Theng Ruth Chew, Le Song, Bryan Kian Hsiang Low, and Patrick Jaillet. Optimistic bayesian optimization with unknown constraints. In *The Twelfth International Conference on Learning Representations*, 2023.

[19] Vincenzo Punzo, Maria Teresa Borzacchiello, and Biagio Ciuffo. On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (ngsim) program data. *Transportation Research Part C: Emerging Technologies*, 19(6):1243–1262, 2011.

[20] Venkatesh G Rao and Dennis S Bernstein. Naive control of the double integrator. *IEEE Control Systems Magazine*, 21(5):86–97, 2001.

[21] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[22] Jeff Schneider. Active learning on non-stationary functions. In *Proceedings of AAAI '95 Fall Symposium on Active Learning*, November 1995.

[23] Jens Schreiter, Duy Nguyen-Tuong, Mona Eberts, Bastian Bischoff, Heiner Markert, and Marc Toussaint. Safe exploration for active learning with gaussian processes. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 133–149. Springer, 2015.

[24] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proc. of the 27th International Conference on Machine Learning*, pages 1015–1022. Omnipress, 2010.

[25] Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *International conference on machine learning*, pages 997–1005. PMLR, 2015.

[26] Yanan Sui, Vincent Zhuang, Joel Burdick, and Yisong Yue. Stagewise safe bayesian optimization with gaussian processes. In *International conference on machine learning*, pages 4781–4789. PMLR, 2018.

[27] Bhavya Sukhija, Lenart Treven, Cansu Sancaktar, Sebastian Blaes, Stelian Coros, and Andreas Krause. Optimistic active exploration of dynamical systems. *Advances in Neural Information Processing Systems*, 36:38122–38153, 2023.

[28] Jörn Tebbe, Christoph Zimmer, Ansgar Steland, Markus Lange-Hegermann, and Fabian Mies. Efficiently computable safety bounds for gaussian processes in active learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1333–1341. PMLR, 2024.

[29] 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.

[30] Jonas Umlauft, Thomas Beckers, Alexandre Capone, Armin Lederer, and Sandra Hirche. Smart forgetting for safe online learning with gaussian processes. In *Learning for dynamics and control*, pages 160–169. PMLR, 2020.

[31] Wenjie Xu, Yuning Jiang, Bratislav Svetozarevic, and Colin Jones. Constrained efficient global optimization of expensive black-box functions. In *International Conference on Machine Learning*, pages 38485–38498. PMLR, 2023.

[32] Xiao Zhang, Hai Zhang, Hongtu Zhou, Chang Huang, Di Zhang, Chen Ye, and Junqiao Zhao. Safe reinforcement learning with dead-ends avoidance and recovery. *IEEE Robotics and Automation Letters*, 9(1):491–498, 2023.

[33] Yangyang Zhao, Mehdi Dastani, Jinchuan Long, Zhenyu Wang, and Shihan Wang. Rescue conversations from dead-ends: Efficient exploration for task-oriented dialogue policy optimization. *Transactions of the Association for Computational Linguistics*, 12:1578–1596, 2024.

[34] Christoph Zimmer, Danny Driess, Mona Meister, and Nguyen-Tuong Duy. Adaptive discretization for evaluation of probabilistic cost functions. In *International Conference on Artificial Intelligence and Statistics*, pages 2098–2108. PMLR, 2020.

[35] Christoph Zimmer, Mona Meister, and Duy Nguyen-Tuong. Safe active learning for time-series modeling with gaussian processes. *Advances in neural information processing systems*, 31, 2018.

## Contents

**A Extended Related Work** **15**

**B Algorithm** **17**

**C Definitions** **19**

**D Proof for auxiliary lemmas** **21**

D.1 Proof of Lemma 2 . . . . . 21

**E Proof for Theorem 1** **21**

E.1 Auxiliary lemmas for Theorem 1 . . . . . 22

**F Proof for Theorem 2** **26**

F.1 Auxiliary lemmas for Theorem 2 . . . . . 28

**G Proof of Corollary 1** **33**

**H Experimental Setup and Reproducibility** **34**

**I Experimental Curves** **35**

I.1 Synthetic datasets . . . . . 35

I.2 Rail-pressure fluid-system dataset . . . . . 36

I.3 NGSIM vehicle-trajectory dataset . . . . . 37

**J Additional Experimental Results** **38**

J.1 Horizon and continuation ablations . . . . . 38

J.2 Sensitivity to Lipschitz constants . . . . . 39

J.3 Periodic GP hyperparameter adaptation . . . . . 40

J.4 Comparison with nominal active-learning MPC . . . . . 41

J.5 NARX window-size ablation . . . . . 42

J.6 Runtime and paired statistical tests . . . . . 42

Table 5: Main notation used in the paper.

| Notation | Meaning |
|---|---|
| $t, T$ | Round index and active-learning round budget. |
| $\mathcal{U}, u_t$ | Domain of manipulated inputs and input applied at round $t$. |
| $y_t, y_t^g$ | Physical output and safety measurement at round $t$. |
| $h_t$ | Observation history $h_t = \{(u_s, y_s, y_s^g)\}_{s=1}^t$. |
| $y_{1:t}, y_{1:t}^g$ | Physical and safety observations collected through round $t$. |
| $\mathcal{X}, x_t$ | NARX regression-state domain and regression state formed from recent outputs and inputs. |
| $d_y, d_u$ | Output and input history window lengths. |
| $D \subset \mathcal{X}$ | Finite domain used for GP confidence and evaluation. |
| $m, U^{\mathrm{cand}}$ | Planning horizon and fixed finite candidate input set. |
| $f : \mathcal{X} \to \mathbb{R}, g : \mathcal{X} \to \mathbb{R}$ | Unknown transition map and safety function; a state is safe if $g(x) \geq 0$. |
| $k_f, k_g$ | Covariance kernels of the transition and safety GPs. |
| $\mathcal{H}_{k_f}, \mathcal{H}_{k_g}$ | RKHSs induced by $k_f$ and $k_g$. |
| $\varepsilon_t, \varepsilon_t^g, \sigma_\varepsilon$ | Physical and safety observation noises, and their common sub-Gaussian parameter. |
| $k_t^f(x), K_t^f$ | Transition-GP covariance vector and Gram matrix. |
| $\mu_t^q(x), \sigma_t^q(x), q \in \{f, g\}$ | GP posterior mean and posterior standard deviation for model $q$. |
| $\beta_t^q$ | Confidence parameter for model $q$. |
| $\pi_t$ | Positive sequence satisfying $\sum_{t \geq 1} \pi_t^{-1} = 1$ in the GP confidence schedule. |
| $\mathcal{E}_t^f, \mathcal{E}_t^g, \delta_f, \delta_g$ | GP confidence events and their failure probabilities. |
| $\hat{f}_t := \mu_t^f$ | Transition surrogate used to generate hallucinated rollouts. |
| $\mathcal{S}$ | True safe set $\mathcal{S} = \{x \in \mathcal{X} : g(x) \geq 0\}$. |
| $\ell_t(x)$ | LCB safety score $\ell_t(x) = \mu_t^g(x) - \sqrt{\beta_t^g}\sigma_t^g(x)$. |
| $S_t^{\mathrm{LCB}}$ | LCB-certified safe set $S_t^{\mathrm{LCB}} = \{x \in \mathcal{X} : \ell_t(x) \geq 0\}$. |
| $\Phi$ | NARX shift/update operator. |
| $\Pi_t(u), u_{t+j+1}^{\mathrm{plan}}(u)$ | Planned continuation sequence and its rollout-step input after candidate $u$. |
| $\hat{x}_{t+j}(u), \hat{y}_{t+j}(u)$ | Hallucinated rollout state and output generated using $\hat{f}_t$. |
| $\bar{x}_{t+j}(u), \bar{y}_{t+j}(u)$ | Noiseless true rollout state and output generated using $f$ under the same continuation. |
| $X_t^{\mathrm{plan}}(u)$ | Planned rollout query locations associated with candidate $u$. |
| $e_{t,j}(u), \bar{e}_{t,j}(u)$ | Rollout mismatch $e_{t,j}(u) = \|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\|$ and its computable upper bound. |
| $L_f, L_g, L_\ell$ | Lipschitz constants of the transition map, safety function, and LCB safety score. |
| $L_x, L_y$ | Bounds on the effects of regressor and output changes under the NARX shift operator with a common input. |
| $U_{t,j}^{\mathrm{safe}}(u)$ | One-step certified continuation set used during rollout construction. |
| $\hat{m}_t(u)$ | Certified rollout depth. |
| $G_t^{(m)}(u)$ | True future-safety score $G_t^{(m)}(u) = \min_{1 \leq j \leq m} g(\bar{x}_{t+j}(u))$. |
| $U_t^{\mathrm{true}}$ | True future-safe candidate set $\{u \in U^{\mathrm{cand}} : G_t^{(m)}(u) \geq 0\}$. |
| $\epsilon$ | Safety margin used to define interior future-safe candidates and separated dead-end inputs. |
| $U_{t,\epsilon}^{\mathrm{int}}$ | Interior future-safe candidate set $\{u \in U^{\mathrm{cand}} : G_t^{(m)}(u) \geq \epsilon\}$. |
| $\widehat{G}_t^{(m)}(u)$ | Certified future-safety score $\min_{1 \leq j \leq m}\{\ell_t(\hat{x}_{t+j}(u)) - L_\ell \bar{e}_{t,j}(u)\}$. |
| $U_{t,m}^{\mathrm{feas}}$ | Candidate inputs whose rollouts are certified over the full horizon $m$. |
| $y(X), y^g(X)$ | Hypothetical physical and safety measurements at query locations $X$. |
| $I_t^f(u), I_t^g(u)$ | Rollout information gains for the transition and safety GPs. |
| $C_\eta := 2(\sigma_\varepsilon^2 + 1)$ | Constant used in the information-gain bounds. |
| $\gamma_t$ | Larger of the two maximal information gains from at most $mt$ observations. |
| $\delta$ | Overall confidence failure probability, allocated between the transition and safety models. |

| Notation | Meaning |
|---|---|
| $\epsilon_t(\delta)$ | Certification-error radius appearing in the interior recovery guarantee. |
| $t^*(\epsilon, \delta)$ | Round threshold for recovering the $\epsilon$-interior future-safe set. |
| $K, z_1, \ldots, z_K$ | Number and locations of fixed reference states in an initially certified safe region within $D$. |
| $A$ | Uniform bound on the Euclidean norms of the chosen reference-feature representation coefficients. |
| $n_t, r_t$ | Number of past rounds using (13), and $2A\sqrt{K}\max_{q\in\{f,g\},\, 1\le i\le K}\sigma_t^q(z_i)$. For a feature represented with coefficient norm at most $A$, $r_t/2$ bounds the posterior standard deviation. |
| $\rho$ | Minimum fraction of the $n_t$ rounds in which a state without such a representation must be queryable on a certified rollout. |
| $\mathcal{I}, v_s$ | Past rounds using (13) and their feasible comparison candidates. |
| $K_s^q(Z), K_s(u)$ | Posterior covariance matrix of model $q$ on the ordered path $Z$ given $h_s$, with $K_s(u) := K_s^f(X_s^{\mathrm{plan}}(u))$. |

# A Extended Related Work

**Safe active learning with Gaussian processes.** Safe active learning with Gaussian processes aims to select informative queries while maintaining safety during data collection [23, 28, 34, 35]. These methods typically use GP posteriors to construct conservative safe sets, and then restrict query selection to points or trajectories that satisfy the corresponding safety criterion. For time-series models, safe active learning becomes more delicate because each query affects not only the next observation but also subsequent regression states. Prior work has developed safe active learning methods for time-series modeling and efficient GP safety bounds [28, 34, 35]. Recent future-aware safe active learning for time-varying systems uses future states as prediction targets in the acquisition objective, for example by reducing posterior variance over current and future system states [15]. Our notion of future safety is different: future states do not only determine where information should be collected, but also which current queries are admissible. In particular, we formulate future viability in a NARX regression-state representation and introduce a margin-certified rollout condition that accounts for transition-model error and excludes immediately safe queries that lead to finite-horizon dead ends.

**Active learning and safe exploration for dynamical systems.** Active learning and safe exploration have also been studied in dynamical systems, online system identification, and control-oriented learning. Work on online and nonstationary active learning focuses on informative sampling under time-varying data distributions [17, 22], while active learning for dynamical models selects informative trajectories or excitation signals to improve transition-model estimation [5, 30]. Control-oriented safe exploration uses learned dynamics, uncertainty estimates, or safety measures to guide interaction while avoiding unsafe states [1, 4, 10, 11, 27, 29]. Our setting is related, but the active-learning objective is formulated at the level of NARX time-series regressors: the learner selects informative inputs for estimating an unknown transition map, while the safety certificate must hold over the future rollout induced by the query. This distinction is important because a query can be safe at the next step while moving the regressor history into a region from which no safe continuation exists.

**Safe and constrained Bayesian optimization.** Safe Bayesian optimization and constrained Bayesian optimization study black-box optimization under unknown safety or feasibility constraints [9, 16, 18, 25, 26, 31]. These methods provide confidence-based mechanisms for avoiding unsafe evaluations and expanding the safe region. However, they are usually formulated around pointwise design variables or optimization objectives. Our setting is different in that the goal is not to optimize an unknown reward, but to actively learn the transition model of a time-series system. Moreover, the safety constraint is finite-horizon viability.

**Dead-ends in planning and reinforcement learning.** Dead-end states have been studied in stochastic shortest path problems, goal-oriented planning, medical decision making, safe reinforcement learning, and dialogue policy optimization [7, 8, 12–14, 32, 33]. In these settings, a state may be locally nonterminal or safe, yet entering it can irreversibly eliminate the possibility of reaching a goal or avoiding failure. We adapt this perspective to safe active learning for time-series systems. Since a NARX state is a shifting history vector rather than an absorbing MDP state, we do not define dead-ends as literal absorbing states. Instead, we define $m$-step dead-end inputs as inputs that are immediately safe but whose induced rollout fails to admit safe continuation over a finite horizon. The proposed future-aware certificate operationalizes this idea by rejecting candidates whose hallucinated rollouts cannot be certified with sufficient margin.

## B Algorithm

**Algorithm 1** Future-Aware Safe Active Learning (FA-SAL)

**Require:** history $h_1$, budget $T$, planning horizon $m$, candidate input set $U^{\mathrm{cand}}$

**for** $t = 1, 2, \ldots, T$ **do**
  Update the GP posteriors of $f$ and $g$
  Set $\ell_t(z) \leftarrow \mu_t^g(z) - \sqrt{\beta_t^g}\,\sigma_t^g(z)$, $U_{t,m}^{\mathrm{feas}} \leftarrow \emptyset$
  **for all** $u \in U^{\mathrm{cand}}$ **do**
    Initialize $\hat{x}_{t+1}(u) \leftarrow \big(y_t, \ldots, y_{t-d_y+1}, u, u_t, \ldots, u_{t-d_u+2}\big)$, $\bar{e}_{t,1}(u) \leftarrow 0$, $\hat{m}_t(u) \leftarrow 0$
    **for** $j = 1, 2, \ldots, m$ **do**
      **if** $\ell_t(\hat{x}_{t+j}(u)) < L_\ell \bar{e}_{t,j}(u)$ **then**
        **break**
      **end if**
      $\hat{m}_t(u) \leftarrow j$
      **if** $j < m$ **then**
        Set $\hat{y}_{t+j}(u) \leftarrow \mu_t^f(\hat{x}_{t+j}(u))$
        Update the error envelope

$$\bar{e}_{t,j+1}(u) \leftarrow (L_x + L_y L_f)\bar{e}_{t,j}(u) + L_y\sqrt{\beta_t^f}\,\sigma_t^f(\hat{x}_{t+j}(u))$$

        Define the one-step certified continuation set

$$U_{t,j}^{\mathrm{safe}}(u) := \big\{v \in \mathcal{U} : \ell_t\big(\Phi(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), v)\big) \geq L_\ell \bar{e}_{t,j+1}(u)\big\}$$

        Choose $u_{t+j+1}^{\mathrm{plan}}(u) \in \arg\max_{v \in U_{t,j}^{\mathrm{safe}}(u)} \sigma_t^f\big(\Phi(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), v)\big)$
        Update $\hat{x}_{t+j+1}(u) \leftarrow \Phi\big(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\big)$
      **end if**
    **end for**
    **if** $\hat{m}_t(u) = m$ **then**
      $U_{t,m}^{\mathrm{feas}} \leftarrow U_{t,m}^{\mathrm{feas}} \cup \{u\}$
      Define $X_t^{\mathrm{plan}}(u) := \{\hat{x}_{t+1}(u), \ldots, \hat{x}_{t+m}(u)\}$
      Set

$$I_t^f(u) = I(f(X_t^{\mathrm{plan}}(u)); y(X_t^{\mathrm{plan}}(u)) \mid h_t),$$
$$I_t^g(u) = I(g(X_t^{\mathrm{plan}}(u)); y^g(X_t^{\mathrm{plan}}(u)) \mid h_t)$$

    **end if**
  **end for**
  Select

$$u_{t+1} \in \arg\max_{u \in U_{t,m}^{\mathrm{feas}}} 2\sqrt{\beta_t^g}\sqrt{mC_\eta I_t^g(u)} + (L_g + L_\ell)L_y\sqrt{\beta_t^f}\sqrt{mC_\eta I_t^f(u)}\sum_{r=0}^{m-2}(L_x + L_y L_f)^r$$

  Apply $u_{t+1}$, observe $y_{t+1}$ and $y_{t+1}^g$, and update $h_{t+1}$
**end for**

**Algorithm 2** Trajectory-Query Variant of FA-SAL

**Require:** history $h_1$, budget $T$, planning horizon $m$, candidate input set $U^{\mathrm{cand}}$, initially certified reference states $z_1, \dots, z_K$

1: Follow Algorithm 1 up to the construction of the certified feasible set $U_{t,m}^{\mathrm{feas}}$ and the certified hallucinated rollout $X_t^{\mathrm{plan}}(u)$ for each feasible $u$.
2: If $\max_{q\in\{f,g\},\, 1\le i\le K} \sigma_t^q(z_i) \le 1/\log(e+t)$, select $u_{t+1}$ by (13).
3: Otherwise, select any $u_{t+1} \in U_{t,m}^{\mathrm{feas}}$ whose queried rollout contains a reference state $z_i$ with $\sigma_t^q(z_i) > 1/\log(e+t)$ for at least one $q \in \{f, g\}$.
4: Query the certified hallucinated trajectory $X_t^{\mathrm{plan}}(u_{t+1})$, observe $y$ and $y^g$ at each queried state, and update the history with both measurements.

## C Definitions

In this section, we collect the main objects used throughout the paper.

**Definition 2** (True safe set)**.** *Let $g : \mathcal{X} \to \mathbb{R}$ denote the unknown safety function. The true safe set is defined by*

$$\mathcal{S} := \{x \in \mathcal{X} : g(x) \geq 0\}. \tag{18}$$

*A state $x$ is called* safe *if $x \in \mathcal{S}$.*

**Definition 3** (LCB safety score and certified safe set)**.** *At round $t$, define the lower confidence bound safety score by*

$$\ell_t(x) := \mu_t^g(x) - \sqrt{\beta_t^g}\,\sigma_t^g(x), \qquad x \in \mathcal{X}. \tag{19}$$

*The corresponding certified safe set is*

$$S_t^{\mathrm{LCB}} := \{x \in \mathcal{X} : \ell_t(x) \geq 0\}. \tag{20}$$

**Definition 4** ($m$-step future-safety score)**.** *Fix a round $t$, a planning horizon $m \geq 1$, and a candidate input $u \in U^{\mathrm{cand}}$. Let $\bar{x}_{t+1}(u), \ldots, \bar{x}_{t+m}(u)$ denote the true rollout induced by $u$ and the continuation rule used by the algorithm. The true $m$-step future-safety score of $u$ is*

$$G_t^{(m)}(u) := \min_{1 \leq j \leq m} g\big(\bar{x}_{t+j}(u)\big). \tag{21}$$

*Thus, $G_t^{(m)}(u) \geq 0$ if and only if the true rollout remains safe for the next $m$ steps.*

**Definition 5** ($m$-step dead-end state)**.** *Fix a planning horizon $m \geq 1$. A state $x_t \in \mathcal{S}$ is called an* $m$-step dead-end state *if there exists no admissible continuation input sequence that keeps the trajectory safe for the next $m$ steps. Equivalently,*

$$x_t \in \mathcal{S} \quad \textit{and} \quad \forall (u_{t+1}, \ldots, u_{t+m}) \in \mathcal{U}^m,\ \exists j \in \{1, \ldots, m\} \textit{ such that } x_{t+j} \notin \mathcal{S}. \tag{22}$$

**Definition 6** ($m$-step dead-end input)**.** *Fix a round $t$ and a candidate input $u \in U^{\mathrm{cand}}$. We say that $u$ is an* $m$-step dead-end input *if*

$$g\big(\bar{x}_{t+1}(u)\big) \geq 0 \quad \textit{but} \quad G_t^{(m)}(u) < 0. \tag{23}$$

*Equivalently, the input is immediately safe at the next step, but the induced rollout violates safety within the next $m$ steps.*

**Definition 7** ($\epsilon$-separated $m$-step dead-end input)**.** *For $\epsilon \geq 0$, a candidate input $u \in U^{\mathrm{cand}}$ is called an* $\epsilon$-separated $m$-step dead-end input *if*

$$g\big(\bar{x}_{t+1}(u)\big) \geq \epsilon \quad \textit{and} \quad G_t^{(m)}(u) \leq -\epsilon. \tag{24}$$

*Thus, the input is safely inside the one-step safe region, yet it is separated by a margin $\epsilon$ from being future-safe over horizon $m$.*

**Definition 8** (Certified rollout depth)**.** *For each candidate input $u \in U^{\mathrm{cand}}$, define the certified rollout depth $\hat{m}_t(u) \in \{0, 1, \ldots, m\}$ by*

$$\hat{m}_t(u) := \max\big\{r \in \{0, 1, \ldots, m\} : \ell_t\big(\hat{x}_{t+j}(u)\big) \geq L_\ell \bar{e}_{t,j}(u),\ \forall j = 1, \ldots, r\big\}. \tag{25}$$

*That is, $\hat{m}_t(u)$ is the largest step up to which the predicted rollout remains inside the safety margin required to absorb rollout mismatch.*

**Definition 9** ($m$-step algorithmic feasible set)**.** *For each round $t$, define the algorithmic feasible set by*

$$U_{t,m}^{\mathrm{feas}} := \{u \in U^{\mathrm{cand}} : \hat{m}_t(u) = m\}. \tag{26}$$

*Thus, a candidate is feasible if and only if the predicted rollout can be certified up to the full planning horizon $m$.*

**Definition 10** (True $m$-step future-safe set)**.** *For each round $t$, define*

$$U_t^{\mathrm{true}} := \{u \in U^{\mathrm{cand}} : G_t^{(m)}(u) \geq 0\}. \tag{27}$$

*This is the set of candidates whose true rollout is safe for all of the next $m$ steps.*

**Definition 11** ($\epsilon$-interior future-safe set)**.** *For $\epsilon > 0$, define*

$$U_{t,\epsilon}^{\mathrm{int}} := \{u \in U^{\mathrm{cand}} : G_t^{(m)}(u) \geq \epsilon\}. \tag{28}$$

*This is the subset of truly future-safe candidates that remain at least $\epsilon$ away from the $m$-step safety boundary.*

**Definition 12** (Certified $m$-step future-safety score)**.** *For each candidate input $u \in U^{\mathrm{cand}}$, define*

$$\widehat{G}_t^{(m)}(u) := \min_{1 \leq j \leq m} \Big( \ell_t\big(\hat{x}_{t+j}(u)\big) - L_\ell \bar{e}_{t,j}(u) \Big). \tag{29}$$

*If $\widehat{G}_t^{(m)}(u) \geq 0$, then the predicted rollout satisfies the margin condition at every step up to horizon $m$.*

**Definition 13** (Planned continuation rollout)**.** *Fix a round $t$, history $h_t$, and a candidate current input $u \in U^{\mathrm{cand}}$. Given $u$, we define a planned continuation sequence*

$$\Pi_t(u) := \big(u_{t+2}^{\mathrm{plan}}(u), \ldots, u_{t+m}^{\mathrm{plan}}(u)\big). \tag{30}$$

*The predicted rollout induced by $u$ and $\Pi_t(u)$ is recursively defined by*

$$\hat{x}_{t+1}(u) := \big(y_t, \ldots, y_{t-d_y+1}, u, u_t, \ldots, u_{t-d_u+2}\big), \tag{31}$$

$$\hat{y}_{t+j}(u) := \mu_t^f(\hat{x}_{t+j}(u)), \qquad j = 1, \ldots, m-1, \tag{32}$$

$$\hat{x}_{t+j+1}(u) := \Phi\big(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\big), \qquad j = 1, \ldots, m-1. \tag{33}$$

*The corresponding true rollout is defined by*

$$\bar{x}_{t+1}(u) := \hat{x}_{t+1}(u), \tag{34}$$

$$\bar{y}_{t+j}(u) := f(\bar{x}_{t+j}(u)), \qquad j = 1, \ldots, m-1, \tag{35}$$

$$\bar{x}_{t+j+1}(u) := \Phi(\bar{x}_{t+j}(u), \bar{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)), \qquad j = 1, \ldots, m-1. \tag{36}$$

**Remark 3** (dead-ends in NARX systems versus MDPs)**.** *In Markov decision processes, dead-ends are often modeled as absorbing states: once entered, no action can recover a desirable outcome [8]. In contrast, under the NARX formulation the state is a shifting history vector of past outputs and inputs, so exact absorption is generally not the right notion. For this reason, we define dead-ends in terms of loss of future safety viability rather than literal state absorption: a state or input is dead-end if it is currently safe but admits no safe continuation over the prescribed planning horizon.*

## D Proof for auxiliary lemmas

### D.1 Proof of Lemma 2

*Proof of Lemma 2.* On $\mathcal{E}_t^g$, for every $z \in D$, we have

$$g(z) \geq \mu_t^g(z) - \sqrt{\beta_t^g}\sigma_t^g(z) = \ell_t(z). \tag{37}$$

Therefore, if $z \in \mathcal{S}_t^{\mathrm{LCB}} \cap D$, then $\ell_t(z) \geq 0$, and hence

$$g(z) \geq \ell_t(z) \geq 0. \tag{38}$$

Thus $z \in \mathcal{S} \cap D$, which proves

$$S_t^{\mathrm{LCB}} \cap D \subseteq \mathcal{S} \cap D. \tag{39}$$

□

**Lemma 3** (Lipschitz continuity of the GP posterior mean)**.** *Let $\mu_t(x) = k_t(x)^\top (K_t + \lambda I)^{-1} y_t$ where $k_t(x) := \big(k(x_1, x), \ldots, k(x_t, x)\big)^\top$. Suppose that for each point $x_i$, the function $x \mapsto k(x_i, x)$ is $L_{k,i}$-Lipschitz. Then, $\mu_t$ is Lipschitz and*

$$|\mu_t(x) - \mu_t(y)| \leq \left(\sum_{i=1}^{t} |a_i| L_{k,i}\right) \|x - y\|, \tag{40}$$

*where $a := (K_t + \lambda I)^{-1} y_t$.*

*Proof.* Define $a = (a_1, \ldots, a_t)^\top := (K_t + \lambda I)^{-1} y_t$. We may write

$$\mu_t(x) = \sum_{i=1}^{t} a_i \, k(x_i, x). \tag{41}$$

Therefore,

$$|\mu_t(x) - \mu_t(y)| = \left|\sum_{i=1}^{t} a_i \big(k(x_i, x) - k(x_i, y)\big)\right| \tag{42}$$

$$\leq \sum_{i=1}^{t} |a_i| \, |k(x_i, x) - k(x_i, y)| \tag{43}$$

$$\leq \left(\sum_{i=1}^{t} |a_i| L_{k,i}\right) \|x - y\|. \tag{44}$$

Hence, $\mu_t$ is Lipschitz. □

**Remark 4** (Lipschitz continuity of the LCB score)**.** *If both the posterior mean $\mu_t$ and posterior standard deviation $\sigma_t$ are Lipschitz on $\mathcal{X}$, then the lower confidence bound*

$$\ell_t(x) := \mu_t(x) - \sqrt{\beta_t}\, \sigma_t(x) \tag{45}$$

*is also Lipschitz, with*

$$\mathrm{Lip}(\ell_t) \leq \mathrm{Lip}(\mu_t) + \sqrt{\beta_t}\, \mathrm{Lip}(\sigma_t). \tag{46}$$

## E Proof for Theorem 1

*Proof of Theorem 1.* From Lemma 4, $\bar{e}_{t,1}(u) = 0$ and

$$\bar{e}_{t,j+1}(u) = (L_x + L_y L_f)\bar{e}_{t,j}(u) + L_y \sqrt{\beta_t^f}\, \sigma_t^f(\hat{x}_{t+j}(u)), \qquad j = 1, \ldots, m-1. \tag{47}$$

Unrolling the recursion gives, for $j \geq 2$,

$$\bar{e}_{t,j}(u) = L_y \sum_{s=1}^{j-1} (L_x + L_y L_f)^{j-1-s} \sqrt{\beta_t^f}\, \sigma_t^f(\hat{x}_{t+s}(u)). \tag{48}$$

Hence, we have

$$\bar{e}_{t,j}(u) \leq L_y \left( \max_{1 \leq s \leq j-1} \sqrt{\beta_t^f}\, \sigma_t^f(\hat{x}_{t+s}(u)) \right) \sum_{q=0}^{j-2} (L_x + L_y L_f)^q. \tag{49}$$

Therefore, if for each $j = 2, \ldots, r$,

$$\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell L_y \left( \max_{1 \leq s \leq j-1} \sqrt{\beta_t^f}\, \sigma_t^f(\hat{x}_{t+s}(u)) \right) \sum_{q=0}^{j-2} (L_x + L_y L_f)^q, \tag{50}$$

then

$$\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell \bar{e}_{t,j}(u), \qquad j = 2, \ldots, r. \tag{51}$$

Since $\bar{e}_{t,1}(u) = 0$, the additional condition $\ell_t(\hat{x}_{t+1}(u)) \geq 0$ is exactly

$$\ell_t(\hat{x}_{t+1}(u)) \geq L_\ell \bar{e}_{t,1}(u). \tag{52}$$

Thus

$$\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell \bar{e}_{t,j}(u), \qquad j = 1, \ldots, r. \tag{53}$$

By Definition 8, this implies

$$\hat{m}_t(u) \geq r. \tag{54}$$

In particular, if the condition holds for all $j = 1, \ldots, m$, then

$$\hat{m}_t(u) = m. \tag{55}$$

The final safety claim then follows from Lemma 5. □

### E.1 Auxiliary lemmas for Theorem 1

In this section, we establish the margin-based certificate that connects the predicted rollout under the NARX model to true finite-horizon safety. We first derive a recursive bound on the mismatch between the true and predicted rollouts, and then show that if the predicted trajectory remains sufficiently inside the LCB safe region, the corresponding true trajectory stays in the certified safe set over the entire horizon. These results provide the main technical step used in the proof of Theorem 1.

**Lemma 4** (Recursive noiseless rollout mismatch bound and computable error envelope)**.** *Fix $t$ and $u \in U^{\mathrm{cand}}$. Suppose that Assumptions 1-3 hold. Let the noiseless rollout $\bar{x}_{t+1}(u), \ldots, \bar{x}_{t+m}(u)$ be generated by the true transition map $f$ under the same planned continuation as the hallucinated rollout:*

$$\bar{x}_{t+1}(u) := \hat{x}_{t+1}(u), \tag{56}$$

$$\bar{y}_{t+j}(u) := f(\bar{x}_{t+j}(u)), \qquad j = 1, \ldots, m-1, \tag{57}$$

$$\bar{x}_{t+j+1}(u) := \Phi\Big(\bar{x}_{t+j}(u), \bar{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\Big), \qquad j = 1, \ldots, m-1. \tag{58}$$

*The hallucinated rollout is generated by*

$$\hat{y}_{t+j}(u) := \mu_t^f(\hat{x}_{t+j}(u)), \qquad j = 1, \ldots, m-1, \tag{59}$$

$$\hat{x}_{t+j+1}(u) := \Phi\Big(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\Big), \qquad j = 1, \ldots, m-1. \tag{60}$$

*Define the rollout mismatch by $e_{t,j}(u) := \|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\|$, for $j = 1, \ldots, m$. Then, on the transition confidence event $\mathcal{E}_t^f$, $e_{t,1}(u) = 0$, and for each $j = 1, \ldots, m-1$,*

$$e_{t,j+1}(u) \leq (L_x + L_y L_f) e_{t,j}(u) + L_y \sqrt{\beta_t^f}\, \sigma_t^f(\hat{x}_{t+j}(u)). \tag{61}$$

*Consequently, if we define the computable error envelope $\bar{e}_{t,j}(u)$ by*

$$\bar{e}_{t,1}(u) := 0, \tag{62}$$

$$\bar{e}_{t,j+1}(u) := (L_x + L_y L_f)\bar{e}_{t,j}(u) + L_y\sqrt{\beta_t^f}\,\sigma_t^f(\hat{x}_{t+j}(u)), \qquad j = 1, \dots, m-1, \tag{63}$$

*then*

$$e_{t,j}(u) = \|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| \le \bar{e}_{t,j}(u), \qquad j = 1, \dots, m, \tag{64}$$

*on $\mathcal{E}_t^f$.*

*Proof.* Fix a round $t$, a candidate input $u \in U^{\mathrm{cand}}$, and a rollout step $j \in \{1, \dots, m-1\}$. We compare two rollouts generated under the same planned continuation

$$u_{t+2}^{\mathrm{plan}}(u), \dots, u_{t+m}^{\mathrm{plan}}(u). \tag{65}$$

The first rollout is the noiseless rollout generated by the true transition map $f$:

$$\bar{y}_{t+j}(u) = f(\bar{x}_{t+j}(u)), \qquad \bar{x}_{t+j+1}(u) = \Phi\Big(\bar{x}_{t+j}(u), \bar{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\Big). \tag{66}$$

The second rollout is the hallucinated rollout generated by the GP posterior mean:

$$\hat{y}_{t+j}(u) = \mu_t^f(\hat{x}_{t+j}(u)), \qquad \hat{x}_{t+j+1}(u) = \Phi\Big(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\Big). \tag{67}$$

Both rollouts use the same initial next regressor, $\bar{x}_{t+1}(u) = \hat{x}_{t+1}(u)$, and the same continuation inputs. Therefore, any mismatch between the two rollouts is caused only by the difference between the true transition output $f(\bar{x}_{t+j}(u))$ and the hallucinated output $\mu_t^f(\hat{x}_{t+j}(u))$. Define

$$e_{t,j}(u) := \|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\|. \tag{68}$$

Since $\bar{x}_{t+1}(u) = \hat{x}_{t+1}(u)$, we immediately have $e_{t,1}(u) = 0$. We now derive the recursive bound. By the definitions of the two rollouts,

$$e_{t,j+1}(u) = \|\bar{x}_{t+j+1}(u) - \hat{x}_{t+j+1}(u)\| \tag{69}$$

$$= \Big\|\Phi\Big(\bar{x}_{t+j}(u), \bar{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\Big) - \Phi\Big(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\Big)\Big\|. \tag{70}$$

The two arguments of $\Phi$ use the same continuation input $u_{t+j+1}^{\mathrm{plan}}(u)$. Hence, applying Lemma 6 gives

$$e_{t,j+1}(u) = \|\bar{x}_{t+j+1}(u) - \hat{x}_{t+j+1}(u)\| \tag{71}$$

$$= \big\|\Phi\big(\bar{x}_{t+j}(u), \bar{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\big) - \Phi\big(\hat{x}_{t+j}(u), \hat{y}_{t+j}(u), u_{t+j+1}^{\mathrm{plan}}(u)\big)\big\| \tag{72}$$

$$\le L_x\|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| + L_y|\bar{y}_{t+j}(u) - \hat{y}_{t+j}(u)| \tag{73}$$

$$= L_x e_{t,j}(u) + L_y|\bar{y}_{t+j}(u) - \hat{y}_{t+j}(u)|. \tag{74}$$

It remains to bound the output mismatch. Since $\bar{y}_{t+j}(u) = f(\bar{x}_{t+j}(u))$ and $\hat{y}_{t+j}(u) = \mu_t^f(\hat{x}_{t+j}(u))$, we have

$$|\bar{y}_{t+j}(u) - \hat{y}_{t+j}(u)| = |f(\bar{x}_{t+j}(u)) - \mu_t^f(\hat{x}_{t+j}(u))| \tag{75}$$

$$\le |f(\bar{x}_{t+j}(u)) - f(\hat{x}_{t+j}(u))| + |f(\hat{x}_{t+j}(u)) - \mu_t^f(\hat{x}_{t+j}(u))|. \tag{76}$$

The first term in (76) is controlled by the Lipschitz continuity of $f$:

$$|f(\bar{x}_{t+j}(u)) - f(\hat{x}_{t+j}(u))| \le L_f\|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| = L_f e_{t,j}(u). \tag{77}$$

The second term is controlled by the transition confidence event $\mathcal{E}_t^f$. On $\mathcal{E}_t^f$, for every $z \in \mathcal{X}$,

$$|f(z) - \mu_t^f(z)| \le \sqrt{\beta_t^f}\sigma_t^f(z). \tag{78}$$

Taking $z = \hat{x}_{t+j}(u)$, we obtain

$$|f(\hat{x}_{t+j}(u)) - \mu_t^f(\hat{x}_{t+j}(u))| \le \sqrt{\beta_t^f}\sigma_t^f(\hat{x}_{t+j}(u)). \tag{79}$$

Substituting the two bounds into (76) yields

$$|\bar{y}_{t+j}(u) - \hat{y}_{t+j}(u)| \le L_f e_{t,j}(u) + \sqrt{\beta_t^f}\sigma_t^f(\hat{x}_{t+j}(u)). \tag{80}$$

Combining (74) and (80), we get

$$e_{t,j+1}(u) \le L_x e_{t,j}(u) + L_y\Big(L_f e_{t,j}(u) + \sqrt{\beta_t^f}\sigma_t^f(\hat{x}_{t+j}(u))\Big) \tag{81}$$

$$= (L_x + L_y L_f)e_{t,j}(u) + L_y\sqrt{\beta_t^f}\sigma_t^f(\hat{x}_{t+j}(u)). \tag{82}$$

This proves the recursive one-step mismatch bound. Now define the computable error envelope by

$$\bar{e}_{t,1}(u) = 0, \tag{83}$$

and, for $j = 1, \dots, m-1$,

$$\bar{e}_{t,j+1}(u) := (L_x + L_y L_f)\bar{e}_{t,j}(u) + L_y\sqrt{\beta_t^f}\sigma_t^f(\hat{x}_{t+j}(u)). \tag{84}$$

We prove by induction that

$$e_{t,j}(u) \le \bar{e}_{t,j}(u), \qquad j = 1, \dots, m. \tag{85}$$

For $j = 1$, the claim holds because

$$e_{t,1}(u) = 0 = \bar{e}_{t,1}(u). \tag{86}$$

Assume that $e_{t,j}(u) \le \bar{e}_{t,j}(u)$ holds for some $j \in \{1, \dots, m-1\}$. Using the recursive mismatch bound derived above,

$$e_{t,j+1}(u) \le (L_x + L_y L_f)e_{t,j}(u) + L_y\sqrt{\beta_t^f}\sigma_t^f(\hat{x}_{t+j}(u)). \tag{87}$$

Since $L_x + L_y L_f \ge 0$, the induction hypothesis implies

$$e_{t,j+1}(u) \le (L_x + L_y L_f)\bar{e}_{t,j}(u) + L_y\sqrt{\beta_t^f}\sigma_t^f(\hat{x}_{t+j}(u)) \tag{88}$$

$$= \bar{e}_{t,j+1}(u). \tag{89}$$

Therefore, by induction,

$$e_{t,j}(u) = \|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| \le \bar{e}_{t,j}(u), \qquad j = 1, \dots, m. \tag{90}$$

This completes the proof. □

**Lemma 5** ($m$-step inside-margin certificate)**.** *Suppose that the rollout mismatch satisfies $e_{t,j}(u) \le \bar{e}_{t,j}(u),\ j = 1, \dots, m$, and that $\ell_t$ is $L_\ell$-Lipschitz. If $\ell_t(\hat{x}_{t+j}(u)) \ge L_\ell \bar{e}_{t,j}(u)$ for $j = 1, \dots, m$, then*

$$\bar{x}_{t+1:t+m}(u) \in \mathcal{S}. \tag{91}$$

*Proof.* Fix any $j \in \{1, \dots, m\}$. By Lemma 4, on the event $\mathcal{E}_t$, we have

$$e_{t,j}(u) = \|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| \le \bar{e}_{t,j}(u). \tag{92}$$

Since $\ell_t$ is $L_\ell$-Lipschitz, it follows that

$$|\ell_t(\bar{x}_{t+j}) - \ell_t(\hat{x}_{t+j})| \le L_\ell\|\bar{x}_{t+j} - \hat{x}_{t+j}\| \le L_\ell \bar{e}_{t,j}(u). \tag{93}$$

Hence,

$$\ell_t(\bar{x}_{t+j}) \ge \ell_t(\hat{x}_{t+j}) - L_\ell \bar{e}_{t,j}(u). \tag{94}$$

By the assumed inside-margin condition,

$$\ell_t(\hat{x}_{t+j}) \ge L_\ell \bar{e}_{t,j}(u), \tag{95}$$

and therefore

$$\ell_t(\bar{x}_{t+j}) \ge 0. \tag{96}$$

Thus, $\bar{x}_{t+j} \in S_t^{\mathrm{LCB}}$. Since $j \in \{1, \dots, m\}$ was arbitrary, we conclude that

$$\bar{x}_{t+j} \in S_t^{\mathrm{LCB}}, \qquad \forall j = 1, \dots, m, \tag{97}$$

which is exactly

$$\bar{x}_{t+1:t+m} \in \big(S_t^{\mathrm{LCB}}\big)^m. \tag{98}$$

Finally, by Lemma 2, $\bar{x}_{t+1:t+m} \in \mathcal{S}$. □

**Lemma 6** (Lipschitz property of the NARX shift operator)**.** *Fix a policy $\pi$, and let $\Phi_\pi(\cdot,\cdot,j)$ denote the regressor update map at rollout step $j$, so that*

$$x_{t+j+1} = \Phi_\pi(x_{t+j}, y_{t+j}, j), \qquad \hat{x}_{t+j+1} = \Phi_\pi(\hat{x}_{t+j}, \hat{y}_{t+j}, j). \tag{99}$$

*Under the NARX construction, $\Phi_\pi$ is Lipschitz in $(x, y)$. In particular, under the Euclidean norm,*

$$\|\Phi_\pi(x, y, j) - \Phi_\pi(x', y', j)\|_2 \le \|x - x'\|_2 + |y - y'|, \qquad \forall x, x' \in \mathcal{X},\ \forall y, y' \in \mathbb{R}. \tag{100}$$

*Therefore the generic Lipschitz recursion condition holds with $L_x = 1$ and $L_y = 1$.*

*Proof.* Under the NARX construction, $\Phi_\pi(x, y, j)$ is obtained by taking the output-history block of $x$, shifting it by one position, inserting the new output $y$, and appending the future input block determined by $\pi$. Hence, for two pairs $(x, y)$ and $(x', y')$, the difference $\Phi_\pi(x, y, j) - \Phi_\pi(x', y', j)$ consists of the shifted output-history differences inherited from $x - x'$, the new scalar difference $y - y'$, and zeros on the appended input block. Therefore,

$$\|\Phi_\pi(x, y, j) - \Phi_\pi(x', y', j)\|_2^2 \le \|x - x'\|_2^2 + |y - y'|^2. \tag{101}$$

Taking square roots gives

$$\|\Phi_\pi(x, y, j) - \Phi_\pi(x', y', j)\|_2 \le \sqrt{\|x - x'\|_2^2 + |y - y'|^2} \le \|x - x'\|_2 + |y - y'|. \tag{102}$$

This proves the claim. □

## F Proof for Theorem 2

For Theorem 2 and its auxiliary lemmas, consider Algorithm 2 with fixed $m \geq 2$, fixed kernels and noise variance $\sigma_\varepsilon^2 > 0$, $U_{t,m}^{\mathrm{feas}} \neq \varnothing$ at each round, and common confidence parameters $\beta_t^f = \beta_t^g = 2\log(|D|\pi^2 t^2/(3\delta))$ satisfying $\mathbb{P}(\bigcap_{t\geq 1}(\mathcal{E}_t^f \cap \mathcal{E}_t^g)) \geq 1-\delta$. This uses the schedule of Lemma 1 with $\pi_t = \pi^2 t^2/6$ and $\delta_f = \delta_g = \delta/2$. Both channels are observed at all queried locations before the next round. Full continuations are defined for all candidates, and all queried, planned, and noiseless true rollout states lie in the fixed finite domain $D$.

If either model's posterior standard deviation exceeds $1/\log(e+t)$ at one or more reference states in round $t$, at least one of those states must be queryable on a certified rollout in that round. For each uncertified $u$ with $G_t^{(m)}(u) > 0$ at round $t$, consider any state on its full planned rollout where the certificate uses model $q \in \{f, g\}$. Either its $q$-feature is a linear combination of the reference features with coefficient vector of Euclidean norm at most $A$, or a certified rollout containing that state was available in at least $\rho n_t$ past rounds using (13), for fixed $A < \infty$ and $\rho \in (0, 1]$. The radius $\epsilon_t(\delta)$ is given by (122) for $n_t \geq 1$ and is $+\infty$ otherwise.

*Proof of Theorem 2.* Let $\mathcal{E}_t := \mathcal{E}_t^f \cap \mathcal{E}_t^g$. By the joint confidence condition, $\bigcap_{t\geq 1} \mathcal{E}_t$ holds with probability at least $1-\delta$, where $\delta = \delta_f + \delta_g$. We prove the theorem on this event. Recall that the certified $m$-step future-safety score is defined by

$$\widehat{G}_t^{(m)}(u') := \min_{1\leq j\leq m} \big(\ell_t(\hat{x}_{t+j}(u')) - L_\ell \bar{e}_{t,j}(u')\big). \tag{103}$$

for each candidate $u' \in U^{\mathrm{cand}}$. Then, by the definition of the feasible set (Definition 9),

$$u' \in U_{t,m}^{\mathrm{feas}} \quad \Longleftrightarrow \quad \widehat{G}_t^{(m)}(u') \geq 0. \tag{104}$$

We first establish a uniform bound between the true and certified future-safety scores. Fix any $u \in U^{\mathrm{cand}}$. Using

$$\min_j a_j - \min_j b_j \leq \max_j (a_j - b_j), \tag{105}$$

we have

$$G_t^{(m)}(u) - \widehat{G}_t^{(m)}(u) = \min_{1\leq j\leq m} g(\bar{x}_{t+j}(u)) - \min_{1\leq j\leq m} \big(\ell_t(\hat{x}_{t+j}(u)) - L_\ell \bar{e}_{t,j}(u)\big) \tag{106}$$

$$\leq \max_{1\leq j\leq m} \big(g(\bar{x}_{t+j}(u)) - \ell_t(\hat{x}_{t+j}(u)) + L_\ell \bar{e}_{t,j}(u)\big). \tag{107}$$

For each $j = 1, \ldots, m$, decompose

$$g(\bar{x}_{t+j}(u)) - \ell_t(\hat{x}_{t+j}(u)) = g(\bar{x}_{t+j}(u)) - g(\hat{x}_{t+j}(u)) + g(\hat{x}_{t+j}(u)) - \ell_t(\hat{x}_{t+j}(u)). \tag{108}$$

On the transition confidence event $\mathcal{E}_t^f$, by Lemma 4, the recursive error envelope satisfies

$$\|\bar{x}_{t+j}(u) - \hat{x}_{t+j}(u)\| \leq \bar{e}_{t,j}(u), \qquad j = 1, \ldots, m. \tag{109}$$

Therefore, by the $L_g$-Lipschitz continuity of $g$,

$$g(\bar{x}_{t+j}(u)) - g(\hat{x}_{t+j}(u)) \leq L_g \bar{e}_{t,j}(u). \tag{110}$$

Moreover, on the safety confidence event $\mathcal{E}_t^g$,

$$g(\hat{x}_{t+j}(u)) \leq \mu_t^g(\hat{x}_{t+j}(u)) + \sqrt{\beta_t^g}\sigma_t^g(\hat{x}_{t+j}(u)). \tag{111}$$

Since

$$\ell_t(\hat{x}_{t+j}(u)) = \mu_t^g(\hat{x}_{t+j}(u)) - \sqrt{\beta_t^g}\sigma_t^g(\hat{x}_{t+j}(u)), \tag{112}$$

we get

$$g(\hat{x}_{t+j}(u)) - \ell_t(\hat{x}_{t+j}(u)) \leq 2\sqrt{\beta_t^g}\sigma_t^g(\hat{x}_{t+j}(u)). \tag{113}$$

Combining (107), (108), (110), and (113) yields

$$G_t^{(m)}(u) - \widehat{G}_t^{(m)}(u) \le 2\sqrt{\beta_t^g} \max_{1\le j\le m} \sigma_t^g(\hat{x}_{t+j}(u)) + (L_g + L_\ell) \max_{1\le j\le m} \bar{e}_{t,j}(u). \tag{114}$$

It remains to bound the error envelope. By the recursion defining $\bar{e}_{t,j}(u)$,

$$\bar{e}_{t,1}(u) = 0, \qquad \bar{e}_{t,j+1}(u) = (L_x + L_y L_f)\bar{e}_{t,j}(u) + L_y\sqrt{\beta_t^f}\,\sigma_t^f(\hat{x}_{t+j}(u)). \tag{115}$$

Unrolling this recursion yields, for every $j \ge 2$,

$$\bar{e}_{t,j}(u) = L_y \sum_{s=1}^{j-1} (L_x + L_y L_f)^{j-1-s} \sqrt{\beta_t^f}\,\sigma_t^f(\hat{x}_{t+s}(u)). \tag{116}$$

Hence, uniformly over $u \in U^{\mathrm{cand}}$,

$$\begin{aligned} G_t^{(m)}(u) - \widehat{G}_t^{(m)}(u) &\le 2\sqrt{\beta_t^g} \max_{1\le j\le m} \sigma_t^g(\hat{x}_{t+j}(u)) \\ &\quad + (L_g + L_\ell) \max_{1\le j\le m} L_y \sum_{s=1}^{j-1} (L_x + L_y L_f)^{j-1-s} \sqrt{\beta_t^f}\,\sigma_t^f(\hat{x}_{t+s}(u)). \end{aligned} \tag{117}$$

Fix $t$ with $n_t \ge 1$ and an uncertified $u \in U_{t,\epsilon}^{\mathrm{int}}$. For $q \in \{f, g\}$, let $x_q$ be a state attaining the corresponding maximum in (119). Choose $\mathcal{I}_q$ to contain at least $\rho n_t$ past rounds using (13), with $v_s^q \in U_{s,m}^{\mathrm{feas}}$ for each $s \in \mathcal{I}_q$. If the $q$-feature at $x_q$ has the bounded reference representation, Lemma 13 gives $\sigma_t^q(x_q) \le r_t/2$ and the choices are arbitrary. Otherwise, the observation condition in Theorem 2 permits choices with $x_q \in X_s^{\mathrm{plan}}(v_s^q)$ for every $s \in \mathcal{I}_q$. Then, from (117), Lemmas 8, 11, 12, and the selection rule (13), we have

$$\begin{aligned} G_t^{(m)}(u) - \widehat{G}_t^{(m)}(u) &\le 2\sqrt{\beta_t^g} \max_{1\le j\le m} \sigma_t^g(\hat{x}_{t+j}(u)) \\ &\quad + (L_g + L_\ell) \max_{1\le j\le m} L_y \sum_{s=1}^{j-1} (L_x + L_y L_f)^{j-1-s} \sqrt{\beta_t^f}\,\sigma_t^f(\hat{x}_{t+s}(u)) \qquad (118) \\ &\le 2\sqrt{\beta_t^g} \max_{1\le j\le m} \sigma_t^g(\hat{x}_{t+j}(u)) \\ &\quad + (L_g + L_\ell) L_y \sqrt{\beta_t^f} \max_{1\le s<m} \sigma_t^f(\hat{x}_{t+s}(u)) \max_{1\le j\le m} \sum_{s=1}^{j-1} (L_x + L_y L_f)^{j-1-s} \qquad (119) \\ &\le 2\sqrt{\beta_t^g}\Big(\frac{r_t}{2} + \frac{1}{|\mathcal{I}_g|} \sum_{s\in\mathcal{I}_g} \max_{1\le j\le m} \sigma_s^g(\hat{x}_{s+j}(v_s^g))\Big) \\ &\quad + (L_g + L_\ell) L_y \sqrt{\beta_t^f}\Big(\frac{r_t}{2} + \frac{1}{|\mathcal{I}_f|} \sum_{k\in\mathcal{I}_f} \max_{1\le s\le m} \sigma_k^f(\hat{x}_{k+s}(v_k^f))\Big) \sum_{r=0}^{m-2} (L_x + L_y L_f)^r \qquad (120) \\ &\le 2\sqrt{\beta_t^g}\Big(\frac{r_t}{2} + \frac{1}{|\mathcal{I}_g|} \sum_{s\in\mathcal{I}_g} \sqrt{m C_\eta I_s^g(v_s^g)}\Big) \\ &\quad + (L_g + L_\ell) L_y \sqrt{\beta_t^f}\Big(\frac{r_t}{2} + \frac{1}{|\mathcal{I}_f|} \sum_{k\in\mathcal{I}_f} \sqrt{m C_\eta I_k^f(v_k^f)}\Big) \sum_{r=0}^{m-2} (L_x + L_y L_f)^r \qquad (121) \\ &\le \Big\{2\sqrt{\beta_t^g} + (L_g + L_\ell) L_y \sqrt{\beta_t^f} \sum_{r=0}^{m-2} (L_x + L_y L_f)^r\Big\}\Big(\frac{r_t}{2} + 2\sqrt{\frac{m C_\eta \gamma_t}{\rho n_t}}\Big) = \epsilon_t(\delta). \qquad (122) \end{aligned}$$

For (120), average over $\mathcal{I}_q$ using the reference bound or $\sigma_t^q(x_q) \le \sigma_s^q(x_q)$ on the chosen paths, as in the proof of Lemma 9. Lemma 11 on each comparison path gives (121). For (122), add the other model's nonnegative acquisition term to each weighted average in (121) and apply Lemma 12 on $\mathcal{I}_g$ and $\mathcal{I}_f$ separately. Each set contains at least $\rho n_t$ rounds, giving the factor 2.

We now prove the two inclusions. Fix $t \geq t^*(\epsilon, \delta)$. First, let $u \in U_{t,m}^{\text{feas}}$. Then $\widehat{G}_t^{(m)}(u) \geq 0$, which implies that

$$\ell_t(\hat{x}_{t+j}(u)) \geq L_\ell \bar{e}_{t,j}(u), \qquad j = 1, \ldots, m. \tag{123}$$

Thus the margin condition holds along the full predicted rollout. By the inside-margin certificate, the corresponding noiseless true rollout satisfies

$$\bar{x}_{t+1:t+m}(u) \in \mathcal{S}^m. \tag{124}$$

Equivalently,

$$g(\bar{x}_{t+j}(u)) \geq 0, \qquad j = 1, \ldots, m. \tag{125}$$

Therefore,

$$G_t^{(m)}(u) = \min_{1 \leq j \leq m} g(\bar{x}_{t+j}(u)) \geq 0, \tag{126}$$

so $u \in U_t^{\text{true}}$. Since $u \in U_{t,m}^{\text{feas}}$ was arbitrary, $U_{t,m}^{\text{feas}} \subseteq U_t^{\text{true}}$. Next, let $u \in U_{t,\epsilon}^{\text{int}}$. By definition, $G_t^{(m)}(u) \geq \epsilon$. Since $t \geq t^*(\epsilon, \delta)$, we have $\epsilon_t(\delta) \leq \epsilon$. If $u$ is already certified, the inclusion holds for $u$. Otherwise, using the uniform certification-gap bound (122),

$$G_t^{(m)}(u) - \widehat{G}_t^{(m)}(u) \leq \epsilon_t(\delta). \tag{127}$$

Therefore,

$$\widehat{G}_t^{(m)}(u) \geq G_t^{(m)}(u) - \epsilon_t(\delta) \geq \epsilon - \epsilon_t(\delta) \geq 0. \tag{128}$$

By the definition of the certified feasible set, this implies $u \in U_{t,m}^{\text{feas}}$. Since $u \in U_{t,\epsilon}^{\text{int}}$ was arbitrary, we obtain $U_{t,\epsilon}^{\text{int}} \subseteq U_{t,m}^{\text{feas}}$. Combining the two inclusions gives

$$U_{t,\epsilon}^{\text{int}} \subseteq U_{t,m}^{\text{feas}} \subseteq U_t^{\text{true}}. \tag{129}$$

It remains to prove the boundary-layer statement. Let $u \in U_t^{\text{true}} \setminus U_{t,m}^{\text{feas}}$. Since $u \in U_t^{\text{true}}$, we have

$$G_t^{(m)}(u) \geq 0. \tag{130}$$

If $G_t^{(m)}(u) \geq \epsilon$, then $u \in U_{t,\epsilon}^{\text{int}}$. However, we have already shown that

$$U_{t,\epsilon}^{\text{int}} \subseteq U_{t,m}^{\text{feas}}, \tag{131}$$

which contradicts $u \notin U_{t,m}^{\text{feas}}$. Therefore, $G_t^{(m)}(u) < \epsilon$. Combining this with $G_t^{(m)}(u) \geq 0$, we get

$$0 \leq G_t^{(m)}(u) < \epsilon. \tag{132}$$

Thus,

$$U_t^{\text{true}} \setminus U_{t,m}^{\text{feas}} \subseteq \left\{ u \in U^{\text{cand}} : 0 \leq G_t^{(m)}(u) < \epsilon \right\}. \tag{133}$$

On the fixed finite domain $D$, $\gamma_t = O(\log t)$ since each Gram matrix has rank at most $|D|$ and trace at most $mt$. With $\beta_t^f = \beta_t^g = O(\log t)$ and fixed $\rho > 0$, Lemma 13 gives $\epsilon_t(\delta) = O((\log t)^{-1/2})$, so $t^*(\epsilon, \delta) < \infty$. Since all arguments hold on $\bigcap_{t \geq 1} \mathcal{E}_t$, and this event holds with probability at least $1 - \delta$, the theorem follows. □

### F.1 Auxiliary lemmas for Theorem 2

**Rollout information-gain acquisition.** For each round $t \geq 1$ and candidate $u \in U_{t,m}^{\text{feas}}$, let

$$X_t^{\text{plan}}(u) = \{z_{t,1}(u), \ldots, z_{t,\hat{m}_t(u)}(u)\} \tag{134}$$

denote the set of GP query locations induced by the predicted rollout under $u$, where $\hat{m}_t(u) \leq m$. For the transition GP, define the rollout information gain by

$$I_t^f(u) := I(f(X_t^{\text{plan}}(u)); y(X_t^{\text{plan}}(u)) \mid h_t). \tag{135}$$

Equivalently, by the entropy identity for mutual information,

$$I_t^f(u) = H(y(X_t^{\text{plan}}(u)) \mid h_t) - H(y(X_t^{\text{plan}}(u)) \mid f(X_t^{\text{plan}}(u)), h_t). \tag{136}$$

Thus, $I_t^f(u)$ measures the information in hypothetical physical outputs about the transition function along the planned rollout. The corresponding safety information gain is

$$I_t^g(u) := I(g(X_t^{\text{plan}}(u)); y^g(X_t^{\text{plan}}(u)) \mid h_t). \tag{137}$$

It conditions the safety GP on hypothetical safety measurements along the planned rollout. Let $K_t(u)$ denote the transition posterior covariance matrix on $X_t^{\text{plan}}(u)$ conditioned on $h_t$. Since $K_t(u)$ is positive semidefinite, let

$$\lambda_1(u), \dots, \lambda_{\hat{m}_t(u)}(u) \tag{138}$$

denote its eigenvalues. Then

$$I_t^f(u) = \frac{1}{2} \log\det\Big(I + \sigma_\varepsilon^{-2} K_t(u)\Big) = \frac{1}{2} \sum_{i=1}^{\hat{m}_t(u)} \log\left(1 + \frac{\lambda_i(u)}{\sigma_\varepsilon^2}\right). \tag{139}$$

Equivalently, by the chain rule for mutual information, this admits the sequential form

$$I_t^f(u) = \frac{1}{2} \sum_{s=1}^{\hat{m}_t(u)} \log\left(1 + \frac{\tilde{\sigma}_{t,s-1}^2(u)}{\sigma_\varepsilon^2}\right), \tag{140}$$

where $\tilde{\sigma}_{t,s-1}^2(u)$ denotes the posterior variance at $z_{t,s}(u)$ conditioned on $h_t$ and the fantasized observations at the preceding rollout points $z_{t,1}(u), \dots, z_{t,s-1}(u)$.

**Cumulative rollout information gain.** Suppose that the algorithm performs $N$ trajectory-query rounds, and that at each round $n$ it evaluates the certified rollout

$$X_n^{\text{plan}}(u_{n+1}) = \{z_{n,1}(u_{n+1}), \dots, z_{n,m}(u_{n+1})\}. \tag{141}$$

Let

$$Z_{1:mN} := \big(z_{1,1}(u_2), \dots, z_{1,m}(u_2), \dots, z_{N,1}(u_{N+1}), \dots, z_{N,m}(u_{N+1})\big) \tag{142}$$

denote the ordered collection of all rollout query locations collected over the $N$ rounds. Here $h_n$ is the history before querying the round-$n$ rollout, as in (6), and each information gain is the Gaussian surrogate design score evaluated at the realized history. With fixed kernels and noise variance, the block determinant identity gives

$$\sum_{n=1}^{N} I_n^f(u_{n+1}) = \frac{1}{2} \log\det\big(I + \sigma_\varepsilon^{-2} K_1^f(Z_{1:mN})\big) \le \gamma_N. \tag{143}$$

The matrix $K_1^f(Z_{1:mN})$ is the covariance on these locations conditioned on $h_1$. Thus, the sum of the per-round rollout information gains is exactly the information gain of the entire ordered set of $mN$ rollout query locations. A proof of this statement is provided below.

**Lemma 7** (Chain decomposition of rollout information gain)**.** *Under the setting of Theorem 2, the cumulative information-gain identity and bound* (143) *hold.*

*Proof.* Fix any realized sequence of query locations, and write $Z_{\le n}$ for the first $n$ rollout blocks. The Schur complement of $K_1^f(Z_{\le n-1}) + \sigma_\varepsilon^2 I$ in $K_1^f(Z_{\le n}) + \sigma_\varepsilon^2 I$ is the posterior covariance of the next observation block given $h_n$. Thus the block determinant identity gives

$$I_n^f(u_{n+1}) = \frac{1}{2} \log \frac{\det(I + \sigma_\varepsilon^{-2} K_1^f(Z_{\le n}))}{\det(I + \sigma_\varepsilon^{-2} K_1^f(Z_{\le n-1}))}. \tag{144}$$

Summing (144) over $n = 1, \dots, N$ yields the equality in (143). Since the initial posterior covariance is bounded by the prior covariance, the final log determinant is at most $\gamma_N$. The argument holds for each realized location sequence, including adaptively selected rollouts. This proves the claim. □

The same determinant argument for the safety observations gives

$$\sum_{s=1}^{N} I_s^g(u_{s+1}) = \frac{1}{2} \log\det\big(I + \sigma_\varepsilon^{-2} K_1^g(Z_{1:mN})\big) \le \gamma_N. \tag{145}$$

**Uniform rollout uncertainty radii.** The following results are stated for the transition model $f$ for notational simplicity, but the same arguments and conclusions hold analogously for the safety function $g$.

**Lemma 8** (Monotonicity of the posterior standard deviation)**.** *Assume Gaussian process regression with kernel $k$ and observation noise variance $\lambda \geq 0$. Then, for every $x \in \mathcal{X}$, the posterior standard deviation is nonincreasing in the number of observations. In particular, for any $1 \leq n \leq N$,*

$$\sigma_N^f(x) \leq \sigma_{n-1}^f(x). \tag{146}$$

*Proof.* For each $t \geq 0$, let $D_t = \{(x_1, y_1), \ldots, (x_t, y_t)\}$ denote the data available after $t$ observations. The GP posterior variance at $x$ given $D_t$ is

$$\big(\sigma_t^f(x)\big)^2 = k(x,x) - k_t(x)^\top (K_t + \lambda I)^{-1} k_t(x), \tag{147}$$

where $k_t(x) := \big(k(x_1,x), \ldots, k(x_t,x)\big)^\top$ and $K_t := [k(x_i,x_j)]_{i,j=1}^t$. It suffices to show that for every $t \geq 1$,

$$\sigma_t^f(x) \leq \sigma_{t-1}^f(x). \tag{148}$$

Indeed, once this is established, repeated application gives

$$\sigma_N^f(x) \leq \sigma_{N-1}^f(x) \leq \cdots \leq \sigma_{n-1}^f(x), \tag{149}$$

which proves the claim. Now fix $t \geq 1$. Write

$$K_t + \lambda I = \begin{pmatrix} K_{t-1} + \lambda I & k_{t-1}(x_t) \\ k_{t-1}(x_t)^\top & k(x_t,x_t) + \lambda \end{pmatrix}, \qquad k_t(x) = \begin{pmatrix} k_{t-1}(x) \\ k(x_t,x) \end{pmatrix}. \tag{150}$$

By Sherman–Morrison-Woodbury matrix identity, the posterior variance update satisfies

$$\big(\sigma_t^f(x)\big)^2 = \big(\sigma_{t-1}^f(x)\big)^2 - \frac{\big(k_{t-1}^f(x_t,x)\big)^2}{\big(\sigma_{t-1}^f(x_t)\big)^2 + \lambda}, \tag{151}$$

where $k_{t-1}^f(x_t,x) = k(x_t,x) - k_{t-1}(x_t)^\top (K_{t-1} + \lambda I)^{-1} k_{t-1}(x)$ is the posterior covariance between $f(x_t)$ and $f(x)$ conditioned on $D_{t-1}$. Since

$$\frac{\big(k_{t-1}^f(x_t,x)\big)^2}{\big(\sigma_{t-1}^f(x_t)\big)^2 + \lambda} \geq 0, \tag{152}$$

it follows that

$$\big(\sigma_t^f(x)\big)^2 \leq \big(\sigma_{t-1}^f(x)\big)^2. \tag{153}$$

Because posterior standard deviations are nonnegative, taking square roots preserves the inequality and yields

$$\sigma_t^f(x) \leq \sigma_{t-1}^f(x). \tag{154}$$

Therefore, for any $1 \leq n \leq N$,

$$\sigma_N^f(x) \leq \sigma_{n-1}^f(x). \tag{155}$$

This completes the proof. □

**Lemma 9** (Averaging by posterior monotonicity)**.** *For every $N \geq 1$ and $u \in U^{\mathrm{cand}}$,*

$$\max_{1 \leq j \leq m} \sigma_N^f\big(\hat{x}_{N+j}(u)\big) \leq \frac{1}{N} \sum_{n=1}^{N} \max_{1 \leq j \leq m} \sigma_{n-1}^f\big(\hat{x}_{N+j}(u)\big). \tag{156}$$

*Proof.* Fix $u \in U^{\mathrm{cand}}$ and $1 \leq j \leq m$. By Lemma 8, we have

$$\sigma_N^f\big(\hat{x}_{N+j}(u)\big) \leq \frac{1}{N} \sum_{n=1}^{N} \sigma_{n-1}^f\big(\hat{x}_{N+j}(u)\big). \tag{157}$$

Taking the maximum over $j$ yields

$$\max_{1 \leq j \leq m} \sigma_N^f\big(\hat{x}_{N+j}(u)\big) \leq \frac{1}{N} \sum_{n=1}^{N} \max_{1 \leq j \leq m} \sigma_{n-1}^f\big(\hat{x}_{N+j}(u)\big) \tag{158}$$

Finally, taking the supremum over $u \in U^{\mathrm{cand}}$ gives the claim. □

**Lemma 10** (Hallucinated variance sum controlled by rollout information gain)**.** *Assume that $k(x,x) \le 1$ for all $x \in \mathcal{X}$. For any round $t$ and any $u \in U_{t,m}^{\mathrm{feas}}$, let $\tilde{\sigma}_{t,s-1}^{f}\big(z_{t,s}(u)\big)$, $s = 1,\dots,\hat{m}_t(u)$, denote the posterior standard deviation along the hallucinated rollout, and define $I_t^f(u) := \frac{1}{2}\sum_{s=1}^{\hat{m}_t(u)} \log\big(1 + \frac{(\tilde{\sigma}_{t,s-1}^{f}(z_{t,s}(u)))^2}{\sigma_\varepsilon^2}\big)$. Then, for every round $t$ and every $u \in U_{t,\hat{m}_t(u)}^{\mathrm{feas}}$,*

$$\sum_{s=1}^{\hat{m}_t(u)} \big(\tilde{\sigma}_{t,s-1}^{f}(z_{t,s}(u))\big)^2 \le 2(\sigma_\varepsilon^2+1)\, I_t^f(u). \tag{159}$$

*The same conclusion holds for either $q \in \{f,g\}$ and any fixed ordered path $Z = (z_1,\dots,z_M)$ when $k_q(x,x) \le 1$:*

$$\sum_{s=1}^{M} (\tilde{\sigma}_{t,s-1}^{q}(z_s))^2 \le 2(\sigma_\varepsilon^2+1) I(q(Z); y^q(Z) \mid h_{t-1}). \tag{160}$$

*Here $\tilde{\sigma}_{t,s-1}^{q}(z_s)$ conditions on $h_{t-1}$ and hallucinated observations at $z_1,\dots,z_{s-1}$, with $y^f(Z) = y(Z)$.*

*Proof.* Fix a round $t$ and $u \in U_{t,m}^{\mathrm{feas}}$. For simplicity, write

$$\hat{m} := \hat{m}_t(u), \qquad \tilde{\sigma}_{s-1}^2 := \Big(\tilde{\sigma}_{t,s-1}^{f}\big(z_{t,s}(u)\big)\Big)^2, \qquad s = 1,\dots,\hat{m}. \tag{161}$$

Since posterior variance is always bounded above by prior variance, and $k(x,x) \le 1$ for all $x \in \mathcal{X}$, we have

$$0 \le \tilde{\sigma}_{s-1}^2 \le 1, \qquad s = 1,\dots,\hat{m}. \tag{162}$$

Therefore,

$$0 \le \frac{\tilde{\sigma}_{s-1}^2}{\sigma_\varepsilon^2} \le \frac{1}{\sigma_\varepsilon^2}. \tag{163}$$

Applying $\log(1+x) \ge \frac{x}{1+1/\sigma_\varepsilon^2}$ for $0 \le x \le 1/\sigma_\varepsilon^2$ at $x = \tilde{\sigma}_{s-1}^2/\sigma_\varepsilon^2$ gives

$$\log\left(1 + \frac{\tilde{\sigma}_{s-1}^2}{\sigma_\varepsilon^2}\right) \ge \frac{\tilde{\sigma}_{s-1}^2/\sigma_\varepsilon^2}{1+1/\sigma_\varepsilon^2} = \frac{\tilde{\sigma}_{s-1}^2}{\sigma_\varepsilon^2+1}. \tag{164}$$

Summing over $s = 1,\dots,\hat{m}$, we obtain

$$I_t^f(u) = \frac{1}{2}\sum_{s=1}^{\hat{m}} \log\left(1 + \frac{\tilde{\sigma}_{s-1}^2}{\sigma_\varepsilon^2}\right) \ge \frac{1}{2}\sum_{s=1}^{\hat{m}} \frac{\tilde{\sigma}_{s-1}^2}{\sigma_\varepsilon^2+1} = \frac{1}{2(\sigma_\varepsilon^2+1)}\sum_{s=1}^{\hat{m}} \tilde{\sigma}_{s-1}^2. \tag{165}$$

Hence

$$\sum_{s=1}^{\hat{m}} \tilde{\sigma}_{s-1}^2 \le 2(\sigma_\varepsilon^2+1)\, I_t^f(u), \tag{166}$$

that is,

$$\sum_{s=1}^{\hat{m}_t(u)} \Big(\tilde{\sigma}_{t,s-1}^{f}\big(z_{t,s}(u)\big)\Big)^2 \le 2(\sigma_\varepsilon^2+1)\, I_t^f(u). \tag{167}$$

This proves the claim. For the fixed path and either GP, the same variance bound and logarithmic inequality apply at each $z_s$. The sequential Gaussian information-gain identity gives the stated bound for any ordering of the fixed path. □

**Lemma 11** (Information gain controls path uncertainty)**.** *Suppose $k_f(x,x), k_g(x,x) \le 1$ for all $x \in \mathcal{X}$, and $\sigma_\varepsilon^2 > 0$. For an ordered path $Z = (z_1,\dots,z_M)$ with $1 \le M \le m$,*

$$\max_{1\le s\le M} \sigma_{t-1}^{f}(z_s) \le \sqrt{m C_\eta I(f(Z); y(Z) \mid h_{t-1})}, \tag{168}$$

$$\max_{1\le s\le M} \sigma_{t-1}^{g}(z_s) \le \sqrt{m C_\eta I(g(Z); y^g(Z) \mid h_{t-1})}. \tag{169}$$

*Here, $C_\eta = 2(\sigma_\varepsilon^2+1)$.*

*Proof.* Fix a point $z_s$ on the path. Since information gain is unchanged by permuting the fixed path, relabel this point as $z_1$. The first hallucinated variance is the original posterior variance, $\tilde{\sigma}^f_{t,0}(z_1) = \sigma^f_{t-1}(z_1)$. Applying Lemma 10 to this order gives

$$\begin{aligned}(\sigma^f_{t-1}(z_1))^2 &= (\tilde{\sigma}^f_{t,0}(z_1))^2 \\ &\le \sum_{j=1}^{M} (\tilde{\sigma}^f_{t,j-1}(z_j))^2 \\ &\le C_\eta I(f(Z); y(Z) \mid h_{t-1}).\end{aligned} \tag{170}$$

Applying the same argument to $g$ and its safety observations $y^g$ gives

$$(\sigma^g_{t-1}(z_1))^2 \le C_\eta I(g(Z); y^g(Z) \mid h_{t-1}). \tag{171}$$

Taking the maximum over $s$, taking square roots, and using $m \ge 1$ proves the claim. □

**Lemma 12** (Average information gain bound)**.** *Under the setting of Theorem 2, let $\mathcal{I}$ be a nonempty set of rounds before $t$ using* (13) *and let $v_s \in U^{\mathrm{feas}}_{s,m}$ for $s \in \mathcal{I}$. Then*

$$\begin{aligned}&\frac{1}{|\mathcal{I}|}\sum_{s\in\mathcal{I}}\Big[2\sqrt{\beta^g_t}\sqrt{mC_\eta I^g_s(v_s)} + (L_g + L_\ell)L_y\sqrt{\beta^f_t}\sqrt{mC_\eta I^f_s(v_s)}\sum_{r=0}^{m-2}(L_x + L_yL_f)^r\Big] \\ &\qquad \le \Big[2\sqrt{\beta^g_t} + (L_g + L_\ell)L_y\sqrt{\beta^f_t}\sum_{r=0}^{m-2}(L_x + L_yL_f)^r\Big]\sqrt{\frac{mC_\eta\gamma_t}{|\mathcal{I}|}}.\end{aligned} \tag{172}$$

*Here, $\gamma_t$ upper-bounds the maximal information gain of each GP over at most $mt$ queried states.*

*Proof.* For $s \in \mathcal{I}$, rule (13) gives

$$\begin{aligned}&2\sqrt{\beta^g_t}\sqrt{mC_\eta I^g_s(v_s)} + (L_g + L_\ell)L_y\sqrt{\beta^f_t}\sqrt{mC_\eta I^f_s(v_s)}\sum_{r=0}^{m-2}(L_x + L_yL_f)^r \\ &\qquad \le 2\sqrt{\beta^g_t}\sqrt{mC_\eta I^g_s(u_{s+1})} + (L_g + L_\ell)L_y\sqrt{\beta^f_t}\sqrt{mC_\eta I^f_s(u_{s+1})}\sum_{r=0}^{m-2}(L_x + L_yL_f)^r.\end{aligned} \tag{173}$$

Indeed, $v_s \in U^{\mathrm{feas}}_{s,m}$, and multiplying the round-$s$ objective by $\sqrt{\beta^f_t/\beta^f_s}$ gives the weights in (173), since $\beta^f_r = \beta^g_r$ at every round $r$. For each $q \in \{f, g\}$, Cauchy-Schwarz inequality gives

$$\frac{1}{|\mathcal{I}|}\sum_{s\in\mathcal{I}}\sqrt{mC_\eta I^q_s(u_{s+1})} \le \sqrt{\frac{mC_\eta}{|\mathcal{I}|}\sum_{s\in\mathcal{I}} I^q_s(u_{s+1})} \le \sqrt{\frac{mC_\eta\gamma_t}{|\mathcal{I}|}}. \tag{174}$$

The second inequality uses Lemma 7, its safety-observation analogue, and the fact that the nonnegative information sum over $\mathcal{I}$ is at most the sum over all past rounds. Average (173) over $\mathcal{I}$ and apply (174). This proves the claim. □

**Lemma 13** (Reference-query variance bound)**.** *Under the reference query rule of Algorithm 2 and the reference availability in Theorem 2, we have $n_t = \Theta(t)$ and $r_t = O(1/\log t)$. For $q \in \{f, g\}$ and any $x \in D$ whose $q$-feature is a linear combination of the reference features with coefficient vector of Euclidean norm at most $A$,*

$$\sigma^q_t(x) \le r_t/2. \tag{175}$$

*Proof.* Fix the realized query locations. Let $w_1, \dots, w_N$ be all queried locations before round $t$, including the initial data. For $q \in \{f, g\}$, let $\phi_q$ be a feature map of $k_q$. For any state $x$ satisfying the representation condition, choose $a^q(x) \in \mathbb{R}^K$ such that

$$\phi_q(x) = \sum_{i=1}^{K} a^q_i(x)\phi_q(z_i), \qquad \|a^q(x)\|_2 \le A. \tag{176}$$

For any $x \in D$, the GP posterior variance has the variational form

$$(\sigma_t^q(x))^2 = \min_{c \in \mathbb{R}^N} \big\{ \|\phi_q(x) - \sum_{l=1}^{N} c_l \phi_q(w_l)\|^2 + \sigma_\varepsilon^2 \|c\|_2^2 \big\}. \tag{177}$$

Let $N_i$ be the number of observations at $z_i$ before a reference round $s$. For $N_i \geq 1$, apply the variational formula at round $s$ with weight $1/N_i$ on each of these observations and zero on all others. Their weighted features sum to $N_i(1/N_i)\phi_q(z_i) = \phi_q(z_i)$, so the feature error is zero and

$$(\sigma_s^q(z_i))^2 \leq \sigma_\varepsilon^2 N_i (1/N_i)^2 = \frac{\sigma_\varepsilon^2}{N_i}, \qquad q \in \{f, g\}. \tag{178}$$

For counting purposes, assign each reference round $s$ to one state $z_i$ on its queried rollout with $\sigma_s^q(z_i) > 1/\log(e+s)$ for some $q$. Combining this selection condition with the variance bound gives, for $N_i \geq 1$,

$$\frac{1}{[\log(e+s)]^2} < (\sigma_s^q(z_i))^2 \leq \frac{\sigma_\varepsilon^2}{N_i}. \tag{179}$$

Multiplying by $N_i[\log(e+s)]^2$ yields $N_i < \sigma_\varepsilon^2[\log(e+s)]^2$. The same bound holds when $N_i = 0$. Each round assigned to $i$ adds an observation at $z_i$. Since $\log(e+s)$ increases with $s$, at most $\lceil \sigma_\varepsilon^2[\log(e+t)]^2 \rceil$ rounds before $t$ can be assigned to each $i$. Thus

$$t - 1 - n_t \leq K \lceil \sigma_\varepsilon^2[\log(e+t)]^2 \rceil = O((\log t)^2), \qquad n_t = \Theta(t). \tag{180}$$

For all sufficiently large $t$, at least one round $s \in \{\lceil t/2 \rceil, \dots, t-1\}$ uses (13). The threshold test at that round and posterior-variance monotonicity give

$$\max_{q \in \{f,g\},\, 1 \leq i \leq K} \sigma_t^q(z_i) \leq \frac{1}{\log(e+s)} \leq \frac{1}{\log(e+t/2)}. \tag{181}$$

The feature representation and the posterior $L^2$ triangle inequality yield

$$\sigma_t^q(x) \leq \sum_{i=1}^{K} |a_i^q(x)| \sigma_t^q(z_i) \leq A\sqrt{K} \max_{1 \leq i \leq K} \sigma_t^q(z_i) \leq r_t/2. \tag{182}$$

The definition of $r_t$ then gives $r_t = O(1/\log t)$, uniformly over the observation history. □

## G Proof of Corollary 1

*Proof.* We prove on the high-probability event of Theorem 2. Fix any $t \geq t^*(\epsilon, \delta)$. By Theorem 2,

$$U_{t,m}^{\text{feas}} \subseteq U_t^{\text{true}}. \tag{183}$$

Let $u \in U^{\text{cand}}$ be an $\epsilon$-separated $m$-step dead-end input. By definition 1,

$$G_t^{(m)}(u) \leq -\epsilon < 0. \tag{184}$$

Therefore $u \notin U_t^{\text{true}}$. Since $U_{t,m}^{\text{feas}} \subseteq U_t^{\text{true}}$, it follows that

$$u \notin U_{t,m}^{\text{feas}}. \tag{185}$$

Finally, since Algorithm 2 selects only from $U_{t,m}^{\text{feas}}$, it cannot select such an input. □

## H Experimental Setup and Reproducibility

**GP models and confidence schedule.** The dynamics and safety models are exact dense GPs with isotropic squared-exponential kernels, unit signal variance, and length scale 1. Kernel and observation-noise hyperparameters remain fixed in the main experiments; the adaptation experiments in Appendix J are reported separately. Table 6 lists the observation-noise standard deviations and confidence levels. For $q \in \{f, g\}$, we use $\beta_t^q = 2\log(|D|\pi^2 t^2/(6\delta_q))$, where $D$ is the finite evaluation domain specified below. The Cholesky factors and posterior coefficients are cached and reused across candidate evaluations within each round.

Table 6: GP noise standard deviations and confidence levels in the main experiments.

| Benchmark | Dynamics noise SD | Safety noise SD | $\delta_f = \delta_g$ |
|---|---|---|---|
| Synthetic base | 0.002 | 0.001 | 0.05 |
| Synthetic penalty-stress | 0.018 | 0.008 | 0.05 |
| Rail pressure | 1.0 | 0.01 | 0.05 |
| NGSIM | 0.15 | 0.02 | 0.1 |

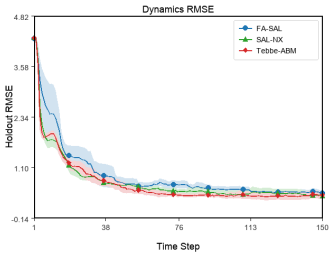
**Candidate construction and finite grids.** For the synthetic benchmark, $U^{\mathrm{cand}}$ contains 11 equally spaced inputs from $-1$ to 1, with spacing 0.2. The state grid contains 31 equally spaced values from $-2$ to 6 in each of the two output coordinates. Its Cartesian product with $U^{\mathrm{cand}}$ defines $D$, with $|D| = 31^2 \cdot 11$. For rail pressure, the endpoint grid is $\{2200, 2230, \ldots, 2800\} \times \{24, 24.6, \ldots, 36\}$, giving 441 candidates. Each endpoint defines a linearly interpolated ramp from the current operating point over $m = 5$ steps. For confidence scheduling, the rail evaluation set is indexed by the 441 endpoints and their five ramp steps, giving $|D| = 2205$. For NGSIM, we enumerate all 625 four-step sequences from the five acceleration choices $\{-4, -2, 0, 1, 2\}\,\mathrm{m/s^2}$. The evaluation set is indexed by these sequences and their four rollout steps, giving $|D| = 2500$. Rail-pressure and NGSIM GP inputs are normalized.

**Lipschitz settings.** The shift constants are $L_x = L_y = 1$ under the Euclidean norm and a common continuation input. For the synthetic dynamics, $L_f$ is the median of the axis-wise finite-difference slopes on the $31 \times 31$ state grid and is then held fixed. We use $L_f = 1$ for rail pressure and NGSIM. At each round, the implementation recomputes $L_\ell$ from coordinate-wise finite differences of the current safety LCB. The local slope estimate is multiplied by $10^{-5}$ for the synthetic base setting (Table 1), 0.14 for penalty-stress (Table 2), and $10^{-4}$ for rail pressure. For NGSIM, $L_\ell$ is $10^{-10}$ times the empirical 0.9-quantile of the local LCB slope norms. These empirical calibrations are not verified global Lipschitz upper bounds, so the theoretical certificate applies only when its confidence and regularity assumptions hold.

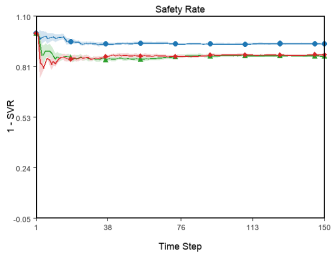
**NGSIM residual simulator and split.** For each recorded transition, we compute the difference between the observed next headway and the nominal kinematic prediction given in Section 6. For a counterfactual regressor, the simulator selects the 16 nearest normalized regressors from the empirical transition library. It estimates the residual by inverse-distance weighting, with weights normalized to sum to one, clips the estimate to $[-10, 10]$, and adds it to the nominal headway prediction. The initial GP training and held-out evaluation sets use a deterministic split by ego-vehicle identity, defined by the pair `(Location, Vehicle_ID)`. No ego vehicle or corresponding trajectory segment is shared across the two sets. We also check that no row index or exact regressor–target pair is shared in any trial. If the next headway is below the minimum threshold of 8 meters, the simulator enters the absorbing state with headway, relative velocity, and ego velocity all equal to zero. All subsequent transitions remain in this state, and the safety value is set to the collision penalty $-10$.

# I Experimental Curves

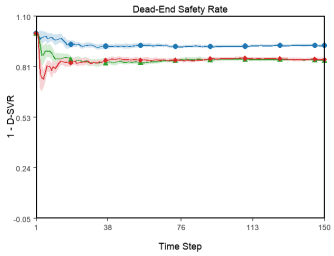
## I.1 Synthetic datasets



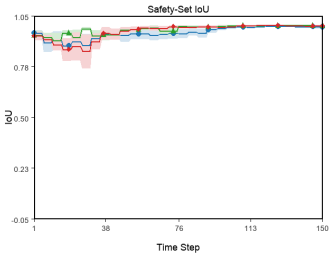

(a) RMSE



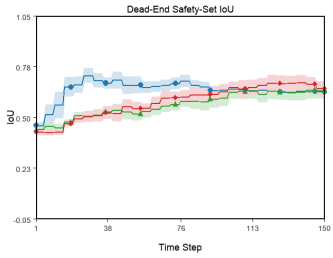

(b) SVR



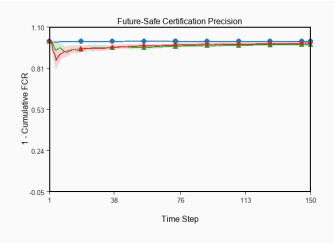

(c) D-SVR




(d) Safety IoU

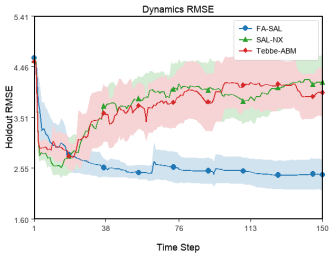



(e) Dead-End IoU

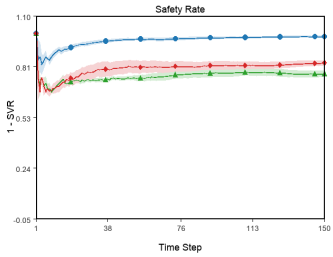



(f) False certification rate

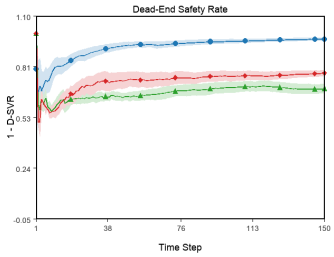
Figure 1: Full round-wise curves corresponding to the summary metrics in Table 1 on the base synthetic benchmark.



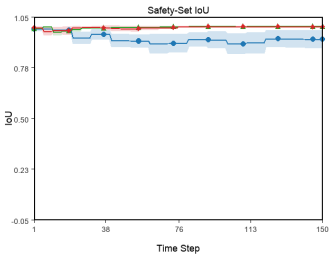

(a) RMSE



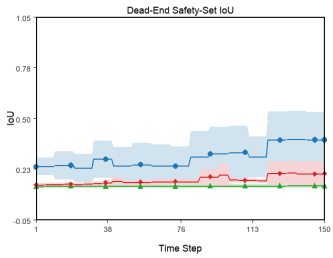

(b) SVR



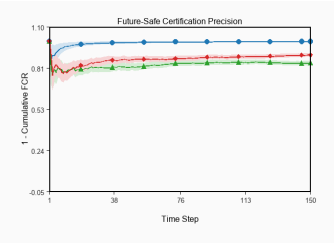

(c) D-SVR




(d) Safety IoU



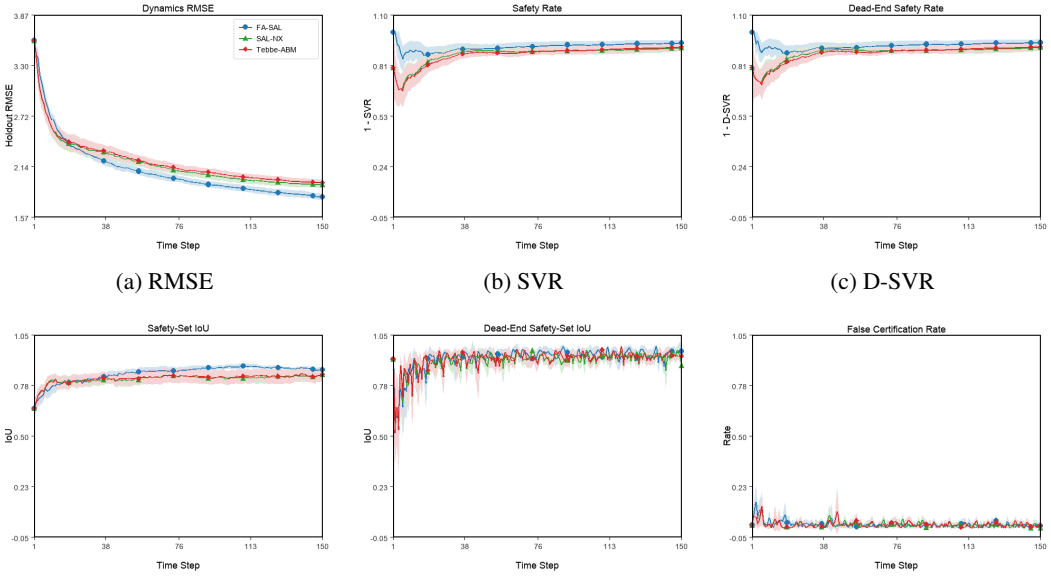

(e) Dead-End IoU




(f) False certification rate

Figure 2: Full round-wise curves corresponding to the summary metrics in Table 2 on the penalty-stress synthetic benchmark.

## I.2 Rail-pressure fluid-system dataset

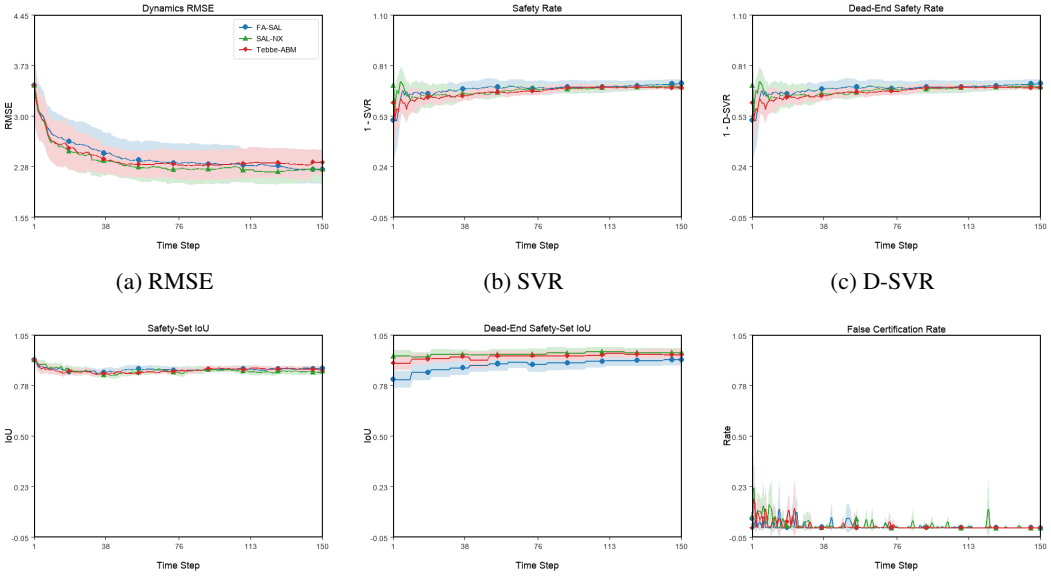



(a) RMSE (b) SVR (c) D-SVR

(d) Safety IoU (e) Dead-End IoU (f) False certification rate

Figure 3: Full round-wise curves corresponding to the summary metrics in Table 3 on the rail-pressure benchmark.

### I.3 NGSIM vehicle-trajectory dataset




(a) RMSE (b) SVR (c) D-SVR

(d) Safety IoU (e) DE-Free IoU (f) False certification rate

Figure 4: Full round-wise curves corresponding to the summary metrics in Table 4 on the NGSIM collision-absorbing benchmark.

# J Additional Experimental Results

This section provides additional experiments to examine how the planning horizon, continuation policy, Lipschitz constants, GP hyperparameter adaptation, and NARX window size affect FA-SAL. We then compare FA-SAL with nominal active-learning model predictive control (AL-MPC) and report runtime and statistical tests. Unless otherwise stated, results are reported as mean $\pm$ standard deviation over 10 random seeds.

## J.1 Horizon and continuation ablations

We first examine the effects of the planning horizon $m$ and continuation policy on model learning, finite-horizon safety, and the availability of certified inputs. For this experiment, we use the penalty-stress benchmark with $m \in \{1, 2, 4, 8\}$ and 150 active-learning decision rounds. We consider three continuation policies: uncertainty-max (UM), random-safe (RS), and greedy-margin (GM). The default FA-SAL policy, UM, chooses the one-step certified continuation with the highest transition-GP uncertainty. In comparison, RS samples from the certified continuation set at random, while GM chooses the continuation with the largest predicted safety margin. When $m = 1$, the policies are identical because no continuation action is needed.

Table 7: Penalty-stress horizon and continuation ablations for $m \in \{1, 2, 4, 8\}$.

| $m$ | Method / policy | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE IoU ↑ |
|---|---|---|---|---|---|---|
| 1 | FA-SAL (UM) | $\mathbf{2.506 \pm 0.215}$ | $0.011 \pm 0.011$ | $\mathbf{0.000 \pm 0.000}$ | $0.960 \pm 0.058$ | $0.797 \pm 0.063$ |
| 1 | FA-SAL (RS) | $\mathbf{2.506 \pm 0.215}$ | $0.011 \pm 0.011$ | $\mathbf{0.000 \pm 0.000}$ | $0.960 \pm 0.058$ | $0.797 \pm 0.063$ |
| 1 | FA-SAL (GM) | $\mathbf{2.506 \pm 0.215}$ | $0.011 \pm 0.011$ | $\mathbf{0.000 \pm 0.000}$ | $0.960 \pm 0.058$ | $0.797 \pm 0.063$ |
| 1 | SAL-NX | $2.860 \pm 0.674$ | $\mathbf{0.009 \pm 0.010}$ | $\mathbf{0.000 \pm 0.000}$ | $0.892 \pm 0.063$ | $\mathbf{0.954 \pm 0.052}$ |
| 1 | Tebbe-ABM | $2.860 \pm 0.674$ | $\mathbf{0.009 \pm 0.010}$ | $\mathbf{0.000 \pm 0.000}$ | $\mathbf{0.988 \pm 0.021}$ | $0.952 \pm 0.054$ |
| 2 | FA-SAL (UM) | $2.431 \pm 0.946$ | $\mathbf{0.024 \pm 0.013}$ | $\mathbf{0.025 \pm 0.014}$ | $0.989 \pm 0.013$ | $0.638 \pm 0.068$ |
| 2 | FA-SAL (RS) | $2.952 \pm 1.264$ | $0.027 \pm 0.015$ | $0.028 \pm 0.015$ | $\mathbf{0.994 \pm 0.006}$ | $\mathbf{0.673 \pm 0.069}$ |
| 2 | FA-SAL (GM) | $2.502 \pm 0.430$ | $0.050 \pm 0.013$ | $0.050 \pm 0.014$ | $0.979 \pm 0.026$ | $\mathbf{0.673 \pm 0.029}$ |
| 2 | SAL-NX | $2.537 \pm 0.502$ | $0.027 \pm 0.014$ | $\mathbf{0.025 \pm 0.013}$ | $0.926 \pm 0.052$ | $0.522 \pm 0.056$ |
| 2 | Tebbe-ABM | $\mathbf{2.307 \pm 0.588}$ | $0.038 \pm 0.013$ | $0.037 \pm 0.012$ | $0.993 \pm 0.010$ | $0.535 \pm 0.094$ |
| 4 | FA-SAL (UM) | $\mathbf{2.640 \pm 0.977}$ | $0.047 \pm 0.022$ | $0.078 \pm 0.035$ | $0.990 \pm 0.012$ | $\mathbf{0.597 \pm 0.176}$ |
| 4 | FA-SAL (RS) | $2.898 \pm 0.645$ | $\mathbf{0.041 \pm 0.025}$ | $\mathbf{0.062 \pm 0.037}$ | $0.996 \pm 0.005$ | $0.454 \pm 0.247$ |
| 4 | FA-SAL (GM) | $3.377 \pm 1.355$ | $0.166 \pm 0.042$ | $0.245 \pm 0.082$ | $0.960 \pm 0.066$ | $0.436 \pm 0.216$ |
| 4 | SAL-NX | $3.709 \pm 1.560$ | $0.180 \pm 0.061$ | $0.288 \pm 0.102$ | $0.976 \pm 0.013$ | $0.446 \pm 0.195$ |
| 4 | Tebbe-ABM | $4.285 \pm 1.344$ | $0.177 \pm 0.026$ | $0.242 \pm 0.027$ | $\mathbf{0.998 \pm 0.002}$ | $0.390 \pm 0.114$ |
| 8 | FA-SAL (UM) | $\mathbf{2.270 \pm 0.572}$ | $\mathbf{0.041 \pm 0.022}$ | $\mathbf{0.289 \pm 0.078}$ | $0.990 \pm 0.016$ | $\mathbf{0.564 \pm 0.151}$ |
| 8 | FA-SAL (RS) | $3.422 \pm 1.099$ | $0.064 \pm 0.028$ | $0.379 \pm 0.052$ | $\mathbf{0.999 \pm 0.002}$ | $0.524 \pm 0.158$ |
| 8 | FA-SAL (GM) | $3.967 \pm 1.240$ | $0.265 \pm 0.126$ | $0.481 \pm 0.175$ | $0.792 \pm 0.168$ | $0.340 \pm 0.151$ |
| 8 | SAL-NX | $3.957 \pm 1.250$ | $0.273 \pm 0.040$ | $0.613 \pm 0.092$ | $0.921 \pm 0.037$ | $0.192 \pm 0.107$ |
| 8 | Tebbe-ABM | $3.013 \pm 0.832$ | $0.213 \pm 0.039$ | $0.520 \pm 0.059$ | $0.993 \pm 0.008$ | $0.239 \pm 0.181$ |

Table 7 shows that the choice of continuation policy matters more as the horizon increases. At $m = 4$, UM yields lower RMSE and higher DE IoU than RS and GM, although RS has slightly lower SVR and D-SVR. At $m = 8$, UM achieves the lowest RMSE, SVR, and D-SVR among the three policies and the two baselines. GM, by contrast, incurs substantially more violations despite maximizing the local safety margin. This difference arises because a continuation action changes the subsequent regression states, so a larger margin at one step need not lead to a more favorable rollout over the full horizon.

The continuation policy also affects how often the certified feasible set becomes empty. Here, the no-feasible rate is the fraction of decision rounds in which no initial candidate passes the full-horizon feasibility test. At $m = 8$, this rate is $0.443$ for UM, $0.215$ for RS, and $0.509$ for GM. For UM, it increases from $0.007$ at $m = 1$ to $0.443$ at $m = 8$, while all three FA-SAL policies show zero observed false certification across the tested horizons. The corresponding runtime for UM increases from $12.68$ to $448.71$ seconds per trial, as reported in Table 15. Together, these results show that a longer horizon captures more delayed safety constraints but can leave fewer inputs available for exploration and increase computation. In practice, the horizon should be long enough to cover the relevant hazard or recovery timescale without making feasible exploration or computation impractical.

### J.2 Sensitivity to Lipschitz constants

We next study how the Lipschitz constants influence the recursive error envelope and the trade-off between safety and exploration. Under the Euclidean norm used here, the NARX shift structure determines $L_x = L_y = 1$. The remaining constants play different roles: $L_f$ controls how transition-model errors propagate through the rollout, while $L_\ell$ converts the resulting state mismatch into a safety margin. To separate these effects, we first vary $L_f$ with the other settings fixed and then vary $L_f$ and $L_\ell$ jointly. These are sensitivity experiments; the scaled values are not independently verified global upper bounds, as required by the theoretical safety guarantee.

We begin with the two shortest horizons, $m = 1$ and $m = 2$. Scaling $L_f$ by $0.5$, $1$, or $2$ produces identical results within each horizon. This follows from the error recursion: because $\bar{e}_{t,1}(u) = 0$, the second-step envelope depends only on the new transition-GP uncertainty. The factor involving $L_f$ affects accumulated error only from the third rollout state, leaving the certificate unchanged for $m \leq 2$.

Table 8: Penalty-stress sensitivity to $L_f$ at $m = 4$ and $m = 8$ using the separate horizon-ablation runs.

| $m$ | $L_f$ scale | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE IoU ↑ | False cert. ↓ | Feasible ratio ↑ |
|---|---|---|---|---|---|---|---|---|
| 4 | $0.5\times$ | $2.537 \pm 0.867$ | $0.048 \pm 0.023$ | $0.077 \pm 0.036$ | $\mathbf{0.994 \pm 0.007}$ | $0.540 \pm 0.189$ | $0.027 \pm 0.086$ | $0.761 \pm 0.133$ |
| 4 | $1\times$ | $2.640 \pm 0.977$ | $0.047 \pm 0.022$ | $0.078 \pm 0.035$ | $0.990 \pm 0.012$ | $0.597 \pm 0.176$ | $\mathbf{0.000 \pm 0.000}$ | $\mathbf{0.765 \pm 0.152}$ |
| 4 | $2\times$ | $2.562 \pm 0.796$ | $0.047 \pm 0.021$ | $0.075 \pm 0.034$ | $0.987 \pm 0.019$ | $0.621 \pm 0.181$ | $\mathbf{0.000 \pm 0.000}$ | $0.751 \pm 0.097$ |
| 4 | $4\times$ | $2.529 \pm 0.969$ | $\mathbf{0.043 \pm 0.011}$ | $\mathbf{0.065 \pm 0.019}$ | $0.979 \pm 0.018$ | $0.608 \pm 0.123$ | $\mathbf{0.000 \pm 0.000}$ | $0.723 \pm 0.162$ |
| 4 | $8\times$ | $\mathbf{2.340 \pm 0.563}$ | $0.046 \pm 0.016$ | $0.075 \pm 0.031$ | $\mathbf{0.994 \pm 0.010}$ | $\mathbf{0.647 \pm 0.162}$ | $\mathbf{0.000 \pm 0.000}$ | $0.653 \pm 0.113$ |
| 8 | $0.5\times$ | $2.811 \pm 1.564$ | $0.074 \pm 0.025$ | $0.349 \pm 0.054$ | $\mathbf{0.994 \pm 0.007}$ | $\mathbf{0.611 \pm 0.143}$ | $\mathbf{0.000 \pm 0.000}$ | $\mathbf{0.566 \pm 0.147}$ |
| 8 | $1\times$ | $\mathbf{2.270 \pm 0.572}$ | $0.041 \pm 0.022$ | $0.289 \pm 0.078$ | $0.990 \pm 0.016$ | $0.564 \pm 0.151$ | $\mathbf{0.000 \pm 0.000}$ | $0.480 \pm 0.151$ |
| 8 | $2\times$ | $2.678 \pm 0.791$ | $0.041 \pm 0.025$ | $0.188 \pm 0.084$ | $0.924 \pm 0.056$ | $0.066 \pm 0.084$ | $\mathbf{0.000 \pm 0.000}$ | $0.080 \pm 0.031$ |
| 8 | $4\times$ | $3.783 \pm 0.741$ | $\mathbf{0.023 \pm 0.036}$ | $\mathbf{0.123 \pm 0.170}$ | $0.820 \pm 0.088$ | $0.000 \pm 0.000$ | $\mathbf{0.000 \pm 0.000}$ | $0.00012 \pm 0.00038$ |
| 8 | $8\times$ | $3.816 \pm 0.768$ | $\mathbf{0.023 \pm 0.036}$ | $\mathbf{0.123 \pm 0.170}$ | $0.820 \pm 0.088$ | $0.000 \pm 0.000$ | $\mathbf{0.000 \pm 0.000}$ | $0.000 \pm 0.000$ |

At longer horizons, the effect of $L_f$ becomes more visible because past prediction errors are propagated over more rollout steps. As shown in Table 8, the violation metrics vary little across scales at $m = 4$, although the $0.5\times$ setting yields an observed false-certification rate of $0.027$. At $m = 8$, increasing the scale from $0.5$ to $8$ reduces D-SVR from $0.349$ to $0.123$, but also reduces DE IoU from $0.611$ to zero and the feasible ratio from $0.566$ to zero. Hence, the lower violation rates at large scales accompany the rejection of nearly all candidate trajectories, rather than better recovery of the future-safe set. These results illustrate the asymmetric effect of misspecification: a value that is too small may leave the error envelope insufficient, whereas an excessively large value makes certification overly conservative.

Table 9: Rail-pressure sensitivity to $L_f$ at $m = 5$ over 10 random seeds.

| $L_f$ scale | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE IoU ↑ | False cert. ↓ | Feasible ratio ↑ |
|---|---|---|---|---|---|---|---|
| $0.5\times$ | $1.814 \pm 0.072$ | $0.039 \pm 0.026$ | $0.037 \pm 0.026$ | $0.813 \pm 0.062$ | $\mathbf{0.955 \pm 0.051}$ | $0.005 \pm 0.010$ | $\mathbf{0.850 \pm 0.199}$ |
| $1\times$ | $1.814 \pm 0.085$ | $0.038 \pm 0.019$ | $0.037 \pm 0.019$ | $\mathbf{0.867 \pm 0.028}$ | $0.940 \pm 0.068$ | $0.013 \pm 0.018$ | $0.712 \pm 0.230$ |
| $2\times$ | $\mathbf{1.810 \pm 0.087}$ | $0.026 \pm 0.015$ | $0.024 \pm 0.016$ | $0.843 \pm 0.042$ | $0.933 \pm 0.065$ | $0.001 \pm 0.002$ | $0.829 \pm 0.209$ |
| $4\times$ | $1.873 \pm 0.097$ | $0.009 \pm 0.009$ | $0.007 \pm 0.009$ | $0.846 \pm 0.048$ | $0.850 \pm 0.145$ | $0.004 \pm 0.006$ | $0.718 \pm 0.253$ |
| $8\times$ | $2.720 \pm 0.454$ | $\mathbf{0.000 \pm 0.000}$ | $\mathbf{0.000 \pm 0.000}$ | $0.756 \pm 0.104$ | $0.273 \pm 0.254$ | $\mathbf{0.000 \pm 0.000}$ | $0.252 \pm 0.247$ |

To test whether this pattern extends beyond the synthetic setting, we repeat the $L_f$ sensitivity study on the rail-pressure benchmark with $m = 5$. Table 9 shows relatively stable results under moderate scaling. At $8\times$, however, SVR and D-SVR both reach zero, while RMSE increases to $2.720$ and the feasible ratio and DE IoU fall to $0.252$ and $0.273$, respectively. Thus, the rail-pressure results show the same trade-off: fewer observed violations at a large $L_f$ can come at the cost of reduced feasible exploration.

We finally consider the combined effects of $L_f$ and $L_\ell$ rather than varying $L_f$ alone. Table 10 uses scales $\{0.5, 1, 2\}$ for both constants on penalty-stress at $m = 4$, following the exact protocol of Table 2. The $(1, 1)$ setting reproduces the main-paper FA-SAL results. Across all nine configurations, FA-SAL has lower mean RMSE, SVR, and D-SVR than SAL-NX and Tebbe-ABM, although certification and feasible-set availability differ across settings. For instance, false certification is $0.100$ at $(L_f, L_\ell)$ scales $(1, 0.5)$ but zero in the other eight settings. Similarly, the no-feasible rate changes from $0.754$ at $(1, 1)$ to $0.369$ at $(0.5, 0.5)$ and $0.845$ at $(2, 2)$. The joint results reinforce the same trade-off: smaller margins may weaken empirical certification, while larger margins restrict

Table 10: Joint sensitivity to $L_f$ and $L_\ell$ on penalty-stress at $m = 4$, using the exact Table 2 protocol.

| Method | $L_f$ scale | $L_\ell$ scale | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE IoU ↑ |
|---|---|---|---|---|---|---|---|
| SAL-NX | – | – | $4.168 \pm 0.979$ | $0.227 \pm 0.036$ | $0.312 \pm 0.051$ | $\mathbf{0.998 \pm 0.002}$ | $0.135 \pm 0.022$ |
| Tebbe-ABM | – | – | $3.971 \pm 0.884$ | $0.167 \pm 0.028$ | $0.224 \pm 0.035$ | $0.996 \pm 0.006$ | $0.200 \pm 0.133$ |
| FA-SAL | 1 | 1 | $2.428 \pm 0.563$ | $\mathbf{0.018 \pm 0.006}$ | $\mathbf{0.033 \pm 0.013}$ | $0.929 \pm 0.095$ | $0.383 \pm 0.298$ |
| FA-SAL | 0.5 | 0.5 | $2.594 \pm 0.463$ | $0.039 \pm 0.016$ | $0.067 \pm 0.032$ | $0.991 \pm 0.012$ | $\mathbf{0.646 \pm 0.254}$ |
| FA-SAL | 0.5 | 1 | $2.647 \pm 0.842$ | $0.034 \pm 0.016$ | $0.057 \pm 0.031$ | $0.954 \pm 0.086$ | $0.465 \pm 0.237$ |
| FA-SAL | 0.5 | 2 | $2.529 \pm 0.504$ | $0.025 \pm 0.014$ | $0.048 \pm 0.026$ | $0.886 \pm 0.112$ | $0.225 \pm 0.188$ |
| FA-SAL | 1 | 0.5 | $\mathbf{2.414 \pm 0.617}$ | $0.043 \pm 0.022$ | $0.077 \pm 0.045$ | $0.971 \pm 0.036$ | $0.582 \pm 0.296$ |
| FA-SAL | 1 | 2 | $2.779 \pm 0.488$ | $0.029 \pm 0.022$ | $0.067 \pm 0.048$ | $0.922 \pm 0.095$ | $0.209 \pm 0.165$ |
| FA-SAL | 2 | 0.5 | $2.618 \pm 0.556$ | $0.037 \pm 0.020$ | $0.072 \pm 0.039$ | $0.898 \pm 0.099$ | $0.379 \pm 0.282$ |
| FA-SAL | 2 | 1 | $3.181 \pm 0.892$ | $0.023 \pm 0.018$ | $0.056 \pm 0.039$ | $0.871 \pm 0.111$ | $0.153 \pm 0.157$ |
| FA-SAL | 2 | 2 | $2.674 \pm 0.436$ | $0.029 \pm 0.016$ | $0.077 \pm 0.040$ | $0.799 \pm 0.157$ | $0.095 \pm 0.089$ |

exploration. They do not establish valid global Lipschitz bounds or extend the safety guarantee to misspecified constants.

### J.3 Periodic GP hyperparameter adaptation

We next evaluate whether periodically updating GP hyperparameters reduces sensitivity to their initialization and how these updates affect safety certification. Both the transition and safety GPs use isotropic squared-exponential kernels. For each GP, we optimize the length scale, signal variance, and observation-noise variance at rounds $50, 60, \ldots, 150$ by minimizing the negative log marginal likelihood over all accumulated observations using L-BFGS-B [6]. This optimization is unconstrained by the Lipschitz certificate. To evaluate sensitivity to initialization, we compare fixed and adaptive settings with initial length scales and signal variances scaled jointly by $0.5$, $1$, or $2$.

Because changing GP hyperparameters also changes the posterior quantities used for certification, we update both posteriors and recompute the affected Lipschitz estimates after each adaptation. Specifically, we estimate $L_f$ using the median absolute coordinate-wise derivative of the transition-GP posterior mean over the evaluation points. For $L_\ell$, we compute forward and backward finite-difference slopes of the updated safety LCB along each coordinate, retain the larger absolute slope, combine these slopes using the Euclidean norm, and take the empirical 90th percentile over the evaluation points. The gradient-based estimate of $L_f$ is applicable to the differentiable kernels used in this experiment. Neither estimate is a verified global Lipschitz upper bound; both are used for empirical calibration.

Table 11: Fixed and periodically adapted GPs on penalty-stress at $m = 4$, with the exact Table 2 protocol and 150 rounds.

| Setting | Init. scale | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE IoU ↑ |
|---|---|---|---|---|---|---|
| Nominal-fixed | 1 | $2.428 \pm 0.563$ | $\mathbf{0.018 \pm 0.006}$ | $\mathbf{0.033 \pm 0.013}$ | $0.929 \pm 0.095$ | $0.383 \pm 0.298$ |
| Nominal-adaptive | 1 | $\mathbf{2.223 \pm 0.485}$ | $0.026 \pm 0.014$ | $0.047 \pm 0.023$ | $\mathbf{0.995 \pm 0.009}$ | $\mathbf{0.527 \pm 0.268}$ |
| Low-fixed | 0.5 | $4.020 \pm 1.247$ | $\mathbf{0.027 \pm 0.027}$ | $\mathbf{0.057 \pm 0.067}$ | $\mathbf{0.984 \pm 0.021}$ | $0.024 \pm 0.038$ |
| Low-adaptive | 0.5 | $\mathbf{2.809 \pm 1.360}$ | $0.029 \pm 0.018$ | $\mathbf{0.057 \pm 0.043}$ | $0.966 \pm 0.053$ | $\mathbf{0.335 \pm 0.248}$ |
| High-fixed | 2 | $2.711 \pm 1.465$ | $0.048 \pm 0.013$ | $0.085 \pm 0.027$ | $0.978 \pm 0.024$ | $\mathbf{0.680 \pm 0.140}$ |
| High-adaptive | 2 | $\mathbf{2.394 \pm 0.408}$ | $\mathbf{0.041 \pm 0.011}$ | $\mathbf{0.077 \pm 0.029}$ | $\mathbf{0.994 \pm 0.009}$ | $0.366 \pm 0.198$ |

We evaluate these updates using the exact $m = 4$, 150-round penalty-stress protocol of Table 2. As Table 11 shows, periodic adaptation reduces RMSE under all three initializations. Under nominal initialization, RMSE decreases from $2.428$ to $2.223$, while Safety IoU increases from $0.929$ to $0.995$, DE IoU increases from $0.383$ to $0.527$, and the no-feasible rate decreases from $0.754$ to $0.621$. These gains, however, come with increases in SVR from $0.018$ to $0.026$ and D-SVR from $0.033$ to $0.047$. The effects also differ across initializations: the no-feasible rate decreases from $0.885$ to $0.680$ for low initialization, but increases from $0.313$ to $0.565$ for high initialization, where DE IoU decreases from $0.680$ to $0.366$. All six fixed and adaptive settings show zero observed false certification. Overall, adaptation improves predictive accuracy but does not consistently improve safety or feasible-set recovery. Moreover, the fixed-hyperparameter theoretical guarantee does not automatically apply to this empirical adaptation procedure.

### J.4 Comparison with nominal active-learning MPC

To assess the role of recursive uncertainty propagation, we next compare FA-SAL with nominal active-learning MPC (AL-MPC). At each decision point, AL-MPC plans an informative $m$-step control sequence under a nominal transition model, checks a flat safety constraint along the predicted trajectory, executes the first action, and replans. Although its objective balances model-information gain and regulation terms, its safety constraint does not incorporate the recursive transition-error margin used by FA-SAL. For a fair comparison, we match the total number of environment interactions and use dedicated MPC runs separate from the continuation ablation.

Table 12: FA-SAL versus AL-MPC on penalty-stress for $m \in \{1, 2, 4, 8\}$ using dedicated MPC-comparison runs, separate from the continuation ablation.

| $m$ | Method | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE IoU ↑ | False cert. ↓ |
|---|---|---|---|---|---|---|---|
| 1 | FA-SAL | $\mathbf{2.506 \pm 0.215}$ | $\mathbf{0.011 \pm 0.011}$ | $\mathbf{0.000 \pm 0.000}$ | $0.960 \pm 0.058$ | $0.797 \pm 0.063$ | $\mathbf{0.000 \pm 0.000}$ |
| 1 | AL-MPC | $2.648 \pm 0.871$ | $0.035 \pm 0.024$ | $\mathbf{0.000 \pm 0.000}$ | $\mathbf{0.987 \pm 0.016}$ | $\mathbf{0.818 \pm 0.014}$ | $\mathbf{0.000 \pm 0.000}$ |
| 2 | FA-SAL | $\mathbf{2.035 \pm 0.527}$ | $\mathbf{0.023 \pm 0.010}$ | $\mathbf{0.023 \pm 0.009}$ | $0.983 \pm 0.016$ | $\mathbf{0.701 \pm 0.064}$ | $\mathbf{0.000 \pm 0.000}$ |
| 2 | AL-MPC | $4.401 \pm 1.247$ | $0.082 \pm 0.038$ | $0.083 \pm 0.038$ | $\mathbf{0.999 \pm 0.002}$ | $0.594 \pm 0.049$ | $0.100 \pm 0.300$ |
| 4 | FA-SAL | $\mathbf{2.586 \pm 0.809}$ | $\mathbf{0.025 \pm 0.016}$ | $\mathbf{0.049 \pm 0.030}$ | $0.920 \pm 0.092$ | $\mathbf{0.405 \pm 0.223}$ | $\mathbf{0.000 \pm 0.000}$ |
| 4 | AL-MPC | $5.357 \pm 1.212$ | $0.139 \pm 0.008$ | $0.412 \pm 0.027$ | $\mathbf{1.000 \pm 0.001}$ | $0.372 \pm 0.057$ | $0.227 \pm 0.395$ |
| 8 | FA-SAL | $\mathbf{2.270 \pm 0.572}$ | $0.041 \pm 0.022$ | $0.289 \pm 0.078$ | $0.990 \pm 0.016$ | $\mathbf{0.564 \pm 0.151}$ | $\mathbf{0.000 \pm 0.000}$ |
| 8 | AL-MPC | $4.659 \pm 1.697$ | $\mathbf{0.032 \pm 0.035}$ | $\mathbf{0.158 \pm 0.090}$ | $\mathbf{1.000 \pm 0.001}$ | $0.174 \pm 0.050$ | $0.100 \pm 0.300$ |

Table 12 shows that FA-SAL achieves lower RMSE at every tested horizon and lower SVR and D-SVR at $m = 2$ and $m = 4$. It also achieves higher DE IoU at $m = 2, 4, 8$ and shows zero observed false certification throughout these comparisons. AL-MPC, however, attains higher Safety IoU at every horizon and lower violation rates at $m = 8$, with SVR of $0.032$ versus $0.041$ and D-SVR of $0.158$ versus $0.289$. At the same horizon, FA-SAL retains lower RMSE ($2.270$ versus $4.659$), higher DE IoU ($0.564$ versus $0.174$), and lower false certification ($0$ versus $0.100$). Thus, neither method dominates across all metrics: the comparison highlights the difference between nominal trajectory screening and uncertainty-aware certification.

Table 13: Feasible-set statistics for uncertainty-max FA-SAL, greedy-margin FA-SAL (GM), and AL-MPC. Set size is the mean number of feasible initial candidates; no-feasible is the fraction of rounds with an empty feasible set. AL-MPC uses nominal feasibility, whereas FA-SAL uses the rollout certificate.

| $m$ | Method | Set size ↑ | Feasible ratio ↑ | No-feasible ↓ |
|---|---|---|---|---|
| 1 | FA-SAL | $10.326 \pm 0.534$ | $0.939 \pm 0.049$ | $0.022 \pm 0.017$ |
| 1 | FA-SAL (GM) | $10.326 \pm 0.534$ | $0.939 \pm 0.049$ | $0.022 \pm 0.017$ |
| 1 | AL-MPC | $\mathbf{10.359 \pm 0.234}$ | $\mathbf{0.942 \pm 0.021}$ | $\mathbf{0.008 \pm 0.009}$ |
| 2 | FA-SAL | $9.863 \pm 0.679$ | $0.897 \pm 0.062$ | $0.053 \pm 0.029$ |
| 2 | FA-SAL (GM) | $9.269 \pm 0.774$ | $0.843 \pm 0.070$ | $0.081 \pm 0.033$ |
| 2 | AL-MPC | $\mathbf{10.027 \pm 0.224}$ | $\mathbf{0.912 \pm 0.020}$ | $\mathbf{0.033 \pm 0.019}$ |
| 4 | FA-SAL | $\mathbf{8.419 \pm 1.667}$ | $\mathbf{0.765 \pm 0.152}$ | $0.167 \pm 0.133$ |
| 4 | FA-SAL (GM) | $7.739 \pm 0.904$ | $0.704 \pm 0.082$ | $0.215 \pm 0.085$ |
| 4 | AL-MPC | $8.039 \pm 0.185$ | $0.731 \pm 0.017$ | $\mathbf{0.078 \pm 0.006}$ |
| 8 | FA-SAL | $5.277 \pm 1.666$ | $0.480 \pm 0.151$ | $0.443 \pm 0.164$ |
| 8 | FA-SAL (GM) | $3.903 \pm 1.142$ | $0.355 \pm 0.104$ | $0.509 \pm 0.158$ |
| 8 | AL-MPC | $\mathbf{6.687 \pm 1.164}$ | $\mathbf{0.608 \pm 0.106}$ | $\mathbf{0.022 \pm 0.016}$ |

To examine how the screening rules affect candidate availability, Table 13 compares the feasible sets of UM-based FA-SAL, GM-based FA-SAL, and AL-MPC over 11 initial candidates per round. The feasible ratio is the fraction of these 11 candidates that pass the method's feasibility test, and No-feasible is the fraction of rounds in which none pass. Within FA-SAL, GM reduces the mean certified feasible-set size relative to UM by approximately $6.0\%$, $8.1\%$, and $26.0\%$ at $m = 2, 4, 8$, respectively. This supports the observation that maximizing the local safety margin need not preserve more full-horizon feasible candidates. At $m = 8$, AL-MPC has a higher feasible ratio than UM-based FA-SAL ($0.608$ versus $0.480$). However, AL-MPC uses nominal feasibility, whereas FA-SAL uses a

margin certificate, so a larger nominal feasible set does not imply stronger safety guarantees. These findings concern the evaluated nominal AL-MPC implementation and do not extend to MPC methods in general.

### J.5 NARX window-size ablation

We also examine the effect of NARX history length on model learning and rollout certification. The output and input window lengths are $d_y$ and $d_u$, giving a regressor dimension of $d_y + d_u$. To isolate the effect of these choices, we vary the two window lengths while retaining a budget of 150 active-learning decision rounds. Each round selects one trajectory and executes up to $m$ environment transitions along it.

Table 14: Performance in the NARX window-size ablation over 150 active-learning decision rounds. Each row gives its output and input window lengths.

| $d_y$ | $d_u$ | $d_y + d_u$ | RMSE ↓ | SVR ↓ | D-SVR ↓ | Safety IoU ↑ | DE IoU ↑ |
|---|---|---|---|---|---|---|---|
| 2 | 1 | 3 | $2.898 \pm 0.645$ | $0.041 \pm 0.025$ | $0.062 \pm 0.037$ | $\mathbf{0.996 \pm 0.005}$ | $0.454 \pm 0.247$ |
| 4 | 1 | 5 | $3.298 \pm 0.334$ | $0.029 \pm 0.016$ | $0.085 \pm 0.063$ | $0.960 \pm 0.048$ | $\mathbf{0.658 \pm 0.055}$ |
| 8 | 1 | 9 | $3.940 \pm 0.651$ | $0.026 \pm 0.015$ | $0.113 \pm 0.151$ | $0.980 \pm 0.035$ | $0.361 \pm 0.113$ |
| 2 | 2 | 4 | $\mathbf{2.261 \pm 0.639}$ | $0.032 \pm 0.019$ | $0.049 \pm 0.031$ | $0.984 \pm 0.033$ | $0.556 \pm 0.140$ |
| 2 | 4 | 6 | $2.932 \pm 0.996$ | $0.025 \pm 0.015$ | $0.043 \pm 0.033$ | $0.810 \pm 0.208$ | $0.401 \pm 0.137$ |
| 2 | 8 | 10 | $3.373 \pm 0.725$ | $\mathbf{0.016 \pm 0.013}$ | $\mathbf{0.030 \pm 0.024}$ | $0.479 \pm 0.246$ | $0.090 \pm 0.079$ |
| 4 | 4 | 8 | $3.973 \pm 0.402$ | $0.023 \pm 0.021$ | $0.089 \pm 0.092$ | $0.956 \pm 0.054$ | $0.345 \pm 0.135$ |
| 8 | 8 | 16 | $4.525 \pm 0.835$ | $0.022 \pm 0.018$ | $0.107 \pm 0.129$ | $0.992 \pm 0.009$ | $0.010 \pm 0.013$ |

Table 14 indicates that increasing history length does not consistently improve prediction or future-safe set recovery. The lowest RMSE, $2.261$, occurs at $(d_y, d_u) = (2, 2)$, whereas the $(8, 8)$ setting yields RMSE of $4.525$ and DE IoU of $0.010$. Even when $d_y$ is fixed at 2, increasing $d_u$ from 1 to 8 reduces Safety IoU from $0.996$ to $0.479$ and DE IoU from $0.454$ to $0.090$, although the observed violation rates decrease. These differences show why violation rates should be interpreted together with the extent of safe-set recovery.

The feasibility statistics show a similar effect. With $d_y = 2$, increasing $d_u$ from 1 to 8 raises the no-feasible rate from $0.107$ to $0.719$, and the rate reaches $0.865$ at $(d_y, d_u) = (8, 8)$. False certification is zero in seven of the eight configurations, but reaches $0.100$ at $(d_y, d_u) = (4, 1)$. Although the error recursion has no explicit factor of $d_y + d_u$, a fixed data budget may cover a higher-dimensional regressor space less densely. The resulting GP uncertainty can enlarge the rollout-error envelope and lower the safety LCB, making certification more difficult. The runtime decreases from $87.09$ seconds at $(d_y, d_u) = (2, 1)$ to $22.84$ seconds at $(8, 8)$, mainly because infeasible rollouts are rejected earlier, not because larger windows reduce GP computational complexity. These observations favor choosing the smallest history orders that adequately capture the dynamics, using safe historical data and temporal validation when available.

### J.6 Runtime and paired statistical tests

Finally, we examine the computational cost of FA-SAL and the statistical significance of its performance differences from SAL-NX and Tebbe-ABM. The experiments were conducted on an Intel Core i7-13700K CPU with 32 GB RAM under Ubuntu 24.04.1; no GPU acceleration was used. We use two exact dense GPs, caching their Cholesky factors and posterior coefficients within each decision round without sparse approximations or incremental Cholesky updates. The experiments evaluate offline active learning rather than real-time control.

Table 15 reports the total wall-clock time per sequential trial. On penalty-stress, FA-SAL takes $104.09$ seconds at $m = 4$ and $448.71$ seconds at $m = 8$, compared with $14.77$ and $44.46$ seconds for SAL-NX and $52.10$ and $74.31$ seconds for Tebbe-ABM. This increase reflects the additional work required by longer rollouts. For fixed numbers of initial candidates and continuation options, the number of GP evaluations grows approximately linearly with $m$, but total runtime also includes feasibility checks, acquisition evaluations, and updates as the training set grows. With $n$ observations, an exact dense-GP refit requires cubic time and quadratic memory, while cached mean-and-variance predictions still require quadratic-time variance evaluation in general. These costs limit scalability

Table 15: Wall-clock seconds per sequential trial. Penalty-stress trials contain 150 rounds and use 10 paired seeds. Dashes denote baseline timings not reported for the other benchmarks.

| Benchmark | $m$ | FA-SAL | SAL-NX | Tebbe-ABM |
|---|---|---|---|---|
| Penalty-stress | 1 | $12.68 \pm 4.66$ | $\mathbf{3.71 \pm 1.10}$ | $13.86 \pm 4.61$ |
| Penalty-stress | 2 | $28.63 \pm 6.72$ | $\mathbf{6.73 \pm 1.13}$ | $32.50 \pm 9.60$ |
| Penalty-stress | 4 | $104.09 \pm 27.30$ | $\mathbf{14.77 \pm 2.05}$ | $52.10 \pm 19.13$ |
| Penalty-stress | 8 | $448.71 \pm 232.62$ | $\mathbf{44.46 \pm 6.24}$ | $74.31 \pm 14.97$ |
| Rail pressure | 5 | $32.9 \pm 0.3$ | – | – |
| NGSIM | 4 | $211.6 \pm 5.7$ | – | – |

at longer horizons; the reported 32 GB specification is not a peak-memory measurement, and the timings do not establish real-time performance.

To assess whether the observed differences are consistent across trials, we also perform paired sign-flip tests using 10 matched random seeds on penalty-stress. Within each horizon, we apply Holm correction separately to the four SVR/D-SVR comparisons and the two RMSE comparisons.

Table 16: Raw and Holm-adjusted $p$-values for paired comparisons of FA-SAL with SAL-NX and Tebbe-ABM on penalty-stress.

| $m$ | Baseline | Metric | Raw $p$ | Adjusted $p$ |
|---|---|---|---|---|
| 4 | SAL-NX | SVR | 0.00195 | 0.00781 |
| 4 | SAL-NX | D-SVR | 0.00391 | 0.00781 |
| 4 | SAL-NX | RMSE | 0.09766 | 0.09766 |
| 4 | Tebbe-ABM | SVR | 0.00195 | 0.00781 |
| 4 | Tebbe-ABM | D-SVR | 0.00195 | 0.00781 |
| 4 | Tebbe-ABM | RMSE | 0.00391 | 0.00781 |
| 8 | SAL-NX | SVR | 0.00195 | 0.00781 |
| 8 | SAL-NX | D-SVR | 0.00195 | 0.00781 |
| 8 | SAL-NX | RMSE | 0.00195 | 0.00391 |
| 8 | Tebbe-ABM | SVR | 0.00195 | 0.00781 |
| 8 | Tebbe-ABM | D-SVR | 0.00195 | 0.00781 |
| 8 | Tebbe-ABM | RMSE | 0.03906 | 0.03906 |

Table 16 gives both raw and adjusted $p$-values. At $m = 4$ and $m = 8$, every SVR and D-SVR comparison with SAL-NX and Tebbe-ABM has Holm-adjusted $p = 0.00781$. The RMSE comparisons are less consistent: the adjusted $p$-value is $0.09766$ against SAL-NX at $m = 4$ and $0.03906$ against Tebbe-ABM at $m = 8$. Thus, the tests provide stronger evidence for the observed safety differences than for uniform RMSE improvements, and their conclusions are limited to the tested penalty-stress protocols rather than the rail-pressure or NGSIM benchmarks.

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes].

   Justification: The abstract and introduction clearly state the proposed future-aware safe active learning framework, the notion of $m$-step dead-end inputs, the theoretical guarantees, and the empirical scope. The claims are supported by the theoretical results in Section 5 and experiments in Section 6.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes].

   Justification: The paper discusses the finite-sample conservativeness of the certificate, the dependence on GP uncertainty, and the role of finite candidate sets and finite-horizon planning. Additional limitations and assumptions are also reflected in the theory and experimental discussion.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes].

   Justification: The paper states the required assumptions in Section 5 and provides complete proofs for the main theorem, interior recovery theorem, dead-end exclusion corollary, and auxiliary lemmas in the appendix.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes].

   Justification: The experimental section specifies the benchmarks, baselines, evaluation metrics, safety functions, horizons, and dataset construction details needed to reproduce the main comparisons. Further implementation details are provided in the appendix.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Code and data are not publicly released at submission time. The paper provides the algorithmic details, benchmark descriptions, and evaluation protocol, and we will release the reproduction materials after the entire process is complete.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes].

   Justification: The paper describes the synthetic, rail-pressure, and NGSIM benchmark settings, safety definitions, candidate/query construction, baselines, and evaluation metrics. Hyperparameter and implementation details are included to the extent needed to interpret the reported results.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes].

Justification: The paper reports mean and standard deviation over trials for all main benchmark tables, and the appendix provides round-wise curves with uncertainty bands.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix J.6 describes the CPU-only setup (Intel Core i7-13700K, 32 GB RAM, Ubuntu 24.04.1), and Table 15 reports wall-clock time per sequential trial.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes].

   Justification: The work is methodological and experimental, uses standard benchmark or publicly available data sources, and is not designed for harmful deployment. The paper preserves anonymity and does not involve deceptive, privacy-invasive, or harmful data collection.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes].

    Justification: The paper motivates positive impacts in safety-critical domains such as robotics, industrial systems, and medical decision support. It also acknowledges that safety certificates are conservative and finite-sample errors may still occur, which is relevant for cautious deployment.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A].

    Justification: The paper does not release high-risk pretrained models, generative models, scraped datasets, or assets with substantial misuse risk.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes].

    Justification: Existing benchmarks and datasets, including the rail-pressure benchmark and NGSIM car-following data, are cited in the paper. The paper does not claim ownership of these external assets.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A].

    Justification: The paper does not introduce or release a new dataset, pretrained model, or benchmark asset as a primary contribution.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing experiments or direct human-subject studies.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing or direct research with human subjects, so IRB approval is not applicable.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A].

Justification: LLMs are not used as an important, original, or non-standard component of the proposed method, experiments, or theoretical analysis.