State of the Art on: Function Approximation for Adversarial Team Games

Luca Carminati, luca5.carminati@mail.polimi.it

1. Introduction to the research topic

Algorithmic Game Theory is a field of study at the intersection of Game Theory and Computer Science. Its objective is to design algorithms capable of operating strategic decisions in complex environments, optimizing a preference score over the possible outcomes. The complexity of the environment derives from uncertainties due to the presence of imperfect information and/or other agents optimizing their own scores.

In particular, the main research areas related to this topic are:

- *Game Theory* for the representation of the environment and the desired properties of proposed solutions;
- *Reinforcement Learning*, *Online Convex Optimization* and *Artificial Intelligence* for what concerns the design of autonomous agents;
- *Theoretical Computer Science* for the evaluation of hardness and complexity of the expected results.

Conferences and Journals

The high degree of multidisciplinarity characterizing the field implies the existance of many different relevant conferences and journals. Those have been selected according to an heterogeneous set of criteria, to ensure a complete evaluation. The main criteria used are:

- GGS[1] and Microsoft Academic rankings[2] to evaluate the quality of conferences;
- IP[3] and Microsoft Academic rankings[4] to evaluate the quality of journals;
- Acceptance rate;
- Number of influential articles and authors in the field;
- Opinion of researchers working in the field.

The most prestigious selected conferences, along with the research area they belong to, are:

- AAAI: Association for the Advancement of Artificial Intelligence - Artificial Intelligence;
- NIPS: Neural Information Processing Systems - Artificial Intelligence;
- IJCAI: International Joint Conference on Artificial Intelligence - Artificial Intelligence;
- AAMAS: Adaptive Agents and Multi-Agents Systems - Game Theory;
- ACM EC: Conference on Economics and Computation - Game Theory, Theoretical Computer Science

The most prestigious selected journals, along with the research area they belong to, are:

[1] The GII-GRIN-SCIE Conference Rating, 2018, http://gii-grin-scie-rating.scie.es/conferenceRating.jsf.
[2] https://academic.microsoft.com/conferences/.
[3] *Impact Factor*: the number of citations received in that year of articles published in a specific journal during the two preceding years,divided by the total number of publications in that journal during the two preceding years
[4] https://academic.microsoft.com/journals/.

- Artificial Intelligence - Artificial Intelligence;
- arXiv: Artificial Intelligence - Artificial Intelligence;
- Journal of Artificial Intelligence Research - Artificial Intelligence;
- Games and Economic Behavior - Game Theory;
- International Journal of Game Theory - Game Theory;
- Algorithmica - Theoretical Computer Science.

1.1. Preliminaries

The main needed concepts to understand the field of work are presented in the following. We will focus on needed game theoretical notions for game representation and online learning tools for learning agents.

1.1.1 Games, Strategies, Solution Concepts

For an informal introduction to the concepts of Games and Solutions, we refer to [33]:

"A game is a description of strategic interaction that includes the constraints on the actions that the players can take and the players' interests, but does not specify the actions that the players do take. A solution is a systematic description of the outcomes that may emerge in a family of games. Game theory suggests reasonable solutions for classes of games and examines their properties."

Many possibilities are available for representing games and strategies. In the following, we focus on *Normal-form* and *Extensive-form* games and thier corresponding strategies, since they are the most relevant to our analysis.

Intuitively, a normal-form game is a matrix-shaped model, characterized by a concurrent choice of actions for all the players, with payoffs defined by the joint tuple of chosen actions.

Definition 1 (Normal-form game)**.** *A normal-form game is a tuple* $<N, A, U>$ *where:*

- $N = \{1, 2, 3, ..., n\}$ *is a finite set of players;*
- $A = \times_{i \in N} A_i$ *is a set of actions profiles where* A_i *is the set of action for player i;*
- $U = (U_1, ..., U_n)$ *is the set of utility functions* $U_i : A \to \mathbb{R}$.

Definition 2 (Normal-form strategy)**.** *A normal-form strategy* s_i *for* $i \in N$ *is defined as a function* $s_i : A_i \to \Delta^{|A_i|}$, *where* $\Delta^n = \{p \in \mathbb{R}^n \mid \sum_j p_j = 1 \text{ and } \forall j \; p_j \in [0, 1]\}$ *is a probability distribution over n elements.*

Intuitively, an imperfect information game in extensive form models a tree-shaped game characterized by a sequential play of actions of each player leading to a terminal node providing a utility. Due to partial information, players may not have full knowledge of the exact past sequence of actions, and in such a case two or more nodes may be identical from a player point of view, thus ending in the same *information set*. Similarly, a behavioral strategy for a player is a mapping that associates each of his information set with a distribution over the possible actions, thus fully characterizing the behavior of a player in a game.

Definition 3 (Imperfect-Information Extensive-form Games [36])**.** *An imperfect-information game in extensive form is a tuple* $\Gamma = (N, A, V, L, \iota, \rho, \chi, U, H) where:$

- $N = \{1, 2, 3, ..., n\}$ *is a finite set of players;*
- A *is a set of actions;*
- V *is a set of nonterminal nodes (also called choice nodes);*

- *L is a set of terminal nodes, disjoint from V;*
- *$\rho : V \to 2^A$ is the action function, which assigns to each choice node a set of possible playable actions;*
- *$\iota : V \to N$ is the player function, which assigns to each non terminal node a player $i \in N$ who chooses an action at that node;*
- *$\chi : V \times A \to V \cup L$ is the successor function, which maps a choice node and an action to a new choice node or terminal node;*
- *$U = (U_1, ..., U_n)$ where $U_i : L \to \mathbb{R}$ is a real-valued utility function for player i on terminal nodes L;*
- *$H = (H_1, ..., H_n)$ is the set of information sets, in which each H_i is a partition of $V_i = \{v \in V | \iota(v) = i\}$ such that for any $x_1, x_2 \in V_i, \rho(x_1) = \rho(x_2)$ whenever there exists a $h \in H_i$ where $x_1 \in h$ and $x_2 \in h$.*

Definition 4 (Behavioral strategy)**.** *A behavioral strategy π_i for player i is a function $\pi_i(v_i) : V_i \to \Delta^{|\rho(v_i)|}$, where $\Delta^n = \{p \in \mathbb{R}^n \mid \sum_j p_j = 1 \text{ and } \forall j \ p_j \in [0, 1]\}$ is a probability distribution over n elements.*

It is to be noted that a classical result is the equivalence of representational power between extensive-form and normal-form games [27], given the *perfect recall condition* on extensive-form games, that implies no player can forget information acquired in earlier stages of the game at each information set. However, the normal-form representation may require an amount of space exponentially larger than the extensive-form to be stored in memory.

A *solution concept* is instead a characterization of the desired notion of equilibrium across strategies of the game. Depending on the context of the game, a solution concept may be preferred over another to characterize the properties of the desired *best* strategy. We'll informally present two such notions: *Nash Equilibria* and *Correlated Equilibria*.

- A *Nash Equilibrium* [32] is a tuple of normal form strategies, one for each player, such that no player has a positive gain in expected value by deviating to a different strategy, against the same fixed strategies for the opponents.
- A *Correlated Equilibrium* [2] is a probability distribution over joint actions of the players, such that each time a player is suggested to play an action, he/she cannot have any positive gain by deviating to another action, given the marginal probabilities of other players' actions

The intuition behind these two solution concepts is that Nash Equilibria describe a *stable* tuple of independent strategies for the players, in the sense that there is no incentive in deviating away, and each player's strategy is defined separately. Correlated equilibria describe a *stable joint* probability distribution over players' actions, where, similarly to NE, no player has incentives to deviate, but where the overall player choices are coordinated by sampling from a common distribution, thus correlating the plays of all the players.

Another needed definition is that of *Zero-sum games*. A game is said Zero-Sum if $\sum_{i \in N} U_i(x) = 0 \ \forall x \in Dom(U_i)$; otherwise, the game is said to be a *General-sum game*.

1.1.2 Online Learning Algorithms for Games

In the following, we present two of the most important algorithms used to iteratively learn a Nash Equilibrium in a game. They are the starting point upon which new algorithms, able to tackle a growing set of games with increasing efficiency, have been developed.

Fictitious Play

Fictitious Play (FP) is a classical game-theoretic algorithm that allows to learn a Nash Equilibrium in many types of games by repeatedly playing the game and adjusting the strategies of all players according to an update procedure. It was first introduced by Brown [5] in 1951 for normal form games.

The algorithm for normal-form 2 players games consists in each player counting the times the opponent chose each action. At time t, each player plays the best action hypothesizing that the opponent will play his average strategy observed until that moment.

If FP is applied by both the agents (i.e. *self-play setting*), convergence to a NE of the average strategy played is guaranteed in two-players zero-sum games.

Regret Matching

The Online Optimization framework allows one to model learning algorithms in the context of repeated games. In this context, the same game is repeatedly played by all players with a potentially different strategy at each timestep, and each player can observe the payoff of the played action and the payoff he/she would have received for each action they could have played. This allows the definition of a *regret* of not having played action *a* in hindsight. Regret-based algorithms leverage information about past regrets to output a strategy for each timestep.

Definition 5 (Cumulative External Regret)**.** *Given l_t, a past history of achieved payoff at each time t, and l_t^i, the payoff that was received if action i was played at time t, then the Cumulative External Regret of action i at time T is defined as* $R_T^i = \sum_{t=0}^{T} l_t^i - l_t$.

Definition 6 (Regret Matching)**.** *Given the cumulative external regret at time $T-1$, the Regret Matching Learning scheme prescribes to play strategy s_T, defined as:*

$$s_T(i) = \begin{cases} \frac{R_T^{i,+}}{\sum_j R_T^{j,+}}, & if \sum_j R_T^{j,+} > 0 \\ \frac{1}{|A|} & otherwise \end{cases} \qquad where\ R_T^{i,+} = \max(0, R_T^i)$$

Intuitively, Regret matching combined with external regret prescribes to play each action proportionally to the positive regret of not having played it always in the past.

Similarly to FP, if RM is applied by both the agents in a normal-form two-players zero-sum game, convergence to a NE of the average strategy played is guaranteed.

1.2. Research topic

In this section, we explain the main motivations and trends of our field of study, focusing on the opportunities presented by the current research status. A more detailed presentation of the topics introduced in this subsection can be found in Section 2.

During the recent fifteen years, the field of Algorithmic Game Theory game theory has undergone a huge transformation: from being able to solve only small instances of simple zero-sum two-players game, to large complex games such as Poker in case of *Libratus*[9] *Pluribus*[11]. These algorithms can then be used outside their original benchmarks to solve complex real games in the field of security, health, economics.

In particular, the introduction of more and more scalable solving techniques, directly working on the extensive-form representation and making use of sampling, substituted the traditional techniques based on Linear Programming. The refinement of these techniques, combined with newly developed abstraction techniques to reduce the size of games and with refinement techniques of the produced strategies, allowed the creation of Libratus and then Pluribus.

Another line of research has been that of using Function Approximation and Deep Learning techniques to produce automatic abstractions without prior knowledge and faster, more generalizing techniques to solve even bigger games. These approaches are strongly linked with the development of *Multiagent Reinforcement Learning (MARL)*, at the intersection of Algorithmic Game Theory and Reinforcement Learning.

In parallel, game theory community shifted part of its interest to other types of games, modeling more complex situations, with three or more players and general-sum payoffs. In these new environments, Nash Equilibria proved fragile in terms of robustness and with poor real-world rational behavior modeling. Thus important research efforts have been made to identify new possible solution concepts and to provide efficient algorithms able to find those solution concepts in reasonable time.

Our research work will be located in the latter research line. Our goal is that of providing a faster and thus more scalable algorithm for one of these equilibrium concepts, by leveraging Function Approximation techniques as happened in the zero-sum field.

2. Main related works

2.1. Classification of the main related works

The relevant literature can be divided according to three categories:

- Literature regarding the characterization of possible Solution Concepts;
- Literature on the equilibrium-finding algorithms;
- Literature regarding the definition of Benchmarks to test these algorithms.

Solution Concepts		Algorithms			Benchmarks
General Games	Team Games	Iterative algorithms for 2-player Zero sum	Approximated 2-player Zero sum	Iterative algorithms for coordinated solution concepts	[3] [19] [26] [29] [37]
[2] [17] [20] [30] [31] [35] [41] [42]	[4] [14]	[7] [8] [10] [22] [24] [25] [28] [34] [39] [40] [43] [44]	[6] [22] [23] [38]	[13] [14] [15] [16] [18] [19] [21]	

2.2. Brief description of the main related works

2.2.1 Solution Concepts

The presence of more than two players and/or a possibly general-sum setting brings several difficulties. Standard algorithms employed in two-players zero-sum games are not guaranteed to converge to a Nash Equilibrium, and Nash Equilibria are unable to model the behaviors of real rational players. Therefore some research lines explored how to define appropriate solution concepts:

- **Multi player, General sum setting**: in this setting, *Correlated Equilibria* [2] proved to be the leading solution concept, which has been then adapted and refined to match specific instances. Some important variants are *Coarse-Correlated Equilibria (CCE)* [31], *Extensive-Form (Coarse-)Correlated Equilibrium (EF(C)CE)* [20] [17] [41], *Agent-Form* [35] *(Coarse-)Correlated Equilibrium (AF(C)CE)* [41]. All these solution concepts are Correlated Equilibria in the sense that they suppose a mediating signal sampling from a distribution, however they differ in the type and timing of reccomendations offered to the players. This versatility of the concept of CE brings a powerful representational power, at the cost of fragmented theory and different algorithms to find them.
  Recently [30] developed a framework called *Hindsight Rationality*, able to propose a single unifying view over the relations across these concepts and the Regret Minimizations algorithms employed. The framework is based on the concepts of hindsight evaluation of different kind of regrets associated to the gain of different types of *deviations* during past plays.

- **Team games**: in this setting, a coalitional structure is present such that players in the same team have the exact same payoffs, while the game is zero-sum. The simplest case considers a two player team facing a single adversary. These type of games have been introduced by Von Stengel and Koller[42], characterizing the *Team Maxmin Equilibrium (TME)* for a normal form game. Basilico et al. in [4], provided an analysis of the solution concept finding from a computational complexity point of view, also proposing algorithms for efficiently finding them in normal form games. Celli and Gatti in [14] introduce the concept of TME on a single-team-single-adversary extensive-form game, considering also the possibility of different forms of communication. *TMEcom* allow the players to communicate before and during the game, whereas a *TMEcor* allows only pre-play coordination across team members.

2.2.2 Algorithms

Great research emphasis has been posed on solving two-players zero-sum extensive games with perfect recall:

- **Tabular Iterative algorithms for equilibrium computation**: The Linear Program proposed by Von Stengel in [40] was the first algorithm able to solve these type of games in a polynomial time with respect to the size of the original game tree. This linear program finds a Nash Equilibrium in an efficient tabular representation of the game called *sequence form*. The main iterative approximating algorithm used nowadays is called Counterfactual Regret Minimization and was proposed by Zinkevich et al. in [44]. Its main point is to map a Regret Matching instance on each information set, and demonstrate that the regret of the overall behavioral strategy played is upper bounded by the sum of the regrets on each information set. Since each RM minimizes the regret on each information set, then CFR converges to a Nash Equilibrium. A great number of variations to improve the efficiency of CFR have been next developed:
  - **Abstraction based variants**: *abstractions* are smaller versions of the original game, with the purpose of capturing the most essential information while allowing a great speedup in the equilibrium finding algorithm. Found strategies will then be mapped onto the original game where a refinement technique might be applied. Important works in this research line are:
    - *CFR-BR*[24] a CFR variant introduced to avoid *abstraction pathologies*[43];
    - a fast method for computation of best response in poker games[25]. This allowed a fast evaluation of goodness of a strategy even in a big game like poker;
    - *regret-based pruning techniques*[7], accelerating CFR iteration by not considering parts of the game tree at each iteration without loss of guarantees;
    - *realtime search techniques*[8] for refining strategies in a game during realtime play with guarantees on imrpoving the overall performance.
  - **Sampling variants**: CFR is an iterative algorithm based on full game tree traversals on the game tree. This may result extremely slow on big games, impacting the scalability of the overall algorithm. A crucial impovement has been made by Lanctot et al. in [28], which introduces Monte Carlo CFR (MCCFR), a version of CFR that allows sampling of actions, thus traversing only one subtree for each information set. Schmid et al. in [34] introduce Variance Reduced MCCFR, a refined MCCFR that uses baselines to estimate value at non explored nodes, thus reducing the variance due to sampling;
  - **Discounting Variants**: in [39], Tammelin found that high inertia due to accumulation of negative cumulative regret impacted negatively CFR's performances. Therefore CFR+ was introduced, avoiding large negative regrets by clipping cumulative regrets at zero. This approach has then been generalized in Discounted CFR[10], introducing a discounting weighting scheme over past iterations.

  Another more recent direction has been opened by Heinrich et al. in [22], through the introduction of Extensive-form Fictitious Play (XFP) an extensive-form algorithm equivalent to the original normal-form formulation of FP, yet more efficient.

- **Deep Learning approximation for equilibrium computation**: traditional variants of CFR/FP are severely constrained by the memory requirements due to the variables maintained at each information set. A possible solution consists in employing Machine Learning techniques to train regressors able to approximate the needed values at each information set, given samples of past experience. The main strength of these approaches is that no prior knowledge is needed for abstraction generation, at the cost of needing an initial exploration phase. In the following we list the main results:
  - **Fictitious Play approximations**: in [22], Heinrich et al. extended the XFP algorithm, to allow Reinforcement Learning and Supervised Learning to be employed for average strategy and best response computation. This new FP variant is called *Fictitious Self Play (FSP)*. A further extension considers Neural Networks as regressors, defining the variant called *Neural Fictitious Self Play (NFSP)*[23]
  - **CFR approximations**: CFR too has been extended with the use of function approximators. Brown et al. in [6] defined *Deep CFR*, a MCCFR variant which employs Neural Networks to estimate cumulative regrets at each information set. The approach has then been extended in DREAM[38], to allow the more aggressive sampling scheme of VR-MCCFR to be used with Neural approximations.
- **Iterative Algorithms for coordinated solution concepts**: in parallel to the development of solution concepts for the team and general sum setting, research addressed the design of no-regret learning schemes to approximate the equilibria defined by those.

  For the team games, the focus has been on the computation of a *TMEcor*, since they can provide arbitrarely larger payoffs to the team with respect to a TME without coordination. Celli and Gatti[14] proposed an *Hybrid Column Generation* algorithm, and Farina and Celli et al.[18] proposed the faster *Team Fictitious Self Play*, an adaptation of FP. Both the algorithms employ mixed integer programming techniques to solve the problem of determining a best response of the Team against the adversary, and LP programs to find a maxmin equilibrium. Another option is represented by the *Soft Team Actor Critic(STAC)*[13] and *Signal Mediated Strategies (SIMS)* [12], adapting Multi-agent Deep Reinforcement Learning techniques to the setting.

  For what concerns the general sum games, algorithms have been developed for each solution concept. Farina et al. in [19], provided a gradient-based algorithm to approximate an EFCE. Celli et al. in [15] and [16] provided suitable CFR formulation named *CFR-Jr* and *ICFR*, to converge to EFCCE and EFCE respectively. Both the works rely upon the use of *internal regret minimizers*, guaranteed to converge to a CE in normal-form games[21], and combined them with a CFR-like per-information-set decompostion.

2.2.3 Benchmarks

In the following, we revise the main environments used to comparatively test the various algorithms we presented.

Traditionally, Poker has represented the main benchmark for two player zero-sum games with imperfect information, thanks to its large dimensions, clear rule and payoff definition, large active professional community. In particular very simple version such as *Kuhn Poker*[26] and *Leduc Poker*[37] have been used as small scale testing scenes, whereas larger instances are *River Poker*, *Heads-up Limit Texas hold'em*, similar to full poker instances, with the only restriction of being one versus one, and with limited betting rounds. Full game instances correspond to *Heads-up No Limit Texas hold'em*, and *No Limit Texas hold'em* in the 6 player version.

In addition to these, other games have been proposed as benchamrk of multiplayer bargaining and coordination capabilities. Those games are often derived by known board games and the most notable examples are *Bridge*, *Diplomacy*, *Goofspiel*, *Hanabi*[3], *Sheriff*[19].

Another class of games is the *Patrolling Games* played on a Gridworld[1]. The idea is that to model real-world problems characterized by a spatial dimension in a discretized world.

To provide a shared,high-performance benchmarking setting, environments exposing many different games with the same interface have been proposed, the most notable of them being *OpenSpiel* [29] by Deepmind.

2.3. Discussion

To conclude our analysis of the state of the art, we present a summary of what have been the traditional topics of research and what are the open issues that need to be analyzed in the future.

Since the beginning of the studies in algorithmic game theory, researchers have been focusing mainly on finding Nash Equilibria two-players zero-sum games, and in the last years, such research process was able to reach outstanding results (e.g. Libratus and Pluribus).

On the other hand, the interest for games with general sum payoffs or with more players has grown only in the last years, leaving many open issues to be tackled. The most important issues regard the scalability and efficiency of exisiting algorithms, that constrain the size of games that can be treated. Another issue regards the robustness of such algorithms to arbitrary structures of games.

In particular, we foresee the opportunity of applying approximation techniques while preserving theoretical guarantee of convergence, as happened for zero-sum games in the last years.

References

[1] Alpern, S., Morton, A., and Papadaki, K. Patrolling games. *Operations Research 59*, 5 (2011), 1246–1257.

[2] Aumann, R. J. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics 1*, 1 (1974), 67–96.

[3] Bard, N., Foerster, J. N., Chandar, S., Burch, N., Lanctot, M., Song, H. F., Parisotto, E., Dumoulin, V., Moitra, S., Hughes, E., and et al. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence 280* (Mar 2020), 103216. arXiv: 1902.00506.

[4] Basilico, N., Celli, A., Nittis, G. D., and Gatti, N. Team-maxmin equilibrium: efficiency bounds and algorithms, 2016.

[5] Brown, G. Iterative solution of games by fictitious play. *Activity Analysis of Production and Allocation 13* (01 1951).

[6] Brown, N., Lerer, A., Gross, S., and Sandholm, T. Deep Counterfactual Regret Minimization. *arXiv:1811.00164 [cs]* (May 2019). arXiv: 1811.00164.

[7] Brown, N., and Sandholm, T. Regret-Based Pruning in Extensive-Form Games. 9.

[8] Brown, N., and Sandholm, T. Safe and nested subgame solving for imperfect-information games. *arXiv preprint arXiv:1705.02955* (2017).

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

[10] Brown, N., and Sandholm, T. Solving imperfect-information games via discounted regret minimization. In *AAAI* (2019).

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

[12] Cacciamani, F., Celli, A., Ciccone, M., and Gatti, N. Multi-agent coordination in adversarial environments through signal mediated strategies. *arXiv:2102.05026 [cs]* (Feb 2021). arXiv: 2102.05026.

[13] Celli, A., Ciccone, M., Bongo, R., and Gatti, N. Coordination in adversarial sequential team games via multi-agent deep reinforcement learning. *CoRR abs/1912.07712* (2019).

[14] Celli, A., and Gatti, N. Computational results for extensive-form adversarial team games. *arXiv:1711.06930 [cs]* (Nov 2017). arXiv: 1711.06930.

[15] Celli, A., Marchesi, A., Bianchi, T., and Gatti, N. Learning to Correlate in Multi-Player General-Sum Sequential Games. *arXiv:1910.06228 [cs]* (Oct. 2019). arXiv: 1910.06228.

[16] Celli, A., Marchesi, A., Farina, G., and Gatti, N. No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. *arXiv:2004.00603 [cs]* (June 2020). arXiv: 2004.00603.

[17] Farina, G., Bianchi, T., and Sandholm, T. Coarse Correlation in Extensive-Form Games. *arXiv:1908.09893 [cs]* (Aug. 2019). arXiv: 1908.09893.

[18] Farina, G., Gatti, N., Celli, A., and Sandholm, T. Ex ante coordination and collusion in zero-sum multi-player extensive-form games. 17.

[19] Farina, G., Ling, C. K., Fang, F., and Sandholm, T. Correlation in extensive-form games: Saddle-point formulation and benchmarks. *arXiv:1905.12564 [cs]* (Oct 2019). arXiv: 1905.12564.

[20] Forges, F., and Von Stengel, B. Computationally efficient coordination in game trees.

[21] Hart, S., and Mas-Colell, A. A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica 68*, 5 (2000), 1127–1150. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00153.

[22] Heinrich, J., and Lanctot, M. Fictitious Self-Play in Extensive-Form Games. 9.

[23] Heinrich, J., and Silver, D. Deep Reinforcement Learning from Self-Play in Imperfect-Information Games. *arXiv:1603.01121 [cs]* (June 2016). arXiv: 1603.01121.

[24] Johanson, M., Bard, N., Burch, N., and Bowling, M. Finding Optimal Abstract Strategies in Extensive-Form Games. 9.

[25] Johanson, M., Waugh, K., Bowling, M., and Zinkevich, M. Accelerating Best Response Calculation in Large Extensive Games. 8.

[26] Kuhn, H. W. A simplified two-person poker. *Contributions to the Theory of Games 1* (1950), 97–103.

[27] Kuhn, H. W. *11. Extensive Games and the Problem of Information*. Princeton University Press, 2016, pp. 193–216.

[28] Lanctot, M. *Monte Carlo Sampling and Regret Minimization for Equilibrium Computation and Decision-Making in Large Extensive Form Games.* Library and Archives Canada = Bibliothèque et Archives Canada, Ottawa, 2013. OCLC: 1019472898.

[29] Lanctot, M., Lockhart, E., Lespiau, J.-B., Zambaldi, V., Upadhyay, S., Pérolat, J., Srinivasan, S., Timbers, F., Tuyls, K., Omidshafiei, S., et al. Openspiel: A framework for reinforcement learning in games. *arXiv preprint arXiv:1908.09453* (2019).

[30] Morrill, D., D'Orazio, R., Sarfati, R., Lanctot, M., Wright, J. R., Greenwald, A., and Bowling, M. Hindsight and Sequential Rationality of Correlated Play. *arXiv:2012.05874 [cs]* (Dec. 2020). arXiv: 2012.05874.

[31] Moulin, H., and Vial, J.-P. Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory 7*, 3-4 (1978), 201–221.

[32] Nash, J. Non-cooperative games. *Annals of mathematics* (1951), 286–295.

[33] Osborne, M. J., and Rubinstein, A. *A course in game theory.* MIT Press, Cambridge, Mass, 1994.

[34] Schmid, M., Burch, N., Lanctot, M., Moravcik, M., Kadlec, R., and Bowling, M. Variance Reduction in Monte Carlo Counterfactual Regret Minimization (VR-MCCFR) for Extensive Form Games using Baselines. *arXiv:1809.03057 [cs]* (Sept. 2018). arXiv: 1809.03057.

[35] Selten, R. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*, 1 (Mar 1975), 25–55.

[36] Shoham, Y. Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. 532.

[37] Southey, F., Bowling, M., Larson, B., Piccione, C., Burch, N., Billings, D., and Rayner, C. Bayes' bluff: opponent modelling in poker. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence* (2005), pp. 550–558.

[38] Steinberger, E., Lerer, A., and Brown, N. DREAM: Deep Regret minimization with Advantage baselines and Model-free learning. *arXiv:2006.10410 [cs, stat]* (Nov. 2020). arXiv: 2006.10410.

[39] Tammelin, O. Solving large imperfect information games using cfr+.

[40] von Stengel, B. Efficient computation of behavior strategies. *Games and Economic Behavior 14*, 2 (1996), 220–246.

[41] von Stengel, B., and Forges, F. Extensive-Form Correlated Equilibrium: Definition and Computational Complexity. *Mathematics of Operations Research 33*, 4 (Nov. 2008), 1002–1022.

[42] von Stengel, B., and Koller, D. Team-maxmin equilibria. *Games and Economic Behavior 21*, 1 (1997), 309–321.

[43] Waugh, K., Schnizlein, D., Bowling, M., and Szafron, D. Abstraction pathologies in extensive games. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2* (Richland, SC, 2009), AAMAS '09, International Foundation for Autonomous Agents and Multiagent Systems, p. 781–788.

[44] Zinkevich, M., Bowling, M., Johanson, M., and Piccione, C. Regret Minimization in Games with Incomplete Information. 14.