Counterfactual Regret Minimization and Domination in Extensive-Form Games

Richard Gibson
Computing Science Department
University of Alberta
Edmonton, Alberta T6G 2E8
rggibson@cs.ualberta.ca

Abstract

The Counterfactual Regret Minimization (CFR) algorithm is state of the art for computing approximate Nash equilibria of large, two-player zero-sum extensive-form games. While the current theory only holds in two-player zero-sum games, strategies produces by CFR for games involving more than two players have performed very well in practice. This short paper provides preliminary theory that attempts to explain the effectiveness of CFR strategies in games with more than two players. We define strictly dominated actions, a new notion of dominance specific to extensive-form games, and present results suggesting that CFR produces strategies free of domination.

1 Introduction

Counterfactual Regret Minimization (CFR) (Zinkevich et al. 2008) is an iterative procedure for computing strategy profiles in extensive-form games. In 2-player zero-sum games, CFR produces an approximate Nash equilibrium profile. Outside of 2-player zero-sum games, however, this guarantee is lost. Even in a small 3-player poker game, CFR fails to generate an approximate equilibrium after millions of iterations (Abou Risk and Szafron 2010, Table 2).

Despite the lack of theoretical guarantees, CFR still tends to produce very strong strategies in games with more than two players. The University of Alberta's entries into the 3-player instant-runoff event of the Annual Computer Poker Competition[1] have won the past three competitions (2009, 2010, and 2011). Each entry was generated primarily by applying CFR to an abstraction of 3-player Limit Texas Hold'em, with the main difference between years being the size of abstraction employed. Currently, there is no good explanation as to why strategies produced by CFR should perform well in this event.

In this short paper, we look to explain why CFR strategies perform well outside of two-player zero-sum games. As empirical evidence indicates that a Nash equilibrium cannot be guaranteed, we instead investigate domination properties of CFR strategies. We define a new notion of dominance specific to extensive-form games, show that CFR avoids strategies with domination properties, and discuss the results.

[1] http://www.computerpokercompetition.org

2 Background

Counterfactual Regret Minimization (CFR) (Zinkevich et al. 2008) is an iterative algorithm resembling self-play. On each iteration T, CFR traverses the entire game tree once per player, computing the expected utility for player i at each information set I under the current profile σ^T, assuming player i plays to reach I. This expectation is the **counterfactual value** for player i,

$$v_i(I,\sigma) = \sum_{z \in Z_I} u_i(z)\pi^\sigma_{-i}(z[I])\pi^\sigma(z[I],z).$$

Here, Z_I is the set of terminal histories passing through I, $u_i(z)$ is the utility for player i at terminal history z, $z[I]$ is the prefix of z contained in I, and $\pi^\sigma_{-i}(z[I])$ and $\pi^\sigma(z[I],z)$ are the probabilities of reaching $z[I]$ under σ assuming player i plays to reach $z[I]$, and of reaching z from $z[I]$ under σ respectively. Let $A(I)$ be the set of all actions available at information set I. For each action $a \in A(I)$, the counterfactual values determine the **counterfactual regret** at iteration T, $r_i^T(I,a) = v_i(I,\sigma^T_{(I\to a)}) - v_i(I,\sigma^T)$, where $\sigma_{(I\to a)}$ is the profile σ except at I, action a is always taken. The regret $r_i^T(I,a)$ measures how much player i would rather play action a at I than play σ^T. These regrets are accumulated to obtain the **immediate counterfactual regret**, $R_i^T(I,a) = \sum_{t=1}^T r_i^t(I,a)$. Finally, the next profile σ^{T+1} is obtained by applying **regret matching** (Hart and Mas-Colell 2000; Zinkevich et al. 2008), where the probability of taking action a at I is

$$\sigma^{T+1}(I,a) = \frac{R_i^{T,+}(I,a)}{\sum_{b\in A(I)} R_i^{T,+}(I,b)}, \tag{1}$$

with $x^+ = \max\{x, 0\}$ and actions being chosen uniformly at random when the denominator is zero. After many iterations T, CFR outputs the average profile $\bar{\sigma}^T$ according to

$$\bar{\sigma}_i^T(I,a) = \frac{\sum_{t=1}^T \pi_i^{\sigma^t}(I)\sigma_i^t(I,a)}{\sum_{t=1}^T \pi_i^{\sigma^t}(I)},$$

where $\pi_i^\sigma(I) = \sum_{h\in I} \pi_i^\sigma(h)$ is the probability of player i playing to reach I under σ.

A *pure strategy* s_i for player i assigns a probability of 1 to a single action at each information set I belonging to player i; denote this action by $s_i(I)$. A pure strategy s_i for player i is **strictly dominated** if there exists another strategy σ_i' such that $u_i(s_i, \sigma_{-i}) < u_i(\sigma_i', \sigma_{-i})$ for all opponent profiles σ_{-i}, where $u_i(\sigma)$ is the expected utility for player i when all players play according to σ. In addition, a pure strategy s_i is recursively defined to be **iteratively strictly dominated** if either s_i is strictly dominated, or if there exists another strategy σ_i' such that $u_i(s_i, \sigma_{-i}) < u_i(\sigma_i', \sigma_{-i})$ for all opponent profiles σ_{-i} containing no iteratively strictly dominated strategies. Clearly, one should never play a strictly dominated strategy. In addition, if we assume our opponents are rational and will not play strictly dominated strategies, one should also avoid iteratively strictly dominated strategies. There are also less restrictive notions, **weak dominance** and **very weak dominance**, where equality of utility is allowed for all but one opponent profile and for all opponent profiles respectively. To simplify our analysis, we focus on strict dominance only.

3 Theoretical Analysis

While strictly dominated strategies are applicable in all types of games, we now introduce a new notion of dominance that is specific to extensive-form games.

Definition 1 *Let I be an information set for player i and let a be an action available at I. We say that a is a **strictly dominated action (at I)** if there exists a player i strategy σ_i' such that for every profile σ, either $\pi_{-i}^\sigma(I) = 0$ or $v_i(I, \sigma_{(I\to a)}) < v_i(I, (\sigma_i', \sigma_{-i}))$. In addition, a is an **iteratively strictly dominated action (at I)** if there exists a player i strategy σ_i' such that for every profile σ containing no iteratively strictly dominated actions, either $\pi_{-i}^\sigma(I) = 0$ or $v_i(I, \sigma_{(I\to a)}) < v_i(I, (\sigma_i', \sigma_{-i}))$.*

In other words, a is strictly dominated at I if no matter how the opponents play to reach I and no matter how everyone plays after taking a at I, there is an alternative strategy (that does not take a at I) that always provides more value. Again, one should never play a strictly dominated action due to the presence of such a better, alternative strategy. Note that any pure strategy that plays an (iteratively) strictly dominated action at I and plays to reach I with positive probability is an (iteratively) strictly dominated strategy. Thus, strictly dominated actions can be considered a subclass of all strictly dominated strategies.

As long as the opponent strategies generated play to reach information set I "often enough," in the limit CFR's average profile will avoid playing strictly dominated actions at I:

Theorem 1 *Let I be an information set for player i. For $\delta > 0$, define $\Sigma_\delta := \{\sigma \in \Sigma \mid \sum_{h\in I} \pi_{-i}^\sigma(h) \geq \delta\}$. If a is a strictly dominated action at I, and if there exists an integer T_0 and real numbers $\delta, \gamma > 0$ such that $\left|\{\sigma^t\}_{t=1}^T \cap \Sigma_\delta\right| / T \geq \gamma$ for all $T \geq T_0$, then the average profile generated by CFR*[2] *satisfies*

$$\bar{\sigma}_i^T(I,a) \to 0 \text{ as } T \to \infty.$$

Moreover, under similar conditions, CFR's average profile avoids all iteratively strictly dominated actions:

Theorem 2 *Let I be an information set for player i, let $a \in A(I)$ be an iteratively strictly dominated action at I, and let $a_1 \in A(I_1), ..., a_k \in A(I_k)$ be all iteratively strictly dominated actions that are required to be played with zero probability for iterative dominance of a to be exhibited. For $\delta > 0$, define $\Sigma_\delta(I') := \{\sigma \in \Sigma \mid \sum_{h\in I'} \pi_{-i}^\sigma(h) \geq \delta\}$. If there exist integers $T_1, ..., T_k$ and real numbers $\delta_1, \gamma_1, ..., \delta_k, \gamma_k > 0$ such that for all $j = 1, ..., k$, $\left|\{\sigma^t\}_{t=1}^T \cap \Sigma_{\delta_j}(I_j)\right| / T \geq \gamma_j$ for all $T \geq T_j$, then*

$$\bar{\sigma}_i^T(I,a) \to 0 \text{ as } T \to \infty.$$

Finally, while we cannot quite guarantee that CFR's average profile will avoid strictly dominated strategies, we can show that the immediate counterfactual regret must eventually be negative for at least some action played by a strategy that is strictly dominated by another pure strategy:

Theorem 3 *Let $s_i, s_i' \in \Sigma_i$ be pure strategies for player i. If s_i is strictly dominated by s_i', then there exists an integer T_0 such that for all $T \geq T_0$, there exists an information set I for player i such that $\pi_i^{s_i}(I) = 1$ and $R_i^T(I, s_i(I)) < 0$.*

4 Discussion

We have provided results suggesting that CFR will avoid producing strategies that exhibit some form of strict dominance. It remains unclear, however, how significant these results actually are. In particular, we do not know how correlated performance is with playing dominated actions

[2]A minor modification to regret matching is required.

or strategies. While Nash equilibria are often considered optimal solutions for 2-player zero-sum games, simply avoiding domination may not be enough to perform well in games with more than two players. Future work will look to compare the performance of strategies containing varying numbers of dominated actions in small, toy games where the dominated actions can more easily be identified.

The output of CFR is the players' average strategies, and as such Theorems 1 and 2 are stated in terms of the result on a player's average strategy. However, stronger results can actually be obtained regarding player i's *current* strategy, σ_i^T. A player's current strategy is actually guaranteed, under the same conditions as Theorem 2, to stop playing iteratively strictly dominated actions after a finite number of iterations. Thus, to simply avoid dominated actions, one can play CFR's final, current strategy rather than the average. Future work will investigate how well CFR's current strategy performs in practice, which may help indicate how important it is to avoid dominated actions.

References

Abou Risk, N., and Szafron, D. 2010. Using counterfactual regret minimization to create competitive multiplayer poker agents. In *Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 159–166.

Hart, S., and Mas-Colell, A. 2000. A simple adaptive procedure leading to correlated equilibrium. *Econometrica* **68**:1127–1150.

Zinkevich, M.; Johanson, M.; Bowling, M.; and Piccione, C. 2008. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems 20 (NIPS)*, 905–912.