Opportunistic Safety Outside the Maximal Controlled Invariant Set

Zexiang Liu[1], Hao Chen[1], Yulong Gao[2], Necmiye Ozay[1]

Abstract— **Existing safety control methods for non-stochastic systems become undefined when the system operates outside the maximal robust controlled invariant set (RCIS), making those methods vulnerable to unexpected initial states or unmodeled disturbances. In this work, we propose a novel safety control framework that can work both inside and outside the maximal RCIS, by identifying the worst-case disturbance that can be handled at each state and constructing the control inputs robust to that worst-case disturbance model. We show that such disturbance models and control inputs can be jointly computed by considering an invariance problem for an auxiliary system. Finally, we demonstrate the efficacy of our method both in simulation and in a drone experiment.**

I. Introduction

Constraints are ubiquitous in control tasks for safety-critical systems, such as lane keeping for autonomous vehicles, overload protection in power systems, and obstacle avoidance for mobile robots. The goal of safety control is to synthesize controllers that can guarantee a system operates under its safety constraints indefinitely. Many methods have been developed over the years that can provide such safety guarantees, such as viability theory [3], reference governers [9], safety supervisory control [14], [16], robust control barrier function [12], and Hamilton-Jacobi reachability [4]. The key behind all those methods is to find a set of states such that if the system starts from this set, the system can be controlled to stay within this set against the worst-case disturbance, without violating any safety constraints. Such a set is called a *robust controlled invariant set (RCIS)* of the system.

Notably, there exists a unique *maximal RCIS* that contains all possible RCISs given some safety constraints. Controllers synthesized by the aforementioned methods are defined only if the system initially operates in the maximal RCIS, since otherwise the worst-case disturbance is able to force the system to violate the safety constraints in finite time. However, in practice, the system may be initialized outside the maximal RCIS or exit the maximal RCIS due to unexpected disturbances. In those cases, the system may still operate safely and even re-enter the maximal RCIS, as long as the disturbance is not completely adversarial (or to put it differently, the disturbance behaves collaboratively to some extent). The core question here is how to synthesize controllers that can seize the opportunity to keep the system safe when the disturbance is not entirely adversarial. Apparently the aforementioned methods do not answer this question since they become undefined outside the maximal RCIS.

Similar issues also arise in the field of reactive synthesis for finite transition systems and have been addressed by recent works [2], [6]. The main idea there is that if a winning strategy robust to all disturbances does not exist, one should pick a strategy at least as good as the other strategies in terms of the amount of disturbances it can be robust to. These ideas are also applied in the context of abstraction-based control [15] and finite-horizon constrained optimal control [8]. Inspired by this line of work, in this letter, we present a novel safety control framework that finds control inputs that are safe against the largest possible disturbance set. Compared with the existing safety control methods restricted by the maximal RCIS, our method has the following benefits:

- The proposed controller provides the same safety guarantees as the existing methods when the system operates inside the maximal RCIS.
- When outside the maximal RCIS, the proposed controller is robustly safe against the largest amount of disturbance within a predefined template set.
- The proposed controller is well-defined as long as a constraint violation is evitable with some possible collaboration from the disturbance.

In addition, we show that the proposed controller can be synthesized by finding the maximal RCIS (or its inner approximation) of an auxiliary system one dimension higher than the original system, using a technique from [17]. Therefore, toolboxes developed for existing safety control methods can be directly applied to synthesize the proposed controllers. Numerical examples show that our method improve the safety of the system outside the maximal RCIS significantly.

In terms of related works, Li et al. in [13] propose a method to extend the domain of reference governer potentially beyond the maximal RCIS by solving a time-optimal control problem when the reference governer is undefined. Shown by the numerical examples in Section IV, our method outperforms the one in [13] significantly when the system is outside the maximal RCIS. If we allow the system of interest to be stochastic, controllers based on the maximal probabilistic RCIS as in [1], [7] can minimize the chance of constraint violation in the infinite horizon even if the system is outside the maximal RCIS, which might be appropriate when the statistics of the disturbance is known. On the other side, our method works without knowing the statistics of the disturbances, and is more computationally tractable since RCISs are much easier to compute than probabilistic RCISs.

Notation: We denote the real line and the set of non-negative numbers by $\mathbb{R}$ and $\mathbb{N}$. For two sets X and Y, $f : X \rightrightarrows$

*This work was supported by NSF grant CNS-1931982 and ONR grant N00014-21-1-2431.

[1] Zexiang Liu, Hao Chen, and Necmiye Ozay are with the Department of Electrical Engineering and Computer Science at University of Michigan, Ann Arbor, MI 48105, USA (e-mail:{zexiang, haochern, necmiye}@umich.edu).

[2] Yulong Gao is with the Department of Computer Science at University of Oxford, Oxford OX1 3QD, UK (e-mail: yulong.gao@cs.ox.ac.uk).

Y denotes a set-valued function from X to Y. The projection of a set X onto its first n coordinates is denoted by $\pi_{1:n}(X)$. The Minkowski sum of two sets X and Y is denoted by $X+Y=\{x+y \mid x\in X, y\in Y\}$. For a singleton set, we denote the sum $\{x\}+Y$ by $x+Y$ for short. For two sets X and Y, their set difference and symmetric set difference are denoted by $X\backslash Y$ and $X\ominus Y := (X\backslash Y)\cup(Y\backslash X)$ respectively. For a matrix $A\in\mathbb{R}^{n\times n}$, a scalar $\alpha\in\mathbb{R}$, and a subset X of $\mathbb{R}^n$, we denote the sets $AX := \{Ax \mid x\in X\}$ and $\alpha X := \{\alpha x \mid x\in X\}$.

II. Preliminaries

We consider a discrete-time linear system Σ

$$\Sigma : x^+ = Ax+Bu+Ed, \tag{1}$$

with state $x\in\mathbb{R}^n$, input $u\in\mathbb{R}^m$, and disturbance $d\in D\subseteq\mathbb{R}^l$. The disturbance set D contains all possible disturbances. Let S_{xu} denote the set of desired state-input pairs, which we call the *safe set* of the system. We assume that both D and S_{xu} are convex polytopes, and moreover D is compact.

A *disturbance model* $\Delta : \mathbb{R}^n \rightrightarrows D$ is a function that assigns a subset $\Delta(x)$ of D to each $x\in\mathbb{R}^n$. Given a disturbance model Δ, if the disturbance input satisfies the constraint $d\in\Delta(x)$ at each time step, we say the disturbance is *generated by* Δ. Given a controller $\mathbf{u}:\mathbb{R}^n\to\mathbb{R}^m$ and a disturbance model Δ, the *k-step forward reachable set* $\mathscr{R}^k_\Sigma(x_0,\mathbf{u},\Delta)$ from the initial state x_0 is defined recursively by

$$\mathscr{R}^{k+1}_\Sigma(x_0,\mathbf{u},\Delta) = \{(x^+,\mathbf{u}(x^+)) \mid \exists (x,u)\in\mathscr{R}^k_\Sigma(x_0,\mathbf{u},\Delta),\\ x^+\in Ax+Bu+E\Delta(x)\}, \tag{2}$$

with $\mathscr{R}^0_\Sigma(x_0,\mathbf{u},\Delta) = \{(x_0,\mathbf{u}(x_0))\}$. Intuitively, $\mathscr{R}^k_\Sigma(x_0,\mathbf{u},\Delta)$ contains all possible state-input pairs reached at time k from x_0 by the closed-loop system when the disturbance d is generated by Δ.

A. Robust Safety Control Framework

Given the system Σ, the safe set S_{xu}, and a disturbance model Δ, the robust safety control problem tries to solve for the set of all the initial states x_0 where there exists $\mathbf{u}:\mathbb{R}^n\to\mathbb{R}^m$ such that

$$\mathscr{R}^k_\Sigma(x_0,\mathbf{u},\Delta)\subseteq S_{xu},\ \forall k\geq 0. \tag{3}$$

Indeed, the maximal set of such initial states is called the *maximal RCIS* C_{max} of Σ with respect to S_{xu} and Δ. There is also an alternative characterization of the maximal RCIS C_{max}: Given the system Σ, the safe set S_{xu}, and a disturbance model Δ, a set $C\subseteq\mathbb{R}^n$ is an RCIS of Σ with respect to S_{xu} and Δ if for all $x\in C$, there exists an input u such that $(x,u)\in S_{xu}$ and $Ax+Bu+E\Delta(x)\subseteq C$. Then, the maximal RCIS C_{max} is the union of all RCIS of Σ with respect to S_{xu} and Δ. Given an RCIS C with respect to S_{xu} and Δ, the *admissible input set* $\mathscr{A}(x,C)$ at a state x is defined as

$$\mathscr{A}(x,C) = \{u \mid (x,u)\in S_{xu},\ Ax+Bu+E\Delta(x)\subseteq C\}. \tag{4}$$

A controller $\mathbf{u}$ satisfies the condition in (3) for any initial state $x_0\in C_{max}$ if and only if for all $x\in C_{max}$,

$$\mathbf{u}(x)\in\mathscr{A}(x,C_{max}). \tag{5}$$

Most of existing works on safety control consider a special case of the robust safety control problem, which we denote as **Problem 1**, where the disturbance model Δ_{all} is such that $\Delta_{all}(x)=D$ for all $x\in\mathbb{R}^n$. We denote the corresponding maximal RCIS as $C_{max,1}$ [5]. As an application of (5), if a reference controller $\mathbf{u}_{ref}$ is given, a robust safety supervisor $\mathbf{u}$ that satisfies (3) with respect to Δ_{all} for all $x_0\in C_{max,1}$ can be synthesized by minimally modifying the reference controller,

$$\mathbf{u}(x) = \min_{u\in\mathscr{A}(x,C_{max,1})} \|u-\mathbf{u}_{ref}(x)\|_2^2,\ \forall x\in C_{max,1}. \tag{6}$$

Note that the robust safety supervisor in (6) is not defined for states outside $C_{max,1}$. In particular, the admissible input set $\mathscr{A}(x,C_{max,1})$ is empty for any $x\notin C_{max,1}$. This is not problematic if the system starts from $C_{max,1}$ and the disturbance is always in D, ensuring that the system stays in $C_{max,1}$ indefinitely. But those assumptions may be unreliable in practice, potentially causing an inadvertent exit from $C_{max,1}$. As a result, the safety control framework described in this subsection is exceedingly susceptible to potential violations of those assumptions.

B. An Opportunistic Safety Control Problem

Let $\mathscr{D}$ be a collection of Borel subsets of the disturbance set D, with $D\in\mathscr{D}$. We call $\mathscr{D}$ the *disturbance template set*. For a given controller $\mathbf{u}$ and an initial state x_0, let $\mathscr{P}(\mathbf{u},x_0)$ be the collection of disturbance models $\Delta:\mathbb{R}^n\to\mathscr{D}$ for which the safety specification in (3) is satisfied, that is

$$\mathscr{P}(\mathbf{u},x_0) := \{\Delta:\mathbb{R}^n\to\mathscr{D} \mid \mathscr{R}^k_\Sigma(x,\mathbf{u},\Delta)\subseteq S_{xu}, \forall k\geq 0\}.$$

We further define $\mathscr{P}(x_0) := \cup_{\mathbf{u}}\mathscr{P}(\mathbf{u},x_0)$, that is the set of disturbance models a controller can possibly be robust to when the system starts at x_0.

With these new notations, Problem 1 can be rephrased as finding $\mathbf{u}$ such that the worst-case disturbance model Δ_{all} is in $\mathscr{P}(\mathbf{u},x_0)$ for a given x_0. Note that Δ_{all} is the worst-case disturbance model in the sense that if the safety specification in (3) is satisfied for Δ_{all}, it is satisfied for any $\Delta:\mathbb{R}^n\to\mathscr{D}$ with respect to the same x_0 and $\mathbf{u}$. Then, it is clear that when Δ_{all} is not contained by $\mathscr{P}(x_0)$ (that is when $x_0\notin C_{max,1}$), Problem 1 has no solutions. Apparently, a better strategy is to find a controller $\mathbf{u}$ robust to the worst-case disturbance model available in $\mathscr{P}(x_0)$ (if Δ_{all} is not)[1]. In this case, we synthesize a controller that is doing its best to keep the system within the safety constraints, as long as $\mathscr{P}(x_0)$ is nonempty.

To formalize this idea, we need to identify the worst-case disturbance model in $\mathscr{P}(x_0)$. Here we use a simple criterion. Let μ be a Borel measure on D. For any disturbance models Δ_1 and Δ_2, we define $\gamma(\Delta_1,\Delta_2)(x) := \sup_{x\in\mathbb{R}^n}|\mu(\Delta_1(x)\ominus\Delta_2(x))|$. This function γ is a pseudometric in the space of disturbance models, measuring the distance between two disturbance models. Then, since we know Δ_{all} is the worst-case disturbance

[1] Under the partial order given by $\Delta_1\leq\Delta_2$ iff $\Delta_1(x)\subseteq\Delta_2(x), \forall x\in\mathbb{R}^n$, it can be shown that Δ_{all} is the unique maximal element in the set of disturbance models, and $\mathscr{P}(\mathbf{u},x)$ is a lower set with respect to this partial order. Thus, being safe against one disturbance model $\Delta\in\mathscr{P}(x_0)$ implies being safe against all disturbance models less than Δ.

model among all disturbance models, we simply consider the nearest point in $\mathscr{P}(x_0)$ to Δ_{all} with respect to γ as the worst-case disturbance model in $\mathscr{P}(x_0)$. As a result, finding a controller robust to the worst-case disturbance model in $\mathscr{P}(x_0)$ is equivalent to finding $\mathbf{u}$ that minimizes the distance $\gamma(\mathscr{P}(\mathbf{u},x_0),\Delta_{all}) := \inf_{\Delta\in\mathscr{P}(\mathbf{u},x_0)}\gamma(\Delta,\Delta_{all})$ between the set $\mathscr{P}(\mathbf{u},x_0)$ and Δ_{all} (by default $\gamma(\mathscr{P}(\mathbf{u},x_0),\Delta_{all}) = +\infty$ if $\mathscr{P}(\mathbf{u},x_0)$ is empty). Based on the above discussion, we pose an opportunistic safety control problem.

Problem 2. *Given the system Σ with its safe set S_{xu}, synthesize a controller $\mathbf{u}^*:\mathbb{R}^n\to\mathbb{R}^m$ such that*

(i) $\Delta_{all}\in\mathscr{P}(\mathbf{u}^*,x)$ for $x\in C_{max,1}$;

(ii) $\mathbf{u}^*$ minimizes[2] the distance between $\mathscr{P}(\cdot,x)$ and Δ_{all} with respect to the pseudometric γ for $x\notin C_{max,1}$.

Point (i) above assures that any solution $\mathbf{u}^*$ to Problem 2 provides safety guarantees as strong as that to Problem 1 when the system operates in the maximal RCIS $C_{max,1}$. When outside $C_{max,1}$, point (ii) assures that $\mathbf{u}^*$ provides extra robustness guarantees compared with solutions to Problem 1. Besides, if $C_{max,1}$ is empty, Problem 1 has no solution, but a solution $\mathbf{u}^*$ to Problem 2 may still exist.

Remark 1. The conservativeness and computational tractability of the solutions $\mathbf{u}^*$ to Problem 2 depend on the measure μ over D and the disturbance template set $\mathscr{D}$. A trivial choice is $\mathscr{D}=\{D\}$, under which Problem 2 degrades to Problem 1 .

III. Construction of $\mathbf{u}^*$

In this section, we show how to construct a solution $\mathbf{u}^*$ to Problem 2. As noted in Remark 1, we need to first specify the measure μ and the disturbance template set $\mathscr{D}$. We choose the measure μ to be the Lebesgue measure on $\mathbb{R}^l$ but restrict it to D. We assume that $\mu(D)>0$ [3]. The disturbance template set is chosen to be

$$\mathscr{D}=\{u_d+\alpha D \mid u_d\in(1-\alpha)D,\ \alpha\in[0,1]\}. \tag{7}$$

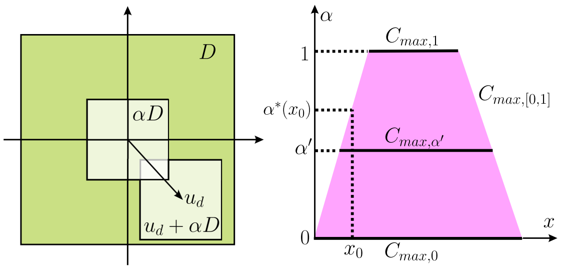
That is, the disturbance template set $\mathscr{D}$ contains all the subsets of D that have the same shape as D. This collection of disturbance sets is rich enough since it contains uncountably many subsets of D scaled to different sizes and positioned at various places, as demonstrated in Fig. 1a. At the same time, $\mathscr{D}$ is simple enough for constructing $\mathbf{u}^*$, which is shown next.

For each $\alpha\in[0,1]$, we define an auxiliary system Σ_α

$$\Sigma_\alpha:\ x^+ = Ax+Bu+E(u_d+d), \tag{8}$$

with A, B, and E the same as in (1), and u_d, $d\in\mathbb{R}^l$. In addition to u, we introduce a new control input u_d. The maximal RCIS of Σ_α with respect to the safe set $S_{xu,\alpha}:=S_{xu}\times(1-\alpha)D$ and the disturbance model $\Delta_\alpha:=\alpha\Delta_{all}$ is denoted by $C_{max,\alpha}$.

Intuitively, in Σ_α, we split the disturbance input in Σ into two parts, namely that $u_d\in(1-\alpha)D$ and $d\in\alpha D$, and turn u_d into a control input. When $\alpha=1$, $C_{max,\alpha}$ is just the maximal RCIS $C_{max,1}$ of Σ with respect to S_{xu} and Δ_{all} defined in



(a) $\mathscr{D}$ and $u_d+\alpha D$ (b) $C_{max,\alpha}$ and $C_{max,[0,1]}$

Fig. 1: Demonstration of the disturbance template set $\mathscr{D}$ in (7) and the maximal RCIS $C_{max,\alpha}$ of Σ_α.

Section II-A. As α goes to 0, Σ_α has more control power and less uncertainty, and thus $C_{max,\alpha}$ monotonically expands as α goes to 0, as demonstrated in Fig. 1b . When $\alpha=0$, we have full control of the disturbances in D. Hence for any initial state x_0 not in $C_{max,0}$, we cannot find a controller and a disturbance model such that (3) is satisfied. In other words, $\mathscr{P}(x_0)$ is empty if and only if $x\notin C_{max,0}$. The following theorem draws a connection between $C_{max,\alpha}$ and solutions $\mathbf{u}^*$ to Problem 2.

Theorem 1. *For any state $x\in C_{max,0}$, let $\alpha^*(x)$ be the maximal $\alpha\in[0,1]$ such that $x\in C_{max,\alpha}$. Then, a controller $\mathbf{u}^*$ is a solution to Problem 2 if and only if for all $x\in C_{max,0}$, there exists $u_d\in\mathbb{R}^l$ such that*

$$(\mathbf{u}^*(x),u_d)\in\mathscr{A}(x,C_{max,\alpha^*(x)}), \tag{9}$$

where $\mathscr{A}(x,C_{max,\alpha^*(x)})$ is the admissible input set of the system Σ_α with respect to $S_{xu,\alpha}$ and Δ_α, with $\alpha=\alpha^*(x)$. In addition, the distance between $\mathscr{P}(\mathbf{u}^*,x)$ and Δ_{all} satisfies

$$\gamma(\mathscr{P}(\mathbf{u}^*,x),\Delta_{all}) = \begin{cases}(1-\alpha^*(x)^l)\mu(D) & x\in C_{max,0},\\ +\infty & o.w.\end{cases} \tag{10}$$

The proof of Theorem 1 is in the appendix. Intuitively, if a state x is in $C_{max,\alpha}$, we can find $\mathbf{u}:\mathbb{R}^n\to\mathbb{R}^m$ and $\mathbf{u}_d:\mathbb{R}^n\to\mathbb{R}^l$ such that the disturbance model $\mathbf{u}_d+\Delta_\alpha$ is in $\mathscr{P}(\mathbf{u},x)$. It can be shown that by taking $\alpha=\alpha^*(x)$, $\mathbf{u}_d+\Delta_\alpha$ is actually the worst-case disturbance model in $\mathscr{P}(x_0)$ and is contained by $\mathscr{P}(\mathbf{u}^*,x)$ for any $\mathbf{u}^*$ satisfying (9). Furthermore, when $\alpha^*(x)=1$, $\mathbf{u}_d+\Delta_\alpha=\Delta_{all}$. Thus, both points in Problem 2 are fulfilled by $\mathbf{u}^*$. Applying Theorem 1, given a reference controller $\mathbf{u}_{ref}$, we propose an opportunistic safety supervisor

$$\mathbf{u}(x) = \min_{(u,u_d)\in\mathscr{A}(x,C_{max,\alpha^*(x)})}\|u-\mathbf{u}_{ref}(x)\|_2^2 \tag{11}$$

This opportunistic safety supervisor is defined over $C_{max,0}$, larger than the domain $C_{max,1}$ of the robust safety supervisor in (6). Recall that when x_0 is not in $C_{max,0}$, it becomes inevitable to violate the safety constraints no matter what the controller and the disturbance do. The opportunistic safety supervisor becomes undefined only in this extreme case.

Remark 2. Our results also hold for Problem 2 with discretized α: Let α_i for i from 0 to N be an increasing sequence of scalars such that $0=\alpha_0<\alpha_1<\cdots<\alpha_N=1$. Let

$$\mathscr{D}=\{u_d+\alpha_i D \mid u_d\in(1-\alpha_i)D, i\in\{0,\cdots,N\}\}. \tag{12}$$

[2] Point (ii) does not necessarily imply point (i) since $\gamma(\mathscr{P}(\mathbf{u},x),\Delta_{all})=0$ does not imply $\Delta_{all}\in\mathscr{P}(\mathbf{u},x)$ (unless $\mathscr{P}(\mathbf{u},x)$ is closed).

[3] Otherwise, D lies in a subspace of $\mathbb{R}^l$, which implies we should lower l.

We redefine $\alpha^*(x)$ to be the maximal α_i such that $x \in C_{max,\alpha_i}$. Theorem 1 with this redefined $\alpha^*(x)$ holds for the disturbance set $\mathscr{D}$ in (12). The only benefit for using $\mathscr{D}$ in (12) instead of (7) is that to construct $\mathbf{u}^*$, we only need $C_{max,\alpha}$ for finitely many α (versus a continuum of α in $[0,1]$).

A. The One-shot Computation of $C_{max,\alpha}$

In this subsection, we show how to compute $C_{max,\alpha}$ for all $\alpha \in [0,1]$ in one shot. Consider a new auxiliary system $\Sigma_{[0,1]}$

$$\Sigma_{[0,1]}: \begin{bmatrix} x^+ \\ \alpha^+ \end{bmatrix} = \begin{bmatrix} Ax + Bu + E(u_d + d) \\ \alpha \end{bmatrix}, \tag{13}$$

where we introduce a new state $\alpha \in [0,1]$ and a new control input $u_d \in \mathbb{R}^l$. Define the safe set $S_{xu,[0,1]}$ of $\Sigma_{[0,1]}$ by

$$S_{xu,[0,1]} = \{(x,\alpha,u,u_d) \mid (x,u,u_d) \in S_{xu,\alpha}, \alpha \in [0,1]\}. \tag{14}$$

Let $\Delta_{[0,1]}$ be the disturbance model such that $\Delta_{[0,1]}(x,\alpha) = \alpha D$ for all $(x,\alpha) \in \mathbb{R}^n \times [0,1]$. We denote the maximal RCIS of $\Sigma_{[0,1]}$ with respect to $S_{xu,[0,1]}$ and $\Delta_{[0,1]}$ by $C_{max,[0,1]}$. Since S_{xu} and D are both polytopes, it can be shown that $S_{xu,[0,1]}$ is a polytope and the maximal RCIS $C_{max,[0,1]}$ can be approximated by the standard iterative method [14], [17]. Once we have $C_{max,[0,1]}$, $C_{max,\alpha'}$ is just equal to the slice of $C_{max,[0,1]}$ through $\alpha = \alpha'$, for any $\alpha' \in [0,1]$, as shown in Fig. 1b . Furthermore, given $C_{max,[0,1]}$ and x, the value $\alpha^*(x)$ can be easily obtained by solving a linear program.

Remark 3 (Computational cost). Compared with the robust safety control framework, our method needs to compute the maximal RCIS $C_{max,[0,1]}$ for a system one dimension higher than the original system Σ (cf., (13) vs. (1)) and thus has a higher offline computational cost.

At runtime, given the current state x and $C_{max,[0,1]}$, we first solve one linear program to check if $x \in C_{max,0}$ and find $\alpha^*(x)$, and then solve the quadratic program in (11). For a comparison, the robust safety control framework solves one linear program to check if $x \in C_{max,1}$ and then solve the quadratic program in (6). The runtime computational cost of the two frameworks should be similar.

Remark 4. If the maximal RCIS $C_{max,[0,1]}$ cannot be computed exactly, one can use any controlled invariant inner approximation of $C_{max,[0,1]}$ in (11), with the cost of extra conservativeness.

IV. Numerical Examples

The maximal RCISs $C_{max,[0,1]}$ in the examples are computed using MPT3 [11] equipped with GUROBI [10] in MATLAB. The code and video can be accessed from `https://haochern.github.io/OpSafe/`.

A. Adaptive Cruise Control

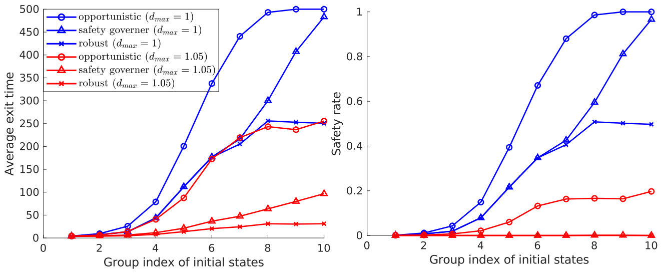
We consider the car-following example in [13]. The goal is to maintain the relative distance Δs and the relative velocity Δv between the ego vehicle and the front vehicle within a safe range. The system is modeled by a discretized double integrator with states $x = (\Delta s, \Delta v)$. The model parameters can be found in [13]. The control input and the disturbance are the acceleration u of the ego vehicle and the acceleration $d \in [-d_{max}, d_{max}]$ of the front vehicle respectively. The safe set is given by $|\Delta s - 15| \leq 5$, $|\Delta v| \leq 5$, and $|u| \leq 2$.

The reference controller $\mathbf{u}_{ref} = 0.2842\Delta s + 0.8056\Delta v$, with a saturation limit at ±2. We implemented the robust and the opportunistic safety supervisors in (6) and (11) and the safety protection and extension governer in [13], assuming $d_{max} = 1$. To evaluate those three safety supervisors at states with different values of $\alpha^*(x)$, we generated 10 groups $\mathscr{X}_{0,i}$ of initial states, where $\mathscr{X}_{0,i}$ contains 1000 states uniformly sampled in $C_{max,0.1*i} \backslash C_{max,0.1*(i+1)}$ for i from 0 to 9. That is, each $x_0 \in \mathscr{X}_{0,i}$ has $\alpha^*(x_0)$ between $0.1i$ and $0.1(i+1)$. Note that $\mathscr{X}_{0,i}$ is disjoint from the maximal RCIS $C_{max,1}$ for all i, since we want to evaluate how the controllers perform outside the maximal RCIS.

For each initial state x_0 in $\mathscr{X}_{0,i}$, we generate a random disturbance sequence in $[-d_{max}, d_{max}]$ and then run simulations for each safety supervisor for 500 steps. During the simulation, if a safety supervisor becomes undefined, we switch to the reference controller. Thus, for each group index i and d_{max}, we have 1000 trajectories starting from $\mathscr{X}_{0,i}$ under each safety supervisor. We evaluate the performance with two metrics: the *average exit time* the system first exits the safe set (taken to be 500 when the system never exits S_{xu}), and the *safety rate*, the ratio of trajectories remaining in S_{xu} through the entire simulation period out of a total of 1000 trajectories. The average exit time and the safety rates of the three safety supervisors for $d_{max} = 1$ and 1.05 are shown in Fig. 2. First note that both metrics of all the safety supervisors grow with the group index i. This is expected since the initial states with a higher value of $\alpha^*(x)$ have worst-case disturbances in $\mathscr{P}(x_0)$ closer to Δ_{all} and thus are easier to be kept within the safe set. Comparing curves in different colors in Fig. 2, we observe that the performance of the safety supervisors degrade as the disturbances is sampled in a range larger than the assumed one. Finally, comparing curves in the same color, we observe that the proposed safety supervisor outperforms the safety supervisors in (6) and in [13] in both metrics for all groups of initial states and all d_{max}. In particular, when unexpected disturbances appear (by increasing d_{max} from 1 to 1.05), the proposed safety supervisor has much larger average exit time than the other two, as shown in Fig.2a, and is the only one among the three that has nonzero safety rates, as shown in Fig. 2b, showing that the proposed method enhances the safety of the system significantly when the system operates outside the maximal RCIS.

B. Lane Keeping Control

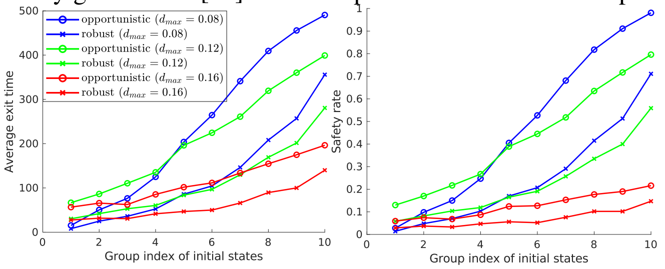
We consider a highway driving scenario where we want to keep the lateral position of a vehicle within given lane boundaries. We use the 4-dimensional linearized bicycle model in [16] with respect to the constant longitudinal velocity $30m/s$, discretized with time step $0.1s$. The states are the lateral displacement y, the lateral velocity v, the yaw angle $\Delta\Psi$, and the yaw rate r. The control input is the steering angle u. The safe set is given by constraints $|y| \leq 0.9$, $|v| \leq 1.2$, $|\Delta\Psi| \leq 0.05$, $|r| \leq 0.3$, and $|u| \leq \pi/2$. The disturbance of the system is the road curvature d with $|d| \leq d_{max}$.



(a) Average exit time

(b) Safety rate

Fig. 2: The average exit time and safety rates of the safety supervisors in (6) (robust) and (11) (opportunistic), and the safety governer in [13] in the adaptive cruise control example.



(a) Average exit time

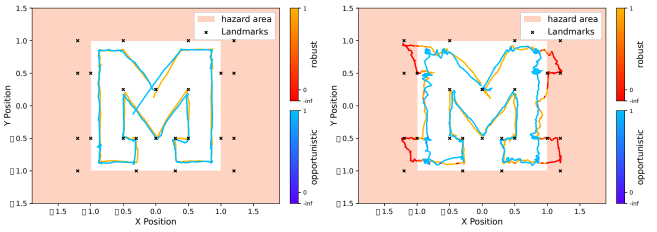
(b) Safety rate

Fig. 3: The average exit time and the safety rate of the robust safety supervisor in (6) and the opportunistic safety supervisor in (11) in the lane keeping example.

The reference controller $\mathbf{u}_{ref}$ is $\mathbf{u}_{ref} = -Kx$ subject to a saturation limit at $\pm\pi/2$, where K is determined through solving an LQR problem (with $Q = I$ and $R = 0$). Then, we implement the proposed safety supervisor in (11) and the robust safety supervisor in (6), assuming $d_{max} = 0.08$. For this example, the safety governer in [13] is infeasible for all x and thus is excluded. We assess the safety supervisors in the same manner as in Section IV-A. Fig. 3 illustrates the average exit time and the safety rate for both safety supervisors under simulations with $d_{max} = 0.08$, 0.12, and 0.16. Similar to the previous example, the performance of the safety supervisors is improved as the initial states have a higher value of $\alpha^*(x)$, and degrades as the disturbance range used in the simulation exceeds that used in control synthesis. Comparing curves in the same color in Fig. 3 , the proposed safety supervisor consistently outperforms the robust safety supervisor across all groups of initial states and all d_{max}. Notably, the performance of our approach at $d_{max} = 0.12$ (50% larger than the assumed d_{max}) is even better than the performance of the robust safety supervisor at $d_{max} = 0.08$, highlighting its enhanced safety and resilience to unexpected disturbances when the system operates beyond the maximal RCIS $C_{max,1}$.

C. Safe Tracking for Aerial Vehicle

We tested our approach on the drone platform Crazyflie 2.1 in a task of cruising around designated waypoints in the horizontal plane while avoiding entering hazardous region (red region in Fig. 4). We use a built-in controller to keep the altitude of the drone constant, and then control its horizontal motion by sending the reference velocities u_x and u_y in the x, y axes to a lower-level controller in the period of $0.1s$. Validated by the flight data, the drone dynamics in x and y axes under the lower-level controller are decoupled, homogeneous, and linear. Thus, we model the dynamics in x and y by two identical 3-dimensional linear systems with states $s_x = (x, v_x, u_{x,-1})$ (or $s_y = (y, v_y, u_{y,-1})$, respectively), where x and v_x are the position and velocity in x axis, and $u_{x,-1}$ is the previous reference velocity u_x. The system matrices are learned from data via least square with the disturbance set the convex hull of the prediction error. The safe set of Σ_x is given by constraints $|x| \leq 1$, $|v_x| \leq 1$, $|u_x| \leq 1$, and $|u_{x,-1} - u_x| \leq 0.5$. The setup for the system in y axis is the same.



(a) No measurement noise

(b) With measurement noise

Fig. 4: The drone trajectories in x-y plane under the safety supervisors in (6) (yellow-red) and (11) (blue-dark blue). The color of the curves reflects the value of $\alpha^*(x)$ along the trajectories.

We synthesize one reference tracking controller for each subsystem in form of $\mathbf{u}_{ref,*} = -K(s_* - s_{ref,*})$ ($*$ is a placeholder for x or y) subject to a saturation limit of ± 1, where K is determined through solving a LQR problem (with $Q = 3I$ and $R = I$). We then implement the safety supervisors in (6) and (11) to supervise $\mathbf{u}_{ref,*}$. For the experiments, we pick 24 waypoints to form a big "M", as shown by the checkmarks in Figure 4. Part of the waypoints is picked outside the safe region such that the reference controller without any supervision would steer the drone to the unsafe region. During the experiments, we switch to the reference controller whenever the safety supervisor is undefined. Since the drone is initialized within $C_{max,1}$, both safety supervisors are able to maintain the drone within the safe region, as shown by Fig. 4a. To make this task more challenging, we repeat this experiment with the state measurements corrupted by an additional Gaussian noise with standard deviation 0.05. Subject to this unexpected measurement noise, our opportunistic safety supervisor still successfully keeps the drone within the safe region, while the robust safety supervisor in (6) fails as shown in Fig. 4b.

V. Conclusion

In this work, we present an opportunistic safety control framework that extends the domain of safety controllers beyond the maximal RCIS. This is achieved by designing a controller that is robustly safe against as much disturbance as possible. Our approach can be trivially extended for nonlinear systems, which we consider in future. We also want to extend these results to probabilistic settings, by using a given or learned probability measure μ instead of the Lebesgue measure used in this work.

Appendix

A. Proof of Theorem 1

In this section, we denote the distance $\gamma(\mathscr{P}(\mathbf{u},x),\Delta_{all})$ between $\mathscr{P}(\mathbf{u},x)$ and Δ_{all} by $r(x,\mathbf{u})$ for short.

Lemma 1. *For any given $(x,\alpha)\in\mathbb{R}^{n+1}$, the minimal distance $\inf_{\mathbf{u}} r(x,\mathbf{u})$ at x is less than or equal to $(1-\alpha^l)\mu(D)$ if and only if $x\in C_{max,\alpha}$. Furthermore, a controller $\mathbf{u}$ satisfies $r(x,\mathbf{u})\leq(1-\alpha^l)\mu(D)$ for all $x\in C_{max,\alpha}$ if and only if for all $x\in C_{max,\alpha}$,*

$$\mathbf{u}(x)\in\mathscr{A}(x,C_{max,\alpha}), \tag{15}$$

where $\mathscr{A}(x,C_{max,\alpha})$ is the admissible input set of Σ_α.

Proof. We first show the "if" direction. Pick an arbitrary $x\in C_{max,\alpha}$. Since $C_{max,\alpha}$ is the maximal RCIS of Σ_α, there exist $\mathbf{u}:\mathbb{R}^n\rightarrow\mathbb{R}^m$ and $\mathbf{u}_d:\mathbb{R}^n\rightarrow(1-\alpha)D$ such that

$$\mathscr{R}^k_{\Sigma_\alpha}((x,\alpha),(\mathbf{u},\mathbf{u}_d),\Delta_\alpha)\subseteq S_{xu,\alpha},\ \forall k\geq 0. \tag{16}$$

Define the disturbance model $\Delta(\bar{x}):=\mathbf{u}_d(\bar{x})+\alpha D\in\mathscr{D}$ for all $\bar{x}\in\mathbb{R}^n$. By the construction of Σ_α and $S_{xu,\alpha}$, (16) implies that $\mathscr{R}^k_\Sigma(x,\mathbf{u},\Delta)\subseteq S_{xu}$ for all $k\geq 0$. That is, $\Delta\in\mathscr{P}(\mathbf{u},x)$. Hence,

$$\inf_{\bar{\mathbf{u}}} r(x,\bar{\mathbf{u}}) \leq r(x,\mathbf{u})\leq\gamma(\Delta,\Delta_{all})=(1-\alpha^l)\mu(D), \tag{17}$$

where the last equality uses the property of Lebesgue measure $\mu(\alpha D)=\alpha^l\mu(D)$ (recall that $D\subseteq\mathbb{R}^l$). Also, by Section II-A, we know that $\mathbf{u}$ satisfies (15). Hence, we proved the "if" direction of both statements in Lemma 1.

Next, we pick an arbitrary (x,α) such that $\inf_{\bar{\mathbf{u}}} r(x,\bar{\mathbf{u}})\leq(1-\alpha^l)\mu(D)$. By the definition of $r(x,\mathbf{u})$, for any integer $i\geq 1/\alpha$, there exist $\mathbf{u}_i:\mathbb{R}^n\rightarrow\mathbb{R}^m$ and $\Delta_i:\mathbb{R}^n\rightarrow\mathscr{D}$ such that $\Delta_i\in\mathscr{P}(\mathbf{u}_i,x)$ and $\gamma(\Delta_i,\Delta_{all})<(1-\alpha_i^l)\mu(D)$, with $\alpha_i:=\alpha-1/i>0$. By the definition of $\mathscr{D}$ in (7), there exists $\mathbf{u}_d:\mathbb{R}^n\rightarrow(1-\alpha_i)D$ such that $\mathbf{u}_d(\bar{x})+\alpha_i D\subseteq\Delta_i(\bar{x})$ for all $\bar{x}\in\mathbb{R}^n$. Hence, $\Delta_i\in\mathscr{P}(\mathbf{u},x)$ implies that the disturbance model $\mathbf{u}_d+\alpha_i D$ is in $\mathscr{P}(\mathbf{u},x)$ as well. That is, $\mathscr{R}^k_\Sigma(x,\mathbf{u}_i,\mathbf{u}_d+\alpha_i D)\subseteq S_{xu}$ for all $k\geq 0$, which is further equivalent to

$$\mathscr{R}^k_{\Sigma_{\alpha_i}}(x,(\mathbf{u}_i,\mathbf{u}_d),\Delta_{\alpha_i})\subseteq S_{xu,\alpha_i},\ \forall k\geq 0. \tag{18}$$

By definition, (18) implies that $x\in C_{max,\alpha_i}$, that is, $(x,\alpha_i)\in C_{max,[0,1]}$. Since $C_{max,[0,1]}$ is closed, we know that $(x,\alpha)=\lim_{i\rightarrow\infty}(x,\alpha_i)\in C_{max,[0,1]}$ as well. This completes the proof for the first statement in Lemma 1.

Now suppose that $\mathbf{u}$ is a controller satisfying $r(x,\mathbf{u})\leq(1-\alpha^l)\mu(D)$ for all $x\in C_{max,\alpha}$. We pick an arbitrary $x\in C_{max,\alpha}$. Clearly, $(x,\mathbf{u}(x))\in S_{xu}$. For all $i\geq 1/\alpha$, there exists $\Delta_i:\mathbb{R}^n\rightarrow\mathscr{D}$ such that $\Delta_i\in\mathscr{P}(\mathbf{u},x)$ and $\gamma(\Delta_i,\Delta_{all})<(1-\alpha_i^l)\mu(D)$. Thus, for all $d\in\Delta_i(x)$,

$$\mathscr{R}^k_\Sigma(Ax+B\mathbf{u}(x)+Ed,\mathbf{u},\Delta_i)\subseteq S_{xu},\ \forall k\geq 0, \tag{19}$$

which implies that $r(Ax+B\mathbf{u}(x)+Ed,\mathbf{u})\leq(1-\alpha_i^l)\mu(D)$ for all $d\in\Delta_i(x)$. Based on the first statement of Lemma 1, we have $(Ax+B\mathbf{u}(x)+Ed,\alpha_i)\in C_{max,[0,1]}$ for all $d\in\Delta_i(x)$. Since $\gamma(\Delta_i,\Delta_{all})<\mu(D)-\mu(\alpha_i D)$, there exists $u_{d,i}\in(1-\alpha_i)D$ such that $u_{d,i}+\alpha_i D\subseteq\Delta_i(x)$. Thus, we have $(Ax+B\mathbf{u}(x)+E(u_{d,i}+d),\alpha_i)\in C_{max,[0,1]}$ for all $d\in\alpha_i D$. Since D is compact, there exists a subsequence of $u_{d,i}$ that converges to a point $u_d\in(1-\alpha)D$. We abuse the notation a bit and denote this subsequence by $u_{d,i}$ again. Then, we have $Ax+B\mathbf{u}(x)+E(u_d+\alpha D)\subseteq C_{max,\alpha}$ and $(x,\mathbf{u}(x))\in S_{xu}$, which implies $\mathbf{u}(x)\in\mathscr{A}(x,C_{max,\alpha})$. □

Proof of Theorem 1 . Point (i) of Problem 2 is trivially satisfied by u^* since (9) implies (5) with respect to $C_{max,1}$ for $x\in C_{max,1}$. For point (ii), according to (10), a controller $\mathbf{u}^*$ minimizes $r(x,\cdot)$ for all $x\in\mathbb{R}^n$ if and only if $r(x,\mathbf{u}^*)\leq(1-\alpha^*(x)^l)\mu(D)$ for all $x\in C_{max,0}$, which is equivalent to the condition in (15) due to Lemma 1. Equation (10) is just a direct application of Lemma 1. □

References

[1] A. Abate, M. Prandini, J. Lygeros, and S. Sastry. Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica*, 44(11):2724–2734, 2008.

[2] B. Aminof, G. De Giacomo, A. Lomuscio, A. Murano, S. Rubin, et al. Synthesizing strategies under expected and exceptional environment behaviors. In *IJCAI*, pages 1674–1680. ijcai. org, 2020.

[3] J.-P. Aubin. A survey of viability theory. *SIAM Journal on Control and Optimization*, 28(4):749–788, 1990.

[4] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin. Hamilton-jacobi reachability: A brief overview and recent advances. In *CDC*, pages 2242–2253. IEEE, 2017.

[5] D. Bertsekas. Infinite time reachability of state-space regions by using feedback control. *IEEE Trans. Autom. Control*, 17(5):604–613, 1972.

[6] B. Finkbeiner and N. Passing. Synthesizing dominant strategies for liveness. In *42nd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[7] Y. Gao, K. H. Johansson, and L. Xie. Computing probabilistic controlled invariant sefts. *IEEE Trans. Autom. Control*, 66(7):3138–3151, 2020.

[8] Y. Gao, C. Liu, and K. H. Johansson. Robust risk-aware model predictive control of linear systems with bounded disturbances. In *CDC*, pages 1148–1155. IEEE, 2022.

[9] E. Garone, S. Di Cairano, and I. Kolmanovsky. Reference and command governors for systems with constraints: A survey on theory and applications. *Automatica*, 75:306–328, 2017.

[10] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2022.

[11] M. Herceg, M. Kvasnica, C. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, pages 502–510, Zürich, Switzerland, July 17–19 2013.

[12] M. Jankovic. Robust control barrier functions for constrained stabilization of nonlinear systems. *Automatica*, 96:359–367, 2018.

[13] N. Li, Y. Li, and I. Kolmanovsky. A unified safety protection and extension governor. *arXiv preprint arXiv:2304.07984*, 2023.

[14] P. Nilsson, O. Hussien, A. Balkan, Y. Chen, A. D. Ames, J. W. Grizzle, N. Ozay, H. Peng, and P. Tabuada. Correct-by-construction adaptive cruise control: Two approaches. *IEEE Transactions on Control Systems Technology*, 24(4):1294–1307, 2015.

[15] S. Samuel, K. Mallik, A.-K. Schmuck, and D. Neider. Resilient abstraction-based controller design. In *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, pages 1–2, 2020.

[16] S. W. Smith, P. Nilsson, and N. Ozay. Interdependence quantification for compositional control synthesis with an application in vehicle safety systems. In *CDC*, pages 5700–5707. IEEE, 2016.

[17] L. Yang, D. Rizzo, M. Castanier, and N. Ozay. Parameter sensitivity analysis of controlled invariant sets via value iteration. In *ACC*, pages 4737–4744. IEEE, 2020.