Preprints, 10th IFAC International Symposium on Advances in Automotive Control Columbus, Ohio, USA, August 28-30, 2022

Justifying Emergency Drift Control for Automated Vehicles

Tong Zhao [*†] **Ekim Yurtsever** [* †] **Giorgio Rizzoni** [*]

[*] *Center for Automotive Research, The Ohio State University, 930 Kinnear Road, Columbus, OH 43212, USA (e-mail: [zhao.1991,yurtsever.2,rizzoni.1]@osu.edu).*

Abstract: Expert human drivers can execute emergency steering actions to avoid sudden events like a deer crossing the road. However, justifying beyond-the-limit emergency maneuvering for automated driving systems is exceptionally challenging. Emergency maneuvering often requires non-linear control policies without stability guarantees. Liability concerns, ethics, lack of safety guarantees, and non-linear system dynamics convolute an already complicated problem. Against this backdrop, we propose a principled approach to justify a particular type of emergency steering in safety-critical situations. A limit-handling controller is justified and deployed to execute the emergency maneuver upon a conventional controller's formally verified incapability to handle. We claim this check justifies the execution of the emergency maneuver as we show failure is mathematically inevitable otherwise. The simulation-based experimental validation shows that using backward reachability analysis, the proposed approach can determine emergencies. The validation justifies using limit-handling controllers for collision avoidance in a scenario where the baseline controllers fail catastrophically.

Keywords: Automated vehicles, safety guarantee, vehicle maneuver, reachability analysis, formal verification

1. INTRODUCTION

Safety is a major concern holding back the widespread deployment of Automated Driving Systems (ADS). Serious issues have been raised by the media, authorities, and academia [Shafaei et al. (2018)]. Contrary to human drivers whose liability could be accounted for by existing law and insurance, the liability of ADS remains unsettled with a myriad of ethical and legal complications.

The need for safety guarantees motivates ADS developers to conduct rigorous design studies. However, we argue that the need for safety guarantees often leads to over-conservative solutions that compromises vehicle performance. Mathematical safety guarantees are often confined to well-defined and conservative control regions. A common example of such a limit is the vehicle maximum yaw rate, which is regulated for getting stability guarantees. However, this may exclude certain beyond-the-limits operations that have the only non-zero chance of saving the vehicle from a collision in emergency situations, i.e., operations that are technically the best to do in an emergency. In other words, having these conservative control constraints is good for liability report but limits the capabilities of the ADS in challenging situations. For example, a sudden deer crossing can quickly turn a safe driving scene into a emergency setting unexpectedly, where only extreme steering control beyond the conservative constraints may lead to collision avoidance. Nevertheless, the common practice is to ignore these edge cases, as extreme situations are rare and generally out of liability concerns.

We approach this problem from a reverse angle: when faced with an edge case, we first try to get a mathematical guarantee that a collision is imminent, if and only if we can get this guarantee, then we seek alternative controllers that have the capability to handle the situation (Fig. 1). We argue that this backward reachability check justifies a limited excursion beyond the conservative control region.

Safety verification can be performed either using sample-based or formal methods[Zhao et al. (2022)]. Backward reachability (BR) analysis is a formal approach for safety verification for dynamical systems [Chen and Tomlin (2018)]. It provides a distribution of states that will absolutely end up in a defined target set of states within certain moments into the future. The guarantee in this verification comes from the fact that BR solves Hamilton-Jacobi-Issac equations to exploit the worst of all possible actions of adversarial disturbance and all possible beneficial ego actions.

Models define belief on what is possible and what is not. Models used for vehicle planning, control and simulation often come in different levels of fidelity. Certain aspects of the modeling (such as tire model) have major influence on the model dynamics outcome for aggressive maneuvers, while some other dynamics (such as yaw dynamics) are more consistent across different modelling levels [Berntorp et al. (2014)]. Also certain high fidelity models allow the capturing of particular vehicle features such as controllable limited-slip differential and individual wheel braking [Subosits and Gerdes (2021)]. In order to make mathematical guarantees on vehicle control, we argue that the different modelling fidelity for planning, control and

[*] † Equal contribution

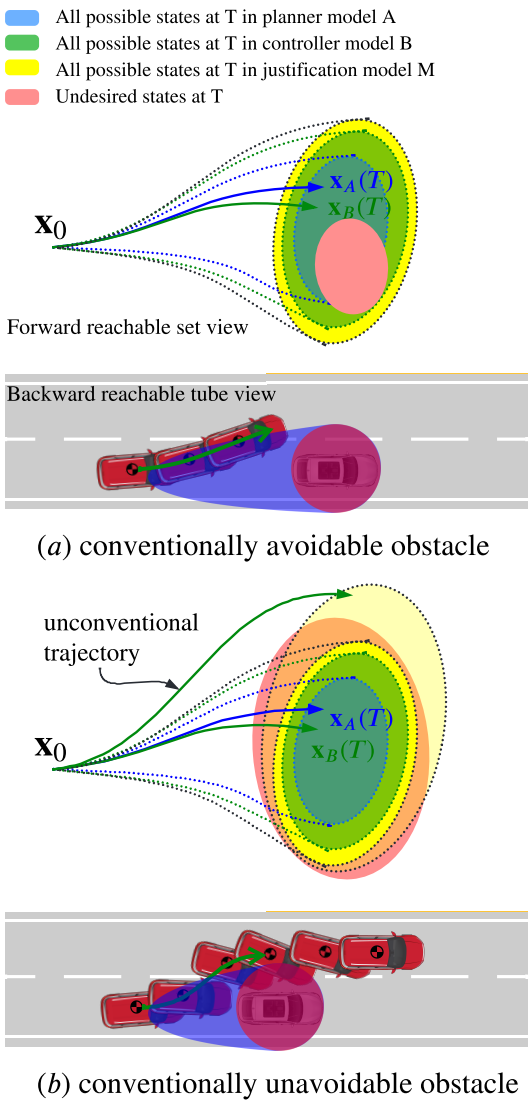

Fig. 1. Overview of emergency maneuver justification. In the backward reachable tube views, the red areas show collision states to avoid and the blue areas are the backward reachable tube of the corresponding collision states at the beginning. The forward reachable set view is an alternative way to visualize the situation: when the obstacle is conventionally avoidable as in (a), the planner, controller and justification models can produce states outside of undesired state set (red area); when the obstacle is conventionally unavoidable as in (b), none of the planner, controller or justification models can produce states outside of undesired state set. However, with a new justification model for the beyond-the-limit controls (transparent yellow area), some safe trajectories can be found.

simulation should be bridged in a certain manner to allow for comparison or judgement. We use differential inclusion as such a means to allow certain assertions to be made on the feasibility of controllers for a specific scenario.

Here we propose a principled approach for justifying emergency maneuvering in emergency situations using BR analysis. The main objective is to ensure that the controller never leaves the conventional control domain unless it is absolutely necessary. To this end, we first use the formal theoretical guarantees of backward reachability to determine whether the system is in an avoidable collision situation. After this step, we use a principled approach to look for a justifiable emergency maneuver beyond the conventional control-based vehicle model. Our analysis shows that it is possible to avoid an accident otherwise unavoidable with this beyond-the-limit control deployment. Experimental validation shows that the proposed framework is safer in emergency situations compared to baselines.

The main contributions can be summarized as follows:

- An investigation of relationship between models used in planner, controller and simulated environments.
- A formal definition of an emergency situation in driving automation using Backward Reachability (BR).
- Justifying whether the current controller (or any controller based on the same system dynamics model) will lead to unavoidable collision by BR analysis.
- A principled approach to justify beyond-the-limit controllers designed under a different set of system dynamics. Our algorithm guarantee that the new controller has the capability to turn the scenario from unavoidable to avoidable.

2. RELATED WORK

Beyond the Limit Driving with Drifting. Besides being a motor-sport stunt, vehicle drifting has been rigorously studied by academia for the past two decades [Ono et al. (1998)]. Drift equilibrium [Velenis et al. (2009)] and various other techniques have been proposed to sustain drift motion, including feedback linearization [Voser et al. (2010)], model inversion [Hindiyeh and Gerdes (2010); Goh et al. (2018)], model predictive control [Acosta et al. (2018); Arab and Yi (2020)], and reinforcement learning [Cai et al. (2020)]. The next challenge would be to handle the transition of such drift motion to other more conservative driving styles. Directly solving nonlinear programming (NLP) problem [Goh (2019)] and feedback-feedforward control [Zhao et al. (2021)] are among such possible solutions. With justified high sideslip motion, road applications such as post-collision recovery [Yin et al. (2020)] can be considered.

However, these beyond-the-limit vehicle drift controllers have not been justified for road use, as they are deemed too risky to be deployed in real traffic.

Backward Reachability Analysis. Backward reachability analysis performs formal verification checks of whether system can reach or avoid certain target sets of states in the near future. The absolute sureness in the verification is guaranteed by the fact that BR solves Hamilton-Jacobi-Issac equations to exploit the worst of all possible actions of adversarial disturbance and all possible beneficial ego actions. BR has been applied to managing collision-free multiple unmanned aerial vehicle (UAV) planning [Chen et al. (2017)], as well as to motion planning of multirotor robots [Seo et al. (2020)], and for vehicle safe parking [Jiang et al. (2020)]. However, BR suffers from the "curse of dimensionality" and has not been used for justifying emergency maneuvering for autonomous driving,

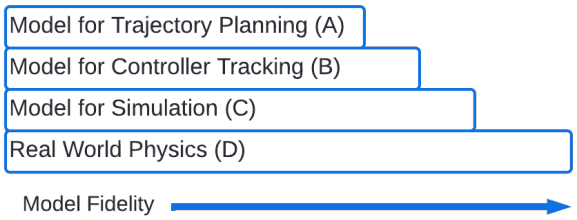

Fig. 2. The model fidelity pyramid: the width qualitatively represent the fidelity level

where enhanced maneuvering skill is equally important as assurance for maneuvering safety. The usual practice for ensuring driving safety is overestimating risks and setting conservative control constraints. In light of the potential of BR, we focus on using lower dimensional BR to search and find a capable maneuver. BR can be applied to the problem of autonomous vehicle planning and control to justify when beyond-the-limit driving is absolutely necessary.

3. MODELS FOR PLANNING AND CONTROL

The operation of specifying and achieving motion tasks for the automated vehicle system is referred to as planning and control. Often the task of planning and the task of control are separated as sequential procedures [Claussmann et al. (2020)]. Here we investigate the hierarchical relationship between planning, control, and the models used for them, and how this relationship interpreted in different forms can produce certain affirmative assertions towards possible planning and control outcome.

3.1 Model Fidelity Hierarchy

Following the commonly accepted sequential relation between planning and control, the different models involved in the vehicle planning control action is illustrated in Fig. 2, where a relatively simple and fast model is used for the planning, and a properly fitted model with relatively low dimensions is used for control. Then in virtual simulation platforms, usually a high fidelity vehicle model is used to simulate the outcome of applying planning and control strategies in targeted scenarios. Such simulations however, still have a fidelity gap between them and the real world physical driving.

3.2 Model-based Assertions

Models define belief on what is possible and what is not. There are limited assertions that we can make about the outcome of planning and control due to the fact that "no models are correct" with respect to the physical world. However, by assigning certain credibility to simulation realism, and certain credibility to the approximate "correctness" of planning and control models, we can generalize some interesting arguments. First, we make certain assumptions to lay the grounds for discussion:

Assumption 1. The model used for planning (A) is an under-approximation of the controller used model (B). In dynamical system terms, this indicates that the exact dynamic inclusion of the controller-used model (B) is always a superset of the exact differential inclusion of the planner used model (A):

$$F_A(\mathbf{x}, t) \subset F_B(\mathbf{x}, t), \forall \mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R} \tag{1}$$

where the exact differential inclusion of a dynamical model A: $\dot{\mathbf{x}} = f_A(\mathbf{x}, \mathbf{u}, t)$ is the exact set of all possible state derivatives $\dot{\mathbf{x}}$ given an arbitrary state $\mathbf{x}$ and any allowed control input $\mathbf{u} \in \mathcal{U}$: $F_A(\mathbf{x}, t) = \{\dot{\mathbf{x}} \in \mathbb{R}^n | \dot{\mathbf{x}} = f_A(\mathbf{x}, \mathbf{u}, t), \forall \mathbf{u} \in \mathcal{U}\}$.

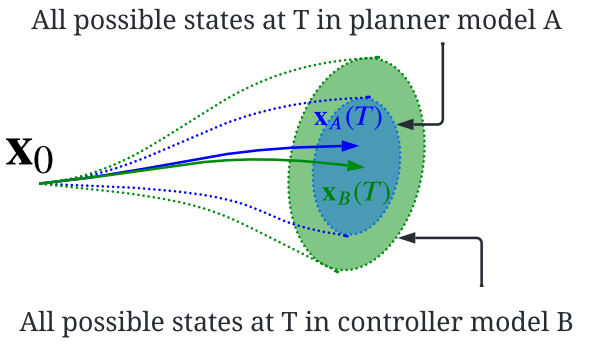

Fig. 3. The planner model is an under-approximation of the controller model (Assumption 1). Starting from the same initial condition, all possible outcomes of the planner model (blue shaded area) is captured by the controller model (green shaded area). The actual outcome in two models under the influence of the same control sequence should be considerably close (Assumption 2).

This assumption is properly justified in the sense that the responsibility of the planner is to only generate feasible trajectories that the corresponding controller can handle, otherwise the risk of controller not being able to catch up with planner can lead to catastrophic consequences such as collisions. This is illustrated in Fig.3. It should be noted that this assumption does not hold true universally, it is possible that planners can generate trajectories that are not feasible for controllers to follow, in which case additional check needs to be performed to ensure the drivability of such trajectories [Schürmann et al. (2017)].

Assumption 2. The models used for planning (A) and control (B) are $T\epsilon$-close to each other, and $T\epsilon$-close to the simulation model (C) in the part $S_0 \subset S$ of the full state space S, such that the modelling error between two models e_{AB}, e_{AC}, e_{BC} represented by state deviation in finite time T starting from the same initial condition $\mathbf{x}_0 \in S$ and control input $\mathbf{u}(t)$ is bounded by a small constant $\epsilon \in \mathbb{R}$:

$$\forall \mathbf{x}_0 \in S_0, e_{i,j} = \|\mathbf{x}_i(T) - \mathbf{x}_j(T)\| \leq \epsilon \tag{2}$$

where $i, j \in \{A, B, C\}$ and $i \neq j$, the states under different models follow their respective model dynamic:

$$\mathbf{x}_i(T) = \int_0^T f_i(\mathbf{x}, \mathbf{u}, t)\mathrm{d}t + \mathbf{x}_0, i \in \{A, B, C\} \tag{3}$$

and the control invariance is implicitly assumed, i.e., as long as the initial condition $\mathbf{x}_0$ stays within the set of modelling interest S_0, then all resulting states following the control strategy $\mathbf{u}$ will also lie within the same set: $\forall \mathbf{x}_0 \in S_0, \mathbf{u} \in \mathcal{U}, t \in \mathbb{R}^+, \mathbf{x}_i(t) \in S_0, i \in \{A, B, C\}$.

This assumption is often the case for vehicle simulations, where the planner and controller usually models well the

linear dynamic behavior around the stable equilibrium of the vehicle system. While the simulation model is often more powerful and is designed to be capable of covering dynamics in non stable regions, those regions are less utilized by conventional planners or controllers.

With the above two assumptions, we can make the following assertions.

Assertion 1. If a viable trajectory exists in the planner model (A) for a specific scenario, then there exists at least a viable trajectory in the controller model (B).

Proof. This assertion directly follows assumption 1, since the planner model (A) under-approximates the controller model (B), by definition of differential inclusion, a trajectory in under-approximated dynamic can always be represented by its counterpart, the approximated dynamic. This assertion is important in guaranteeing feasible planner trajectory feed into the controller, and therefore is a good practice for planning and control engineers.

Assertion 2. If no viable trajectories exist in the planner model (A) for a specific scenario, then there is almost no viable trajectories in the controller model (B).

This "almost" no viable trajectories argument indicates that the planner model has captured the majority of possible outcomes that can happen in the controller model. This however does not rule out the possibility that there exists a few viable trajectories in the controller model which are not captured by the planner model. In order to have a stronger argument about the planner result, we introduce the following corollary on a justification model.

Corollary 1. Given a model (M) that over-approximates the controller model (B) in the system region of interest $S_0 \subset S$: $F_B(\mathbf{x}, t) \subset F_M(\mathbf{x}, t)$, if no viable trajectories exist in the model (M), then there is certainly no viable trajectories in the controller model (B).

Proof. this is again a direct result from Assumption 1 and the property of differential inclusion, and can be proved by contradiction. If such a viable trajectory η indeed exists in the controller model (B), then since $F_B(\mathbf{x}, t) \subset F_M(\mathbf{x}, t)$, the same trajectory should exist in the model (M). This contradicts the starting assumption. Therefore the original statement is true.

We call the model (M) the justification model. And we will show in the next section how this corollary can be used to justify unconventional emergency maneuvers when such a model (M) predicts the failure of any controllers modelled under the linear region of the vehicle dynamics.

4. PROPOSED METHOD

The main objective of this work is to justify a beyond-the-limit vehicle controller in constant-speed emergency scenarios at the absolute necessity. The constant-speed constraint is assumed to simplify the model and reachability calculations, and has application in cases of high-speed heavy vehicles [Liu et al. (2016)] as well as low tire-road friction scenarios.

4.1 Problem formulation

The problem can be broken into two parts: (1) estimating the potential failure of the conservative control policy and (2) finding a new controller with which the emergency might be mitigated. We are not looking for a stability guarantee for the new controller. Instead, we argue that selecting a new controller with potential for collision avoidance instead of certainty for failure is still a good strategy.

Problem 1. *Given a driving scene with a set of observations O_t, ego vehicle state $\mathbf{x_t}$, and a backward event horizon $[t_0, t_f]$, we are interested in finding whether the current state $\mathbf{x_t}$ is in the backward reachability tube of a set of undesired states $\mathcal{T} = p(O_t, t_f)$. Where p is a perception function that maps the observations and the time horizon to a set of undesired states in the ego-vehicle coordinate system. If the check fails, then we are also interested in finding a new controller that would push the vehicle state out of the backward reachability tube.*

4.2 Vehicle models

A baseline controller, Stanley lateral control [Thrun et al. (2006)] is based on a linear bicycle model with infinite cornering stiffness:

$$\begin{aligned} \dot{x} &= v\cos(\phi) \\ \dot{y} &= v\sin(\phi) \\ \dot{\phi} &= \frac{v}{L}\delta \end{aligned} \tag{4}$$

where δ is the steering angle, L is the distance between the front axle and rear axle. The stable operation of the Stanley controller is restricted to a region where this linear bicycle model largely holds for each speed v. For the constant speed that we are assuming for this scenario, the stable performance range for Stanley controller is empirically tested to be: $|\dot{\phi}| \leq 0.20$[rad/s] in simulation for the particular vehicle configuration used in this work, which can be found in Table I in [Zhao et al. (2021)].

In order to justify this baseline controller for incoming situations, we need a justification model (M) for it. The Dubin's car model is a well-studied model for backward reachibilty analysis [Chen et al. (2018)]. The system dynamics is represented with the following equations:

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} v\cos\phi \\ v\sin\phi \\ \omega \end{bmatrix} \tag{5}$$

where $\dot{x}$ and $\dot{y}$ are the time derivatives of the x and y positions respectively. ω is the angular velocity and also the control input, v is velocity. This model can serve as our justification model (M) for the conservative baseline controller, since the Dubins model is an over-approximation of the controller-based model with an assumption that the vehicle speed is constant (proof in Appendix A), and that the control input ω is over-approximated by the range $|\omega| \leq 0.21$[rad/s]:

$$F_B(\mathbf{x}, t) \subset F_M(\mathbf{x}, t) \tag{6}$$

therefore based on Corollary 1, we can use the Dubins car model as a justification model (M) to justify whether the

baseline controller will be in an unavoidable collision using the minimal backward reachable tube.

4.3 Minimal backward reachable tube

The above dynamical system consists of system states $\mathbf{x}$ and the system equation:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, \mathbf{d}) \tag{7}$$

where $f(\cdot)$ is the system dynamics, $\mathbf{u}$ and $\mathbf{d}$ are control inputs by player I (ego actor) and player II (adversarial actor). It is assumed that $f(\cdot)$ is uniformly continuous, bounded and Lipschitz continous in $\mathbf{x}$ for fixed $\mathbf{u}$ and $\mathbf{d}$.

A system trajectory from time t_0 to t_f under the inputs $\mathbf{u}$ and $\mathbf{d}$ can be conveniently represented by $\zeta(\cdot) \in \mathbb{R}^{\mathrm{n}}$:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t; \mathbf{x}, t_0, t_f, \mathbf{u}(\cdot), \mathbf{d}(\cdot)) \\ &= f(\zeta(t; \mathbf{x}, t_0, t_f, \mathbf{u}(\cdot), \mathbf{d}(\cdot)), \mathbf{u}(t), \mathbf{d}(t)) \end{aligned} \tag{8}$$

The minimal backward reachable tube (minBRT) [Chen and Tomlin (2018)] represents the set of states $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ at $t = t_0$ from which the system can be driven into some target set $\mathcal{T}$ within a time horizon $t \in [t_0, t_f]$ regardless of any action $\mathbf{u}$ taken by the ego actor. The minBRT for non-anticipative adversarial actor strategies $\mathbf{d} = \gamma \in \Gamma(t)$ can be expressed as:

$$\begin{aligned} \overline{\mathcal{A}}(t_0) = \{\mathbf{x} : \exists \gamma \in \Gamma(t), \forall \mathbf{u}(\cdot) \in \mathbb{U}, \exists t \in [t_0, t_f], \\ \zeta(t; \mathbf{x}, t_0, t_f, \mathbf{u}(\cdot), \gamma[\mathbf{u}](\cdot)) \in \mathcal{T}\} \end{aligned} \tag{9}$$

where $\Gamma(t)$ is the set of all non-anticipative strategies, γ is a strategy belonging to $\Gamma(t)$ and reacting to $\mathbf{u}(\cdot)$, $\zeta(\cdot)$ is the trajectory of system state starting from time t_0 till time t_f for the control policy $\mathbf{u}(\cdot)$ of ego actor and $\mathbf{d}(\cdot) = \gamma[\mathbf{u}](\cdot)$ of adversarial actor, $\mathcal{T}$ is the target set defined at time t_f. In this work, the adversarial input (disturbance $\mathbf{d}$) can be included during verification to account for modeling error or extra safety margin, see Fig.4.

Property 1. The minBRT only grows as $t_f - t_0$ increases (assuming t_f is fixed):

$$\overline{\mathcal{A}}(t_0) \subset \overline{\mathcal{A}}(t_0') \quad \text{for} \quad t_0' < t_0 < t_f \tag{10}$$

This property can be easily proven based on Theorem 2 in [Mitchell (2002)]. The implication of this property is that one needs not look further than the minBRT with a slightly over-estimated time interval $[t_0, t_f]$ from current time t_0 to over-estimated time of reach t_f into the target set, if the control goal is to avoid going into the target set.

To calculate the minBRT, an abstract value function $V(\mathbf{x})$ is constructed such that the states leading to negative values in $V(\mathbf{x})$ corresponds to the minBRT $\overline{\mathcal{A}}(t_0)$:

$$\overline{\mathcal{A}}(t_0) = \{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : V(\mathbf{x}) < 0\} \tag{11}$$

A toolbox [Bansal et al. (2017)] has been developed to calculate the minBRT by solving the Hamilton-Jacobi-Issac equations leading to $V(\mathbf{x})$, and is used here for the backward reachability (BR) verification.

4.4 Justifying emergency maneuvering

With the definition of minBRT in the previous section, it is now possible to combine minBRT and the justification

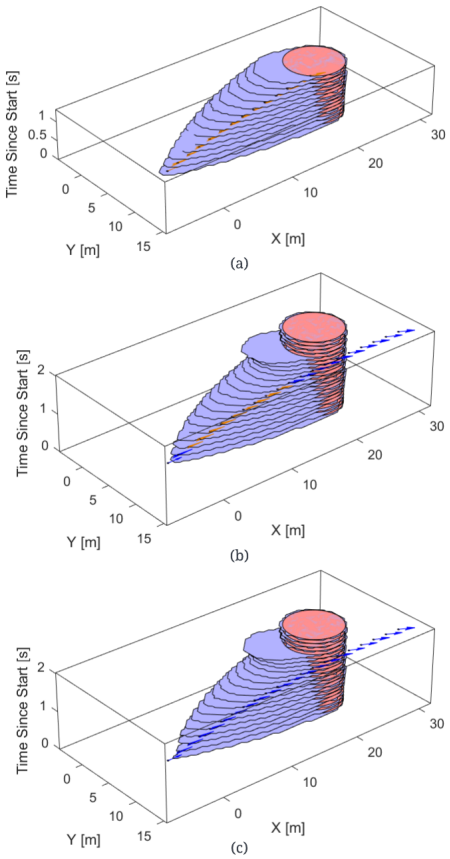

Fig. 4. (a) A vehicle with constant speed (orange arrows) $v = 15$m/s takes maximum heading rate under constraint $\dot{\phi}_{\max} = 0.20$rad/s and collides with the undesired target set (red) by staying inside the minBRT (cyan). (b) A vehicle with constant speed (orange arrows) $v = 15$m/s takes maximum $\dot{\phi}$ under constraint $\dot{\phi}_{\max} = 0.25$rad/s and is able to steer away from the undesired target set (red) by nearly touching the minBRT (cyan). The vehicle is initially in minBRT then comes out of minBRT because of disturbance $\mathbf{d} = [0.25, 0.25, 0]^T$ playing a larger propagated role when vehicle is further away from the target. (c) If disturbance $\mathbf{d}$ is removed from (b), then the vehicle is always outside of the minBRT.

model (M) to instantiate Corollary 1 and produce the following proposition to prove the certainty of failure for controllers whose controller-based models (B) are over-approximated by the justification model (M).

Proposition 1. *If the current vehicle state* $\mathbf{x}$ *is in the the minBRT:* $\mathbf{x} \in \overline{\mathcal{A}}(t_0)$ *of the justification model (M), all possible control policies* Π *whose dynamics are over-approximated by the justification model (M) will fail, and the scene will end in an undesired state within horizon* t_f*.*

Proof. By definition, $\mathbf{x}_0$ is in $\overline{\mathcal{A}}(t_0)$. Then, following the BRT assumptions and proofs [Mitchell (2007)], regardless of how the control policy $\pi \in \Pi$ is selected, with the worst of all possible actions of adversarial disturbance and all possible beneficial ego actions, it indicates that there is no solution that the ego-vehicle will end up in a safe state in the event horizon t_f.

An example of unavoidable collision is illustrated in Fig. 4(a). The vehicle follows a simplified Dubin's car dynamics, with a fixed v and a $\dot{\phi}$ range of $[-\dot{\phi}_{\max}, \dot{\phi}_{\max}] = [-0.2, 0.2]$(rad/s). To avoid hitting undesired target set (red circle), maximum $\dot{\phi}_{\max}$ is adopted. However, since the vehicle is inside the minBRT since the beginning, the collision with the undesired target set is inevitable.

With a new (and more capable) controller candidate whose controller based dynamics is under-approximated by another justification model M' (Dubins car model with $\dot{\phi}_{\max} = 0.25$rad/s), if the vehicle's initial state is outside the minBRT in M', then the new controller will be able to produce a viable control to avoid the undesired target. The illustration is shown in Fig. 4(b) and 4(c). As can be seen in the figure, with the new justification model, a collision-free trajectory can be generated.

4.5 Beyond-the-limit controller

Controllers that operate past the linear operating range of vehicles are considered beyond-the-limit controllers. Car racing maneuvers such as trail-braking, pendulum turn [Velenis et al. (2007); Velenis et al. (2008)] and drifting are such examples. These controllers are often without stability guarantees. However they do provide new pathways to safety. In the event of an emergency, if the current vehicle state $\mathbf{x}$ finds itself inside the minBRT of undesired target set $\mathbf{x} \in \overline{\mathcal{A}}(t_0)$, then conservative controllers, if executed, will lead to unavoidable collision as per Proposition 1. If the system regions on which the beyond-the-limit controllers operate allow vehicle state to escape out of the minBRT, then these controllers have undeniable advantage over the conservative controller, and should be selected to potentially avoid a previously unavoidable collision. In practice however, such controllers should also be further verified and validated to minimize the additional risk they can generate.

5. SIMULATED VALIDATION

5.1 Simulation Environment

The simulation is implemented with MATLAB$^{\text{TM}}$ Simulink$^{\text{TM}}$ using the Vehicle Dynamics Blockset$^{\text{TM}}$ [MathWorks® (2020)]. The HelperOC toolbox from [Bansal et al. (2017)] is used for BR calculations. A vehicle modelled with 3 degrees of freedom (3DOF) body dynamics and combined slip magic tire model [Gillespie (1992)] is used as the "ground truth". The vehicle parameters are listed in Table I in [Zhao et al. (2021)].

5.2 Scenarios

Scenario 1 includes an ego vehicle ($v = 15$m/s) and a pop-up obstacle in front of the ego vehicle at a fairly long distance ($D = 30$m). The obstacle is detected when ego vehicle reaches $x = 0$m. Scenario 2 includes the same setup except that this time the pop-up obstacle comes much closer to the ego vehicle when detected($D' = 22$m). To avoid the obstacles, each scenario is associated with a pre-planned path that avoids the collision, but the path may or may not be suitable for the controllers. The freedom here

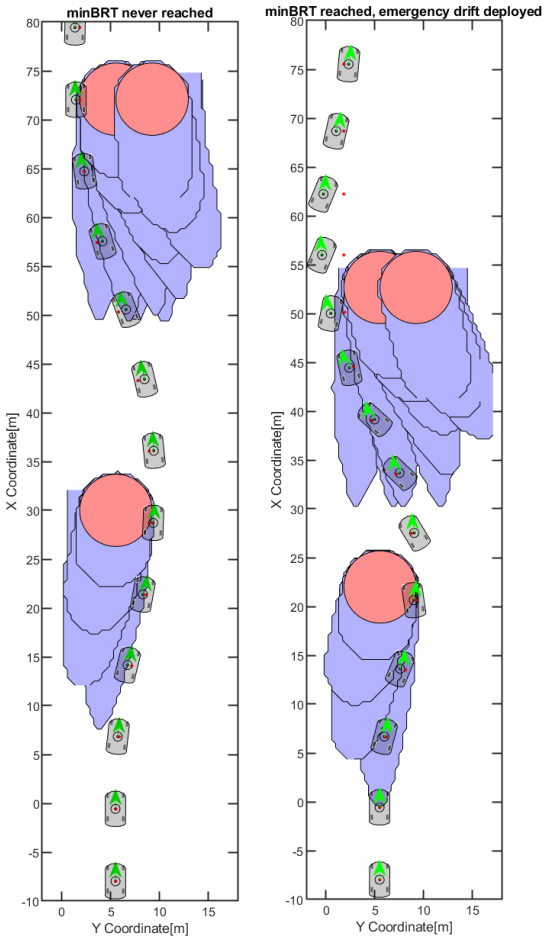

Fig. 5. (a) scenario 1: ego vehicle is able to use normal driving to steer through a more forgiving pop-up obstacle; (b) scenario 2: ego vehicle deploys beyond the limit driving after an inside-BRT event is triggered.

is the choice to switch controllers at the desired moment, and our proposed method will make exactly such decisions. The predefined paths are made up of piece-wise G^1-smooth straight lines and circular arcs.

5.3 Implementation

As a simplified model, the target set(obstacle) $\mathcal{T}$ is a circle with a radius equal to lane width $R = 3.7$m. Predefined paths are provided for controller to avoid the obstacles. When it comes to BR check of collision with obstacles, a simple constant speed Dubins car model (Eq.(5)) is used with randomized disturbances of magnitude $[0.25\text{m/s}, 0.25\text{m/s}, 0\text{rad/s}]^{\text{T}}$ on $[\dot{x}, \dot{y}, \dot{\phi}]^{\text{T}}$ to model uncertainties.

The justification process constantly computes the minBRT $\overline{\mathcal{A}}_{\mathcal{T}_i}(t_0)$ of ego vehicle using the justification model dynamics (5) for each detected obstacle $\mathcal{T}_i$ at a frequency of 0.5 seconds. And an overestimated time-to-collision $t_f - t_0$ is used to feed the minBRT calculation for assurance:

$$t_f - t_0 = \frac{D}{v} \tag{12}$$

where D is the distance from vehicle CG to obstacle CG, and v is the current vehicle speed. Under the constant

speed assumption, this value of D/v is the time for ego vehicle to not only reach the boundary of, but enough to arrive at the center of the undesired target. Therefore it is a proper overestimation of time-to-collision.

Two controller variants are considered: (1) a conservative baseline controller with optimal preview longitudinal control [MacAdam (1981)] and Stanley lateral control [Hoffmann et al. (2007)]; (2) a beyond-the-limit controller (HOTDOG [Zhao et al. (2021)]) capable of both sustaining drift and transitioning to and from conservative driving.

The HOTDOG controller already implements a controller mode switching logic based on the product of speed v and curvature κ to arbitrate between controller (1) and (2) above, and that switching logic is seamlessly updated with the minBRT-based controller justification algorithm: the product $v\kappa$ corresponds to the vehicle's heading rate $\dot{\phi}$. The maximum operation regions for controllers (1) and (2) are $|\dot{\phi}|_{drift,\max} = 0.26$(rad/s) and $|\dot{\phi}|_{drive,\max} = 0.20$(rad/s).

The successful avoidance of an obstacle is marked by the vehicle CG driving past the obstacle CG without getting inside the obstacle area or vehicle destabilization (i.e., spin-out).

6. RESULTS

In scenario 1 (Fig.5(a)), while the ego vehicle is driving forward with the default conservative controller, BR is constantly checked with the initial justification model M1 (Dubins car model with $|\phi|_{\max} = 0.21$rad/s). Since the obstacle is far enough when it pops up, the ego vehicle is not inside minBRT, thus just by tracking the pre-planned trajectory with the conventional controller the collision could be avoided. In scenario 2 (Fig.5(b)), since the obstacle is much closer when it popped up, the ego vehicle is immediately in the minBRT when it detects the obstacle. As a result, the conservative controller is determined to be unsafe for the scenario, and alternative controllers need to be deployed. In this case, a beyond-the-limit controller (HOTDOG in this case) is justified by its verification model M2(Dubins car model with $|\phi|_{\max} = 0.26$rad/s), since it can handle situation up to 0.26rad/s according to Section 5.3, and it passes the BR check with its verification model M2. In summary, the two scenarios demonstrate the cases when the minBRT based verification is either triggered or not triggered. And in the triggered case, it is demonstrated how a beyond-the-limit controller is justified to be deployed.

Because the planned path for the two scenarios is not calculated by a specified planner model A that satisfies Assumption 1, there is no guarantee (like the one in Assertion 1) that either the conservative controller or the beyond-the-limit controller can follow the planned path stably. This imperfect situation, however, signifies the importance of a verification model M that satisfies Corollary 1. Such a model M can detect the vulnerability of the existing controller following a specific path/trajectory. Upon such detection, an alternative controller, if available, can be tried with its associated new verification model M' and can be deployed to handle the situation if it passes the verification under M'.

7. CONCLUSIONS

In this work a method is proposed to address the justification of beyond-the-limit driving in emergency driving scenarios. Suddenly detected obstacles may force the automated vehicle to take drastic limit-handling measures to go back to safety. By checking the backward reachability (BR) of incoming collision objects and by controller arbitration, the proposed justification method ensures the beyond-the-limit controller is deployed in absolute urgency to steer vehicle away from collision. The experimental results demonstrate the ability to turn an unavoidable collision of baselines into an avoidable event. This method is not necessarily limited to collision avoidance, but can also be applied to formally justify any hazard prevention operation when conventional approaches fail, as long as the system and operations of interest can be properly modelled.

REFERENCES

Acosta, M., Kanarachos, S., and Fitzpatrick, M.E. (2018). On full magv lateral dynamics exploitation: Autonomous drift control. In *2018 IEEE 15th International Workshop on Advanced Motion Control (AMC)*, 529–534.

Arab, A. and Yi, J. (2020). Safety-guaranteed learning-predictive control for aggressive autonomous vehicle maneuvers. In *2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 1036–1041.

Bansal, S., Chen, M., Herbert, S., and Tomlin, C.J. (2017). Hamilton-jacobi reachability: A brief overview and recent advances. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2242–2253.

Berntorp, K., Olofsson, B., Lundahl, K., and Nielsen, L. (2014). Models and methodology for optimal trajectory generation in safety-critical road–vehicle manoeuvres. *Vehicle System Dynamics*, 52(10), 1304–1332.

Cai, P., Mei, X., Tai, L., Sun, Y., and Liu, M. (2020). High-speed autonomous drifting with deep reinforcement learning. *IEEE Robotics and Automation Letters*, 5(2), 1247–1254.

Chen, M., Herbert, S.L., Vashishtha, M.S., Bansal, S., and Tomlin, C.J. (2018). Decomposition of reachable sets and tubes for a class of nonlinear systems. *IEEE Transactions on Automatic Control*, 63(11), 3675–3688.

Chen, M., Hu, Q., Fisac, J.F., Akametalu, K., Mackin, C., and Tomlin, C.J. (2017). Reachability-based safety and goal satisfaction of unmanned aerial platoons on air highways. *Journal of Guidance, Control, and Dynamics*, 40(6), 1360–1373.

Chen, M. and Tomlin, C.J. (2018). Hamilton–jacobi reachability: Some recent theoretical advances and applications in unmanned airspace management. *Annual Review of Control, Robotics, and Autonomous Systems*, 1(1), 333–358.

Claussmann, L., Revilloud, M., Gruyer, D., and Glaser, S. (2020). A review of motion planning for highway autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 21(5), 1826–1848.

Gillespie, T. (1992). *Fundamentals of Vehicle Dynamics*. SAE International.

Goh, J., Goel, T., and Gerdes, J. (2018). A controller for automated drifting along complex trajectories. In *14th International Symposium on Advanced Vehicle Control*, 1–6.

Goh, J.Y.M. (2019). *Automated Vehicle Control Beyond the Stability Limits*. Ph.D. thesis, Stanford University.

Hindiyeh, R.Y. and Gerdes, J.C. (2010). Design of a dynamic surface controller for vehicle sideslip angle during autonomous drifting. *IFAC Proceedings Volumes*, 43(7), 560 – 565. 6th IFAC Symposium on Advances in Automotive Control.

Hoffmann, G.M., Tomlin, C.J., Montemerlo, M., and Thrun, S. (2007). Autonomous automobile trajectory tracking for off-road driving: Controller design, experimental validation and racing. In *2007 American Control Conference*, 2296–2301.

Jiang, F.J., Gao, Y., Xie, L., and Johansson, K.H. (2020). Ensuring safety for vehicle parking tasks using hamilton-jacobi reachability analysis. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 1416–1421.

Liu, J., Jayakumar, P., Stein, J.L., and Ersal, T. (2016). A study on model fidelity for model predictive control-based obstacle avoidance in high-speed autonomous ground vehicles. *Vehicle System Dynamics*, 54(11), 1629–1650.

MacAdam, C.C. (1981). Application of an optimal preview control for simulation of closed-loop automobile driving. *IEEE Transactions on Systems, Man, and Cybernetics*, 11(6), 393–399.

MathWorks® (2020). *Vehicle Dynamics Blockset™*. Release version 1.5.

Mitchell, I. (2002). *Application of Level Set Methods to Control and Reachability Problems in Continuous and Hybrid Systems*. Ph.D. thesis, Stanford University.

Mitchell, I.M. (2007). Comparing forward and backward reachability as tools for safety analysis. In *International Workshop on Hybrid Systems: Computation and Control*, 428–443. Springer.

Ono, E., Hosoe, S., Hoang D. Tuan, and Doi, S. (1998). Bifurcation in vehicle dynamics and robust front wheel steering control. *IEEE Transactions on Control Systems Technology*, 6(3), 412–420.

Schürmann, B., Heß, D., Eilbrecht, J., Stursberg, O., Köster, F., and Althoff, M. (2017). Ensuring drivability of planned motions using formal methods. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 1–8. IEEE.

Seo, H., Youngdong Son, C., Lee, D., and Jin Kim, H. (2020). Trajectory planning with safety guaranty for a multirotor based on the forward and backward reachability analysis. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 7142–7148.

Shafaei, S., Kugele, S., Osman, M.H., and Knoll, A. (2018). Uncertainty in machine learning: A safety perspective on autonomous driving. In *International Conference on Computer Safety, Reliability, and Security*, 458–464. Springer.

Subosits, J.K. and Gerdes, J.C. (2021). Impacts of model fidelity on trajectory optimization for autonomous vehicles in extreme maneuvers. *IEEE Transactions on Intelligent Vehicles*, 6(3), 546–558.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., et al. (2006). Stanley: The robot that won the darpa grand challenge. *Journal of field Robotics*, 23(9), 661–692.

Velenis, E., Tsiotras, P., and Lu, J. (2007). Aggressive maneuvers on loose surfaces: Data analysis and input parametrization. In *2007 Mediterranean Conference on Control Automation*, 1–6. doi: 10.1109/MED.2007.4433885.

Velenis, E., Frazzoli, E., and Tsiotras, P. (2009). On steady-state cornering equilibria for wheeled vehicles with drift. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, 3545–3550.

Velenis, E., Tsiotras, P., and Lu, J. (2008). Optimality properties and driver input parameterization for trail-braking cornering. *European Journal of Control*, 14(4), 308–320.

Voser, C., Hindiyeh, R.Y., and Gerdes, J.C. (2010). Analysis and control of high sideslip manoeuvres. *Vehicle System Dynamics*, 48(sup1), 317–336.

Yin, Y., Li, S.E., Li, K., Yang, J., and Ma, F. (2020). Self-learning drift control of automated vehicles beyond handling limit after rear-end collision. *Transportation Safety and Environment*, 2(2), 97–105.

Zhao, T., Yurtsever, E., Chladny, R., and Rizzoni, G. (2021). Collision avoidance with transitional drift control. In *2021 International Conference on Intelligent Transportation Systems(ITSC)*.

Zhao, T., Yurtsever, E., Paulson, J., and Rizzoni, G. (2022). Formal certification methods for automated vehicle safety assessment. *IEEE Transactions on Intelligent Vehicles*, 1–18.

Appendix A. PROOF OF MODEL OVER-APPROXIMATION

In this appendix, we prove that the controller based model B in (4) with $|\dot{\phi}| \leq 0.20[rad/s]$ is over-approximated by the Dubin's car model M in (5) with $|\omega| \leq 0.21[rad/s]$, assuming additionally that the velocity is constant $v = v_0$.

Proof. The exact differential inclusion of model B at an arbitrary state $\mathbf{x} = [x, y, \phi]^T$ at time t is:

$$\begin{aligned} F_A(\mathbf{x}, t) &= \{\dot{\mathbf{x}} \in \mathbb{R}^3 | \dot{\mathbf{x}} = f_A(\mathbf{x}, \mathbf{u}, t)\} \\ &= \left\{ \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} \in \mathbb{R}^3 \middle| \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} v_0 \cos\phi \\ v_0 \sin\phi \\ \dot{\phi} \end{bmatrix}, \forall |\dot{\phi}| \leq 0.20 \right\} \end{aligned} \tag{A.1}$$

While the exact differential inclusion of model M at the same arbitrary state $\mathbf{x} = [x, y, \phi]^T$ at time t is:

$$\begin{aligned} F_B(\mathbf{x}, t) &= \{\dot{\mathbf{x}} \in \mathbb{R}^3 | \dot{\mathbf{x}} = f_B(\mathbf{x}, \mathbf{u}, t)\} \\ &= \left\{ \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} \in \mathbb{R}^3 \middle| \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} v_0 \cos\phi \\ v_0 \sin\phi \\ \omega \end{bmatrix}, \forall |\omega| \leq 0.21 \right\} \end{aligned} \tag{A.2}$$

By comparing the expression of (A.1) and (A.2), it is apparent that $\forall \dot{\mathbf{x}} \in F_B(\mathbf{x}, t)$, one can find the equivalence in $F_M(\mathbf{x}, t)$. Therefore,

$$F_B(\mathbf{x}, t) \subset F_M(\mathbf{x}, t) \tag{A.3}$$

that is, the controller based model B is over-approximated by the Dubin's car model M.