Sound Algorithms in Imperfect Information Games

Extended Abstract

Michal Šustr
Czech Technical University
DeepMind
michal.sustr@aic.fel.cvut.cz

Martin Schmid
DeepMind
mschmid@google.com

Matej Moravčík
DeepMind
moravcik@google.com

Neil Burch
DeepMind
burchn@google.com

Marc Lanctot
DeepMind
lanctot@google.com

Michael Bowling
DeepMind
bowlingm@google.com

ABSTRACT

Search has played a fundamental role in computer game research since the very beginning. And while online search has been commonly used in perfect information games such as Chess and Go, online search methods for imperfect information games have only been introduced relatively recently. This paper addresses the question of what is a sound online algorithm in an imperfect information setting of two-player zero-sum games? We argue that the fixed-strategy definitions of exploitability and epsilon-Nash equilibria are ill suited to measure the worst-case performance of an online algorithm. We thus formalize epsilon-soundness, a concept that connects the worst-case performance of an online algorithm to the performance of an epsilon-Nash equilibrium. Our definition of soundness and the consistency hierarchy finally provide appropriate tools to analyze online algorithms in repeated imperfect information games. We thus inspect some of the previous online algorithms in a new light, bringing new insights into their worst case performance guarantees.

KEYWORDS

Game Theory; Imperfect Information; Nash Equilibrium; Online Algorithm; Exploitability; Soundness

ACM Reference Format:
Michal Šustr, Martin Schmid, Matej Moravčík, Neil Burch, Marc Lanctot, and Michael Bowling. 2021. Sound Algorithms in Imperfect Information Games: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 INTRODUCTION

Online methods for approximating Nash equilibria in sequential imperfect information games appeared only in the last few years [2–4, 7, 8]. We thus investigate what it takes for an online algorithm to be sound in imperfect information settings. While it has been known that search with imperfect information is more challenging than with perfect information [5, 7], the problem is maybe more complex than previously thought. Online algorithms "live" in a fundamentally different setting, and they need to be evaluated appropriately.

We introduce a framework for evaluating the performance of an online algorithm. Within this framework, we introduce the definition of a sound and ϵ-sound algorithm. Like the exploitability of a strategy in the offline setting, the soundness of an algorithm is a measure of if its performance against a worst case adversary. Importantly, this notion collapses to the previous notion of exploitability when the algorithm follows a fixed strategy profile.

We then introduce a consistency framework, a hierarchy that allows us to formally state in what sense an online algorithm plays "consistently" with an ϵ-equilibrium. This allows to state multiple bounds on the soundness of the algorithm, based on the ϵ-equilibrium and the type of consistency. The stronger the consistency in our hierarchy, the stronger the bounds.

A complete version of this paper can be found on arXiv [9].

2 MOTIVATIONAL EXAMPLE

Consider a simple online algorithm for (r)ock (p)aper (s)cissors game that simply produces the sequence $(r, p, s, r, p, s, \ldots)$. While this algorithm is clearly highly exploitable, the worst-case performance of this algorithm can only be properly analysed through the repeated game settings as no offline policy captures the dynamics of this algorithm.

3 BACKGROUND

We present our results using the recent formalism of factored-observations games [6], where factored-observations game is a tuple $\mathcal{G} = \langle \mathcal{N}, \mathcal{W}, w^o, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O} \rangle$

3.1 Online Settings

The **repeated game** p consists of a finite sequence of k individual matches $m = (z_1, z_2, \ldots, z_k)$, where each **match** $z_i \in \mathcal{Z}$ is a sequence of world states and actions $z_i = (w_i^0, a_i^0, w_i^1, a_i^1 \ldots, a_i^{l_i - 1}, w_i^{l_i})$. An online algorithm Ω then simply maps an information state observed during a match to a strategy, while possibly using its internal algorithm state (Def. 1).

Given two players Ω_1, Ω_2, we use $P^k_{\Omega_1,\Omega_2}$ to denote the distribution over all the possible repeated games m of length k when these two players face each other. The average reward of m is $\mathcal{R}_n(m) = 1/k \sum_{i=1}^k u_n(z_i)$ and we denote $\mathbb{E}_{m \sim P^k_{\Omega_1,\Omega_2}} [\mathcal{R}_n(m)]$ to be the expected average reward when the players play k matches.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

Definition 1. *Online algorithm Ω is a function $\mathcal{S}_n \times \Theta \mapsto \Delta(\mathcal{A}_n(s_n)) \times \Theta$, that maps an information state $s_n \in \mathcal{S}_n$ to the strategy $\sigma_n(s_n) \in \Delta(\mathcal{A}_n(s_n))$, while possibly making use of algorithm's state $\boldsymbol{\theta} \in \Theta$ and updating it.*

4 SOUNDNESS OF ONLINE ALGORITHM

Exploitability / ϵ-equilibrium considers the expected utility of a fixed strategy against a worst-case adversary in a single match. We thus define a similar concept for the settings of an online algorithm in a repeated game: (k, ϵ)-soundness. Intuitively, an online algorithm is (k, ϵ)-sound if and only if it is guaranteed the same reward as if it followed a fixed ϵ-equilibrium after k matches.

Definition 2. *For an (k, ϵ)-sound online algorithm Ω, the expected average reward against any opponent is at least as good as if it followed an ϵ-Nash equilibrium fixed strategy σ for any number of matches k':*

$$\forall k' \geq k \ \forall \Omega_2 \ : \mathbb{E}_{m \sim P^{k'}_{\Omega,\Omega_2}} \left[\mathcal{R}(m)\right] \geq \mathbb{E}_{m \sim P^{k'}_{\sigma,\Omega_2}} \left[\mathcal{R}(m)\right]. \quad (1)$$

If algorithm Ω is (k, ϵ)-sound for $\forall k \geq 1$, we say the algorithm is ϵ-sound.

5 CONSISTENCY HIERARCHY

As the (k, ϵ)-soundness can often be infeasible to compute, we introduce the concept of consistency. This allows to formalize that an algorithm plays "consistently" with an ϵ-equilibrium, directly bounding the (k, ϵ)-soundness.

5.1 Local Consistency

Local consistency simply guarantees that every time we query the online algorithm, there is an ϵ-equilibrium that has the same local behavioral strategy $\sigma(s)$ for the queried state s.

Definition 3. *Algorithm Ω is locally consistent with ϵ-equilibria if*

$$\forall k \ \forall m = (z_1, z_2, \ldots, z_k) \ \ \forall h \in \mathcal{H}(z_k) \ \ \exists \sigma \in \mathcal{NE}^{\epsilon}_n$$
$$\textit{holds that} \ \ \Omega^{(z_1, \ldots, z_{k-1})}(s(h)) = \sigma(s(h)).$$

We then show that this link has different implication for perfect and imperfect information games.

Theorem 4. *An algorithm that is locally consistent with ϵ-equilibria might not be (k, ϵ)-sound.*

Theorem 5. *In perfect information games, an algorithm that is locally consistent with a subgame perfect equilibrium is sound.*

A particularly interesting example of an algorithm that is only locally consistent is Online Outcome Sampling [7] (OOS). In the full paper version [9] we show that although this algorithm was previously considered to be sound, it can produce highly exploitable strategies in imperfect information games.

5.2 Global Consistency

Local consistency guaranteed consistency only for individual states. Global consistency is a stronger criterion that guarantees consistency with some equilibria for all the states in combination.

Definition 6. *Algorithm Ω is globally consistent with ϵ-equilibria if*

$$\forall k \ \forall m = (z_1, z_2, \ldots, z_k) \ \ \exists \sigma \in \mathcal{NE}^{\epsilon}_n \ \ \forall h \in \mathcal{H}(z_i)$$
$$\textit{holds that} \ \ \Omega^{(z_1, \ldots, z_{i-1})}(s(h)) = \sigma(s(h)) \ \ \textit{for} \ \forall i \in \{1, \ldots, k\}.$$

However:

Theorem 7. *An algorithm that is globally consistent with ϵ-equilibria might not be ϵ-sound.*

But what if the algorithm keeps on playing the repeated game? While the global consistency with equilibria does not guarantee soundness, it guarantees that the expected average reward converges to the game value in the limit.

Theorem 8. *For an algorithm Ω that is globally consistent with ϵ-equilibria,*

$$\forall k \ \forall \Omega_2 \ : \mathbb{E}_{m \sim P^{k}_{\Omega,\Omega_2}} \left[\mathcal{R}(m)\right] \geq u^* - \epsilon - \frac{|\mathcal{S}_1|\Delta}{k}. \quad (2)$$

Corollary 9. *An algorithm Ω that is globally consistent with ϵ-equilibria is (k, ϵ)-sound as $k \to \infty$.*

5.3 Strong Global Consistency

Strong global consistency additionally guarantees that the game-play *itself* is generated consistently with an equilibrium; and as in global consistency, the partial strategies for this game-play also correspond to an ϵ-equilibrium. In other words, the online algorithm simply exactly follows a predefined equilibrium.

Definition 10. *Online algorithm Ω is strongly globally consistent with ϵ-equilibrium if*

$$\exists \sigma \in \mathcal{NE}^{\epsilon}_n \ \ \forall k \ \forall m = (z_1, z_2, \ldots, z_k) \ \ \forall h \in \mathcal{H}(z_k)$$
$$\textit{holds that} \ \ \Omega^{(z_1, \ldots, z_{k-1})}(s(h)) = \sigma(s(h)).$$

Theorem 11. *Online algorithm Ω that is strongly globally consistent with ϵ-equilibrium is ϵ-sound.*

Canonical examples of strongly globally consistent online algorithms are DeepStack/Libratus. In general, an algorithm that uses a notion of safe (continual) resolving is strongly globally consistent as it essentially re-solves some ϵ-equilibrium (albeit an unknown one) that it follows. Another, more recent example is ReBeL [1], as it essentially imitates CFR-D iterations in conjunction with a neural network at a test time.

6 REGRET AND (k, ϵ)-SOUNDNESS

Finally, the introduced notion of soundness has tight connection to the popular concept of regret when a regret minimizer is used as the online algorithm.

Corollary 12. *Any regret minimizer with a regret bound of R_k is $(k, \frac{R_k}{2k})$-sound.*

ACKNOWLEDGMENTS

Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140). This work was supported by Czech science foundation grant no. 18-27483Y.

REFERENCES

[1] Noam Brown, Anton Bakhtin, Adam Lerer, and Qucheng Gong. 2020. Combining Deep Reinforcement Learning and Search for Imperfect-Information Games. *arXiv preprint arXiv:2007.13544* (2020).

[2] Noam Brown and Tuomas Sandholm. 2017. Safe and nested subgame solving for imperfect-information games. In *Advances in Neural Information Processing Systems*. 689–699.

[3] Noam Brown and Tuomas Sandholm. 2018. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science* 359, 6374 (2018), 418–424.

[4] Noam Brown and Tuomas Sandholm. 2019. Superhuman AI for multiplayer poker. *Science* 365, 6456 (2019), 885–890.

[5] Ian Frank and David Basin. 1998. Search in games with incomplete information: A case study using bridge card play. *Artificial Intelligence* 100, 1-2 (1998), 87–123.

[6] Vojtěch Kovařík, Martin Schmid, Neil Burch, Michael Bowling, and Viliam Lisý. 2019. Rethinking formal models of partially observable multiagent decision making. *arXiv preprint arXiv:1906.11110* (2019).

[7] Viliam Lisý, Marc Lanctot, and Michael Bowling. 2015. Online Monte Carlo counterfactual regret minimization for search in imperfect information games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 27–36.

[8] Matej Moravcik, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. 2017. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* 356, 6337 (2017), 508–513.

[9] Michal Šustr, Martin Schmid, Matej Moravčík, Neil Burch, Marc Lanctot, and Michael Bowling. 2020. Sound search in imperfect information games. *arXiv preprint arXiv:2006.08740* (2020).