Discriminative Barrier Functions for Safe Adversarial Imitation Learning from Observation

Author Names Omitted for Anonymous Review. Paper-ID [add your ID here]

***Abstract*—Inverse Reinforcement Learning (IRL) algorithms are powerful tools for replicating expert behavior, but they rely on unconstrained exploration that renders them unsafe for real-world deployment. Meanwhile, Control Barrier Functions (CBFs) guarantee the safety of dynamical systems, but analytically designing them for complex environments remains tedious and challenging. In this work, we address these limitations jointly by leveraging CBFs as a structural inductive bias to constrain the search space for the reward function within an inverse reinforcement learning framework, specifically Adversarial Imitation Learning (AIL). This formulation yields a policy that exhibits safe online control while continuously improving through experience. Crucially, this framework enables the data-driven recovery of barrier functions directly from unlabeled expert observations. We demonstrate that the recovered barrier function is robust to unsafe states entirely absent from the expert data. Furthermore, we benchmark our method against standard IRL baselines in a simulated navigation environment, demonstrating improved safety performance. Finally, we investigate the trade-offs of planning-based versus policy-based AIL methods across both simulation and a real world obstacle avoidance task.**

I. Introduction

Adversarial Imitation Learning (AIL) is a principled framework for recovering reward functions that yield robust imitation policies [8, 15]. However, most AIL methods rely on action-labeled demonstrations, which are often unavailable or expensive to collect in real-world robotics, limiting their scalability [6, 20]. In contrast, humans are capable of learning to imitate behaviors directly from visual observation without access to the demonstrator's precise actions, improving through experience; in robot learning, this setting is known as Learning from Observation (LfO).

Moreover, humans also exhibit the ability to avoid undesirable states encountered during learning to imitate. We refer to such undesirable states as *unsafe observations*, in contrast to expert demonstrations representing *safe observations*. In cognitive science, the suppression of actions at unsafe states is associated with inhibitory control [23, 37, 18]. However, existing LfO methods focus solely on behavior fidelity and provide no explicit mechanism for avoiding unsafe states [34, 35, 13]. To address this limitation, we study the following problem within the LfO setting: *Given expert observational trajectories, how can a robot learn an imitative policy that also performs inhibitory control when encountering unsafe observations?*

To instantiate this notion of inhibitory control, we look to control theory, in particular, Control Barrier Functions (CBFs) [33]. CBF theory consists of a barrier function and a dynamics model, and, under certain conditions, it can render a set control-invariant (i.e., there exists a control that ensures the system remains inside the (safe) set indefinitely). Owing to these guarantees, CBFs have been widely employed as explicit safety layers in applications such as adaptive cruise control [4], aerial robotics [14], and legged locomotion [19]. However, manually designing task-specific barrier functions to form CBFs for safety-critical systems remains challenging, analogous to reward engineering in reinforcement learning (RL) [3, 9, 41].

In this work, we assert that constraining the reward function learning with the forward invariance constraint induced by CBFs in LfO algorithms leads to the recovery of both the barrier function and a policy that learns to inhibit its control near unsafe observations through experience. In particular, we employ a discrete form of control barrier functions to relax the continuity and differentiability requirements of the barrier function during training. *In this work, we propose DBFs (Discriminative Barrier Functions) for AIL, which recovers a learned barrier function and an imitative policy jointly from state-only expert observations, enabling safe online imitation learning without action and labeled observations.* We summarize our key contributions as follows:

Learning Interpretable Barrier Functions: We demonstrate that restricting the cost function class within the AIL framework to Discrete Control Barrier Functions [2] yields highly interpretable and generalizable barrier functions; this constrained objective to synthesize cost function through AIL is called DBFs. These functions successfully capture expert intent while inherently discriminating against unsafe observations, such as obstacles completely absent from the expert demonstrations.

Safe Online Learning from Observation: We show that incorporating DBF regularization into the reward function search significantly reduces collisions in online navigation settings. This mitigates the risks of environment interaction during LfO, paving the way toward safe online deployment in the real world.

Empirical Validation of the DBF Formulation: We test and compare DBF across both policy-based and planning-based AIL frameworks, demonstrating its efficacy in obstacle avoidance tasks through both simulation and real-world robotic experiments unseen during training time, thereby showcasing the generalization capability of DBF.

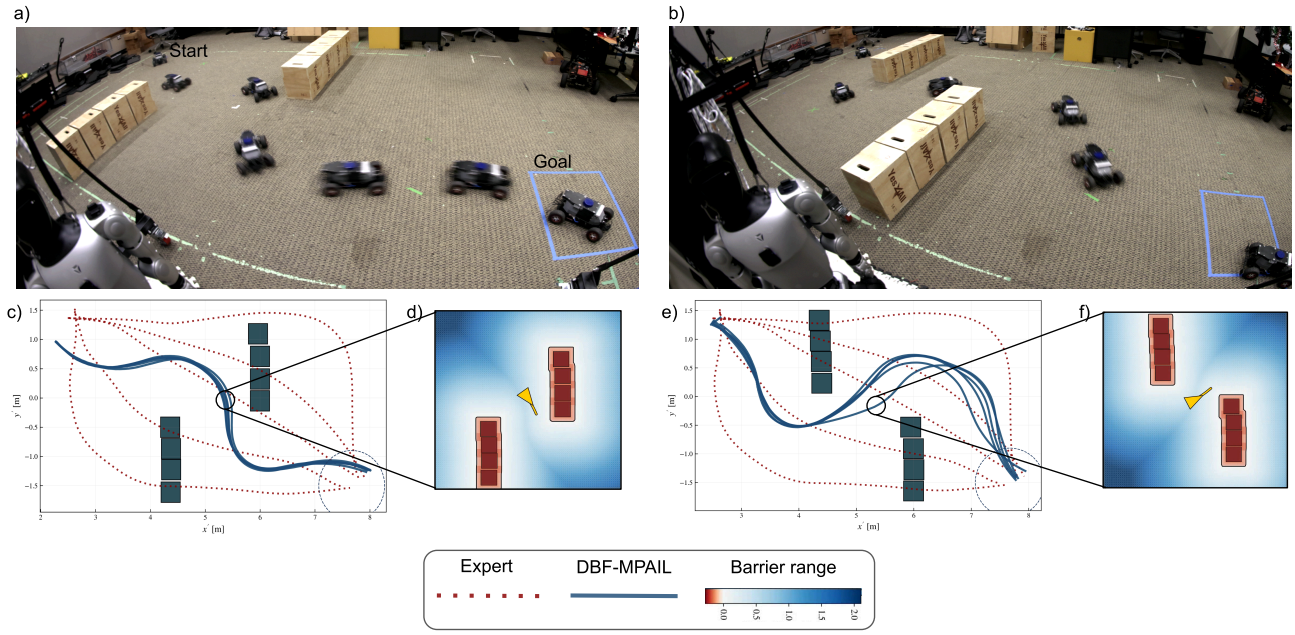

Fig. 1: Hardware deployment across two obstacle configurations unseen during training (a,b), demonstrating generalization to unseen unsafe observations. Robots are shown mid-inference navigating toward the goal. (c,e) Closed-loop trajectories for DBF based planning method with expert reference (dotted). (d,f) bird's-eye view map of the robot representing the barrier function range for the map region. See Figure 7 and Table II for performance on additional unseen scenarios.

II. Related Work

A. Learning Control Barrier Functions

CBFs can be challenging to design by hand for general, high-dimensional systems, yet their safety desiderata are intuitive: forward invariance within the safe set and recovery enforcement when violated [5]. These properties make CBFs an attractive structural object for robot learning, where they have already garnered substantial research attention [5, 11, 28]. A common approach to learning CBFs from data exploits discriminative boundaries between safe and unsafe regions [31, 26]. For instance, Srinivasan et al. [31] optimizes for a separating hyperplane in a kernel-induced feature space. Along a related line, Yang et al. [39, 40] employ Adversarial Inverse Reinforcement Learning (AIRL) as a pre-training step to obtain a policy, which is subsequently used to recover a CBF. More recently, Nakamura et al. [22] learn a value-function-based CBF from action-labeled demonstrations, similarly requiring a pre-specified safe set and offline supervision [29]. In contrast, a parallel line of work learns barrier-like certificates directly from demonstration data without explicit safe/unsafe labels [7, 26], or imposes conservatism toward out-of-distribution states [16, 21]. However, these methods uniformly require action-labeled demonstrations, operate strictly offline, and rely on a pre-specified unsafe set — limiting their applicability to settings where safety structure must be inferred from unlabeled observations alone. Our work addresses this gap by learning a discrete barrier function directly within the adversarial imitation loop, requiring neither action labels, safety labels, nor offline pretraining.

B. Safe Reinforcement and Imitation Learning

Towards high-dimensional systems that learn from interaction, safe RL methods aim to adhere to safety-oriented constraints while simultaneously maximizing reward [10]. These approaches naturally depend on a pre-specified constraint set and do not address settings where safety structure must be derived from data. SafeDICE [17] addresses offline safe imitation learning by correcting stationary distributions using labeled non-preferred demonstrations, avoiding constraint-violating behavior without explicit reward learning. However, it operates purely offline, requires action-labeled demonstrations and explicit non-preferred labels, and lacks any mechanism for online adaptation — assumptions and limitations our framework removes entirely. Closer to our method, CBF-RL [38] incorporates a hand-specified CBF as a constraint term in the policy loss during RL training, internalizing safety without requiring a runtime safety filter at inference. However, the CBF itself must be provided a priori and is not learned from data. While our DBF cost plays an analogous role — discouraging actions that push toward unsafe regions — it is learned entirely from unlabeled observations rather than derived from a hand-specified model, and operates within an adversarial imitation loop rather than a pure RL setting.

III. Discriminative Barrier Functions

This section describes the methodology of learning barrier function through AIL framework.

A. Preliminaries:

Consider a Partially Observable Markov Decision Process (POMDP). In an unknown world state $s_w \in S_w$, the agent makes an observation $o \sim p(o|s_w)$. From a history of observations $\mathbf{o}_{0:t}$, the agent perceives its *state* $s_t \sim p(s|\mathbf{o}_{0:t})$ and performs *action* $a \in \mathcal{A}$. Through a predictive model $f : S \times \mathcal{A} \to S$, the agent self-predicts forward in time for a model-based method. Partial observations of S are desirably used for demonstrations to perform IRL and AIL, as full observation history would quickly become intractable [36]. The expert π_E may be acting to optimize some reward based on current and next state $r_w(s, s')$ in the POMDP, which may also potentially encode safety constraints. While in this work we do not assume its availability, known constraints can be employed seamlessly as we discuss in Section III-E.

B. Inverse Reinforcement Learning from Observation

While there are many IRL formulations from which to begin, Ho and Ermon [15] show how apprenticeship learning with regularization unifies various perspectives on IRL. In short, this objective seeks cost functions $c(\cdot)$ under which the performance of the expert is maximized while other *optimal* policy (i.e. learner, π) candidates are minimized:

$$\begin{aligned}\text{IRLfO}_\psi(\pi_E) = &\arg\max_{c\in\mathbb{R}^{S\times S}} -\psi(c) \\ &+ \underbrace{\min_{\pi\in\Pi} -\lambda \text{H}(\pi) + \mathbb{E}_\pi[c(s,s')]}_{\text{Entropy-regularized Policy Optimization}} \\ &- \mathbb{E}_{\pi_E}[c(s,s')], \end{aligned} \tag{1}$$

where the observation-only cost $c(s, s')$ [35] is introduced, reflecting the absence of expert action data in our problem setting.

It is through the cost regularizer, $\psi : \mathbb{R}^{S\times S} \to \mathbb{R}$, that Ho and Ermon instantiate other IRL algorithms. Likewise, to introduce our specific form of regularization for our DBF formulation, a brief introduction to CBFs is required.

C. Discrete-Time Control Barrier Functions

Let safe states be the set $\mathcal{S}$ and unsafe states $\mathcal{U}$. Towards barrier functions, we also introduce the *specification function* $h(\cdot)$ such that, $\mathcal{S} := \{s \in \mathcal{S} \mid h(s) \geq 0\}$. Finally, h can be certified as a Discrete-Time CBF if

$$\sup_{a\in\mathcal{A}} h(f(s,a)) - h(s) \geq -\alpha(h(s)), \tag{2}$$

where α is a class $\mathcal{K}$ function. Under this condition, CBFs guarantee that there is an action which maintains forward invariance and stability. For instance: if $h(s_t) < 0$ (i.e. s_t is unsafe), then there is an action for which the agent can increase h for s_{t+1} thereby returning to safety; if $h(s_t) \to 0^+$ (i.e. agent approaches unsafe from safe states), the allowable rate at which the agent may continue approaching $\mathcal{U}$ goes to 0 thereby stably avoiding unsafe states.

D. Discriminative Barrier Functions via Cost Regularization

Now equipped with the IRL optimization objective and CBF safety constraints, we can finally proceed with the Discriminative Barrier Function (DBF) formulation. First, we observe a simple relationship between AIL and CBFs, which will then allow us to view the CBF as a constraint on the AIL objective.

In AIL, an ideal discriminator D^* achieves a boundary which perfectly separates expert states $s \in \mathcal{D}_E$ from learner states $s \in \mathcal{D}_\pi$. Meanwhile for CBFs, the barrier function h can be viewed as separating safe states $s \in \mathcal{S}$ from unsafe states $s \in \mathcal{U}$ 2. These binary classification objectives appear adjacent: *can a barrier function act as a discriminator?* Rather than only learning to discriminate expert and learner data, we wish to also discriminate safe from unsafe data through enforcing control-theoretic structure in the discriminator. Namely, we wish to classify a transition (s, s') based on the condition in Equation (2), through the *dynamic constraint Lagrangian* $q(s, s')$:

$$q_h(s,s') := h(s') - h(s) + \alpha(h(s)) \tag{3}$$

which should be more positive for safe transitions that satisfy the constraint and more negative for unsafe transitions that violate the constraint.

Returning our attention to the IRL objective in Equation (1), consider restricting the cost function space to R^S than in standard AIL dependent on $R^{S\times S}$.

$$\begin{aligned}\mathcal{C}_{\text{CBF}} := \{c \mid c(s,s') = q_h(s,s'),\ h \in \mathbb{R}^S \\ \text{s.t.}\quad h(s) > 0,\ \forall s \in \mathcal{S} \\ h(s) < 0,\ \forall s \in \mathcal{U}\}.\end{aligned} \tag{4}$$

That is, we wish to search through cost function candidates that are *also Control Barrier Functions* (thereby also imposing control-theoretic structure on the cost function). By combining this restriction on c with the convex regularization from Ho and Ermon [15], as derived in Appendix A, we obtain the *DBF objective*,

$$\begin{aligned}\arg\max_{h\in\mathbb{R}^S} \quad &\underbrace{\mathbb{E}_{\tau_{\text{safe}}}[D_h(s,s')] - \mathbb{E}_{\tau_{\text{unsafe}}}[D_h(s,s')]}_{\mathcal{L}_{\text{Wasserstein-GAN}}} \\ &- \underbrace{\mathbb{E}_{\hat{\mathcal{T}}}\left[\left(\left\|\nabla_{(s,s')} D_h(s,s')\right\|_2 - 1\right)^2\right]}_{\mathcal{L}_{\text{Gradient Penalty}}} \\ &\text{s.t.} \quad \begin{cases} h(s) > 0, & \forall s \in \mathcal{S}_{\text{safe}} \\ h(s) < 0, & \forall s \in \mathcal{S}_{\text{unsafe}} \end{cases}\end{aligned} \tag{5}$$

where $D_h(s, s') = q_h(s, s')$, $\hat{\mathcal{T}} \sim \eta\mathcal{T}_{safe} + (1-\eta)\mathcal{T}_{unsafe}$ and $\sigma(x) := (1 + e^{-x})^{-1}$ is the sigmoid function. Note that for the barrier function, to achieve stable training of GAN's we use a Wasserstein GAN formulation, which encourage 1-Lipschitz continuity via a gradient penalty.

E. Learning DBFs via Adversarial Imitation Learning

What remains is how this objective should be computed. Since we have not assumed that safety constraints are known,

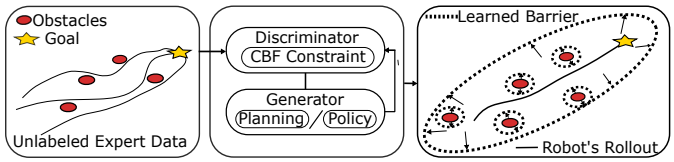

Fig. 2: Overview of DBF: a barrier function and imitation policy are recovered jointly from unlabeled expert observations within an AIL framework. The discriminator enforces CBF constraints, classifying expert states as safe and learner rollouts as unsafe, while the generator (planning or policy) optimizes against the learned barrier.

we make the simplifying assumptions that (i) learner data is unsafe, $\mathcal{T}_{\text{unsafe}} \sim \pi$ and (ii) expert data is safe, $\mathcal{T}_{\text{safe}} \sim \pi_E$. Meaning, we can then employ Adversarial Imitation Learning to compute the expectations in Equation (5). Altogether, the AIL objective with DBFs can be summarized:

$$\begin{aligned}\min_{\pi\in\Pi}\max_{h\in\mathbb{R}^S} \quad & \lambda_{\text{wgan}}\Big[\mathbb{E}_{\pi_E}[D_h(s,s')] - \mathbb{E}_{\pi}[D_h(s,s')]\Big] \\ & - \lambda_{\text{GP}}\,\mathbb{E}_{\hat{\pi}}\Big[(\|\nabla_{(s,s')}D_h(s,s')\|_2 - 1)^2\Big] - \mathrm{H}(\pi) \\ & + \lambda_{\text{sign}}\Big[\underbrace{\mathbb{E}_{s\sim\pi_E}\big[(\delta + h(s))_+\big]}_{\textbf{hinge loss}} \\ & \qquad + \underbrace{\mathbb{E}_{s\sim\pi}\big[(\delta - h(s))_+\big]}\Big] \end{aligned} \tag{6}$$

This formulation optimizes the DBF directly from data without requiring explicit, pre-defined safety labels, leveraging the adversarial framework to dynamically separate safe and unsafe states.

In the setting where safety constraints *are known*, it is straightforward to incorporate this additional information, for example, by also populating $\mathcal{T}_{\text{safe}}$ conditionally from the learner or by decomposing h(s) = $\min\{\bar{h}(s), h_\theta(s)\}$ where $\bar{h}$ and h_θ are prior and learned barrier functions, respectively. In addition, all demonstrations need not be safe as assumed by $\mathcal{T}_{\text{safe}} \sim \pi_E$. It is straightforward to include demonstrations that are unsafe directly in $\mathcal{T}_{\text{unsafe}}$. While we do not explore these extensions here due to their assumption of additional safety knowledge, they remain as promising future work in more applied settings.

As in Robey et al. [26], our intuition behind states being drawn from closed and non-empty safety sets requires notions of ϵ-nets. Put simply, each demonstrated state subsumes an ϵ ball of states around it, implying that the expert operates with some minimum distance to the unsafe boundary. This is equivalent to constraining the value of the barrier function at safe and unsafe states while limiting the Lipschitz constant of the function, which (along with gradient penalties) is coincidentally common practice for stabilizing AIL algorithms [25].

IV. Experimental Results

We evaluate the DBF objective with different AIL algorithms, including GAIL [15], AIRL [8], and MPAIL [13].

Algorithm	Reward	Learner
GAIL	$\log D$	PPO (Policy)
AIRL	$\sigma^{-1} \circ D$	PPO (Policy)
MPAIL	$\sigma^{-1} \circ D$	MPPI (Planner)

TABLE I: AIL Algorithm Summaries

These algorithms are primarily differentiated by their reward structure and whether they are policy-based or planning-based methods table I. GAIL and AIRL are model-free AIL algorithms, and the recently introduced MPAIL [12] is a model-based AIL method that performs well in out-of-distribution recovery through approximated policy value function bootstrapping at the terminal state, with a reward function similar to AIRL used as a stage cost. In experiments we aim to answer the following questions:

- **Q1** Can DBFs learn barrier functions that generalize to novel environments from unlabeled demonstrations unseen during training?
- **Q2** How do DBF-augmented AIL algorithms compare against existing AIL baselines in terms of safety?
- **Q3** Can DBFs improve safety in real-world and safety-critical environments?

A. Navigation In Simulated Environment:

Safe exploration is critical in online IRL, where the agent must learn both the policy and the reward function while robustly distinguishing safe from unsafe states [1, 39]. To this end, we first qualitatively analyze the learned boundary around expert trajectory in a simple no-obstacle simulation setting, where unsafe states can be defined as those not present in the expert data. We use MUSHR [30], a 10-DOF ego vehicle, as a testbed in the Isaac Lab simulation environment for navigation to a goal point. Figure 3 compares the barrier function learned by DBF-based methods, $h(s)$, with the single-state (position-only) reward function, $r(s)$, used by baseline AIL methods for learning the discriminative boundary around the expert trajectory. This comparison is necessary because, in standard AIL, the reward function may also encode safety constraints. We retain the circling motion near the goal in the expert trajectory for qualitative comparison, since this behavior is more difficult to imitate from state-only observations, as corroborated by [24]. We observe that in Figure 3, AIL algorithms equipped with DBFs generalize substantially better around the expert demonstrations, capturing the occupancy of a single trajectory within the boundary while also correctly modeling the difficult-to-classify [24] turn near the goal. In contrast, the baselines collapse and overestimate the discriminative boundary.

B. Q1: Barrier Function Robustness to unsafe states not present in expert data

Having observed that DBF-based methods learn tighter discriminative boundaries, we next analyze the obstacle-avoidance task using unlabeled expert trajectories, as shown in Figure 8, to address the above question. For this study,

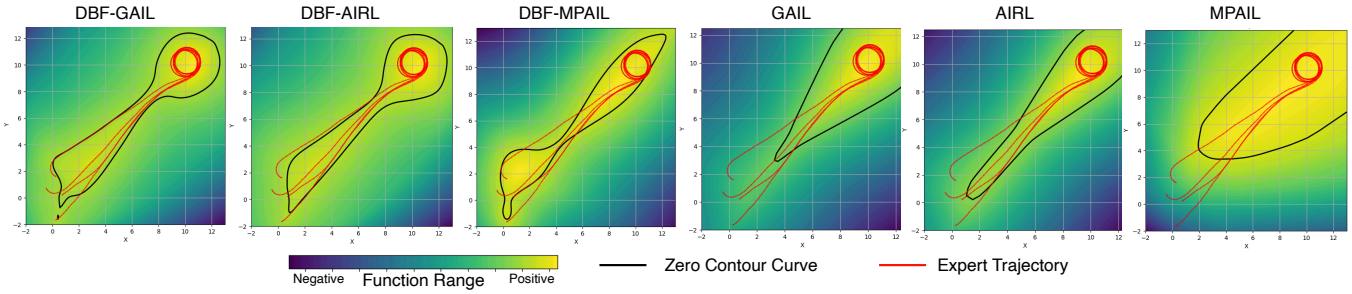

Fig. 3: Qualitative Comparison of GAIL, AIRL, and MPAIL with and without DBF structure in learned reward function.

we use a maze environment enclosing the expert trajectories, with horizontal wall positions on the left and right sides randomized during training, as shown in Figure 4b, and a vertical wall placed at the center of the environment during test time to investigate generalization, as shown in Figure 4d. During training, the expert data is classified as safe; however, while learning to imitate in the simulation environment, the robot may encounter unseen obstacles that hinder the path caused by the walls. As shown in Figure 4 parts a and c, DBF-based methods preserve meaningful notions of safety even under these distribution shifts. Although portions of the expert trajectories intersect obstacle regions, the learned barrier function adapts by rejecting unsafe transitions while maintaining connectivity to the demonstrated safe set. The continuity imposed by the barrier objective forces the learned safe set to remain topologically consistent around the obstacle. As a result, the DBF typically commits to one side of the obstacle as safe while classifying the other side as unsafe, even though both routes are geometrically feasible. This behavior suggests that DBFs are not merely memorizing demonstrated states, but instead learning coherent safety structure that extrapolates beyond the expert-data distribution.

C. Q2: Performance Comparison and Safety Analysis

We compare all algorithms in a maze environment (Figure 4b) with randomized wall positions, evaluating both imitation performance via normalized reward and safety via cost rate [1] — a metric that quantifies constraint-violation regret throughout learning (Equation (7)).

$$\text{CR} = \frac{\sum_{i=0}^{N_{\text{iter}}} \sum_{t=0}^{N_{\text{int}}} c_{it}}{(1+n_{\text{iter}})(1+n_{\text{int}})} \tag{7}$$

As shown in Figure 5, DBF-based methods consistently reduce cost rate as learning progresses, while GAIL collapses entirely — failing to imitate the expert and accumulating high constraint violations. MPAIL, as a planning-based method, shows greater robustness to out-of-distribution obstacle states than policy-based GAIL, but achieves higher cost rate than DBF-based methods. This is because DBF-based methods constrain the reward function search to satisfy the CBF dynamic constraint (Equation (5)), encouraging both policy and planning-based methods to learn safe imitative behavior with minimal regret. Notably, DBF-GAIL achieves the lowest cost rate at convergence, while DBF-MPAIL trades marginally higher cost rate for significantly better reward than DBF-GAIL. Although MPAIL achieves the highest reward, it does so at the cost of higher constraint violations compared to DBF-based methods, highlighting the safety-performance trade-off that DBFs are designed to address.

Effect of different $\mathcal{K}$-class functions: AIL addresses the occupancy matching problem relative to the expert distribution, with the learned reward serving as a dynamic constraint for the Control Barrier Function (CBF). Because this reward directly affects safe exploration, we quantify its impact using the success rate and cost rate. In our experiments, discriminator logit outputs remained in the range $[-0.5, 0.0]$ across all considered $\mathcal{K}$-class functions (linear, cubic, and exponential) in training environment Figure 4b. Within this range, the slopes of these functions follow the order linear $>$ cubic $>$ exponential. As the reward is learned, the cost-function logits for generated trajectories shift from large negative values toward zero. This progression is reflected in the safety metrics, including cost rate and total collisions per environment, as shown in Figure 6. The figure indicates that cost rate decreases fastest for linear and slowest for exponential, consistent with their slope ordering. While total collisions per environment continue to accumulate throughout learning, linear accumulates them at the lowest rate, further corroborating its superior safety performance.

D. Q3: Real world Deployment

Our hardware experiment evaluates GAIL, MPAIL, DBF-GAIL, and DBF-MPAIL through Real-to-Sim-to-Real: 1) a small number of state-only (position) observations are collected from the real-world, 2) the method is trained using interactions from simulation with randomized obstacle positions, finally 3) the method is deployed zero-shot to the real-world for evaluation. We find that DBF-MPAIL consistently synthesizes trajectories not present in the expert demonstrations as shown in Figure 1 to avoid obstacle whilst imitating to reach the goal. Unlike DBF-MPAIL, which plans over the barrier landscape, policy-based methods take positional state and a BEV map directly as input to a black-box policy. In addition, Figure 1 and Figure 7 illustrate a key advantage of planning-based DBF-AIL – success in imitating goal-reaching behavior whilst avoiding obstacles unseen in expert data and unseen positions

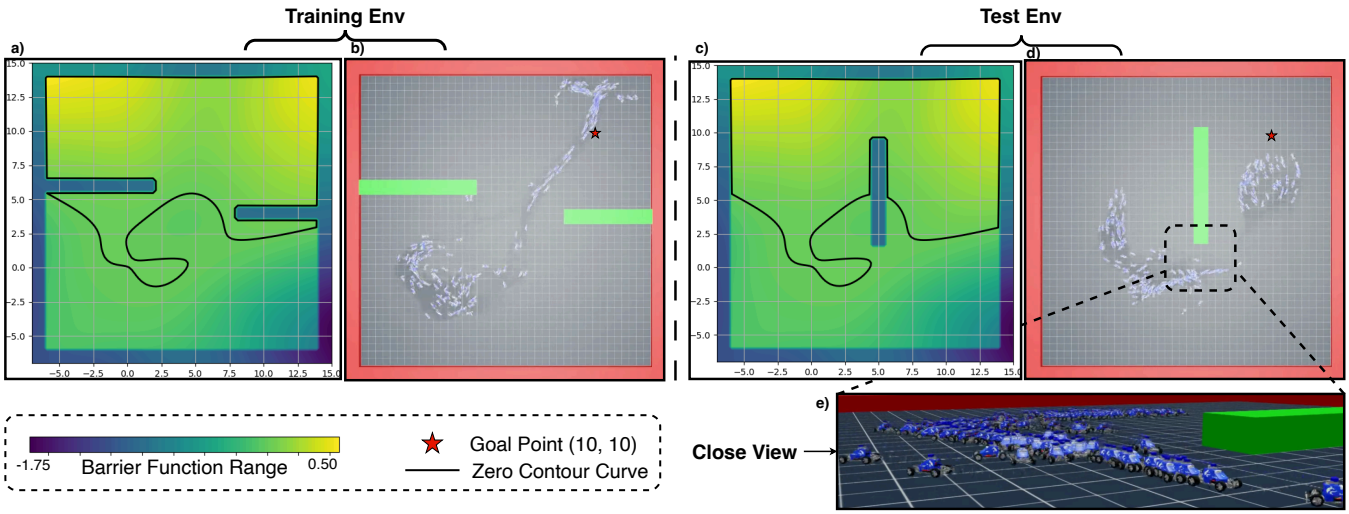

Fig. 4: Qualitative analysis of barrier function generalization synthesized via DBF-MPAIL. (a) Learned barrier function in the training environment using expert trajectories from Figure 8. **(b) (b)** Top view of the training environment showing robot trajectories clustering around the expert-demonstrated safe set near the goal. **(c)** The same barrier function transferred to a novel test environment, demonstrating topological adaptation around the unseen obstacle. **(d)** Top view of the test environment showing agent trajectories avoiding the novel obstacle, with the dashed box indicating the zoomed region shown in (e). **(e)** Close-up view of robots inhibiting their control inputs to avoid an obstacle in an unseen environment.

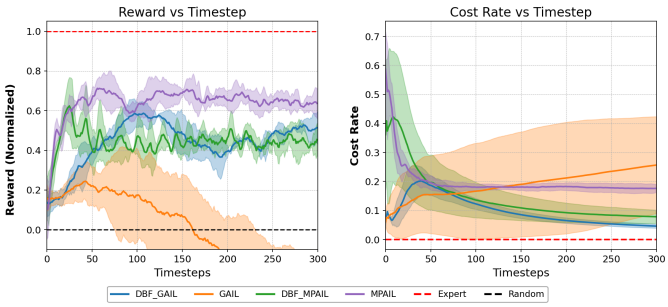

Fig. 5: Comparison of AIL algorithms with and without DBF.

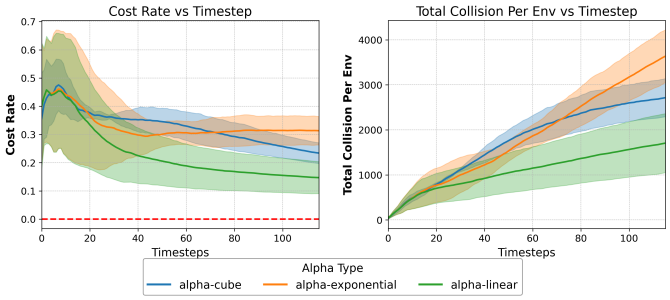

Fig. 6: Effect of $\mathcal{K}$-class functions, tested on DBF-MPAIL.

in the training environment. By granting access to the agent's optimization landscape, DBF-MPAIL significantly improves on the interpretability of agents trained through unlabeled and suboptimal data when compared to black-box policies. We find that although DBF-GAIL success rate in reaching the goal is low due to the sim-real gap and lack of structure in planning,

TABLE II: Hardware deployment results (5 trials per variant across four obstacle configurations (Figures 1 and 7)). DBF variants achieve zero collisions across all trials. d (m) denotes the mean minimum distance to obstacles throughout each trial.

Variant	Succ.	Coll.	Min d (m)
Without DBF			
GAIL	0/20 (0%)	10/20 (50%)	0.035 ± 0.274
MPAIL	0/20 (0%)	6/20 (30%)	0.337 ± 0.534
With DBF			
DBF-GAIL	1/20 (5%)	**0/20 (0%)**	0.500 ± 0.437
DBF-MPAIL	**13/20 (65%)**	**0/20 (0%)**	0.188 ± 0.138

it achieves zero collisions across all obstacle configurations, unlike non-DBF baselines as quantified in Table II. During deployment, GAIL's policy consistently hits obstacles or veer off the expert demonstrated path, limiting imitation objective. While literature on the evaluation of AIL methods in the real-world are sparse, we find that AIL policies can be extremely poor performing in the real-world, as corroborated by [32].

V. Conclusion

In this work, we introduce Discriminative Barrier Functions (DBFs), a framework for incorporating Control Barrier Function (CBF) theory directly into the AIL framework. By constraining the learned discriminator through barrier-function objectives, DBFs enable AIL methods to recover notions of safety and inhibitory control directly from unlabeled demonstrations without requiring explicit unsafe labels or hand-engineered safety costs. Across both policy-based and planning-based AIL algorithms, we demonstrate that DBFs improve the generalization of learned safe sets beyond the expert

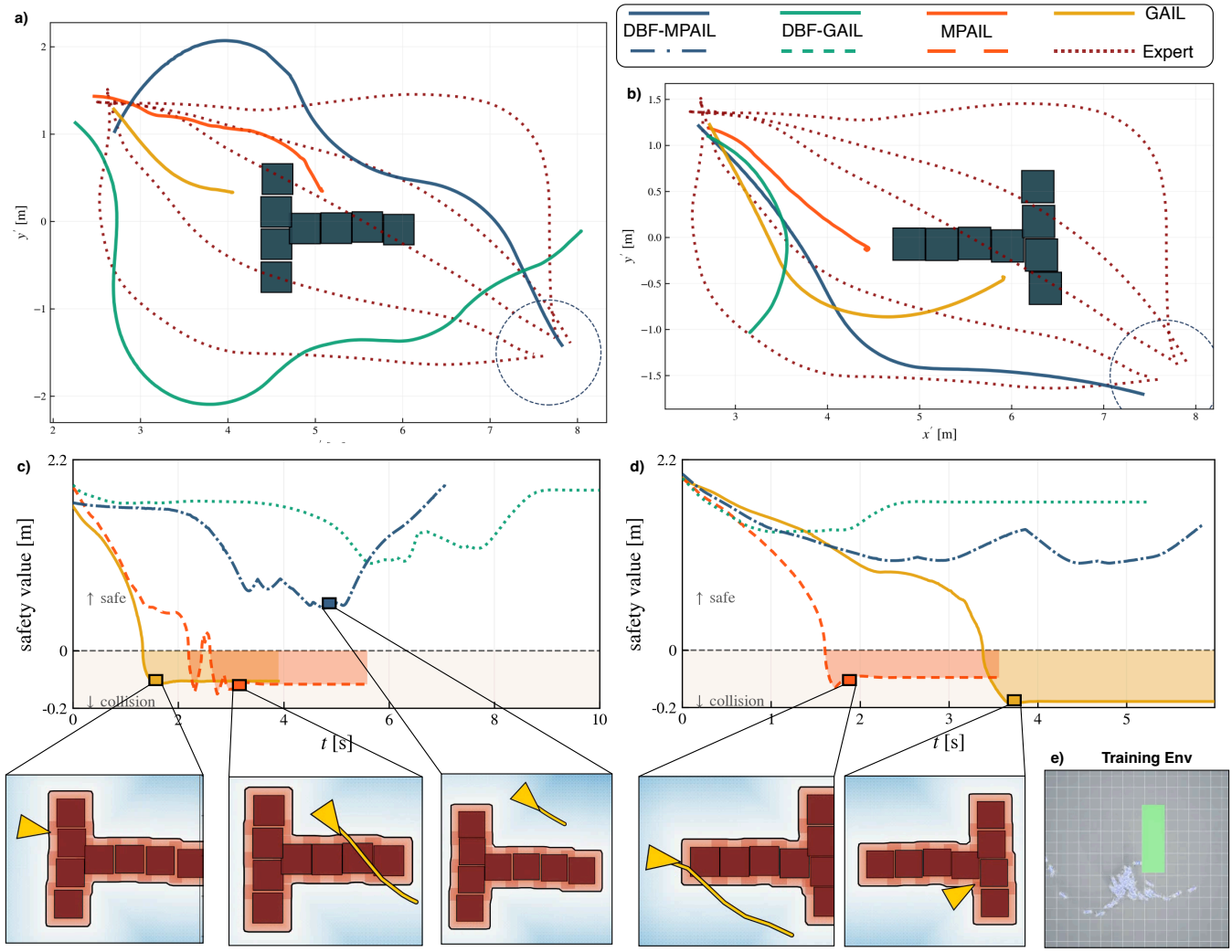

Fig. 7: Hardware deployment across two additional obstacle configurations beyond Figure 1. (a,b) Closed-loop trajectories for all variants with expert reference (dotted). (c,d) Safety value represented by the minimum distance to all obstacles. Insets show robot state at near-collision timesteps. (e) Training environment used to train all methods for Real-Sim-Real experiments.

data distribution without sacrificing imitation performance. DBFs recover coherent safety boundaries, exhibit robustness to previously unseen obstacles and environmental changes, and induce safer exploration during online learning in both simulation and the real world. Overall, our results suggest that incorporating control-theoretic structure into adversarial imitation learning provides a promising direction toward safer and more robust online learning from observation.

VI. Limitations and Future Work

Although DBFs effectively capture low-coverage regions of the expert distribution, they may become overly conservative when expert data covers only a narrow subset of valid strategies. In such cases, the learned inhibitory structure may incorrectly suppress behaviors that are safe but underrepresented in the data. Future work may investigate uncertainty-aware DBFs, active data collection, or mechanisms for reasoning about demonstration coverage and ambiguity.

The $\mathcal{K}$-class function is currently specified by hand, which provides a simple and stable mechanism for shaping inhibitory behavior but can influence optimization dynamics and the resulting safe set geometry. Prior work suggests these functions may themselves be learnable [27], potentially enabling more adaptive barrier representations across tasks with complex or heterogeneous safety constraints — a direction we leave for future work.

Acknowledgments

References

[1] J. Achiam A. Ray and D. Amodei. Benchmarking safe exploration in deep reinforcement learning. URL https://cdn.openai.com/safexp-short.pdf.

[2] Ayush Agrawal and Koushil Sreenath. Discrete control barrier functions for safety-critical control of discrete

systems with application to bipedal robot navigation. In *Robotics: Science and Systems*, volume 13, pages 1–10. Cambridge, MA, USA, 2017.

[3] Prithvi Akella, Apurva Badithela, Richard M. Murray, and Aaron D. Ames. Lipschitz continuity of signal temporal logic robustness measures: Synthesizing control barrier functions from one expert demonstration, 2023. URL https://arxiv.org/abs/2304.03849.

[4] Aaron D. Ames, Jessy W. Grizzle, and Paulo Tabuada. Control barrier function based quadratic programs with application to adaptive cruise control. In *53rd IEEE Conference on Decision and Control*, pages 6271–6278, 2014. doi: 10.1109/CDC.2014.7040372.

[5] Aaron D. Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control Barrier Functions: Theory and Applications. In *2019 18th European Control Conference (ECC)*, pages 3420–3431, June 2019. doi: 10.23919/ECC.2019.8796030. URL https://ieeexplore.ieee.org/abstract/document/8796030.

[6] Returaj Burnwal, Hriday Mehta, Nirav Pravinbhai Bhatt, and Balaraman Ravindran. Learning from observation: A survey of recent advances, 2025. URL https://arxiv.org/abs/2509.19379.

[7] Rowan McAllister Koushil Sreenath Adrien Gaidon Fernando Castañeda, Haruki Nishimura. In-distribution barrier functions: Self-supervised policy filters that avoid out-of-distribution states, January 2023. URL https://arxiv.org/abs/2301.12012. arXiv:2301.12012 [cs.RO].

[8] Justin Fu, Katie Luo, and Sergey Levine. Learning Robust Rewards with Adverserial Inverse Reinforcement Learning. February 2018. URL https://openreview.net/forum?id=rkHywl-A-.

[9] Kunal Garg, James Usevitch, Joseph Breeden, Mitchell Black, Devansh Agrawal, Hardik Parwana, and Dimitra Panagou. Advances in the theory of control barrier functions: Addressing practical challenges in safe control synthesis for autonomous and robotic systems, 2023. URL https://arxiv.org/abs/2312.16719.

[10] Shangding Gu, Long Yang, Yali Du, Guang Chen, Florian Walter, Jun Wang, and Alois Knoll. A Review of Safe Reinforcement Learning: Methods, Theories, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):11216–11235, December 2024. ISSN 1939-3539. doi: 10.1109/TPAMI.2024.3457538. URL https://ieeexplore.ieee.org/document/10675394/.

[11] Maeva Guerrier, Hassan Fouad, and Giovanni Beltrame. Learning Control Barrier Functions and their application in Reinforcement Learning: A Survey, April 2024. URL http://arxiv.org/abs/2404.16879. arXiv:2404.16879 [cs].

[12] Tyler Han, Alex Liu, Anqi Li, Alexander Spitzer, Guanya Shi, and Byron Boots. Model Predictive Control for Aggressive Driving Over Uneven Terrain. In *Robotics: Science and Systems XX*. Robotics: Science and Systems Foundation, July 2024. ISBN 9798990284807. doi: 10.15607/RSS.2024.XX.022. URL http://www.roboticsproceedings.org/rss20/p022.pdf.

[13] Tyler Han, Yanda Bao, Bhaumik Mehta, Gabriel Guo, Anubhav Vishwakarma, Emily Kang, Sanghun Jung, Rosario Scalise, Jason Zhou, Bryan Xu, and Byron Boots. Model Predictive Adversarial Imitation Learning for Planning from Observation, July 2025. URL http://arxiv.org/abs/2507.21533. arXiv:2507.21533 [cs].

[14] Marvin Harms, Mihir Kulkarni, Nikhil Khedekar, Martin Jacquet, and Kostas Alexis. Neural control barrier functions for safe navigation, 2024. URL https://arxiv.org/abs/2407.19907.

[15] Jonathan Ho and Stefano Ermon. Generative Adversarial Imitation Learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/hash/cc7e2b878868cbae992d1fb743995d8f-Abstract.html.

[16] Hussein Sibai Ihab Tabbara. Learning neural control barrier functions from offline data with conservatism, September 2025. URL https://arxiv.org/abs/2505.00908v1. arXiv:2505.00908 [cs.LG].

[17] Youngsoo Jang, Geon-Hyeong Kim, Jongmin Lee, Sungryull Sohn, Byoungjip Kim, Honglak Lee, and Moontae Lee. Safedice: Offline safe imitation learning with non-preferred demonstrations. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 74921–74951. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/ed2fb79f2664c3d9ba878be7e575b2af-Paper-Conference.pdf.

[18] Armin Lederer, Erfaun Noorani, John S. Baras, and Sandra Hirche. Risk-sensitive inhibitory control for safe reinforcement learning. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 1040–1045, 2023. doi: 10.1109/CDC49753.2023.10383524.

[19] Qiayuan Liao, Zhongyu Li, Akshay Thirugnanam, Jun Zeng, and Koushil Sreenath. Walking in narrow spaces: Safety-critical locomotion control for quadrupedal robots with duality-based optimization, 2023. URL https://arxiv.org/abs/2212.14199.

[20] YuXuan Liu, Abhishek Gupta, Pieter Abbeel, and Sergey Levine. Imitation from observation: Learning to imitate behaviors from raw video via context translation, 2018. URL https://arxiv.org/abs/1707.03374.

[21] Aditya Singh Shishir Kolathaya Ravi Prakash Mumuksh Tayal, Manan Tayal. V-ocbf: Learning safety filters from offline data via value-guided offline control barrier functions, April 2026. URL https://arxiv.org/abs/2512.10822. arXiv:2512.10822 [cs.AI].

[22] Kensuke Nakamura, Arun L. Bishop, Steven Man, Aaron M. Johnson, Zachary Manchester, and Andrea Bajcsy. How to train your latent control barrier function: Smooth safety filtering under hard-to-model constraints.

In *Learning for Dynamics and Control*, 2026.

[23] Joel T. Nigg. On Inhibition/Disinhibition in Developmental Psychopathology: Views from Cognitive and Personality Psychology and a Working Inhibition Taxonomy. *Psychological Bulletin*, 126(2):220–246, 2000.

[24] Manu Orsini, Anton Raichuk, Leonard Hussenot, Damien Vincent, Robert Dadashi, Sertan Girgin, Matthieu Geist, Olivier Bachem, Olivier Pietquin, and Marcin Andrychowicz. What matters for adversarial imitation learning? In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 14656–14668. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/7b647a7d88f4d6319bf0d600d168dbeb-Paper.pdf.

[25] Manu Orsini, Anton Raichuk, Leonard Hussenot, Damien Vincent, Robert Dadashi, Sertan Girgin, Matthieu Geist, Olivier Bachem, Olivier Pietquin, and Marcin Andrychowicz. What Matters for Adversarial Imitation Learning? In *Advances in Neural Information Processing Systems*, volume 34, pages 14656–14668. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/7b647a7d88f4d6319bf0d600d168dbeb-Abstract.html.

[26] Alexander Robey, Haimin Hu, Lars Lindemann, Hanwen Zhang, Dimos V. Dimarogonas, Stephen Tu, and Nikolai Matni. Learning Control Barrier Functions from Expert Demonstrations. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3717–3724, December 2020. doi: 10.1109/CDC42340.2020.9303785. URL https://ieeexplore.ieee.org/document/9303785/. ISSN: 2576-2370.

[27] Vivek Sharma, Negar Mehr, and Naira Hovakimyan. Learning differentiable and safe multi-robot control for generalization to novel environments using control barrier functions, 2024.

[28] Oswin So, Zachary Serlin, Makai Mann, Jake Gonzales, Kwesi Rutledge, Nicholas Roy, and Chuchu Fan. How to train your neural control barrier function: Learning safety filters for complex input-constrained systems, 2023. URL https://arxiv.org/abs/2310.15478.

[29] Oswin So, Zachary Serlin, Makai Mann, Jake Gonzales, Kwesi Rutledge, Nicholas Roy, and Chuchu Fan. How to train your neural control barrier function: Learning safety filters for complex input-constrained systems. In *2024 IEEE International Conference on Robotics and Automation (ICRA) (Under Review)*, 2024.

[30] Siddhartha S. Srinivasa, Patrick Lancaster, Johan Michalove, Matt Schmittle, Colin Summers, Matthew Rockett, Joshua R. Smith, Sanjiban Chouhury, Christoforos Mavrogiannis, and Fereshteh Sadeghi. MuSHR: A low-cost, open-source robotic racecar for education and research. *CoRR*, abs/1908.08031, 2019.

[31] Mohit Srinivasan, Amogh Dabholkar, Samuel Coogan, and Patricio A. Vela. Synthesis of Control Barrier Functions Using a Supervised Machine Learning Approach. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7139–7145, October 2020. doi: 10.1109/IROS45743.2020.9341190. URL https://ieeexplore.ieee.org/document/9341190/. ISSN: 2153-0866.

[32] Jiankai Sun, Lantao Yu, Pinqian Dong, Bo Lu, and Bolei Zhou. Adversarial Inverse Reinforcement Learning With Self-Attention Dynamics Model. *IEEE Robotics and Automation Letters*, 6(2):1880–1886, April 2021. ISSN 2377-3766. doi: 10.1109/LRA.2021.3061397. URL https://ieeexplore.ieee.org/document/9361118.

[33] Aaron D. Ames; Samuel Coogan; Magnus Egerstedt; Gennaro Notomista; Koushil Sreenath; Paulo Tabuada. Control barrier functions: Theory and applications. In *IEEE*, 2019.

[34] Faraz Torabi, Garrett Warnell, and Peter Stone. Behavioral cloning from observation, 2018. URL https://arxiv.org/abs/1805.01954.

[35] Faraz Torabi, Garrett Warnell, and Peter Stone. Generative Adversarial Imitation from Observation, June 2019. URL http://arxiv.org/abs/1807.06158. arXiv:1807.06158 [cs].

[36] Samuel Triest, Mateo Guaman Castro, Parv Maheshwari, Matthew Sivaprakasam, Wenshan Wang, and Sebastian Scherer. Learning risk-aware costmaps via inverse reinforcement learning for off-road navigation, 2023. URL https://arxiv.org/abs/2302.00134.

[37] Núñez D. E. Gaspar P. A. Brass M. Ulloa, J. L. Imitative inhibitory control is associated with psychotic experiences in a sample from the general population. *Frontiers in psychiatry, 15, 1470030*, 126, 2024.

[38] Lizhi Yang, Blake Werner, Massimiliano de Sa, and Aaron D. Ames. Cbf-rl: Safety filtering reinforcement learning in training with control barrier functions, 2026. URL https://arxiv.org/abs/2510.14959.

[39] Yue Yang, Letian Chen, and Matthew Gombolay. Safe Inverse Reinforcement Learning via Control Barrier Function, March 2023. URL http://arxiv.org/abs/2212.02753. arXiv:2212.02753 [cs].

[40] Yue Yang, Letian Chen, Zulfiqar Zaidi, Sanne van Waveren, Arjun Krishna, and Matthew Gombolay. Enhancing Safety in Learning from Demonstration Algorithms via Control Barrier Function Shielding. In *2024 19th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 820–829, March 2024. URL https://ieeexplore.ieee.org/document/10660943/. ISSN: 2167-2121.

[41] Songyuan Zhang, Oswin So, Kunal Garg, and Chuchu Fan. Gcbf+: A neural graph control barrier function framework for distributed safe multiagent control. *IEEE Transactions on Robotics*, 41:1533–1552, 2025. ISSN 1941-0468. doi: 10.1109/tro.2025.3530348. URL http://dx.doi.org/10.1109/TRO.2025.3530348.

APPENDIX

a) DBF as Cost-Regularized Observation-Only IRL.: We now show how the DBF objective in Equation (5) can be interpreted as a restricted form of the entropy-regularized observation-only adversarial imitation learning formulation in Equation (1).

To start, suppose that

$$\mathcal{C}_{\text{CBF}} := \{c : c(s,s') = q_h(s,s'),\ h \in \mathcal{H}\} \subset \mathbb{R}^{S\times S} \tag{8}$$

is a nonempty, closed, convex class of cost functions parametrized by h.

Now recall that Proposition 3.2 in [15] states that for any proper, closed, convex regularizer ψ,

$$\text{RL} \circ \text{IRL}_\psi(\pi_E) = \arg\min_\pi [-H(\pi) + \psi^*(\rho_\pi - \rho_E)] \tag{9}$$

where $\text{IRL}_\psi(\pi_E) = \arg\max_{c\in\mathbb{R}^{\mathbb{S}\times\mathbb{A}}} -\psi(c) + \min_{\pi\in\Pi} -\lambda\mathbb{H}(\pi) + \mathbb{E}_\pi[c(s,a)] - \mathbb{E}_{\pi_E}[c(s,a)]$ and ρ_π, ρ_E are the state-action occupancy measures of the policy and expert respectively. Following [35], we instead consider the state-transition occupancy measure

$$\rho_\pi(s,s') = \sum_a P(s' \mid s,a)\,\pi(a \mid s) \sum_{t=0}^{\infty} \gamma^t P(s_t = s \mid \pi). \tag{10}$$

While [35] derives the observation-only analogue of Equation 9 in the absence of entropy regularization, in practice we optimize the corresponding entropy-regularized objective

$$\text{RL} \circ \text{IRLfO}_\psi(\pi_E) = \arg\min_\pi [-H(\pi) + \psi^*(\rho_\pi - \rho_E)] \tag{11}$$

where IRLfO is defined in Equation 1 and ρ_π and ρ_E are now state-transition occupancy measures. This can be viewed as the natural entropy-regularized extension of the GAIfO objective.

Now consider the indicator cost function regularizer

$$\psi_{\text{CBF}}(c) = \begin{cases} 0, & \text{if} \quad c \in \mathcal{C}_{\text{CBF}} \\ \infty, & \text{otherwise} \end{cases} \tag{12}$$

Note that this regularizer is proper, closed, and convex by assumption on $\mathcal{C}_{\text{CBF}}$. Now recall that

$$\psi^*_{\text{CBF}}(u) = \sup_{c\in\mathbb{R}^{S\times S}} [\langle u, c\rangle - \psi_{\text{CBF}}(c)] \tag{13}$$

$$\tag{14}$$

But since $\psi_{\text{CBF}}(c) = 0$ for all $c \in \mathcal{C}_{\text{CBF}}$ and is ∞ everywhere else, we obtain that

$$\psi^*_{\text{CBF}}(u) = \sup_{c\in\mathcal{C}_{\text{CBF}}} \langle u, c\rangle \tag{15}$$

Interpreting the discriminator as a parametrization of the cost function, we let $D_h := q_h$. Substituting $u = \rho_\pi - \rho_E$ and reparameterizing by $h \in \mathcal{H}$ yields

$$\psi^*_{\text{CBF}}(\rho_\pi - \rho_E) = \sup_{h\in\mathcal{H}} [\mathbb{E}_{\rho_\pi}[D_h(s,s')] - \mathbb{E}_{\rho_E}[D_h(s,s')]] \tag{16}$$

From our earlier observation that ψ_{CBF} is proper, closed and convex we may apply Equation 11 and then obtain that

$$\text{RL} \circ \text{IRLfO}_\psi(\pi_E) = \arg\min_\pi \max_{h\in\mathcal{H}} [-H(\pi) + \mathbb{E}_{\rho_\pi}[D_h(s,s')] - \mathbb{E}_{\rho_E}[D_h(s,s')]] \tag{17}$$

Furthermore, if we restrict the class of cost functions we consider to be 1-Lipschitz $\mathcal{C}_{\text{CBF}} \subset \{c : \|c\|_L \le 1\}$ then the objective becomes an integral probability metric over the restricted CBF class. If additionally $\mathcal{C}_{\text{CBF}}$ equals the full 1-Lipschitz ball then by Kantorovich-Rubinstein duality we get that

$$\sup_{c\in\mathcal{C}_{\text{CBF}}} [\mathbb{E}_{\rho_\pi}[c(s,s')] - \mathbb{E}_{\rho_E}[c(s,s')]] = W_1(\rho_\pi, \rho_E)$$

where W_1 is the Wasserstein distance between the two distributions. Thus the discriminator objective corresponds to the Kantorovich dual formulation underlying WGANs.

Remark: In practice, we use the penalty term

$$\lambda_{\text{GP}} \mathbb{E}_{\hat{\pi}} \left[(\|\nabla_{(s,s')} D_h(s,s')\|_2 - 1)^2 \right] \tag{18}$$

to encourage approximate 1-Lipschitz continuity. Also, the exact parametrized neural network class used to represent $\mathcal{C}_{\text{CBF}}$ is not guaranteed to be convex or closed. However, the idealized function class can be approximated by:

$$\mathcal{H}_{\text{CBF},\epsilon} := \{h \in \mathbb{R}^S : h(s) \ge \epsilon\ \forall s \in \mathcal{S},\ h(s) \le -\epsilon\ \forall s \in \mathcal{U}\} \tag{19}$$

$$\alpha(r) = \kappa r + \beta \tag{20}$$

$$\mathcal{Q}_{\text{CBF}} := \{c : c(s,s') = q_h(s,s'),\ h \in \mathcal{H}_{\text{CBF},\epsilon}\} \cap \{c : \|c\|_L \le 1\} \tag{21}$$

Then $\mathcal{H}_{\text{CBF},\epsilon}$ is convex and closed as the intersection of half-spaces. It is also non-empty as long as $\mathcal{S}$ and $\mathcal{U}$ are disjoint and ϵ is small enough. Expanding the definition of q_h yields $q_h(s,s') = h(s') + (\kappa - 1)h(s) + \beta$ which is an affine map. We already know the Lipschitz ball is closed and convex and the parametrized cost function set is also closed and convex since $h \mapsto q_h$ is affine. Thus, $\mathcal{Q}_{\text{CBF}}$ is also closed and convex. Assuming there exists at least one feasible h such that the induced q_h is 1-Lipschitz, a mild assumption given the continuous setting of many robotics tasks, the class is nonempty so the assumptions on $\mathcal{C}_{\text{CBF}}$ are satisfied.

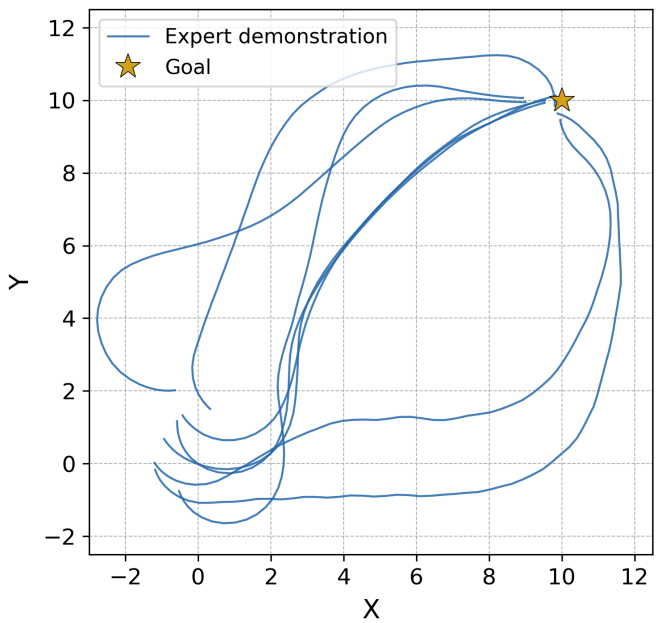

Fig. 8: Expert Trajectories used in benchmarking in maze environment.