arXiv:2403.02741v2 [cs.GT] 4 Jun 2024

State-Constrained Zero-Sum Differential Games with One-Sided Information

Mukesh Ghimire [1] **Lei Zhang** [1] **Zhe Xu** [1] **Yi Ren** [1]

Abstract

We study zero-sum differential games with state constraints and one-sided information, where the informed player (Player 1) has a categorical payoff type unknown to the uninformed player (Player 2). The goal of Player 1 is to minimize his payoff without violating the constraints, while that of Player 2 is to violate the state constraints if possible, or to maximize the payoff otherwise. One example of the game is a man-to-man matchup in football. Without state constraints, Cardaliaguet (2007) showed that the value of such a game exists and is convex to the common belief of players. Our theoretical contribution is an extension of this result to games with state constraints and the derivation of the primal and dual subdynamic principles necessary for computing behavioral strategies. Different from existing works that are concerned about the scalability of no-regret learning in games with discrete dynamics, our study reveals the underlying structure of strategies for belief manipulation resulting from information asymmetry and state constraints. This structure will be necessary for scalable learning on games with continuous actions and long time windows. We use a simplified football game to demonstrate the utility of this work, where we reveal player positions and belief states in which the attacker should (or should not) play specific random deceptive moves to take advantage of information asymmetry, and compute how the defender should respond.

1. Introduction

We study fixed-time zero-sum differential games with state constraints and one-sided information, where Player 1 holds a private type (e.g., an intent or preference) that defines the payoffs of the game. The goal of Player 1 (resp. Player 2) is to minimize (resp. maximize) the cost. Since violation of the state constraint results in $+\infty$ penalty to Player 1, Player 2 resorts to violating the constraints when possible; and Player 1 resigns when state violation is inevitable. At the beginning of the game, Nature draws a type from a distribution known to both players and assigns the type only to Player 1. Initialized as Nature's distribution, the common belief about Player 1's type is updated dynamically during the game based on observations, and shared between the players. A stochastic state trajectory is produced based on the initial state and belief, the deterministic system dynamics, and the behavioral strategies of the two players. The value of the game, when exists, follows a Hamilton-Jacobi (HJ) equation and is a function of time, state, and belief. Importantly, Player 1 may control the release of information at the equilibrium to manipulate the common belief and take advantage of information asymmetry.

We use Hexner's game (Hexner, 1979) as a minimal example to demonstrate information control by Player 1: Consider two players with linear dynamics

$$\dot{x}_i = A_i x_i + B_i u_i,$$

for $i = 1, 2$, where $x_i(t) \in \mathbb{R}^{d_x}$ are system states, $u_i(t) \in \mathcal{U}$ are control inputs belonging to the admissible set $\mathcal{U}$, $A_i, B_i \in \mathbb{R}^{d_x \times d_x}$. Let $\theta \in \{-1, 1\}$ be Player 1's type unknown to Player 2[1]. Let p_θ be Nature's probability distribution of θ. Consider that the game is to be played infinite many times, the payoff is an expectation over θ:

$$J(u_1, u_2) = \mathbb{E}_\theta \Big[\int_0^T \Big(\|u_1\|^2_{R_1} - \|u_2\|^2_{R_2} \Big) dt + \|x_1(T) - z\theta\|^2_{K_1(T)} - \|x_2(T) - z\theta\|^2_{K_2(T)} \Big], \tag{1}$$

where $z \in \mathbb{R}^{d_x}$, R_1 and R_2 are positive-definite, continuous matrix functions, and $K_1(T)$ and $K_2(T)$ are positive semi-definite matrices. All parameters are common knowledge except θ. Essentially Player 1's goal is to get closer to $z\theta$ than Player 2. Since Player 2 can infer the target based on Player 1's control, Player 1 may play a non-revealing strategy for some time, i.e., as if he also only knows p_θ rather than the actual θ, before he eventually reveals.

[1]Department of Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ, USA. Correspondence to: Yi Ren <yiren@asu.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

[1]Hexner's analysis is applicable to $\theta \in \mathbb{R}^{d_x}$, but is not generalizable to games with arbitrary dynamics and payoff functions. Here we adopt Cardaliaguet's setting where types are categorical (Cardaliaguet, 2007).

The equilibrium of this game is as follows: First, it can be shown that players' control has a 1D representation, denoted by $\tilde{\theta}_i \in \mathbb{R}$, through:

$$u_i = -R_i^{-1} B_i^T K_i x_i + R_i^{-1} B_i^T K_i \Phi_i z \tilde{\theta}_i,$$

for $i = 1, 2$, where $\dot{\Phi}_i = A_i \Phi_i$ with boundary condition $\Phi_i(T) = I$, and

$$\dot{K}_i = -A_i^T K_i - K_i A_i + K_i^T B_i R_i^{-1} B_i^T K_i.$$

Then by introducing

$$d_i = z^T \Phi_i^T K_i B_i R_i^{-1} B_i^T K_i^T \Phi_i z, \tag{2}$$

and defining the critical time as

$$t_r = \arg\min_t \int_0^t (d_1(s) - d_2(s)) ds,$$

one can derive Player 1's strategy as $\tilde{\theta}_1(t) = 0$ for $t \in [0, t_r]$ and $\tilde{\theta}_1(t) = \theta$ for $t \in (t_r, T]$, i.e., Player 1 reveals its type at t_r. Player 2's strategy turns out to be to strictly follow Player 1: $\tilde{\theta}_2(t) = \tilde{\theta}_1(t)$. The original analysis by Hexner exploits the fact that both players solve linear-quadratic regulators parameterized by θ. We will revisit this game after introducing the differential game theory for one-sided information games (Cardaliaguet, 2007; 2009), which arrives at Hexner's solution but can also solve games with arbitrary dynamics and payoff functions, subject to continuity assumptions. This paper extends the unconstrained settings in Cardaliaguet (2009) and Souquiere (2010): We prove that value exists for differential games *with state constraints* and one-sided information, and derive the primal-dual HJ equations necessary for computing player strategies.

Different from existing works that focus on scalable no-regret learning on imperfect-information games with discrete dynamics (Brown et al., 2020; Perolat et al., 2022), this paper builds on top of repeated games and incomplete-information differential games (Cardaliaguet, 2007; 2009) to reveal the underlying mechanism of belief manipulation resulted from information asymmetry and state constraints. Specifically, we show that in any subgame, Player 1 plays a behavioral strategy (i.e., probability distributions over the action space for all his types) that convexifies his value with respect to the common belief. As a result, the common belief "splits" to vertices of the value convex hull with probabilities that are optimal for Player 1. See Fig. 1 for an illustration using Hexner's game. Importantly, the number of splits for Player 1 is no more than the number of possible player types. On the other hand, Player 2 counters Player 1 by playing a dual game where her behavioral strategy is determined by the convexification of the conjugate value. See Sec. 3 and 4 for details.

Within this context, it becomes clear that understanding whether and how belief should be manipulated relies on knowing the value landscape over the belief space at any time and state. In addition to the curse of dimensionality (CoD) commonly present for games with non-trivial state/action/belief spaces and time horizons, we also experience computational challenges due to value discontinuity and the need for convexification and splitting. We discuss in Sec. 5 a set of solutions, including using physics-informed neural network to characterize the backward reachable set to smooth value approximation, and using an input convex architecture (Amos et al., 2017) for predicting convex values.

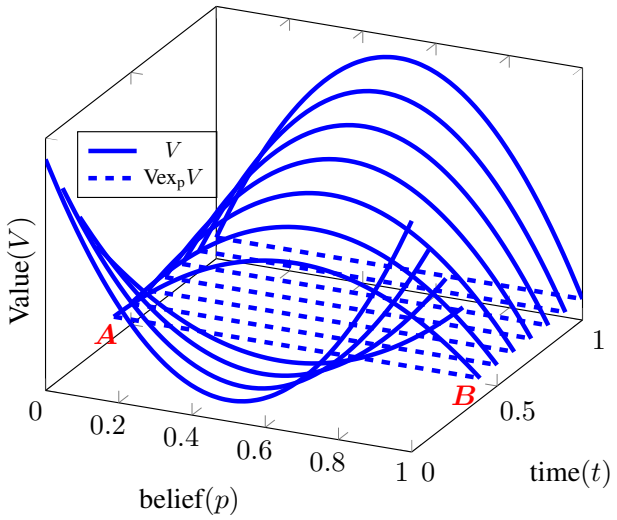

Figure 1: Value along belief (p) and time (t) in Hexner's game. Belief splits to A ($p = 0$) and B ($p = 1$) depending on the true type of Player 1, when the value becomes concave should Player 1 play a non-revealing strategy. In other words, Player 1 delays the release of his type until a critical time. In more general cases, belief splitting may not fully reveal Player 1's type, leading to belief manipulation.

To summarize, we claim the following contributions:

- We extend the theory of zero-sum differential games with one-sided information to games with state constraints by proving value existence of such games and deriving the primal and dual subdynamic principles;
- We elucidate, with detailed examples, how the subdynamic principles lead to the derivation of behavioral strategies;
- We develop numerical tools to alleviate CoD in value approximation and to infer behavioral strategies from values. In Sec. 6, we solve an 8D man-to-man matchup game and reveal player positions in which the attacker can take advantage of information asymmetry by playing specific deceptive moves, and to derive the defender's best response in the lack of information. See Fig. 2.

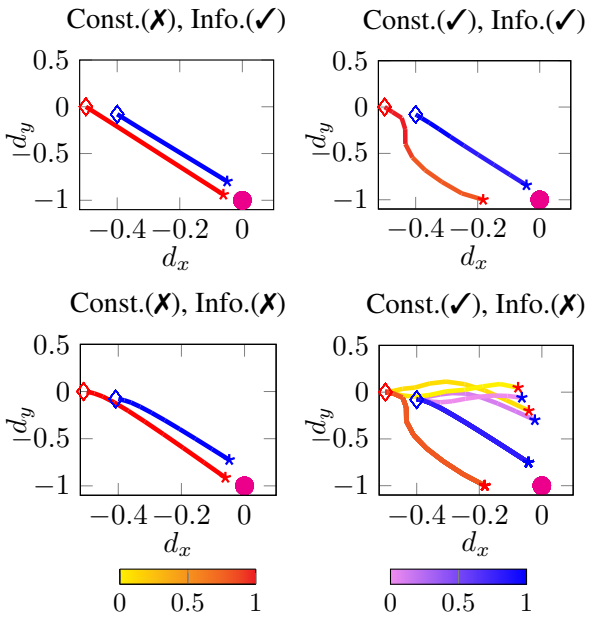

Figure 2: Trajectories of informed Player 1 (red) and uninformed Player 2 (blue) in an 8D Hexner's game w/ and w/o a state constraint or information asymmetry. Color shades indicate probabilities. When constrained, Player 1 stays away from Player 2 while trying to be closer to the target (the circle) than Player 2. Diamonds indicate initial states and stars indicate final states. See Sec. 6 for details.

2. Related Work

Incomplete-information repeated and differential games Harsanyi (1967) first formalized information asymmetry in a stage game by introducing a private player types. Aumann et al. (1995) provided a framework to study repeated games with incomplete information on one side. De Meyer (1996) introduced dual games from where strategies of the uninformed player can be derived from a recursive structure of the conjugate value. Extending these results to differential games with Markov rewards, Cardaliaguet (2007) and Souquiere (2010) confirmed the structures of incomplete-information games with one sided information on player type: (1) the game enjoys a primal-dual decomposition so that the informed player does not need to know the equilibrium of the uninformed player to compute his own; and (2) the value is convexified by belief splitting at the equilibrium. Recently, Hu et al. (2023) proposed independently a belief-space HJ formulation for zero-sum differential games with one-sided information. While their framework can incorporate state constraints, it does not reveal the above structure of the equilibrium strategies of such games.

Imperfect-information dynamic games Since player types can be considered as static private states, our work belongs to the category of imperfect-information dynamic games, where more general dynamics and information structures (e.g., disturbances, partial observability, and delayed information sharing) are considered. Nayyar et al. (2013) showed that the game can be reformulated as perfect-information by introducing a common belief state, provided that the belief is strategy independent. This strategy-independence assumption is relaxed in Kartik and Nayyar (2021) for zero-sum dynamic games by introducing past strategies as part of the players' information state. The general setting of Kartik and Nayyar (2021), however, does not facilitate a value existence proof. A significant amount of recent work build on top of common belief to approximate values of imperfect-information dynamic games (e.g., ReBeL (Brown et al., 2020), DeepNash (Perolat et al., 2022), and SoG (Schmid et al., 2023)). Following Nayyar et al. (2013), these algorithms model behavioral strategies as *prescriptions*, i.e., belief-conditioned action distributions. In addition, by taking advantage of the equivalence between local regret matching and Nash equilibrium in two-player zero-sum games (Zinkevich et al., 2007), no-regret algorithms (Brown & Sandholm, 2018; 2019; Brown et al., 2020) have been developed for more scalably solving games with large action spaces and long time horizons than linear programming based methods (Koller & Pfeffer, 1995). It should be noted that these algorithms scale linearly to the square-root of the action space, and thus induce high costs as the action space grows. While often disconnected, the studies on imperfect-information dynamic (or extensive-form) games and those on incomplete-information differential games are consistent in theory. Specifically, regret matching, i.e., solving subgame minimax problems with respect to behavioral strategies in the former, leads to strategies that satisfy the subdynamic programming principles stated in Cardaliaguet (2007), due to the fact that the behavioral strategies intrinsically convexify values. The key difference, however, is that regret matching algorithms do not enforce belief splitting. In practice, this means that the resultant strategy, often as a result of manually chosen action discretization, does not explicitly *explain* whether a certain random action is to be taken in a given belief state in order to delay information release or to manipulate the belief in a specific way.

3. State Constrained Zero-Sum Differential Games with One-Sided Information

Preliminaries We consider a time-invariant deterministic dynamical system that defines the dynamics of the combined state x of Players 1 and 2, whose control inputs are u and v, respectively:

$$\begin{cases} \dot{x}(t) = f(x(t), u(t), v(t)), \quad u(t) \in \mathcal{U},\ v(t) \in \mathcal{V} \\ x(t_0) = x_0 \end{cases} \tag{3}$$

The game starts at $t_0 \in [0, T]$ with an initial state $x_0 \in \mathbb{R}^{d_x}$. Denote $g_i : \mathbb{R}^{d_x} \to \mathbb{R}$ the terminal payoff functions for $i \in [I]$, each corresponding to a Player 1 type drawn from

Nature's distribution $p = \{p_1, ..., p_I\} \in \Delta(I)$, where $\Delta(I)$ is an I-dimensional simplex; denote $\mathcal{C} = \{x \in \mathbb{R}^{d_x} | c(x) \leq 0\}$ the set of feasible states. The goal of Player 1 is to minimize the expected payoff while keeping the state in $\mathcal{C}$. Player 1 receives $+\infty$ if state violation occurs; the goal of Player 2 is to maximize the expected payoff and hence may resort to violating the state constraint. We omit instantaneous payoffs (e.g., effort losses due to control) for conciseness, and discuss in Sec. 4 modifications to the Bellman backup when common-knowledge instantaneous payoffs exist.

The following assumptions will be used:

1. $\mathcal{U}$ and $\mathcal{V}$ are compact and finite-dimensional sets;
2. $f : \mathbb{R}^{d_x} \times \mathcal{U} \times \mathcal{V} \to \mathbb{R}^{d_x}$ is bounded, continuous, and uniformly Lipschitz continuous with respect to x;
3. $g_i : \mathbb{R}^{d_x} \to \mathbb{R}$ for $i = 1, ..., I$ and $c : \mathbb{R}^{d_x} \to \mathbb{R}$ are Lipschitz continuous and bounded.
4. Isaacs' condition holds for the Hamiltonian $H : \mathbb{R}^{d_x} \times \mathbb{R}^{d_x} \to \mathbb{R}$:
$$\begin{aligned} H(x, \xi) &:= \min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} f(x, u, v)^T \xi \\ &= \max_{v \in \mathcal{V}} \min_{u \in \mathcal{U}} f(x, u, v)^T \xi. \end{aligned} \tag{4}$$
5. Control inputs and states of both players are fully observable by all. The dynamics, the payoff set, and the equilibrium strategies are common knowledge to all.

Behavioral strategy Let $\mathcal{A}(t)$ (resp. $\mathcal{D}(t)$) be the set of open-loop controls for Player 1 (resp. Player 2):

$$\begin{aligned} \mathcal{A}(t) &:= \{\alpha : [t, T] \to \mathcal{U} \mid \text{Lebesgue measurable}\}, \\ \mathcal{D}(t) &:= \{\delta : [t, T] \to \mathcal{V} \mid \text{Lebesgue measurable}\}. \end{aligned}$$

Following (Cardaliaguet, 2007), we introduce $\mathcal{H}(t)$ (resp. $\mathcal{Z}(t)$) as the set of non-anticipative pure strategies with delay for Player 1 (resp. Player 2) (Elliott & Kalton, 1972):

$$\begin{aligned} \mathcal{H}(t) := \{\eta : \mathcal{D}(t) \to \mathcal{A}(t) \mid \exists \tau > 0 \text{ such that} \\ \forall s \in (t, T - \tau) \text{ and } \delta, \bar{\delta} \in \mathcal{D}(t), \text{ if } \delta = \bar{\delta} \text{ a.e.} \\ \text{in } [t, s], \text{then } \eta(\delta) = \eta(\bar{\delta}) \text{ a.e. in } [t, s + \tau]\}. \\ \mathcal{Z}(t) := \{\zeta : \mathcal{A}(t) \to \mathcal{D}(t) \mid \exists \tau > 0 \text{ such that} \\ \forall s \in (t, T - \tau) \text{ and } \alpha, \bar{\alpha} \in \mathcal{A}(t), \text{ if } \alpha = \bar{\alpha} \text{ a.e.} \\ \text{in } [t, s], \text{then } \zeta(\alpha) = \zeta(\bar{\alpha}) \text{ a.e. in } [t, s + \tau]\} \end{aligned}$$

A behavioral (mixed) strategy for Player 1 is defined by a pair $((\Omega_\eta, \mathcal{F}_\eta, \mathbf{P}_\eta), \eta)$, where $(\Omega_\eta, \mathcal{F}_\eta, \mathbf{P}_\eta)$ is a probability space and the strategy $\eta : \Omega_\eta \times \mathcal{D}(t) \to \mathcal{A}(t)$ is measurable and non-anticipative with delay, i.e., there is some $\tau > 0$ such that, for any $s \in (t, T - \tau)$ and $\delta, \bar{\delta} \in \mathcal{D}(t)$, if $\delta = \bar{\delta}$ a.e. in $[t, s]$ then $\eta(\omega, \delta) = \eta(\omega, \bar{\delta})$ a.e. in $[t, s + \tau]$ for any $\omega \in \Omega_\eta$. We denote the sets of behavioral strategies of Player 1 by $(\mathcal{H}_r(t))^I$ and the behavioral strategy of Player 2 by $\mathcal{Z}_r(t)$. With mild notational abuse, we will denote by $(\eta_i) \in (\mathcal{H}_r(t))^I$ a particular set of behavioral strategies of Player 1, and by $\zeta \in \mathcal{Z}_r(t)$ a particular behavioral strategy of Player 2. Lastly, we assume that η_i for $i = 1, ..., I$ are defined on the same probability space.

Remarks. Nonanticipative strategies *with delay* are used, as opposed to ones without delay that are often used in complete-information games (Elliott & Kalton, 1972), in order to enable Lemma 1 that associates random strategies with open-loop controls. This association will become useful in proving the existence of value of incomplete-information differential games and in value characterization (see discussions in (Cardaliaguet, 2007)):

Lemma 1. *(Lemma 2.2 of (Cardaliaguet, 2007)) For any pair $(\eta, \zeta) \in \mathcal{H}_r(t) \times \mathcal{Z}_r(t)$ and any $\omega := (\omega_1, \omega_2) \in \Omega_\eta \times \Omega_\zeta$, there is a unique pair $(\alpha_\omega, \delta_\omega) \in \mathcal{A}(t) \times \mathcal{D}(t)$ such that*

$$\eta(\omega_1, \delta_\omega) = \alpha_\omega \text{ and } \zeta(\omega_2, \alpha_\omega) = \delta_\omega. \tag{5}$$

Furthermore the map $\omega \to (\alpha_\omega, \delta_\omega)$ is measurable from $\Omega_\eta \times \Omega_\zeta$ endowed with $\mathcal{F}_\eta \otimes \mathcal{F}_\zeta$ into $\mathcal{A}(t) \times \mathcal{D}(t)$ endowed with the Borel σ-field associated with the L^1 distance.

Backward reachable set Let $X_\tau^{t_0, x_0, \alpha, \delta}$ be the solution of Eq. (3) at $t = \tau$ when starting at (t_0, x_0) and following (α, δ). With behavioral strategies (η, ζ) and initials (t_0, x_0), we denote by $\mathcal{X}_\tau^{t_0, x_0, \alpha, \delta}$ the trajectory of states reachable by (α, δ) within $[t_0, \tau]$, and $\mathcal{X}_\tau^{t_0, x_0, \eta, \zeta}$ as states reachable by (η, ζ) within $[t_0, \tau]$:

$$\mathcal{X}_\tau^{t_0, x_0, \eta, \zeta} := \bigcup_{\omega \in \Omega_\eta \times \Omega_\zeta} \mathcal{X}_\tau^{t_0, x_0, \alpha_\omega, \delta_\omega}$$

where $(\alpha_\omega, \delta_\omega)$ is defined by Eq. (5). Introduce $\rho(\mathcal{S}) = 1$ if $\mathcal{S} \subseteq \mathcal{C}$, and otherwise $\rho(\mathcal{S}) = +\infty$; and the backward reachable (infeasible) set as

$$\begin{aligned} \bar{\mathcal{Q}}(t) := \{x \in \mathbb{R}^{d_x} \mid \forall \eta \in \mathcal{H}_r(t), \exists \zeta \in \mathcal{Z}_r(t), \tau \in (t, T], \\ s.t., \rho\left(\mathcal{X}_\tau^{t, x, \eta, \zeta}\right) = +\infty\}. \end{aligned}$$

$\mathcal{Q}(t) := \mathbb{R}^{d_x} \setminus \bar{\mathcal{Q}}(t)$ is the set of *feasible states*. $\mathcal{Q}(T) = \mathcal{C}$. Lastly, we use $\bar{\rho}(t, x) = 1$ if $x \in \mathcal{Q}(t)$ and otherwise $\bar{\rho}(t, x) = +\infty$.

Payoff and value We define the expected payoff of player type i for taking behavioral strategies (η, ζ) as

$$\begin{aligned} G_i(t_0, x_0, \eta, \zeta) &:= \mathbb{E}_{\eta, \zeta}\left[g_i(X_T^{t_0, x_0, \eta, \zeta}) \rho(\mathcal{X}_\tau^{t_0, x_0, \eta, \zeta})\right] \\ &= \int_{\Omega_\eta \times \Omega_\zeta} g_i\left(X_T^{t_0, x_0, \alpha_\omega, \delta_\omega}\right) \\ &\qquad \rho(\mathcal{X}_\tau^{t_0, x_0, \alpha_\omega, \delta_\omega}) d\mathbf{P}_\eta \otimes \mathbf{P}_\zeta(\omega). \end{aligned}$$

The payoff of Player 1 is $\sum_{i=1}^I p_i G_i(t_0, x_0, \eta_i, \zeta)$. With strategys $(\eta_i) \in (\mathcal{H}_r(t_0))^I$ and $\zeta \in \mathcal{Z}_r(t_0)$, the upper value function is defined by

$$V^+(t_0, x_0, p) := \inf_{(\eta_i)} \sup_{\zeta} \sum_{i=1}^I p_i G_i(t_0, x_0, \eta_i, \zeta),$$

and the lower value function is given by

$$V^-(t_0, x_0, p) := \sup_{\zeta} \inf_{(\eta_i)} \sum_{i=1}^{I} p_i G_i(t_0, x_0, \eta_i, \zeta).$$

The existence of the value While the existence of value is proven for both zero-sum complete-information state-constrained differential games (Lee, 2022) and zero-sum differential games with one-sided information (Cardaliaguet, 2009), the proof for games with both one-sided information and state constraints is missing. Our main theoretical result fills in this gap (see Appendix A for the proof):

Theorem 1. *If assumptions 1-5 hold, we have* $V^+(t,x,p) = V^-(t,x,p)$ *for all* $(t,x,p) \in [0,T] \times \mathbb{R}^{d_x} \times \Delta(I)$.

Characterization of the value We need to first characterize the value of the unconstrained game since this value will later appear in that of the state-constrained game.

Let $U : [0,T] \times \mathbb{R}^{d_x} \times \Delta(I) \to \mathbb{R}$ be the value of the unconstrained version of the game, and $U^* : [0,T] \times \mathbb{R}^{d_x} \times \mathbb{R}^I \to \mathbb{R}$ its convex conjugate:

$$U^*(t,x,\hat{p}) := \sup_{p \in \Delta(I)} \hat{p}^T p - U(t,x,p) \quad \forall (t,x,\hat{p}) \in [0,T] \times \mathbb{R}^{d_x} \times \mathbb{R}^I.$$

We have the following properties for U and U^*:

1. U is Lipschitz continuous in (t,x,p) and convex to p. $U(T,x,p) = \sum_{i=1}^{I} p_i g_i(x)$, $\forall (x,p) \in \mathbb{R}^{d_x} \times \Delta(I)$. U^* is Lipschitz continuous in $(t,x,\hat{p})$ and convex to $\hat{p}$. $U^*(T,x,\hat{p}) = \max_{i \in [I]} \hat{p}_i - g_i(x)$, $\forall (x,\hat{p}) \in \mathbb{R}^{d_x} \times \mathbb{R}^I$.
2. For any $p \in \Delta(I)$, $(t,x) \to U(t,x,p)$ is a viscosity subsolution to the primal HJ equation
$$w_t + H(x, Dw) = 0,$$
where H is defined by Eq. (4).
3. For any $\hat{p} \in \mathbb{R}^I$, $(t,x) \to U^*(t,x,p)$ is a viscosity subsolution to the dual HJ equation
$$w_t + H^*(x, Dw) = 0,$$
where $H^*(x,\xi) = -H(x,-\xi)\ \forall\ (x,\xi) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_x}$.

The conjugate U^* defines the value of a *dual game* where Player 2 minimizes her payment, $U^*(T,x,\hat{p})$, to Player 1 where $\hat{p}$ is common knowledge. We note that by definition (see (De Meyer, 1996)), the dual variables $\hat{p}$ are the info-state values defined in Brown et al. (2020), i.e., $\hat{p}[i]$ captures the value of Player 1 if he is of type i and plays the best response to Player 2's equilibrium strategy. De Meyer (1996) showed that when $\hat{p} \in \partial_p U(0,x,p)$, Player 2's strategy in the dual game is her equilibrium in the primal game. We note that Player 2's strategy cannot be derived from the primal subdynamic principle because her best response is dependent on Player 1's type which is unknown to her. The introduction of the dual game allows us to derive a subdynamic principle of the conjugate value from where her equilibrium strategy can be computed.

For the state-constrained game, the following corollary is a result of the subdynamic principles derived from Theorem 1, and will guide the numerical approximation of values for the state-constrained game (Sec. 4):

Corollary 1.1. *Under assumptions 1-5, the value function* $V := V^+ = V^-$ *(resp.* V^**) is a unique function defined on* $[0,T] \times \mathbb{R}^{d_x} \times \Delta(I)$ *(resp.* $[0,T] \times \mathbb{R}^{d_x} \times \mathbb{R}^I$*) such that:*

1. V *is convex respect to* p *and*
$$V(T,x,p) = \rho(x) U(T,x,p)\ \forall (x,p) \in \mathbb{R}^{d_x} \times \Delta(I); \tag{6}$$
V^* *is convex respect to* $\hat{p}$ *and*
$$V^*(T,x,\hat{p}) = \max_{i \in [I]} \hat{p}_i - \rho(x) g_i(x)\ \forall (x,\hat{p}) \in \mathbb{R}^{d_x} \times \mathbb{R}^I \tag{7}$$
2. *For all* $p \in \Delta(I)$, $(t,x) \to V(t,x,p)$ *is a viscosity subsolution to the primal HJ equation*
$$\min\{\rho(x) U(t,x,p) - w, w_t + H(x, Dw)\} = 0. \tag{8}$$
3. *For all* $\hat{p} \in \mathbb{R}^I$, $(t,x,z) \to V^*(t,x,z,p)$ *is a viscosity subsolution to the dual HJ equation*
$$\min\{\rho(x) U^*(t,x,\hat{p}/\rho(x)), w_t + H^*(x, Dw)\} = 0. \tag{9}$$

4. Bellman Backup and Behavioral Strategies

Discrete-time Bellman backup computes an approximated value $V_\tau(t_k, \cdot, \cdot) : \mathbb{R}^{d_x} \times \Delta(I) \to \mathbb{R}$, with time step $\tau = T/L$ for some large L and $t_k = k\tau$ for $k = 0, ..., L$:

(i) At the terminal time, set $V_\tau(T,x,p) = \rho(x) \sum_i p_i g_i(x)$.

(ii) At $k \in \{0, ..., L-1\}$
$$V_\tau(t_k,x,p) = \rho(x) \text{Vex}_p \left(\min_u \max_v V_\tau(t_{k+1}, x', p) \right), \tag{10}$$
where $x' = x + \tau f(x,u,v)$ and $\text{Vex}_p(\cdot)$ is the convex hull with respect to p.

Let $l : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ be a Lipschitz continuous and bounded function that represents the instantaneous payoff of the game. To incorporate l, Eq. (10) becomes

$$V_\tau(t_k,x,p) = \rho(x) \text{Vex}_p \left(\min_u \max_v V_\tau(t_{k+1}, x', p) + \tau l(u,v) \right) \tag{11}$$

Theorem 2 states that V_τ uniformly converges to V as $\tau \to 0^+$ (see proof in Appendix B):

Theorem 2. *If assumptions 1-5 hold, V_τ converges uniformly to V on compact subsets of $[0,T]\times\mathbb{R}^{d_x}\times\Delta(I)$:*

$$\lim_{\substack{\tau\to 0^+,\ t_k\to t,\\ x'\to x, p'\to p}} V_\tau(t_k,x',p') = V(t,x,p)$$
$$\forall (t,x,p)\in[0,T]\times\mathbb{R}^{d_x}\times\Delta(I).$$

Behavioral strategy for Player 1 and the belief dynamics Player 1's behavioral strategy is a probability distribution over $\mathcal{U}$ conditioned on (t,x,p). At time t_k, Player 1 resigns if $x_k\in\bar{\mathcal{Q}}(t_k)$; otherwise, he determines his strategy using the following steps: First he finds $\lambda=\{\lambda_1,...,\lambda_I\}\in\Delta(I)$ and $p^j\in\Delta(I)$ for $j=1,...,I$, such that

$$V_\tau(t_k,x_k,p_k)=\sum_{j=1}^{I}\lambda_j\Big(\min_{u\in\mathcal{U}}\max_{v\in\mathcal{V}} V_\tau(t_{k+1},x_k',p^j)\Big),$$
$$\sum_{j=1}^{I}\lambda_j p^j = p_k. \tag{12}$$

Then he computes u^j as the minimax solution corresponding to p^j, and chooses $u_k=u^j$ with probability $\lambda_j p^j[i]/p_k[i]$, where i is its true type. It is proved that this behavioral strategy of Player 1 is ϵ-optimal for small enough τ (Cardaliaguet, 2009). Importantly, $\{p^j\}_{j=1}^I$ are vertices of the value convex hull. Thus by announcing his strategy, and by assuming that players use the same Bayes belief update, Player 1 controls the belief dynamics to follow a martingale that optimizes his gain, i.e., $p_{k+1}=p^j$ if u^j is chosen. Note that the introduction of state constraints changes the minimax solutions, the value convex hulls, and thus the behavioral strategies. Lastly, Eq. (12) will be modified according to Eq. (11) when instantaneous loss is present.

The dual game and behavioral strategy for Player 2 Player 2's strategy is determined by a dual game for which the conjugate value is approximated by $V_\tau^*(t_k,\cdot,\cdot):\mathbb{R}^{d_x}\times\mathbb{R}^I\to\mathbb{R}$. Specifically,

(i) At the terminal time, set $V_\tau^*(T,x,\hat{p})=\max_i\{\hat{p}_i-\rho(x)g_i(x)\}$.

(ii) At $k\in\{0,...,L-1\}$

$$V_\tau^*(t_k,x,\hat{p})=\mathrm{Vex}_{\hat{p}}\left(\min_v\max_u V_\tau^*(t_{k+1},x',\hat{p})\right), \tag{13}$$

if $\bar{\rho}(t_k,x)=1$; otherwise $V_\tau^*(t_k,x,\hat{p})=-\infty$.

Similar to Theorem 2, Theorem 3 proves that V_τ^* uniformly converges to V^* as $\tau\to 0^+$ (proof omitted).

Theorem 3. *If assumptions 1-5 holds, V_τ^* converges uniformly to V^* on compact subsets of $[0,T]\times\mathbb{R}^{d_x}\times\mathbb{R}^I$:*

$$\lim_{\substack{\tau\to 0^+,\ t_k\to t,\\ x'\to x, \hat{p}'\to\hat{p}}} V_\tau^*(t_k,x',\hat{p}') = V^*(t,x,\hat{p})$$
$$\forall (t,x,\hat{p})\in[0,T]\times\mathbb{R}^{d_x}\times\mathbb{R}^I.$$

With instantaneous loss l, the Bellman backup in Eq. (13) becomes

$$V_\tau^*(t_k,x,\hat{p})=\mathrm{Vex}_{\hat{p}}\left(\min_v\max_u V_\tau^*(t_{k+1},x',\hat{p}-\tau l(u,v))\right) \tag{14}$$

We explain this modification in detail in Appendix C. An intuitive explanation is as follows: Recall that each element of $\hat{p}$ represents Player 1's value for the corresponding type in the primal game. Hence $\hat{p}$ at the next time step should discount the common instantaneous loss incurred at the current time step.

The behavioral strategy of Player 2 defines a probability distribution over $\mathcal{V}$ conditioned on $(t,x(t),\hat{p}(t))$, with the dual variable $\hat{p}(t_0)\in\partial_p V(t_0,x_0,p(t_0))$. At time t_k, if $x_k\in\bar{\mathcal{Q}}(t_k)$, Player 2 plays according to a pursuit-evasion game since she can always catch Player 1 according to the definition of $\bar{\mathcal{Q}}(t_k)$; otherwise, Player 2 determines her strategy using the following steps: First she finds $\lambda=\{\lambda_1,...,\lambda_{I+1}\}\in\Delta(I+1)$ and $\hat{p}^j\in\mathbb{R}^I$ for $j=1,...,I+1$, such that

$$V_\tau^*(t_k,x_k,\hat{p})=\sum_{j=1}^{I+1}\lambda_j\left(\min_{v\in\mathcal{V}}\max_{u\in\mathcal{U}} V_\tau^*(t_{k+1},x_k',\hat{p}^j)\right),$$
$$\sum_{j=1}^{I+1}\lambda_j\hat{p}^j=\hat{p}_k. \tag{15}$$

Then she computes the minimax solution v^j, and chooses $v_k=v^j$ with probability λ_j. It is proved that this behavioral strategy of Player 2 is ϵ-optimal for small enough τ (Cardaliaguet, 2009). $\hat{p}$ follows a martingale $\hat{p}_{k+1}=\hat{p}^j$ if v^j is chosen by Player 2, or $\hat{p}_{k+1}=\hat{p}^j-\tau l(u^j,v^j)$ if l is present, where u^j is the best response to v^j in the dual game. Notice that due to her lack of information, Player 2 solves harder value approximation and control synthesis problems of belief dimension $I+1$ rather than I.

To help readers better understand the mechanisms described in this section, we provide detailed derivation of behavioral strategies for two sample problems in Appendix D (D.1 for a zero-sum version of the beer-quiche game and D.2 for Hexner's game).

5. Numerical Methods

5.1. Primal and dual value approximation

We use backward induction to solve Eq. (10) and (15), and discuss treatments that alleviate error propagation. We focus the discussion on the primal problem for brevity.

Value discontinuity At each time step t, $V_\tau(t,\cdot,\cdot)$ (resp. $V_\tau^*(t,\cdot,\cdot)$) can be approximated separately in $\bar{\mathcal{Q}}(t)$ and $\mathcal{Q}(t)$: the primal (resp. dual) value in the former is set to $+\infty$ (resp. $-\infty$) and value in the latter will be approximated using a neural network $\hat{V}_\tau(t,\cdot,\cdot)$ (resp. $\hat{V}_\tau^*(t,\cdot,\cdot)$). This avoids fitting the value networks to functions with

large Lipschitz constants during numerical implementation. $\bar{\mathcal{Q}}(t)$ for $t \in [0,T]$ can be approximated by a physics-informed neural network (PINN) solver (Bansal & Tomlin, 2021) (see details in Appendix E), by recognizing that $\bar{\mathcal{Q}}(t)$ can be defined by pure strategies instead of behavioral ones using Lemma 1. PINN alleviates CoD in solving HJ equations with Lipschitz continuous solutions (Shin et al., 2020), and here it results in a separate value network $\hat{V}(\cdot,\cdot) : [0,T] \times \mathbb{R}^{d_x}$ that approximates $\bar{\mathcal{Q}}(t)$ as $\{x \in \mathbb{R}^{d_x} | \hat{V}(t,x) \leq 0\}$.

Partially convex values At each t_k and for uniformly sampled $\mathcal{S}(t) \subset \mathcal{Q}(t)$, we scan a lattice $\mathcal{P} \subset \Delta(I)$ to obtain the minimax solution of the RHS of Eq. (10) (denoted by $\vartheta^0(t,x,p)$ for $(x,p) \in \mathcal{S}(t) \times \mathcal{P}$), resulting in a dataset $\{(p, \vartheta^0(t,x,p))\}_{p\in\mathcal{P}}$. Value convexification is then obtained by applying the Monotone Chain Convex Hull algorithm to this dataset for each $x \in \mathcal{S}(t)$ and taking the lower hull of the resulting convex hull. Let $\vartheta(t,x,p)_{\mathcal{S}(t)}$ be the resultant value on the convex hull. A value network $\hat{V}_\tau(t,\cdot,\cdot)$ is then trained using data $\{(x, \vartheta_{\mathcal{S}(t)}(t,x,p) | (x,p) \in \mathcal{S}(t) \times \mathcal{P}\}$ so that during the Bellman backup at $t-1$, we can predict convexified values at previously unvisited states at t. We use a Partially Input Convex Neural Network (PICNN) (Amos et al., 2017) to ensure that $\hat{V}_\tau(t,\cdot,\cdot)$ is convex in p. Alg. 1 summarizes the value approximation algorithm. Optionally, we also train a separate value network to predict the minimax values using $\{(p, \vartheta^0(t,x,p))\}_{p\in\mathcal{P}}$. This network helps remove the nested minimax problem during control synthesis.

Convexification error. Backward induction suffers from error propagation, where errors at each time step are originated from (i) value approximation through neural networks, (ii) backward reachable set approximation, (iii) convex hull approximation, and (iv) finite time discretization (and Euler method for ODE). Here we discuss control of the error resulted from convex hull approximation, which is unique to incomplete-information games. We leave a full analysis of data complexity for error control to future studies. At each $t \in [0,T]$ and $x \in \mathcal{Q}(t)$, let $\vartheta(t,x,\cdot)$ be the RHS of Eq. (10) after convexification, and the convexification error be $\varepsilon_{vex}(t,x) := \max_{p\in\Delta(I)} \|\vartheta(t,x,p) - \vartheta_{\mathcal{S}(t)}(t,x,p)\|_\infty$. Proposition 1 shows that $\varepsilon_{vex}(t,x)$ can be controlled by refining $\mathcal{P}$ (see proof in Appendix F):

Proposition 1. *For given (t,x), let the Lipschitz constant of $\vartheta(t,x,\cdot)$ be L, and $d_{\mathcal{P}}$ be the minimum distance between two neighboring nodes of the lattice $\mathcal{P}$. $\varepsilon_{vex}(t,x) \leq 2d_{\mathcal{P}}L$.*

Approximation of the conjugate value. Recall that the dual game is initialized by the dual variables $\hat{p} \in \partial_p V(0,x,p)$ when the primal game starts at (x,p). Since $\hat{V}(0,x,\cdot)$ is a differentiable neural network defined on a simplex, subgradients can be found using $\hat{p}^T p = V(0,x,p)$ and $\hat{p}^T q \leq V(0,x,q)$ for all $q \in \Delta(I)$ and $q \neq p$. Specific to the case study where $I=2$ and $\hat{V} := \hat{V}_\tau(0,x,p[1])$ is modeled to be a function of the first element of p to reduce dimensionality, we have $\hat{p} = (\hat{V} + \nabla_{p[1]}\hat{V}(1-p[1]), \hat{V} - \nabla_{p[1]}\hat{V}p[1])^T$.

5.2. Synthesis of strategies

Given $(t,x,p) \in [0,T]\times\mathbb{R}^{d_x}\times\Delta(I)$, Player 1 computes his behavioral strategy by finding $\lambda \in \Delta(I)$ and splitting beliefs $\{p^j \in \Delta(I)\}_{j=1}^I$ that best satisfy Eq. (12) in L^2, if $x \in \mathcal{Q}(t)$ (otherwise he surrenders). Given $(t,x,\hat{p}) \in [0,T] \times \mathbb{R}^{d_x} \times \mathbb{R}^I$, Player 2 finds $\lambda \in \Delta(I+1)$ and $\{\hat{p}^j \in \mathbb{R}^I\}_{j=1}^{I+1}$ that best satisfy Eq. (15) in L^2. When $I=2$ as in the case study, the splitting beliefs and resultant strategies for Player 1 can be approximated through sweeping $p[1] \in [0,1]$. For Player 2, we use gradient descent to solve a 6D optimization problem with the initial guess $\lambda = [1/3, 1/3, 1/3]^T$ and $\hat{p}^j = \hat{p}$ for $j = [3]$.

Algorithm 1 Value Approximation

Inputs: current time step t, time discretization τ, sample size N, admissible action spaces $(\mathcal{U}(t), \mathcal{V}(t))$, value approximation at $t+\tau$: $\hat{V}_{\text{next}}(\cdot,\cdot) := \hat{V}_\tau(t+\tau,\cdot,\cdot)$, feasible state set $\mathcal{Q}(t)$, instantaneous loss $l(\cdot,\cdot)$, terminal loss in Eq. (6)
Initialize: $\hat{V}_\tau(t,\cdot,\cdot)$, $\vartheta^0 = \emptyset$

```
S(t) ← sample N states from Q(t)
for x in S do
    for p in P do
        v(x,p) ← min_{u∈U(t)} max_{v∈V} V̂_next(x',p) + τ l(U,V) {if t+τ = T, V̂_next is given by Eq. (6)}
        append v(x,p) to ϑ^0
    end for
    ϑ_{S(t)}(t,x,·) ← compute Vex_p(ϑ^0(x,·)) via Eq. (10)
end for
Update V̂_τ to match ϑ_{S(t)}
```

6. Case Study

Setup We study a state-constrained version of Hexner's game that represents a simplified football game: Player 1 (P1)'s goal is to move closer to one of the two targets than P2 without being caught during the interaction (see Fig. 3); P2's goal is to catch P1 if possible, or otherwise move close to P1's target. Each player has 4 state variables: x- and y- position and velocity; and their actions encode x- and y-acceleration. The parameters $R_A = diag([0.05, 0.025])$ and $R_D = diag([0.05, 0.1])$ are chosen so that P1 can afford to accelerate faster in the y-direction than P2. The state constraint is $c(x_1,x_2) = r - \|(d_{x_1}, d_{y_1}) - (d_{x_2}, d_{y_2})\|_2$, where $r = 0.05$. We note that due to the introduction of an (instantaneous) effort loss, the backward induction is modified as: $V_\tau(t_k,x,p) = \text{Vex}_p(\min_u \max_v V_\tau(t_{k+1}, x + \tau(f,x,u,v),p) + \tau l(u,v))$, where $l(u,v)$ is the integral term in $[t_k, t_{k+1}]$ in Eq. (1).

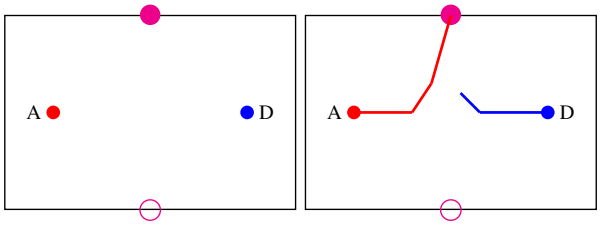

Figure 3: Schematics of a simplified football game with Player 1 (red) and Player 2 (blue). Left: the initial configuration. Right: equilibrium trajectory. Magenta circles: two goals. The filled is the current type private to Player 1. Players move in a 2D space bounded by $[-1, 1] \times [-1, 1]$.

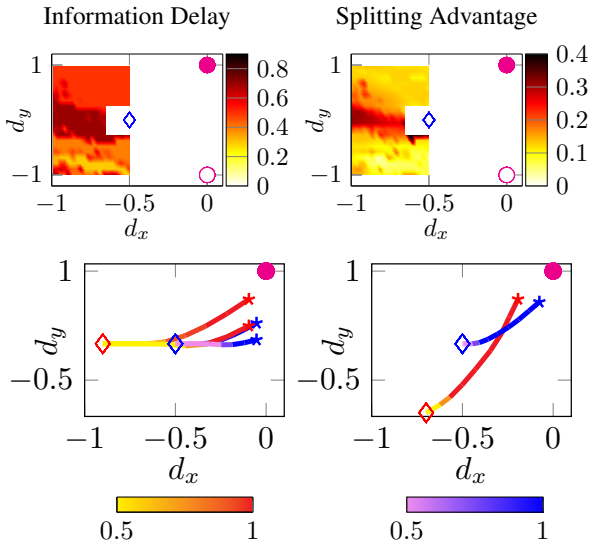

Figure 4: **Top:** Average delay ($\mathcal{T}$) in information reveal (left) and average maximum advantage of playing the revealing strategy (right), keeping P2's location fixed at (-0.5, 0) and changing P1's location. **Bottom:** Trajectory with high delay and advantage (left) and with low delay and advantage (right). Color shades indicate current belief.

Value network architecture and training The value network uses PICNN with 5 hidden layers and 256 neurons each and has 9-dimensional inputs (state and belief). We train 10 separate networks for each time step starting from $t = 0.9$ with $\tau = 0.1$, each being trained for 10 epochs. For each epoch, $\mathcal{S}(t)$ includes 5000 states sampled from $\mathcal{Q}(t)$. Since $I = 2$, value networks can be considered as functions of $p[1]$ and thus we set $\mathcal{P} = \{p[1] = 0, 0.01, \ldots, 0.99, 1\}$. $\hat{V}_\tau(t, \cdot, \cdot)$ is trained on data collected from $\mathcal{S}(t) \times \mathcal{P}$. For the conjugate $\hat{V}_\tau^*$, we set $\hat{\mathcal{P}} = \{\hat{p}[1] = \{-14, \ldots, 14\}, \hat{p}[2] = \{-14, \ldots, 14]\}$. More details can be found in Appendix G.

Constrained vs. unconstrained strategies Fig. 2 compares strategies with and without the state constraint, visualizing the equilibrium strategies of P1 and the best responses of P2 given P1's strategies. Note that the best responses of P2 give P2 an advantage since she does not know actions to be taken by P1 in reality. The analytical solution to the unconstrained game is given by Hexner's analysis, where P1's strategy is to start moving to the goal after the critical time $t_r = 0.4$. This strategy no longer holds in the constrained case as it violates the state constraint. Instead, P1 actively tries to stay clear of P2 while pursuing the goal (see 2nd col. of Fig. 2). Note that in this case, P1 resorts to a random strategy with the presence of incomplete information and state constraints, as the two contribute to value nonconvexity with respect to the belief.

Information delay and advantage of random strategies To understand how P1 uses information asymmetry, we examine the delay in information reveal, measured by the time at which the belief converges to the true type, i.e. $\mathcal{T} = \inf\{k \in [L] : p_k = \{0, 1\}\}$. We then take average of $\mathcal{T}$ for each initial state over 10 simulations. In Fig. 4, we visualize $\mathcal{T}$ over the space of P1's starting positions, while fixing P2's starting position and setting both players' initial speed to 0. The trajectories represent cases where P1 delays (bottom left) and does not delay (bottom right) the release of information. We also compute the advantage of following a belief manipulating strategy (that convexifies the value) as opposed to taking the non-revealing strategy (i.e. never split), expressed as $[\min_u \max_v V(t+\tau, x', p) - V(t, x, p)]$. Overall, P1 tends to conceal and deceive when it has equal distances to the possible targets.

Equilibrium strategy of Player 2 Fig. 5 shows sample trajectories when both players play their equilibrium strategies. Note that compared with P2's best responses to P1 in previous examples, P2's equilibrium strategy is more conservative, due to her lack of knowledge about P1's type. We also note that P2's dual game is one dimension higher than P1's primal game, and thus encounters higher numerical errors in value and strategy approximation (see Appendix G).

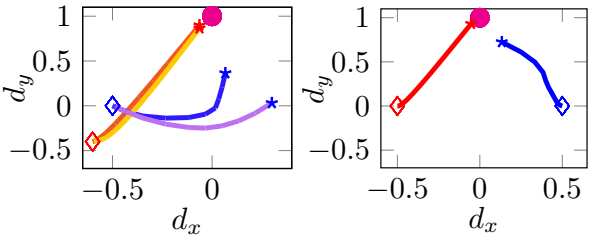

Figure 5: Trajectories where both players use their respective behavioral strategies. P1 keeps track of p, whereas P2 keeps track of $\hat{p}$.

7. Conclusion and Future Work

We proved the existence of value for zero-sum differential games with state constraints and one-sided information and developed a backward induction scheme to approximate the value. Our method enables mechanistic synthesis of behavioral strategies and allows explanation of the resultant splitting of strategies and beliefs. Future work will investigate more efficient learning+search methods that take advantage of value convexity and alleviate error propagation.

Acknowledgements

This work was in part supported by NSF CMMI-1925403 and NSF CNS-2101052. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation or the U.S. Government.

Impact Statement

This paper is concerned with advancing the field of differential game theory and artificial intelligence. The developed theories and tools help understand how players play strategies by taking into account their information advantage and disadvantage. Similar to the development of imperfect-information games, protocols to mitigate the potential societal consequences and risks (e.g., deception by robots and machines) shall be comprehensively investigated.

References

Amos, B., Xu, L., and Kolter, J. Z. Input convex neural networks. In *International Conference on Machine Learning*, pp. 146–155. PMLR, 2017.

Aumann, R. J., Maschler, M., and Stearns, R. E. *Repeated games with incomplete information*. MIT press, 1995.

Bansal, S. and Tomlin, C. J. Deepreach: A deep learning approach to high-dimensional reachability. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1817–1824. IEEE, 2021.

Brown, N. and Sandholm, T. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

Brown, N. and Sandholm, T. Superhuman ai for multiplayer poker. *Science*, 365(6456):885–890, 2019.

Brown, N., Bakhtin, A., Lerer, A., and Gong, Q. Combining deep reinforcement learning and search for imperfect-information games. *Advances in Neural Information Processing Systems*, 33:17057–17069, 2020.

Cardaliaguet, P. Differential games with asymmetric information. *SIAM journal on Control and Optimization*, 46 (3):816–838, 2007.

Cardaliaguet, P. Numerical approximation and optimal strategies for differential games with lack of information on one side. *Advances in Dynamic Games and Their Applications: Analytical and Numerical Developments*, pp. 1–18, 2009.

De Meyer, B. Repeated games, duality and the central limit theorem. *Mathematics of Operations Research*, 21(1): 237–251, 1996.

Elliott, R. J. and Kalton, N. J. *The existence of value in differential games*, volume 126. American Mathematical Soc., 1972.

Harsanyi, J. C. Games with incomplete information played by "bayesian" players, i–iii part i. the basic model. *Management science*, 14(3):159–182, 1967.

Hexner, G. A differential game of incomplete information. *Journal of Optimization Theory and Applications*, 28: 213–232, 1979.

Hu, H., Zhang, Z., Nakamura, K., Bajcsy, A., and Fisac, J. F. Learning-aware safety for interactive autonomy. *arXiv preprint arXiv:2309.01267*, 2023.

Kartik, D. and Nayyar, A. Upper and lower values in zero-sum stochastic games with asymmetric information. *Dynamic Games and Applications*, 11:363–388, 2021.

Koller, D. and Pfeffer, A. Generating and solving imperfect information games. In *IJCAI*, pp. 1185–1193. Citeseer, 1995.

Lee, D. *Safety-Guaranteed Autonomy under Uncertainty*. University of California, Berkeley, 2022.

Nayyar, A., Gupta, A., Langbort, C., and Başar, T. Common information based markov perfect equilibria for stochastic games with asymmetric information: Finite games. *IEEE Transactions on Automatic Control*, 59(3):555–570, 2013.

Perolat, J., De Vylder, B., Hennes, D., Tarassov, E., Strub, F., de Boer, V., Muller, P., Connor, J. T., Burch, N., Anthony, T., et al. Mastering the game of stratego with model-free multiagent reinforcement learning. *Science*, 378(6623): 990–996, 2022.

Schmid, M., Moravčík, M., Burch, N., Kadlec, R., Davidson, J., Waugh, K., Bard, N., Timbers, F., Lanctot, M., Holland, G. Z., et al. Student of games: A unified learning algorithm for both perfect and imperfect information games. *Science Advances*, 9(46):eadg3256, 2023.

Shin, Y., Darbon, J., and Karniadakis, G. E. On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type pdes. *arXiv preprint arXiv:2004.01806*, 2020.

Souquière, A. Approximation and representation of the value for some differential games with asymmetric information. *International Journal of Game Theory*, 39: 699–722, 2010.

Zinkevich, M., Johanson, M., Bowling, M., and Piccione, C. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20, 2007.

A. Proof of Theorem 1

The following proofs extend results and techniques from Cardaliaguet (2007) to zero-sum differential games with one-sided information and state constraints. To overview, we start by showing that the upper and lower values $V^{\pm}$ are Lipschitz continuous within the safe and unsafe state sets (Lemma 2) and convex with respect to p (Lemma 3). We then show that: (1) V^{-*} satisfies a subdynamic principle and is therefore a subsolution of a dual HJ equation and hence V^{-} is a dual supersolution of the corresponding primal HJ (Lemma 4, Lemma 4, Lemma 5); and (2) V^{+} also satisfies a subdynamic principle (Lemma 6) and is therefore a dual subsolution of the primal HJ (Lemma 7). We can then use a comparison principle (see (Cardaliaguet, 2007)) to show that since V^{-} is a dual supersolution and V^{+} is a dual subsolution of the primal HJ while both share the same terminal value, $V^{-} \geq V^{+}$. On the other hand, $V^{-} \leq V^{+}$ by definition and hence $V^{-} = V^{+}$.

We start with the following regularity result (see proof of Lemma 3.1 in (Cardaliaguet, 2007)):

Lemma 2. *(regularity of $V^{\pm}$). $V^{\pm}(t_0, x_0, p)$ are Lipschitz continuous for all $x_0 \in \mathcal{Q}(t_0)$. $V^{\pm}(t_0, x_0, p) = +\infty$ for all $x_0 \in \bar{\mathcal{Q}}(t_0)$.*

The following convexity result was originally developed for repeated games with incomplete information (De Meyer, 1996) and was later extended to differential games (Cardaliaguet, 2007). The same convexity result holds for imperfect-information dynamic games (Brown et al., 2020).

Lemma 3. *(convexity property of $V^{\pm}$). For any $(t, x) \in [0, T] \times \mathbb{R}^{d_x}$, $V^{\pm}$ are convex in p on $\Delta(I)$.*

Proof. Let $p^{\lambda} = (1-\lambda)p^0 + \lambda p^1$ for some $p^0, p^1 \in \Delta(I)$. Let $((\eta_i^0), \zeta^0)$ and $((\eta_i^1), \zeta^1)$ be the equilibrial strategies for $V(t, x, p^0)$ and $V(t, x, p^1)$, respectively. Introduce a set of "splitting" behavioral strategies (η_i^{λ}) for (t, x, p^{λ}) such that for any type i, $\eta_i^{\lambda} = \eta_i^0$ with probability $(1-\lambda)p_i^0/p_i^{\lambda}$ and $\eta_i^{\lambda} = \eta_i^1$ with probability $\lambda p_i^1/p_i^{\lambda}$. Then we have

$$\begin{aligned}
&\sup_{\zeta} \sum_i p_i^{\lambda} G_i(t, x, \eta_i^{\lambda}, \zeta) \\
=&\sup_{\zeta} \sum_i \left(p_i^{\lambda} \frac{(1-\lambda)p_i^0}{p_i^{\lambda}} G_i(t, x, \eta_i^0, \zeta) + p_i^{\lambda} \frac{\lambda p_i^1}{p_i^{\lambda}} G_i(t, x, \eta_i^1, \zeta) \right) \\
\leq&(1-\lambda) \sup_{\zeta} \sum_i p_i^0 G_i(t, x, \eta_i^0, \zeta) + \lambda \sup_{\zeta} \sum_i p_i^1 G_i(t, x, \eta_i^1, \zeta).
\end{aligned} \tag{16}$$

Since the inequality in Eq. (16) holds for any "splitting" (η_i^{λ}), we have

$$V^{\pm}(t, x, p^{\lambda}) \leq (1-\lambda) V^{\pm}(t, x, p^0) + \lambda V^{\pm}(t, x, p^1) \tag{17}$$

for any $t \in [0, T]$ and $x \in \mathcal{Q}(t)$. For $x \in \bar{\mathcal{Q}}(t)$, the equality holds since $V^{\pm}(t, x, \cdot) = +\infty$. □

Remarks. (1) The proof says that by playing a "splitting" strategy, the value at (t, x, p^{λ}) should at least be as good as a linear interpolation between those at (t, x, p^0) and at (t, x, p^1). Hence the value is a convex hull in $\Delta(I)$ at any $(t, x) \in [0, T] \times \mathbb{R}^{d_x}$. (2) Assuming that the "splitting" strategy of Player 1 is known by Player 2, then the latter can perform Bayesian inference on Player 1's type based on his actions. For any type i, let u_i^0 and u_i^1 be two distinct actions to be taken at (t, x) following η_i^0 and η_i^1, respectively, and let p^{λ} be the current common belief. Then under observation of u_i^0 (resp. u_i^1), the common belief becomes p^0 (resp. p^1) with probability $(1-\lambda)$ (resp. λ). Since $p^{\lambda} = (1-\lambda)p^0 + \lambda p^1$, common belief is a martingale.

Next, we introduce a reformulation of V^{-*} to facilitate the derivation of its subdynamic principle. The proof of this reformulation is extended from Lemma 4.1 of Cardaliaguet (2007) to incorporate the state constraints.

Lemma 4. *(reformulation of V^{-*}). We have*

$$V^{-*}(t_0, x_0, \hat{p}) = \inf_{\zeta \in \mathcal{Z}_r(t_0)} \sup_{\eta \in \mathcal{H}_r(t_0)} \max_i \{\hat{p}_i - G_i(t_0, x_0, \eta, \zeta)\} \tag{18}$$

Proof. For later use, we first note that

$$\begin{aligned}
V^{-}(t_0, x_0, p) &= \sup_{\zeta} \inf_{(\eta_i)} \sum_i p_i G_i(t_0, x_0, \eta, \zeta) \\
&= \sup_{\zeta} \sum_i p_i \inf_{\eta} G_i(t_0, x_0, \eta, \zeta).
\end{aligned} \tag{19}$$

Let the right-hand side of Eq. (18) be $z = z(t_0, x_0, \hat{p})$. We can show that z is convex with respect to $\hat{p}$ using a technique similar to that of Lemma 3.

Then by the definition of z:

$$\begin{aligned} z^*(t_0,x_0,p) &= \sup_{\hat{p}} p^T\hat{p} - \inf_{\zeta}\sup_{\eta}\max_i \{\hat{p}_i - G_i(t_0,x_0,\eta,\zeta)\} \\ &= \sup_{\hat{p}} p^T\hat{p} - \inf_{\zeta}\max_i \left\{\hat{p}_i - \inf_{\eta} G_i(t_0,x_0,\eta,\zeta)\right\} \\ &= \sup_{\zeta}\sup_{\hat{p}}\min_i \left\{p^T\hat{p} - \hat{p}_i + \inf_{\eta} G_i(t_0,x_0,\eta,\zeta)\right\}. \end{aligned} \tag{20}$$

In this last expression, $\sup_{\hat{p}}$ is attained by setting $\hat{p}_i = \inf_\eta G_i(t_0,x_0,\eta,\zeta)$, in which case we have

$$z^*(t_0,x_0,p) = \sup_{\zeta}\sum_i p_i \inf_{\eta} G_i(t_0,x_0,\eta,\zeta) = V^-(t_0,x_0,p,z). \tag{21}$$

Since z is convex with respect to $\hat{p}$, we have $V^{-*} = z^{**} = z$. □

Next, to introduce the subdynamic principle of V^{-*}, we first introduce

$$U^{-*}(t_0,x_0,\hat{p}) := \inf_{\zeta\in\mathcal{Z}_r(t_0)}\sup_{\eta\in\mathcal{H}_r(t_0)}\max_i \left\{\hat{p}_i - \mathbb{E}_{\eta,\zeta}\left[g_i(X_T^{t_0,x_0,\eta,\zeta})\right]\right\} \tag{22}$$

as the conjugate lower value of the *unconstrained* version of the game, and $U^{\pm}$ as the corresponding upper and lower values. From Lemma 2 and Lemma 4, U^{-*} is Lipschitz continuous and convex in $\hat{p}$.

Lemma 5. *(subdynamic principle for V^{-*}). For any $(t_0,x_0,\hat{p}) \in [0,T)\times\mathbb{R}^{d_x}\times\mathbb{R}^I$ and any $t_1 \in (t_0,T]$, denote $x_1 = X_{t_1}^{t_0,x_0,\eta,\zeta}$ and $\mathcal{X}_1 = \mathcal{X}_{t_1}^{t_0,x_0,\eta,\zeta}$. We have*

$$V^{-*}(t_0,x_0,\hat{p}) \le \inf_{\zeta\in\mathcal{Z}(t_0)}\sup_{\eta\in\mathcal{H}(t_0)}\min\left\{\rho(\mathcal{X}_1)U^{-*}\left(t_1,x_1,\frac{\hat{p}}{\rho(\mathcal{X}_1)}\right), V^{-*}(t_1,x_1,\hat{p})\right\} \tag{23}$$

Proof. Denote $V_1^{-*}(t_0,t_1,x_0,\hat{p}) := \inf_{\zeta\in\mathcal{Z}(t_0)}\sup_{\eta\in\mathcal{H}(t_0)} V^{-*}\left(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta}, \hat{p}\right)$. U_1^{-*} is similarly defined. We need the following preparations for the proof.

Player 1 plays a pure strategy in V^{-*}. We show below that best responses are always pure. In particular, Player 1 can play in pure strategies in V^{-*}, namely,

$$V^{-*}(t_0,x_0,\hat{p}) = \inf_{\zeta\in\mathcal{Z}_r(t_0)}\sup_{\eta\in\mathcal{H}(t_0)}\max_i\{\hat{p}_i - G_i(t_0,x_0,\eta,\zeta)\} \tag{24}$$

for any $(t_0,x_0,\hat{p})$. First from Theorem 4 and using $\mathcal{H}(t)\subset\mathcal{H}_r(t)$, we have

$$\begin{aligned} V^{-*}(t_0,x_0,\hat{p}) &= \inf_{\zeta\in\mathcal{Z}_r(t_0)}\sup_{\eta\in\mathcal{H}_r(t_0)}\max_i\{\hat{p}_i - G_i(t_0,x_0,\eta,\zeta)\} \\ &\ge \inf_{\zeta\in\mathcal{Z}_r(t_0)}\sup_{\eta\in\mathcal{H}(t_0)}\max_i\left\{\hat{p}_i - \mathbb{E}_\zeta\left[g_i(X_T^{t_0,x_0,\eta,\zeta})\rho(\mathcal{X}_T^{t_0,x_0,\eta,\zeta})\right]\right\}. \end{aligned} \tag{25}$$

For the reverse inequality, we first note that for any $\eta\in\mathcal{H}_r(t_0)$ and $\omega_1\in\Omega_\eta$, $\eta(\omega_1,\cdot)\in\mathcal{H}(t_0)$. With a fixed $\zeta\in\mathcal{Z}_r(t_0)$, and by using the convexity of $\max_i$ (i.e., $\max_i \mathbb{E}_\omega[f_i(\omega)] \le \mathbb{E}_\omega[\max_i f_i(\omega)]$), we have

$$\begin{aligned} &\sup_{\eta\in\mathcal{H}_r(t_0)}\max_i\{\hat{p}_i - G_i(t_0,x_0,\eta,\zeta)\} \\ &\le \sup_{\eta\in\mathcal{H}_r(t_0)}\int_{\Omega_\eta}\max_i\left\{\hat{p}_i - \mathbb{E}_\zeta\left[g_i(X_T^{t_0,x_0,\eta(\omega_1,\cdot),\zeta})\rho(\mathcal{X}_T^{t_0,x_0,\eta(\omega_1,\cdot),\zeta})\right]\right\} d\mathbf{P}_\eta(\omega_1) \\ &\le \sup_{\eta\in\mathcal{H}_r(t_0)}\sup_{\omega_1\in\Omega_\eta}\max_i\left\{\hat{p}_i - \mathbb{E}_\zeta\left[g_i(X_T^{t_0,x_0,\eta(\omega_1,\cdot),\zeta})\rho(\mathcal{X}_T^{t_0,x_0,\eta(\omega_1,\cdot),\zeta})\right]\right\} \\ &\le \sup_{\eta\in\mathcal{H}(t_0)}\max_i\left\{\hat{p}_i - \mathbb{E}_\zeta\left[g_i(X_T^{t_0,x_0,\eta(\omega_1,\cdot),\zeta})\rho(\mathcal{X}_T^{t_0,x_0,\eta(\omega_1,\cdot),\zeta})\right]\right\}. \end{aligned} \tag{26}$$

Since Eq. (26) holds for any ζ, together with Eq. (25), we have

$$V^{-*}(t,x,\hat{p}) = \inf_{\zeta\in\mathcal{Z}_r(t_0)} \sup_{\eta\in\mathcal{H}(t_0)} \max_i \left\{ \hat{p}_i - \mathbb{E}_\zeta \left[g_i(X_T^{t_0,x_0,\eta,\zeta}) \rho(\mathcal{X}_T^{t_0,x_0,\eta,\zeta}) \right] \right\}. \tag{27}$$

Note that one can reach the same conclusion for U^{-*}.

ϵ-optimal strategy of Player 2. Let $\epsilon > 0$ and $\zeta^0 \in \mathcal{Z}(t_0)$ be some pure ϵ-optimal strategy for $V_1^{-*}(t_0,t_1,x,\hat{p})$. For any $x_1 \in \mathbb{R}^{d_x}$, we can find some ϵ-optimal strategy $\zeta^{x_1} \in \mathcal{Z}_r(t_1)$ for Player 2 in the game $V^{-*}(t_1,x_1,\hat{p})$. Let $B_\rho(x)$ be a ball around x with radius ρ, and let $\partial\mathcal{Q}(t)$ be the boundary of $\mathcal{Q}(t)$, i.e., for any $x \in \partial\mathcal{Q}(t)$ and $\rho > 0$, there exist $y \in B_\rho(x) \bigcap \mathcal{Q}(t)$ and $y' \in B_\rho(x) \bigcap \bar{\mathcal{Q}}(t)$.

For $x_1, y \in \mathcal{Q}(t_1) \setminus \partial\mathcal{Q}(t_1)$, from Lipschitz continuity of the map $y \to V^{-*}(t_1,y,\hat{p})$, ζ^{x_1} is also (2ϵ)-optimal for $V^{-*}(t_1,y,\hat{p})$ if $y \in B_\rho(x_1)$ for some radius $\rho > 0$. The same applies to $x_1, y \in \bar{\mathcal{Q}}(t_1)$ and $y \in B_\rho(x_1)$ since $V^{-*}(t_1,x,\hat{p}) = +\infty$ is constant for $x \in \bar{\mathcal{Q}}(t_1)$.

Since the dynamics f is bounded, we also know that the reachable states from (t_0,x_0) is bounded in some ball $B_R(0)$. Let us set $M = \|f\|_\infty$ and some small $\sigma > 0$ such that $M\sigma \le \rho/2$. Then we choose $(x_l)_{l=1,\dots,l_0}$ such that $\bigcup_{l=1}^{l_0} B_{\rho/2}(x_l)$ contains $B_R(0)$. Let $(E_l)_{l=1,\dots,l_0}$ be a Borel partition of $B_R(0)$ such that, for any l, $E_l \subset B_{\rho/2}(x_l)$. We also require (x_l) to be chosen properly so that $E_l \subset \mathcal{Q}(t_1)$ or $E_l \subset \bar{\mathcal{Q}}(t_1)$.

We set

$$\zeta^l = \zeta^{x_l},\ \Omega^l = \Omega_{\zeta^{x_l}},\ \mathcal{F}^l = \mathcal{F}_{\zeta^{x_l}},\ \text{and } \mathbf{P}^l = \mathbf{P}_{\zeta^{x_l}} \tag{28}$$

for $l = 1,\dots,l_0$. We choose some delay $\tau \in (0,\sigma]$ for all the strategies ζ^l. Note that if for some open-loop control $(\alpha,\delta) \in \mathcal{A}(t_0) \times \mathcal{D}(t_0)$ and for some l, we have $X_{t_1-\tau}^{t_0,x_0,\alpha,\delta} \in E_l$, then

$$|X_{t_1-\tau}^{t_0,x_0,\alpha,\delta} - X_{t_1}^{t_0,x_0,\alpha,\delta}| \le \|f\|_\infty \tau \le M\sigma \le \rho/2, \tag{29}$$

so that $X_{t_1}^{t_0,x_0,\alpha,\delta}$ belongs to $B_\rho(x_l)$. Hence ζ^l is (2ϵ)-optimal for V^{-*} at $(t_1, X_{t_1}^{t_0,x_0,\alpha,\delta}, \hat{p})$.

Let us now define a new strategy $\zeta \in \mathcal{Z}_r(t_0)$ by setting

$$\Omega_\zeta = \prod_{l=1}^{l_0} \Omega^l,\ \mathcal{F}_\zeta = \mathcal{F}^1 \otimes \dots \otimes \mathcal{F}^{l_0},\ \text{and } \mathbf{P}_\zeta = \mathbf{P}^1 \otimes \dots \otimes \mathbf{P}^{l_0}. \tag{30}$$

For any $\omega = (\omega^1,\dots,\omega^{l_0}) \in \Omega_\zeta$ and $\alpha \in \mathcal{A}(t_0)$, set

$$\zeta(\omega,\alpha) = \begin{cases} \zeta^0(\alpha)(\tau) & \text{if } \tau \in [t_0,t_1) \\ \zeta^l(\omega^l,\alpha)(\tau) & \text{if } \tau \in [t_1,T] \text{ and } X_{t_1-\tau}^{t_0,x_0,\alpha,\zeta^0} \in E_l. \end{cases} \tag{31}$$

For any pure strategy $\eta \in \mathcal{H}(t_0)$, we have

$$g_i(X_T^{t_0,x_0,\eta,\zeta}) = \sum_{l=1}^{l_0} g_i\left(X_T^{t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0}, \eta, \zeta^l} \right) \mathbf{1}_{O^l}, \tag{32}$$

where $O^l := \left\{ X_{t_1-\tau}^{t_0,x_0,\eta,\zeta^0} \in E_l \right\}$.

Property of $\rho(\cdot)$. Let $\rho_0 := \rho\left(\mathcal{X}_{t_1}^{t_0,x_0,\eta,\zeta^0}\right)$ and $\rho_1^l := \rho\left(\mathcal{X}_T^{t_1,x_1^l,\eta,\zeta^l}\right)$, where $x_1^l := X_{t_1}^{t_0,x_0,\eta,\zeta^0}$ if the state falls in E_l following pure strategies (η,ζ^0). Then we have

$$\begin{aligned} g_i\left(X_T^{t_0,x_0,\eta,\zeta}\right) \rho\left(\mathcal{X}_T^{t_0,x_0,\eta,\zeta}\right) &= g_i(X_T^{t_0,x_0,\eta,\zeta}) \sum_l \max\{\rho_0,\rho_1^l\} \mathbf{1}_{O^l} \\ &= \sum_l g_i(X_T^{t_1,x_1^l,\eta,\zeta}) \mathbf{1}_{O^l} \sum_l \max\{\rho_0,\rho_1^l\} \mathbf{1}_{O^l} \\ &= \sum_l g_i(X_T^{t_1,x_1^l,\eta,\zeta}) \max\{\rho_0,\rho_1^l\} \mathbf{1}_{O^l}. \end{aligned} \tag{33}$$

For any set $(\omega^l) \in (\Omega^l)$, let $a^l := g_i(X_T^{t_1,x_1^l,\eta,\zeta(\omega^l,\cdot)}) \geq 0$. Also note that ρ_0 and ρ_1^l only take values from $\{1,+\infty\}$. One can show that the following always holds:

$$\sum_l a^l \max\{\rho_0,\rho_1^l\}\mathbf{1}_{O^l} = \max\left\{\sum_l a^l\rho_0\mathbf{1}_{O^l}, \sum_l a^l\rho_1^l\mathbf{1}_{O^l}\right\}. \tag{34}$$

Similarly we have

$$\begin{aligned}
&\int_{\Omega^l}\sum_l\left[g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta,\zeta^l(\omega^l,\cdot)}\right)\max\{\rho_0,\rho_1^l\}\mathbf{1}_{O^l}\right]d\mathbf{P}^l(\omega^l)\\
=&\max\left\{\begin{array}{l}\int_{\Omega^l}\sum_l\left[g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta,\zeta^l(\omega^l,\cdot)}\right)\rho_0\mathbf{1}_{O^l}\right]d\mathbf{P}^l(\omega^l),\\ \int_{\Omega^l}\sum_l\left[g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta,\zeta^l(\omega^l,\cdot)}\right)\rho_1^l\mathbf{1}_{O^l}\right]d\mathbf{P}^l(\omega^l)\end{array}\right\}
\end{aligned} \tag{35}$$

Now we derive an upper bound of $\max_i\left\{\hat{p}_i - \mathbb{E}_\zeta\left[g_i(X_T^{t_0,x_0,\eta,\zeta})\rho(\mathcal{X}_T^{t_0,x_0,\eta,\zeta})\right]\right\}$:

$$\begin{aligned}
&\max_i\left\{\hat{p}_i - \mathbb{E}_\zeta\left[g_i(X_T^{t_0,x_0,\eta,\zeta})\rho(\mathcal{X}_T^{t_0,x_0,\eta,\zeta})\right]\right\},\\
=&\max_i\left\{\hat{p}_i - \int_{\Omega^l}\sum_l\left[g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta,\zeta^l(\omega^l,\cdot)}\right)\max\{\rho_0,\rho_1^l\}\mathbf{1}_{O^l}\right]d\mathbf{P}^l(\omega^l)\right\},\\
\leq&\sup_{\eta'\in\mathcal{H}(t_1)}\max_i\left\{\hat{p}_i - \sum_l\left[\int_{\Omega^l}g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\max\{\rho_0,\rho_1^l\}d\mathbf{P}^l(\omega^l)\mathbf{1}_{O^l}\right]\right\},\\
=&\sup_{\eta'\in\mathcal{H}(t_1)}\max_i\left\{\min\left\{\begin{array}{l}\hat{p}_i - \sum_l\left[\int_{\Omega^l}g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\rho_0 d\mathbf{P}^l(\omega^l)\mathbf{1}_{O^l}\right],\\ \hat{p}_i - \sum_l\left[\int_{\Omega^l}g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\rho_1^l d\mathbf{P}^l(\omega^l)\mathbf{1}_{O^l}\right]\end{array}\right\}\right\},\\
\leq&\min\left\{\begin{array}{l}\sup_{\eta'\in\mathcal{H}(t_1)}\max_i\left\{\hat{p}_i - \sum_l\left[\int_{\Omega^l}g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\rho_0 d\mathbf{P}^l(\omega^l)\mathbf{1}_{O^l}\right]\right\},\\ \sup_{\eta'\in\mathcal{H}(t_1)}\max_i\left\{\hat{p}_i - \sum_l\left[\int_{\Omega^l}g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\rho_1^l d\mathbf{P}^l(\omega^l)\mathbf{1}_{O^l}\right]\right\}\end{array}\right\},\\
\leq&\min\left\{\begin{array}{l}\sum_l\left[\rho_0\sup_{\eta'\in\mathcal{H}(t_1)}\max_i\left\{\hat{p}_i/\rho_0 - \int_{\Omega^l}g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)d\mathbf{P}^l(\omega^l)\right\}\mathbf{1}_{O^l}\right],\\ \sum_l\left[\sup_{\eta'\in\mathcal{H}(t_1)}\max_i\left\{\hat{p}_i - \int_{\Omega^l}g_i\left(X_T^{t_1,X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\rho_1^l d\mathbf{P}^l(\omega^l)\right\}\mathbf{1}_{O^l}\right]\end{array}\right\}.
\end{aligned} \tag{36}$$

Shorten the first and second terms in the above upper bound $\min\{\cdot,\cdot\}$ as A and B, respectively. For B:

$$\begin{aligned}
&\sum_l \left[\sup_{\eta'\in\mathcal{H}(t_1)} \max_i \left\{\hat{p}_i - \int_{\Omega^l} g_i\left(X_T^{t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\rho_1^l d\mathbf{P}^l(\omega^l)\right\}\mathbf{1}_{O^l}\right]\\
\leq&\sum_l \left(V^{-*}(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0},\hat{p}) + 2\epsilon\right)\mathbf{1}_{O^l}\\
&\text{(because } \zeta^l \text{ is } (2\epsilon)\text{-optimal for } V^{-*} \text{ at } (t_1,x_1,\hat{p}) \text{ for any } x_1\in E_l.)\\
=&V^{-*}(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0},\hat{p}) + 2\epsilon\\
\leq&V_1^{-*}(t_0,t_1,x_0,\hat{p}) + 3\epsilon\\
&\text{(because } \zeta^0 \text{ is } \epsilon\text{-optimal for } V_1^{-*}(t_0,t_1,x_0,\hat{p}).)
\end{aligned} \tag{37}$$

For A, we consider the following scenarios: (1) If $\inf_{\eta'\in\mathcal{H}(t_1)}\sum_l \rho_1^l\mathbf{1}_{O^l} = +\infty$, i.e., there is always a chance for Player 2 to achieve constraint violation when the game starts at $(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0})$, and $\rho_0 = 1$, i.e., (η,ζ^0) does not induce constraint violation, then

$$\begin{aligned}
A =&\sum_l \left[\sup_{\eta'\in\mathcal{H}(t_1)} \max_i \left\{\hat{p}_i - \inf_{\eta'\in\mathcal{H}(t_1)}\int_{\Omega^l} g_i\left(X_T^{t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)d\mathbf{P}^l(\omega^l)\right\}\mathbf{1}_{O^l}\right]\\
>&\sum_l \left[\sup_{\eta'\in\mathcal{H}(t_1)} \max_i \left\{\hat{p}_i - \inf_{\eta'\in\mathcal{H}(t_1)}\int_{\Omega^l} g_i\left(X_T^{t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0},\eta',\zeta^l(\omega^l,\cdot)}\right)\rho_1^l d\mathbf{P}^l(\omega^l)\right\}\mathbf{1}_{O^l}\right]\\
=&B = -\infty.
\end{aligned} \tag{38}$$

If $\inf_{\eta'\in\mathcal{H}(t_1)}\sum_l \rho_1^l\mathbf{1}_{O^l} = \rho_0 = 1$, then $A = B$. This is the scenario where the game reduces to its unconstrained version. Applying the same analysis from B to have $A \leq \rho_0(U_1^{-*}(t_0,t_1,x_0,\hat{p}/\rho_0) + 3\epsilon)$. If $\inf_{\eta'\in\mathcal{H}(t_1)}\sum_l \rho_1^l\mathbf{1}_{O^l} = \rho_0 = +\infty$, then $A = B = -\infty$.

If $\inf_{\eta'\in\mathcal{H}(t_1)}\sum_l \rho_1^l\mathbf{1}_{O^l} = 1$ and $\rho_0 = +\infty$, the game starting from $(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta^0})$ will be played as an unconstrained one. $A < B$. Hence $A \leq \rho_0(U_1^{-*}(t_0,t_1,x_0,\hat{p}/\rho_0) + 3\epsilon) = -\infty$.

Combining these scenarios we have

$$\min\{A,B\} \leq \min\{\rho_0(U_1^{-*}(t_0,t_1,x_0,\hat{p}/\rho_0) + 3\epsilon), V_1^{-*}(t_0,t_1,x_0,\hat{p}) + 3\epsilon\}. \tag{39}$$

Since ϵ can be arbitrarily small, we have

$$\begin{aligned}
V^{-*}(t_0,x_0,\hat{p}) \leq& \min\left\{\rho_0 U_1^{-*}(t_0,t_1,x_0,\hat{p}/\rho_0), V_1^{-*}(t_0,t_1,x_0,\hat{p})\right\}\\
=&\inf_{\zeta\in\mathcal{Z}(t_0)}\min\left\{\sup_{\eta\in\mathcal{H}(t_0)}\rho_0 U^{-*}(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta},\hat{p}/\rho_0), \sup_{\eta\in\mathcal{H}(t_0)} V^{-*}(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta},\hat{p})\right\}
\end{aligned} \tag{40}$$

Lastly, if $x_0 \in \bar{\mathcal{Q}}(t_0)$, $\inf_\zeta \sup_\eta \rho_0 = -\infty$ by definition; otherwise, $\rho_0 = 1$ and $U^{-*} = V^{-*}$ at t_1. In both cases, the RHS of Eq. (40) becomes

$$\inf_{\zeta\in\mathcal{Z}(t_0)}\sup_{\eta\in\mathcal{H}(t_0)}\min\left\{\rho_0 U^{-*}(t_0, X_{t_1}^{t_0,x_0,\eta,\zeta},\hat{p}/\rho_0), V^{-*}(t_0, X_{t_1}^{t_0,x_0,\eta,\zeta},\hat{p})\right\}. \tag{41}$$

□

Theorem 4. *(V^{-*} is a subsolution of HJ). For any $\hat{p}\in\mathbb{R}^I$, the map $(t,x)\to V^{-*}(t,x,\hat{p})$ is a viscosity subsolution of the dual Hamilton-Jacobi equation:*

$$\min\left\{\rho(x)U^{-*}(t,x,\hat{p}/\rho(x)) - w,\ w_t + H^*(x,Dw)\right\} = 0 \text{ in } [0,T]\times\mathbb{R}^{d_x}, \tag{42}$$

where H is defined by Eq. (4) and $H^*(x,\xi) = -H(x,-\xi)$. $\bar{\rho}(t_0,x_0) = 1$ if $x_0\in\mathcal{Q}(t_0)$.

Proof. Let $\hat{p} \in \mathbb{R}^I$ be fixed, and let ϕ be a smooth test function such that

$$\phi(t,x) \geq V^{-*}(t,x,\hat{p}) \quad \forall (t,x) \in [0,T] \times \mathbb{R}^{d_x}, \tag{43}$$

with an equality at (t_0, x_0), where $t_0 \in [0,T)$. For any $v \in \mathcal{V}$, define a pure strategy $\zeta \in \mathcal{Z}(t_0)$ by setting

$$\zeta(\alpha)(t) = v \quad \forall \alpha \in \mathcal{A}(t_0),\ t \in [t_0, T]. \tag{44}$$

Since V^{-*} satisfies the subdynamic programming principle of Lemma 5, these exist $\epsilon > 0$, $h > 0$, and a pure strategy $\eta_h \in \mathcal{H}(t_0)$ such that

$$\begin{aligned} V^{-*}(t_0,x_0,\hat{p}) \leq \min \Big\{ & \rho\left(\mathcal{X}_{t_0+h}^{t_0,x_0,\eta_h,\zeta}\right) U^{-*}\left(t_0+h, X_{t_0+h}^{t_0,x_0,\eta_h,\zeta}, \hat{p}/\rho\left(\mathcal{X}_{t_0+h}^{t_0,x_0,\eta_h,\zeta}\right)\right), \\ & V^{-*}\left(t_0+h, X_{t_0+h}^{t_0,x_0,\eta_h,\zeta}, \hat{p}\right)\Big\} + \epsilon h, \end{aligned} \tag{45}$$

or equivalently

$$\rho\left(\mathcal{X}_{t_0+h}^{t_0,x_0,\eta_h,\zeta}\right) U^{-*}\left(t_0+h, X_{t_0+h}^{t_0,x_0,\eta_h,\zeta}, \hat{p}/\rho\left(\mathcal{X}_{t_0+h}^{t_0,x_0,\eta_h,\zeta}\right)\right) - V^{-*}(t_0,x_0,\hat{p}) + \epsilon h \geq 0, \tag{46}$$

and

$$V^{-*}\left(t_0+h, X_{t_0+h}^{t_0,x_0,\eta_h,\zeta}, \hat{p}\right) - V^{-*}(t_0,x_0,\hat{p}) + \epsilon h \geq 0. \tag{47}$$

Set the open-loop control $\alpha_h(s) := \eta_h(v)(s)$ and the trajectory $x_h(s) = X_s^{t_0,x_0,\eta_h,\beta} = X_s^{t_0,x_0,\alpha_h,v}$. Then

$$x_h(t_0+h) = x_0 + \int_{t_0}^{t_0+h} f(x_h(s), \alpha_h(s), v) ds = x_0 + \int_{t_0}^{t_0+h} f(x_0, \alpha_h(s), v) ds + h\epsilon(h), \tag{48}$$

where $\epsilon(h) \to 0$ as $h \to 0^+$. For Eq. (46), let $\epsilon \to 0^+$ and $h \to 0^+$, we have

$$\rho(x_0) U^{-*}(t_0, x_0, \hat{p}/\rho(x_0)) - \phi(t_0,x_0) \geq 0. \tag{49}$$

For Eq. (47), we have

$$\begin{aligned} 0 &\leq V^{-*}\left(t_0+h, X_{t_0+h}^{t_0,x_0,\eta_h,\zeta}, \hat{p}\right) - V^{-*}(t_0,x_0,\hat{p}) + \epsilon h \\ &\leq \phi\left(t_0+h, x_0 + \int_{t_0}^{t_0+h} f(x_0,\alpha_h(s),v)ds + h\epsilon(h), z\right) - \phi(t_0,x_0) + \epsilon h \\ &\leq h\phi_t(t_0,x_0) + \int_{t_0}^{t_0+h} D\phi(t_0,x_0)^T f(x_0,\alpha_h(s),v) ds + h\epsilon_1(h) + \epsilon h \\ &\leq h\phi_t(t_0,x_0) + h \sup_{u\in\mathcal{U}} D\phi(t_0,x_0)^T f(x_0,u,v) + h\epsilon_1(h) + \epsilon h, \end{aligned} \tag{50}$$

where $\epsilon_1(h) \to 0$ as $h \to 0^+$. Dividing the last inequality by h, letting $h \to 0^+$, $\epsilon \to 0^+$, and taking the infimum over $v \in \mathcal{V}$ to have

$$\phi_t(t_0,x_0) + \inf_{v\in\mathcal{V}} \sup_{u\in\mathcal{U}} D\phi(t_0,x_0)^T f(x_0,u,v) \geq 0. \tag{51}$$

Now notice that by definition

$$H^*(x, D\phi) = -H(x,-D\phi) = \inf_{v\in\mathcal{V}} \sup_{u\in\mathcal{U}} f(x,u,v)^T D\phi. \tag{52}$$

Hence

$$\phi_t(t_0,x_0) + H^*(x_0, D\phi(t_0,x_0)) \geq 0. \tag{53}$$

Hence

$$\min\left\{\rho(x)U^{-*}(t,x,\hat{p}/\rho(x)) - \phi,\ \phi_t + H^*(x,D\phi)\right\} \geq 0 \text{ in } [0,T]\times\mathbb{R}^{d_x}. \tag{54}$$

□

Using the same proof techniques we can derive the subdynamic principle for V^+ (Lemma 6) and prove that V^+ is a viscosity subsolution to the primal HJ (Lemma 7):

Lemma 6. *(subdynamic principle for V^+). We have for any $(t_0, x_0, p) \in [0,T) \times \mathbb{R}^{d_x} \times \Delta(I)$ and any $t_1 \in (t_0, T]$*

$$V^+(t_0,x_0,p) \leq \inf_{\eta\in\mathcal{H}(t_0)} \sup_{\zeta\in\mathcal{Z}(t_0)} \max\left\{\rho(\mathcal{X}_{t_1}^{t_0,x_0,\eta,\zeta})U^+\left(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta}, p\right), V^+\left(t_1, X_{t_1}^{t_0,x_0,\eta,\zeta}, p\right)\right\} \tag{55}$$

Lemma 7. *(V^+ is a subsolution of HJ). For any $p \in \Delta(I)$, the map $(t,x) \to V^+(t,x,p)$ is a viscosity subsolution of the primal Hamilton-Jacobi equation:*

$$\min\left\{\rho(x)U^+(t,x,p) - w,\ w_t + H(x,Dw)\right\} = 0 \text{ in } [0,T]\times\mathbb{R}^{d_x}, \tag{56}$$

where H is defined by Eq. (4).

B. Proof of Theorem 2

Proof. The following proof follows that of (Cardaliaguet, 2009), with the additional treatment of the state constraint. Note that in numerical approximation, we use $K > 0$ to replace infinite values so that we can introduce bounded test functions. The HJ equations thus become

$$\begin{cases} w_t + H(x,Dw) = 0 & (t,x) \in \Omega \\ \min\{K - w, w_t + H(x,Dw)\} = 0 & (t,x) \in \bar{\Omega}, \end{cases}$$

where Ω (resp. $\bar{\Omega}$) contains all (t,x) such that $V(t,x) < K$ (resp. $V(t,x) = K$).

Consider w be any cluster point in the topology of uniform convergence on compact subsets of $[0,T]\times\mathbb{R}^{d_x}\times\Delta(I)$ of V_τ as $\tau \to 0^+$. w is convex with respect to p and satisfies:

$$w(T,x,p) = \sum_{i=1}^{I} p_i g_i(x), \tag{57}$$

for any $(T,x,p) \in \Omega \times \Delta(I)$ and $w(T,x,p) = K$ for any $(T,x,p) \in \bar{\Omega}\times\Delta(I)$. Let ϕ be a test function such that $w(\cdot,\cdot,p) - \phi$ has a strict local maximum at (t_0,x_0), and $w(t_0,x_0,p) = \phi(t_0,x_0)$. Then there are (t_k,x_k) converging to (t_0,x_0) such that $V_\tau(\cdot,\cdot,p) - \phi$ has a local maximum at (t_k,x_k).

First consider $(t_k,x_k) \in \Omega$. For any $x \in \mathbb{R}^{d_x}$,

$$V_\tau(t_{k+1},x,p) - \phi(t_{k+1},x) \leq V_\tau(t_k,x_k,p) - \phi(t_k,x_k)$$

Then, rearranging (10) to have

$$\begin{aligned} 0 &= \mathrm{Vex}_p\left(\min_u \max_v V_\tau(t_{k+1}, x_k + \tau f(x_k,u,v), p)\right) - V_\tau(t_k,x_k,p) \\ &\leq \min_u \max_v V_\tau(t_{k+1}, x_k + \tau f(x_k,u,v), p) - V_\tau(t_k,x_k,p) \\ &\leq \min_u \max_v \phi(t_{k+1}, x_k + \tau f(x_k,u,v)) - \phi(t_k,x_k) \end{aligned}$$

Then, from standard arguments (see (Cardaliaguet, 2009) and references therein)

$$\frac{\partial\phi}{\partial t}(t_0,x_0) + \min_{u\in U}\max_{v\in V} f(x_0,u,v)\frac{\partial\phi}{\partial x}(t_0,x_0) \geq 0. \tag{58}$$

Now consider $(t_k,x_k) \in \bar{\Omega}$, in which case $\phi(t_k,x_k) = V_\tau(t_k,x_k,p) = K$. When $\tau \to 0^+$,

$$\min_u \max_v \phi(t_{k+1}, x_k + \tau f(x_k,u,v)) = K,$$

hence

$$\frac{\partial \phi}{\partial t}(t_0, x_0) + \min_{u \in U} \max_{v \in V} f(x_0, u, v) \frac{\partial \phi}{\partial x}(t_0, x_0) = 0.$$

Then we have

$$\min\left\{K - \phi,\ \phi_t + H(x, D\phi)\right\} = 0 \text{ in } \bar{\Omega}.$$

Hence, w is a dual subsolution of the HJI. We can follow the same technique to show that w is a supersolution in the dual sense, and therefore $w = V$. □

C. Bellman Backup of the Conjugate Value with the Presence of Instantaneous Loss

Here we extend the subdynamic principle for V^{-*} (Lemma 5) when instantaneous loss is present. Since we will use letter l to index possible states reached at t_1 from t_0, we denote the instantaneous loss for pure strategies (η, ζ) at time s as $L(\eta, \zeta, s)$ instead. For conciseness, let us consider $t_0 \in [0, T]$ and $(x_0, \hat{p}) \in \mathcal{Q}(t_0) \times \mathbb{R}^I$, i.e., states for which Player 1 can play to avoid state constraint violation. To recap, let η be a pure strategy of Player 1; let ζ be such that $\zeta = \zeta^0$ in $[t_0, t_1]$ where ζ^0 is pure and ϵ-optimal for $V_1^{-*}(t_0, t_1, x_0, \hat{p}) :=:= \inf_{\zeta \in \mathcal{Z}(t_0)} \sup_{\eta \in \mathcal{H}(t_0)} V^{-*}\left(t_1, X_{t_1}^{t_0, x_0, \eta, \zeta}, \hat{p} - \int_{t_0}^{t_1} L(\eta, \zeta, s) ds\right)$, and $\zeta = \zeta^l$ in $[t_1, T]$ where ζ^l is mixed and (2ϵ)-optimal for V^{-*} at $(t_1, x_1, \hat{p})$ for any $x_1 \in E_l$.

By definition and using the fact that Player 1 plays a pure strategy in the dual game, we have

$$V^{-*}(t_0, x_0, \hat{p}) = \inf_{\zeta} \sup_{\eta} \max_i \left\{ \hat{p}_i - \int_\omega \left(g_i(X_T^{t_0, x_0, \eta, \zeta(\omega, \cdot)}) + \int_{t_0}^T L(\eta, \zeta(\omega, \cdot), s) ds \right) d\mathbf{P}(\omega) \right\}. \tag{59}$$

Here

$$\begin{aligned}
&\max_i \left\{ \hat{p}_i - \int_\omega \left(g_i(X_T^{t_0, x_0, \eta, \zeta(\omega, \cdot)}) + \int_{t_0}^T L(\eta, \zeta(\omega, \cdot), s) ds \right) d\mathbf{P}(\omega) \right\} \\
=&\max_i \left\{ \hat{p}_i - \sum_l \left(\int_{\omega^l} \left(g_i(X_T^{t_1, X_{t_1}^{t_0, x_0, \eta, \zeta^0}, \eta, \zeta^l(\omega, \cdot)}) + \int_{t_0}^{t_1} l(\eta(s), \zeta^0(\omega, \cdot)(s)) ds + \int_{t_1}^T L(\eta, \zeta^l(\omega, \cdot), s) ds \right) d\mathbf{P}^l(\omega^l) \right) \mathbf{1}_{O^l} \right\} \\
\leq& \sum_l \sup_{\eta'} \max_i \left\{ \hat{p}_i - \left(\int_{\omega^l} g_i(X_T^{t_1, X_{t_1}^{t_0, x_0, \eta, \zeta^0}, \eta', \zeta^l(\omega^l, \cdot)}) + \int_{t_0}^{t_1} L(\eta, \zeta^0, s) ds + \int_{t_1}^T L(\eta', \zeta^l(\omega^l, \cdot), s) ds d\mathbf{P}^l(\omega^l) \right) \right\} \mathbf{1}_{O^l} \\
\leq& \sum_l \left(V^{-*}\left(t_1, X_{t_1}^{t_0, x_0, \eta, \zeta^0}, \hat{p} - \int_{t_0}^{t_1} L(\eta, \zeta^0, s) ds \right) + 2\epsilon \right) \mathbf{1}_{O^l} \\
&\text{(because } \zeta^l \text{ is } (2\epsilon)\text{-optimal for } V^{-*} \text{ at } (t_1, x_1, \hat{p}) \text{ for any } x_1 \in E_l.) \\
=& V^{-*}\left(t_1, X_{t_1}^{t_0, x_0, \eta, \zeta^0}, \hat{p} - \int_{t_0}^{t_1} L(\eta, \zeta^0, s) ds \right) + 2\epsilon \\
\leq& V_1^{-*}(t_0, t_1, x_0, \hat{p}) + 3\epsilon \\
&\text{(because } \zeta^0 \text{ is } \epsilon\text{-optimal for } V_1^{-*}(t_0, t_1, x_0, \hat{p}).)
\end{aligned} \tag{60}$$

Since ϵ can be arbitrarily small, we have

$$V^{-*}(t_0, x_0, \hat{p}) \leq \inf_{\zeta \in \mathcal{Z}(t_0)} \sup_{\eta \in \mathcal{H}(t_0)} V^{-*}(t_1, X_T^{t_0, x_0, \eta, \zeta}, \hat{p} - \int_{t_0}^{t_1} L(\eta, \zeta, s) ds) \tag{61}$$

D. Examples of Zero-Sum Games with One-Sided Information

Here we discuss two games in detail, namely, the zero-sum beer-quiche game which is extensive-form with sequential actions, and Hexner's game (Hexner, 1979) which is differential and with simultaneous actions. Both games have one-sided information and all analytically solved. We show that the characterization of value proposed by Cardaliaguet (Cardaliaguet, 2007) leads to the true equilibrium behavioral strategies for both games.

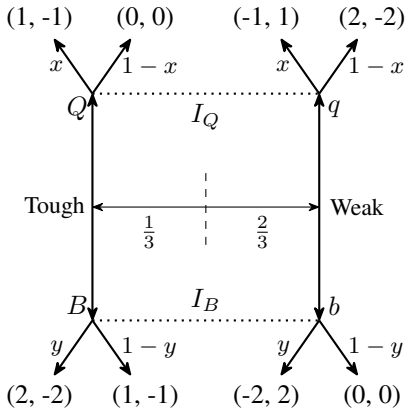

Figure 6: Zero-Sum Variant of the Beer-Quiche Game

D.1. Zero-sum beer-quiche game

We present a zero-sum variant of the classic beer-quiche game [2], which is an incomplete-information game with a perfect Bayesian equilibrium.

Game settings. In this sequential game, Player 1 first chooses to take either quiche (Q) or beer (B), and based on his choice, Player 2 chooses to either defer (d) or bully (b). Player 1 has a probability of p_T to be tough (T) and $p_W = 1 - p_T$ to be weak (W). The exact type is unknown to Player 2 but $p = [p_T, p_W]^T$ is common knowledge. The payoffs to be maximized by Player 1 follow Table 1. For example, if Player 1 is tough and chooses to eat quiche (Q) while Player 2 chooses to bully (b), then Player 1 receives a payoff of 1.

Table 1: Payoff table for a zero-sum beer-quiche game

Tough

	b	d
B	2	1
Q	1	0

Weak

	b	d
B	-2	0
Q	-1	2

Perfect Bayesian equilibrium. The standard approach finds the behavioral strategies of both players for a particular p. Consider the extensive form of the game as shown in Fig. 6. Dotted lines represent info sets that Player 2 cannot distinguish. Here, Player 1 has $p_T = \frac{1}{3}$ to be Tough. The behavioral strategies for each player are derived as follows:

Let Q, B, q, b represent probabilities of Player 1 choosing quiche given he is tough, beer given he is tough, quiche given he is weak, and beer given he is weak, respectively. Assume x and y be the probability of Player 2 bullying Player 1 who chooses quiche and beer, respectively. First, we find the beliefs of Player 2 when Player 1 chooses quiche or beer (info-set $\mathcal{I}_\mathcal{Q}$ and $\mathcal{I}_\mathcal{B}$, respectively):

$$\text{if } (Q,q) \neq (0,0), \quad \mu_2(T|\mathcal{I}_\mathcal{Q}) = \frac{\frac{1}{3}Q}{\frac{1}{3}(Q) + \frac{2}{3}(q)} = \frac{Q}{Q+2q} \quad \text{and}$$

$$\mu_2(W|\mathcal{I}_\mathcal{Q}) = \frac{\frac{2}{3}q}{\frac{1}{3}Q + \frac{2}{3}q} = \frac{2q}{Q+2q}$$

$$\text{if } (B,b) \neq (0,0), \quad \mu_2(T|\mathcal{I}_\mathcal{B}) = \frac{B}{B+2b}, \quad \text{and} \quad \mu_2(W|\mathcal{I}_\mathcal{B}) = \frac{2b}{B+2b}$$

[2] For more information about the original beer-quiche game, please see https://gametheory101.com/courses/game-theory-101/the-beer-quiche-game/

Then, the expected payoffs for bully and defer at $\mathcal{I}_\mathcal{Q}$ are:

$$E_2(\text{bully}|\mathcal{I}_\mathcal{Q}) = \frac{Q}{Q+2q}(-1) + \frac{2q}{Q+2q}(1) = -\frac{Q-2q}{Q+2q}$$
$$E_2(\text{defer}|\mathcal{I}_\mathcal{Q}) = \frac{Q}{Q+2q}(0) + \frac{2q}{Q+2q}(-2) = -\frac{4q}{Q+2q}$$

Given Player 2's strategy at $\mathcal{I}_\mathcal{Q}$, his expected payoff can be expressed as:

$$\begin{aligned} E_2(\mathcal{I}_\mathcal{Q}) &= -\frac{Q-2q}{Q+2q}x - (1-x)\frac{4q}{Q+2q} \\ &= \frac{-(Q-6q)x - 4q}{Q+2q} \end{aligned}$$

The value of x that maximizes $E_2(\mathcal{I}_\mathcal{Q})$ is:

$$x = \begin{cases} \text{any} & \text{if} \quad (Q,q) = (0,0) \\ 1 & \text{if} \quad Q < 6q \\ \text{any} & \text{if} \quad Q = 6q \\ 0 & \text{if} \quad Q > 6q \end{cases}$$

Applying the same reasoning to info-set $\mathcal{I}_\mathcal{B}$, we find the value of y that maximizes $E_2(\mathcal{I}_\mathcal{B})$ as:

$$y = \begin{cases} \text{any} & \text{if} \quad (B,b) = (0,0) \\ 0 & \text{if} \quad 4b < B \\ \text{any} & \text{if} \quad 4b = B \\ 1 & \text{if} \quad 4b > B \end{cases}$$

Given Player 2's strategy, the expected payoffs to Player 1 for his strategies are:

$$E_1(Q) = x, \quad E_1(B) = y + 1, \quad E_1(q) = 2 - 3x, \quad E_1(b) = -2y$$

As a result the expected value for each of P1's type are:

$$E_1(T) = B(1+y) + (1-B)x$$

$$E_1(W) = b(-2y) + (1-b)(2-3x)$$

Assume $B \geq 4b$. Then,

$$\begin{aligned} & 1 - Q = B \geq 4b = 4(1-q) \\ \implies & 1 - Q \geq 4 - 4q \\ \implies & 4q \geq Q + 3 \\ \implies & 6q > Q \end{aligned}$$

Hence, $x = 1$. Thus, $E_1(Q) = 1 < E_1(B) = y + 1$. As a result, $B = 1$, and $Q = 0$.

Assuming $B > 4b$, following the process as above, we reach to a contradiction. Therefore,

$$B = 4b \implies b = \frac{B}{4} = \frac{1}{4}$$

Then,

$$\frac{\partial E_1(W)}{\partial b} = 1 - 2y$$

Hence, for $b = 1/4$ to be feasible, we need:

$$y = \frac{1}{2}$$

Therefore, we find an equilibrium with

$$x = 1,\ y = \frac{1}{2},\ q = \frac{3}{4},\ b = \frac{1}{4},\ \text{and } B = 1, Q = 0$$

To summarize, Player 2 always bullies the person who eats quiche and bullies the person drinking beer half the time. The tough guy always drinks beer while the weak guy drinks beer a quarter of the time and eats quiche three-quarters of the time. One can easily check that the $B < 4b$ case also leads to a contradiction, resulting in a unique equilibrium for the game.

Solution using primal and dual backward induction. Now we solve the game through backward induction of its primal and dual values (denoted by V and C respectively). Here we introduce discrete-time $t = 0, 1, 2$: Players 1 and 2 make their respective decisions at $t = 0$ and $t = 1$, and the game ends at $t = 2$. We describe the states of the game as the decisions being made up to the corresponding time, e.g., $x = (B, b)$ at $t = 2$ means that Player 1 has chosen beer and Player 2 to defer. **Primal game:** At the terminal time step ($t = 2$), based on the payoff table, we have

$$V(2, x, p) = \begin{cases} 4p_T - 2 & \text{if } x = (B, b) \\ p_T & \text{if } x = (B, d) \\ 2p_T - 1 & \text{if } x = (Q, b) \\ 2 - 2p_T & \text{if } x = (Q, d) \end{cases}. \tag{62}$$

At the intermediate time step ($t = 1$), we have

$$V(1, x, p) = \min_{v \in \{b,d\}} V(2, (x, v), p). \tag{63}$$

We can find the best responses of Player 2 for both actions of Player 1. This leads to

$$V(1, x, p) = \begin{cases} p_T & \text{if } x = B,\ 3p_T - 2 \geq 0 \quad (v^* = d) \\ 4p_T - 2 & \text{if } x = B,\ 3p_T - 2 < 0 \quad (v^* = b) \\ 2 - 2p_T & \text{if } x = Q,\ 4p_T - 3 \geq 0 \quad (v^* = d) \\ 2p_T - 1 & \text{if } x = Q,\ 4p_T - 3 < 0 \quad (v^* = b) \end{cases}. \tag{64}$$

Note that since Player 1 does not take an action in this time step, we do not need to take a concave hull of $V(1, x, \cdot)$. At the beginning of the game ($t = 0$), we have

$$V(0, \emptyset, p) = \text{Cav}\left(\max_{u \in \{B,Q\}} V(1, u, p)\right). \tag{65}$$

By taking the concave hull with respect to p_T (see Fig. 7), we get

$$V(0, \emptyset, p) = \begin{cases} 5p_T/2 - 1 & \text{if } p_T < 2/3 \\ p_T & \text{if } p_T \geq 2/3 \end{cases}. \tag{66}$$

Note that from Fig. 7, when $p_T \in [0, 2/3)$, $V(0, \emptyset, p) = \lambda \max_u V(1, u, p^1) + (1-\lambda) \max_u V(1, u, p^2)$, where $p^1 = [0, 1]^T$, $p^2 = [2/3, 1/3]^T$, and $\lambda p^1 + (1 - \lambda)p^2 = p$. When $p_T = 1/3$, $\lambda = 1/2$, Player 1's strategy is thus

$$\begin{aligned} &\Pr(u = Q|T) = \frac{\lambda p^1[1]}{p[1]} = 0, && \Pr(u = Q|W) = \frac{\lambda p^1[2]}{p[2]} = 3/4, \\ &\Pr(u = B|T) = \frac{(1-\lambda)p^2[1]}{p[1]} = 1, && \Pr(u = B|W) = \frac{(1-\lambda)p^2[2]}{p[2]} = 1/4. \end{aligned} \tag{67}$$

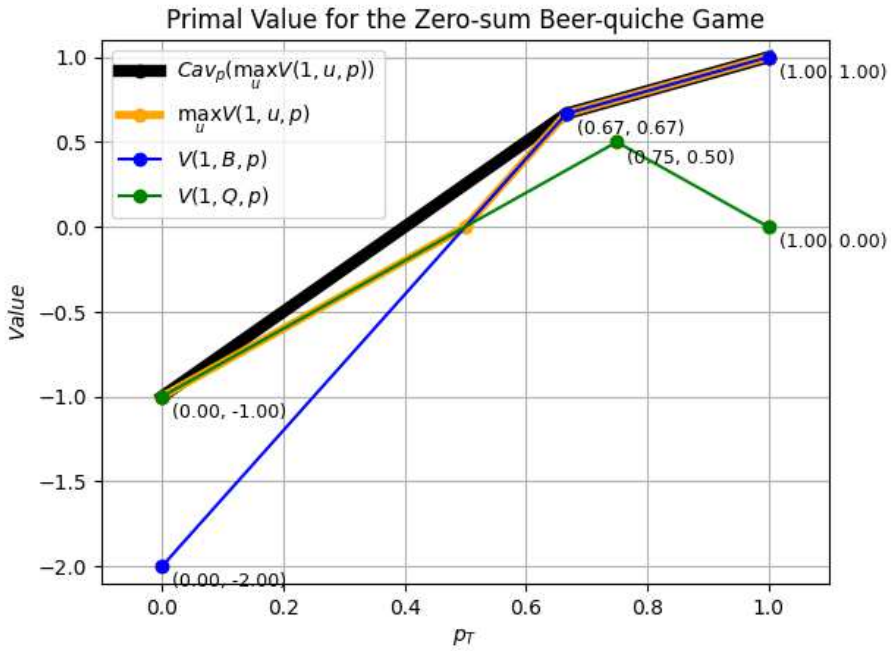

Figure 7: Primal value $V(0, \emptyset, p_T)$ at $t = 0$.

This result is consistent with the true perfect Bayesian equilibrium we previously derived.

Dual game: To solve for Player 2's equilibrium, we first derive the dual variable $\hat{p} \in \partial_p V(0, \emptyset, p)$ for $p = [1/3, 2/3]^T$. By definition, $\hat{p}^T p$ defines the concave hull of $V(0, \emptyset, p)$, and therefore we have

$$\begin{aligned} [1/3, 2/3]\hat{p} &= V(0, \emptyset, p) = -1/6 \\ [0, 1]\hat{p} &= V(0, \emptyset, [0, 1]) = -1. \end{aligned} \tag{68}$$

This leads to $\hat{p} = [3/2, -1]^T$.

At the terminal time, we have

$$\begin{aligned} C(2, x, \hat{p}) &= \min\{\hat{p}[1] - g_T(x), \hat{p}[2] - g_W(x)\} \\ &= \begin{cases} \min\{\hat{p}[1] - 2, \hat{p}[2] + 2\} & \text{if } x = (B, b) \\ \min\{\hat{p}[1] - 1, \hat{p}[2]\} & \text{if } x = (B, d) \\ \min\{\hat{p}[1] - 1, \hat{p}[2] + 1\} & \text{if } x = (Q, b) \\ \min\{\hat{p}[1], \hat{p}[2] - 2\} & \text{if } x = (Q, d) \end{cases} \end{aligned} \tag{69}$$

At $t = 1$, we have

$$C(1, u, \hat{p}) = \text{Cav}_{\hat{p}} \left(\max_v C(2, (u, v), \hat{p}) \right). \tag{70}$$

When $u = B$, the conjugate value is a concave hull of a piece-wise linear function:

$$\begin{aligned} C(1, B, \hat{p}) &= \text{Cav}_{\hat{p}} \left(\begin{cases} \hat{p}[1] - 1 & \text{if } \hat{p}[2] \geq \hat{p}[1] - 1 & (v^* = d) \\ \hat{p}[2] & \text{if } \hat{p}[2] \in [\hat{p}[1] - 2, \hat{p}[1] - 1) & (v^* = b) \\ \hat{p}[1] - 2 & \text{if } \hat{p}[2] \in [\hat{p}[1] - 4, \hat{p}[1] - 2) & (v^* = d) \\ \hat{p}[2] + 2 & \text{if } \hat{p}[2] < \hat{p}[1] - 4 & (v^* = b) \end{cases} \right) \\ &= \begin{cases} \hat{p}[1] - 1 & \text{if } \hat{p}[2] \geq \hat{p}[1] - 1 & (v^* = d) \\ 2/3\hat{p}[1] + 1/3\hat{p}[2] - 2/3 & \text{if } \hat{p}[2] \in [\hat{p}[1] - 4, \hat{p}[1] - 1) & \text{(mixed strategy)} \\ \hat{p}[2] + 2 & \text{if } \hat{p}[2] < \hat{p}[1] - 4 & (v* = b) \end{cases} \end{aligned} \tag{71}$$

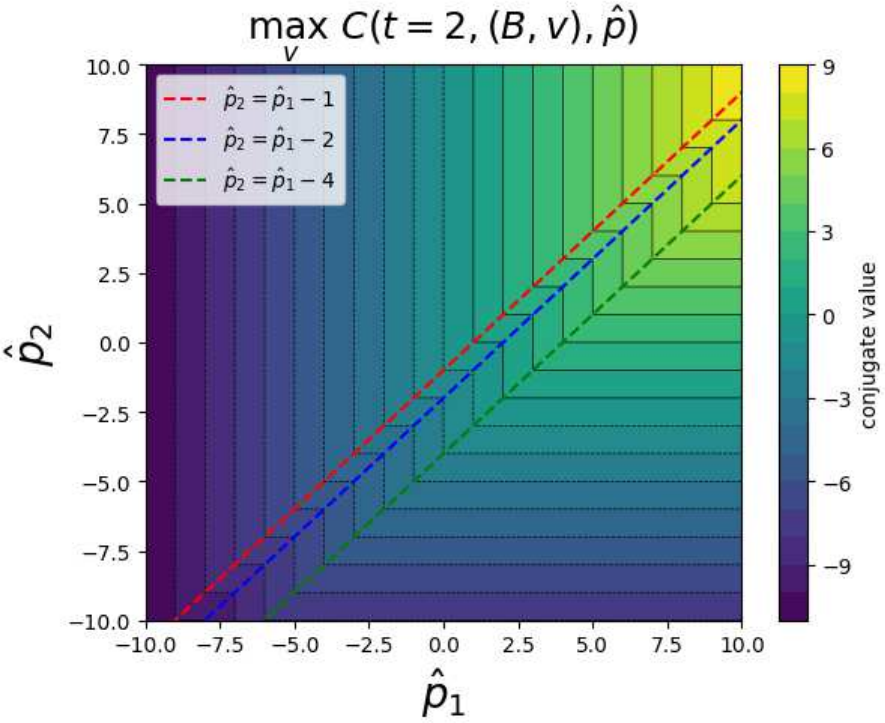

Figure 8: Conjugate value $\max_v\ C(2, B, \hat{p})$ at $t = 2$.

Fig. 8 visualizes $C(1, B, \hat{p})$. For $\hat{p} = [3/2, -1]^T$ which satisfies $\hat{p}[2] \in [\hat{p}[1] - 4, \hat{p}[1] - 1)$, Player 2 follows a mixed strategy determined based on $\{\lambda_1, \lambda_2, \lambda_3\} \in \Delta(3)$ and $\hat{p}^j \in \mathbb{R}^2$ for $j = 1, 2, 3$ such that

(i) At least one of $\hat{p}^j$ for $j = 1, 2, 3$ should satisfy $\hat{p}[2] = \hat{p}[1] - 1$ (denoted as line 1) and another $\hat{p}[2] = \hat{p}[1] - 4$ (denoted as line 2). The last could be on either line 1 or 2. These conditions are necessary for $C(1, B, \hat{p})$ to be a concave hull:

$$C(1, B, \hat{p}) = \sum_{j=1}^{3} \lambda_j \max_v C(2, (B, v), \hat{p}^j). \tag{72}$$

Without loss of generality, we will set $\hat{p}^1$ on line 1 and both $\hat{p}^2$ and $\hat{p}^3$ on line 2;

(ii) $\sum_{j=1}^{3} \lambda_j \hat{p}^j = \hat{p}$.

These conditions leads to $\lambda_1 = 1/2$ and $\lambda_2 + \lambda_3 = 1/2$. Therefore Player 2 chooses to defer and bully with equal chance when Player 1 takes beer.

When $u = Q$, we similarly have

$$C(1, Q, \hat{p}) = \begin{cases} \hat{p}[1] & \text{if } \hat{p}[2] \geq \hat{p}[1] + 2 & (v^* = d) \\ \dots & \text{if } \hat{p}[2] \in [\hat{p}[1] - 2, \hat{p}[1] + 2) & (\text{mixed strategy}) \\ \hat{p}[2] + 1 & \text{if } \hat{p}[2] < \hat{p}[1] - 2 & (v* = b) \end{cases} \tag{73}$$

We omitted the derivation of the concave hull when $\hat{p}[2] \in [\hat{p}[1] - 2, \hat{p}[1] + 2)$ because for $\hat{p} = [3/2, -1]^T$, $C(1, Q, \hat{p}) = \hat{p}[2] + 1 = 0$ while $v^* = b$, i.e. if Player 1 takes quiche, Player 2 chooses to bully with certainty.

The value and its conjugate provide behavioral strategies for Player 1 (informed) and Player 2 (non-informed), respectively, for arbitrary initial belief p. Moreover, the convexity of the value reveals subsets of p where Player 1 should use a mixed strategy that manipulates the belief in order to improve its value. Similarly, the convexity of the conjugate value reveals subsets of dual variables $\hat{p}$ where Player 2 should use a mixed strategy to mitigate risks due to its uncertainty about Player 1.

D.2. Hexner's game

Here we discuss the solution to Hexner's game using Cardaliaguet's method based on the reformulation proposed by Hexner. To recap, the payoff to be minimized by Player 1 is

$$J(t,\tilde{\theta}_1,\tilde{\theta}_2)=\mathbb{E}_\theta\left[\int_{\tau=t}^{T}(\tilde{\theta}_1(\tau)-\theta)^2 d_1(\tau)-(\tilde{\theta}_2(\tau)-\theta)^2 d_2(\tau)d\tau\right], \tag{74}$$

where d_1, d_2, p_θ are common knowledge; θ is only known to Player 1; the scalar $\tilde{\theta}_1$ (resp. $\tilde{\theta}_2$) is Player 1's (resp. Player 2's) strategy. We consider two player types $\theta\in\{-1,1\}$ and therefore $p_\theta\in\Delta(2)$. Since the reformulation contains no system state, the strategies are functions of only time. Hexner's solution is as follows:

$$\tilde{\theta}_1(s)=\tilde{\theta}_2(s)=0 \quad \forall s\in[0,t_r] \tag{75}$$

$$\tilde{\theta}_1(s)=\tilde{\theta}_2(s)=\theta \quad \forall s\in(t_r,T], \tag{76}$$

where

$$t_r=\arg\min_t \int_0^t (d_1(s)-d_2(s))ds, \tag{77}$$

and (d_1,d_2) are defined in Eq. (2).

We will need the following preparation before introducing Cardaliaguet's solution. First, introduce time stamps $[T_k]_{k=1}^{2r}$ as roots of the time-dependent function d_1-d_2, with $T_0=0$, $T_{2q+1}=t_r$, and $T_{2r+1}=T$. Without loss of generality, we assume that:

$$d_1-d_2<0 \quad \forall t\in(T_{2k},T_{2k+1})\ \forall k=0,...,r, \tag{78}$$

$$d_1-d_2\geq 0 \quad \forall t\in[T_{2k-1},T_{2k}]\ \forall k=1,...,r. \tag{79}$$

We also introduce $D_k:=\int_{T_k}^{T_{k+1}}(d_1-d_2)ds$ and

$$\tilde{D}_k=\begin{cases}\tilde{D}_{k+1}+D_k & \text{if } \tilde{D}_{k+1}+D_k<0\\ 0 & \text{otherwise}\end{cases}, \tag{80}$$

with $\tilde{D}_{2r+1}=0$.

Lemma 8. *(Properties of D_k and $\tilde{D}_k$) The following properties will be useful:*

1. $\int_k^{2q+1}(d_1-d_2)ds=\sum_k^{2q}D_k<0$, $\forall k=0,...,2q$;
2. $\int_{2q+1}^{k}(d_1-d_2)ds=\sum_{2q+1}^{k-1}D_k>0$, $\forall k=2q+2,...,2r+1$;
3. $\tilde{D}_{2q+2}+D_{2q+1}>0$;
4. $\tilde{D}_k<0$, $\forall k<2q+1$.

Proof. Properties 1 and 2 are results directly from the definition of D_k.

For property 3, if $\tilde{D}_{2q+2}+D_{2q+1}\leq 0$, then $\tilde{D}_{2q+2}=\tilde{D}_{2q+3}+D_{2q+2}\leq -D_{2q+1}$, then $\tilde{D}_{2q+3}\leq -(D_{2q+2}+D_{2q+1})<0$ (property 2). This leads to $\tilde{D}_{2q+k}\leq -\sum_{i=1}^{k-1}D_{2q+i}<0$ for $k=1,...,2r-2q$. Thus $\tilde{D}_{2r}<0$. Contradiction.

For property 4, first we have $\tilde{D}_{2q+1}=0$ (property 3). Since $D_{2q}<0$ (property 1), $\tilde{D}_{2q}=D_{2q}<0$. □

Primal game. We start with $V(T,p)=0$ where we use $p:=p_\theta[1]$ as the probability of $\theta=-1$. The Hamiltonian can be derived as

$$\begin{aligned}H(p)&=\min_{\tilde{\theta}_1}\max_{\tilde{\theta}_2}\mathbb{E}_\theta\left[(\tilde{\theta}_1-\theta)^2 d_1-(\tilde{\theta}_2-\theta)^2 d_2\right]\\&=4p(1-p)(d_1-d_2).\end{aligned}$$

The optimal actions for the Hamiltonian are $\tilde{\theta}_1 = \tilde{\theta}_2 = 1 - 2p$. From Bellman backup, we can get

$$V(T_k, p) = 4p(1-p)\tilde{D}_k.$$

Therefore, at T_{2q+1}, we have

$$\begin{aligned} V(T_{2q+1}, p) &= Vex_p \left(V(T_{2q+2}, p) + 4p(1-p)D_{2q+1} \right) \\ &= Vex_p \left(4p(1-p)(\tilde{D}_{2q+2} + D_{2q+1}) \right). \end{aligned}$$

Notice that $\tilde{D}_{2q+2} + D_{2q+1} > 0$ (property 3) and $\tilde{D}_k < 0$ for all $k < 2q+1$ (property 4), T_{2q+1} is the first time such that the right-hand side term inside the convexification operator, i.e., $4p(1-p)(\tilde{D}_{2q+2} + D_{2q+1})$, becomes concave. Therefore, splitting of belief happens at T_{2q+1} with $p^1 = 0$ and $p^2 = 1$. Player 1 plays $\tilde{\theta}_1 = -1$ (resp. $\tilde{\theta}_1 = 1$) with probability 1 if $\theta = -1$ (resp. $\theta = 1$), i.e., Player 1 reveals its type. This result is consistent with Hexner's.

Dual game. To find Player 2's strategy, we need to derive the conjugate value which follows

$$C(t, \hat{p}) = \begin{cases} \max_{i \in \{1,2\}} \hat{p}[i] & \forall t \geq T_{2q+1} \\ \hat{p}[2] - \tilde{D}_t \left(1 - \frac{\hat{p}[1] - \hat{p}[2]}{4\tilde{D}_t}\right)^2 & \forall t < T_{2q+1},\ 4\tilde{D}_t \leq \hat{p}[1] - \hat{p}[2] \leq -4\tilde{D}_t \\ \hat{p}[1] & \forall t < T_{2q+1},\ \hat{p}[1] - \hat{p}[2] \geq 4\tilde{D}_t \\ \hat{p}[2] & \forall t < T_{2q+1},\ \hat{p}[1] - \hat{p}[2] < 4\tilde{D}_t \end{cases}$$

Here $\hat{p} \in \nabla_{p_\theta} V(0, p_\theta)$ and $V(0, p_\theta) = 4p[1]p[2]\tilde{D}_0$. For any particular $p_* \in \Delta(2)$, from the definition of subgradient, we have $\hat{p}[1]p_*[1] + \hat{p}[2]p_*[2] = 4p_*[1]p_*[2]\tilde{D}_0$ and $\hat{p}[1] - \hat{p}[2] = 4(p_*[2] - p_*[1])\tilde{D}_0$. Solving these to get $\hat{p} = [4p_*[2]^2\tilde{D}_0, 4p_*[1]^2\tilde{D}_0]^T$. Therefore $\hat{p}[1] - \hat{p}[2] = 4\tilde{D}_0(1 - 2p_*[1]) \in [4\tilde{D}_0, -4\tilde{D}_0]$, and

$$C(0, \hat{p}) = \hat{p}[2] - \tilde{D}_0 \left(1 - \frac{\hat{p}[1] - \hat{p}[2]}{4\tilde{D}_0}\right)^2.$$

Notice that $C(t, \hat{p})$ is convex to $\hat{p}$ since $\tilde{D}_0 < 0$ (property 4) for all $t \in [0, T]$. Therefore, there is no splitting of $\hat{p}$ during the dual game, i.e., $\tilde{\theta}_2 = 1 - 2p$. This result is also consistent with Hexner's.

E. Backward Reachable Tube

The computation of the Backward Reachable Tube (BRT) allows us to classify the state space into feasible and infeasible regions at different times from Player 1's perspective.

Computation of BRT. For the simplified football game, the state constraint is defined as $c(x) := \|(d_{x_1}, d_{y_1}) - (d_{x_2}, d_{y_2})\|_2 - r$, and $\mathcal{C} = \{x : c(x) \leq 0\}$. The Hamilton-Jacobi-Isaacs Variational Inequality (HJI VI) is denoted by L and satisfies the boundary condition D (Bansal & Tomlin, 2021):

$$\begin{aligned} &L(\tilde{V}, t, x) = \min\{\nabla_t \tilde{V}(t, x) + H(t, x), c(x) - \tilde{V}(t, x)\} = 0, \\ &D(\tilde{V}, x) = \tilde{V}(T, x) - c(x) = 0, \end{aligned} \tag{81}$$

where H is the Hamiltonian:

$$H(t, x) = \max_u \min_v \langle \nabla_x \tilde{V}(t, x), f(x, u, v) \rangle. \tag{82}$$

We use Physics-Informed Neural Network (PINN) to learn the value function $\tilde{V}(t, x)$, the sub-zero level set of which represents the BRT:

$$\bar{\mathcal{Q}}(t) = \{x \in \mathbb{R}^{d_x} : \tilde{V}(t, x) \leq 0\}. \tag{83}$$

We denote PINN dataset $\mathcal{D} = \left\{ \left(t^{(k)}, x^{(k)}\right) \right\}_{k=1}^K$ containing uniformly sampled data points in $[0, T] \times \mathbb{R}^{d_x}$ and define the loss function as:

$$\min_{\tilde{V}} \quad \mathcal{L}\left(\tilde{V}\right) = \sum_{k=1}^K \left\| L(\tilde{V}^{(k)}, t^{(k)}, x^{(k)}) \right\|_1 + C_1 \left\| D(\tilde{V}^{(k)}, x^{(k)}) \right\|_1, \tag{84}$$

where $\tilde{V}^{(k)}$ is an abbreviation for $\tilde{V}\left(t^{(k)}, x^{(k)}\right)$ and C_1 is the hyperparameter that balances the loss term $\|L\|_1$ and $\|D\|_1$.

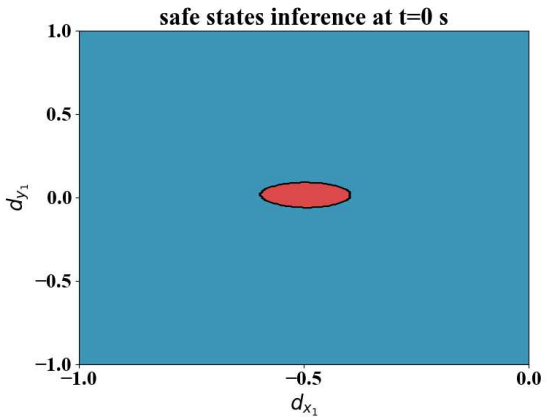

Figure 9: A visualization of safe/unsafe initial position for the attacker when the defender is fixed at $(-0.05, 0)$. The initial velocities for both players are zero in the beginning. The red (blue) region represents unsafe (safe) states.

Training. We uniformly sample 60k input states $x \in [-1, 1]$ (specifically, 10k states $x \in \mathcal{C}$) and use curriculum learning proposed in (Bansal & Tomlin, 2021) to improve the training convergence. The rest of the dynamics parameters are chosen as: $T = 1$, $r = 0.05$, $u_x \in [-6, 6]$, $u_y \in [-12, 12]$, $v_x \in [-6, 6]$, $v_y \in [-4, 4]$, velocities are sampled as mentioned in Sec. G.2 and are normalized between $[-1, 1]$. The PINN utilizes a fully-connected network with 3 hidden layers, each comprising 512 neurons with `sin` activation function. The network adopts the Adam optimizer with a fixed learning rate of 2×10^{-5}. We first pretrain the network over 10k iterations to satisfy the boundary condition D and then refine the network through 100k gradient descent steps, with states sampled from an expanding time window starting from the terminal. Fig. 9 shows the visualization of BRT in a 2D frame given $t = 0$ and fixed states except (d_{x_1}, d_{y_1}).

F. Proof of Proposition 1

Proof. Let $f^0 : [0, 1]^{I-1} \to \mathbb{R}$ be a bounded and Lipschitz continuous function, $\mathcal{P} \subset [0, 1]^{I-1}$ be a lattice, and f be a convex hull computed from the data $\{f(p), p\}_{p\in\mathcal{S}}$. Let the true convex hull of f^0 be $Vex(f^0)$: $Vex(f^0)(p) \leq f(p)$ for all $p \in [0, 1]^{I-1}$, with equality reached at least for $p \in \mathcal{S}$.

Introduce a set $P^0 = \{p^{(i)} \in \mathcal{S}\}_{i=1}^I$ and a space $\mathcal{P}^0 = \{p \in [0, 1]^{I-1} \mid \exists\lambda \in \Delta(I) \text{ s.t. } p = \sum_{i=1}^I \lambda[i]p^{(i)},\ p^{(i)} \in P^0\}$ so that $f(p) = \sum_{i=1}^I \lambda[i]f(p^{(i)})$ for all $p \in \mathcal{P}^0$, i.e., P^0 are vertices of a segment $\mathcal{P}^0$ of $[0, 1]^{I-1}$ within which f is affine.

Let $U := \{u^{(i)}\}_{i=1}^N = \mathcal{P} \bigcap \mathcal{P}^0$ be the set of lattice nodes contained in $\mathcal{P}^0$. Since f is a convex hull of f^0, we have $f(u^{(i)}) \leq f^0(u^{(i)})$ for all $i = 1, ..., N$. U defines a segmentation $\mathcal{E}$ of $\mathcal{P}^0$: Each $e \in \mathcal{E}$ is associated with $U_e := \{u^{(e_i)}\}_{i=1}^I \subset U$ such that $e = \{p \in [0, 1]^{I-1} \mid \exists\lambda \in \Delta(I) \text{ s.t. } p = \sum_{i=1}^I \lambda[i]u^{(e_i)},\ u^{(e_i)} \in U_e\}$ and $u \notin e$ for any $u \in U \setminus U_e$.

For any $p \in e$, we have the following loose lower bound on $f^0(p)$:

$$f^0(p) \geq \min_i f^0(u^{e_i}) - \Delta L \geq \min_i f(u^{e_i}) - \Delta_e L, \tag{85}$$

where $\Delta := \max_{i,j} \|u^{e_i} - u^{e_j}\|_2$, and L is the Lipschitz constant of f^0. Δ is a constant for a given lattice $\mathcal{P}$.

Therefore within e, the convexification error is lower bounded by

$$\max_{\lambda\in\Delta(I)} \left\{ f(\sum_i \lambda[i]u^{(e_i)}) - \min_i f(u^{e_i}) + \Delta L \right\} = \max_i f(u^{e_i}) - \min_i f(u^{e_i}) + \Delta L \leq 2\Delta L. \tag{86}$$

Since this error is constant, and $f - 2\Delta L$ is a convex lower bound of $Vex(f^0)$, we have $\varepsilon_{vex} \leq 2\Delta L$. □

G. Details on Case Studies

The code for the implementation is available at `https://github.com/ghimiremukesh/OSIIG`.

G.1. Hexner's Strategy

For the unconstrained simplified football game discussed in Sec. 6, the strategies depend on the trajectory of the $d_1 - d_2$. In Fig.10, we plot the trajectory and determine the critical time from Eq.(77). For the choices of parameters, we determine $t_r = 0.4$s. We set $R_A = \texttt{diag}(0.05, 0.025)$, and $R_D = \texttt{diag}(0.05, 0.1)$.

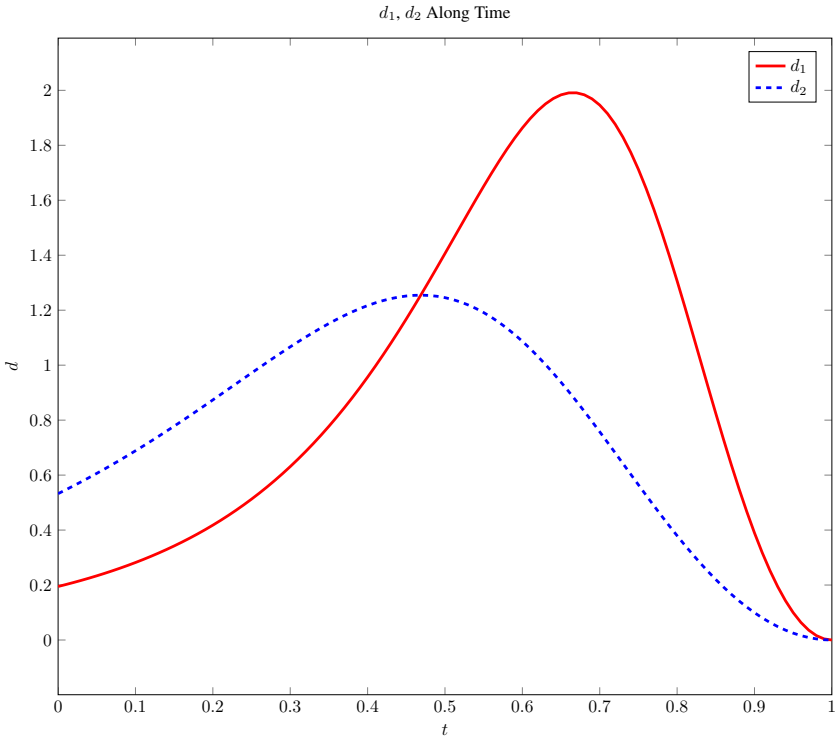

Figure 10: Plot of d_1 and d_2 along time. The critical time t_r occurs at $t \approx 0.4$. The attacker will conceal its type until t_r and reveal it after t_r.

G.2. Data Sampling

Unconstrained Game. For the unconstrained game, we sample positions (d_x, d_y) and velocities $(\dot{d}_x, \dot{d}_y)$ for both players. As the arena is bounded between $[-1, 1]$ in both x and y directions, we sample the positions of the two players in $[-1, 1]$. However, when it comes to velocities, we experimentally determine the range from the LQR problem as the following: $\dot{d}_{x_1} \in [-6, 6]$, $\dot{d}_{y_1} \in [-4, 4]$, $\dot{d}_{x_2} \in [-6, 6]$, and $\dot{d}_{y_2} \in [-4, 4]$. We then normalize the velocities between $[-1, 1]$ and compute the values as described in algorithm 1. The resulting normalized joint states $(\mathcal{X})$ and values (V) are stored for training the value network. At each time step we sample 10000 states and set $|\mathcal{P}| = 100$. This brings the total training data at each time step to 1M for the unconstrained case.

Constrained Primal Game. For the constrained game, we sample the positions between $[-1, 1]$ and all velocities between the ranges discussed above. As in the unconstrained case, these are normalized to $[-1, 1]$ before computing the values and storing the training data. With the same $\mathcal{P}$, we sample 5000 states from the feasible set $\mathcal{Q}(t)$, resulting in 500,000 training data at each time step. Solving constrained game requires evaluating $\min_u \max_v V(t, x + \tau f(x, u, v), p)$ over all possible pairs of x'(i.e. $x + \tau f(x, u, v)$), which is memory intensive. Based on the available resources we set the total number of initial states to be sampled to 5000. To speed up the calculation, and capture a wide range of data, we divide the state space into 50 uniform intervals, and distribute the computation to 56 CPU cores, with 515,271 MB of total memory. Each minimax computation is independent and hence can be evaluted in parallel.

Constrained Dual Game. In the dual game, the uninformed player keeps track of the process $\hat{p} \in \mathbb{R}^I$. As a result, the dual value is a 10-D function, which increases the complexity of the computation due to the need for convexification of the

value along I dimensions (here, $I = 2$). We follow the same procedure as in the primal game and collect 250,000 samples for training. The range of $\hat{p}$ was determined to be $[-14, 14]$ based on the primal value at the initial time as discussed in Sec. 5. Furthermore, due to the additional input dimension in the dual value network, the dual value approximation suffers from relatively higher error compared to the primal value. Ultimately this affects the strategy of the uninformed player (Player 2). We compare the resulting strategy of Player 2 from the dual value with that of the ground truth strategy in the unconstrained game.

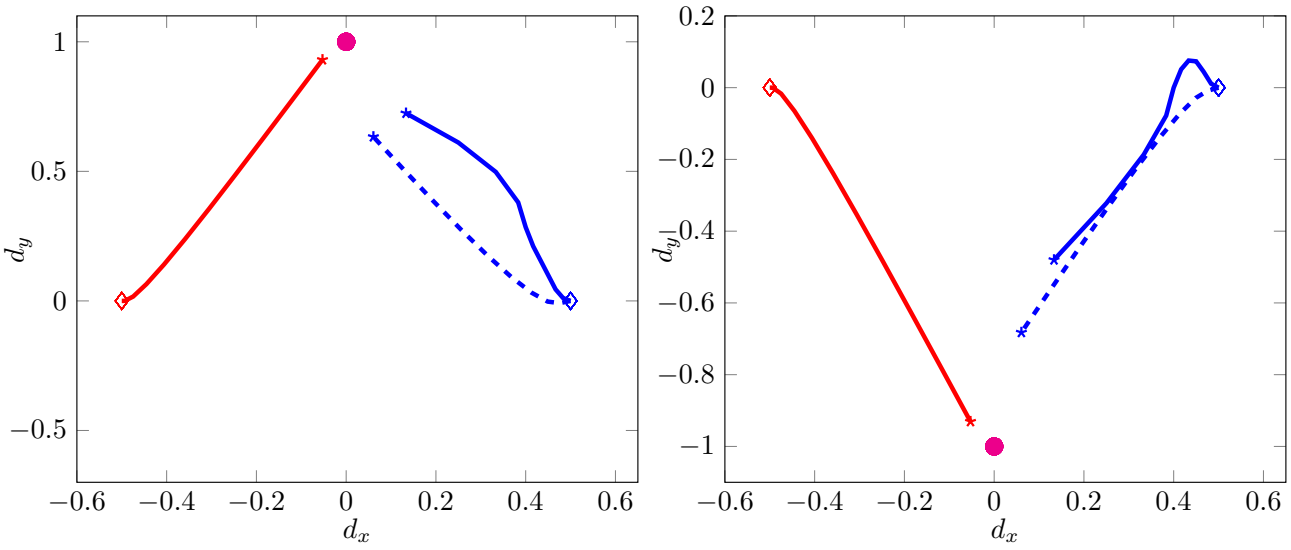

Figure 11: Comparison between the P2's ground truth strategy and the strategy synthesized from the dual value. P1's trajectory is shown red and P2's in blue. Solid trajectories correspond to that obtained when P2 plays its equilibrium strategy. Dotted trajectories represent the ground truth solution.