Two-Player Zero-Sum Differential Games with One-Sided Information

Mukesh Ghimire[1], **Zhe Xu**[1], **Yi Ren**[1]

[1]Arizona State University
{mghimire, xzhe1, yiren}@asu.edu

Abstract

Unlike Poker where the action space $\mathcal{A}$ is discrete, differential games in the physical world often have continuous action spaces not amenable to discrete abstraction, rendering no-regret algorithms with $\mathcal{O}(|\mathcal{A}|)$ complexity not scalable. To address this challenge within the scope of two-player zero-sum (2p0s) games with one-sided information, we show that (1) a computational complexity independent of $|\mathcal{A}|$ can be achieved by exploiting the convexification property of incomplete-information games and the Isaacs' condition that commonly holds for dynamical systems, and that (2) the computation of the two equilibrium strategies can be decoupled under one-sidedness of information. Leveraging these insights, we develop an algorithm that successfully approximates the optimal strategy in a homing game. Code available in github[1].

Introduction

The strength of game solvers has grown rapidly in the last decade, beating elite-level human players in Chess (Silver et al. 2017a), Go (Silver et al. 2017b), Poker (Brown and Sandholm 2019; Brown et al. 2020a), Diplomacy (FAIR† et al. 2022), Stratego (Perolat et al. 2022), among others with increasing complexity. These successes motivated recent interests in solving differential games in continuous time and space, e.g., competitive sports (Wang et al. 2024; Ghimire et al. 2024), where critical strategic plays should be executed precisely within the continuous action space and at specific moments in time (e.g., consider set piece scenarios in soccer). However, existing regret minimization algorithms, e.g., CFR+ (Tammelin 2014) and its variants (Burch, Johanson, and Bowling 2014; Moravčík et al. 2017; Brown et al. 2020a; Lanctot et al. 2009), and last-iterate online learning algorithms, e.g., variants of follow the regularized leader (FTRL) (McMahan 2011; Perolat et al. 2021) and of mirror descent (Sokota et al. 2022; Cen, Wei, and Chi 2021; Vieillard et al. 2020), are designed for discrete actions and have computational complexities increasing with respect to the size of the action space $\mathcal{A}$. Thus applying these algorithms to differential games would require either insightful action and time abstraction or enormous compute, neither of which are readily available.

[1]https://github.com/ghimiremukesh/cams/tree/workshop

As a step towards addressing this challenge, our study focuses on games with one-sided information, which represent a variety of attack-defence scenarios: Both players have common knowledge about the finite set of I possible payoff types and nature's distribution over these types p_0. At the beginning of the game, nature draws a type and informs Player 1 (P1) about the type but not P2. As the game progresses, the public belief about the chosen type is updated from p_0 based on the action sequence taken by P1 via the Bayes rule. P1's goal is to minimize the expected cost over p_0. This game is proved to have a value under Isaacs' condition (Cardaliaguet 2009). Due to the zero-sum nature, P1 may need to delay information release or manipulate P2's belief to take full advantage of information asymmetry, and P2's strategy is to optimize the worst-case payoff. Real-world examples of the game include man-on-man matchup in sports where the attacker has private information about which play is to be executed, and defense games where multiple potential targets are concerned.

The two differences between our game and commonly studied imperfect-information extensive-form games (IIEFGs) (Sandholm 2010; Perolat et al. 2022; FAIR† et al. 2022) are that: (1) IIEFGs often have belief spaces (e.g., belief about opponent's cards in Poker) larger than their abstracted action spaces (e.g., betting categories in Poker), and (2) information asymmetry in our games is only one-sided. This paper investigates the potential computational advantages from exploiting these differences via the following insights: (1) At any infostate, P1's (resp. P2's) behavioral strategy is I (resp. $I+1$)-atomic and convexifies the primal (resp. dual) value with respect to the public belief (Fig. 1). With this, we can reformulate the convex-concave minimax problem of size $\mathcal{O}(|\mathcal{A}|)$ at each infostate into a nonconvex-nonconcave problem of size $\mathcal{O}(I^2)$. When $I^2 \ll |\mathcal{A}|$, and in particular when $|\mathcal{A}| = \infty$, the latter becomes more efficient to solve in practice. (2) Due to the one-sidedness of information, the equilibrium behavioral strategies of P1 and

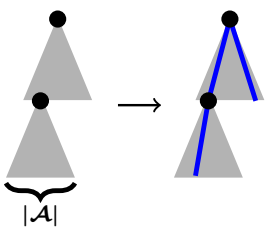

Figure 1: SOTA algorithms like CFR require expanding over entire action space (left), whereas our algorithm only requires expanding over at most I actions for P1 ($I+1$ for P2) at each decision node (right).

P2 can be solved separately through primal and dual formulations of the game, in each of which the opponent plays pure best responses. This decoupling avoids recurrent learning dynamics between the pair of strategies without regularization (Perolat et al. 2021).

To summarize, this work has two contributions: (1) familiarizing the broader AI community with the connections between computational game theory and differential game theory, and (2) providing the first algorithm with scalable convergence to the equilibrium of differential games with continuous action spaces and one-sided information.

Related Work

2p0s games with incomplete information. (Harsanyi 1967) introduced a Bayesian game framework to solve incomplete-information normal-form games by transforming the game into an imperfect-information one involving a chance mechanism. The seminal work of (Aumann, Maschler, and Stearns 1995) extended this idea to repeated games and established the connection between value convexification and belief manipulation. Within the same framework, Blackwell's approachability theorem (Blackwell 1956) naturally becomes the theoretical support for the optimal strategy of the uninformed player (P2). Building on top of (Aumann, Maschler, and Stearns 1995), (De Meyer 1996) introduced the concept of a dual game in which the behavioral strategy of the uninformed player becomes Markov. This concept later helped (Cardaliaguet 2007; Ghimire et al. 2024) to establish the value existence proof for 2p0s differential games with incomplete information. Unlike repeated games in which belief manipulation occurs only in the first round of the game, differential games may have multiple critical collocation points in the joint space of time, state, and public belief where belief manipulations are necessary to achieve Nash equilibrium, depending on the specifications of system dynamics, payoffs, and state constraints (Ghimire et al. 2024). For this reason, scalable value and strategy approximation for 2p0s differential games with incomplete information has not yet been achieved.

Imperfect information extensive-form games. IIEFGs represent the more general set of simultaneous or sequential multi-agent decision-making problems with finite horizons. Since any 2p0s IIEFG with finite action sets has a normal-form formulation, a unique Nash equilibrium always exists in the space of mixed strategies. Significant efforts have been taken to find equilibrium of large IIEFGs such as poker (Koller and Megiddo 1992; Billings et al. 2003; Gilpin and Sandholm 2006; Gilpin et al. 2007; Sandholm 2010; Brown and Sandholm 2019), with a converging set of algorithms that are no-regret, average- or last-iterate converging, and with sublinear or linear convergence rates (Zinkevich et al. 2007; Abernethy, Bartlett, and Hazan 2011; McMahan 2011; Tammelin 2014; Johanson et al. 2012; Lanctot et al. 2009; Brown et al. 2019, 2020b; Perolat et al. 2021; Sokota et al. 2022; Perolat et al. 2022; Schmid et al. 2023) (see summary in Tab. 1). These algorithms all have computational complexities increasing with $|\mathcal{A}|$, provided that the equilibrium behavioral strategy lies in the interior of the simplex $\Delta(|\mathcal{A}|)$. Critically, this assumption does not hold for differential games equipped with the Isaacs' condition, in which case the equilibrium strategy is mostly pure along the game tree, and is atomic on $\mathcal{A}$ when mixed.

Table 1: Solver computational complexity with respect to action space $\mathcal{A}$ and equilibrium error ε

Algorithm	Complexity
CFR variants (Zinkevich et al. 2007; Lanctot et al. 2009; Brown et al. 2019; Tammelin 2014; Johanson et al. 2012)	$\mathcal{O}\left(\lvert\mathcal{A}\rvert\varepsilon^{-2}\right)$ to ε-Nash
FTRL variants & MMD (McMahan 2011; Perolat et al. 2021; Sokota et al. 2022)	$\mathcal{O}\left(\frac{\ln(\lvert\mathcal{A}\rvert)}{\varepsilon}\ln\left(\frac{1}{\varepsilon}\right)\right)$ to ε-QRE

Descent-ascent algorithms for nonconvex-nonconcave minimax problems. Existing developments in IIEFGs focused on convex-concave minimax problems due to the bilinear form of the expected payoff through the conversion of games to their normal forms. This paper, on the other hand, investigates the nonconvex-nonconcave minimax problems to be solved at every infostate when actions are considered continuous. To this end, we use the doubly smoothed gradient descent ascent method (DS-GDA) which has a worst-case complexity of $\mathcal{O}(\varepsilon^{-4})$ (Zheng et al. 2023).

2p0s Differential Games w/ One-Sided Info.

Notations and preliminaries. We use $\Delta(I)$ as the simplex in $\mathbb{R}^I$, $[K] := \{1, ..., K\}$ for $K \in \mathbb{Z}_+$, $a[i]$ as the ith element of vector a, ∂V as the subgradient of function V, and $\langle \cdot, \cdot \rangle$ for vector product. Consider a time-invariant dynamical system that defines the evolution of the joint state $x \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ of P1 and P2 with control inputs $u \in \mathcal{U}$ and $v \in \mathcal{V}$, respectively:

$$\dot{x}(t) = f(x(t), u, v). \tag{1}$$

The game starts at $t_0 \in [0, T]$ from some initial state $x(t_0) = x_0$. The initial belief $p_0 \in \Delta(I)$ is set to nature's distribution. P1 of type i accumulates a running cost $l_i(u, v)$ during the game and receives a terminal cost $g_i(x(T))$, where $i \sim p_0$. The goal of P1 is to minimize the expected sum of the running and terminal costs, while P2 aims to maximize it. A behavioral strategy pair (η, ζ) is a Nash equilibrium (NE) of a zero-sum game if and only if

$$\inf_{\eta}\sup_{\zeta}\mathbb{E}_{\eta,\zeta,i\sim p_0}\int_0^T l_i dt + g_i = \sup_{\zeta}\inf_{\eta}\mathbb{E}_{\eta,\zeta,i\sim p_0}\int_0^T l_i dt + g_i, \tag{2}$$

and we call this common value the *value of the game*. A NE is called *pure* if the strategies (η, ζ) are deterministic, specifying a definite action for every decision point. It is called *mixed* if the strategies are probabilistic, involving randomization over action spaces. When information is one-sided, $\eta = \{\eta_i\}^I$ since P1 prepares one strategy for each possible game type. We introduce the following assumptions under which mixed NE exists for Eq. (2) (Cardaliaguet 2009):

1. $\mathcal{U} \subseteq \mathbb{R}^{d_u}$ and $\mathcal{V} \subseteq \mathbb{R}^{d_v}$ are compact and finite-dimensional sets.

2. $f : \mathcal{X} \times \mathcal{U} \times \mathcal{V} \to \mathcal{X}$ is bounded, continuous, and uniformly Lipschitz continuous with respect to x.
3. $g_i : \mathcal{X} \to \mathbb{R}$ and $l_i : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ are Lipschitz continuous and bounded.
4. Isaacs' condition holds for the Hamiltonian $H : \mathcal{X} \times \mathbb{R}^{d_x} \to \mathbb{R}$:

$$\begin{aligned} H(x,\xi) &:= \min_{u\in\mathcal{U}} \max_{v\in\mathcal{V}} f(x,u,v)^\top \xi - l_i(u,v) \\ &= \max_{v\in\mathcal{V}} \min_{u\in\mathcal{U}} f(x,u,v)^\top \xi - l_i(u,v). \end{aligned} \tag{3}$$

5. Both players have full knowledge about f, $\{g_i\}_{i=1}^I$, $\{l_i\}_{i=1}^I$, p_0, and the NE of the game. Control inputs and states are fully observable and we assume perfect recall.

Critically, the Isaacs' condition ensures that 2p0s differential games with complete information have pure NE.

Behavioral strategy of P1. A behavioral strategy prescribes distributions over the action space at every subgame (t, x, p). In order to determine the strategy, it is necessary to first characterize the value function. From Cardaliaguet (2009), we obtain the following backward induction to approximate the value given a sufficiently fine time-discretization $\tau \to 0^+$:

$$\begin{aligned} V_\tau(t,x,p) = \text{Vex}_p\Big(&\min_{u\in\mathcal{U}} \max_{v\in\mathcal{V}} V_\tau(t+\tau, x+\tau f(x,u,v), p) \\ &+ \tau \mathbb{E}\, l(u,v)\Big);\ V_\tau(T,x,p) = \sum_i p_i g_i(x) \end{aligned} \tag{4}$$

where Vex is the convexification operator. The behavioral strategy of P1 is computed as follows: P1 first finds $\lambda = [\lambda^1, \ldots, \lambda^I] \in \Delta(I)$ and $p^k \in \Delta(I)$ for $k \in [I]$ such that:

$$\begin{aligned} V_\tau(t,x,p) = \sum_k \lambda_k \Big(&\min_u \max_v \Big(V_\tau(t+\tau, x+\tau f(x,u,v), p) \\ &+ \mathbb{E}\, l(u,v)\Big)\Big);\ \sum_k \lambda_k p^k = p \end{aligned} \tag{5}$$

He then chooses u^k with $\Pr(u = u^k|i) = \lambda^k p^k[i]/p[i]$ if he is of type i and updates the belief to p^k. This is famously known as the splitting mechanism in repeated game, and is a consequence of the "Cav u" theorem (Aumann, Maschler, and Stearns 1995; De Meyer 1996).

Behavioral strategy of P2. For P2, the idea is to reformulate the game so that we can compute the value using P2's behavioral strategies and P1's pure best responses. This can be achieved by introducing the Fenchel conjugate V^* of V:

$$\begin{aligned} V^*(t_0,x_0,\hat{p}) &:= \max_p p\cdot\hat{p} - V(t_0,x_0,p) \\ &= \inf_\zeta \sup_\eta \max_{i\in\{1,\ldots,I\}} \Big\{ \hat{p}_i - \mathbb{E}_{\eta,\zeta}\Big[g_i\left(X_T^{t_0,x_0,\eta,\zeta}\right) \\ &\quad + \int_{t_0}^T l_i(\eta(s),\zeta(s))ds \Big]\Big\}, \end{aligned} \tag{6}$$

which describes a dual game with complete information in which P2's goal is to minimize some worst-case dual payoff. It is proved that P2's equilibrium in the dual game starting from some $(t_0, x_0, \hat{p})$ is also an equilibrium for the primal game if $\hat{p} \in \partial_p V(t_0, x_0, p)$ (Cardaliaguet 2007).

P2's strategy can be obtained through the dual game using a procedure similar to that of P1's: We first obtain the backward induction for the dual value:

$$\begin{aligned} V_\tau^*(T,x,\hat{p}) &= \max_i\{\hat{p}_i - g_i(x(T))\} \\ V_\tau^*(t,x,\hat{p}) &= \text{Vex}_{\hat{p}}\Big(\min_{v\in\mathcal{V}} \max_{u\in\mathcal{U}} V_\tau^*(t+\tau, x+\tau f(x,u,v), \\ &\qquad \hat{p} - \tau l(u,v))\Big), \end{aligned} \tag{7}$$

where, $l = [l_1, ..., l_I]^T$. Then at any $(t, x, \hat{p})$, P2 finds $\lambda = [\lambda^1, \ldots, \lambda^{I+1}]$ and $\hat{p}^k \in \mathbb{R}^I$ for $k \in [I+1]$ such that:

$$\begin{aligned} V_\tau^*(t,x,\hat{p}) = \sum_k^{I+1} \lambda_k \Big(&\min_v \max_u \Big(V_\tau^*(t+\tau, x+\tau f(x,u,v), \\ &\hat{p} - \tau l(u,v))\Big);\ \sum_k^{I+1} \lambda_k \hat{p}^k = \hat{p} \end{aligned} \tag{8}$$

where $l = [l_1, ..., l_I]^T$. P2's strategy is to compute the minimax solution v^k corresponding to $\hat{p}^k$ and chooses $v = v^k$ with probability λ^k.

Methods

Reformulation of the primal and dual games. To recap, at any (t, x), P1 computes actions u^k and their type-conditioned probabilities $\alpha_{ki} := \Pr(u = u^k|i)$ such that $\sum_{k=1}^I \alpha_{ki} = 1$ for $i \in [I]$. Then, $\lambda^k = \sum_{i=1}^I \alpha_{ki} p[i]$ and $p^k[i] = \alpha_{ki} p[i]/\lambda_k$ are both functions of α_{ki}. We can now reformulate (5) as follows:

$$\begin{aligned} &\min_{\{u^k\},\{\alpha_{ki}\}} \max_{\{v^k\}} \sum_{k=1}^I \lambda^k \left(V(t+\tau, x^k, p^k) + \tau \mathbb{E}_{i\sim p^k}[l_i(u^k, v^k)] \right) \\ &\text{s.t. } u^k \in \mathcal{U}, \quad x^k = \text{ODE}(x, \tau, u^k, v^k; f), \quad v^k \in \mathcal{V}, \quad \alpha_{ki} \in [0,1], \\ &\sum_{k=1}^I \alpha_{ki} = 1, \quad \lambda^k = \sum_{i=1}^I \alpha_{ki} p[i], \quad p^k[i] = \frac{\alpha_{ki} p[i]}{\lambda^k}, \quad \forall i,k \in [I]. \end{aligned} \tag{P_1}$$

P_1 is in general a nonconvex-nonconcave minimax problem of size $(\mathcal{O}(I(I+d_u)), \mathcal{O}(Id_v))$ that needs to be solved at all sampled infostates $(t, x, p) \in [0, T] \times \mathcal{X} \times \Delta(I)$. The resultant minimax objective is by definition the convexified value of the primal game.

P2, on the other hand, keeps track of the dual variable $\hat{p} \in \mathbb{R}^I$ instead of the public belief p during the dual game and solves the following problem at all sampled infostates $(t, x, \hat{p})$:

$$\begin{aligned} &\min_{\{v^k\},\{\lambda^k\},\{\hat{p}^k\}} \max_{\{u^k\}} \sum_{k=1}^{I+1} \lambda^k \left(V^*(t+\tau, x^k, \hat{p}^k - \tau l(u^k, v^k)) \right) \\ &\text{s.t. } u^k \in \mathcal{U}, \quad v^k \in \mathcal{V}, \quad x^k = \text{ODE}(x, \tau, u^k, v^k; f), \quad \lambda^k \in [0,1], \\ &\sum_{k=1}^{I+1} \lambda^k \hat{p}^k = \hat{p}, \quad \sum_{k=1}^{I+1} \lambda^k = 1, \quad k \in [I+1]. \end{aligned} \tag{P_2}$$

P_2 is in general nonconvex-nonconcave of size $(\mathcal{O}(I(I + d_v)), \mathcal{O}(Id_u))$.

Game solver. We propose a continuous-action mixed-strategy (CAMS) solver for 2p0s differential games with

one-sided information. Our algorithm performs Bellman backup through P_1 (resp. P_2) starting from the terminal condition in (4) (resp. (8)) at discretized time stamps $t \in \{T, T-\tau, ..., 0\}$ and (x, p) (resp. $(x, \hat{p})$) uniformly sampled in $\mathcal{X} \times \Delta(I)$ (resp. $\mathcal{X} \times \mathbb{R}^I$). Specifically, at any t, with a value approximation model $\hat{V}_{t+1} : \mathcal{X} \times \Delta(I) \to \mathbb{R}$, we solve P_1 using DS-GDA at N collocation points $(x, p) \in \mathcal{X} \times \Delta(I)$ and collect a dataset $\mathcal{D}_t := \{(x^{(i)}, p^{(i)}, \tilde{V}^{(i)})\}_{i=1}^N$ where $\tilde{V}$ is the numerical approximation of the convexified value at $(t, x^{(i)}, p^{(i)})$ for the minimax problem. Then we fit a model $\hat{V}_t(x, p)$ to $\mathcal{D}_t$ and go to $t - \tau$. Alg. 1 summarizes the solver for the primal game. The dual game solver is similarly defined.

Algorithm 1: Continuous Action Mixed Strategy Solver (CAMS)

Require: τ, $V(T, \cdot, \cdot)$, N, minimax solver $\mathbb{O}$
1: Initialize $\{\hat{V}_t\}_{t=0}^{T-\tau}$, $\mathcal{D} \leftarrow \emptyset$
2: $\mathcal{S} \leftarrow$ sample N states $(x, p) \in \mathcal{X} \times \Delta(I)$
3: **for** $t \in \{T - \tau, \ldots, 0\}$ **do**
4: **for** $(x, p) \in \mathcal{S}$ **do**
5: Append $\{(t, x, p), \mathbb{O}(t, x, p)\}$ to $\mathcal{D}$
6: **end for**
7: Fit $\hat{V}_t$ to $\mathcal{D}$
8: **end for**

Empirical Validation

We introduce Hexner's homing game (Hexner 1979) that has an analytical Nash equilibrium. We use variants of this game to compare CAMS with baselines (MMD, CFR+, and Deep-CFR) on solution quality and computational cost. As shown in Fig. 2, it is a two-player game, in which P1's goal is to get closer to the target Θ unknown to P2, while keeping P2 away and minimizing running costs. The cost to P1 is the expected value of the total cost:

$$J = \int_0^T (u^\top R_1 u - v^\top R_2 v)dt + [x_1(T) - \Theta]^\top K_1 [x_1(T) - \Theta] - [x_2(T) - \Theta]^\top K_2 [x_2(T) - \Theta], \tag{9}$$

where R_1 , $R_2 \succ 0$ and K_1, $K_2 \succeq 0$ are control and state-penalty matrices respectively. Due to the quadratic cost and decoupled dynamics, this game can be solved analytically as done in Hexner (1979).

Comparison on 1- and 4-stage Hexner's Games. We first use a normal-form Hexner's game with $\tau = T$ and a fixed initial state x_0 to demonstrate that IIEFG algorithms suffer from increasing costs along $|\mathcal{A}|$ while CAMS does not. We consider CFR+ (Tammelin 2014), MMD (Sokota et al. 2022), and a modified CFR-BR (Johanson et al. 2012) (dubbed CFR-BR-Primal, where we only focus on solving P1's optimal strategy) as baselines. Each player's state consists of 2D position and velocity. For baselines, we discretize the action sets $\mathcal{A}_1$ and $\mathcal{A}_2$ with sizes $\{16, 36, 64, 144\}$. All algorithms terminate when a threshold of NashConv (see Lanctot et al. (2017) for definition) is met. For conciseness, we only consider solving P1's strategy and thus use P1's δ in NashConv. We then use Deep-CFR as a baseline for a Hexner's game with 4 time-steps, where $T = 1$ and $\tau = 0.25$. DeepCFRs were run for 1000 CFR iterations (resp. 100) with 10 (resp. 5) traversals for $|\mathcal{A}| = 9$ (resp. 16). We compare the computational cost and the expected action error ε (and average action error at each time-step, $\bar{\varepsilon}_t$ for 4-stage game) from the ground-truth action of P1. Fig. 3 summarizes the comparisons. For the normal-form game, all baselines have complexities increasing with $\mathcal{A}$, while CAMS is invariant. In the 4-stage game, CAMS achieves significantly better strategies than DeepCFR, as visualized in Fig. 4.

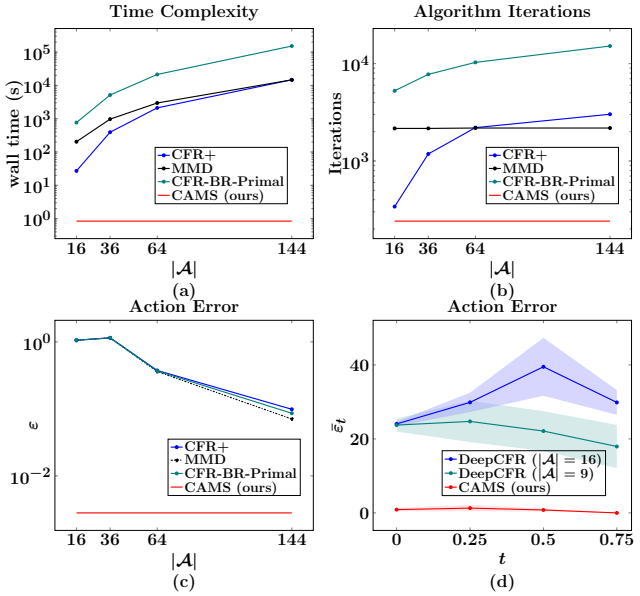

Figure 3: Comparisons b/w CAMS and baseline algorithms.

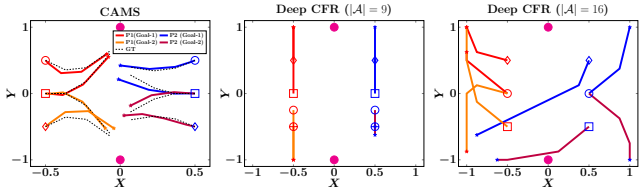

Figure 4: Trajectories using strategies from CAMS and DeepCFR. Markers indicate initial position.

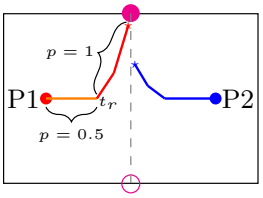

Figure 2: Hexner's game with a sample equilibrium trajectory. P1 starts to move to its target after t_r.

Conclusion

This work highlights the need for a scalable algorithm for solving incomplete-information differential games which are structurally similar to imperfect-information games such as poker. We demonstrated that SOTA IIEFG solvers are intractable when it comes to solving differential games. To the authors' best knowledge, this is the first method to provide tractable solution for incomplete-information differential games with continuous action spaces without problem-specific abstraction and discretization.

Acknowledgment

This work is partially supported by NSF CNS 2304863, CNS 2339774, IIS 2332476, and ONR N00014-23-1-2505.

References

Abernethy, J.; Bartlett, P. L.; and Hazan, E. 2011. Blackwell approachability and no-regret learning are equivalent. In *Proceedings of the 24th Annual Conference on Learning Theory*, 27–46. JMLR Workshop and Conference Proceedings.

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

Billings, D.; Burch, N.; Davidson, A.; Holte, R.; Schaeffer, J.; Schauenberg, T.; and Szafron, D. 2003. Approximating game-theoretic optimal strategies for full-scale poker. In *IJCAI*, volume 3, 661.

Blackwell, D. 1956. An analog of the minimax theorem for vector payoffs.

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

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

Brown, N.; Lerer, A.; Gross, S.; and Sandholm, T. 2019. Deep counterfactual regret minimization. In *International conference on machine learning*, 793–802. PMLR.

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

Burch, N.; Johanson, M.; and Bowling, M. 2014. Solving imperfect information games using decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.

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

Cardaliaguet, P. 2009. 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*, 1–18.

Cen, S.; Wei, Y.; and Chi, Y. 2021. Fast policy extragradient methods for competitive games with entropy regularization. *Advances in Neural Information Processing Systems*, 34: 27952–27964.

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

FAIR†, M. F. A. R. D. T.; Bakhtin, A.; Brown, N.; Dinan, E.; Farina, G.; Flaherty, C.; Fried, D.; Goff, A.; Gray, J.; Hu, H.; et al. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, 378(6624): 1067–1074.

Ghimire, M.; Zhang, L.; Xu, Z.; and Ren, Y. 2024. State-Constrained Zero-Sum Differential Games with One-Sided Information. In Salakhutdinov, R.; Kolter, Z.; Heller, K.; Weller, A.; Oliver, N.; Scarlett, J.; and Berkenkamp, F., eds., *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, 15512–15539. PMLR.

Gilpin, A.; Hoda, S.; Pena, J.; and Sandholm, T. 2007. Gradient-based algorithms for finding Nash equilibria in extensive form games. In *Internet and Network Economics: Third International Workshop, WINE 2007, San Diego, CA, USA, December 12-14, 2007. Proceedings 3*, 57–69. Springer.

Gilpin, A.; and Sandholm, T. 2006. Finding equilibria in large sequential games of imperfect information. In *Proceedings of the 7th ACM conference on Electronic commerce*, 160–169.

Harsanyi, J. C. 1967. Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model. *Management science*, 14(3): 159–182.

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

Johanson, M.; Bard, N.; Burch, N.; and Bowling, M. 2012. Finding optimal abstract strategies in extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, 1371–1379.

Koller, D.; and Megiddo, N. 1992. The complexity of two-person zero-sum games in extensive form. *Games and economic behavior*, 4(4): 528–552.

Lanctot, M.; Waugh, K.; Zinkevich, M.; and Bowling, M. 2009. Monte Carlo sampling for regret minimization in extensive games. *Advances in neural information processing systems*, 22.

Lanctot, M.; Zambaldi, V.; Gruslys, A.; Lazaridou, A.; Tuyls, K.; Pérolat, J.; Silver, D.; and Graepel, T. 2017. A unified game-theoretic approach to multiagent reinforcement learning. *Advances in neural information processing systems*, 30.

McMahan, B. 2011. Follow-the-Regularized-Leader and Mirror Descent: Equivalence Theorems and L1 Regularization. In Gordon, G.; Dunson, D.; and Dudík, M., eds., *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, 525–533. Fort Lauderdale, FL, USA: PMLR.

Moravčík, M.; Schmid, M.; Burch, N.; Lisỳ, V.; Morrill, D.; Bard, N.; Davis, T.; Waugh, K.; Johanson, M.; and Bowling, M. 2017. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337): 508–513.

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. 2022. Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science*, 378(6623): 990–996.

Perolat, J.; Munos, R.; Lespiau, J.-B.; Omidshafiei, S.; Rowland, M.; Ortega, P.; Burch, N.; Anthony, T.; Balduzzi, D.;

De Vylder, B.; et al. 2021. From poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *International Conference on Machine Learning*, 8525–8535. PMLR.

Sandholm, T. 2010. The state of solving large incomplete-information games, and application to poker. *Ai Magazine*, 31(4): 13–32.

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

Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Lai, M.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; et al. 2017a. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*.

Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. 2017b. Mastering the game of go without human knowledge. *nature*, 550(7676): 354–359.

Sokota, S.; D'Orazio, R.; Kolter, J. Z.; Loizou, N.; Lanctot, M.; Mitliagkas, I.; Brown, N.; and Kroer, C. 2022. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. *arXiv preprint arXiv:2206.05825*.

Tammelin, O. 2014. Solving large imperfect information games using CFR+. *arXiv preprint arXiv:1407.5042*.

Vieillard, N.; Kozuno, T.; Scherrer, B.; Pietquin, O.; Munos, R.; and Geist, M. 2020. Leverage the average: an analysis of kl regularization in reinforcement learning. *Advances in Neural Information Processing Systems*, 33: 12163–12174.

Wang, Z.; Veličković, P.; Hennes, D.; Tomašev, N.; Prince, L.; Kaisers, M.; Bachrach, Y.; Elie, R.; Wenliang, L. K.; Piccinini, F.; et al. 2024. TacticAI: an AI assistant for football tactics. *Nature communications*, 15(1): 1906.

Zheng, T.; Zhu, L.; So, A. M.-C.; Blanchet, J.; and Li, J. 2023. Universal gradient descent ascent method for nonconvex-nonconcave minimax optimization. *Advances in Neural Information Processing Systems*, 36: 54075–54110.

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