

Optimal Parameter Adaptation for Safety-Critical Control via Safe Barrier Bayesian Optimization

Shengbo Wang, Ke Li, *Senior Member, IEEE*, Zheng Yan, Zhenyuan Guo, Song Zhu, Guanghui Wen, *Senior Member, IEEE*, and Shiping Wen, *Senior Member, IEEE*

Abstract—Safety is of paramount importance in control systems to avoid costly risks and catastrophic damages. The control barrier function (CBF) method, a promising solution for safety-critical control, poses a new challenge of enhancing control performance due to its direct modification of original control design and the introduction of uncalibrated parameters. In this work, we shed light on the crucial role of configurable parameters in the CBF method for performance enhancement with a systematical categorization. Based on that, we propose a novel framework combining the CBF method with Bayesian optimization (BO) to optimize the safe control performance. Considering feasibility/safety-critical constraints, we develop a safe version of BO using the barrier-based interior method to efficiently search for promising feasible configurable parameters. Furthermore, we provide theoretical criteria of our framework regarding safety and optimality. An essential advantage of our framework lies in that it can work in model-agnostic environments, leaving sufficient flexibility in designing objective and constraint functions. Finally, simulation experiments on swing-up control and high-fidelity adaptive cruise control are conducted to demonstrate the effectiveness of our framework.

Index Terms—Safety-critical control, control barrier functions, Bayesian optimization, safe optimization.

I. INTRODUCTION

IN practical control scenarios, safety is of paramount importance during the design and implementation of control strategies. However, ensuring safety often comes at the expense of degrading control performance. Striking a good balance between maintaining satisfactory control performance and adhering to safety constraints remains an ongoing challenging task. In literature, there are three competitive frameworks for safe control, include receding-horizon-based model predictive control (MPC) [1] and reinforcement learning (RL) [2], stability-inspired barrier Lyapunov function (BLF) [3], [4] and control barrier function (CBF) methods [5], as well as reachability-based methods [6]. Among these, the CBF

method is becoming a promising solution for safe control due to its high computational efficiency and versatility for easy integration with other control modules [7]. Unfortunately, as also a conservative solution [8], the CBF method involves various configurable parameters that are *implicitly* related to both control performance and system safety [9]. In this work, we aim to develop a framework for parameter adaptation for CBF-based safety-critical control tasks (SC2Ts), focusing on safely and efficiently optimizing the safe control performance.

There are two main issues hindering the utilization of existing optimization techniques, such as evolutionary algorithms [10] and derivative-based programming [11], [12], to calibrate the configurable parameters in SC2Ts. The first issue lies in the optimization efficiency under a black-box environment. The numerical control performance, such as the linear quadratic regulator (LQR) performance [13], may not be analytically dependent on the configurable parameters. However, measuring control performance can be computationally expensive, especially for high-fidelity simulations or real systems. As a result, we can only leverage zeroth-order and sample-efficient techniques for optimization [14]. The second issue is the feasibility of candidate configurations in safety-critical environments. Since a poorly calibrated configuration can lead to violations of safety constraints or pose higher risks, the trial-and-error approaches commonly used in safe RL are not suitable for SC2Ts. To address these challenges, we plan to leverage Bayesian optimization (BO), a recognized query-efficient framework for black-box optimization [14]. Furthermore, we will propose a safe version of BO that can explicitly incorporate feasibility-critical constraints during the parameter adaptation processes. While existing research on safe BO methods relies on the assumptions of Lipschitz-continuity of constraints [15], the proposed safe BO do not require these rigid assumptions.

Although considerable research has been conducted on parameter calibration and performance enhancement of control systems [16]–[19], this paper proposes a principled performance optimization framework tailored for SC2Ts.

- 1) We highlight the role of configurable parameters in CBF-based SC2Ts, and introduce a high-level structure of parameter adaptation as a constrained black-box optimization problem. For the first time, we provide a systematic categorization of involved parameters, offering valuable guidance to practitioners in parameter selection.
- 2) For efficient and safe parameter adaptation, we propose a novel safe BO method that combines the barrier-based interior-point method and pessimistic bounds of

The statements made herein are solely the responsibility of the authors. (*Corresponding author: Shiping Wen.*)

S. Wang and K. Li are with School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and K. Li is also with the Department of Computer Science, University of Exeter, North Park Road, Exeter, EX4 4QF, UK (e-mail: shnbo.wang@foxmail.com, k.li@exeter.ac.uk).

Z. Yan and S. Wen are with Australian AI Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, NSW 2007, Australia (e-mail: {yan.zheng; shiping.wen}@uts.edu.au).

Z. Guo is with College of Mathematics and Econometrics, Hunan University, Changsha, 410082, China (e-mail: zyguo@hnu.edu.cn).

S. Zhu is with School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China (e-mail: songzhu@cumt.edu.cn).

G. Wen is with the Department of Systems Science, School of Mathematics, Southeast University, Nanjing, China (e-mail: wenguanghui@gmail.com).

Gaussian processes (GPs) to ensure feasibility during optimization. We also give theoretical criteria for feasibility and optimality of this method.

- 3) We develop a novel framework that integrates the lower-level CBF method and the high-level safe barrier BO for optimal parameter adaptation. To illustrate the efficacy of this framework, we present two toy examples including tracking with obstacle avoidance and solving constrained LQR. Additionally, we conduct simulation experiments on swing-up control and adaptive cruise control (ACC) tasks to demonstrate the effectiveness.

Notations: Throughout this paper, \mathbb{R} denotes the real value space, while \mathbb{R}^n and \mathbb{R}_+^n denote n -dimensional real space and positive real space, respectively. For vector x and matrix A , $\|x\|$ and $\|A\|$ are the Euclidean norm and induced matrix norm, respectively. For $x \in \mathbb{R}^n$ and a continuously differentiable function $h(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, we denote its Lie derivative regarding $f(x) \in \mathbb{R}^n$ as $L_h f(x) = \frac{\partial h}{\partial x}^\top f(x)$. For a compact set \mathcal{C} , $\text{Int}(\mathcal{C})$ and $\partial\mathcal{C}$ denote the interior and boundary of \mathcal{C} .

II. PRELIMINARIES

A. Problem Formulation

Consider the following autonomous systems:

$$\dot{x} = f(x) + g(x)u, \quad (1)$$

where $x \in \mathcal{X} \subset \mathbb{R}^n$ and $u \in \mathcal{U} \subset \mathbb{R}^m$ are the system state and control input respectively, $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denotes the drift dynamics, and $g(x) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ is the input dynamics. Both f and g are locally Lipschitz continuous. The initial condition of (1) is $x_0 = x(t_0) \in \mathcal{X}$ for some starting time t_0 . For autonomous systems, we simplify the notation to x and u instead of $x(t)$ and $u(t)$. In practical control scenarios, both system state and input should be constrained in task-specific spaces. Specifically, any violation of state constraints, i.e. $x \notin \mathcal{X}$, may lead to a catastrophic damage. In addition, we consider $\mathcal{U} = \{u \in \mathbb{R}^m | A_u u \leq b_u\}$. Consequently, the safety-critical control tasks (SC2Ts) aims to synthesize a controller from \mathcal{U} that enforces $x \in \mathcal{X}$ during control processes.

In this work, we are going to propose a high-level optimization framework for SC2Ts, which safely and efficiently optimizes the configurable parameters in SC2Ts to improve the practical performance across different scenarios. This is formulated as a constrained optimization problem:

$$\text{minimize}_{\mathbf{z} \in \Omega} r(\mathbf{z}) \quad \text{s.t.} \quad \vec{g}(\mathbf{z}) \leq 0, \quad (2)$$

where $\mathbf{z} = (z_1, \dots, z_p)^\top \in \Omega$ denotes the decision variable, i.e., the configurable parameters in SC2Ts, $\Omega = [z_i^L, z_i^U]_{i=1}^p \subset \mathbb{R}^p$ is the search space, z_i^L and z_i^U are the lower and upper bounds of z_i respectively. The objective function $r(\mathbf{z})$ and q constraint functions $\vec{g}(\mathbf{z}) = (g_1(\mathbf{z}), \dots, g_q(\mathbf{z}))^\top$ are: *i) analytically unknown*, we do not have access to the formulation of r or \vec{g} , but are given N observations $\langle \mathbf{z}^i, r(\mathbf{z}^i), \vec{g}(\mathbf{z}^i) \rangle_{i=1}^N$; *ii) computational expensive*, the number of observations N is limited; and *iii) feasibility-critical*, we should ensure that candidate solutions to (2) are feasible with high probability. Under these harsh requirements, existing methods including differentiable convex programming (DOP), safe RL, and constrained BO, may not work well, as illustrated in Table I.

TABLE I
THE COMPARISON OF DIFFERENT OPTIMIZATION TECHNIQUES

Methods	Black-box	Efficiency	Feasibility
DOP [11], [12]	×	-	✓
Safe RL [2], [4]	✓	×	-
Constrained BO [19], [20]	✓	✓	×
Safe BO [15], [21]	✓	✓	✓

B. Control Barrier Function Method

The CBF method paves a way to effectively mediate the control objective and safety requirements when synthesizing a safe controller. We present a brief review of the CBF method for SC2Ts as follows, and the involved configurable parameters will be discussed later.

Definition 2.1 (Control Barrier Function [5]): For a continuously differentiable function $h(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ and a compact set $\mathcal{C} = \{x \in \mathcal{X} | h(x) \geq 0\}$ with $h(x) > 0, \forall x \in \text{Int}(\mathcal{C})$, and $h(x) = 0, \forall x \in \partial\mathcal{C}$, $h(x)$ is called a (zeroing) CBF defined on \mathcal{X} for systems (1) with a relative degree 1 if there exists an extended class \mathcal{K} function $\bar{\alpha}$ such that $\forall x \in \mathcal{X}$,

$$\sup_{u \in \mathcal{U}} (L_f h(x) + L_g h(x)u) \geq -\bar{\alpha}(h(x)). \quad (3)$$

A practical implementation of the CBF method is based on quadratic programming (QP), where the values of control input are computed in real-time by enforcing the safety requirements in (3) and integrating with nominal controllers, formulated by

$$u_{\text{safe}} = \arg \min_{u \in \mathbb{R}^m} l_\vartheta(u, x) \quad (4)$$

$$\text{s.t. } L_f h(x) + L_g h(x)u + \alpha h(x) \geq 0, \quad (4a)$$

$$A_u u \leq b_u. \quad (4b)$$

Herein, the objective function $l : \mathcal{U} \times \mathcal{X} \rightarrow \mathbb{R}_+$ denotes the myopic loss for mediating control performance, the subscript ϑ denotes the configurable variable in the design of $l_\vartheta(u, x)$. For instance, given a nominal controller $k_{\text{ref}}(x)$, the loss can be defined by $l_\vartheta(u, x) = \|u - k_{\text{ref}}(x)\|^2 + \lambda \|u\|^2$ with $\lambda \geq 0$ the configurable parameter for regularization. Note that the choice of function $\bar{\alpha}$ is also configurable. For brevity, we stipulate $\bar{\alpha}(h(x)) = \alpha h(x)$ with $\alpha > 0$ to be determined.

Remark 1: When a system has a relative degree more than one, i.e., $L_g h(x) = 0$, the exponential CBF can be introduced for a more conservative design of safe controllers [22], [23]. Similar to (4a), the exponential CBF method will introduce additional parameters [22] or functions [23] that need to be determined, thereby increasing the burden of calibration.

C. Gaussian Processes and Bayesian Optimization

BO serves as a query-efficient framework for black-box optimization, consisting of surrogate modeling for unknown objectives and constraints and acquisition functions for efficient search. The most commonly used surrogate models are Gaussian processes (GPs) [24]. Given the dataset $\mathcal{D}_r = \langle \mathbf{z}^i, r(\mathbf{z}^i) \rangle_{i=1}^N$ of the objective function in (2), a GP model with mean $m(\mathbf{z})$ and noise-free likelihood predicts a distribution $\tilde{r}(\tilde{\mathbf{z}}) \sim \mathcal{N}(\mu_r(\tilde{\mathbf{z}}), \sigma_r^2(\tilde{\mathbf{z}}))$ of a candidate solution $\tilde{\mathbf{z}} \in \Omega$ by:

$$\begin{aligned} \mu_r(\tilde{\mathbf{z}}) &= m(\tilde{\mathbf{z}}) + \mathbf{k}^{*\top} K^{-1} \mathbf{r}', \\ \sigma_r^2(\tilde{\mathbf{z}}) &= k(\tilde{\mathbf{z}}, \tilde{\mathbf{z}}) - \mathbf{k}^{*\top} K^{-1} \mathbf{k}^*, \end{aligned} \quad (5)$$

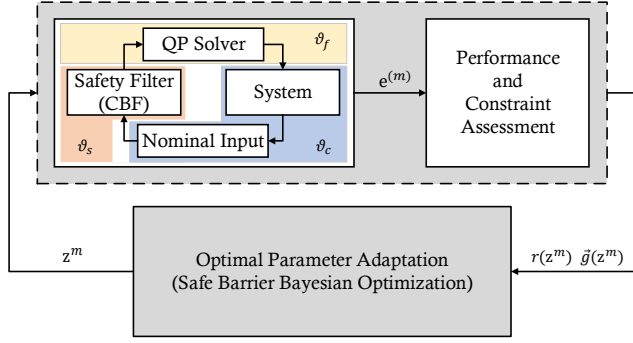


Fig. 1. An overview of the CBF-SB2O framework for SC2Ts.

in which \mathbf{k}^* is the covariance matrix between Z and $\tilde{\mathbf{z}}$, K is the covariance matrix of Z , $Z = (\mathbf{z}^1, \dots, \mathbf{z}^N)^\top$ and $\mathbf{r} = (r(\mathbf{z}^1) - m(\mathbf{z}^1), \dots, r(\mathbf{z}^N) - m(\mathbf{z}^N))^\top$. In this paper, we use a trainable constant mean function and the Matérn 5/2 as the kernel function [24]. The i -th constraint in (2) will be modeled by an independent GP $\tilde{g}_i(\tilde{\mathbf{z}})$ whose predictive mean and variance are denoted by $\mu_g^i(\tilde{\mathbf{z}})$ and $\sigma_g^i(\tilde{\mathbf{z}})$, respectively.

The search process of BO is driven by an acquisition function that automatically strikes a balance between exploiting the predicted optimum and exploring the uncertain areas that potentially contain better solutions. One of the most commonly used acquisition functions is the lower confidence bound (LCB) [25], which takes the following form:

$$\text{LCB}(\tilde{\mathbf{z}}|\mathcal{D}_r) = \mu_r(\tilde{\mathbf{z}}) - \beta_r \sigma_r(\tilde{\mathbf{z}}). \quad (6)$$

Herein, $\beta_r > 0$ controls the inclination of exploration during search. The constrained version of BO is referred to [14], [15].

III. MAIN RESULTS

In this section, we will develop a safe and efficient parameter adaptation framework for SC2Ts. Firstly, we study the relationship between configurable parameters and control performance in SC2Ts, meanwhile presenting a systematical categorization of the involved parameters. Secondly, we propose a novel safe barrier Bayesian optimization (SB2O) method that ensures feasibility with high probability during search. Finally, we design an integrated framework of parameter adaptation for SC2Ts, dubbed CBF-SB2O, with detailed investigations. The overall framework is depicted in Fig. 1.

A. Performance and Configurable Parameters of SC2Ts

1) *Relationship between Performance and Configurations:* The control performance of SC2Ts defined by reward functions in (2) is dependent not only on the configurable parameters but also dependent on x , u , and some inherent parameters \mathbf{k} , thereby we have $r(\mathbf{z}, \mathbf{k}, x, u)$. As an important example, the analytical format of the LQR performance [2] is

$$r(\mathbf{z}, \mathbf{k}, x, u) = \int_{t_0}^{\infty} (x^\top Q x + u^\top R u) dt, \quad (7)$$

where Q and R are predefined matrix parameters, i.e., \mathbf{k} . For a linear system, determining the optimal controller to (7) is equivalent to computing the optimal coefficients of the

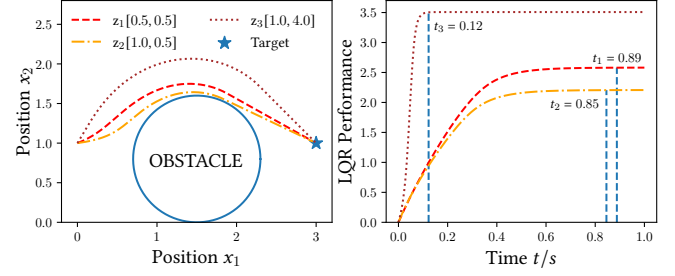


Fig. 2. A safe tracking control task with three different configurations. (Left) The moving trajectories with different \mathbf{z}_i . (Right) the LQR performance ($Q = I$ and $R = 2I$) and the response time $t_i = \inf_{t>t_0} \{\|x(t) - x_d\| \leq 0.1\}$.

linear feedback gain matrix, i.e. \mathbf{z} . For a nonlinear system, the performance can be optimized through adaptive dynamic programming (ADP) [26], [27]. For both cases, the configurable parameters play a critical role in enhancing the performance. To get rid of the influence of x and u , we study autonomous systems with fixed initial state x_0 .

In what follows, we present an illustrative example to show the influence of configurations on the control performance.

Illustrative Example 1: Consider a tracking control task with obstacle avoidance requirements. A 2-D moving system, described by $\dot{x} = u$ with position x and velocity input u , starts from $x_0 = [0, 1]^\top$ and moves to a target position $x_d = [3, 1]^\top$. For tracking control, a nominal controller is designed by $k_{\text{ref}}(x) = k_p(x_d - x)$ with a coefficient $k_p > 0$. There is an obstacle on the way, marked by a circle centered at $x_c = [1.5, 0.8]^\top$ with radius $r = 0.8$. By setting $h(x) = \|x - x_c\| - r$, we obtained a valid CBF constraint as (4a) with any $\alpha > 0$. Additionally, we set $l_\vartheta(u, x) = \|u - k_{\text{ref}}(x)\|^2$ in the QP (4) for a basic implementation of the CBF method [28]. Overall, the involved configurable parameters are $\mathbf{z} = [\alpha, k_p]$ with $\Omega = \mathbb{R}_+^2$. In Fig. 2, three runs were conducted with different \mathbf{z} , all tracked successfully without collision. However, the ranking of the three configurations varies according to different definitions of $r(\mathbf{z})$ in (2), as demonstrated below.

- Considering the LQR performance (7), \mathbf{z}_2 should be the best configuration among three. We may conclude that a larger α can lead to a better performance in this scenario.
- In terms of response time, \mathbf{z}_3 is a better choice since a larger k_p results in faster convergence. If a response time $t > 0.85$ s is forbidden, \mathbf{z}_1 is infeasible and should not be tested in feasibility-critic environment (2).
- For conservativeness, \mathbf{z}_2 is considered to be less conservative as it allows closer distance to the obstacle [29], while \mathbf{z}_1 and \mathbf{z}_3 maintain a higher safety margin.

In summary, the control performance of a SC2T is highly related to the configurable parameters, and the optimal configuration vary according to different objectives [30]. Furthermore, most objectives and constraints, such as latency and robustness, may not have analytical formats. In general, we consider problem (2) a black-box optimization problem. Since simulations and physical experiments can be costly, the optimization problem is expensive to solve.

2) *Categorization of configurable parameters in SC2Ts:* Based upon the above discussions, we provide a system-

atic categorization of configurable parameters in CBF-based SC2Ts, consisting of three aspects as follows.

- *Control parameters* ϑ_c related to specific control tasks. Although the CBF method enables a decoupled design of safety and control objective, for a well-calibrated nominal controller, the mediation between safety and control objective may degrade its performance and make it sub-optimal for SC2Ts. Therefore, ϑ_c contains the control parameters such as the proportional-integral-derivative (PID) [20] coefficients and the linear matrix inequality (LMI) gains [31]. Besides, implicit control parameters such as that in control Lyapunov function (CLF) [5] are included in this set as well.
- *Safety parameters* ϑ_s in CBF method. The real-time control synthesis framework (4) necessitates the design of both the objective function, such as the regularization and balancing among control objectives, and the safety constraints, such as α in (4a) and additional parameters in adaptive and robust CBF methods [19], [32]. Despite knowledge on the roles of some parameters, their impact towards control performance are rarely studied and can be significantly different across scenarios.
- *Deployment parameters* ϑ_f from practical algorithm implementation. To reduce the real-time computational burden, numerical solutions to (4) are required, such as discretization and approximation, which will in turn lead to the performance degradation and even unsafe behaviors. These parameters include configurations of a specific QP solver such as the error tolerance and learning rate [33], [34], and alternatives numerical solutions to (4) such as gradient-flow based method [28].

In all, the decision variable \mathbf{z} in (2) is a subset of the union group $\Pi = \{\vartheta_c, \vartheta_s, \vartheta_f\}$. Solving (2) at a high level requires low level evaluation for the performance measurement and constraint violations. While a re-calibration in this way has the potential to obtain good control performance, it poses a risk of unexpected output such as high evaluation costs and even unsafe behaviors. To this end, we introduce feasibility-critical constraints in (2) to circumvent this issue as much as possible, which will be studied in the next section.

Remark 2: The criteria and principles for designing ϑ_c and ϑ_s , such as LMI, can be further utilized as a feasibility guidance when searching for a better solution \mathbf{z} . For instance, policy search can be used to find LQR solution under semi-definite programming [35], [36]. In effect, it narrows the configuration space Ω , thereby improves optimization efficiency.

B. Safe Barrier Bayesian Optimization

To cope with unknown feasibility-critical constraints, we propose a safe version of BO that ensures candidate solutions feasible with high probability. We make the following assumptions commonly used in BO and constrained optimization.

Assumption 3.1: The objective r and all constraints in \vec{g} are independent from each other. In addition, the i -th constraint, $g_i : \Omega \rightarrow \mathbb{R}$, is in a Reproducing Kernel Hilbert Space (RKHS) of the kernel k and have a bounded RKHS norm $\|g_i\|_k$.

Lemma 3.1 (Calibrated uncertainty [25]): Let $\delta \in (0, 1)$. If Assumption 3.1 holds, then the following inequality holds

$$|\mu_g^i(\tilde{\mathbf{z}}) - g_i(\tilde{\mathbf{z}})| \leq \bar{\beta}_{N+1}^i \sigma_g^i(\tilde{\mathbf{z}}) \quad (8)$$

with probability at least $1 - \delta$ for all $\tilde{\mathbf{z}} \in \Omega$, where

$$\bar{\beta}_N^i = \left(2 \|g_i\|_k^2 + 300 \gamma_N^i \ln^3\left(\frac{N}{\delta}\right) \right)^{0.5}, \quad (9)$$

and γ_N^i is the maximum information gain with $\langle \mathbf{z}^j, g_i(\mathbf{z}^j) \rangle_{j=1}^N$.

Based upon the above lemma, to avoid evaluating an infeasible configuration, we introduce a pessimistic estimation of each constraint and integrate it into (6) as follows:

$$\begin{aligned} & \text{minimize}_{\tilde{\mathbf{z}} \in \Omega} \text{LCB}(\tilde{\mathbf{z}} | \mathcal{D}_r) \\ & \text{s.t. } \mu_g^i(\tilde{\mathbf{z}}) - \bar{\beta}_N^i \sigma_g^i(\tilde{\mathbf{z}}) \leq 0, \quad i = 1, \dots, q. \end{aligned} \quad (10)$$

Proposition 3.1: Let $\delta \in (0, 1)$. If Assumption 3.1 holds, the optimal solution $\tilde{\mathbf{z}}^*$ to (10) is safe/feasible in original problem (2) with probability at least $(1 - \delta)^q$.

The above result can be proven from the joint probability of independent events. In essence, solutions to (10) inherit the merits of LCB for efficient search and ensure feasibility against uncertainty. Despite mathematically appealing, it is indeed another intractable nonlinear constrained optimization problem, different from (2), with known objective and constraints. In order to mitigate the intractability, we modify (10) and propose a barrier-based acquisition function as

$$\text{SB2O}(\tilde{\mathbf{z}} | \mathcal{D}_{r,g}) = \text{LCB}(\tilde{\mathbf{z}} | \mathcal{D}_r) + \sum_{i=1}^q \phi_g^i(\tilde{\mathbf{z}}), \quad (11)$$

where $\phi_g^i(\tilde{\mathbf{z}}) = -c^{-1} \log(-\mu_g^i(\tilde{\mathbf{z}}) + \bar{\beta}_N^i \sigma_g^i(\tilde{\mathbf{z}}))$ with $c > 0$, and $\mathcal{D}_{r,g} = \langle \mathbf{z}^i, r(\mathbf{z}^i), \vec{g}(\mathbf{z}^i) \rangle_{i=1}^N$ denotes the collection of N evaluated objective and constraint pairs at the $(N+1)$ th round. Noted that solving (11) is as trivial as solving the original LCB problem (6). and the GP models enable us to leverage gradients for more efficient search. We quantify the feasibility and suboptimality of SB2O as follows.

Proposition 3.2: Let $\delta \in (0, 1)$ and the Assumption 3.1 hold. Assume also that the safe BO starts from a feasible point. At each following round, the optimal candidate solution $\tilde{\mathbf{z}}^*$ to (11) is safe/feasible with probability at least $(1 - \delta)^q$. Moreover, the solution is q/c -suboptimal as

$$|\text{LCB}(\tilde{\mathbf{z}}^* | \mathcal{D}_r) - \text{SB2O}(\tilde{\mathbf{z}}^* | \mathcal{D}_{r,g})| \leq \frac{q}{c}. \quad (12)$$

Proof. We will leverage the theoretical results from the interior-point method [37] to show the feasibility and suboptimality of the candidate solutions obtained by solving (11). First, we introduce an equivalent transformation of (10) as

$$\text{minimize}_{\tilde{\mathbf{z}} \in \Omega} \text{LCB}(\tilde{\mathbf{z}} | \mathcal{D}_r) + \sum_{i=1}^q I_- (\mu_g^i(\tilde{\mathbf{z}}) - \bar{\beta}_N^i \sigma_g^i(\tilde{\mathbf{z}})) \quad (13)$$

in which $I_- : \mathbb{R} \rightarrow \mathbb{R}$ denotes the indicator function satisfying

$$I_-(a) = \begin{cases} 0, & a \leq 0, \\ +\infty, & a > 0. \end{cases}$$

Based on Proposition 3.1, with probability at least $(1 - \delta)^q$, the optimal solutions to (13) is feasible regarding the original

optimization problem (2). To mitigate the intractability of the non-differentiable indicator function, a logarithmic approximation is used as $\phi(a) = -c^{-1} \log(-a)$, where $c > 0$ determines the accuracy of the approximation, revealed by the fact that $\lim_{c \rightarrow +\infty} \phi(a) = I_-(a)$, $\forall a \in \mathbb{R}$. As a result, (11) leverages this transformation for a differentiable approximation. Note that there should be a known feasible solution at the beginning to ensure this optimization problem is well-defined. As long as the initial search point is feasible, the candidate solutions to (11) is also feasible to (13), therefore with probability at least $(1 - \delta)^q$ feasible to (2) in the context of SC2Ts.

As a compromise, the approximation introduces a bias from optima. We see the fact that, with any gradient-based optimizer [24], all solutions $\tilde{\mathbf{z}}^*$ to (11) have zero value on their gradients. Denoting $\tilde{g}_i(\tilde{\mathbf{z}}) = -\mu_g^i(\tilde{\mathbf{z}}) + \tilde{\beta}_N^i \sigma_g^i(\tilde{\mathbf{z}})$, it implies

$$\begin{aligned} 0 &= \nabla \text{LCB}(\tilde{\mathbf{z}}^* | \mathcal{D}_r) + \sum_{i=1}^q \nabla \phi_g^i(\tilde{\mathbf{z}}^*) \\ &= \nabla \text{LCB}(\tilde{\mathbf{z}}^* | \mathcal{D}_r) - c^{-1} \sum_{i=1}^q \frac{1}{\tilde{g}_i(\tilde{\mathbf{z}}^*)} \nabla \tilde{g}_i(\tilde{\mathbf{z}}^*). \end{aligned} \quad (14)$$

Then we introduce the dual variables $\lambda_i = -1/(c\tilde{g}_i(\tilde{\mathbf{z}}))$, the solutions $\tilde{\mathbf{z}}^*$ are also solutions to the following Lagrangian

$$L(\tilde{\mathbf{z}}, \lambda_i) = \text{LCB}(\tilde{\mathbf{z}} | \mathcal{D}_r) + \sum_{i=1}^q \lambda_i \tilde{g}_i(\tilde{\mathbf{z}}). \quad (15)$$

Specifically, the dual function $l(\lambda_i)$ at $\tilde{\mathbf{z}}^*$ satisfies

$$l(\lambda_i^*) = \text{LCB}(\tilde{\mathbf{z}}^* | \mathcal{D}_r) + \sum_{i=1}^q \lambda_i^* \tilde{g}_i(\tilde{\mathbf{z}}^*) = \text{LCB}(\tilde{\mathbf{z}}^* | \mathcal{D}_r) - \frac{q}{c}.$$

The above equation quantifies the duality gap on any solution satisfying (14), i.e., any solution ultimately recommended by SB2O at each round. Therefore, we obtained the worst-case sub-optimality quantification in (12). \square

Remark 3: There are two basic safe BO frameworks in literature. The most popular one in machine learning ensures probabilistic safety with regret analysis [15]. To conduct safe exploration, additional conditions on Lipschitz continuity of both objective and constraints were needed in this framework, and the discretization of the search space Ω was also required in most of its extensions. In comparison, SB2O focuses optimization, mediating the exploration/exploitation balance by setting β and c . In addition, since equation (11) is differentiable with analytical gradients, we do not need discretization. The other safe BO framework also adopted the interior-point method, which was claimed to be compatible with any acquisition functions [21]. Unfortunately, it lacks the theoretical analysis on both safety and sub-optimality. In comparison, employing LCB enables SB2O to have good characteristics for both theoretical and pragmatic perspectives.

Illustrative Example 2: Consider a second-order system $\ddot{x} = u$. Initially, $x(t_0) = \dot{x}(t_0) = 1$. The objective is to find the optimal control input gain $K = [k_1, k_2]$ such that a controller $u = K[x, \dot{x}]^\top$ yields the best LQR performance defined in (7). Let $Q = I$ and $R = I$. Conventionally, the optimal gain K_r^* can be obtained by solving a well-defined Riccati function [38]. Alternately, we can conduct the optimal parameter

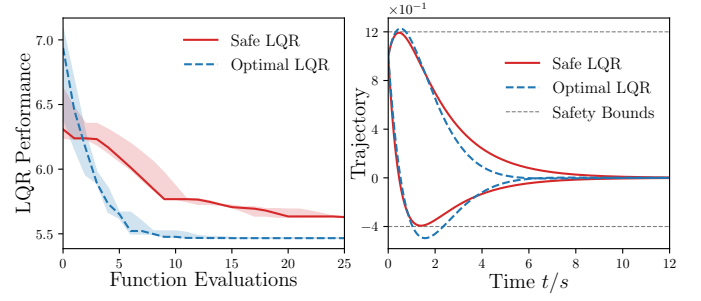


Fig. 3. An example of optimal control for linear systems via vanilla BO (optimal LQR) and SB2O (safe LQR). The search space is $[-5, -0.001]^2$. The best solution for optimal LQR is $[-1.0, -1.732]$ (same as the solution to the Riccati equation), and the one for safe LQR is $[-1.004, -2.324]$. (Left) The mean and variance of best observations during search with 5 repetitive experiments. (Right) Trajectories of system state with different gains.

adaptation with a few evaluations through BO. The results are shown in Fig. 3, where the optimal gain can be found within 15 evaluations. With a safety constraint such that x and \dot{x} should be bounded by $[-0.4, 1.2]$, conventional routines need more complex computations [39]. In this situation, we can utilize the SB2O. The results are also presented in Fig. 3. Likewise, the optimal parameter adaptation can be efficiently obtained within 30 evaluations. As shown in the right panel of Fig. 3, in addition to the good performance, this solution is safe regarding the state constraint, which is significant in SC2Ts.

IV. EXPERIMENTS

We conduct two simulation experiments on swing-up control of a cart-pole system [40] and ACC for autonomous vehicles in CARLA simulator [41]. All illustrative examples and experiments are performed on a desktop with Intel(R) Xeon(R) CPU E5-2620 v4 (2.10GHz) and NVIDIA GeForce GTX 1080Ti GPU. The GP models and CBO algorithm are implemented based on GPflow and GPflowOPT [42]. Two comparative algorithms are considered. The first one is random search strategy (dubbed CBF-RS) to mimic manual adjustment, where LHS is used to uniformly sample the candidate solutions. The other algorithm is a powerful black-box optimization method, the covariance matrix adaptation evolution strategy (CMA-ES) dubbed CBF-CMA [10], which is implemented based on pycma [43]. To obtain acceptable solutions in limited evaluations, we set the population size as 2 and initial deviation as 1. Another safe BO framework discussed in Remark 3 is not included in our experiments due to the impracticality of discretizing the searching space in SC2Ts [15]. For all algorithms, 20 repetitive experiments are conducted with different initialization.

A. Swing-Up Control for Cart-Pole Systems

1) *System Descriptions:* The physical structure of a 2-D cart-pole system is depicted in Fig. 4, whose dynamics can be formulated by

$$\dot{\mathbf{p}} = \begin{bmatrix} \dot{x} \\ \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} v \\ f_v(\theta, \omega) \\ \omega \\ f_\omega(\theta, \omega) \end{bmatrix} + \begin{bmatrix} 0 \\ g_v(\theta) \\ 0 \\ g_\omega(\theta) \end{bmatrix} u, \quad (16)$$

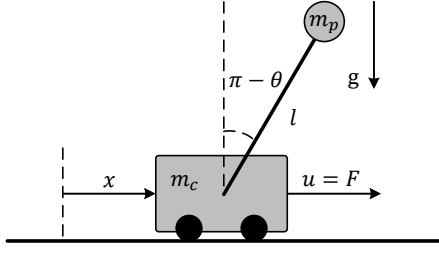


Fig. 4. The physical structure of a 2-D cart-pole system.

where

$$\begin{aligned} f_v(\theta, \omega) &= \frac{m_p \sin \theta (l\omega^2 + g \cos \theta)}{m_c + m_p \sin^2 \theta}, \\ f_\omega(\theta, \omega) &= -\frac{m_p l \omega^2 \cos \theta \sin \theta + (m_c + m_p) g \sin \theta}{l (m_c + m_p \sin^2 \theta)}, \\ g_v(\theta) &= \frac{1}{m_c + m_p \sin^2 \theta}, \quad g_\omega(\theta) = -\frac{1}{l (m_c + m_p \sin^2 \theta)}. \end{aligned}$$

The objective of the swing-up control is to manipulate the pendulum, starting from the initial angle $\theta = 0$ and cart position $x = x_0$, to swing above the cart such that θ reaches the desired angle $\theta_d = (2n+1)\pi$, where $n \in \mathbb{N}$. Simultaneously, the cart should be positioned at $x = x_d$. In what follows, we introduce the safe control framework of this task, and then optimize its configurable parameters.

2) *Nominal controller*: According to the energy shaping method [40], we define the error of control energy as

$$\tilde{E} = \frac{1}{2} m_p l^2 \omega^2 - m_p g l - m_p g l \cos \theta. \quad (17)$$

To ensure convergence of $|\tilde{E}|$, we introduce an additional proportional-differential (PD) controller and define \ddot{x}_d as

$$\ddot{x}_d = k_E \omega \cos \theta \tilde{E} - k_p x - k_d v, \quad (18)$$

where $k_E > 0$, $k_p \geq 0$ and $k_d \geq 0$ are configurable parameters to be calibrated. The nominal controller in (16) is designed by

$$\begin{aligned} u_d &= (m_c + m_p - m_p \cos^2 \theta) \ddot{x}_d \\ &\quad - m_p g \sin \theta \cos \theta - m_p l \omega^2 \sin \theta. \end{aligned} \quad (19)$$

To improve robustness, an LQR controller with feedback gain K_{LQR} is switched on when system state approaches to the target state [44]. In all, given a switching threshold $\eta > 0$ and $p_d = [x_d, 0, \theta_d, 0]^\top$, we design the control input as

$$u = \begin{cases} u_d, & \text{initially,} \\ -K_{LQR}(p - p_d), & \text{once } \|p^{2-4} - p_d^{2-4}\| \leq \eta. \end{cases} \quad (20)$$

As a safety constraint, the position x should be limited in $[-x_l, x_l]$. Accordingly, a CBF can be defined by $h_x = x_l^2 - x^2$, whose derivative is $\dot{h}_x = -2xv$. Since \dot{h}_x does not explicitly contain control input u , an exponential control barrier function (ECBF) [22] is constructed by $h_x^e = \dot{h}_x + \mu h_x$ with $\mu > 0$. For $\alpha > 0$, the safety constraint derived from h_x^e is

$$\begin{aligned} \dot{h}_x^e &= \ddot{h}_x + \mu \dot{h}_x \\ &= -2x(f_v(\theta, \omega) + g_v(\theta)u) - 2v^2 - 2\mu xv \\ &\geq -\alpha(-2xv + \mu(x_l^2 - x^2)). \end{aligned} \quad (21)$$

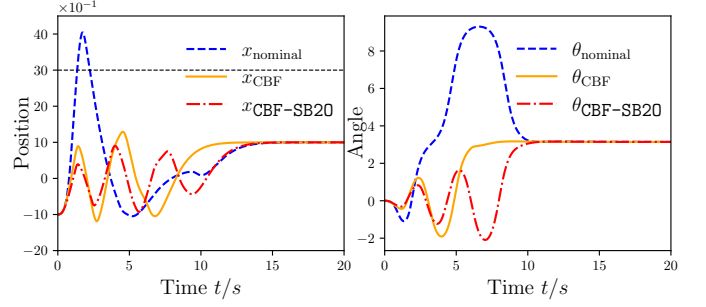


Fig. 5. Preliminary test on the swing up control using an unsafe nominal controller with $[k_E, k_p, k_d] = [0.3, 0.8, 1.0]$. We manually set the CBF parameters as $[\alpha, \mu] = [1.0, 1.0]$, with LQR performance 25.74. After 40 function evaluations by CBF-SB2O, we have $[\alpha, \mu] = [0.3, 7.7]$, with performance 22.44. (Left) Trajectories of position of systems with different controllers. (Right) Trajectories of angle under different controllers.

TABLE II
NUMERICAL RESULTS OF SIMULATION EXPERIMENTS.

Simulation	Method	Control Performance ^a	Feasibility Rate
Swing up	CBF-RS	21.05 (0.04%)	54.92%
	CBF-CMA	19.96 (0.81%)	89.13%
	CBF-SB2O	19.22[†]	100%
ACC	CBF-RS	200.11 (0.02%)	12.33%
	CBF-CMA	194.31 (2.07%)	67.25%
	CBF-SB2O	189.91[†]	90.33%

^a The superscript [†] indicates that the best algorithm is significantly better according to the Wilcoxon rank-sum test at 5% significance level. The computing result is given in brackets.

Finally, the safe controller is derived by the following QP:

$$\min_u \frac{1}{2} (u - u_d)^2 \quad (22)$$

$$s.t. \dot{h}_x^e(u) + \alpha(-2xv + \mu(x_l^2 - x^2)) \geq 0. \quad (22a)$$

3) *Experiment settings*: We consider two kinds of configurable parameters considered in this task, including the control parameters k_E , k_p , k_d , and safety parameters α and μ , i.e., $\mathbf{z} = [k_E, k_p, k_d, \alpha, \mu]$. It is noted that the K_{LQR} is calculated by first conducting local linearization and then solving the corresponding Riccati function. In this task, we consider the LQR performance (7) as the objective in (2). Two constraints are considered. The first one is failure constraint considering the limitation in update frequency of QP solver. The other one is timeout constraint to save experiment costs.

The cart-pole system is instantiated by $g = 10.0 \text{ m/s}^2$, $l = 2.0 \text{ m}$, $m_c = 5 \text{ kg}$, $m_p = 1 \text{ kg}$, the control input $u \in [-100 \text{ N}, 100 \text{ N}]$. Initially, system is stationary at $x = -1.0 \text{ m}$, $\theta = 0 \text{ rad}$. The target position is $x_d = 1.0 \text{ m}$ and the maximal allowable position is $x_l = 3.0 \text{ m}$. Let $Q = I$ and $R = 10^{-2}$ in LQR performance. Besides, the switch threshold is $\eta = 1.0$. A preliminary test is given in Fig. 5 with manually-determined or optimized CBF parameters. The configuration parameters ϑ_f follow default settings in the QP toolbox [45], where the update period of the filtered control input is fixed by 0.01s. The search space is $\Omega = [0.01, 100]^5$ (log scaled). Each experiment of CBF-SB2O and CBF-RS contains 50 initial solutions and 100 function evaluations, while CBF-CMA takes the best solution found in initialization as the starting point.

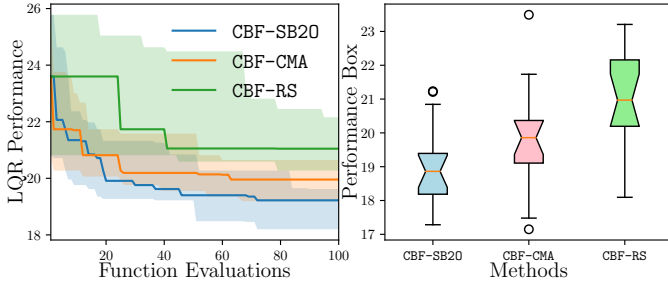


Fig. 6. The optimization result for swing up control via different algorithms. (Left) Performance optimization during 100 function evaluations, where the median and 25/75 percentiles of the best feasible objective value are plotted. (Right) Box plot of the final best performance of different algorithms. The median parameters found by CBF-SB20 is $[0.18, 0.03, 0.75, 0.56, 20.50]$.

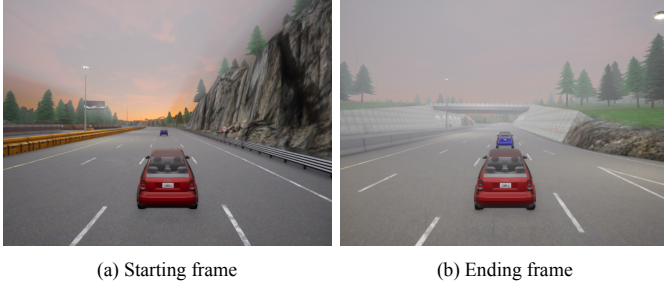


Fig. 7. Snapshots of ACC simulation in 'Town04'. The ego vehicle (red) should track the leader vehicle (blue) and keep a safe distance.

4) *Experiment Results:* The simulation results are shown in Fig. 6. Compared to a random safe configuration in Fig 5, the control performance is improved by all optimization methods. Differently, CBF-SB20 locates a better solution more efficiently throughout limited evaluations. As shown in Table II, CBF-SB20 is significantly better than other two algorithms. In terms of feasibility, it is shown that all configurations suggested by CBF-SB20 are feasible, while CBF-RS have a much higher probability to evaluate an infeasible solution. Similarly, the optimization efficiency of CBF-CMA is better than the random strategy, however, with more risks of conducting infeasible evaluations.

B. Adaptive Cruise Control for Autonomous Vehicles

Consider a leader vehicle (LV) starts and runs at a desired speed $v_0 = 64\text{km/h}$, while an ego vehicle (EV), initially 31.4m away from LV, should follow the leader while keeping a safe distance $d_0 = 10\text{m}$. The task environment is given in Fig. 7. Note that in CARLA, the vehicle models are inaccessible. We assume no lane change happens and adopt PID controllers with fixed parameters for lane keeping (steering) [46]. For longitudinal control (throttle), we use the following dynamics:

$$\dot{v} = -\frac{1}{m}F_r(v) + \frac{1}{m}u, \quad \dot{d} = v_0 - v, \quad (23)$$

where v , d and $u \in [-0.4\text{mg}, 0.4\text{mg}]$ are the velocity of EV, distance between EV and LV, and EV throttle input, respectively. We use the blueprint `audi` vehicles for both EV and LV. We set parameters as $m = 1650\text{kg}$, $f_0 = 2$, $f_1 = 5$,

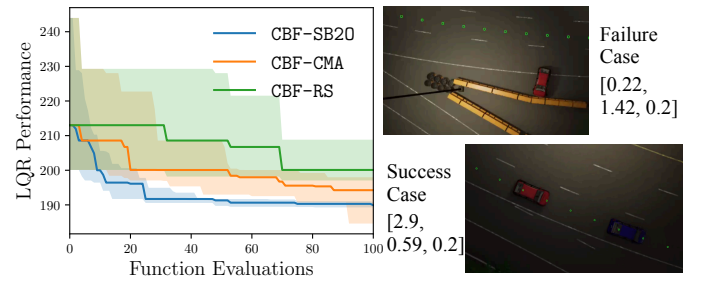


Fig. 8. (Left) Performance enhancement trajectories (median and 25/75 percentiles) for ACC via different algorithms during 100 function evaluations. The median parameters found by CBF-SB20 is $[3.51, 5.67, 0.1]$. (Right) Two successful and failed cases with the same f , where a smaller α is more likely to lead to the lane change.

and $f_2 = 3$ (standard unit) combining manual adjustment and estimation techniques [8], and fixed $g = 9.81\text{m/s}^2$.

Following the settings in [47], the ACC controller consists of a CLF $V = (v - v_0)^2$ and a CBF $h = d_0 - d_0$. Given a slack variable δ , we use QP with CBF&CLF constraints and the following objective function

$$l_\vartheta(u, \delta) = \frac{1}{m^2} (u - F_r)^2 + k\delta^2, \quad (24)$$

with $k > 0$. We also consider safety parameter α and QP computation frequency f as the deployment parameter. Overall, the parameter to be optimized is $\mathbf{z} = [\alpha, k, f]$. The objective function for ACC is designed as

$$r = \int_{t_0}^{\infty} \frac{1}{v_0^2} (v - v_0)^2 + \frac{1}{d_0^2} (d - d_0)^2 + \frac{1}{m^2} (u - F_r)^2 dt. \quad (25)$$

We use Town04 as the ACC environment. The search space is $\Omega = [0.1, 10]^2 \times [0.1, 1]$ (log scaled). The comparative results are presented in Fig. 8 and Table II, where CBF-SB20 remains competitive against other two algorithms. Notably, the successful rate of CBF-SB20 is maintained beyond 90%, while other algorithms are more likely to fail due to model mismatches and noises in the simulator. The failure occurs in two scenarios. First, there is a chance that QP becomes infeasible during the ACC process. We also observed that α is more important than f in influencing feasibility. The failure can also occur (with very small probability) when EV changed lane while turning due to the combination of *fixed* PID-based steering and *varying* CBF-based throttle input. This further highlights the importance of parameter adaptation in SC2Ts.

V. CONCLUSIONS

We proposed a systematic framework for optimal parameter adaptation of CBF-based controllers for SC2Ts. In consideration of the expensive black-box objective and constraints, a novel safe version of BO was developed for efficient and safe optimization. Two simulation experiments on swing-up control and high-fidelity adaptive cruise control demonstrated the competitiveness of CBF-SB20 in safe and efficient performance optimization of SC2Ts. One limitation of CBF-SB20 lies in that it can only work for episodic control tasks, making it not suitable for online parameter adaptation yet, which is likely to be solved by combining with DOP [12] or other approximation flow-based optimization techniques [28].

REFERENCES

- [1] X. Zhang, J. Ma, Z. Cheng, S. Huang, S. S. Ge, and T. H. Lee, "Trajectory generation by chance-constrained nonlinear MPC with probabilistic prediction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3616–3629, 2021.
- [2] J. García and F. Fernández, "A comprehensive survey on safe reinforcement learning," *J. Mach. Learn. Res.*, vol. 16, pp. 1437–1480, 2015.
- [3] K. P. Tee, S. S. Ge, and E. H. Tay, "Barrier lyapunov functions for the control of output-constrained nonlinear systems," *Autom.*, vol. 45, no. 4, pp. 918–927, 2009.
- [4] Y. Chow, O. Nachum, E. A. Duéñez-Guzmán, and M. Ghavamzadeh, "A Lyapunov-based approach to safe reinforcement learning," in *NeurIPS'18: Annual Conference on Neural Information Processing Systems*, 2018, pp. 8103–8112.
- [5] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Trans. Autom. Control.*, vol. 62, no. 8, pp. 3861–3876, 2017.
- [6] S. Bansal, M. Chen, S. L. Herbert, and C. J. Tomlin, "Hamilton-Jacobi reachability: A brief overview and recent advances," in *CDC'17: IEEE Annual Conference on Decision and Control*, 2017, pp. 2242–2253.
- [7] W. Shaw-Cortez, D. Oetomo, C. Manzie, and P. Choong, "Control barrier functions for mechanical systems: Theory and application to robotic grasping," *IEEE Trans. Control. Syst. Technol.*, vol. 29, no. 2, pp. 530–545, 2021.
- [8] S. Wang, B. Lyu, S. Wen, K. Shi, S. Zhu, and T. Huang, "Robust adaptive safety-critical control for unknown systems with finite-time elementwise parameter estimation," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 53, no. 3, pp. 1607–1617, 2023.
- [9] B. Li, S. Wang, Z. Guo, S. Zhu, J. Huang, J. Sun, G. Wen, and S. Wen, "Safe control framework of multi-agent systems from a performance enhancement perspective," *IEEE Transactions on Automation Science and Engineering*, pp. 1–10, 2024.
- [10] D. V. Arnold and N. Hansen, "A (1+1)-CMA-ES for constrained optimisation," in *GECCO'12: Genetic and Evolutionary Computation Conference*. ACM, 2012, pp. 297–304.
- [11] H. Parwana and D. Panagou, "Recursive feasibility guided optimal parameter adaptation of differential convex optimization policies for safety-critical systems," in *ICRA'22: International Conference on Robotics and Automation*. IEEE, 2022, pp. 6807–6813.
- [12] H. Ma, B. Zhang, M. Tomizuka, and K. Sreenath, "Learning differentiable safety-critical control using control barrier functions for generalization to novel environments," in *ECC'22: European Control Conference*. IEEE, 2022, pp. 1301–1308.
- [13] S. Wang, S. Wen, K. Shi, X. Zhou, and T. Huang, "Approximate optimal control for nonlinear systems with periodic event-triggered mechanism," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 6, pp. 2722–2731, 2023.
- [14] S. Wang and K. Li, "Constrained Bayesian optimization under partial observations: Balanced improvements and provable convergence," in *AAAI'24: Thirty-Eighth AAAI Conference on Artificial Intelligence*, 2024, pp. 15 607–15 615.
- [15] F. Berkenkamp, A. Krause, and A. P. Schoellig, "Bayesian optimization with safety constraints: Safe and automatic parameter tuning in robotics," *Machine Learning*, pp. 1–35, 2021.
- [16] R. Martinez-Cantin, "Funneled Bayesian optimization for design, tuning and control of autonomous systems," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1489–1500, 2019.
- [17] A. Marco, D. Baumann, M. Khadiv, P. Hennig, L. Righetti, and S. Trimpe, "Robot learning with crash constraints," *IEEE Robotics Autom. Lett.*, vol. 6, no. 2, pp. 1439–1446, 2021.
- [18] M. Zhu, D. Piga, and A. Bemporad, "C-glisp: Preference-based global optimization under unknown constraints with applications to controller calibration," *IEEE Trans. Control. Syst. Technol.*, vol. 30, no. 5, pp. 2176–2187, 2022.
- [19] R. K. Cosner, M. Tucker, A. J. Taylor, K. Li, T. G. Molnár, W. Ubelacker, A. Alan, G. Orosz, Y. Yue, and A. D. Ames, "Safety-aware preference-based learning for safety-critical control," in *Learn. Dynamics and Control Conference, L4DC*, vol. 168, 2022, pp. 1020–1033.
- [20] M. Khosravi, C. Koenig, M. Maier, R. S. Smith, J. Lygeros, and A. Rupenyan, "Safety-aware cascade controller tuning using constrained Bayesian optimization," *IEEE Trans. Ind. Informatics*, pp. 1–1, 2022.
- [21] D. Krishnamoorthy and F. J. Doyle, "Safe Bayesian optimization using interior-point methods - applied to personalized insulin dose guidance," *IEEE Control. Syst. Lett.*, vol. 6, pp. 2834–2839, 2022.
- [22] Q. Nguyen and K. Sreenath, "Exponential control barrier functions for enforcing high relative-degree safety-critical constraints," in *ACC'16: American Control Conference*. IEEE, 2016, pp. 322–328.
- [23] W. Xiao and C. Belta, "High-order control barrier functions," *IEEE Trans. Autom. Control.*, vol. 67, no. 7, pp. 3655–3662, 2022.
- [24] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005.
- [25] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," in *ICML'10: Proceedings of the 27th International Conference on Machine Learning*, 2010, pp. 1015–1022.
- [26] K. Zhang, H. Zhang, Y. Mu, and C. Liu, "Decentralized tracking optimization for partially unknown fuzzy interconnected systems via reinforcement learning method," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 4, pp. 917–926, 2021.
- [27] Q. Wei, Z. Yang, H. Su, and L. Wang, "Online adaptive dynamic programming for optimal self-learning control of VTOL aircraft systems with disturbances," *IEEE Transactions on Automation Science and Engineering*, pp. 1–10, 2022.
- [28] S. Wang, S. Wen, Y. Yang, Y. Cao, K. Shi, and T. Huang, "Suboptimal safety-critical control for continuous systems using prediction-correction online optimization," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 53, no. 7, pp. 4091–4101, 2023.
- [29] B. T. Lopez, J. E. Slotine, and J. P. How, "Robust adaptive control barrier functions: An adaptive and data-driven approach to safety," *IEEE Control. Syst. Lett.*, vol. 5, no. 3, pp. 1031–1036, 2021.
- [30] T. Huang, S. Wang, and K. Li, "Direct preference-based evolutionary multi-objective optimization with dueling bandits," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [31] H. K. Khalil, *Nonlinear systems*, 3rd ed. Prentice Hall, 2002.
- [32] 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 *CoRL'20: 4th Conference on Robot Learning*, vol. 155. PMLR, 2020, pp. 654–670.
- [33] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. Academic Press, 1981.
- [34] M. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, "Prediction-correction interior-point method for time-varying convex optimization," *IEEE Trans. Autom. Control.*, vol. 63, no. 7, pp. 1973–1986, 2018.
- [35] V. Balakrishnan and L. Vandenbergh, "Semidefinite programming duality and linear time-invariant systems," *IEEE Trans. Autom. Control.*, vol. 48, no. 1, pp. 30–41, 2003.
- [36] M. Fazel, R. Ge, S. M. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for the linear quadratic regulator," in *ICML'18: Proceedings of the 35th International Conference on Machine Learning*, vol. 80. PMLR, 2018, pp. 1466–1475.
- [37] S. P. Boyd and L. Vandenbergh, *Convex Optimization*. Cambridge University Press, 2014.
- [38] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. John Wiley & Sons, 2012.
- [39] D. Limón, T. Alamo, F. Salas, and E. F. Camacho, "On the stability of constrained MPC without terminal constraint," *IEEE Trans. Autom. Control.*, vol. 51, no. 5, pp. 832–836, 2006.
- [40] R. Tedrake, *Underactuated Robotics*. Course Notes for MIT 6.832, 2022. [Online]. Available: <http://underactuated.mit.edu>
- [41] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.
- [42] A. G. de G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrà, Z. Ghahramani, and J. Hensman, "Gpflow: A Gaussian process library using tensorflow," *J. Mach. Learn. Res.*, vol. 18, pp. 40:1–40:6, 2017.
- [43] N. Hansen, Y. Akimoto, and P. Baudis, "CMA-ES/pycma on Github," Zenodo, DOI:10.5281/zenodo.2559634, Feb. 2019.
- [44] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.
- [45] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.
- [46] L. Lindemann, A. Robey, L. Jiang, S. Das, S. Tu, and N. Matni, "Learning robust output control barrier functions from safe expert demonstrations," *IEEE Open Journal of Control Systems*, vol. 3, pp. 158–172, 2024.
- [47] W. Xiao and C. Belta, "Control barrier functions for systems with high relative degree," in *58th IEEE Conference on Decision and Control, CDC 2019, Nice, France, December 11-13, 2019*. IEEE, 2019, pp. 474–479.