Learning to Correlate in Multi-Player General-Sum Sequential Games*

Andrea Celli[1,†], Alberto Marchesi[1,†], Tommaso Bianchi[2], Nicola Gatti[1]
Politecnico di Milano, Milan, Italy
[1]{name.surname}@polimi.it [2]tommaso4.bianchi@mail.polimi.it

Abstract

In the context of multi-player, general-sum games, there is an increasing interest in solution concepts modeling some form of communication among players, since they can lead to socially better outcomes with respect to Nash equilibria, and may be reached through learning dynamics in a decentralized fashion. In this paper, we focus on *coarse correlated equilibria* (CCEs) in sequential games. Simple arguments show that CFR—working with behavioral strategies—may not converge to a CCE. First, we devise a simple variant (CFR-S) which provably converges to the set of CCEs, but may be empirically inefficient. Then, we design a variant of the CFR algorithm (called CFR-Jr) which approaches the set of CCEs with a regret bound sub-linear in the size of the game, and is shown to be dramatically faster than CFR-S and the state-of-the-art algorithms to compute CCEs.

1 Introduction

A number of recent studies explore relaxations of the classical notion of equilibrium (*i.e.*, the *Nash equilibrium* (NE) [22]), allowing to model communication among the players [2, 12, 26]. Communication naturally brings about the possibility of playing correlated strategies. These are customarily modeled through the *correlated equilibrium* (CE) [1]. A popular variation of the CE is the *coarse correlated equilibrium* (CCE), which only prevents deviations before knowing the recommendation [21]. In sequential games, CEs and CCEs are well-suited for scenarios where the players have limited communication capabilities and can only communicate before the game starts, such as, *e.g.*, military settings where field units have no time or means of communicating during a battle, collusion in auctions where communication is illegal during bidding, and, in general, any setting with costly communication channels or blocking environments. CCEs present a number of appealing properties. A CCE can be reached through simple (no-regret) learning dynamics in a decentralized fashion [14, 16], and, in several classes of games (such as, *e.g.*, normal-form and succinct games [23, 18]), it can be computed exactly in poly-time. Furthermore, an optimal (*i.e.*, social-welfare-maximizing) CCE may provide arbitrarily larger welfare than an optimal CE or NE [10]. However, the problem of computing CCEs has been addressed only for some specific games with particular structures [2, 17]. In this work, we study how to compute CCEs in the general class of games which are sequential, general-sum, and multi-player.

In sequential games, it is known that, when there are two players without chance moves, an optimal CCE can be computed in polynomial time [10]. Celli et al. [10] also provide an algorithm (with no polynomiality guarantees) to compute solutions in multi-player games, using a column-generation procedure with a MILP pricing oracle. As for computing approximate CCEs, in the normal-form setting, any *Hannan consistent* regret-minimizing procedure for simplex decision spaces may be employed to approach the set of CCEs [4, 11]—the most common of such techniques is *regret*

[*] The extended version of the paper has been accepted to NeurIPS-19.

[†] Equal contribution.

Abstract for: Smooth Games Optimization and Machine Learning Workshop (NeurIPS 2019), Vancouver, Canada.

matching (RM) [3, 16]. However, approaching the set of CCEs in sequential games is more demanding. One could represent the sequential game with its equivalent normal form and apply RM to it. However, this would result in a guarantee on the cumulative regret which would be exponential in the size of the game tree (see Section 2). Thus, reaching a good approximation of a CCE could require an exponential number of iterations. The problem of designing learning algorithms avoiding the construction of the normal form has been successfully addressed in sequential games for the two-player, zero-sum setting. This is done by decomposing the overall regret locally at the information sets of the game [13]. The most widely adopted of such approaches are *counterfactual regret minimization* (CFR) [33] and CFR+ [30, 29], which originated variants such as [7, 9]. These techniques were the key for many recent remarkable results [5, 6, 8, 20]. However, these algorithms work with players' behavioral strategies rather than with correlated strategies, and, thus, they are not guaranteed to approach CCEs in general-sum games, even with two players. The only known theoretical guarantee of CFR when applied to multi-player, general-sum games is that it excludes dominated actions [15]. Some works also attempt to apply CFR to multi-player, zero-sum games, see, *e.g.*, [24].

2 Preliminaries

We focus on *extensive-form games* (EFGs) with imperfect information and perfect recall. We denote the set of players as $\mathcal{P} \cup \{c\}$, where c is the *Nature* (*chance*) player (representing exogenous stochasticity). H is the set of nodes of the game tree, and a node $h \in H$ is identified by the ordered sequence of actions from the root to the node. $Z \subseteq H$ is the set of terminal nodes. For every $h \in H \setminus Z$, we let $P(h)$ be the unique player who acts at h and $A(h)$ be the set of actions available at h. For each player $i \in \mathcal{P}$, $u_i : Z \to \mathbb{R}$ is the payoff function. We denote by Δ the maximum range of payoffs in the game. We represent imperfect information using *information sets* (from here on, infosets). Any infoset I belongs to a unique player i, and it groups nodes which are indistinguishable for that player, *i.e.*, $A(h) = A(h') \; \forall h, h' \in I$. $\mathcal{I}_i$ denotes the set of all player i's infosets. We denote by $A(I)$ the set of actions available at I. We denote with π_i a *behavioral strategy* of player i, which is a vector defining a probability distribution at each player i's infoset. Given π_i, we let $\pi_{i,I}$ be the (sub)vector representing the probability distribution at $I \in \mathcal{I}_i$, with $\pi_{i,I,a}$ denoting the probability of choosing action $a \in A(I)$. An EFG has an equivalent tabular (*normal-form*) representation. A *normal-form plan* for player i is a vector $\sigma_i \in \Sigma_i = \bigtimes_{I \in \mathcal{I}_i} A(I)$ which specifies an action for each player i's infoset. Then, an EFG is described through a $|\mathcal{P}|$-dimensional matrix specifying a utility for each player at each *joint normal-form plan* $\sigma \in \Sigma = \bigtimes_{i \in \mathcal{P}} \Sigma_i$. The expected payoff of player i, when she plays $\sigma_i \in \Sigma_i$ and the opponents play normal-form plans in $\sigma_{-i} \in \Sigma_{-i} = \bigtimes_{j \neq i \in \mathcal{P}} \Sigma_j$, is denoted, with an overload of notation, by $u_i(\sigma_i, \sigma_{-i})$. Finally, a *normal-form strategy* x_i is a probability distribution over Σ_i. We denote by $\mathcal{X}_i$ the set of the normal-form strategies of player i. Moreover, $\mathcal{X}$ denotes the set of joint probability distributions defined over Σ. We also let ρ^{π_i} be a vector in which each component $\rho^{\pi_i}_z$ is the probability of reaching the terminal node $z \in Z$, given that player i adopts the behavioral strategy π_i and the other players play so as to reach z. Similarly, given a normal-form plan $\sigma_i \in \Sigma_i$, we define the vector ρ^{σ_i}. Finally, $Z(\sigma_i) \subseteq Z$ is the subset of terminal nodes which are (potentially) reachable if player i plays according to $\sigma_i \in \Sigma_i$.

The classical notion of CE by Aumann [1] models correlation via the introduction of an external mediator who, before the play, draws $\sigma^* \in \Sigma$ according to a publicly known $x^* \in \mathcal{X}$, and privately communicates each *recommendation* σ^*_i to the corresponding player. After observing their recommended plan, each player decides whether to follow it or not. A CCE is a relaxation of the CE, defined by Moulin and Vial [21], which enforces protection against deviations which are independent from the sampled joint normal-form plan. Formally, a CCE of an EFG is a probability distribution $x^* \in \mathcal{X}$ such that, for every $i \in \mathcal{P}$, and $\sigma'_i \in \Sigma_i$, it holds: $\sum_{\sigma_i \in \Sigma_i} \sum_{\sigma_{-i} \in \Sigma_{-i}} x^*(\sigma_i, \sigma_{-i}) \left(u_i(\sigma_i, \sigma_{-i}) - u_i(\sigma'_i, \sigma_{-i}) \right) \geq 0$. CCEs differ from CEs in that a CCE only requires that following the suggested plan is a best response in expectation, before the recommended plan is actually revealed. An NE [22] is a CCE which can be written as a product of players' normal-form strategies $x^*_i \in \mathcal{X}_i$. In conclusion, an ε-CCE is a relaxation of a CCE in which every player has an incentive to deviate less than or equal to ε (the same holds for ε-CE and ε-NE).

In the *online convex optimization* framework [32], each player i plays repeatedly against an unknown environment by making a series of decisions $x^1_i, x^2_i, \ldots, x^t_i$. In the basic setting, the decision space of player i is the whole normal-form strategy space $\mathcal{X}_i$. At iteration t, after selecting x^t_i, player i observes a utility $u^t_i(x^t_i)$. The *cumulative external regret* of player i up to iteration T is defined as $R^T_i = \max_{\hat{x}_i \in \mathcal{X}_i} \sum_{t=1}^T u^t_i(\hat{x}_i) - \sum_{t=1}^T u^t_i(x^t_i)$. A *regret minimizer* is a function providing the next player i's strategy x_i^{t+1} on the basis of the past history of play and the observed utilities up to iteration t. In an EFG, the regret can be defined at each infoset. After T iterations, the cumulative regret for not having selected action $a \in A(I)$ at $I \in \mathcal{I}_i$ (denoted by $R_I^T(a)$) is the cumulative difference in utility that player i would have experienced by selecting a at I instead of following the behavioral strategy π_i^t at each iteration t up to T. Then, the regret for player i at infoset $I \in \mathcal{I}_i$ is defined as $R_I^T = \max_{a \in A(I)} R_I^T(a)$. *Regret matching* (RM) [16] is the most widely adopted regret-minimizing scheme when the decision space is $\mathcal{X}_i$ (*e.g.*, in normal-form games). In the context of EFGs, RM is usually applied locally at each infoset, where the player selects a distribution over available actions proportionally to their positive regret. Playing according to RM at each iteration guarantees, on iteration T, $R_I^T \leq \Delta\sqrt{|A(I)|}/\sqrt{T}$ [11]. CFR [33] is an anytime algorithm to compute ε-NEs in two-player, zero-sum EFGs. CFR minimizes the external regret R_i^T by employing RM locally at each infoset. In two-player, zero-sum games, if both players have cumulative regrets such that $\frac{1}{T}R_i^T \leq \varepsilon$, then their average behavioral strategies are a 2ε-NE [31].

3 CFR in multi-player general-sum sequential games

It is well known that, when players follow strategies recommended by a regret minimizer, the *empirical frequency of play* approaches the set of CCEs [11]. Suppose that, at time t, the players play a joint normal-form plan $\sigma^t \in \Sigma$ drawn according to their current strategies. Then, the empirical frequency of play after T iterations is defined as the joint probability distribution $\bar{x}^T \in \mathcal{X}$ such that $\bar{x}^T(\sigma) := \frac{|t \leq T: \sigma^t = \sigma|}{T}$ for every $\sigma \in \Sigma$. However, vanilla CFR and its most popular variations [30, 7] do not keep track of the empirical frequency of play, as they only keep track of the players' average behavioral strategies. This ensures that the strategies are compactly represented, but it is not sufficient to recover a CCE in multi-player, general-sum games. Indeed, even in normal-form games, if the players play according to some regret-minimizing strategies, then the product distribution $x \in \mathcal{X}$ resulting from players' (marginal) average strategies may not converge to a CCE. In order to see this, consider the two-player game depicted on the left in Figure 1. At iteration t, let players' strategies x_1^t, x_2^t be such that $x_1^t(\sigma_L) = x_2^t(\sigma_L) = (t+1) \mod 2$. Clearly, $u_1^t(x^t) = u_2^t(x^t) = 1$ for any t. For both players, at iteration t, the regret of not having played σ_L is 0, while the regret of σ_R is -1 if and only if t is even, otherwise it is 0. As a result, after T iterations, $R_1^T = R_2^T = 0$, and, thus, x_1^t and x_2^t minimize the cumulative external regret. Players' average strategies converge to $(\frac{1}{2}, \frac{1}{2})$ as $T \to \infty$. However, $x \in \mathcal{X}$ s.t., $\forall \sigma \in \Sigma$, $x(\sigma) = \frac{1}{4}$ is not a CCE of the game. This example employs handpicked regret-minimizing strategies, but similar behaviors can be easily found when applying common regret minimizers. As an illustrative case, Figure 1 shows, on the right, that, even with a simple variation of the Shapley game the outer product of the average strategies $\bar{x}_1^T \otimes \bar{x}_2^T$ obtained via RM does not converge to a CCE as $T \to \infty$.

	σ_L	σ_R
σ_L	$1, 1$	$1, 0$
σ_R	$0, 1$	$1, 1$

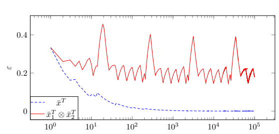

Figure 1: *Left*: Game where $\bar{x}_1^T \otimes \bar{x}_2^T$ does not converge to a CCE. *Right*: Approximation attained by $\bar{x}^T$ and $\bar{x}_1^T \otimes \bar{x}_2^T$.

The previous examples suggest a simple variation of CFR guaranteeing approachability to the set of CCEs even in multi-player, general-sum EFGs, which we call CFR *with sampling* (CFR-S). The key ingredient of CFR-S is a way to keep track of the empirical frequency of play. At each iteration t and for each player i, CFR-S draws a normal-form plan σ_i^t according to the current strategy π_i^t and updates the regrets using utilities computed according to the sampled plans σ_{-i}^t (rather than current behavioral strategies). Joint normal-form plans $\sigma^t := (\sigma_i^t, \sigma_{-i}^t)$ can be easily stored to compute the empirical frequency of play. It is possible to show that the empirical frequency of play $\bar{x}^T$ obtained with CFR-S converges to a CCE almost surely, for $T \to \infty$. We show that it is possible to achieve better perfomances via a smarter reconstruction technique that keeps CFR deterministic, avoiding any sampling step.

4 CFR with joint distribution reconstruction (CFR-Jr)

We design a new method—called *CFR with joint distribution reconstruction* (CFR-Jr)—to enhance CFR so as to approach the set of CCEs in multi-player, general-sum EFGs. Differently from the naive CFR-S algorithm, CFR-Jr does not sample normal-form plans, thus avoiding any stochasticity. The main idea behind CFR-Jr is to keep track of the average joint probability distribution $\bar{x}^T \in \mathcal{X}$ arising from the regret-minimizing strategies built with CFR. Formally, $\bar{x}^T = \frac{1}{T}\sum_{t=1}^{T} x^t$, where $x^t \in \mathcal{X}$ is the joint probability distribution defined as the product of the players' normal-form strategies

Game	Tree size #infosets	CFR-S $\alpha = 0.05$	CFR-S $\alpha = 0.005$	CFR-S $\alpha = 0.0005$	CFR-S $\mathrm{sw_{APX}}/\mathrm{sw_{OPT}}$	CFR-Jr $\alpha = 0.05$	CFR-Jr $\alpha = 0.005$	CFR-Jr $\alpha = 0.0005$	CFR-Jr $\mathrm{sw_{APX}}/\mathrm{sw_{OPT}}$	CG
K3-6	72	1.41s	9h15m	> 24h	-	1.03s	13.41s	11m21s	-	3h47m
K3-7	84	4.22s	17h11m	> 24h	-	2.35s	14.33s	51m27s	-	14h37m
K3-10	120	22.69s	> 24h	> 24h	-	7.21s	72.78s	4h11m	-	> 24h
L3-4	1200	10m33s	> 24h	> 24h	-	1m15s	6h10s	> 24h	-	> 24h
L3-6	2664	2h5m	> 24h	> 24h	-	2m40s	11h19m	> 24h	-	> 24h
L3-8	4704	13h55m	> 24h	> 24h	-	20m22s	> 24h	> 24h	-	> 24h
G3-4-A*	98508	1h33m	> 24h	> 24h	0.996	1h3m	4h13m	> 24h	0.999	> 24h
G3-4-DA*	98508	1h13m	> 24h	> 24h	0.987	12m18s	1h50m	> 24h	1.000	> 24h
G3-4-DH*	98508	47m33s	19h40m	> 24h	0.886	16m38s	4h8m	15h27m	1.000	> 24h
G3-4-AL*	98508	32m34s	15h32m	17h30m	0.692	1h21m	5h2s	> 24h	0.730	> 24h

Table 1: Run time and the social welfare of CFR-S, CFR-Jr (for various levels of accuracy α), and CG. General-sum instances are marked with $\star$. Results of CFR-S are averaged over 50 runs.

at iteration t. At each t, CFR-Jr computes π_i^t with CFR's update rules, and then constructs, via Algorithm 1, a strategy $x_i^t \in \mathcal{X}_i$ which is realization equivalent (*i.e.*, it induces the same probability distribution on the terminal nodes) to π_i^t. We do this efficiently by directly working on the game tree, without resorting to the normal-form representation. Strategies x_i^t are then employed to compute x^t.

Algorithm 1 `Reconstruct` x_i `from` π_i

```
1: function NF-STRATEGY-RECONSTRUCTION(π_i)
2:     X ← ∅                          ▷ X is a dictionary defining x_i
3:     ω_z ← ρ_z^{π_i}   ∀z ∈ Z
4:     while ω > 0 do
5:         σ̄_i ← arg max_{σ_i ∈ Σ_i} min_{z ∈ Z(σ_i)} ω_z
6:         ω̄ ← min_{z ∈ Z(σ̄_i)} ω_i(z)
7:         X ← X ∪ (σ̄_i, ω̄)
8:         ω ← ω − ω̄ ρ^{σ̄_i}
       return x_i built from the pairs in X
```

Algorithm 1 maintains a vector ω which is initialized with the probabilities of reaching the terminal nodes by playing π_i (Line 3), and it works by iteratively assigning probability to normal-form plans so as to induce the same distribution of ω over Z. At each iteration, the algorithm must pick $\bar{\sigma}_i \in \Sigma_i$ which maximizes the minimum (remaining) probability ω_z over $z \in Z(\bar{\sigma}_i)$ (Line 5). Then, the probabilities ω_z for $z \in Z(\bar{\sigma}_i)$ are decreased by the minimum (remaining) probability $\bar{\omega}$ corresponding to $\bar{\sigma}_i$, and $\bar{\sigma}_i$ is assigned probability $\bar{\omega}$ in x_i. The algorithm terminates when the vector ω is zeroed, returning a normal-form strategy x_i realization equivalent to π_i. It is possible to show that Algorithm 1 outputs a normal-form strategy $x_i \in \mathcal{X}_i$ realization equivalent to a given behavioral strategy π_i, and it runs in time $O(|Z|^2)$. Moreover, x_i has support size at most $|Z|$. Finally, it is possible to show that, if $\frac{1}{T}R_i^T \leq \varepsilon$ for each player $i \in \mathcal{P}$, then $\bar{x}^T$ obtained with CFR-Jr is an ε-CCE.

5 Experimental evaluation and discussion

We experimentally evaluate CFR-Jr, comparing its performance with that of CFR-S, CFR, and the state-of-the-art algorithm for computing optimal CCEs (denoted by CG) [10]. A direct application of RM on the normal form is not feasible, as $|\Sigma| > 10^{20}$ even for the smallest instances. We conduct experiments on parametric instances of 3-player Kuhn poker games [19], 2/3-player Leduc hold'em poker games [28], and 3-player Goofspiel games [25]. Each instance is identified by parameters p and r, which denote, respectively, the number of players and the number of ranks in the deck of cards. For example, a 3-player Kuhn game with rank 4 is denoted by Kuhn3-4, or K3-4. We use different tie-breaking rules for the Goofspiel instances. We evaluate the run time required by the algorithms to find an approximate CCE. The results are provided in Table 1, which reports the run time needed by CFR-S, CFR-Jr to achieve solutions with different levels of accuracy, and the time needed by CG for reaching an equilibrium. The accuracy α of the ε-CCEs reached is defined as $\alpha = \frac{\varepsilon}{\Delta}$. CFR-Jr consistently outperforms both CFR-S and CG, being orders of magnitude faster.

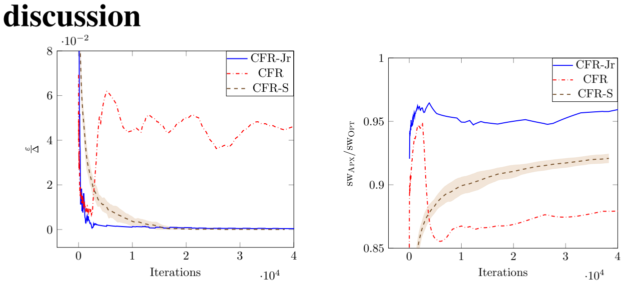

Figure 2: Convergence rate (*left*) and social welfare (*right*).

Figure 2, on the left, shows the performance of CFR-Jr, CFR-S (mean plus/minus standard deviation), and CFR over G2-4-DA in terms of ε/Δ approximation. CFR performs dramatically worse than CFR-S and CFR-Jr. Table 1 shows, for the general-sum games, the social welfare approximation ratio between the social welfare of the solutions returned by the algorithms ($\mathrm{sw_{APX}}$) and the optimal social welfare ($\mathrm{sw_{OPT}}$). The social welfare guaranteed by CFR-Jr is always nearly optimal, which makes it a good heuristic to compute optimal CCEs. Reaching a *socially good* equilibrium is crucial, in practice, to make correlation credible. Figure 2, on the right, shows the performance of CFR-Jr, CFR-S (mean plus/minus standard deviation), and CFR over G2-4-DA in terms of social welfare approximation ratio. Also in this case, CFR performs worse than the other two algorithms.

References

[1] Robert J Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.

[2] Siddharth Barman and Katrina Ligett. Finding any nontrivial coarse correlated equilibrium is hard. In *ACM Conference on Economics and Computation (EC)*, pages 815–816, 2015.

[3] David Blackwell et al. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[4] Avrim Blum and Yishay Monsour. Learning, regret minimization, and equilibria. 2007.

[5] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218):145–149, 2015.

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

[7] Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. *arXiv preprint arXiv:1809.04040*, 2018.

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

[9] Noam Brown, Adam Lerer, Sam Gross, and Tuomas Sandholm. Deep counterfactual regret minimization. *arXiv preprint arXiv:1811.00164*, 2018.

[10] Andrea Celli, Stefano Coniglio, and Nicola Gatti. Computing optimal coarse correlated equilibria in sequential games. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2019.

[11] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[12] Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. Ex ante coordination and collusion in zero-sum multi-player extensive-form games. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[13] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Online convex optimization for sequential decision processes and extensive-form games. *arXiv preprint arXiv:1809.03075*, 2018.

[14] Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40, 1997.

[15] Richard Gibson. Regret minimization in non-zero-sum games with applications to building champion multiplayer computer poker agents. *CoRR*, abs/1305.0034, 2013.

[16] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[17] Jason Hartline, Vasilis Syrgkanis, and Eva Tardos. No-regret learning in bayesian games. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 3061–3069, 2015.

[18] Albert Xin Jiang and Kevin Leyton-Brown. Polynomial-time computation of exact correlated equilibrium in compact games. *Games and Economic Behavior*, 91:347–359, 2015.

[19] Harold W Kuhn. A simplified two-person poker. *Contributions to the Theory of Games*, 1: 97–103, 1950.

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

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

[22] J. Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

[23] Christos H Papadimitriou and Tim Roughgarden. Computing correlated equilibria in multi-player games. *Journal of the ACM (JACM)*, 55(3):14, 2008.

[24] Nick Abou Risk and Duane Szafron. Using counterfactual regret minimization to create competitive multiplayer poker agents. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 159–166, 2010.

[25] Sheldon M Ross. Goofspiel—the game of pure strategy. *Journal of Applied Probability*, 8(3): 621–625, 1971.

[26] Tim Roughgarden. Intrinsic robustness of the price of anarchy. In *ACM symposium on Theory of computing (STOC)*, pages 513–522, 2009.

[27] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[28] Finnegan Southey, Michael H. Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and D. Chris Rayner. Bayes' bluff: Opponent modelling in poker. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 550–558, 2005.

[29] Oskari Tammelin. Solving large imperfect information games using cfr+. *arXiv preprint arXiv:1407.5042*, 2014.

[30] Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit texas hold'em. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 645–652, 2015.

[31] Kevin Waugh. Abstraction in large extensive games. 2009.

[32] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning (ICML)*, pages 928–936, 2003.

[33] Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1729–1736, 2008.