

Safety Under Uncertainty: Tight Bounds with Risk-Aware Control Barrier Functions

Mitchell Black¹ Georgios Fainekos² Bardh Hoxha² Danil Prokhorov² Dimitra Panagou³

Abstract— We propose a novel class of risk-aware control barrier functions (RA-CBFs) for the control of stochastic safety-critical systems. Leveraging a result from the stochastic level-crossing literature, we deviate from the martingale theory that is currently used in stochastic CBF techniques and prove that a RA-CBF based control synthesis confers a tighter upper bound on the probability of the system becoming unsafe within a finite time interval than existing approaches. We highlight the advantages of our proposed approach over the state-of-the-art via a comparative study on a mobile-robot example, and further demonstrate its viability on an autonomous vehicle highway merging problem in dense traffic.

I. INTRODUCTION

A safe autonomous system is the gate to a fully autonomous system. Since the arrival of control barrier functions (CBFs) as a tool for safe control design and system verification [1]–[5], the field of safety-critical systems has drawn nearer to passing through this gate. Intuitively, a valid CBF constitutes a certificate that all system trajectories beginning within the set of safe states shall remain safe for all future time. For a class of control-affine dynamical systems, it is popular to include CBFs as linear constraints in a quadratic program (QP) based control law. And while the potential of this set-theoretic approach to render, preserve, and/or verify safety in a system has been demonstrated in applications like autonomous driving [5]–[7], quadrotor control [8], [9], and multi-agent systems [2], [10], its guarantees of safety may be lost in the absence of a complete system model.

In the deterministic setting, various works have addressed this problem by modifying standard CBF conditions. For example, robust-CBF approaches were proposed for safe control design under bounded perturbations to the system dynamics [4], [11], [12] or measurement error [13], [14], though the worst-case assumptions beget conservatism. Under linearly parametric model uncertainty, adaptive-CBF approaches have been used to both learn [9] and compensate for the effect of [15] unknown parameters in the system dynamics. For more general nonlinear uncertainty, Gaussian processes have been used to learn non-parametric, probabilistic models of the system [10], [16], [17], residual terms in the CBF [18], and barrier functions directly [8], among others. Additional works have adopted chance-constrained

¹Dept. of Aerospace Engineering, Univ. of Michigan, 1320 Beal Ave, Ann Arbor, MI 48109, USA; {mblackjr}@umich.edu.

²Toyota North America Research & Development, 1555 Woodridge Ave, Ann Arbor, MI 48105, USA; {georgios.fainekos, bardh.hoxha, danil.prokhorov}@toyota.com.

³Dept. of Robotics and Dept. of Aerospace Engineering, Univ. of Michigan, Ann Arbor, MI 48109, USA; {dpanagou}@umich.edu.

CBF conditions for probabilistic models [19]–[21], but do not consider safety over a time interval.

In many practical applications, the system behavior instead may be modeled by a class of stochastic differential equations (SDEs). Beginning with [22] and the stochastic barrier certificate, CBF development in the stochastic setting has leaned heavily on martingale theory for both discrete- and continuous-time stochastic processes. Stochastic CBFs (S-CBFs), introduced in [23] and adapted for risk-bounded control in [5], [24], leverage martingales to bound the probability that a system becomes unsafe over a finite time interval. While useful in theory, in practice the probability of safety is severely limited by the initial condition. In [25] this problem is addressed via reciprocal and zeroing CBFs for stochastic systems, with claims of safety with probability one, though the required level of conservatism is unclear.

In this paper, we deviate from martingale theory and make the following contributions:

- We introduce a new class of risk-aware control barrier functions (RA-CBFs) that uses a generator¹ condition derived from the stochastic level-crossing literature to obtain an upper bound on the probability that the system becomes unsafe over a finite time interval.
- We derive conditions under which our RA-CBF controller guarantees a smaller upper bound on the risk of the system becoming unsafe than existing S-CBF methods, and further show via a 100,000 trial numerical study that our controller results in less conservative behavior despite this stronger guaranteed risk protection.
- We consider an autonomous vehicle highway merging problem and demonstrate the efficacy of our proposed RA-CBF based controller in successfully merging amongst dense traffic under a required safety probability of 99%.

II. PRELIMINARIES AND PROBLEM FORMULATION

The uniform distribution supported by a and b is $U[a, b]$. A bolded \boldsymbol{x}_t denotes a vector stochastic process at time t . The Gauss error function is $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$, and $\text{erf}^{-1}(\cdot)$ is its inverse. The trace of a matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ is $\text{Tr}(\boldsymbol{M})$. The Lie derivative of a function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ along a vector field $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ at a point $x \in \mathbb{R}^n$ is $L_f \phi(x) \triangleq \frac{\partial \phi}{\partial x} f(x)$.

¹The (infinitesimal) generator of a stochastic process is analogous to the Lie derivative for deterministic systems.

A. Preliminaries

We consider the following class of nonlinear, control-affine, stochastic differential equations (SDE),

$$d\mathbf{x}_t = (f(\mathbf{x}_t) + g(\mathbf{x}_t)\mathbf{u}_t)dt + \sigma(\mathbf{x}_t)d\mathbf{w}_t, \quad (1)$$

where $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ denotes the state, $\mathbf{u} \in \mathcal{U} \subseteq \mathbb{R}^m$ the control input, and $\mathbf{w} \in \mathbb{R}^q$ a standard q -dimensional Wiener process (i.e., Brownian motion) defined over the complete probability space (Ω, \mathcal{F}, P) for sample space Ω , σ -algebra \mathcal{F} over Ω , and probability measure $P : \mathcal{F} \rightarrow [0, 1]$. We consider a class of memoryless, state-feedback controllers such that the control signal is $\mathbf{u}_t = k(\mathbf{x}_t)$, with $f : \mathcal{X} \rightarrow \mathbb{R}^n$, $g : \mathcal{X} \rightarrow \mathbb{R}^{n \times m}$, and $k : \mathcal{X} \rightarrow \mathcal{U}$ known, locally Lipschitz, and bounded on \mathcal{X} , which is assumed to be bounded. We consider that $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times q}$ from (1) also satisfies these regularity conditions, and thus assume that for all $\mathbf{u} \in \mathcal{U}$ and $\mathbf{x}_0 \in \mathcal{X}_0 \subset \mathbb{R}^n$ the process $\{\mathbf{x}_t : t \in [0, \infty)\}$ is a strong solution to (1) (see [26, Ch. 5, Def. 2.1]). For strong solutions, the generator is defined as follows.

Definition 1. [27, Def. 7.3.1] The (infinitesimal) generator \mathcal{A} of \mathbf{x}_t is defined by

$$\mathcal{A}\phi(\mathbf{y}) = \lim_{t \downarrow 0} \frac{\mathbb{E}[\phi(\mathbf{x}_t) | \mathbf{x}_0 = \mathbf{y}] - \phi(\mathbf{y})}{t}, \quad (2)$$

where $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ belongs to $\mathcal{D}_{\mathcal{A}}$, the set of all functions such that the limit exists for all $\mathbf{x} \in \mathbb{R}^n$.

The generator is the stochastic analog to the Lie derivative for deterministic systems in that it characterizes the derivative of ϕ over the trajectories of (1) in expectation. By [27, Thm. 7.3.3], for a twice continuously differentiable function ϕ with compact support, i.e., $\phi \in \mathcal{C}_0^2(\mathbb{R}^n) \subset \mathcal{D}_{\mathcal{A}}$, the generator \mathcal{A} of \mathbf{x}_t is described by

$$\mathcal{A}\phi(\mathbf{x}) = L_f\phi(\mathbf{x}) + L_g\phi(\mathbf{x})k(\mathbf{x}) + \frac{1}{2}\text{Tr}\left(\sigma(\mathbf{x})^T \frac{\partial^2 \phi}{\partial \mathbf{x}^2} \sigma(\mathbf{x})\right),$$

which we denote $\Gamma_\phi(\mathbf{x}, \mathbf{u}) := \mathcal{A}\phi(\mathbf{x})$ by using $\mathbf{u} = k(\mathbf{x})$.

Consider the set S defined by a twice continuously differentiable, positive semi-definite function $B : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$S = \{\mathbf{x} \in \mathbb{R}^n : 0 \leq B(\mathbf{x}) < 1\}, \quad (3)$$

and assume that is also known that for some $\gamma \in [0, 1]$,

$$B(\mathbf{x}) \leq \gamma, \quad \forall \mathbf{x} \in \mathcal{X}_0. \quad (4)$$

In the deterministic setting, the set S is said to be *forward-invariant* if $\mathbf{x}_0 \in S \implies \mathbf{x}_t \in S, \forall t \geq 0$. In this paper, we assume that S denotes the set of safe states for (1) and therefore use the notions of forward invariance and safety interchangeably. In the stochastic setting, however, there may be failure cases in which \mathbf{x}_t exits S , i.e., the system becomes unsafe. We therefore consider the stopped process, $\tilde{\mathbf{x}}_t$, and probabilistic forward invariance, adapted from [28].

Definition 2. [29] Suppose that $\tau > 0$ is the first time of exit of \mathbf{x}_t from the open set S . The stopped process $\tilde{\mathbf{x}}_t$ is

$$\tilde{\mathbf{x}}_t = \begin{cases} \mathbf{x}_t; & t < \tau, \\ \mathbf{x}_\tau; & t \geq \tau. \end{cases}$$

Definition 3. Let $0 < p \leq 1$, and consider the stopped process over an interval of length $T > 0$, i.e., $\{\tilde{\mathbf{x}}_t : t \in [0, T]\}$, w.r.t. the set S defined by (3). The set $S \subset \mathcal{X} \subseteq \mathbb{R}^n$ is a **probabilistic forward-invariant set with probability p** for system (1) over the interval $[0, T]$ if $P(\tilde{\mathbf{x}}_t \in S, \forall t \in [0, T]) \geq p$.

Thus, S is **safe** with probability p over the interval $[0, T]$ if it is a probabilistic forward-invariant set with probability p over $[0, T]$. Alternatively, the probability ρ of the system becoming unsafe over $[0, T]$, i.e., $\rho := P(\exists t \in [0, T] : \tilde{\mathbf{x}}_t \notin S)$, is bounded by $\rho \leq 1 - p$. We refer to ρ in the remainder as the "system risk". One approach to bounding the system risk of (1) is to use S-CBFs in the control design [5], [23].

Definition 4. Consider a set $S \subset \mathbb{R}^n$ defined by (3) for a twice continuously differentiable, positive semi-definite function B satisfying (4). The function B is a **stochastic control barrier function (S-CBF)** defined on the set S if there exist $\alpha, \beta \geq 0$ such that for the system (1) the generator $\Gamma_B(\mathbf{x}, \mathbf{u})$ satisfies the following condition, for all $\mathbf{x} \in S$,

$$\inf_{\mathbf{u} \in \mathcal{U}} \Gamma_B(\mathbf{x}, \mathbf{u}) \leq -\alpha B(\mathbf{x}) + \beta. \quad (5)$$

A valid S-CBF guarantees that the system risk is bounded from above, as shown in the following [23, Prop. 1].

Theorem 1. Consider a stochastic system of the form (1), a set of safe states S implicitly defined by a function B as in (3), and the interval $[0, T]$ for $T > 0$. Let the probability that $\{\tilde{\mathbf{x}}_t : t \in [0, T]\}$ exits S be denoted $\rho_{S\text{-CBF}} := P(\exists t \in [0, T] : \tilde{\mathbf{x}}_t \notin S \mid \tilde{\mathbf{x}}(0) \in \mathcal{X}_0)$. If B is a stochastic control barrier function for (1) over the set S , then

$$\rho_{S\text{-CBF}} \leq \begin{cases} 1 - (1 - \gamma)e^{-\beta T}; & \alpha > 0 \text{ and } \alpha \geq \beta, \\ \left(\gamma + (e^{\beta T} - 1)\frac{\beta}{\alpha}\right)e^{-\beta T}; & \alpha > 0 \text{ and } \alpha < \beta, \\ \gamma + \beta T; & \alpha = 0. \end{cases} \quad (6)$$

Remark 1. A S-CBF controller can certify that at best a fraction of $1 - \gamma$ of the trajectories will be safe over a time interval for any choice of $\alpha, \beta, T \geq 0$. Note that, due to the martingale origins of S-CBFs, the strength of the process noise ($\sigma(\mathbf{x})$) in (1) does not appear in (6). This motivates the problem formalized in Section II-B.

For control design, it has become popular to synthesize a variety of CBFs (e.g. S-CBFs [5], chance-constrained CBFs [19], [20], robust CBFs [4], etc.) with a nominal controller via quadratic program (QP) based control laws of the form

$$\begin{aligned} \mathbf{u}^* = \arg \min_{\mathbf{u} \in \mathcal{U}} & \frac{1}{2}\|\mathbf{u} - \mathbf{u}_0\|^2 + \frac{1}{2}w\delta^2 \\ \text{s.t.} \end{aligned} \quad (7a)$$

$$A\mathbf{u} + b + c\delta \leq 0, \quad (7b)$$

where \mathbf{u}_0 is the nominal control law, δ is a slack variable with weight $w \geq 0$, and (7b) represents a generic CBF constraint, with $b, c \in \mathbb{R}$, $A \in \mathbb{R}^{1 \times m}$. In the remainder, we use (7) to compare emergent behaviors of systems under S-CBFs and our proposed risk-aware CBF.

B. Problem Formulation

Based on Remark 1, we hypothesize that S-CBFs may introduce unnecessary conservatism into risk-aware control design. We use an illustrative example to show that this is indeed true, and thereby motivate the problem.

Example 1. Consider a mobile robot seeking to achieve the following objective: visit a circular region of radius $R_g > 0$ centered on $s_g = [x_g \ y_g]^T$, defined with respect to the origin s_0 of an inertial frame \mathcal{I} , while remaining inside a circular region of radius R_c centered on s_0 . The goal specification may be thought of as visiting a point of interest, while the constraint may model e.g. limited communication range. We choose $s_g = [2 \ 2]^T$, $R_c = 1$, $R_g = 0.25$ such that the goal and safe sets do not intersect. Assume that the robot may be modeled as a 2D stochastic single-integrator,

$$d\mathbf{z}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} dt + \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix} d\mathbf{w}_t, \quad (8)$$

where $\mathbf{z} = [x \ y]^T$ denotes the robot's position (in m) with respect to s_0 , the control $\mathbf{u} = [v_x \ v_y]^T$ consists of velocities along x and y axes (in m/s), and $\sigma_x, \sigma_y \in \mathbb{R}$ dictate the strength of noise introduced by the Wiener process $\mathbf{w} \in \mathbb{R}^2$.

For control we use (7) with a nominal input of $\mathbf{u}_0 = -k[(x - x_g)(y - y_g)]^T$ with $k > 0$. The input constraints are $|v_x, v_y| \leq v_{max} = 10$, and (7b) is the S-CBF condition (5), with $B(\mathbf{z}) = \frac{x^2 + y^2}{R_c^2}$. An upper bound on the risk of the system becoming unsafe under the S-CBF-QP controller is then given by (6). We fixed $\mathbf{z}_0 = [1/\sqrt{2}, 0]^T$ such that $B(\mathbf{z}_0) = 0.5 = \gamma$ and considered a time horizon of $T = 1$ sec at a time-step of $\Delta t = 0.001$ sec. We then simulated the trajectories over $N = 100,000$ trials with $\sigma_x, \sigma_y = 0.3v_{max} \cdot \Delta t$, i.e., a strength of 30% of the maximum control input.

The results (shown in Table I) confirmed our hypothesis: the S-CBF constraint may yield a theoretical system risk bound that significantly overestimates the actual fraction of unsafe outcomes. Despite bounded failure rates of 0.505 and 0.990, the S-CBF based controller preserved safety in 100% of the 100,000 trials (0 failures) in both cases over the $T = 1$ sec intervals. It is clear from this example that the S-CBF risk bounds may not, and certainly here do not, provide any meaningful guidelines.

TABLE I

STOCHASTIC CBF TRIALS $N = 100,000$

| Theoretical ρ | Measured ρ | α | β | γ | T |
|--------------------|-----------------|----------|---------|----------|-----|
| 0.505 | 0 | 0.1 | 0.01 | 0.50 | 1.0 |
| 0.990 | 0 | 10.0 | 4.0 | 0.50 | 1.0 |

As such, we seek to design a stochastic control framework that bounds the system risk over a finite time interval while bridging the gap between results derived in theory and those observed in practice. We now formally define the problem.

Problem 1. Consider the stochastic dynamical system of the form (1) and an associated safe set S defined by a twice continuously differentiable, positive semi-definite function B

satisfying (4). Design a feedback controller $\mathbf{u}_t = k(\mathbf{x}_t)$ such that under certain conditions $\rho := P(\exists t \in [0, T] : \tilde{\mathbf{x}}_t \notin S) < \rho_{S\text{-CBF}}$, where $\rho_{S\text{-CBF}}$ is given by (6), and identify the conditions under which this relation holds.

III. RISK-AWARE CONTROL BARRIER FUNCTION

In this section, we propose solving Problem 1 via a novel class of risk-aware control barrier functions (RA-CBFs). First, we require the following.

Lemma 1. Suppose that $w : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ is a standard Wiener process, and $T > 0$ and $a > 0$ are constants. Then, the probability that $w_t < a$, for all $t \in [0, T]$ is given by

$$P\left(\sup_{0 \leq t \leq T} w(t) < a\right) = \operatorname{erf}\left(\frac{a}{\sqrt{2T}}\right). \quad (9)$$

Proof. The proof follows directly from [30, Section 3]. \square

In what follows, we denote the integral of the generator of $\tilde{\mathbf{x}}_t$ as

$$I_L(t) := \int_0^t \Gamma_B(\tilde{\mathbf{x}}_s, \mathbf{u}_s) ds, \quad (10)$$

which may be included as an integrator state in an augmented system of dimension $n+1$. We now formally introduce the notion of the risk-aware control barrier function.

Definition 5. Consider a set $S \subset \mathbb{R}^n$ defined by (3) for a twice continuously differentiable, positive semi-definite function B satisfying (4). The function B is a **risk-aware control barrier function** on the set S if there exists a Lipschitz continuous function $\alpha \in \mathcal{K}_\infty$ such that for the system (1) the following holds for all $\mathbf{x} \in S$,

$$\inf_{\mathbf{u} \in \mathcal{U}} \Gamma_B(\mathbf{x}, \mathbf{u}) \leq \alpha(h(I_L(t))), \quad (11)$$

where

$$h(I_L(t)) = 1 - \gamma - (\sqrt{2\eta T}) \operatorname{erf}^{-1}(1 - \rho_d) - I_L(t), \quad (12)$$

with $I_L(t)$ given by (10), $\rho_d \in \left[1 - \operatorname{erf}\left(\frac{1-\gamma}{\sqrt{2\eta T}}\right), 1\right]$ a design parameter, and

$$\eta = \sup_{\mathbf{x} \in S} \|L_\sigma B(\mathbf{x})\|. \quad (13)$$

In the following theorem, our main result, we prove that RA-CBFs bound the risk that a system of the form (1) becomes unsafe over a finite time interval.

Theorem 2. Let $T > 0$, and denote the system risk as $\rho := P(\exists t \in [0, T] : \tilde{\mathbf{x}}_t \notin S \mid \tilde{\mathbf{x}}_0 \in \mathcal{X}_0)$. If B is a risk-aware control barrier function on the set S , then,

$$\rho \leq \rho_d, \quad (14)$$

where $\rho_d \in \left[1 - \operatorname{erf}\left(\frac{1-\gamma}{\sqrt{2\eta T}}\right), 1\right]$ is a design parameter with η given by (13).

Proof. Let $\tau > 0$ be the first time of exit of \mathbf{x}_t from the open set S . With $\{\mathbf{x}_t : t \in [0, \infty)\}$ a strong solution to (1), we have via Itô's Formula [27, Theorem 4.2.1] that $\forall t < \tau$,

$$dB(\tilde{\mathbf{x}}_t) = \Gamma_B(\tilde{\mathbf{x}}_t, \mathbf{u}_t) dt + L_\sigma B(\tilde{\mathbf{x}}_t) d\mathbf{w}_t,$$

which leads to the integral equation $B(\tilde{\mathbf{x}}_t) = B(\tilde{\mathbf{x}}_0) + I_L(t) + I_S(t)$, where $I_L(t)$ is a Lebesgue integral defined by (10) and $I_S(t)$ is a stochastic integral defined by

$$I_S(t) = \int_0^t L_\sigma B(\tilde{\mathbf{x}}_s) d\mathbf{w}_s. \quad (15)$$

While (10) can be evaluated deterministically, the stochastic integral (15) is an Itô integral [27, Def. 3.1.6] and thus induces a distribution on $B(\tilde{\mathbf{x}}_t)$ based on

$$I_S(t) \sim \mathcal{N} \left(0, \mathbb{E} \left[\left(\int_0^t L_\sigma B(\tilde{\mathbf{x}}_s) d\mathbf{w}_s \right)^2 \right] \right).$$

With \mathbf{w} the q -dim. standard Wiener process, it follows from the q -dim. Itô isometry ([31, Lemma 18] (an extension of the 1-dim. Itô isometry [27, Lemma 3.1.5])) that $\mathbb{E} \left[\left(\int_0^t L_\sigma B(\tilde{\mathbf{x}}_s) d\mathbf{w}_s \right)^2 \right] = \int_0^t \|L_\sigma B(\tilde{\mathbf{x}}_s)\|^2 ds$, and thus that $B(\tilde{\mathbf{x}}_t) \sim \mathcal{N}(\mu_B(t), \sigma_B^2(t))$, where $\mu_B(t) = B(\tilde{\mathbf{x}}_0) + I_L(t)$ and $\sigma_B^2(t) = \int_0^t \|L_\sigma B(\tilde{\mathbf{x}}_s)\|^2 ds$. As such,

$$\rho = 1 - P \left(\sup_{0 \leq t \leq T} B(\tilde{\mathbf{x}}_t) < 1 \mid B(\tilde{\mathbf{x}}_0) \leq \gamma \right).$$

Now, let $\bar{I}_S(t) \sim \mathcal{N}(0, \int_0^t \eta^2 ds)$ induce a probability distribution on $\bar{B}(\tilde{\mathbf{x}}_t)$, i.e., $\bar{B}(\tilde{\mathbf{x}}_t) = B(\tilde{\mathbf{x}}_0) + I_L(t) + \bar{I}_S(t)$. Then, since by (13) $\int_0^t \eta^2 ds \geq \sigma_B^2(t)$, for all $t \geq 0$, it follows that

$$\rho \leq \bar{\rho} := 1 - P \left(\sup_{0 \leq t \leq T} \bar{B}(\tilde{\mathbf{x}}_t) < 1 \mid B(\tilde{\mathbf{x}}_0) \leq \gamma \right). \quad (16)$$

However, we observe that $\int_0^t \eta^2 ds = \eta^2 t$, and thus by Gaussian linearity $\bar{I}_S(t) = \eta \sqrt{t} w(t)$, where $w(t)$ is the 1-dimensional standard Wiener process, which implies that $\bar{B}(\tilde{\mathbf{x}}_t) = B_0 + I_L(t) + \eta \sqrt{t} w(t)$. Therefore,

$$\begin{aligned} \bar{\rho} &= 1 - P \left(\sup_{0 \leq t \leq T} w(t) < \frac{1 - \gamma - I_L(t)}{\eta \sqrt{t}} \mid B_0 \leq \gamma \right), \\ &\leq 1 - P \left(\sup_{0 \leq t \leq T} w(t) < \frac{1 - \gamma - \bar{I}_L}{\eta \sqrt{T}} \mid B_0 \leq \gamma \right), \end{aligned} \quad (17)$$

where $\bar{I}_L = \sup_{0 \leq t \leq T} I_L(t)$. Thus, from (16), (17), and Lemma 1 we have

$$\rho \leq \bar{\rho} \leq 1 - \text{erf} \left(\frac{1 - \gamma - \bar{I}_L}{\sqrt{2}\eta T} \right). \quad (18)$$

Now, in order for (18) to be true it must hold that $\bar{I}_L \leq 1 - \gamma - (\sqrt{2}\eta T) \text{erf}^{-1}(1 - \bar{\rho})$, a sufficient condition for which is that $I_L(t) \leq 1 - \gamma - (\sqrt{2}\eta T) \text{erf}^{-1}(1 - \bar{\rho})$, $\forall t \in [0, T]$. We then define a set $S_I = \{I_L \in \mathbb{R} : h(I_L) \geq 0\}$, where $h(I_L) = 1 - \gamma - (\sqrt{2}\eta T) \text{erf}^{-1}(1 - \bar{\rho}) - I_L$, and observe that if h is a valid CBF for the set S_I , i.e., if there exists $\alpha \in \mathcal{K}_\infty$ such that, $\forall I_L \in S_I$ and $\forall t \in [0, T]$, (11) holds then the set S is probabilistically forward-invariant with probability $p = 1 - \rho \geq 1 - \bar{\rho}$. Thus, from (18) it follows that since $I_L(0) = 0$ by definition, $\bar{\rho}_0 \leq 1 - \text{erf} \left(\frac{1 - \gamma}{\sqrt{2}\eta T} \right)$ where $\bar{\rho}_0$ is $\bar{\rho}$ at $t = 0$. Therefore, for $h(I_L) \geq 0$ it must hold that $\rho_d \in [\bar{\rho}_0, 1]$. This completes the proof. \square

Remark 2. Under an RA-CBF controller, the upper bound ρ on the system risk is a function only of the initial condition γ , the length of the time interval T , and the effect of the stochastic noise on B , i.e., η . The function h measures how closely the controller has taken the system to the tolerable risk threshold ρ via actions integrated to form I_L .

We now present conditions under which the bound on the system risk guaranteed by Theorem 2 is strictly less than the bound guaranteed under the S-CBF control framework.

Theorem 3. Let the premises of Theorem 1 hold, and let ρ_d be as defined in Theorem 2. If B is a risk-aware control barrier function, then

$$\min_{\rho_d \in \mathcal{R}} \rho_d < \rho_{S\text{-CBF}} \quad (19)$$

whenever

$$\eta < \frac{1 - \gamma}{\sqrt{2}T \text{erf}^{-1}(1 - \gamma)}, \quad (20)$$

where η is given by (13) and $\mathcal{R} = [1 - \text{erf} \left(\frac{1 - \gamma}{\sqrt{2}\eta T} \right), 1]$.

Proof. The proof follows immediately from the observation in Remark 1, i.e., that $\rho_{S\text{-CBF}} \geq \gamma$, $\forall \alpha, \beta, T \geq 0$. Then, from Theorem 2, $\min \rho_d < \rho_{S\text{-CBF}}$ whenever $1 - \text{erf} \left(\frac{1 - \gamma}{\sqrt{2}\eta T} \right) < \gamma$. By rearranging terms, we recover (20). \square

This result provides guidelines as to when a RA-CBF controller would predict lower levels of risk than a S-CBF controller, or vice versa. In the robot motion problem from Section II-B, with dynamics (8) and barrier function $B(\mathbf{z}) = \frac{x^2 + y^2}{R_c^2}$ we have $\eta \approx 0.009$. As such, $\min \rho_d \geq \rho_{S\text{-CBF}}$ over the $T = 1$ sec time interval would have required either $\gamma < 1e-15$ given σ_x, σ_y or $\sigma_x, \sigma_y \approx 50v_{max} \cdot \Delta t$ given $\gamma = 0.5$, both of which are unrealistic for the problem.

When η given by (13) is large, however, the allowable risk specifications using a RA-CBF (based on $\min \rho_d$) may not be acceptable. In this case, it may be more useful to design the controller to remain inside a smaller sub-level set $S_\mu = \{\mathbf{x} \in \mathbb{R}^n : 0 \leq B(\mathbf{x}) < \mu\}$, or to derive a total risk of the system becoming unsafe by cascading sets $S_{\mu_1}, \dots, S_{\mu_k}$, as shown in the following result.

Theorem 4. Suppose that the premises of Theorem 2 hold. Consider a sequence μ_0, \dots, μ_k such that $\gamma = \mu_0 < \mu_1 < \dots < \mu_k = 1$ with sub-level sets $S_{\mu_i} = \{\mathbf{x} \in \mathbb{R}^n : 0 \leq B(\mathbf{x}) < \mu_i\} \subseteq S$, $\forall i \in \{1, \dots, k\}$, each of which has η_i defined by (13) over S_{μ_i} . If B is a RA-CBF on each set S_{μ_i} , then $\rho \leq \rho_d$, where ρ_d is a design parameter bounded by

$$\prod_{i=1}^k \left(1 - \text{erf} \left(\frac{\mu_i - \mu_{i-1}}{\sqrt{2}T\eta_i} \right) \right) \leq \rho_d \leq 1. \quad (21)$$

Proof. First, observe that by Definition 5 the function B is a RA-CBF on the set S_{μ_i} if (11) holds for all $\mathbf{x} \in S_{\mu_i}$, where the $1 - \gamma$ term in (12) is replaced by $\mu_i - \gamma$. Let $\rho_{\mu_i} := P(\exists t \in [0, T] : \tilde{\mathbf{x}}_t \notin S_{\mu_i} \mid \tilde{\mathbf{x}}_0 \in S_{\mu_i} \setminus S_{\mu_{i-1}})$. Then, with B a RA-CBF on S_{μ_i} , it follows from Theorem 2 that

$$\rho_{\mu_i} \leq 1 - \text{erf} \left(\frac{\mu_i - \mu_{i-1}}{\sqrt{2}T\eta_i} \right), \quad (22)$$

where η_{μ_i} is defined by (13) over the set S_{μ_i} . By (4) and Bayes' rule, we then obtain that $\rho \leq \prod_{i=1}^k \rho_{\mu_i}$ and thus by (22) we recover (21). \square

The bound in (22) is particularly useful when $\eta_1 < \dots < \eta_k$, as this is the best reduction in the conservatism in using η over all S . The number of partitions k is a design choice, and should be adjusted according to the desired system risk and each η_i . For control design, the RA-CBF condition (11) must be satisfied on each S_{μ_i} with a choice of $\rho_{d_i} \geq \rho_{\mu_i}$.

IV. NUMERICAL CASE STUDIES

In this section, we highlight the efficacy of our RA-CBF controller in solving two illustrative examples: the robot problem from Section II-B, and a highway merging problem.

A. Single-Integrator Robot

The problem setup is identical to that in Section II-B, with the robot's dynamics given by (8) and its controller of the form (7) with RA-CBF condition (11). The results were a striking departure from the S-CBF based controller. When an upper bound on system risk was set to 0.505 to match the S-CBF trial, a fraction of 0.458 of the trials violated the safety condition, as shown in Table II. When the RA-CBF

TABLE II
RISK-AWARE CBF TRIALS $N = 100,000$

| Predicted ρ | Measured ρ | γ | η |
|------------------|-----------------|----------|--------|
| 0.010 | 10^{-4} | 0.50 | 0.006 |
| 0.505 | 0.458 | 0.50 | 0.006 |

controller was used at a maximum system risk of $\rho = 0.01$, however, not only did the measured ρ satisfy this bound ($1e-4$), but the system trajectories took more aggressive actions toward the boundary of the safe set than the S-CBF controller even when its risk level was set to $\rho_{S-CBF} = 0.505$, as shown in Figure 1.

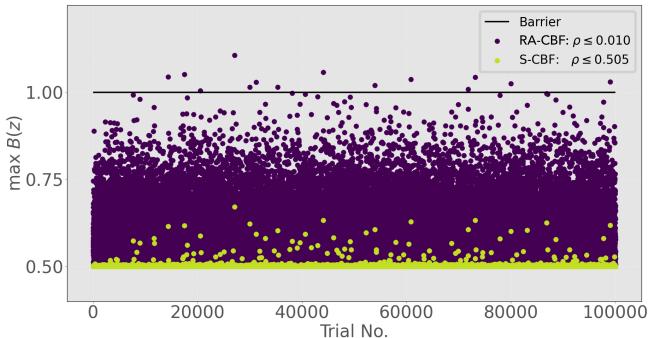


Fig. 1. Maximum barrier function values ($\max_{0 \leq t \leq T} B(z_t)$) over each trial for RA-CBF (resp. S-CBF) with system risk bounded by $\rho \leq 0.01$ (resp. $\rho_{S-CBF} \leq 0.505$).

B. Highway Merging

Let \mathcal{I} be an inertial frame with an origin point s_0 . Consider a collection of automobiles \mathcal{A} , a subset of which travel on a two-lane highway near an on-ramp (i.e., $\mathcal{A}_H \subset \mathcal{A}$), and the remainder of which seek to merge onto the highway via

the on-ramp (i.e., $\mathcal{A}_M \subset \mathcal{A}$). Suppose that the dynamics of vehicle $i \in \mathcal{A}$ obey a stochastic bicycle model of the form (1) whose deterministic component is described by [32, Ch. 2] (omitted due to space) and used to model cars in [7]. The stochastic term is $\sigma_i(z_i) = \sigma^T \mathbf{I}_{5 \times 5}$ with $\sigma = [0 \ 0 \ \sigma_a \ \sigma_\omega]^T$. The state is $z_i = [x_i \ y_i \ \psi_i \ v_i \ \beta_i]^T$, where x_i and y_i denote the longitudinal and lateral positions (in m) of the center of gravity (c.g.) of vehicle i with respect to s_0 , ψ_i is the orientation (in rad) of its body-fixed frame, \mathcal{B}_i , with respect to \mathcal{I} , v_i is the velocity (in m/s) of the rear wheel with respect to s_0 , and β_i is the slip angle² (assume $|\beta_i| < \frac{\pi}{2}$) of the c.g. relative to \mathcal{B}_i (in rad). The front and rear wheelbases are l_f and l_r . See [7, Figure 1] for a model diagram. The control is $u_i = [a_i \ \omega_i]^T$, with a_i the linear acceleration of the rear wheel (in m/s²) and ω_i the angular velocity (in rad/s) of β_i . The vector $w \in \mathbb{R}^5$ is the 5D standard Wiener process.

We simulated 1000 trials of a $T = 4$ s highway merging scenario with 11 vehicles, where $\mathcal{A}_M = \{0\}$ was the ego vehicle and $\mathcal{A}_H = \{1, \dots, 10\}$. Highway vehicles $i \in \mathcal{A}_H$ were initialized 15m apart in the x direction, and distributed evenly between lanes 1 ($y = 0$) and 2 ($y = 3$). Their initial velocities were distributed according to $v_{i,0} \sim U[29, 31]$. The ego vehicle was initialized 98.75m down the on-ramp with an initial velocity $v_{e,0} \sim U[24, 26]$ such that (deterministically) under its nominal acceleration policy it would collide directly with vehicle 2. The noise components were $\sigma_a = A_{drag}\Delta t$ and $\sigma_\omega = \sigma_a \frac{\bar{\omega}}{\bar{a}}$, where $\bar{\omega} = \frac{\pi}{16}$ (rad/s) and $\bar{a} = 2.0$ (m/s²) define the input constraints $a_i \in [-\bar{a}, \bar{a}]$ and $\omega_i \in [-\bar{\omega}, \bar{\omega}]$, and $A_{drag} = 0.1 + 5\bar{v} + 0.25\bar{v}^2$, with $\bar{v} = 35$ (m/s), such that the noise at one standard deviation represents the acceleration due to aerodynamic drag [1] traveling at 35m/s. The ego vehicle was controlled using (7) with 11 RA-CBF constraints corresponding to the occupied sub-level set S_{μ_i} , where we selected $\mu_i = i/5$ for $i \in \{1, \dots, 5\}$. The 10 ego collision avoidance constraints were encoded via $B_{ei}(z_e, z_i) = e^{-h_{ei}(z_e, z_i)}$, where $h_{ei}(z_e, z_i)$ is the relaxed future-focused CBF (rff-CBF) (introduced for collision avoidance in [7]) with $\gamma(h_0) = 0.1h_0$. The road constraints were encoded with a rff-CBF of the form $B_r(z_e) = e^{-h_r(z_e)}$ for $h_r(z_e) = h_{r,0}(z_e, 0) + h_{r,0}(z_e, 1)$, where

$$h_{r,0}(z_e, \tau) = - \left(x_e \tan(\theta) + \frac{w_l}{2 \cos(\theta)} - (y_e + \dot{y}_e \tau - y_l) \right)^2$$

with θ the road angle with respect to the x -axis, w_l the width of a lane, and y_l the lane center. For all RA-CBFs, the corresponding η_i values were determined numerically by simulating 1000 trials and taking $\eta_i = \max_{x \in S_{\mu_i}} \|L_\sigma B(x)\|$ over all trials and all time. The resulting η_i values are provided in Table III. For the on-ramp, the angle of attack was $\theta = 3^\circ$ ($\theta = 0^\circ$ for highway lanes). The ego nominal control u_0 was the LQR law detailed in [7, Appendix 1] based on the desired lane and velocity ($v_d = 30$ m/s). For naturalistic driving behavior, we used the intelligent driver model (IDM) [33] to compute acceleration inputs a_i of the

² β_i is related to the steering angle δ_i via $\tan \beta_i = \frac{l_r}{l_r + l_f} \tan \delta_i$.

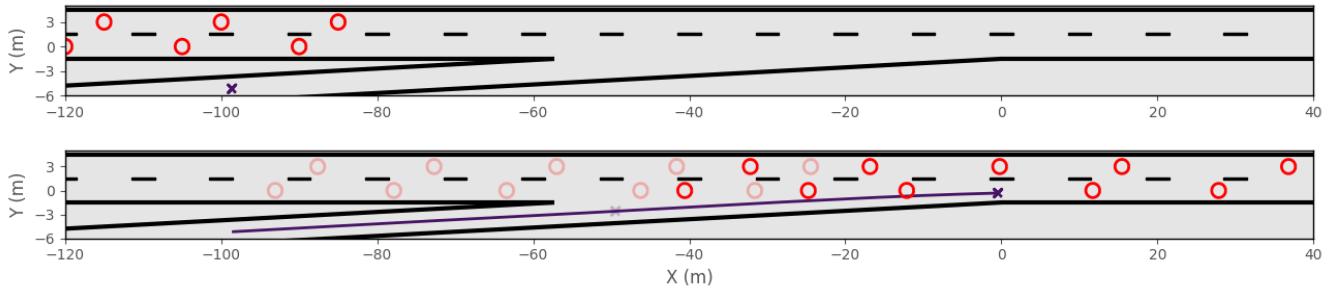


Fig. 2. Snapshots at $t = 0.0$ s (a) and $t = 4.0$ s (with $t = 2.0$ s translucent) (b) of one trial from the empirical study on the RA-CBF-QP controller in the highway merging scenario. Traffic flows left to right, the ego vehicle is a blue X, and highway vehicles are red circles.

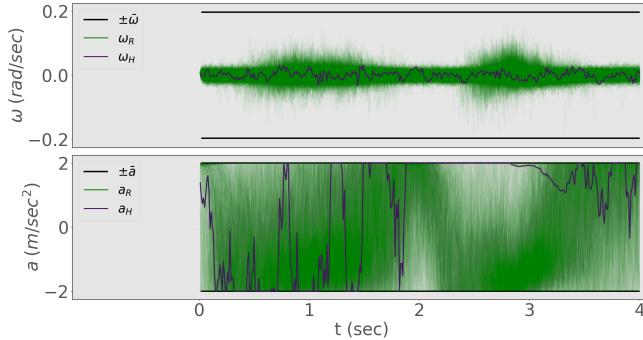


Fig. 3. Ego vehicle control inputs from both the highlighted trial in Figure 2 (subscript H) and remaining trials (subscript R).

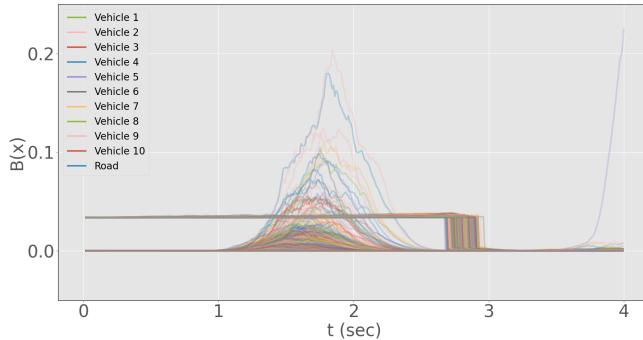


Fig. 4. RA-CBF trajectories over 1000 highway merging trials.

TABLE III
 η_i VALUES DERIVED EMPIRICALLY

| CBF | η_1 | η_2 | η_3 | η_4 | η_5 |
|-----------|----------|----------|----------|----------|----------|
| Road | 0.012 | 0.025 | 0.035 | 0.046 | 0.067 |
| Collision | 0.018 | 0.031 | 0.049 | 0.063 | 0.076 |

highway vehicles $i \in \mathcal{A}_H$. For varying driver aggression, we randomized the vehicles' desired time gaps in the IDM according to $\tau \sim U[0.25, 0.75]$. Their steering inputs ω_i were computed using LQR based on the desired heading ($\psi_d = 0$).

Based on η_i from Table III, a simulation length of $T = 4$ sec, known γ for all $B(z_0)$, the minimum specifiable risks associated with leaving each sub-level set S_{μ_i} for road safety and collision avoidance are provided in Table IV. We chose the $\rho_{d,i}$ values provided in Table V such that the probability

TABLE IV

ρ_i VALUES FOR SUB-LEVEL SETS S_{μ_i}

| CBF | ρ_1 | ρ_2 | ρ_3 | ρ_4 | ρ_5 |
|-----------|----------|----------|----------|----------|----------|
| Road | 8.58e-4 | 0.046 | 0.153 | 0.277 | 0.456 |
| Collision | 0.026 | 0.107 | 0.308 | 0.427 | 0.511 |

of remaining safe with respect to the road is 0.99999, the probability of remaining safe with respect to all 10 highway vehicles combined is 0.991, and thus the total probability of safety is $p \geq 0.99$, which yields $\rho \leq 0.01$.

TABLE V
SPECIFIED RISK BOUNDS $\rho_{d,i}$ FOR SUB-LEVEL SETS S_{μ_i}

| CBF | $\rho_{d,1}$ | $\rho_{d,2}$ | $\rho_{d,3}$ | $\rho_{d,4}$ | $\rho_{d,5}$ |
|-----------|--------------|--------------|--------------|--------------|--------------|
| Road | 0.001 | 0.1 | 0.25 | 0.5 | 0.6 |
| Collision | 0.05 | 0.15 | 0.4 | 0.5 | 0.6 |

Over 1000 simulated trials, the RA-CBF based controller safely merged 1000 times, satisfying the risk bound of $\rho \leq 0.01$. Figure 2 highlights one of these safe merges in which the ego vehicle merges behind vehicle 2, where the applied control inputs are shown in Figure 3. In this study, we observed that in all 1000 trials the ego vehicle merged behind vehicle 2. In another study, in which a risk of $\rho \leq 0.12$ was specified, we observed that the ego vehicle safely merged in 914 of the 1000 trials ($p = 0.914$). Interestingly, of these 914 safe trials, the ego vehicle merged behind vehicle 2 at a rate of 0.749 and merged ahead of it the remaining 0.251 fraction of safe trials, an indicator of the willingness of the ego vehicle in the second study to take on additional risk.

V. CONCLUSION

In this paper, we proposed a new class of RA-CBFs for stochastic safety-critical systems. We introduced a new CBF condition for a class of stochastic, nonlinear, control-affine systems and proved that its use for control synthesis guarantees an upper bound on the risk that the system becomes unsafe over a finite time interval. We then derived conditions under which our RA-CBF controller results in a smaller system risk than existing methods, and conducted a direct comparative study on a mobile robot example. We demonstrated our control strategy under a 99.1% safety guarantee on an autonomous vehicle highway merging problem in the midst of dense traffic. In the future, we will

consider measurement noise and investigate applications of our control framework to recovery problems in the context of safe, predictive control.

REFERENCES

- [1] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, “Control barrier function based quadratic programs for safety critical systems,” *IEEE Trans. on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2017.
- [2] Y. Chen, A. Singletary, and A. D. Ames, “Guaranteed obstacle avoidance for multi-robot operations with limited actuation: A control barrier function approach,” *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 127–132, 2021.
- [3] W. Shaw Cortez, D. Oetomo, C. Manzie, and P. Choong, “Control barrier functions for mechanical systems: Theory and application to robotic grasping,” *IEEE Transactions on Control Systems Technology*, vol. 29, no. 2, pp. 530–545, 2021.
- [4] K. Garg and D. Panagou, “Robust control barrier and control lyapunov functions with fixed-time convergence guarantees,” in *2021 American Control Conference (ACC)*, 2021, pp. 2292–2297.
- [5] S. Yaghoubi, K. Majd, G. Fainekos, T. Yamaguchi, D. Prokhorov, and B. Hoxha, “Risk-bounded control using stochastic barrier functions,” *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1831–1836, 2021.
- [6] T. D. Son and Q. Nguyen, “Safety-critical control for non-affine nonlinear systems with application on autonomous vehicle,” in *58th IEEE Conference on Decision and Control*, 2019, pp. 7623–7628.
- [7] M. Black, M. Jankovic, A. Sharma, and D. Panagou, “Intersection crossing with future-focused control barrier functions,” *arXiv preprint arXiv:2204.00127*, 2022.
- [8] L. Wang, E. A. Theodorou, and M. Egerstedt, “Safe learning of quadrotor dynamics using barrier certificates,” in *2018 IEEE International Conference on Robotics and Automation*, 2018, pp. 2460–2465.
- [9] M. Black, E. Arabi, and D. Panagou, “Fixed-time parameter adaptation for safe control synthesis,” *arXiv preprint arXiv:2204.10453*, 2022.
- [10] R. Cheng, M. J. Khojasteh, A. D. Ames, and J. W. Burdick, “Safe multi-agent interaction through robust control barrier functions with learned uncertainties,” in *59th IEEE Conference on Decision and Control*, 2020, pp. 777–783.
- [11] M. Jankovic, “Robust control barrier functions for constrained stabilization of nonlinear systems,” *Automatica*, vol. 96, pp. 359–367, 2018.
- [12] S. Yaghoubi, G. Fainekos, and S. Sankaranarayanan, “Training neural network controllers using control barrier functions in the presence of disturbances,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.
- [13] 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,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 6286–6291.
- [14] S. Dean, A. Taylor, R. Cosner, B. Recht, and A. Ames, “Guaranteeing safety of learned perception modules via measurement-robust control barrier functions,” in *Proceedings of the 2020 Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 155. PMLR, 16–18 Nov 2021, pp. 654–670.
- [15] A. J. Taylor and A. D. Ames, “Adaptive safety with control barrier functions,” in *2020 American Control Conf.*, 2019, pp. 1399–1405.
- [16] P. Jagtap, G. J. Pappas, and M. Zamani, “Control barrier functions for unknown nonlinear systems using gaussian processes,” in *59th IEEE Conference on Decision and Control*, 2020, pp. 3699–3704.
- [17] F. Castañeda, J. J. Choi, B. Zhang, C. J. Tomlin, and K. Sreenath, “Pointwise feasibility of gaussian process-based safety-critical control under model uncertainty,” in *60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 6762–6769.
- [18] M. Khan, T. Ibusi, and A. Chatterjee, “Safety uncertainty in control barrier functions using gaussian processes,” in *2021 IEEE International Conference on Robotics and Automation*, 2021, pp. 6003–6009.
- [19] M. J. Khojasteh, V. Dhiman, M. Franceschetti, and N. Atanasov, “Probabilistic safety constraints for learned high relative degree system dynamics,” in *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, vol. 120, Jun 2020, pp. 781–792.
- [20] Y. Lyu, W. Luo, and J. M. Dolan, “Probabilistic safety-assured adaptive merging control for autonomous vehicles,” in *2021 IEEE International Conference on Robotics and Automation*, 2021, pp. 10764–10770.
- [21] W. Luo, W. Sun, and A. Kapoor, “Multi-robot collision avoidance under uncertainty with probabilistic safety barrier certificates,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 372–383, 2020.
- [22] S. Prajna, A. Jadbabaie, and G. J. Pappas, “A framework for worst-case and stochastic safety verification using barrier certificates,” *IEEE Trans. on Automatic Control*, vol. 52, no. 8, pp. 1415–1428, 2007.
- [23] C. Santoyo, M. Dutreix, and S. Coogan, “A barrier function approach to finite-time stochastic system verification and control,” *Automatica*, vol. 125, p. 109439, 2021.
- [24] S. Yaghoubi, G. Fainekos, T. Yamaguchi, D. Prokhorov, and B. Hoxha, “Risk-bounded control with kalman filtering and stochastic barrier functions,” in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 5213–5219.
- [25] A. Clark, “Control barrier functions for stochastic systems,” *Automatica*, vol. 130, p. 109688, 2021.
- [26] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer New York, NY, 1998.
- [27] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed. Springer, 2003.
- [28] E. Kofman, J. A. De Doná, and M. M. Seron, “Probabilistic set invariance and ultimate boundedness,” *Automatica*, vol. 48, no. 10, pp. 2670–2676, 2012.
- [29] H. J. Kushner, *Stochastic Stability and Control*, ser. Mathematics in Science and Engineering. Elsevier, 1967, vol. 33.
- [30] I. Blake and W. Lindsey, “Level-crossing problems for random processes,” *IEEE Trans. on Information Theory*, vol. 19, no. 3, pp. 295–315, 1973.
- [31] K. S. Zhang, G. Peyré, J. Fadili, and M. Pereyra, “Wasserstein control of mirror langevin monte carlo,” in *Proceedings of 33rd Conference on Learning Theory*, vol. 125, 09–12 Jul 2020, pp. 3814–3841.
- [32] R. Rajamani, *Vehicle Dynamics and Control*. Springer US, 2012.
- [33] M. Treiber, A. Hennecke, and D. Helbing, “Congested traffic states in empirical observations and microscopic simulations,” *Phys. Rev. E*, vol. 62, pp. 1805–1824, Aug 2000.