Learning Autonomous Vehicle Safety Concepts from Demonstrations

Karen Leung⋆†, Sushant Veer⋆, Edward Schmerling ⋆‡, Marco Pavone⋆‡

Abstract—Evaluating the safety of an autonomous vehicle (AV) depends on the behavior of surrounding agents which can be heavily influenced by factors such as environmental context and informally-defined driving etiquette. A key challenge is in determining a minimum set of assumptions on what constitutes reasonable foreseeable behaviors of other road users for the development of AV safety models and techniques. In this paper, we propose a data-driven AV safety design methodology that first learns “reasonable” behavioral assumptions from data, and then synthesizes an AV safety concept using these learned behavioral assumptions. We borrow techniques from control theory, namely high-order control barrier functions and Hamilton-Jacobi reachability, to provide inductive bias to aid interpretability, verifiability, and tractability of our approach. In our experiments, we learn an AV safety concept using demonstrations collected from a highway traffic-weaving scenario, compare our learned concept to existing baselines, and showcase its efficacy in evaluating real-world driving logs.

I. INTRODUCTION

As autonomous vehicle (AV) operations grow, developing appropriate methods for evaluating AV safety becomes ever more imperative. The question of “is a vehicle in an unsafe state?” is relevant for AV system developers, policymakers, and the general public alike (see Figure 1). While guaranteeing safety may not be practical in the face of the myriad uncertainties and complexities that come with real-world driving, there is still a broad desire to codify, to some extent, collectively agreed-upon notions of safety [1]. Should safety be defined using data-driven methods that can account for the complexities of the AV’s environment but lack interpretability and formal guarantees, or leverage control theoretic techniques derived from first principles which are interpretable and rigorous but not as scalable or expressive as their learned counterparts? In this work, we strike a middle-ground by using a learning safety-critical driving behavior model and integrating it within a robust control framework to develop an interpretable and rigorous AV safety model that is informed by data.

Towards this goal of building safe and trustworthy AVs, various stakeholders have advanced safety concepts consisting, in general, of two functions mapping world state (e.g., joint state of all agents and environmental context) to (i) a scalar measure of safety, and (ii) a set of allowable (safe) agent actions [2]. Numerous uses of such safety concepts have been proposed throughout AV pipelines, e.g., a criterion to prune away unsafe plans [3], a safety monitor to determine when evasive action must be taken [4], a component in the planning objective [5], or for perception safety evaluation metrics [6], [7]. Critically, different behavioral assumptions lead to different safety concepts which in turn affects realized AV safety and performance. To mitigate over conservatism and enable maximum flexibility, we contend that the design of an effective safety concept hinges upon characterizing reasonable foreseeable behaviors of other agents, in a way conducive to efficiently evaluating the safety of driving scenes and producing associated safe controls.

To address this challenge, we propose designing novel safety concepts by learning from data what are controls, specifically, control sets, that humans operate with when their safety is threatened, and then using these learned control sets to inform AVs of what are reasonable foreseeable behaviors of other agents in safety-critical scenarios. Equipped with such a learned “human behavior collision avoidance model,” we perform safety concept synthesis by using robust control theory, specifically Hamilton-Jacobi reachability [8], as a powerful inductive bias for interpretability, verifiability, and tractability. Our control set learning approach differs from reward learning paradigms (i.e., inverse reinforcement learning / inverse optimal control [9], [10], [11]) which strive to learn high-level human intentions from demonstrations, often in the absence of constraints [3], [12]. Reward learning approaches aim to model a human’s high-level planning objective which encapsulates more than just safety considerations. Whereas in this work, we are interested in learning control constraint boundaries associated with collision avoidance behaviors as opposed to nuanced behaviors (e.g., aggressive versus passive).

Structure and Contributions. We provide a literature review in Section II, give an overview on Hamilton-Jacobi (HJ) reachability in Section III, and formally state our safety concept learning problem in Section IV. Then we describe the details of our key contributions: (i) We propose a data-driven approach to learn humans’ collision avoidance behaviors in the control space to capture “reasonable driving behaviors” (Section V). Specifically, we learn safe control sets from demonstrations via a high-order control barrier function (HOCBF) [13] framework. (ii) We develop a con-
strained game-theoretic optimization problem derived from a HJ reachability formulation to synthesize novel data-driven safety concepts that are robust to other agents’ behaviors while respecting the learned collision avoidance behaviors (Section VI). (iii) We demonstrate our proposed learning framework using highway driving data and show that the resulting data-driven safety concept is less conservative than other common safety concepts (due to the way it captures constraints on reasonable agent behavior) and thus is useful as a “responsibility-aware” evaluation metric for the safety of AV interactions (Section VII).

II. RELATED WORK

We use the term safety concept to help unify existing safety theory prevalent in various robot planning and control algorithms, such as velocity obstacles [14], [15], forward reachable sets [16], [17], contingency planning [18], [19], backward reachability [8], and other methods that make static assumptions on agent behavior [20], [21]. The differences between various safety concepts stem from the assumptions about the behavior of other interacting agents, ranging from worst-case assumptions [8] to presuming agents follow fixed open-loop trajectories (e.g., braking [21], constant velocity [15]). Indeed, a core challenge is in selecting behavioral assumptions that balances conservatism, tractability, interpretability, and compatibility with real-world interactions.

Recent works propose dynamically changing the conservatism of the safety concept based on online estimations of how confident the robot’s human behavior prediction model is. If a human agent is behaving as expected (i.e., high model confidence), then the worst-case assumptions in the safety concept can be relaxed, and vice versa [22], [23]. However, the integrity of the adaptive safety concept depends on the quality of the prediction model where obtaining an accurate human behavior prediction model is in general quite challenging and, indeed, is an active research field [24].

Another data-driven safe control technique is to use expert demonstrations and learn a control barrier function (CBF) [25] to describe unsafe regions in the state space [26], [27], [28], [29]. The learned CBF is then directly used as the core safety mechanism in synthesizing a safe policy. However, CBFs are not well-suited for interactive settings where there is uncertainty in how other interacting agents may behave. In our work, we too consider learning CBFs (specifically, high order CBFs (HOCBFs) [13]), but instead use the learned HOCBF as an intermediate step towards formulating a more rigorous notion of safety rooted in robust control theory which enjoys interpretability and verifiability benefits.

III. SAFETY CONCEPT VIA HAMILTON-JACOBI REACHABILITY

We define a safety concept as a combination of two functions mapping world state to (i) a scalar measure of safety, and (ii) a set of allowable actions for each agent in which to preserve safety. A family of safety concepts can be described via a HJ reachability formulation [2].

HJ reachability is a mathematical formalism used for characterizing the safety properties of dynamical systems [8], [30]. The outputs of a HJ reachability computation are (i) a HJ value function, a scalar-valued function that measures “distance” to collision, and (ii) a set of controls that prevent the safety measure from decreasing further—precisely the components needed for a safety concept.

Consider a target set \( \mathcal{T} \) which is the set of collision states between agents \( A \) (ego agent) and \( B \) (contender). The HJ reachability formulation describes a two-player differential game to determine whether it is possible for the ego agent to avoid entering \( \mathcal{T} \) under any family of closed-loop policies of the contender, as well as the ego agent’s appropriate control policy for ensuring safety. It is assumed that the contender follows an adversarial policy and has the advantage with respect to the information pattern. Using the principle of dynamic programming, the collision avoidance problem reduces to solving the Hamilton-Jacobi-Isaacs (HJI) partial differential equation (PDE)[8],

\[
\frac{\partial \mathcal{V}(x,t)}{\partial t} + \min \left\{ 0, \max_{u_A \in \mathcal{U}^A} \min_{u_B \in \mathcal{U}^B} \nabla_x \mathcal{V}(x,t)^T f(x,u_A,u_B) \right\} = 0
\]

\( \mathcal{V}(x,0) = \ell(x) \)

where \( x \in \mathcal{X} \) denotes the joint state of agents \( A \) and \( B \), \( u_A \in \mathcal{U}^A \) and \( u_B \in \mathcal{U}^B \) are the available (bounded) controls\(^1\) of agents \( A \) and \( B \), respectively, and \( f(\cdot, \cdot, \cdot) \) is the joint dynamics assumed to be measurable in \( u_A \) and \( u_B \) for each \( x \), and uniformly continuous, bounded, and Lipschitz continuous in \( x \) for fixed \( u_A \) and \( u_B \).\(^2\) The boundary condition is defined by a function \( \ell : \mathcal{X} \to \mathbb{R} \) whose zero sub-level set encodes the target set, i.e., \( \mathcal{T} = \{ x \mid \ell(x) < 0 \} \). The solution \( \mathcal{V}(x,t), t \in [-T,0] \), called the HJ value function, captures the lowest value of \( \ell(\cdot) \) along the system trajectory within \( |t| \) seconds if the system starts at \( x \) and both agents \( A \) and \( B \) act optimally, that is, \( u_A^*(x), u_B^*(x) = \arg \max_{u_A \in \mathcal{U}^A(\arg)} \min_{u_B \in \mathcal{U}^B} \nabla_x \mathcal{V}(x,t)^T f(x,u_A,u_B) \).

Thus the HJ value function fulfills the first aspect of a safety concept. After obtaining the HJ value function, we can also consider the set of controls that prevent the HJ value function from decreasing over time. That is, we can compute the safety-preserving control set, \( \mathcal{U}_{safe}^A(x) = \{ u_A \in \mathcal{U}^A \mid \min_{u_B \in \mathcal{U}^B} \frac{\partial \mathcal{V}(x,t)}{\partial t} \geq 0 \} \), thus fulfilling the second aspect of a safety concept. By varying the problem parameters, i.e., control sets, behavior type, and a choice of \( \ell(\cdot) \), we can synthesize a family of safety concepts via HJ reachability.

Although solving HJI PDE suffers from the curse of dimensionality, it is performed offline. For reasonably sized problems, such as the pairwise car-car system considered in the work, solving the HJI PDE is tractable on modern computers. Equipped with the HJ value function and a controllers, it is uncertainty in how other interacting agents may behave. In our work, we too consider learning CBFs (specifically, high order CBFs (HOCBFs) [13]), but instead use the learned HOCBF as an intermediate step towards formulating a more rigorous notion of safety rooted in robust control theory which enjoys interpretability and verifiability benefits.

IV. TOWARDS DATA-DRIVEN SAFETY CONCEPT SYNTHESIS

A key limitation to the standard HJ reachability setup is in the assumption the contender will choose optimal collision-seeking controls from \( \mathcal{U}^B \), the set of admissible controls of

\(^1\) The control sets \( \mathcal{U}^A \) and \( \mathcal{U}^B \) are typically chosen to reflect the physically feasible limits of the system.

\(^2\) This assumption ensures that trajectories are generated by a unique control sequence.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
the system (typically corresponding to the system’s physical limits). While a worst-case assumption is key in providing robustness guarantees for the safety of the system, the assumption that a contener agent will execute any available, including worst-case, controls is unrealistic in practice.

To reduce the conservativeness in how we model a contender’s behavior within the HJ formulation (and hence resulting safety concept), we aim to use data to learn state-dependent control sets to capture agent behaviors reflecting practical real-world operations. That is, the problem we seek to solve consists of two parts: (i) learning state-dependent control sets from data, and (ii) integration of the state-dependent control sets into the HJ reachability formulation for safety concept synthesis. Building upon the notation introduced in Section III, consider a nonlinear control-disturbance-affine system describing the joint dynamics between agent $A$ and $B$,

$$
\dot{x} = f(x) + g(x)u + h(x)u_B
$$

where $x \in X \subset \mathbb{R}^n$, $u_A \in U_k \subset \mathbb{R}^{m_A}$ and $u_B \in U_s \subset \mathbb{R}^{m_B}$, and $f : X \to \mathbb{R}^n$, $g : X \to \mathbb{R}^{n \times m_A}$, and $h : X \to \mathbb{R}^{n \times m_B}$ are locally Lipschitz continuous.

**Learning state-dependent control sets.** Consider an offline dataset $S = \{(x^{(i)}, u^{(i)})\}_{i=1}^N$ of joint states between agent $A$ and $B$, and agent $A$‘s controls collected from safe interactions between agents $A$ and $B$. We assume agent $A$ does not observe the controls of agent $B$. We seek to find a mapping $\phi : X \to 2^U$ such that $u_A^{(i)} \in \phi(x^{(i)})$ for all $(x^{(i)}, u^{(i)}) \in S$. Of course $\phi$ can be a trivial mapping, i.e., $\phi(x^{(i)}) = U^A$ for all $x \in X$; thus we additionally require the learned state-dependent sets to be minimal in the sense that the set should contain the relevant points as tightly as possible.

**Safety concept synthesis.** Equipped with $\phi$, we seek to integrate the learned control sets into the HJ reachability framework to restrict the feasible control set that an agent will operate with.

V. LEARNING STATE-DEPENDENT CONTROL SETS FROM DATA

In this section, we describe our proposed approach to describing human collision avoidance behaviors by learning state-dependent control sets from data. First, we give a self-contained overview of HOCBFs (Section V-A), then follow by our proposed learning procedure (Section V-B).

A. High-Order Control Barrier Functions

HOCBF provides a method to represent and reason about control invariant sets [13, Theorem 4]. Before introducing HOCBFs, we first introduce control barrier functions (CBFs), a special case of HOCBFs. Consider a control-affine system of the form,

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

where $x \in X \subset \mathbb{R}^n$, $u \in U \subset \mathbb{R}^m$, (3)

where $f : X \to \mathbb{R}^n$ and $g : X \to \mathbb{R}^{n \times m}$ are locally Lipschitz continuous. Suppose there is a set that the system wants to stay inside of (e.g., a set of safe states). A control invariant set is a set of states for which it is possible for a system (3) to stay inside of indefinitely. In many settings, it is often desirable to find the largest control invariant set that is contained within the safe set.

### Definition 1 (Control invariant set).

A set $C$ is control invariant for a system $\dot{x} = f(x, u)$ if for all $x_0 \in C$, there exists a control law $u = k(x) \in U$ such that $x(t) \in C$, $x(0) = x_0$, $\dot{x}(t) = f(x(t), u)$ for all $t \in [0, \infty)$.

One way to represent a control invariant set is to use Control Barrier Functions, a scalar-valued function that satisfies a set of certain conditions which we describe next.

### Definition 2 (Extended Class $K_\infty$ function).

A continuous function $\alpha : (-\infty, \infty) \to (-\infty, \infty)$, belongs to extended class $K_\infty$ if it is strictly increasing, $\alpha(0) = 0$, and $\lim_{r \to \pm \infty} \alpha(r) = \pm \infty$.

### Definition 3 (Control Barrier Function [25]).

Given a system (3), consider a continuously differentiable function $b : X \to \mathbb{R}$ and a set $C \subset X$ where $C = \{x \in X \mid b(x) \geq 0\}$. Then $b$ is a candidate control barrier function (CBF) if there exists a class $K_\infty$ function $\alpha$ such that

$$
\sup_{x \in U} [L_f b(x) + L_g b(x)u] \geq -\alpha(b(x)) \quad \forall x \in X,
$$

where $L_f b(x) = \nabla b(x)^T f(x)$ and $L_g b(x) = \nabla b(x)^T g(x)$ are Lie derivatives describing the change in $b(x)$ along a vector field described by $f(x)$ and $g(x)$, respectively.

A valid CBF ensures that there exists a feasible control that prevents $b(x(t))$ from decreasing faster than a rate of $-\alpha(b(x))$ along system trajectories. If $L_g b(x) = 0$, the control input $u$ has no influence in ensuring that (4) holds.

In fact, if the relative degree $m_r$ of the system and $b$ is greater than one, then $L_g b(x) = 0$ for all $x \in X$.

### Definition 4 (Relative degree).

The relative degree $m_r$ of a (sufficiently) differentiable function $b : \mathbb{R}^n \to \mathbb{R}$ with respect to dynamics (3) is when $L_g L_f^{m_r-1} b(x) \neq 0$ but $L_g L_f^{m_r-2} b(x) = 0$.

When $m_r > 1$, the standard CBF formulation (Definition 3) cannot be used to determine what are safe control actions. Thus we consider HOCBFs [13] instead which addresses this issue, and we show later that HOCBFs are a more natural paradigm to encode obstacle sets.

### Definition 5 (High-order Control Barrier Function [13]).

Given a system (3), and an $m_r$th order differentiable function $b : X \to \mathbb{R}$ with respect to dynamics (3) is when $L_g L_f^{m_r-1} b(x) \neq 0$ but $L_g L_f^{m_r-2} b(x) = 0$.

When $m_r > 1$, the standard CBF formulation (Definition 3) cannot be used to determine what are safe control actions. Thus we consider HOCBFs [13] instead which addresses this issue, and we show later that HOCBFs are a more natural paradigm to encode obstacle sets.

### Theorem 1 ([13], Theorem 4).

Given a valid HOCBF $b(x)$ for a system (3) and sets $C_i$, $i = 1, \ldots, m_r$, $C_i = \{x \in X \mid \psi_i(x) \geq 0\}$. Then any Lipschitz controller $u \in$...
Fig. 2: Two different $\alpha_i$ functions parameterized by different parameters, $p_A$ and $p_B$, result in different trajectory behaviors and effective CBFs. Left: More “aggressive” behaviors are observed with using $p_B$ compared to $p_A$. Right: The ground truth effective CBF for each set of trajectories.

$K_{\text{HOCBF}}(x)$ renders the set $\cap_{i}^{m}C_i$ forward invariant, where

$$K_{\text{HOCBF}}(x) = \{ u \in U \mid \mathcal{L}_{f}^{m_{i}}b(x) + \mathcal{L}_{g}^{m_{i}}b(x)u + \mathcal{O}(b(x)) + \alpha_{m_{i}}(\psi_{m_{i}}(x)) \geq 0 \}.$$ (6)

**Definition 6** (Effective CBF). Given a HOCBF $b$ and corresponding $\alpha_i$, $i = 1, \ldots, m_i$, we say $\psi_{m_{i}}$ is the “effective” CBF since (5) is equivalent to (4) with $b(x) = \psi_{m_{i}}(x)$.

In short, given system dynamics and an unsafe set, HOCBFs can describe control invariant sets. The size and shape of a control invariant set is influenced by the $\alpha_i$ functions which imposes a bound on the allowable rate at which the system is allowed to approach the unsafe set.

**B. State-dependent Control Set Learning Procedure**

In this section, we detail our state-dependent control set learning procedure. Roughly, we use HOCBF as inductive bias in learning a mapping from state to control set.

1) **Learning set-up and assumptions**

We make the assumption that (safe and reasonable) drivers do not put themselves in situations where it is not possible for them to avoid a collision. That is, drivers execute controls that ensure they continually stay inside a “safe set” because leaving that set may lead to collision. For example, (typical) safe drivers avoid approaching obstacles at high speeds because they may not be able to steer and slow down to avoid the obstacle. This so-called “safe-set” is precisely a control invariant set (see Definition 1). Thus, given a dataset of safe collision-free trajectories, the goal is to learn a control invariant set and corresponding control sets that contain the observed samples. We represent the control invariant set using HOCBFs [13, Theorem 4] which provides an inductive bias that helps make the learning process tractable and a direct and interpretable way to obtain state-dependent control sets (see (6)) amenable for safety concept synthesis.

For collision avoidance, agents are aiming to avoid the same obstacle set but only differ in how they avoid it. Thus we use HOCBFs to avoid reasoning about obstacle sets in velocity states (or higher derivatives) which is difficult to construct, and instead reason about $\alpha_i$’s.

2) **Illustrative example**

Figure 2 illustrates how different parameters for the HOCBF $\alpha_i$ functions result in different collision avoidance behaviors, thereby motivating our goal of learning HOCBFs, specifically parameters for $\alpha_i$, from data. Suppose a car needs to avoid a circular obstacle centered at $x_{\text{ob}}$ with radius $r_{\text{ob}}$ while reaching a goal state $x_{\text{goal}}$. We use a HOCBF $b(x) = ||x - x_{\text{ob}}||_2 - r_{\text{ob}}^2$, to represent the obstacle (i.e., the unsafe set). Figure 2 illustrates two sets of trajectories generated from the HOCBF-CLF planner described in [13], each set constructed using different parameters, $p_A$ and $p_B$ for $\alpha_i$. We see that $p_A$ (blue) results in more conservative behaviors than $p_B$ (orange) since the blue trajectories start turning away from the obstacle earlier. Correspondingly, we see the zero sub-level of the set of effective CBF with $p_A$ is larger than the one for $p_B$.

3) **HOCBF learning algorithm**

Next, we present our HOCBF learning algorithm whereby we aim to learn the parameters for $\alpha_i$ from data. Let $S = \{(x^{(j)}, u^{(j)})\}_{j=1}^{N}$ be a dataset containing states and controls collected from humans operating in the target environment exhibiting safe collision avoidance behaviors. Since negative examples are rare and their explicit collection is impractical, one of the key strengths of our method is its ability to learn control invariant sets from just positive examples. Although our method can also consider negative examples if available.

Suppose we have dynamic data given by (3). Let $C$ describe the set of obstacle states in position space (we consider $\mathbb{R}^2$ but this analysis can extend to $\mathbb{R}^3$). Let $b : \mathcal{X} \rightarrow \mathbb{R}$ be a function where $b(x) \leq 0 \Leftrightarrow \text{pos}(x) \in C$, and $\text{pos} : \mathcal{X} \rightarrow \mathcal{R}^2$ is a function that extracts the position states. Let $m_i \geq 1$ be the relative degree of (3) and $b$. We seek to find $\alpha_i$, $i = 1, \ldots, m_i$ such that $u^{(j)} \in K_{\text{HOCBF}}(x^{(j)})$ for all $(x^{(j)}, u^{(j)}) \in S$. More concretely, let $\alpha_i^p$ indicate that $\alpha_i$ is parameterized by a vector of parameters $p_i$, and let $p = [p_1, \ldots, p_{m_i}]$. Considering that (i) $\alpha_i^p$ can be chosen to be arbitrarily large (i.e., steep) which will trivially lead to (6) being true for all datapoints, (ii) we want the HOCBF control constraint (6) to be satisfied, (iii) the states should be inside the (safe) control invariant set, and (iv) the dataset could contain outliers or noise, we seek to find $p^*$ such that,

$$p^* = \arg\min_{p} \frac{1}{N} \sum_{j=1}^{N} -\beta_3 \min \{ G_x^j u^{(j)} + F_x^j(p), \mathcal{R} \} +$$

$$\beta_4 \max \{ \text{tanh}(G_x^j u^{(j)} + F_x^j(p)), 0 \} +$$

$$\beta_5 \text{Effective CBF violation}$$

$$\beta_6 \text{Effective CBF satisfaction (saturated)}$$

$$\mathcal{R} \geq \mathcal{G}$$

$$\alpha_i \text{ regularization}$$

where $G_x^j = \mathcal{L}_{g}^{m_i}b(x^{(j)})$, $F_x^j(p) = \mathcal{L}_{f}^{m_i}b(x^{(j)}) + \mathcal{O}(b(x^{(j)}); p) + \alpha_{m_{i}}(\psi_{m_{i}}(x^{(j)}))$, and $\mathcal{O}(b(x); p) = \sum_{i=1}^{m_i} \mathcal{L}_{f}(\alpha_{m_{i}} (\psi_{m_{i}} - 1)(x))$ (c.f., (6)). We highlight some important considerations for this optimization problem. **Saturation:** We apply tanh on the second and fourth terms in (7) because while we desire all the data points to satisfy the HOCBF conditions, we are not too interested in how much each data point satisfies them. That is, we want to reduce the influence of extremely safe points (e.g., states very far and moving away from the obstacle).

**Parameterization:** The $\alpha_i$, $i = 1, \ldots, m_i$ functions must belong to extended class $\mathcal{K}_\infty$. As such, we can parameterize $\alpha_i$ as a linear combination of extended class $\mathcal{K}_\infty$ basis functions. In future work, we can consider using a monotonic
Pairwise joint dynamics: In the case with pairwise joint dynamics, as presented in (2), there is an additional term corresponding to the contender’s input (i.e., a disturbance term), and ultimately, it will show up in (7). In particular, the HOCBF constraint becomes

\[ \mathcal{L}_g \mathcal{L}_e^{m-1} b(x) u_A + \mathcal{L}_h \mathcal{L}_e^{m-1} b(x) u_B + \]

\[ Ego's \ influence \] \[ \text{Contender's influence} \]

\[ \mathcal{L}_f^{m} b(x) + \mathcal{O}(b(x); p) + \alpha_m \psi_m(x) \geq 0 \]

which is affine in \( u_A \) and \( u_B \). To select a value for \( u_B \), depending on what information is available, we could either (i) use the ground truth value, (ii) assume the worst-case, or (iii) predict a distribution over agent \( B \)’s controls and take the worst-case with respect to the predictions.

Differentiability: If the dynamics and \( b \) are differentiable, then (7) can be optimized via standard gradient descent algorithms. In this work, we use JAX [33], a Python software library for automatic differentiation.

Negative examples: If negative examples are available (e.g., from real-world driving logs or synthetically generated [34]), we can include additional terms to (7) that encourage the negative examples to violate the control constraint and have negative effective CBF values. The negative examples will help refine the delineation of the control invariant set.

VI. SAFETY CONCEPT SYNTHESIS VIA CONSTRAINED HJ REACHABILITY

We incorporate the learned HOCBF into the HJ reachability formulation to synthesize a safety concept that assumes reasonable human behaviors in safety-critical situations. Directly adding the learned HOCBF control constraint (8) to (1) places the burden of constraint satisfaction on the agent that acts second (i.e., agent \( B \)), thus imposing a strong assumption that the other (uncontrolled) agents will always behave responsibly with respect to HOCBF constraint satisfaction. To address this, we swap the ordering in which the agents act, and the following proposition states that this swapping results in a more conservative HJ value function, i.e., it is more advantageous for a player to act first.

**Proposition 1.** Consider a coupled affine constraint \( p(u_A, u_B) := au_A + bu_B + c \geq 0 \) in \( u_A \) and \( u_B \), and a linear objective \( q(u_A, u_B) = Au_A + Bu_B \). Let \( \bar{U}^A = \{ u_A \in \bar{U}^A | \exists u_B \in \bar{U}^B, p(u_A, u_B) \geq 0 \} \) and \( \bar{U}^B = \{ u_B \in \bar{U}^B | \exists u_A \in \bar{U}^A, p(u_A, u_B) \geq 0 \} \). For each agent which ensures the other agent can satisfy the constraint \( p(u_A, u_B) \geq 0 \). Let \( U^A(u_B) = \{ u_A \in U^A | p(u_A, u_B) \geq 0 \} \), and \( U^B(u_A) = \{ u_B \in U^B | p(u_A, u_B) \geq 0 \} \) describe an agent’s feasible control set with the other agent’s control fixed. Then

\[ \max_{u_B \in \bar{U}^B} \min_{u_A \in \bar{U}^A(u_B)} q(u_A, u_B) \geq \min_{u_A \in \bar{U}^A} \max_{u_B \in \bar{U}^B(u_A)} q(u_A, u_B). \]

That is, the player that acts first has the advantage assuming the second player is provided feasible options.
Fig. 4: Comparison between the ground truth control (green), distribution of predicted control (blue), three standard deviation bounds, the learned distribution, three standard deviation bounds, the ground truth, and control limits.  

with $(\delta, a)$ as steering and acceleration control inputs, and $\ell$ is the wheelbase length of the car. For the HOCBF learning, we use the following parametric form $\alpha_i^{\beta}(a) = p_{i,1} a^{p_{i,2}}$, $p_1 = [0.54, 1.16]$, $p_2 = [0.68, 1.11]$ to generate a set of trajectories with various random initial conditions. The parameter values serve as the ground truth (GT). We applied the HOCBF learning procedure using various parameterizations of $\alpha_i^{\beta}$: power function (same parameterization as GT), linear function $\alpha_i^{\beta} = p_i a$, cubic function $\alpha_i^{\beta} = p_i a^3$, and a linear combination (LC) of a linear, cubic, and tanh function $\alpha_i^{\beta}(x) = p_{i,1} x + p_{i,2} \text{tanh}(p_{i,3} x) + p_{i,4} x^3$ where $p_i = [p_{i,1}, p_{i,2}, p_{i,3}, p_{i,4}]$, all with $i = 1, 2$.

When learning the parameters of $\alpha$ (i.e., performing gradient descent on (7)), we used $\beta_1 = 1$, $\beta_2 = \beta_4 = 0.001$, $\beta_3 = 1$, $\beta_5 = 0.001$, step size/learning rate $= 0.001$, and 150000 gradient steps (though it converged much earlier). Figure 3 visualizes the zero level set of the learned effective CBF (left) and resulting trajectory from running the same reach-avoid planner used to generate the dataset but with the different learned $\alpha_i$ parameterizations (right). We see that the learned HOCBFs except the cubic parameterization is able to closely recover the ground truth behavior (black line).

B. Highway driving

In this example, we consider a highway driving scenario, synthesize a novel safety concept based on a highway driving dataset, and evaluate on real traffic data.

1) Data collection

We use the traffic-weaving dataset [36], and assume the control inputs of the other cars are not observable (this is often true in real-world settings). Instead, we learn a generative model to construct distributions over likely contender controls (i.e., disturbances) and use them as ground truth for our HOCBF learning procedure. We trained a conditional variational autoencoder with a continuous latent space (latent dimension size of 8) with a history length of ten time steps, and a prediction horizon of five. Although we predict five time steps into the future, we only consider the first (i.e., current) control of the other agent. We used LSTM cells for the encoder and decoder with hidden dimension 8. Given the prediction, we define $D_{\text{pred}}$ as the interval three standard deviations away from the mean, and then compute $u_{\text{pred}} \in D_{\text{pred}}$ that minimizes the LHS of (8) during the HOCBF learning process. Figure 4 shows an example that compares the learned distribution, three standard deviation bounds, the ground truth, and control limits.

2) Safety concept analysis

We consider a HOCBF $b(x) = \frac{\Delta x^2}{\alpha^2} + \frac{\Delta y^2}{\beta^2} - 1$, $\alpha = 5.4$, $\beta = 2.4$, and consider relative dynamics between two vehicles modeled by the simple car model (11), resulting in a relative degree of two. We applied the learning procedure outlined in Section V-B; we performed a hyperparameter sweep over the following parameters defined in (7). The values are presented in Table I, and the bolded values are the hyperparameters that we found yielded the best results.

With our learned HOCBF, we then solved the corresponding HOCBF-constrained HJ reachability problem described by (10). As our problem is relatively low-dimensional, we can solve the constrained HJ reachability problem exactly by enumerating over the vertices of the feasible domain (but adds significant computational costs). We consider a two second horizon when solving the constrained HOCBF-HJ reachability problem and compare against various safety concepts (also synthesized via a HJ computation).

Figure 5 shows the unsafe region and optimal controls for different safety concepts. Right: Plots of control sets with infeasible regions according to the HOCBF constraint shaded gray.

(a) Zero level-set of HJ value function evaluated at $\Delta x - \Delta y$ slice. $v_A = 18 \text{ms}^{-1}$, $v_B = 25 \text{ms}^{-1}$, $\theta_A = \theta_B = 0\text{rad}$.

(b) HOCBF-HJ and WC-HJ optimal controls for agents $A$ and $B$ when agent $B$ is approaching from behind.  

| Parameter       | Values                      |
|-----------------|-----------------------------|
| $\beta_1$       | $\{0.0, 0.001, 0.01, 0.1\}$ |
| $\beta_2$       | $\{0.0, 0.01, 1.0\}$       |
| $\beta_3$       | $\{0.1, 0.001\}$           |
| $\beta_4$       | $\{p_{1,1}, p_{1,2}, p_{1,3}, p_{1,4}\}$ |
| Learning rate   | $\{0.001\}$               |
| Number of steps | $\{10000\}$                |

Table I: Values of a hyperparameter sweep. The values chosen for the rest of the experiments are shown in bold.
TABLE II: Volumetric “confusion matrices” compared against the unconstrained (standard) HJ value function. Values represent the percentage of grid points in state space.

|                  | WC-HJ | HOCBF-HJ | Brake | Constant |
|------------------|-------|----------|-------|----------|
|                  | Safe  | Unsafe   | Safe  | Unsafe   | Safe  | Unsafe   |
| WC-HJ Safe       | 89.08 | 0        | 88.93 | 0.15     | 86.56 | 2.52     | 86.28 | 2.79   |
| WC-HJ Unsafe     | 10.92 | 1.52     | 9.40  | 9.43     | 1.30  | 9.62     | 1.52  | 9.40   |

TABLE III: HJ value statistics corresponding to various safety concepts evaluated on the NGSIM dataset.

| Safety Concept  | Mean | 0th | 5th | 50th | 95th | 100th |
|-----------------|------|-----|-----|------|------|-------|
| WC-HJ           | 4.3  | -0.38 | 0.24 | 4.04 | 9.36 | 13.67 |
| HOCBF-HJ        | 4.42 | -0.37 | 0.23 | 4.17 | 9.54 | 13.62 |
| Brake           | 4.96 | 0.13  | 0.65 | 4.52 | 10.94| 14.93 |
| Constant        | 4.52 | -0.01 | 0.55 | 4.21 | 9.52 | 13.83 |

restrictions imposed by the HOCBF constraints, the HOCBF-HJ unsafe set can be significantly less “wide” compared to WC-HJ. The skinniness of the HOCBF-HJ unsafe set reflects practical driving assumptions where we expect nearby drivers to stay within their lanes. Figure 5a also highlights the over-conservatism and impracticality of WC-HJ which would render cars traveling alongside agent A in adjacent lanes as unsafe. A volumetric comparison⁴ is provided in Table II where we compare the size of the safe/unsafe regions. We see that compared to the Brake and Constant safety concepts, the HOCBF-HJ unsafe set is the most contained within the WC-HJ unsafe set—the regions which WC-HJ deems unsafe but HOCBF-HJ / Brake / Constant deems safe are of a similar ballpark (between 1.30% to 1.49%), but the size of the region which WC-HJ deems safe and HOCBF-HJ unsafe is significantly lower compared to Brake and Constant (0.15% versus 2.52% and 2.79% respectively). Figure 5b illustrates the optimal control for a car-following scenario depicted in Figure 5a. The gray shaded region denotes the infeasible controls determined by the HOCBF constraint, and we see that the optimal HOCBF-HJ control (disturbance), shown by the green dot, maximizes (minimizes) \( \nabla V(x, t)^T g(x)u_A \) \( (\nabla V(x, t)^T h(x)u_HJ) \) while satisfying the HOCBF constraint. While the optimal control and disturbance for WC-HJ, denoted by the black dot, are the extremes of the control bounds. Figure 5b reflects intuitive and expected driving behaviors—it is unreasonable for agent A to assume that agent B will aggressively accelerate from behind.

3) Offline evaluation on real traffic data:
We evaluate how various safety concepts perform ex post facto on the NGSIM dataset [38], a highway driving dataset containing vehicles traveling along the US 101 highway in Los Angeles, USA. There are no collisions observed, and therefore we expect all the datapoints to be considered “safe”. Table III presents the mean and various percentile values of the HJ value of each safety concept. While it is difficult to definitively say that a particular safety concept is better than another based on these statistics since there are a lot of nuances in interpreting the HJ value, we make the following observations: (i) The Brake safety concept presents the highest values in all the statistics, indicating that it is the most optimistic out of the safety concepts studied. Being overly optimistic can lead to over-confident driving where the AV can overestimate how safe a situation is. (ii) For situations when the HJ value is low (i.e., for low percentiles), both the HOCBF-HJ and WC-HJ safety concept present very similar statistics, but the HOCBF-HJ statistics gradually become more similar to the Constant statistics as we consider higher percentile values. This suggests that (perhaps not too surprisingly) HOCBF-HJ behaves like a hybrid between the WC-HJ and Constant safety concepts.

VIII. LIMITATIONS, FUTURE WORK, AND CONCLUSIONS
We conclude by highlighting some limitations of this work, laying out exciting future directions, and summarizing our key contributions.

A. Limitations
A major limitation of this work stems from (i) the curse of dimensionality in solving the HJ PDE explicitly (as opposed to implicit representations [39]), and (ii) multi-agent extensions to the HOCBF and HJ reachability theory. Although we are currently using pairwise decomposition when considering multi-agent scenarios, pairwise decomposition for multi-agent analysis is effective in practice [5], [20], and in some settings, multi-agent analysis reduces to a pairwise decomposition [40]. While the eventual goal is to investigate using implicit representation (e.g., neural networks) when solving the HJ PDE to account for high dimensional scenarios, the (H)OCBF and HJ reachability formulation nonetheless provides useful inductive biases in building responsibility-aware safety concepts.

Our HOCBF-constrained safety concepts run the risk of encountering behaviors not abiding by the HOCBF constraint and therefore invalidating assurances provided by the synthesized safety concept. While we can further improve the behavior learning step to help reduce model mismatch, there have been complementary work in out-of-distribution (OOD) detection to catch anomalous events [41]. Such OOD detectors are designed to be used as a complementary safety filter and catch rare tail events that are not nominally considered by the system.

B. Future work
An area primed for future work is in investigating the division of responsibility for the constraint satisfaction of (10). A similar consideration is studied in [42] which analyzes how much each agent should contribute to constraint satisfaction in a multi-agent setting. Fruitful future work entails learning appropriate responsibility allocation within a game theoretic context [43], [44], and combining offline safety concept synthesis with online responsibility learning.

C. Conclusions
We have proposed a data-driven safety concept that is robust yet reflective of real-world driving interaction behaviors. We first learn control sets describing collision avoidance

⁴We considered regions where the vehicle speeds were \([15, 30] \text{ms}^{-1}\) and headings between \([-0.4\pi, 0.4\pi]\).
behaviors by leveraging high-order control barrier functions, and then use them to constrain the HJ reachability concepts in closed-loop AV operations, and extensions to multi-agent interactive scenarios, especially where the division of responsibility for collision avoidance is ambiguous.

REFERENCES

[1] “IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems,” IEEE Std 2846-2022, 2022.

[2] K. Leung, A. Bajcsy, E. Schmerling, and M. Pavone, “Towards the unification and data-driven synthesis of autonomous vehicle safety concepts,” Available at https://arxiv.org/abs/2107.14412, 2022.

[3] T. Phan-Minh, F. Howington, T.-S. Chu, S. U. Lee, M. S. Tomov, N. Li, C. Dicle, S. Findler, F. Suarez-Ruiz, R. Beaudoin, B. Yang, S. Omari, and E. M. Wolff, “Driving in real life with inverse reinforcement learning,” Available at https://arxiv.org/abs/2206.03004, 2022.

[4] K. Leung, E. Schmerling, M. Zhang, M. Chen, J. Talbot, J. C. Gerdes, and M. Pavone, “On infusing reachability-based safety assurance within planning frameworks for human-robot vehicle interactions,” Int. Journal of Robotics Research, vol. 39, pp. 1326–1345, 2020.

[5] X. Wang, K. Leung, and M. Pavone, “Infusing reachability-based safety into planning and control for multi-agent interactions,” in IEEE/RSJ Int. Conf. on Intelligent Robots & Systems, 2020.

[6] F. Oboril, C. Buerkle, A. Sussmann, S. Bitton, and S. Fabris, “MTBF Model for AVs – from perception errors to vehicle-level failures,” in IEEE Intelligent Vehicles Symposium, 2022.

[7] S. Topan, K. Leung, Y. Chen, P. Tüpek, E. Schmerling, J. Nilsson, M. Cox, and M. Pavone, “Interaction-dynamics-aware perception zones for obstacle detection safety evaluation,” in IEEE Intelligent Vehicles Symposium, 2022.

[8] L. M. Mitchell, A. M. Bayen, and C. J. Tomlin, “A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games,” IEEE Transactions on Automatic Control, vol. 50, no. 7, pp. 947–957, 2005.

[9] P. Abbeel and A. Y. Ng, “Apprenticeship learning via inverse reinforcement learning,” in Int. Conf. on Machine Learning, 2004.

[10] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, “Maximum entropy inverse reinforcement learning,” in Proc. AAAI Conf. on Artificial Intelligence, 2008.

[11] S. Levine and V. Koltun, “Continuous inverse optimal control with learned optimal samples,” in Int. Conf. on Machine Learning, 2012.

[12] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, “Planning for autonomous cars that leverage effects on human actions,” in Robotica: Science and Systems, 2016.

[13] W. Xiao and C. Beta, “High order control barrier functions,” IEEE Transactions on Automatic Control, vol. 67, no. 7, pp. 3655–3662, 2021.

[14] P. Fiorini and Z. Shiller, “Motion planning in dynamic environments using velocity obstacles,” Int. Journal of Robotics Research, vol. 17, no. 10-11, pp. 760–772, 1998.

[15] D. Wilkie, J. Van Den Berg, and D. Manocha, “Generalized velocity obstacles,” in IEEE/RSJ Int. Conf. on Intelligent Robots & Systems, 2009.

[16] P. Holmes, S. Kousik, B. Zhang, D. Raz, C. Barbalata, M. Johnson-Roberson, and R. Vasudevan, “Reachable sets for safe, real-time manipulator trajectory design,” in Robotics: Science and Systems, 2020.

[17] M. Althoff and J. Dolan, “Online verification of automated road vehicles using reachability analysis,” IEEE Transactions on Robotics, vol. 30, no. 4, pp. 903–918, 2014.

[18] Y. Kuwata, J. Teo, G. Fiore, S. Karaman, E. Frazzoli, and J. P. How, “Real-time motion planning with applications to autonomous urban driving,” IEEE Transactions on Control Systems Technology, vol. 17, no. 5, pp. 1105–1118, 2009.

[19] L. Janson, T. Hu, and M. Pavone, “Safe motion planning in unknown environments: Optimality benchmarks and tractable policies,” in Robotics: Science and Systems, 2012.

[20] D. Nisté, H.-L. Lee, J. Ng, and Y. Wang, “An introduction to the safety force field,” Available at https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/the-safety-force-field.pdf, 2019.

[21] S. Shalev-Shwartz, S. Shammah, and A. Shashua, “On a formal model of safe and scalable self-driving cars,” Available at https://arxiv.org/abs/1708.06374, 2017.

[22] D. Friedovich-Keil, A. Bajcsy, F. Fisac, J. K. M. Herbert, S. Wang, A. D. Dragan, and C. J. Tomlin, “Confidence-aware motion prediction for real-time collision avoidance,” 2019.

[23] R. Tian, L. Sun, A. Bajcsy, M. Tomizuka, and A. D. Dragan, “Safety assurances for human-robot interaction via confidence-aware game-theoretic human models,” in Proc. IEEE Conf. on Robotics and Automation, 2022.

[24] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, “Human motion trajectory prediction: A survey,” IEEE Trans. on Robotics, vol. 39, no. 8, pp. 1895–1906, 2021.

[25] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, “Control barrier functions: Theory and applications,” in European Control Conference, 2019.

[26] A. Robey, H. Hu, L. Lindemann, H. Zhang, D. Dimarogonas, S. T. and N. Matni, “Learning control barrier functions from expert demonstrations,” in Proc. IEEE Conf. on Decision and Control, 2020.

[27] A. Robey, L. Lindemann, S. T. and N. Matni, “Learning robust hybrid control barrier functions for uncertain systems,” in Hybrid Systems: Computation and Control, 2021.

[28] Z. Qin, K. Zhang, Y. Chen, C. Jingkai, and C. Fan, “Learning safe multi-agent control with decentralized neural barrier certificates,” in Int. Conf. on Learning Representations, 2021.

[29] Y. Lyu, W. Luo, and J. M. Dolan, “Adaptive safe merging control for heterogeneous autonomous vehicles using parametric control barrier functions,” Available at https://arxiv.org/abs/2202.09936, 2022.

[30] K. Margellos and J. Lygeros, “Hamilton-Jacobi formulation for reach-avoid differential games,” IEEE Transactions on Automatic Control, vol. 56, no. 8, pp. 1849–1861, 2011.

[31] J. Sill, “Monotonic networks,” in Conf. on Neural Information Processing Systems, 1997.

[32] A. Wehenkel and G. Louppe, “Unconstrained monotonic neural networks,” in Conf. on Neural Information Processing Systems, 2019.

[33] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, “JAX: composable transformations of Python+NumPy programs,” Available at http://github.com/google/jax, 2018.

[34] Y. Chen, S. Dathathri, T. Phan-Minh, and R. M. Murray, “Counter-example guided learning of bounds on environment behavior,” in Conf. on Robot Learning, 2019.

[35] J. von Neumann, “Zur Theorie der Gesellschaftsspiele,” Mathematische Annalen, vol. 100, no. 1, pp. 95–320, 1927.

[36] E. Schmerling, K. Leung, W. Vollprecht, and M. Pavone, “Multimodal probabilistic model-based planning for human-robot interaction,” in Proc. IEEE Conf. on Robotics and Automation, 2018.

[37] J. Dahl, G. Rodrigues de Campos, C. Olsson, and J. Fredriksson, “Collision Avoidance: A Literature Review on Threat-Assessment Techniques,” IEEE Transactions on Intelligent Vehicles, vol. 4, no. 1, pp. 101–113, 2019.

[38] USDOT, “Next generation simulation (NGSIM) vehicle trajectories and supporting data,” U.S. Department of Transportation, Intelligent Transportation Systems Joint Program, 2016, Available at: https://data.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Tracker/sect-6qj.

[39] S. Bansal and C. Tomlin, “DeepReach: A deep learning approach to high-dimensional reachability,” in Proc. IEEE Conf. on Robotics and Automation, 2021.

[40] Y. Chen, A. Singletary, and A. D. Ames, “Guaranteed obstacle avoidance for multi-robot operations with limited actuation: A control barrier function approach,” vol. 5, no. 1, pp. 127–132, 2021.

[41] A. Farid, S. Veer, and A. Majumdar, “Task-driven output-of-distribution detection with statistical guarantees for robot learning,” in Conf. on Robot Learning, 2021.

[42] K. Guo, D. Wang, T. Fan, and J. Pan, “VR-ORCA: Variable responsibility optimal reciprocal collision avoidance,” IEEE Robotics and Automation Letters, vol. 6, no. 3, pp. 4520–4527, 2021.

[43] I. Tsaknakis, M. Hong, and S. Zhang, “Minimax problems with coupled linear constraints: Computational complexity, duality and solution methods,” Available at https://arxiv.org/abs/2110.11210, 2021.

[44] D. Goktas and A. Greenwald, “Convex-concave min-max Stackelberg games,” in Conf. on Neural Information Processing Systems, 2021.