

Safety-Aware Preference-Based Learning for Safety-Critical Control

Ryan K. Cosner¹

Maegan Tucker¹

Andrew J. Taylor¹

Kejun Li¹

Tamás G. Molnár¹

Wyatt Ubellacker¹

Anil Alan²

Gábor Orosz²

Yisong Yue¹

Aaron D. Ames¹

RKCOSNER@CALTECH.EDU

MTUCKER@CALTECH.EDU

AJTAYLOR@CALTECH.EDU

KLI5@CALTECH.EDU

TMOLNAR@CALTECH.EDU

WUBELLAC@CALTECH.EDU

ANILALAN@UMICH.EDU

OROSZ@UMICH.EDU

YYUE@CALTECH.EDU

AMES@CALTECH.EDU

¹ *California Institute of Technology, Pasadena, CA, USA*

² *University of Michigan, Ann Arbor, MI, USA*

Abstract

Bringing dynamic robots into the wild requires a tenuous balance between performance and safety. Yet controllers designed to provide robust safety guarantees often result in conservative behavior, and tuning these controllers to find the ideal trade-off between performance and safety typically requires domain expertise or a carefully constructed reward function. This work presents a design paradigm for systematically achieving behaviors that balance performance and robust safety by integrating *safety-aware* Preference-Based Learning (PBL) with Control Barrier Functions (CBFs). Fusing these concepts—safety-aware learning and safety-critical control—gives a robust means to achieve safe behaviors on complex robotic systems in practice. We demonstrate the capability of this design paradigm to achieve safe and performant perception-based autonomous operation of a quadrupedal robot both in simulation and experimentally on hardware.

Keywords: Preference-Based Learning, Safety-Critical Control, Robotics

1. Introduction

The increasing demands of modern engineering problems have required a commensurate increase in the complexity of the underlying control systems being used. The process of designing these complex control systems is often accomplished by separating the design into individual subsystems such as sensing, planning, and low-level control, which are later integrated. A principal challenge in the integration of such complex systems is balancing safety with performance at the system level. When each individual subsystem is designed using over-approximations of worst-case scenarios, the integrated system becomes extremely conservative and exhibits poor performance (Singletary et al., 2021; Alan et al., 2022). The commonly employed alternative is tuning the safety-performance trade-off of each component to achieve the desired system-level behavior (Ma et al., 2017), which can be challenging even for domain experts as the tuning is often done via qualitative assessments.

For instance, for complex safety-critical systems, Control Barrier Functions (CBFs) have become a popular tool for the constructive synthesis of model-based controllers that endow nonlinear systems with rigorous guarantees of safety (Ames et al., 2014, 2019; Hobbs et al., 2021). As these safety guarantees are susceptible to inaccuracies in the models of a system’s dynamics, actuators, and sensors, approaches have been proposed to deal with model uncertainty (Wang et al., 2018;

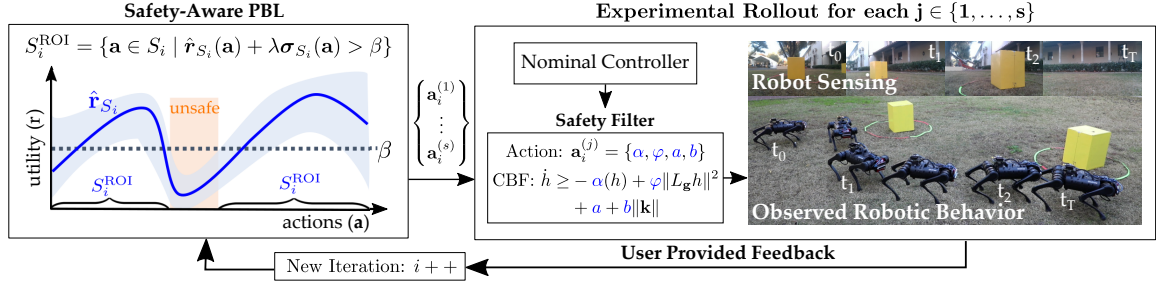


Figure 1. An overview of design paradigm with Safety-Aware Preference-Based Learning.

Taylor and Ames, 2020; Castañeda et al., 2020; Taylor et al., 2020), disturbances (Jankovic, 2018; Kolathaya and Ames, 2018; Clark, 2019; Santoyo et al., 2019; Alan et al., 2022; Choi et al., 2021), and measurement errors (Takano and Yamakita, 2018; Dean et al., 2020; Cosner et al., 2021). These approaches can work well when deployed independently, but can often lead to extremely conservative systems with poor performance when used in conjunction. In practice, achieving performant system behaviors with these methods is accomplished by conceding theoretical safety guarantees for over-approximations of worst-case scenarios and tuning controller robustness parameters.

To reduce the burden on experts in controller tuning, we seek to incorporate Preference-Based Learning (PBL) into the design of complex safety-critical control systems. PBL has shown to be a powerful tool for converting subjective user preferences of system behavior (e.g., behavior A has a superior performance-safety trade-off than behavior B) into quantitative adjustments to design parameters. The main advantage of online PBL is its ability to interactively infer a user’s latent utility function using only subjective feedback such as pairwise preferences and ordinal labels (Yue et al., 2012; Shivaswamy and Joachims, 2012). This methodology has been demonstrated in application for exoskeleton gait optimization (Tucker et al., 2020b), bipedal locomotion (Tucker et al., 2021), spinal cord stimulation (Sui et al., 2018), trajectory planning (Sadigh et al., 2017; Bıyık et al., 2020; Jain et al., 2015), search engines (Raman et al., 2013), and recommender systems (De Gemmis et al., 2009). For applications with actions that may be classified as safe or unsafe, *safety-critical* PBL algorithms have been demonstrated to prevent unsafe actions from being sampled (Sui et al., 2015, 2018; Berkenkamp et al., 2016). However, these safety-critical algorithms require approximating worst-case scenarios, and over-approximations may yield the characterization of performant and safe actions as catastrophically unsafe. Thus, we seek to formulate a *safety-aware* approach to PBL that balances avoiding sampling unsafe actions with being overly conservative.

In this work we propose a design paradigm for synthesizing performant and robust safety-critical controllers on real systems via safety-aware online PBL (illustrated in Fig. 1). The contributions of this work are threefold. First, we propose Safety-Aware LineCoSpar (SA-LineCoSpar), a modified version of LineCoSpar (Tucker et al., 2020a) capable of high-dimensional preference-based Bayesian optimization while also accounting for safety. Second, we combine the robustness properties of Measurement-Robust CBFs (MR-CBFs) (Dean et al., 2020) to measurement uncertainty and Input-to-State Safe CBFs (ISSf-CBFs) (Kolathaya and Ames, 2018) to disturbances with reduced-order multi-layer safety-critical control (Molnar et al., 2021) to achieve provable safety guarantees in a parametric form amenable to SA-LineCoSpar. Finally, we deploy these two methods together as a design paradigm for a safety-critical controller on a quadrupedal robot in simulation and on hardware in laboratory and outdoor settings. Additionally, this work is the first time that PBL has been used to tune a CBF-based controller, and the first time these CBF methods have been combined.

2. Preference-Based Learning

Preference-Based Learning (PBL) provides an approach for searching complex parameter spaces via subjective feedback, eliminating the need to explicitly define an underlying reward function. This is particularly relevant for safety-critical systems, as quantifying the user-preferred trade-off between robustness and performance is generally difficult. Moreover, poorly defined reward functions often result in “reward hacking” (Amodei et al., 2016), in which undesirable actions achieve high rewards. The PBL approach taken in this work is motivated by the CoSpar framework (Tucker et al., 2020b), which can elicit preference feedback from a human expert and quickly converge to (near-)optimal behaviors. In this work we propose Safety-Aware LineCoSpar (SA-LineCoSpar), outlined in Alg. 1. This algorithm is a modification of the LineCoSpar algorithm (Tucker et al., 2020a), which iteratively selects actions, queries the user for subjective feedback, and updates its belief of the user’s underlying utility function via Bayesian inference.

Problem Setup: Let \mathbf{a} denote an action, such as a collection of l parameters used in a feedback controller, that takes values in a finite search space $A \subset \mathbb{R}^l$. We assume that each action $\mathbf{a} \in A$ has an unknown utility to the user, defined by a function $r(\mathbf{a}) : A \rightarrow \mathbb{R}$. These utilities can be written in a vectorized form $\mathbf{r}_A = [r(\mathbf{a}_1), \dots, r(\mathbf{a}_{|A|})]^\top \in \mathbb{R}^{|A|}$. In each iteration, $s \in \mathbb{N}$ actions are sampled from A . After these actions are executed, the user is queried for two forms of feedback: pairwise preferences and ordinal labels, in this work describing *performance* and *safety*, respectively. This feedback is collected into the combined dataset D .

Modeling the Utility Function: We model the unknown utilities \mathbf{r}_A as a Gaussian process with prior $\mathcal{N}(\mathbf{0}, \Sigma^{\text{pr}})$, where each element of the covariance matrix $\Sigma^{\text{pr}} \in \mathbb{S}_{>0}^{|A| \times |A|}$ is computed as $\Sigma_{ij}^{\text{pr}} = k(\mathbf{a}_i, \mathbf{a}_j)$ with a kernel function $k : A \times A \rightarrow \mathbb{R}$ and $\mathbf{a}_i \in A$ denoting the i^{th} action in A . We select k to be the squared exponential kernel, yielding a prior given by the multivariate Gaussian:

$$\mathcal{P}(\mathbf{r}_A) = \frac{1}{(2\pi)^{|A|/2} |\Sigma^{\text{pr}}|^{1/2}} \exp \left(-\frac{1}{2} \mathbf{r}_A^\top (\Sigma^{\text{pr}})^{-1} \mathbf{r}_A \right). \quad (1)$$

Given a dataset D , the posterior is proportional to the likelihood and the prior by Bayes’ theorem, i.e., $\mathcal{P}(\mathbf{r}_A \mid D) \propto \mathcal{P}(D \mid \mathbf{r}_A) \mathcal{P}(\mathbf{r}_A)$. We denote the maximum a posterior (MAP) estimate of the posterior by $\hat{\mathbf{r}}_A \in \mathbb{R}^{|A|}$, which is defined as $\hat{\mathbf{r}}_A \triangleq \operatorname{argmax}_{\mathbf{r}_A \in A} \mathcal{P}(\mathbf{r}_A \mid D)$, noting that $\hat{\mathbf{r}}_A$ is equivalent to the minimizer of $\mathcal{S}(\mathbf{r}_A) = -\ln(\mathcal{P}(D \mid \mathbf{r}_A)) + \frac{1}{2} \mathbf{r}_A^\top (\Sigma^{\text{pr}})^{-1} \mathbf{r}_A$. We approximate the posterior as a multivariate Gaussian centered at $\hat{\mathbf{r}}_A$ with the covariance $\Sigma_A \in \mathbb{S}_{>0}^{|A| \times |A|}$ defined as $\Sigma_A = (\frac{\partial^2 \mathcal{S}}{\partial \mathbf{r}_A^2}(\hat{\mathbf{r}}_A))^{-1}$. This is known as the Laplace approximation of the distribution $\mathcal{P}(\mathbf{r}_A \mid D)$, i.e., $\mathcal{P}(\mathbf{r}_A \mid D) \approx \mathcal{N}(\hat{\mathbf{r}}_A, \Sigma_A)$. For details, refer to Chu and Ghahramani (2005). Lastly, since this process is intractable for high-dimensional action spaces, we perform it over a subset $S \subset A$ of the action space, forming a partial characterization of the utilities denoted by $\mathcal{P}(\mathbf{r}_S \mid D) \approx \mathcal{N}(\hat{\mathbf{r}}_S, \Sigma_S)$, with $\mathbf{r}_S, \hat{\mathbf{r}}_S \in \mathbb{R}^{|S|}$.

Preference Likelihood Function: A pairwise preference is defined as a relation between two actions $\mathbf{a}_1, \mathbf{a}_2 \in A$, where $\mathbf{a}_1 \succ \mathbf{a}_2$ if action \mathbf{a}_1 is preferred to \mathbf{a}_2 . Since user preferences are expected to be corrupted by noise, we model individual pairwise preferences via a likelihood function:

$$\mathcal{P}(\mathbf{a}_1 \succ \mathbf{a}_2 \mid r(\mathbf{a}_1), r(\mathbf{a}_2)) = g_p \left(\frac{r(\mathbf{a}_1) - r(\mathbf{a}_2)}{c_p} \right) \quad (2)$$

where $g_p : \mathbb{R} \rightarrow [0, 1]$ is any monotonically-increasing link function, and $c_p \in \mathbb{R}_{>0}$ accounts for preference noise. We select g_p to be the sigmoid function, i.e., $g_p(x) = 1/(1 + e^{-x})$. Assuming

conditional independence, the likelihood function for a collection of $K \in \mathbb{N}$ preferences, D_p , can be modeled as the product of each individual preference likelihood:

$$\mathcal{P}(D_p | r(\mathbf{a}_{11}), r(\mathbf{a}_{12}), \dots, r(\mathbf{a}_{K2})) = \prod_{k=1}^K \mathcal{P}(\mathbf{a}_{k1} \succ \mathbf{a}_{k2} | r(\mathbf{a}_{k1}), r(\mathbf{a}_{k2})), \quad (3)$$

where $\mathbf{a}_{k1}, \mathbf{a}_{k2} \in A$ are the preferred and non-preferred actions, respectively, in the k^{th} preference.

Ordinal Likelihood Function: We partition the action space into “unsafe” and “safe” actions by leveraging the ordinal nature of these definitions (i.e., unsafe actions are always considered worse than safe actions). A user provides this feedback as an ordinal label, which assigns an action to a discrete ordered category such as “bad” and “good” (Chu et al., 2005). While ordinal labels can be generalized to any number of ordinal categories (c.f. Li et al. (2021)), we utilize just two categories to represent “unsafe” and “safe”. In this case, the action space is decomposed into two disjoint sets, $A = O_1 \cup O_2$, with $\mathbf{a} \in O_1$ if $r(\mathbf{a}) < \beta$ and $\mathbf{a} \in O_2$ if $r(\mathbf{a}) \geq \beta$, with the ordinal threshold $\beta \in \mathbb{R}$. As with preferences, we assume that ordinal label feedback is corrupted by noise and is modeled as:

$$\mathcal{P}(\mathbf{a} \in O_1 | r(\mathbf{a})) = g_o\left(\frac{\beta - r(\mathbf{a})}{c_o}\right), \quad \mathcal{P}(\mathbf{a} \in O_2 | r(\mathbf{a})) = 1 - g_o\left(\frac{\beta - r(\mathbf{a})}{c_o}\right), \quad (4)$$

where $g_o : \mathbb{R} \rightarrow [0, 1]$ is any monotonically-increasing link function and c_o quantifies the noise in the ordinal label feedback. Again, we select g_o to be the sigmoid function $g_o(x) = 1/(1 + e^{-x})$. Assuming conditional independence of ordinal label queries, the likelihood function for a collection of $M \in \mathbb{N}$ ordinal labels, D_o , can be modeled as the product of each individual ordinal likelihood:

$$\mathcal{P}(D_o | r(\mathbf{a}_1), \dots, r(\mathbf{a}_k)) = \prod_{k=1}^M \mathcal{P}(\mathbf{a}_k \in O_{o(k)} | r(\mathbf{a}_k)), \quad (5)$$

where $\mathbf{a}_k \in A$ refers to the action corresponding to the k^{th} ordinal label, $o(k) \in \{1, 2\}$. For our simulation and experiments, the hyperparameters c_p, c_o, β are determined in advance. Lastly, assuming conditional independence of the feedback mechanisms, the combined likelihood function is calculated as the product of the individual likelihoods, $\mathcal{P}(D | r) = \mathcal{P}(D_p | r)\mathcal{P}(D_o | r)$.

Sampling New Actions: In the first iteration ($i = 1$), $s \in \mathbb{N}$ actions are sampled randomly from A , recorded as the set of visited actions $V_1 = \{\mathbf{a}_1^{(1)}, \dots, \mathbf{a}_1^{(s)}\}$, executed on the system, and the preferences and ordinal labels are collected into a dataset D_1 . In each subsequent iteration ($i > 1$), s new actions are sampled using Thompson sampling, which is shown to have desirable regret minimization properties (Chapelle and Li, 2011). Ideally, Thompson sampling draws s samples from the posterior $\mathcal{P}(\mathbf{r}_A | D_{i-1})$, i.e. $\mathbf{r}^{(j)} \sim \mathcal{P}(\mathbf{r}_A | D_{i-1})$ for $j \in \{1, \dots, s\}$, and the action $\mathbf{a}_i^{(j)} \in A$ maximizing each $\mathbf{r}^{(j)}$ is selected to execute on the system. These sampled actions $\{\mathbf{a}_i^{(1)}, \dots, \mathbf{a}_i^{(s)}\}$ are concatenated with V_{i-1} to produce V_i , executed on the system, and the resulting preferences and ordinal labels are concatenated with D_{i-1} to produce D_i . However, since it is intractable to approximate $\mathcal{P}(\mathbf{r}_A | D)$ for high-dimensional action spaces, we utilize a dimensionality-reduction technique introduced in (Tucker et al., 2020a) that instead updates the posterior over a subset $S_i \subset A$. Motivated by (Kirschner et al., 2019), we construct the subset as $S_i = L_i \cup V_{i-1}$, where $L_i \subset A$ is the collection of $e \in \mathbb{N}$ actions in A closest to a randomly drawn line $\ell_i \subset \mathbb{R}^l$. This line is drawn to intersect with the believed best action, computed as $\hat{\mathbf{a}}_{i-1}^* = \arg\max_{\mathbf{a} \in V_{i-1}} \hat{\mathbf{r}}_{V_{i-1}}(\mathbf{a})$ where $\hat{\mathbf{r}}_{V_{i-1}}$ is the MAP estimate of the posterior $\mathcal{P}(\mathbf{r}_{V_{i-1}} | D_i)$. See Tucker et al. (2020a) for more details.

Safety-Aware Sampling: It is important to avoid unsafe actions during the sequential decision making process for certain applications, such as learning robotic controllers on hardware, where low-reward actions might lead to physical damage of the platform. Safe exploration algorithms such as (Sui et al., 2015, 2018; Berkenkamp et al., 2016), considered the setting where actions below some prespecified safety threshold are catastrophic and need to be avoided at all cost. In our work, since we construct controllers that account for safety, we adopt a more optimistic learning approach called *safety-aware*. In this case, actions labeled by a human as “unsafe” are not catastrophic but undesirable. Thus, the algorithm *avoids* these actions; whereas the safe exploration algorithms guarantee that no such actions are sampled and can be sometimes exceedingly conservative in settings like ours.

To achieve this safety-awareness, we leverage the approach introduced in (Li et al., 2021), which uses ordinal labels to identify a *region of interest* (ROI) in A . In this work, the ROI is defined to be the actions labeled as “safe”. In each iteration i we estimate an ROI within the set S_i as:

$$S_i^{\text{ROI}} = \{\mathbf{a} \in S_i \mid \hat{\mathbf{r}}_{S_i}(\mathbf{a}) + \lambda \sigma_{S_i}(\mathbf{a}) > \beta\}, \quad (6)$$

where $\hat{\mathbf{r}}_{S_i}(\mathbf{a})$ and $\sigma_{S_i}(\mathbf{a})$ denote the posterior mean and the posterior standard deviation, respectively, evaluated at the action $\mathbf{a} \in S_i$. The variable $\lambda \in \mathbb{R}$ determines how conservative the algorithm would be in estimating the safety region, as illustrated in Figure 2. We see that lower values of λ result in fewer unsafe actions being sampled, with only a slight effect on sample-efficiency. The restriction to S_i^{ROI} is added to LineCoSpar by only considering actions from within S_i^{ROI} during Thompson sampling. We refer to this version of the LineCoSpar algorithm as Safety-Aware LineCoSpar (SA-LineCoSpar), with the full algorithm outlined in Alg. 1.

Algorithm 1: Safety-Aware LineCoSpar

input: s uniform random actions ($V_1 \subset A$),
corresponding feedback (D_1),

for $i = 2, \dots, N$ **do**

 Update posterior over V_{i-1}

$\hat{\mathbf{a}}_{i-1}^* \leftarrow \operatorname{argmax}_{\mathbf{a} \in V_{i-1}} \hat{\mathbf{r}}_{V_{i-1}}(\mathbf{a})$

$L_i \leftarrow$ New linear subspace intersecting $\hat{\mathbf{a}}_{i-1}^*$

 Construct subspace $S_i = L_i \cup V_{i-1}$

 Update the model posterior over S_i

 Determine region of interest S_i^{ROI}

for $j = 1, \dots, s$ **do**

$r^{(j)} \sim \mathcal{N}(\hat{\mathbf{r}}_{S_i}, \Sigma_{S_i})$

$\mathbf{a}_i^{(j)} \leftarrow \operatorname{argmax}_{\mathbf{a} \in S_i^{\text{ROI}}} r^{(j)}$

end

 Deploy $\{\mathbf{a}_i^{(1)}, \dots, \mathbf{a}_i^{(s)}\}$ on system

$V_i \leftarrow V_{i-1} \cup \{\mathbf{a}_i^{(1)}, \dots, \mathbf{a}_i^{(s)}\}$

$D_i \leftarrow D_{i-1} \cup$ new prefs. \cup new ord. labels

end

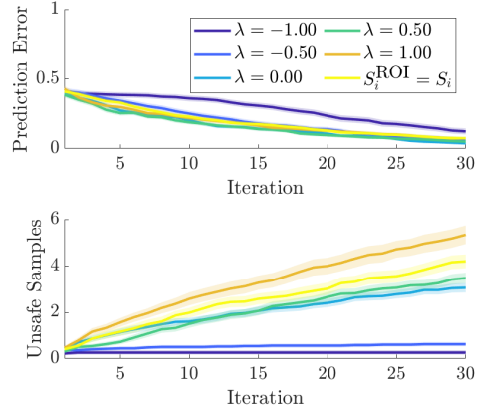


Figure 2. A comparison of SA-LineCoSpar and standard LineCoSpar on a synthetic utility function (drawn from the Gaussian prior) averaged over 50 runs with standard error shown by the shaded region. The safety-aware criteria reduces the number of sampled unsafe actions with a minimal effect on the prediction error, defined as $|\hat{\mathbf{a}}_i^* - \mathbf{a}^*|$ with $\hat{\mathbf{a}}_i^* \triangleq \operatorname{argmax}_{\mathbf{a}} \hat{\mathbf{r}}_{S_i}$ and $\mathbf{a}^* \triangleq \operatorname{argmax}_{\mathbf{a}} r(\mathbf{a})$.

3. Robust Safety-Critical Control

In this section, we formalize robust safety and discuss safe controller synthesis through the use of Control Barrier Functions (CBFs), that ultimately yield controllers whose parameters are to be

updated with SA-LineCoSpar. Consider the following nonlinear control-affine system:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})(\mathbf{v} + \mathbf{d}(t)), \quad (7)$$

with state $\mathbf{x} \in \mathbb{R}^n$, input $\mathbf{v} \in \mathbb{R}^m$, functions $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ assumed to be locally Lipschitz continuous on their domains, and piecewise continuous disturbance signal $\mathbf{d} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^m$ for which we define $\|\mathbf{d}\|_\infty \triangleq \sup_{t \geq 0} \|\mathbf{d}(t)\|$. Specifying the input via a controller $\mathbf{k} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ that is locally Lipschitz continuous on its domain yields the closed-loop system:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})(\mathbf{k}(\mathbf{x}) + \mathbf{d}(t)). \quad (8)$$

We assume for any initial condition $\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^n$ and disturbance \mathbf{d} , this system has a unique solution $\mathbf{x}(t)$ for all $t \in \mathbb{R}_{\geq 0}$. We consider this system safe if its state $\mathbf{x}(t)$ remains in a *safe set* $\mathcal{C} \subset \mathbb{R}^n$, defined as the 0-superlevel set of a continuously differentiable function $h : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$:

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}, \boldsymbol{\rho}) \geq 0\}, \quad (9)$$

where $\boldsymbol{\rho} \in \mathbb{R}^p$ are constant application-specific parameters. We say the set $\mathcal{C} \subset \mathbb{R}^n$ is *forward invariant* if for every $\mathbf{x}_0 \in \mathcal{C}$ the solution $\mathbf{x}(t)$ to (8) satisfies $\mathbf{x}(t) \in \mathcal{C}$ for all $t \geq 0$. The system (8) is *safe* with respect to \mathcal{C} if \mathcal{C} is forward invariant. Ensuring the safety of the set \mathcal{C} in the absence of disturbances can be achieved through *Control Barrier Functions (CBFs)*:

Definition 1 (Control Barrier Functions (CBF) (Ames et al., 2014)) *The function h is a Control Barrier Function (CBF) for (7) on \mathcal{C} if there exists $\alpha \in \mathcal{K}_\infty^e$ ¹ such that for all $\mathbf{x} \in \mathbb{R}^n$:*

$$\sup_{\mathbf{v} \in \mathbb{R}^m} \underbrace{\frac{\partial h}{\partial \mathbf{x}}(\mathbf{x}, \boldsymbol{\rho}) \mathbf{f}(\mathbf{x})}_{L_f h(\mathbf{x}, \boldsymbol{\rho})} + \underbrace{\frac{\partial h}{\partial \mathbf{x}}(\mathbf{x}, \boldsymbol{\rho}) \mathbf{g}(\mathbf{x}) \mathbf{v}}_{L_g h(\mathbf{x}, \boldsymbol{\rho})} > -\alpha(h(\mathbf{x}, \boldsymbol{\rho})). \quad (10)$$

While it may be possible to synthesize controllers that render a given set \mathcal{C} safe in the presence of disturbances (Jankovic, 2018), this may result in overly-conservative behavior. Instead, we consider how safety properties degrade with disturbances via the following definition.

Definition 2 (Input-to-State Safety (Kolathaya and Ames, 2018)) *The system (8) is Input-to-State Safe (ISSf) with respect to \mathcal{C} if there exists $\gamma \in \mathcal{K}_\infty$ such that for all $\delta \in \mathbb{R}_{\geq 0}$ and disturbances $\mathbf{d} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^m$ satisfying $\|\mathbf{d}\|_\infty \leq \delta$, the set $\mathcal{C}_\delta \subset \mathbb{R}^n$ defined as:*

$$\mathcal{C}_\delta = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}, \boldsymbol{\rho}) \geq -\gamma(\delta)\} \quad (11)$$

is forward invariant. The function h is an Input-to-State Safe Control Barrier Function (ISSf-CBF) for (7) on \mathcal{C} with parameter $\varphi \in \mathbb{R}_{\geq 0}$ if there exists $\alpha \in \mathcal{K}_\infty^e$ such that for all $\mathbf{x} \in \mathbb{R}^n$:

$$\sup_{\mathbf{v} \in \mathbb{R}^m} L_f h(\mathbf{x}, \boldsymbol{\rho}) + L_g h(\mathbf{x}, \boldsymbol{\rho}) \mathbf{v} - \varphi \|L_g h(\mathbf{x}, \boldsymbol{\rho})\|^2 > -\alpha(h(\mathbf{x}, \boldsymbol{\rho})). \quad (12)$$

1. We say that a continuous function $\alpha : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is *class \mathcal{K}_∞* ($\alpha \in \mathcal{K}_\infty$) if $\alpha(0) = 0$, α is strictly monotonically increasing, and $\lim_{r \rightarrow \infty} \alpha(r) = \infty$. We say that a continuous function $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ is *class \mathcal{K}_∞^e* ($\alpha \in \mathcal{K}_\infty^e$) if $\alpha(0) = 0$, α is strictly monotonically increasing, $\lim_{r \rightarrow \infty} \alpha(r) = \infty$, and $\lim_{r \rightarrow -\infty} \alpha(r) = -\infty$.

The parameter $\rho \in \mathbb{R}^p$ contains information about the system's environment that designates a given state as safe or unsafe, such as the location and size of obstacles. In highly unstructured environments the system may need to estimate ρ , denoted by $\hat{\rho} \in \mathbb{R}^p$ from measurements of the environment, such as camera data. The process of converting complex measurements into the representation $\hat{\rho}$ is often imperfect, leading to error between the estimator $\hat{\rho}$ and the true value of ρ . In this setting, safety can be achieved via *Measurement-Robust Control Barrier Functions (MR-CBFs)*:

Definition 3 (Measurement-Robust Control Barrier Functions (Dean et al., 2020)) *The function h is a Measurement-Robust Control Barrier Function (MR-CBF) for (7) on \mathcal{C} with parameters $a, b \in \mathbb{R}_{\geq 0}$ if there exists $\alpha \in \mathcal{K}_{\infty}^e$ such that for all $\hat{\rho} \in \mathbb{R}^p$ and $\mathbf{x} \in \mathbb{R}^n$:*

$$\sup_{\mathbf{v} \in \mathbb{R}^m} L_{\mathbf{f}}h(\mathbf{x}, \hat{\rho}) + L_{\mathbf{g}}h(\mathbf{x}, \hat{\rho})\mathbf{v} - a - b\|\mathbf{v}\| > -\alpha(h(\mathbf{x}, \hat{\rho})). \quad (13)$$

The following theorem summarizes the safety results achieved with these various types of CBFs:

Theorem 1 *Consider the set \mathcal{C} defined in (9).*

1. *If h is a CBF for (7) on \mathcal{C} , $\mathbf{d}(t) = \mathbf{0}$ for $t \in \mathbb{R}_{\geq 0}$ and $\hat{\rho} = \rho$, then there exists a controller \mathbf{k} such that (8) is safe with respect to \mathcal{C} .*
2. *If h is an ISSf-CBF for (7) on \mathcal{C} with parameter φ and $\hat{\rho} = \rho$, then there exists a controller \mathbf{k} such that (8) is ISSf with respect to \mathcal{C} with $\gamma(\delta) = -\alpha^{-1}(-\delta^2/4\varphi)$ where $\alpha^{-1} \in \mathcal{K}_{\infty}^e$.*
3. *Assume $L_{\mathbf{f}}h$, $L_{\mathbf{g}}h$, and $\alpha \circ h$ are Lipschitz continuous on their domains, and assume that $\|\hat{\rho} - \rho\| \leq \epsilon$ for some $\epsilon \in \mathbb{R}_{\geq 0}$. Then there exists $\underline{a}, \underline{b} \in \mathbb{R}_{\geq 0}$ such that if h is an MR-CBF for (7) on \mathcal{C} with parameters $a, b \in \mathbb{R}_{\geq 0}$ satisfying $a \geq \underline{a}$ and $b \geq \underline{b}$, and $\mathbf{d}(t) = \mathbf{0}$ for $t \in \mathbb{R}_{\geq 0}$, then there exists a controller \mathbf{k} such that (8) is safe with respect to \mathcal{C} .*

4. Preference-Based Learning for Safety-Critical Control

In this section we propose a design paradigm that leverages SA-LineCoSpar (1) to select parameters for a CBF controller that is both performant and safe.

Multi-Layered System Dynamics: Many real-life engineering systems have high-dimensional state spaces and complex dynamics. Hence control systems are often designed as a set of interconnected subsystems, such as a low-dimensional subsystem that provides reference signals capturing safe behavior and a high-dimensional subsystem that tracks these reference signals. In particular, consider the following cascaded nonlinear control-affine system resulting as a modification of (7):

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\kappa(\boldsymbol{\xi}), \quad \dot{\boldsymbol{\xi}} = \mathbf{f}_{\boldsymbol{\xi}}(\mathbf{x}, \boldsymbol{\xi}) + \mathbf{g}_{\boldsymbol{\xi}}(\mathbf{x}, \boldsymbol{\xi})\mathbf{u}, \quad (14)$$

with additional states $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$, control input $\mathbf{u} \in \mathbb{R}^{m_{\boldsymbol{\xi}}}$ and functions $\kappa : \mathbb{R}^{n_{\boldsymbol{\xi}}} \rightarrow \mathbb{R}^m$, $\mathbf{f}_{\boldsymbol{\xi}} : \mathbb{R}^n \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \rightarrow \mathbb{R}^{n_{\boldsymbol{\xi}}}$, and $\mathbf{g}_{\boldsymbol{\xi}} : \mathbb{R}^n \times \mathbb{R}^{n_{\boldsymbol{\xi}}} \rightarrow \mathbb{R}^{n_{\boldsymbol{\xi}} \times m_{\boldsymbol{\xi}}}$ assumed to be locally Lipschitz continuous on their domains. We note that the input \mathbf{v} from (7) was replaced by the term $\kappa(\boldsymbol{\xi})$. These dynamics may represent systems described by the Euler-Lagrange equations, such as mobile robots, where \mathbf{x} reflects floating-base position, $\boldsymbol{\xi}$ captures both floating-base velocities and joint positions and velocities, and the input \mathbf{u} represents the forces and torques applied to the system.

Given this cascaded system, we utilize the low-dimensional subsystem to ensure that \mathcal{C} is ISSf by making two assumptions. First, we assume the safe set \mathcal{C} can be described as in (9), such that it

only depends on the states \mathbf{x} and parameters $\boldsymbol{\rho}$, and does not explicitly depend on the states $\boldsymbol{\xi}$. For example, in the context of a mobile robotic system, this assumption is justified if safety is described as keeping the floating base position of the robot away from obstacles. Second, we assume that there exists a controller $\boldsymbol{\pi} : \mathbb{R}^n \times \mathbb{R}^{n_\xi} \times \mathbb{R}^m \rightarrow \mathbb{R}^{m_\xi}$ and $\mu \in \mathbb{R}_{\geq 0}$ such that for any continuous, bounded signal $\mathbf{s} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^m$, the closed-loop system:

$$\dot{\boldsymbol{\xi}} = \mathbf{f}_\xi(\mathbf{x}, \boldsymbol{\xi}) + \mathbf{g}_\xi(\mathbf{x}, \boldsymbol{\xi})\boldsymbol{\pi}(\mathbf{x}, \boldsymbol{\xi}, \mathbf{s}(t)) \quad (15)$$

satisfies the following implication:

$$\|\boldsymbol{\kappa}(\boldsymbol{\xi}(0)) - \mathbf{s}(0)\| \leq \mu \implies \|\boldsymbol{\kappa}(\boldsymbol{\xi}(t)) - \mathbf{s}(t)\| \leq \mu, \quad t \in \mathbb{R}_{\geq 0}. \quad (16)$$

This assumption reflects that a separate controller may be designed for the high-dimensional dynamics to track well-behaved reference signals synthesized via the low-dimensional model. In particular, if a continuous controller $\mathbf{k} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is designed for the low-dimensional system (7) and $\|\boldsymbol{\kappa}(\boldsymbol{\xi}(0)) - \mathbf{k}(\mathbf{x}(0))\| \leq \mu$, then we have that the controller $\boldsymbol{\pi}$ ensures $\|\boldsymbol{\kappa}(\boldsymbol{\xi}(t)) - \mathbf{k}(\mathbf{x}(t))\| \leq \mu$ for $t \in \mathbb{R}_{\geq 0}$. With this assumption in mind, we may study the ISSf behavior of the closed-loop system:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})(\mathbf{k}(\mathbf{x}) + \mathbf{d}(t)), \quad \dot{\boldsymbol{\xi}} = \mathbf{f}_\xi(\mathbf{x}, \boldsymbol{\xi}) + \mathbf{g}_\xi(\mathbf{x}, \boldsymbol{\xi})\boldsymbol{\pi}(\mathbf{x}, \boldsymbol{\xi}, \mathbf{k}(\mathbf{x})), \quad (17)$$

with the disturbance defined as $\mathbf{d}(t) = \boldsymbol{\kappa}(\boldsymbol{\xi}(t)) - \mathbf{k}(\mathbf{x}(t))$ satisfying $\|\mathbf{d}\|_\infty \leq \mu$.

Combined Robust CBFs for PBL: We now combine the robustness properties of MR- and ISSf-CBFs to account for measurement uncertainty and the disturbance, \mathbf{d} , allowing us to make robust safety guarantees for the full system (17). This is formalized in the following theorem:

Theorem 2 *Given the set \mathcal{C} defined in (9), suppose the functions $L_{\mathbf{f}}h$, $L_{\mathbf{g}}h$, $\|L_{\mathbf{g}}h\|^2$, and $\alpha \circ h$ are Lipschitz continuous on their domains, and assume that $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\| \leq \epsilon$ for some $\epsilon \in \mathbb{R}_{\geq 0}$. Then there exists $\underline{a}, \underline{b} \in \mathbb{R}_{\geq 0}$ such that if h satisfies:*

$$\sup_{\mathbf{v} \in \mathbb{R}^m} L_{\mathbf{f}}h(\mathbf{x}, \hat{\boldsymbol{\rho}}) + L_{\mathbf{g}}h(\mathbf{x}, \hat{\boldsymbol{\rho}})\mathbf{v} - \varphi\|L_{\mathbf{g}}h(\mathbf{x}, \hat{\boldsymbol{\rho}})\|^2 - a - b\|\mathbf{v}\| > -\alpha(h(\mathbf{x}, \hat{\boldsymbol{\rho}})), \quad (18)$$

for all $\mathbf{x} \in \mathbb{R}^n$ and some $a, b \in \mathbb{R}_{\geq 0}$ satisfying $a \geq \underline{a}$ and $b \geq \underline{b}$, then there exists a controller $\mathbf{k} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that (17) is ISSf with respect to \mathcal{C} with $\gamma(\delta) = -\alpha^{-1}(-\delta^2/(4\varphi))$ where $\alpha^{-1} \in \mathcal{K}_\infty^e$.

In this work we refer to CBFs that satisfy (18) as Tunable Robustified CBFs (TR-CBFs) where the tunable parameters are α, φ, a , and b . The proof of this theorem can be found in Appendix A. Constraint (18) can now be incorporated into a safety filter on a locally Lipschitz continuous nominal controller $\mathbf{k}_{\text{nom}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ to ensure the safe behavior of the system:

$$\begin{aligned} \mathbf{k}(\mathbf{x}) &= \arg\min_{\mathbf{v} \in \mathbb{R}^m} \|\mathbf{v} - \mathbf{k}_{\text{nom}}(\mathbf{x})\|^2 & (\text{TR-CBF-OP}) \\ \text{s.t. } & L_{\mathbf{f}}h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i) + L_{\mathbf{g}}h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i)\mathbf{v} - \varphi\|L_{\mathbf{g}}h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i)\|^2 - a - b\|\mathbf{v}\| \geq -\alpha h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i), \\ & \forall i \in \{1, \dots, N_o\}. \end{aligned}$$

Here a linear class \mathcal{K}_∞^e function is chosen with coefficient $\alpha \in \mathbb{R}_{>0}$. In the case where we wish to enforce multiple safety conditions, such as in obstacle avoidance with several obstacles, $\hat{\boldsymbol{\rho}}_i$ can be used to indicate the measured parameters of obstacle i with $N_o \in \mathbb{N}$ being the total number of

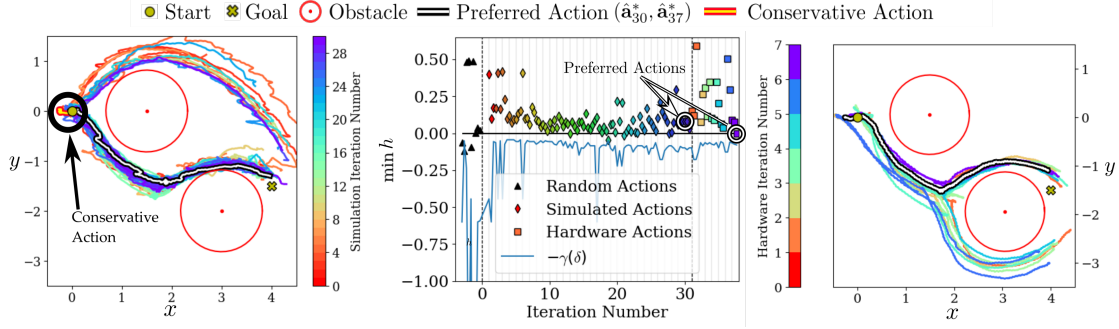


Figure 3. (Left) 30 samples of 3 actions generated via simulation. The preferred action $\hat{\mathbf{a}}_{30} = (3, 0.6, 0.5, 0.015)$, is shown in black and white. A conservative action $\mathbf{a} = (2, 0.5, 0.0651, 0.485)$ is shown in red and yellow, where a and b were determined by estimating the Lipschitz coefficients present in the proof of Theorem 2. The conservative action fails to progress whereas SA-LineCoSpar provides an action which successfully navigates between obstacles. **(Center)** The minimum value of h that occurred in each iteration. Triangles, diamonds, and squares represent actions that are randomly sampled, simulated and tested on hardware in the laboratory, respectively, and colors correlate to iteration number. The lower bound, $-\gamma(\delta)$, for the expanded set \mathcal{C}_δ is plotted for $\delta = 1$ m/s. The preferred actions for simulation and hardware experiments are circled. **(Right)** 7 additional samples of 3 actions generated on hardware in a laboratory setting. The preferred action $\hat{\mathbf{a}}_{37}^* = (4, 0.6, 0.4, 0)$ successfully traverses between the obstacles.

obstacles we wish to consider. Enforcing this constraint for $N_o > 1$ can be viewed as Boolean composition of safe sets (Glotfelter et al., 2018). Additionally, this safety filter is a Second-Order Cone Program (SOCP) (Boyd and Vandenberghe, 2004) for which an array of solvers exist including ECOS (Domahidi et al., 2013).

Integrating Learning to Tune the Control Barrier Function: The parameter selection process of TR-CBF-OP is particularly important since the parameters \underline{a} and \underline{b} guaranteed to exist by Theorem 2 are worst-case approximations of the uncertainty generated using Lipschitz constants. Such approximations often lead to undesired conservatism and may render the system incapable of performing its goal (as seen in Figure (3)). Thus, we propose a unified design, as illustrated in Figure 1, where we utilize SA-LineCoSpar to identify user-preferred parameters of TR-CBF-OP that relax this worst-case approximation to experimentally realize performant and safe behavior.

Note that this design paradigm relies on the tunable construction of TR-CBF-OP where we define the actions to be $\mathbf{a} = (\alpha, \varphi, a, b)$. Also, the construction of TR-CBF-OP assures that unsafe actions are not catastrophic, since for any $\alpha, \varphi, a, b > 0$, it provides robustness to some amount of disturbances and measurement error. This assurance allows us to utilize a safety-aware approach as opposed to more conservative notions of learning.

5. Experimental Results

We applied the proposed design paradigm to an obstacle avoidance task with a Unitree A1 quadrupedal robot (Figure 1) in simulation and on hardware for both indoor and outdoor environments (video: vid). The action space \mathcal{A} and hyperparameters of PBL are defined in Table 1. We used the unicycle

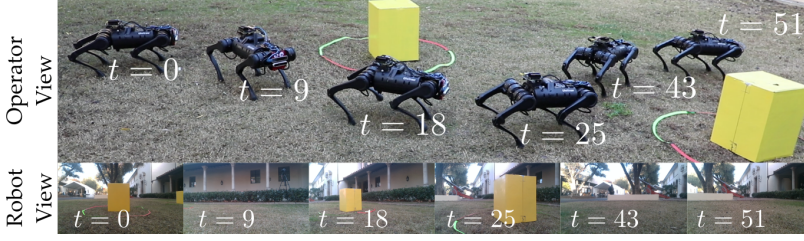


Figure 4. The preferred action $\hat{\mathbf{a}}_{40} = (5, 0.1, 0.4, 0.02)$ after simulation, indoor experiments, and 3 additional iterations of 3 actions in an outdoor environment is shown alongside views from the onboard camera.

| hyperparameter | | value | |
|----------------|-----|-------|----------|
| λ | | -0.5 | |
| β | | 0 | |
| name | min | max | Δ |
| α | 0.5 | 5 | 0.5 |
| φ | 0 | 1 | 0.1 |
| a | 0 | 1 | 0.1 |
| b | 0 | 0.05 | 0.005 |

Table 1. The hyperparameters and action space A for the experiments where Δ is the discretization step.

model as our simplified model (7) with the nominal controller \mathbf{k}_{nom} :

$$\underbrace{\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\psi} \end{bmatrix}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{f}(\mathbf{x})} + \underbrace{\begin{bmatrix} \cos \psi & 0 \\ \sin \psi & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{g}(\mathbf{x})} \left(\underbrace{\begin{bmatrix} v \\ \omega \end{bmatrix}}_{\mathbf{v}} + \mathbf{d}(t) \right), \quad \mathbf{k}_{\text{nom}}(\mathbf{x}) = \begin{bmatrix} K_v d_g + C \\ -K_\omega (\sin \psi - (y_g - y)/d_g) \end{bmatrix}, \quad (19)$$

where (x, y) is the planar position of the robot, ψ is the yaw angle, (x_g, y_g) is the goal position of the robot, $d_g = \|(x_g - x, y_g - y)\|$ is the distance to the goal, and K_v, K_ω , and C are positive constants. Obstacle avoidance is encoded via the 0-superlevel set of the function:

$$h(\mathbf{x}, \boldsymbol{\rho}_i) = d_{\text{obs},i} - r - \zeta \cos(\psi - \theta_i), \quad (20)$$

where $\boldsymbol{\rho}_i = [x_{\text{obs},i}, y_{\text{obs},i}]$ is the location of the i^{th} obstacle, $d_{\text{obs},i} = \|(x_{\text{obs},i} - x, y_{\text{obs},i} - y)\|$ and $\theta_i = \arctan((y_{\text{obs},i} - y)/(x_{\text{obs},i} - x))$ are the distance and angle from the i^{th} obstacle, r_{obs} is the sum of the radii of the obstacle and robot, and $\zeta > 0$ determines the effect of the heading angle on safety. The controller used to drive the system is the **TR-CBF-OP** with the nominal controller \mathbf{k}_{nom} from (19). In practice, infeasibilities of this safety filter were considered unsafe and the inputs were saturated such that $v \in [-0.2, 0.3]$ m/s and $\omega \in [-0.4, 0.4]$ rad/s. The velocity command \mathbf{v} is computed at 20 Hz and error introduced by this sampling scheme is captured by the tracking error $\mathbf{d}(t)$. Tracking of \mathbf{v} is performed by an inverse dynamics quadratic program (ID-QP) walking controller designed using the concepts in Buchli et al. (2009), which realizes a stable walking gait for (17) at 1 kHz.

Simulation results: We simulated the quadruped executing the proposed controller with parameters provided by SA-LineCoSpar, with the resulting trajectories, and the position of the obstacles shown in Figure 3. We ran 30 iterations, with 3 new actions sampled in each iteration ($s = 3$), and obtained user preferences and ordinal labels in between each set of actions. To simulate perception error, the measurements of the obstacles were shifted by -0.1 m in the y -direction. The parameters found with SA-LineCoSpar allows the robot to navigate between obstacles. For comparison, a conservative action is also shown, which is safe but fails to progress towards the goal. SA-LineCoSpar eliminates this conservatism with only minor safety violations and determines a parameter set which is both safe and performant.

Hardware results: After simulation, we continued learning on hardware experiments in a laboratory setting for 7 additional iterations until the user was satisfied with the experimental behavior.

The robot and obstacle positions were estimated using Intel Realense T265 and D415 cameras to perform SLAM and segmentation. Centroids of segmented clusters in the occupancy map were used as the measured obstacle positions $\hat{\rho}_i$. The true robot and obstacle positions were obtained for comparison using an OptiTrack motion capture system. The results of these experiments can be seen in Figure 3. Afterwards, 3 additional iterations were conducted outdoors on grass until again the user was satisfied with the experimental behavior. The resulting best trajectory can be seen in Figure 4. The preferred action was also tested on a variety of other obstacle arrangements to confirm its generalizability. The performance of the final preferred action for these obstacle configurations can be seen in the supplementary video: [vid](#).

6. Conclusion

In this work we proposed a design paradigm for control systems in which the robust safety requirements of a provably safe, but conservative controller are relaxed and controller parameters are instead chosen using a Preference-Based Learning algorithm called SA-LineCoSpar. Using our algorithm, we were able to learn a set of parameters that leads to user-preferred balance between safety and robustness on a quadrupedal robot platform. Future work includes applying this work to other platforms such as bipedal robots and including more complicated environments like obstacles with time-varying parameters.

References

Supplementary video, <https://vimeo.com/649746749>.

- A. Alan, A. J. Taylor, C. R. He, G. Orosz, and A. D. Ames. Safe controller synthesis with tunable input-to-state safe control barrier functions. *Control Systems Letters*, 6:908–913, 2022.
- A. Ames, J. Grizzle, and P. Tabuada. Control barrier function based quadratic programs with application to adaptive cruise control. In *Conference on Decision & Control (CDC)*, pages 6271–6278. IEEE, 2014.
- A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada. Control barrier functions: Theory and applications. In *European Control Conference (ECC)*, pages 3420–3431. IEEE, 2019.
- D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Man. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*, 2016.
- F. Berkenkamp, A. P. Schoellig, and A. Krause. Safe controller optimization for quadrotors with Gaussian processes. In *International Conference on Robotics and Automation (ICRA)*, pages 491–496. IEEE, 2016.
- E. Bıyık, N. Huynh, M. J. Kochenderfer, and D. Sadigh. Active preference-based Gaussian process regression for reward learning. *arXiv preprint arXiv:2005.02575*, 2020.
- S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- J. Buchli, M. Kalakrishnan, M. Mistry, P. Pastor, and S. Schaal. Compliant quadruped locomotion over rough terrain. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 814–820. IEEE, 2009.
- F. Castañeda, J. J. Choi, B. Zhang, C. J. Tomlin, and K. Sreenath. Gaussian process-based min-norm stabilizing controller for control-affine systems with uncertain input effects. *arXiv preprint arXiv:2011.07183*, 2020.
- O. Chapelle and L. Li. An empirical evaluation of Thompson sampling. *Advances in Neural Information Processing Systems*, 24:2249–2257, 2011.
- J. J. Choi, D. Lee, K. Sreenath, C. J. Tomlin, and S. L. Herbert. Robust control barrier-value functions for safety-critical control. *arXiv preprint arXiv:2104.02808*, 2021.
- W. Chu and Z. Ghahramani. Preference learning with Gaussian processes. In *International Conference on Machine Learning (ICML)*, pages 137–144, 2005.
- W. Chu, Z. Ghahramani, and C. K. I. Williams. Gaussian processes for ordinal regression. *Journal of Machine Learning Research*, 6(7), 2005.
- A. Clark. Control barrier functions for complete and incomplete information stochastic systems. In *American Control Conference (ACC)*, pages 2928–2935. IEEE, 2019.

- R. K. Cosner, A. W. Singletary, A. J. Taylor, T. G. Molnar, K. L. Bouman, and A. D. Ames. Measurement-robust control barrier functions: Certainty in safety with uncertainty in state. *arXiv preprint arXiv:2104.14030*, 2021.
- M. De Gemmis, L. Iaquinta, P. Lops, C. Musto, F. Narducci, and G. Semeraro. Preference learning in recommender systems. *Preference Learning*, 41:41–55, 2009.
- S. Dean, A. J. Taylor, R. K. Cosner, B. Recht, and A. D. Ames. Guaranteeing safety of learned perception modules via measurement-robust control barrier functions. In *Conference on Robotics Learning (CoRL)*, 2020.
- A. Domahidi, E. Chu, and S. Boyd. ECOS: An SOCP solver for embedded systems. In *European Control Conference (ECC)*, pages 3071–3076. IEEE, 2013.
- P. Glotfelter, J. Cortés, and M. Egerstedt. Boolean composability of constraints and control synthesis for multi-robot systems via nonsmooth control barrier functions. In *Conference on Control Technology and Applications (CCTA)*, pages 897–902. IEEE, 2018.
- K. Hobbs, M. Mote, M. Abate, S. Coogan, and E. Feron. Run time assurance for safety-critical systems: An introduction to safety filtering approaches for complex control systems. *arXiv preprint arXiv:2110.03506*, 2021.
- A. Jain, S. Sharma, T. Joachims, and A. Saxena. Learning preferences for manipulation tasks from online coactive feedback. *The International Journal of Robotics Research*, 34(10):1296–1313, 2015.
- M. Jankovic. Robust control barrier functions for constrained stabilization of nonlinear systems. *Automatica*, 96:359–367, 2018.
- J. Kirschner, M. Mutny, N. Hiller, R. Ischebeck, and A. Krause. Adaptive and safe Bayesian optimization in high dimensions via one-dimensional subspaces. In *International Conference on Machine Learning*, pages 3429–3438. PMLR, 2019.
- S. Kolathaya and A. D. Ames. Input-to-state safety with control barrier functions. *Control Systems Letters*, 3(1):108–113, 2018.
- K. Li, M. Tucker, E. Bıyık, E. Novoseller, J. W. Burdick, Y. Sui, D. Sadigh, Y. Yue, and A. D. Ames. Roial: Region of interest active learning for characterizing exoskeleton gait preference landscapes. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3212–3218. IEEE, 2021.
- W. Ma, S. Kolathaya, E. R. Ambrose, C. M. Hubicki, and A. D. Ames. Bipedal robotic running with DURUS-2D: Bridging the gap between theory and experiment. In *International Conference on Hybrid Systems: Computation and Control (HSCC)*, pages 265–274. ACM, 2017.
- T. G. Molnar, R. K. Cosner, A. W. Singletary, W. Ubellacker, and A. D. Ames. Model-free safety-critical control for robotic systems. *arXiv preprint arXiv:2109.09047*, 2021.
- K. Raman, T. Joachims, P. Shivaswamy, and T. Schnabel. Stable coactive learning via perturbation. In *International conference on machine learning*, pages 837–845. PMLR, 2013.

- D. Sadigh, A. D. Dragan, S. Sastry, and S. A. Seshia. Active preference-based learning of reward functions. In *Robotics: Science and Systems*, 2017.
- C. Santoyo, M. Dutreix, and S. Coogan. Verification and control for finite-time safety of stochastic systems via barrier functions. In *2019 IEEE Conference on Control Technology and Applications (CCTA)*, pages 712–717. IEEE, 2019.
- P. Shivaswamy and T. Joachims. Online structured prediction via coactive learning. *arXiv preprint arXiv:1205.4213*, 2012.
- A. Singletary, K. Klingebiel, J. R. Bourne, N. A. Browning, P. Tokumaru, and A. D. Ames. Comparative analysis of control barrier functions and artificial potential fields for obstacle avoidance. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.
- Y. Sui, A. Gotovos, J. Burdick, and A. Krause. Safe exploration for optimization with Gaussian processes. In *International Conference on Machine Learning (ICML)*, pages 997–1005, 2015.
- Y. Sui, V. Zhuang, J. W. Burdick, and Y. Yue. Stagewise safe Bayesian optimization with Gaussian processes. In *International Conference on Machine Learning (ICML)*, 2018.
- R. Takano and M. Yamakita. Robust constrained stabilization control using control Lyapunov and control barrier function in the presence of measurement noises. In *Conference on Control Technology and Applications (CCTA)*, pages 300–305. IEEE, 2018.
- A. J. Taylor and A. D. Ames. Adaptive safety with control barrier functions. In *American Control Conference (ACC)*, pages 1399–1405. IEEE, 2020.
- A. J. Taylor, A. Singletary, Y. Yue, and A. D. Ames. Learning for safety-critical control with control barrier functions. *Proceedings of Machine Learning Research (PMLR)*, 120:708–717, 2020.
- M. Tucker, M. Cheng, E. Novoseller, R. Cheng, Y. Yue, J. W. Burdick, and A. D. Ames. Human preference-based learning for high-dimensional optimization of exoskeleton walking gaits. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3423–3430. IEEE, 2020a.
- M. Tucker, E. Novoseller, C. Kann, Y. Sui, Y. Yue, J. W. Burdick, and A. D. Ames. Preference-based learning for exoskeleton gait optimization. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2351–2357. IEEE, 2020b.
- M. Tucker, N. Csomay-Shanklin, W.-L. Ma, and A. D. Ames. Preference-based learning for user-guided HZD gait generation on bipedal walking robots. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2804–2810. IEEE, 2021.
- L. Wang, E. A. Theodorou, and M. Egerstedt. Safe learning of quadrotor dynamics using barrier certificates. In *International Conference on Robotics and Automation (ICRA)*, pages 2460–2465. IEEE, 2018.
- Y. Yue, J. Broder, R. Kleinberg, and T. Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

Appendix A. Proof of Theorem 2

In this appendix we present the proof of Theorem 2 which establishes the input-to-state safety of \mathcal{C} in the context of disturbances and measurement uncertainties. Recall that Theorem 2 is given as:

Theorem 2 *Given the set \mathcal{C} defined in (9), suppose the functions $L_f h$, $L_g h$, $\|L_g h\|^2$, and $\alpha \circ h$ are Lipschitz continuous on their domains, and assume that $\|\hat{\rho} - \rho\| \leq \epsilon$ for some $\epsilon \in \mathbb{R}_{\geq 0}$. Then there exists $\underline{a}, \underline{b} \in \mathbb{R}_{\geq 0}$ such that if h satisfies:*

$$\sup_{\mathbf{v} \in \mathbb{R}^m} L_f h(\mathbf{x}, \hat{\rho}) + L_g h(\mathbf{x}, \hat{\rho})\mathbf{v} - \varphi \|L_g h(\mathbf{x}, \hat{\rho})\|^2 - a - b\|\mathbf{v}\| > -\alpha(h(\mathbf{x}, \hat{\rho})), \quad (21)$$

for all $\mathbf{x} \in \mathbb{R}^n$ and some $a, b \in \mathbb{R}_{\geq 0}$ satisfying $a \geq \underline{a}$ and $b \geq \underline{b}$, then there exists a controller $\mathbf{k} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that (17) is ISSf with respect to \mathcal{C} with $\gamma(\delta) = -\alpha^{-1}(-\delta^2/(4\varphi))$ where $\alpha^{-1} \in \mathcal{K}_{\infty}^e$.

Proof. First, we show that satisfying (21) for a particular set of $\underline{a}, \underline{b}$ implies satisfaction of (12). For this we choose:

$$\underline{a} = \epsilon(\mathcal{L}_{L_f h} + \mathcal{L}_{\alpha \circ h} + \mathcal{L}_{\varphi \|L_g h\|^2}), \quad \underline{b} = \epsilon \mathcal{L}_{L_g h}, \quad (22)$$

where \mathcal{L} indicates the Lipschitz coefficient of the subscripted function with respect to argument ρ . Let us define the function $c : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \rightarrow \mathbb{R}$ such that:

$$c(\mathbf{x}, \rho, \mathbf{v}) \triangleq L_f h(\mathbf{x}, \rho) + L_g h(\mathbf{x}, \rho)\mathbf{v} - \varphi \|L_g h(\mathbf{x}, \rho)\|^2 + \alpha(h(\mathbf{x}, \rho)). \quad (23)$$

Using this definition we have for all $\mathbf{v} \in \mathbb{R}^m$ that:

$$c(\mathbf{x}, \rho, \mathbf{v}) = c(\mathbf{x}, \hat{\rho}, \mathbf{v}) + c(\mathbf{x}, \rho, \mathbf{v}) - c(\mathbf{x}, \hat{\rho}, \mathbf{v}), \quad (24)$$

$$\geq c(\mathbf{x}, \hat{\rho}, \mathbf{v}) - \underbrace{\epsilon(\mathcal{L}_{L_f h} + \mathcal{L}_{\alpha \circ h} + \mathcal{L}_{\varphi \|L_g h\|^2})}_{\underline{a}} - \underbrace{\epsilon \mathcal{L}_{L_g h} \|\mathbf{v}\|}_{\underline{b}} \quad (25)$$

$$\geq c(\mathbf{x}, \hat{\rho}, \mathbf{v}) - a - b\|\mathbf{v}\|. \quad (26)$$

Above we added zero in (24) and used the Lipschitz coefficients and the worst case uncertainty ϵ to achieve the bound in (25). Since $\sup_{\mathbf{v} \in \mathbb{R}^m} c(\mathbf{x}, \hat{\rho}, \mathbf{v}) - a - b\|\mathbf{v}\| > 0$ holds based on (21), inequality (26) implies that (12) holds for the true parameters, ρ . Since (12) holds, the conditions of Theorem 1 point 2 are satisfied and thus \mathcal{C} is ISSf with $\gamma(\delta) = -\alpha^{-1}(-\delta^2/(4\varphi))$ where $\alpha^{-1} \in \mathcal{K}_{\infty}^e$. \square