Strategy-Based Warm Starting for Regret Minimization in Games

Noam Brown
Computer Science Department
Carnegie Mellon University
noamb@cs.cmu.edu

Tuomas Sandholm
Computer Science Department
Carnegie Mellon University
sandholm@cs.cmu.edu

Abstract

Counterfactual Regret Minimization (CFR) is a popular iterative algorithm for approximating Nash equilibria in imperfect-information multi-step two-player zero-sum games. We introduce the first general, principled method for warm starting CFR. Our approach requires only a strategy for each player, and accomplishes the warm start at the cost of a single traversal of the game tree. The method provably warm starts CFR to as many iterations as it would have taken to reach a strategy profile of the same quality as the input strategies, and does not alter the convergence bounds of the algorithms. Unlike prior approaches to warm starting, ours can be applied in all cases.

Our method is agnostic to the origins of the input strategies. For example, they can be based on human domain knowledge, the observed strategy of a strong agent, the solution of a coarser abstraction, or the output of some algorithm that converges rapidly at first but slowly as it gets closer to an equilibrium. Experiments demonstrate that one can improve overall convergence in a game by first running CFR on a smaller, coarser abstraction of the game and then using the strategy in the abstract game to warm start CFR in the full game.

Introduction

Imperfect-information games model strategic interactions between players that have access to private information. Domains such as negotiations, cybersecurity and physical security interactions, and recreational games such as poker can all be modeled as imperfect-information games. Typically in such games, one wishes to find a *Nash equilibrium*, where no player can do better by switching to a different strategy. In this paper we focus specifically on two-player zero-sum games. Over the last 10 years, tremendous progress has been made in solving increasingly larger two-player zero-sum imperfect-information games; for reviews, see (Sandholm 2010; 2015). Linear programs have been able to solve games up to 10^7 or 10^8 nodes in the game tree (Gilpin and Sandholm 2005). Larger games are solved using iterative algorithms that converge over time to a Nash equilibrium. The most popular iterative algorithm for this is *Counterfactual Regret Minimization (CFR)* (Zinkevich et al. 2007). A variant of CFR was recently used to essentially solve Limit Texas Hold'em, which at 10^{15} nodes (after lossless abstraction (Gilpin and Sandholm 2007)) is the largest imperfect-information game ever to be essentially solved (Bowling et al. 2015).

One of the main constraints in solving such large games is the time taken to arrive at a solution. For example, essentially solving Limit Texas Hold'em required running CFR on 4,800 cores for 68 days (Tammelin et al. 2015). Even though Limit Texas Hold'em is a popular human game with many domain experts, and even though several near-Nash equilibrium strategies had previously been computed for the game (Johanson et al. 2011; 2012), there was no known way to leverage that prior strategic knowledge to speed up CFR. We introduce such a method, enabling user-provided strategies to warm start convergence toward a Nash equilibrium.

The effectiveness of warm starting in large games is magnified by *pruning*, in which some parts of the game tree need not be traversed during an iteration of CFR. This results in faster iterations and therefore faster convergence to a Nash equilibrium. The frequency of pruning opportunities generally increases as equilibrium finding progresses (Lanctot et al. 2009). This may result in later iterations being completed multiple orders of magnitude faster than early iterations. This is especially true with the recently-introduced *regret-based pruning* method, which drastically increases the opportunities for pruning in a game (Brown and Sandholm 2015a). Our warm starting algorithm can "skip" these early expensive iterations that might otherwise account for the bulk of the time spent on equilibrium finding. This can be accomplished by first solving a coarse abstraction of the game, which is relatively cheap, and using the equilibrium strategies computed in the abstraction to warm start CFR in the full game. Experiments presented later in this paper show the effectiveness of this method.

Our warm start technique also opens up the possibility of constructing and refining abstractions during equilibrium finding. Current abstraction techniques for large imperfect-information games are domain specific and rely on human expert knowledge because the abstraction must be set before any strategic information is learned about the game (Brown, Ganzfried, and Sandholm 2015; Ganzfried and Sandholm 2014; Johanson et al. 2013; Billings et al. 2003). There are some exceptions to this, such as work that refines parts of the game tree based on the computed strategy of a coarse abstraction (Jackson 2014; Gibson 2014). However, in these cases either equilibrium finding had to be restarted from scratch after the modification, or the final strategy was not guaranteed to be a Nash equilibrium. Recent work has also considered feature-based abstractions that allow the abstraction to change during equilibrium finding (Waugh et al. 2015). However, in this case, the features must still be determined by domain experts and set before equilibrium finding begins.

In contrast, the recently introduced *simultaneous abstraction and equilibrium finding (SAEF)* algorithm does not rely on domain knowledge (Brown and Sandholm 2015b). Instead, it iteratively refines an abstraction based on the strategic information gathered during equilibrium finding. When

an abstraction is refined, SAEF warm starts equilibrium finding in the new abstraction using the strategies from the previous abstraction. However, previously-proposed warm-start methods only applied in special cases. Specifically, it was possible to warm start CFR in one game using the results of CFR in another game that has identical structure but where the payoffs differ by some known parameters (Brown and Sandholm 2014). It was also possible to warm start CFR when adding actions to a game that CFR had previously been run on, though a $O(1)$ warm start could only be achieved under limited circumstances. In these prior cases, warm starting required the prior strategy to be computed using CFR. In contrast, the method presented in this paper can be applied in all cases, is agnostic to the origin of the provided strategy, and costs only a single traversal of the game tree. This expands the scope and effectiveness of SAEF.

The rest of the paper is structured as follows. The next section covers background and notation. After that, we introduce the method for warm starting. Then, we cover practical implementation details that lead to improvements in performance. Finally, we present experimental results showing that the warm starting method is highly effective.

Background and Notation

In an imperfect-information extensive-form game there is a finite set of players, $\mathcal{P}$. H is the set of all possible histories (nodes) in the game tree, represented as a sequence of actions, and includes the empty history. $A(h)$ is the actions available in a history and $P(h) \in \mathcal{P} \cup c$ is the player who acts at that history, where c denotes chance. Chance plays an action $a \in A(h)$ with a fixed probability $\sigma_c(h, a)$ that is known to all players. The history h' reached after an action is taken in h is a child of h, represented by $h{\cdot}a = h'$, while h is the parent of h'. If there exists a sequence of actions from h to h', then h is an ancestor of h' (and h' is a descendant of h). $Z \subseteq H$ are terminal histories for which no actions are available. For each player $i \in \mathcal{P}$, there is a payoff function $u_i : Z \to \Re$. If $P = \{1, 2\}$ and $u_1 = -u_2$, the game is two-player zero-sum.

Imperfect information is represented by *information sets* for each player $i \in \mathcal{P}$ by a partition $\mathcal{I}_i$ of $h \in H : P(h) = i$. For any information set $I \in \mathcal{I}_i$, all histories $h, h' \in I$ are indistinguishable to player i, so $A(h) = A(h')$. $I(h)$ is the information set I where $h \in I$. $P(I)$ is the player i such that $I \in \mathcal{I}_i$. $A(I)$ is the set of actions such that for all $h \in I$, $A(I) = A(h)$. $|A_i| = \max_{I \in \mathcal{I}_i} |A(I)|$ and $|A| = \max_i |A_i|$. We define Δ_i as the range of payoffs reachable by player i. Formally, $\Delta_i = \max_{z \in Z} u_i(z) - \min_{z \in Z} u_i(z)$ and $\Delta = \max_i \Delta_i$. We similarly define $\Delta(I)$ as the range of payoffs reachable from I. Formally, $\Delta(I) = \max_{z \in Z, h \in I : h \sqsubseteq z} u_{P(I)}(z) - \min_{z \in Z, h \in I : h \sqsubseteq z} u_{P(I)}(z)$.

A strategy $\sigma_i(I)$ is a probability vector over $A(I)$ for player i in information set I. The probability of a particular action a is denoted by $\sigma_i(I, a)$. Since all histories in an information set belonging to player i are indistinguishable, the strategies in each of them must be identical. That is, for all $h \in I$, $\sigma_i(h) = \sigma_i(I)$ and $\sigma_i(h, a) = \sigma_i(I, a)$. We define σ_i to be a probability vector for player i over all available strategies Σ_i in the game. A strategy profile σ is a tuple of strategies, one for each player. $u_i(\sigma_i, \sigma_{-i})$ is the expected payoff for player i if all players play according to the strategy profile $\langle \sigma_i, \sigma_{-i} \rangle$. If a series of strategies are played over T iterations, then $\bar{\sigma}_i^T = \frac{\sum_{t \in T} \sigma_i^t}{T}$.

$\pi^\sigma(h) = \Pi_{h' \cdot a \sqsubseteq h} \sigma_{P(h)}(h, a)$ is the joint probability of reaching h if all players play according to σ. $\pi_i^\sigma(h)$ is the contribution of player i to this probability (that is, the probability of reaching h if all players other than i, and chance, always chose actions leading to h). $\pi_{-i}^\sigma(h)$ is the contribution of all players other than i, and chance. $\pi^\sigma(h, h')$ is the probability of reaching h' given that h has been reached, and 0 if $h \not\sqsubset h'$. In a *perfect-recall* game, $\forall h, h' \in I \in \mathcal{I}_i$, $\pi_i(h) = \pi_i(h')$. In this paper we focus on perfect-recall games. Therefore, for $i = P(I)$ we define $\pi_i(I) = \pi_i(h)$ for $h \in I$. We define the average strategy $\bar{\sigma}_i^T(I)$ for an information set I to be

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

Counterfactual Regret Minimization (CFR)

Counterfactual regret minimization (CFR) is an equilibrium finding algorithm for extensive-form games that independently minimizes regret in each information set (Zinkevich et al. 2007). While any regret-minimizing algorithm can be used in the information sets, *regret matching (RM)* is the most popular option (Hart and Mas-Colell 2000).

Our analysis of CFR makes frequent use of *counterfactual value*. Informally, this is the expected utility of an information set given that player i tries to reach it. For player i at information set I given a strategy profile σ, this is defined as

$$v^\sigma(I) = \sum_{h \in I} \left(\pi_{-i}^\sigma(h) \sum_{z \in Z} \left(\pi^\sigma(h, z) u_i(z) \right) \right) \tag{2}$$

and the counterfactual value of an action a is

$$v^\sigma(I, a) = \sum_{h \in I} \left(\pi_{-i}^\sigma(h) \sum_{z \in Z} \left(\pi^\sigma(h \cdot a, z) u_i(z) \right) \right) \tag{3}$$

Let σ^t be the strategy profile used on iteration t. The *instantaneous regret* on iteration t for action a in information set I is $r^t(I, a) = v^{\sigma^t}(I, a) - v^{\sigma^t}(I)$. The *regret* for action a in I on iteration T is

$$R^T(I, a) = \sum_{t=1}^T r^t(I, a) \tag{4}$$

Additionally, $R_+^T(I, a) = \max\{R^T(I, a), 0\}$ and $R^T(I) = \max_a\{R_+^T(I, a)\}$. Regret for player i in the entire game is

$$R_i^T = \max_{\sigma_i' \in \Sigma_i} \sum_{t=1}^T \left(u_i(\sigma_i', \sigma_{-i}^t) - u_i(\sigma_i^t, \sigma_{-i}^t) \right) \tag{5}$$

In RM, a player in an information set picks an action among the actions with positive regret in proportion to the positive regret on that action. Formally, on each iteration $T + 1$, player i selects actions $a \in A(I)$ according to probabilities

$$\sigma_i^{T+1}(I,a) = \begin{cases} \frac{R_+^T(I,a)}{\sum_{a' \in A(I)} R_+^T(I,a')}, & \text{if } \sum_{a' \in A_i} R_+^T(I,a') > 0 \\ \frac{1}{|A(I)|}, & \text{otherwise} \end{cases} \tag{6}$$

If player i plays according to RM in information set I on iteration T, then

$$\sum_{a \in A(I)} \left(R_+^T(I,a)\right)^2 \leq \sum_{a \in A(I)} \left(\left(R_+^{T-1}(I,a)\right)^2 + \left(r^T(I,a)\right)^2 \right) \tag{7}$$

This leads us to the following lemma.[1]

Lemma 1. *After T iterations of regret matching are played in an information set I,*

$$\sum_{a \in A(I)} \left(R_+^T(I,a)\right)^2 \leq \pi_{-i}^{\bar{\sigma}^T}(I) \left(\Delta(I)\right)^2 |A(I)| T \tag{8}$$

Most proofs are presented in an extended version of this paper.

In turn, this leads to a bound on regret

$$R^T(I) \leq \sqrt{\pi_{-i}^{\bar{\sigma}^T}(I)} \Delta(I) \sqrt{|A(I)|} \sqrt{T} \tag{9}$$

The key result of CFR is that $R_i^T \leq \sum_{I \in \mathcal{I}_i} R^T(I) \leq \sum_{I \in \mathcal{I}_i} \sqrt{\pi_{-i}^{\bar{\sigma}^T}} \Delta(I) \sqrt{|A(I)|} \sqrt{T}$. So, as $T \to \infty$, $\frac{R_i^T}{T} \to 0$.

In two-player zero-sum games, regret minimization converges to a *Nash equilibrium*, i.e., a strategy profile σ^* such that $\forall i,\ u_i(\sigma_i^*, \sigma_{-i}^*) = \max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \sigma_{-i}^*)$. An *$\epsilon$-equilibrium* is a strategy profile σ^* such that $\forall i,\ u_i(\sigma_i^*, \sigma_{-i}^*) + \epsilon \geq \max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \sigma_{-i}^*)$. Since we will reference the details of the following known result later, we reproduce the proof here.

Theorem 1. *In a two-player zero-sum game, if $\frac{R_i^T}{T} \leq \epsilon_i$ for both players $i \in \mathcal{P}$, then $\bar{\sigma}^T$ is a $(\epsilon_1 + \epsilon_2)$-equilibrium.*

Proof. We follow the proof approach of Waugh et al. (2009). From (5), we have that

$$\max_{\sigma_i' \in \Sigma_i} \frac{1}{T} \Big(\sum_{t=1}^{T} u_i(\sigma_i', \sigma_{-i}^t) - u_i(\sigma_i^t, \sigma_{-i}^t) \Big) \leq \epsilon_i \tag{10}$$

Since σ_i' is the same on every iteration, this becomes

$$\max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \bar{\sigma}_{-i}^T) - \frac{1}{T} \sum_{t=1}^{T} u_i(\sigma_i^t, \sigma_{-i}^t) \leq \epsilon_i \tag{11}$$

Since $u_1(\sigma) = -u_2(\sigma)$, if we sum (11) for both players

$$\max_{\sigma_1' \in \Sigma_1} u_1(\sigma_1', \bar{\sigma}_2^T) + \max_{\sigma_2' \in \Sigma_2} u_2(\bar{\sigma}_1^T, \sigma_2') \leq \epsilon_1 + \epsilon_2 \tag{12}$$

$$\max_{\sigma_1' \in \Sigma_1} u_1(\sigma_1', \bar{\sigma}_2^T) - \min_{\sigma_2' \in \Sigma_2} u_1(\bar{\sigma}_1^T, \sigma_2') \leq \epsilon_1 + \epsilon_2 \tag{13}$$

Since $u_1(\bar{\sigma}_1^T, \bar{\sigma}_2^T) \geq \min_{\sigma_2' \in \Sigma_2} u_1(\bar{\sigma}_1^T, \sigma_2')$ so we have $\max_{\sigma_1' \in \Sigma_1} u_1(\sigma_1', \bar{\sigma}_{-2}^T) - u_1(\bar{\sigma}_1^T, \bar{\sigma}_2^T) \leq \epsilon_1 + \epsilon_2$. By symmetry, this is also true for Player 2. Therefore, $\langle \bar{\sigma}_1^T, \bar{\sigma}_2^T \rangle$ is a $(\epsilon_1 + \epsilon_2)$-equilibrium. □

[1] A tighter bound would be $\sum_{t=1}^{T} \left(\pi_{-i}^{\sigma^t}(I)\right)^2 \left(\Delta(I)\right)^2 |A(I)|$. However, for reasons that will become apparent later in this paper, we prefer a bound that uses only the average strategy $\bar{\sigma}^T$.

Warm-Starting Algorithm

In this section we explain the theory of how to warm start CFR and prove the method's correctness. By warm starting, we mean we wish to effectively "skip" the first T iterations of CFR (defined more precisely later in this section). When discussing intuition, we use normal-form games due to their simplicity. Normal-form games are a special case of games in which each player only has one information set. They can be represented as a matrix of payoffs where Player 1 picks a row and Player 2 simultaneously picks a column.

The key to warm starting CFR is to correctly initialize the regrets. To demonstrate the necessity of this, we first consider an ineffective approach in which we set only the starting strategy, but not the regrets. Consider the two-player zero-sum normal-form game defined by the payoff matrix $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right]$ with payoffs shown for Player 1 (the row player). The Nash equilibrium for this game requires Player 1 to play $\langle \frac{2}{3}, \frac{1}{3} \rangle$ and Player 2 to play $\langle \frac{2}{3}, \frac{1}{3} \rangle$. Suppose we wish to warm start regret matching with the strategy profile σ^* in which both players play $\langle 0.67, 0.33 \rangle$ (which is very close to the Nash equilibrium). A naïve way to do this would be to set the strategy on the first iteration to $\langle 0.67, 0.33 \rangle$ for both players, rather than the default of $\langle 0.5, 0.5 \rangle$. This would result in regret of $\langle 0.0023, -0.0067 \rangle$ for Player 1 and $\langle -0.0023, 0.0067 \rangle$ for Player 2. From (6), we see that on the second iteration Player 1 would play $\langle 1, 0 \rangle$ and Player 2 would play $\langle 0, 1 \rangle$, resulting in regret of $\langle 0.0023, 1.9933 \rangle$ for Player 1. That is a huge amount of regret, and makes this warm start no better than starting from scratch. Intuitively, this naïve approach is comparable to warm starting gradient descent by setting the initial point close to the optimum, but not reducing the step size. The result is that we overshoot the optimal strategy significantly. In order to add some "inertia" to the starting strategy so that CFR does not overshoot, we need a method for setting the regrets as well in CFR.

Fortunately, it is possible to efficiently calculate how far a strategy profile is from the optimum (that is, from a Nash equilibrium). This knowledge can be leveraged to initialize the regrets appropriately. To provide intuition for this warm starting method, we consider warm starting CFR to T iterations in a normal-form game based on an arbitrary strategy σ. Later, we discuss how to determine T based on σ.

First, the average strategy profile is set to $\bar{\sigma}^T = \sigma$. We now consider the regrets. From (4), we see regret for action a after T iterations of CFR would normally be $R_i^T(a) = \sum_{t=1}^{T} \Big(u_i(a, \sigma_{-i}^t) - u_i(\sigma^t) \Big)$. Since $\sum_{t=1}^{T} u_i(a, \sigma_{-i}^t)$ is the value of having played action a on every iteration, it is the same as $T u_i(a, \bar{\sigma}_{-i}^T)$. When warm starting, we can calculate this value because we set $\bar{\sigma}^T = \sigma$. However, we cannot calculate $\sum_{t=1}^{T} u_i(\sigma^t)$ because we did not define individual strategies played on each iteration. Fortunately, it turns out we can substitute another value we refer to as $T v_i'^{\bar{\sigma}^T}$, chosen from a range of acceptable options. To see this, we first observe that the value of $\sum_{t=1}^{T} u_i(\sigma^t)$ is not relevant to the proof of Theorem 1. Specifically, in (12), we see it cancels out. Thus, if we choose $v_i'^{\bar{\sigma}^T}$ such that $v_1'^{\bar{\sigma}^T} + v_2'^{\bar{\sigma}^T} \leq 0$, Theorem 1 still holds. This is our first constraint.

There is an additional constraint on our warm start. We must ensure that no information set violates the bound on regret guaranteed in (8). If regret exceeds this bound, then convergence to a Nash equilibrium may be slower than CFR guarantees. Thus, our second constraint is that when warm starting to T iterations, the initialized regret in every information set must satisfy (8). If these conditions hold and CFR is played after the warm start, then the bound on regret will be the same as if we had played T iterations from scratch instead of warm starting. When using our warm start method in extensive-form games, we do not directly choose $v_i'^{\bar{\sigma}^T}$ but instead choose a value $u'^{\bar{\sigma}^T}(I)$ for every information set (and we will soon see that these choices determine $v_i'^{\bar{\sigma}^T}$).

We now proceed to formally presenting our warm-start method and proving its effectiveness. Theorem 2 shows that we can warm start based on an arbitrary strategy σ by replacing $\sum_{t=1}^{T} v^{\sigma^t}(I)$ for each I with some value $Tv'^{\sigma}(I)$ (where $v'^{\sigma}(I)$ satisfies the constraints mentioned above). Then, Corollary 1 shows that this method of warm starting is lossless: if T iterations of CFR were played and we then warm start using $\bar{\sigma}^T$, we can warm start to T iterations.

We now define some terms that will be used in the theorem. When warm starting, a *substitute information set value* $u'^{\sigma}(I)$ is chosen for every information set I (we will soon describe how). Define $v'^{\sigma}(I) = \pi^{\sigma}_{-P(I)}(I)u'^{\sigma}(I)$ and define $v_i'^{\sigma}(h)$ for $h \in I$ as $\pi^{\sigma}_{-i}(h)u'^{\sigma}(I)$. Define $v_i'^{\sigma}(z)$ for $z \in Z$ as $\pi^{\sigma}_{-i}u_i(z)$.

As explained earlier in this section, in normal-form games $\sum_{t=1}^{T} u_i(a, \sigma^t_{-i}) = Tu_i(a, \bar{\sigma}^T_{-i})$. This is still true in extensive-form games for information sets where a leads to a terminal payoff. However, it is not necessarily true when a leads to another information set, because then the value of action a depends on how the player plays in the next information set. Following this intuition, we will define substitute counterfactual value for an action. First, define $Succ_i^{\sigma}(h)$ as the set consisting of histories h' that are the earliest reachable histories from h such that $P(h') = i$ or $h' \in Z$. By "earliest reachable" we mean $h \sqsubseteq h'$ and there is no h'' in $Succ^{\sigma}(h)$ such that $h'' \sqsubset h'$. Then the *substitute counterfactual value* of action a, where $i = P(I)$, is

$$v'^{\sigma}(I, a) = \sum_{h \in I} \Big(\sum_{h' \in Succ_i^{\sigma}(h \cdot a)} v_i'^{\sigma}(h') \Big) \tag{14}$$

and *substitute value* for player i is defined as

$$v_i'^{\sigma} = \sum_{h' \in Succ_i^{\sigma}(\emptyset)} v_i'^{\sigma}(h') \tag{15}$$

We define *substitute regret* as

$$R'^{T}(I, a) = T\big(v'^{\sigma}(I, a) - v'^{\sigma}(I)\big)$$

and

$$R'^{T,T'}(I, a) = R'^{T}(I, a) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}}(I, a) - v^{\sigma^{t'}}(I)\big)$$

Also, $R'^{T,T'}(I) = \max_{a \in A(I)} R'^{T,T'}(I, a)$. We also define the *combined strategy profile*

$$\sigma'^{T,T'} = \frac{T\sigma + T'\bar{\sigma}^{T'}}{T + T'}$$

Using these definitions, we wish to choose $u'^{\sigma}(I)$ such that

$$\sum_{a \in A(I)} \big(v'^{\sigma}(I, a) - v'^{\sigma}(I)\big)_+^2 \leq \frac{\pi^{\sigma}_{-i}(I)\big(\Delta(I)\big)^2 |A(I)|}{T} \tag{16}$$

We now proceed to the main result of this paper.

Theorem 2. *Let σ be an arbitrary strategy profile for a two-player zero-sum game. Choose any T and choose $u'^{\sigma}(I)$ in every information set I such that $v_1'^{\sigma} + v_2'^{\sigma} \leq 0$ and (16) is satisfied for every information set I. If we play T' iterations according to CFR, where on iteration T^*, $\forall I\ \forall a$ we use substitute regret $R'^{T,T^*}(I, a)$, then $\sigma'^{T,T'}$ forms a $(\epsilon_1 + \epsilon_2)$-equilibrium where $\epsilon_i = \frac{\sum_{I \in \mathcal{I}_i} \sqrt{\pi^{\sigma'^{T,T'}}_{-i}(I)}\Delta(I)\sqrt{|A(I)|}}{\sqrt{T+T'}}$.*

Theorem 2 allows us to choose from a range of valid values for T and $u'^{\sigma}(I)$. Although it may seem optimal to choose the values that result in the largest T allowed, this is typically not the case in practice. This is because in practice CFR converges significantly faster than the theoretical bound. In the next two sections we cover how to choose $u'^{\sigma}(I)$ and T within the theoretically sound range so as to converge even faster in practice.

The following corollary shows that warm starting using (16) is lossless: if we play CFR from scratch for T iterations and then warm start using $\bar{\sigma}^T$ by setting $u'^{\bar{\sigma}^T}(I)$ to even the lowest value allowed by (16), we can warm start to T.

Corollary 1. *Assume T iterations of CFR were played and let $\sigma = \bar{\sigma}^T$ be the average strategy profile. If we choose $u'^{\sigma}(I)$ for every information set I such that $\sum_{a \in A(I)} \big(v'^{\sigma}(I, a) - v'^{\sigma}(I)\big)_+^2 = \frac{\pi^{\sigma}_{-i}(I)\big(\Delta(I)\big)^2 |A(I)|}{T}$, and then play T' additional iterations of CFR where on iteration T^*, $\forall I\ \forall a$ we use $R_i'^{T,T^*}(I, a)$, then the average strategy profile over the $T + T'$ iterations forms a $(\epsilon_1 + \epsilon_2)$-equilibrium where $\epsilon_i = \frac{\sum_{I \in \mathcal{I}_i} \sqrt{\pi^{\sigma'^{T,T'}}_{-i}(I)}\Delta(I)\sqrt{|A(I)|}}{\sqrt{T+T'}}$.*

Choosing Number of Warm-Start Iterations

In this section we explain how to determine the number of iterations T to warm start to, given only a strategy profile σ. We give a method for determining a theoretically acceptable range for T. We then present a heuristic for choosing T within that range that delivers strong practical performance.

In order to apply Theorem 1, we must ensure $v_1'^{\sigma} + v_2'^{\sigma} \leq 0$. Thus, a theoretically acceptable upper bound for T would satisfy $v_1'^{\sigma} + v_2'^{\sigma} = 0$ when $u'^{\sigma}(I)$ in every information set I is set as low as possible while still satisfying (16).

In practice, setting T to this theoretical upper bound would perform very poorly because CFR tends to converge much faster than its theoretical bound. Fortunately, CFR also tends to converge at a fairly consistent rate within a game. Rather than choose a T that is as large as the theory allows, we can instead choose T based on how CFR performs over a short run in the particular game we are warm starting.

Specifically, we generate a function $f(T)$ that maps an iteration T to an estimate of how close $\bar{\sigma}^T$ would be to a Nash equilibrium after T iterations of CFR starting from scratch.

This function can be generated by fitting a curve to the first few iterations of CFR in a game. $f(T)$ defines another function, $g(\sigma)$, which estimates how many iterations of CFR it would take to reach a strategy profile as close to a Nash equilibrium as σ. Thus, in practice, given a strategy profile σ we warm start to $T = g(\sigma)$ iterations. In those experiments that required guessing an appropriate T (namely Figures 2 and 3) we based $g(\sigma)$ on a short extra run (10 iterations of CFR) starting from scratch. The experiments show that this simple method is sufficient to obtain near-perfect performance.

Choosing Substitute Counterfactual Values

Theorem 2 allows for a range of possible values for $u'^{\sigma}(I)$. In this section we discuss how to choose a particular value for $u'^{\sigma}(I)$, assuming we wish to warm start to T iterations.

From (14), we see that $v'^{\sigma}(I,a)$ depends on the choice of $u'^{\sigma}(I')$ for information sets I' that follow I. Therefore, we set $u'^{\sigma}(I)$ in a bottom-up manner, setting it for information sets at the bottom of the game tree first. This method resembles a best-response calculation. When calculating a best response for a player, we fix the opponent's strategy and traverse the game tree in a depth-first manner until a terminal node is reached. This payoff is then passed up the game tree. When all actions in an information set have been explored, we pass up the value of the highest-utility action.

Using a best response would likely violate the constraint $v_1'^{\sigma} + v_2'^{\sigma} \leq 0$. Therefore, we compute the following response instead. After every action in information set I has been explored, we set $u'^{\sigma}(I)$ so that (16) is satisfied. We then pass $v'^{\sigma}(I)$ up the game tree.

From (16) we see there are a range of possible options for $u'^{\sigma}(I)$. In general, lower regret (that is, playing closer to a best response) is preferable, so long as $v_1'^{\sigma} + v_2'^{\sigma} \leq 0$ still holds. In this paper we choose an information set-independent parameter $0 \leq \lambda_i \leq 1$ for each player and set $u'^{\sigma}(I)$ such that

$$\sum_{a \in A(I)} \left(v'^{\sigma}(I) - v'^{\sigma}(I,a)\right)_+^2 = \frac{\lambda_i \pi^{\sigma}_{-i}(I)\left(\Delta(I)\right)^2 |A(I)|}{T}$$

Finding λ_i such that $v_1'^{\sigma} + v_2'^{\sigma} = 0$ is difficult. Fortunately, performance is not very sensitive to the choice of λ_i . Therefore, when we warm start, we do a binary search for λ_i so that $v_1'^{\sigma} + v_2'^{\sigma}$ is close to zero (and not positive).

Using λ_i is one valid method for choosing $u'^{\sigma}(I)$ from the range of options that (16) allows. However, there may be heuristics that perform even better in practice. In particular, $\pi^{\sigma}_{-i}\left(\Delta(I)\right)^2$ in (16) acts as a bound on $\left(r^t(I,a)\right)^2$. If a better bound, or estimation, for $\left(r^t(I,a)\right)^2$ exists, then substituting that in (16) may lead to even better performance.

Experiments

We now present experimental results for our warm-starting algorithm. We begin by demonstrating an interesting consequence of Corollary 1. It turns out that in two-player zero-sum games, we need not store regrets at all. Instead, we can keep track of only the average strategy played. On every iteration, we can "warm start" using the average strategy to directly determine the probabilities for the next iteration. We tested this algorithm on random 100x100 normal-form games, where the entries of the payoff matrix are chosen uniformly at random from $[-1, 1]$. On every iteration $T > 0$, we set $v_1'^{\bar{\sigma}^T} = v_2'^{\bar{\sigma}^T}$ such that

$$\frac{|\Delta_1|^2|A_1|}{\sum_{a_1}\left(u_1(a_1, \bar{\sigma}_2^T) - v_1'^{\bar{\sigma}^T}\right)_+^2} = \frac{|\Delta_2|^2|A_2|}{\sum_{a_2}\left(u_2(a_2, \bar{\sigma}_1^T) - v_2'^{\bar{\sigma}^T}\right)_+^2}$$

Figure 1 shows that warm starting every iteration in this way results in performance that is virtually identical to CFR.

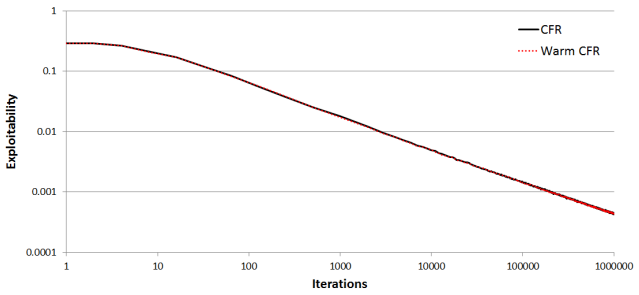

Figure 1: Comparison of CFR vs warm starting every iteration. The results shown are the average over 64 different 100x100 normal-form games.

The remainder of our experiments are conducted on a game we call *Flop Texas Hold'em* (FTH). FTH is a version of poker similar to Limit Texas Hold'em except there are only two rounds, called the pre-flop and flop. At the beginning of the game, each player receives two private cards from a 52-card deck. Player 1 puts in the "big blind" of two chips, and Player 2 puts in the "small blind" of one chip. A round of betting then proceeds, starting with Player 2, in which up to three bets or raises are allowed. All bets and raises are two chips. Either player may fold on their turn, in which case the game immediately ends and the other player wins the pot. After the first betting round is completed, three community cards are dealt out, and another round of betting is conducted (starting with Player 1), in which up to four bets or raises are allowed. At the end of this round, both players form the best five-card poker hand they can using their two private cards and the three community cards. The player with the better hand wins the pot.

The second experiment compares our warm starting to CFR in FTH. We run CFR for some number of iterations before resetting the regrets according to our warm start algorithm, and then continuing CFR. We compare this to just running CFR without resetting. When resetting, we determine the number of iterations to warm start to based on an estimated function of the convergence rate of CFR in FTH, which is determined by the first 10 iterations of CFR. Our projection method estimated that after T iterations of CFR, $\bar{\sigma}^T$ is a $\frac{10.82}{T}$-equilibrium. Thus, when warm starting based on a strategy profile with exploitability x, we warm start to $T = \frac{10.82}{x}$. Figure 2 shows performance when warm starting at 100, 500, and 2500 iterations. These are three separate runs, where we warm start once on each run. We compare them to a run of CFR with no warm starting. Based

on the average strategies when warm starting occurred, the runs were warm started to 97, 490, and 2310 iterations, respectively. The figure shows there is almost no performance difference between warm starting and not warm starting.[2]

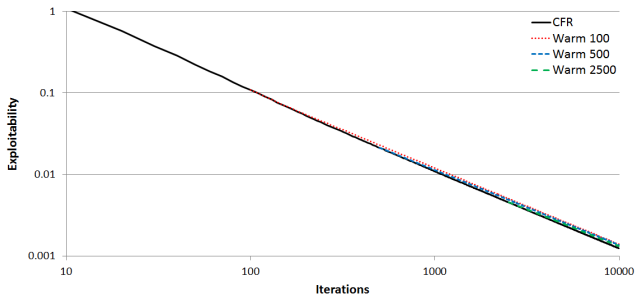

Figure 2: Comparison of CFR vs warm starting after 100, 500, or 2500 iterations. We warm started to 97, 490, and 2310 iterations, respectively. We used $\lambda = 0.08, 0.05, 0.02$ respectively (using the same λ for both players).

The third experiment demonstrates one of the main benefits of warm starting: being able to use a small coarse abstraction and/or quick-but-rough equilibrium-finding technique first, and starting CFR from that solution, thereby obtaining convergence faster. In all of our experiments, we leverage a number of implementation tricks that allow us to complete a full iteration of CFR in FTH in about three core minutes (Johanson et al. 2011). This is about four orders of magnitude faster than vanilla CFR. Nevertheless, there are ways to obtain good strategies even faster. To do so, we use two approaches. The first is a variant of CFR called External-Sampling Monte Carlo CFR (MCCFR) (Lanctot et al. 2009), in which chance nodes and opponent actions are sampled, resulting in much faster (though less accurate) iterations. The second is abstraction, in which several similar information sets are bucketed together into a single information set (where "similar" is defined by some heuristic). This constrains the final strategy, potentially leading to worse long-term performance. However, it can lead to faster convergence early on due to all information sets in a bucket sharing their acquired regrets and due to the abstracted game tree being smaller. Abstraction is particularly useful when paired with MCCFR, since MCCFR can update the strategy of an entire bucket by sampling only one information set.

In our experiment, we compare three runs: CFR, MCCFR in which the 1,286,792 flop poker hands have been abstracted into just 5,000 buckets, and CFR that was warm started with six core minutes of the MCCFR run. As seen in Figure 3, the MCCFR run improves quickly but then levels off, while CFR takes a relatively long time to converge, but eventually overtakes the MCCFR run. The warm start run combines the benefit of both, quickly reaching a good strategy while converging as fast as CFR in the long run.

[2]Although performance between the runs is very similar, it is not identical, and in general there may be differences in the convergence rate of CFR due to seemingly inconsequential differences that may change to which equilibrium CFR converges, or from which direction it converges.

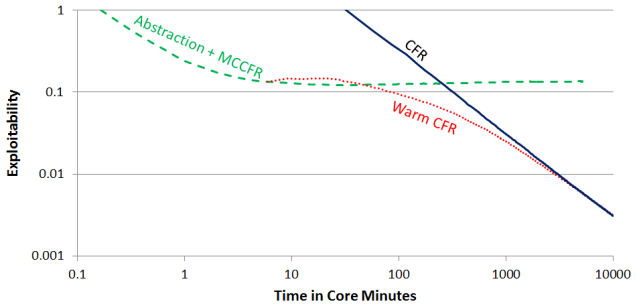

Figure 3: Performance of full-game CFR when warm started. The MCCFR run uses an abstraction with 5,000 buckets on the flop. After six core minutes of the MCCFR run, its average strategy was used to warm start CFR in the full to $T = 70$ using $\lambda = 0.08$.

In many extensive-form games, later iterations are cheaper than earlier iterations due to the increasing prevalence of pruning, in which sections of the game tree need not be traversed. In this experiment, the first 10 iterations took 50% longer than the last 10, which is a relatively modest difference due to the particular implementation of CFR we used and the relatively small number of player actions in FTH. In other games and implementations, later iterations can be orders of magnitude cheaper than early ones, resulting in a much larger advantage to warm starting.

Conclusions and Future Research

We introduced a general method for warm starting RM and CFR in zero-sum games. We proved that after warm starting to T iterations, CFR converges just as quickly as if it had played T iterations of CFR from scratch. Moreover, we proved that this warm start method is "lossless." That is, when warm starting with the average strategy of T iterations of CFR, we can warm start to T iterations.

While other warm start methods exist, they can only be applied in special cases. A benefit of ours is that it is agnostic to the origins of the input strategies. We demonstrated that this can be leveraged by first solving a coarse abstraction and then using its solution to warm start CFR in the full game.

Our warm start method expands the scope and effectiveness of SAEF, in which an abstraction is progressively refined during equilibrium finding. SAEF could previously only refine public actions, due to limitations in warm starting. The method presented in this paper allows SAEF to potentially make arbitrary changes to the abstraction.

Recent research that finds close connections between CFR and other iterative equilibrium-finding algorithms (Waugh and Bagnell 2015) suggests that our techniques may extend beyond CFR as well. There are a number of equilibrium-finding algorithms with better long-term convergence bounds than CFR, but which are not used in practice due to their slow initial convergence (Kroer et al. 2015; Hoda et al. 2010; Nesterov 2005; Daskalakis, Deckelbaum, and Kim 2015). Our work suggests that a similar method of warm starting in these algorithms could allow their faster asymptotic convergence to be leveraged later in the run while CFR is used earlier on.

Acknowledgments

This material is based on work supported by the National Science Foundation under grants IIS-1320620 and IIS-1546752, as well as XSEDE computing resources provided by the Pittsburgh Supercomputing Center.

References

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 *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI)*.

Bowling, M.; Burch, N.; Johanson, M.; and Tammelin, O. 2015. Heads-up limit hold'em poker is solved. *Science* 347(6218):145–149.

Brown, N., and Sandholm, T. 2014. Regret transfer and parameter optimization. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Brown, N., and Sandholm, T. 2015a. Regret-based pruning in extensive-form games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*.

Brown, N., and Sandholm, T. 2015b. Simultaneous abstraction and equilibrium finding in games. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

Brown, N.; Ganzfried, S.; and Sandholm, T. 2015. Hierarchical abstraction, distributed equilibrium computation, and post-processing, with application to a champion no-limit Texas Hold'em agent. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Daskalakis, C.; Deckelbaum, A.; and Kim, A. 2015. Near-optimal no-regret algorithms for zero-sum games. *Games and Economic Behavior* 92:327–348.

Ganzfried, S., and Sandholm, T. 2014. Potential-aware imperfect-recall abstraction with earth mover's distance in imperfect-information games. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Gibson, R. 2014. *Regret Minimization in Games and the Development of Champion Multiplayer Computer Poker-Playing Agents*. Ph.D. Dissertation, University of Alberta.

Gilpin, A., and Sandholm, T. 2005. Optimal Rhode Island Hold'em poker. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 1684–1685. Pittsburgh, PA: AAAI Press / The MIT Press. Intelligent Systems Demonstration.

Gilpin, A., and Sandholm, T. 2007. Lossless abstraction of imperfect information games. *Journal of the ACM* 54(5).

Gilpin, A.; Peña, J.; and Sandholm, T. 2012. First-order algorithm with $\mathcal{O}(\ln(1/\epsilon))$ convergence for ϵ-equilibrium in two-person zero-sum games. *Mathematical Programming* 133(1–2):279–298. Conference version appeared in AAAI-08.

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

Hoda, S.; Gilpin, A.; Peña, J.; and Sandholm, T. 2010. Smoothing techniques for computing Nash equilibria of sequential games. *Mathematics of Operations Research* 35(2):494–512. Conference version appeared in WINE-07.

Jackson, E. 2014. A time and space efficient algorithm for approximately solving large imperfect information games. In *AAAI Workshop on Computer Poker and Imperfect Information*.

Johanson, M.; Waugh, K.; Bowling, M.; and Zinkevich, M. 2011. Accelerating best response calculation in large extensive games. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

Johanson, M.; Bard, N.; Burch, N.; and Bowling, M. 2012. Finding optimal abstract strategies in extensive-form games. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Johanson, M.; Burch, N.; Valenzano, R.; and Bowling, M. 2013. Evaluating state-space abstractions in extensive-form games. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Kroer, C.; Waugh, K.; Kılınç-Karzan, F.; and Sandholm, T. 2015. Faster first-order methods for extensive-form game solving. In *Proceedings of the ACM Conference on Economics and Computation (EC)*.

Lanctot, M.; Waugh, K.; Zinkevich, M.; and Bowling, M. 2009. Monte Carlo sampling for regret minimization in extensive games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, 1078–1086.

Nesterov, Y. 2005. Excessive gap technique in nonsmooth convex minimization. *SIAM Journal of Optimization* 16(1):235–249.

Sandholm, T. 2010. The state of solving large incomplete-information games, and application to poker. *AI Magazine* 13–32. Special issue on Algorithmic Game Theory.

Sandholm, T. 2015. Solving imperfect-information games. *Science* 347(6218):122–123.

Tammelin, O.; Burch, N.; Johanson, M.; and Bowling, M. 2015. Solving heads-up limit texas hold'em. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.

Waugh, K., and Bagnell, D. 2015. A unified view of large-scale zero-sum equilibrium computation. In *Computer Poker and Imperfect Information Workshop at the AAAI Conference on Artificial Intelligence (AAAI)*.

Waugh, K.; Schnizlein, D.; Bowling, M.; and Szafron, D. 2009. Abstraction pathologies in extensive games. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Waugh, K.; Morrill, D.; Bagnell, D.; and Bowling, M. 2015. Solving games with functional regret estimation. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Zinkevich, M.; Bowling, M.; Johanson, M.; and Piccione, C. 2007. Regret minimization in games with incomplete information. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*.

Appendix

Projection Experiment

In this experiment, we demonstrate the accuracy of our projection method for determining the number of iterations to warm start to. We plot the convergence rate of CFR in FTH, and also plot what the convergence rate is projected to be based on data from the first 10 iterations of CFR in FTH. Specifically, it predicts the rate of convergence as $\frac{10.82}{T}$, where T is the number of iterations. Although it bases this projection on just 10 iterations, the results show it to be very accurate even up to 10,000 iterations.

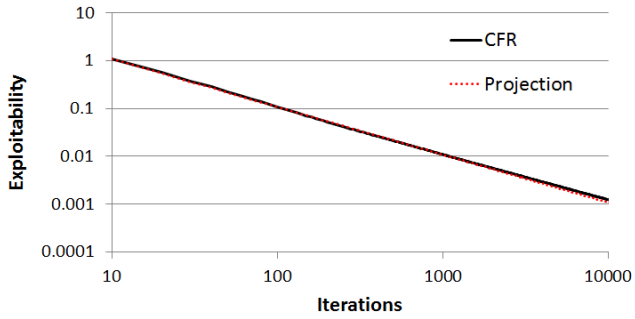

Figure 4: Actual convergence of CFR compared to a projection of convergence based on the first 10 iterations of CFR.

Choice of λ

This experiment demonstrates that performance is not particularly sensitive to the choice of λ. Figure 5 shows the results of warm starting after 500 iterations of CFR (warm starting to $T = 500$) when using various choices of λ (same λ for both players). Performance is virtually identical for $\lambda = 0, 0.05$, and 0.1, though $\lambda = 0.05$ performs the best by a small margin. Nevertheless, performance degrades drastically when choosing a value such as $\lambda = 0.5$.

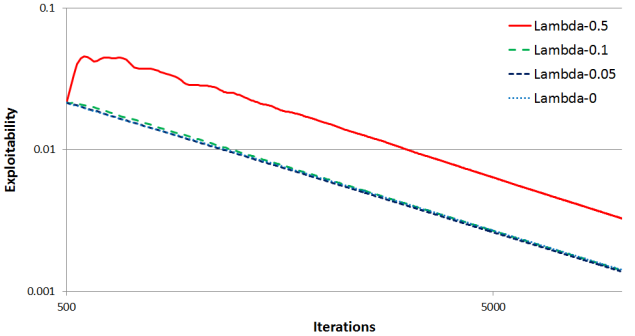

Figure 5: Comparison of different choices for λ when warm starting (using the same λ_i for both players).

Proof of Lemma 1

Proof. From (7) we see that

$$\sum_{a\in A(I)} \left(R^T_+(I,a)\right)^2 \le \sum_{a\in A(I)} \sum_{t=1}^{T} \left(r^t(I,a)\right)^2$$

From (2) and (3), we see that $r^t(I,a) \le \pi^{\sigma^t}_{-i}(I)\Delta(I)$, so

$$\sum_{a\in A(I)} \left(R^T_+(I,a)\right)^2 \le |A(I)|\left(\Delta(I)\right)^2 \sum_{t=1}^{T} \left(\pi^{\sigma^t}_{-i}(I)\right)^2$$

We know $0 \le \pi^{\sigma^t}_{-i}(I) \le 1$. Therefore, $\sum_{t=1}^{T} \left(\pi^{\sigma^t}_{-i}(I)\right)^2 \le \sum_{t=1}^{T} \pi^{\sigma^t}_{-i}(I) \le T\pi^{\bar{\sigma}^T}_{-i}(I)$. Thus, we have

$$\sum_{a\in A(I)} \left(R^T_+(I,a)\right)^2 \le \pi^{\bar{\sigma}^T}_{-i}(I)\left(\Delta(I)\right)^2 |A(I)|T$$

□

Lemma 2

Lemma 2 proves that the growth of substitute regret has the same bound as the growth rate of normal regret. This lemma is used in the proof of Theorem 2.

Lemma 2. *If*

$$\sum_{a\in A} \left(R'^T(I,a)\right)^2_+ \le \left(\pi^{\sigma}_{-i}(I)\right)\left(\Delta(I)\right)^2 |A(I)|T$$

and strategies are chosen in I according to CFR using $R'^{T,T^*}(I,a)$ for all a on every iteration T^*, then $R'^{T,T'}(I) \le \sqrt{\pi^{\sigma'^{T,T'}}_{-i}(I)}\Delta(I)\sqrt{|A(I)|}\sqrt{T+T'}$.

Proof. After T' iterations of CFR, we are guaranteed that $\left(R'^{T,T'}(I,a)\right)^2_+ \le \left(R'^T(I,a)\right)^2_+ + \sum_{t'=1}^{T'} \left(r^{t'}(I,a)\right)^2$. On iteration $t' \le T'$, from (2) and (3), we know that $\left(r^{t'}(I,a)\right)^2 \le \left(\pi^{\sigma^{t'}}_{-i}(I)\Delta(I)\right)^2$. Thus,

$$\left(R'^{T,T'}(I,a)\right)^2_+ \le \left(R'^T(I,a)\right)^2_+ + \sum_{t'=1}^{T'} \left(\pi^{\sigma^{t'}}_{-i}(I)\Delta(I)\right)^2$$

$$\left(R'^{T,T'}(I,a)\right)^2_+ \le \left(T\pi^{\sigma}_{-i}(I) + \sum_{t'=1}^{T'} \left(\pi^{\sigma^{t'}}_{-i}(I)\right)^2\right)\left(\Delta(I)\right)^2$$

Since $0 \le \pi^{\sigma^{t'}}_{-i}(I) \le 1$, we know that $\sum_{t'=1}^{T'} \left(\pi^{\sigma^{t'}}_{-i}(I)\right)^2 \le \sum_{t'=1}^{T'} \pi^{\sigma^{t'}}_{-i}(I)$. Also, $\sum_{t'=1}^{T'} \pi^{\sigma^{t'}}_{-i}(I) = T'\pi^{\bar{\sigma}^{T'}}_{-i}(I)$. So

$$\left(R'^{T,T'}(I,a)\right)^2_+ \le \left(T\pi^{\sigma}_{-i}(I) + T'\pi^{\bar{\sigma}^{T'}}_{-i}(I)\right)\left(\Delta(I)\right)^2$$

$$\left(R'^{T,T'}(I,a)\right)^2_+ \le \left(\pi^{\sigma'^{T,T'}}_{-i}(I)\right)\left(\Delta(I)\right)^2(T+T')$$

Since $R'^{T,T'}(I) \le \sqrt{\sum_{a\in A(I)} \left(R'^{T,T'}(I,a)\right)^2_+}$ so we get

$$R'^{T,T'}(I) \le \sqrt{\pi^{\sigma'^{T,T'}}_{-i}(I)}\Delta(I)\sqrt{|A(I)|}\sqrt{T+T'}$$

□

Lemma 3

Lemma 3 proves that we can use substitute regret to prove convergence to a Nash equilibrium just as we could use normal regret. Specifically, it proves that if average substitute regret is bounded for both players in the whole game, then the strategy profile is as close to a Nash equilibrium as if average normal regret were similarly bounded. This lemma is used in the proof of Theorem 2.

Lemma 3. *Define* $R_i'^{T,T'}$ *as*

$$\max_{\sigma_i^* \in \Sigma_i} \Big(T\big(u_i(\sigma_i^*, \sigma_{-i}) - v_i'^{\sigma}\big) + \sum_{t'=1}^{T'} \big(u_i(\sigma_i^*, \sigma^{t'}) - u_i(\sigma^{t'})\big)\Big) \tag{17}$$

In a two-player zero-sum game, if $v_1'^{\sigma} + v_2'^{\sigma} \leq 0$ *and* $\frac{R'^{T,T'}}{T+T'} \leq \epsilon_i$*, then* $\sigma'^{T,T'}$ *is a* $(\epsilon_1 + \epsilon_2)$*-equilibrium.*

Proof. Since σ_i^* is the same on every iteration,

$$R_i'^{T,T'} = (T+T') \max_{\sigma_i^* \in \Sigma_i} u_i(\sigma_i^*, \sigma_{-i}'^{T,T'}) - T v_i'^{\sigma} - \sum_{t'=1}^{T'} u_i(\sigma^{t'})$$

Since $v_1'^{\sigma} + v_2'^{\sigma} \leq 0$ and $u_1(\sigma^{t'}) = -u_2(\sigma^{t'})$, so

$$\max_{\sigma_1^* \in \Sigma_1} u_1(\sigma_1^*, \sigma_2'^{T,T'}) + \max_{\sigma_2^* \in \Sigma_2} u_2(\sigma_1'^{T,T'}, \sigma_2^*) \leq \epsilon_1 + \epsilon_2$$

$$\max_{\sigma_1^* \in \Sigma_1} u_1(\sigma_1^*, \sigma_2'^{T,T'}) - \min_{\sigma_2^* \in \Sigma_2} u_1(\sigma_1'^{T,T'}, \sigma_2^*) \leq \epsilon_1 + \epsilon_2$$

Since $u_1(\sigma_1'^{T,T'}, \sigma_2'^{T,T'}) \geq \min_{\sigma_2' \in \Sigma_2} u_1(\sigma_1'^{T,T'}, \sigma_2^*)$ so we have $\max_{\sigma_1^* \in \Sigma_1} u_1(\sigma_1^*, \sigma_2'^{T,T'}) - u_1(\sigma_1'^{T,T'}, \sigma_2'^{T,T'}) \leq \epsilon_1 + \epsilon_2$. By symmetry, this is also true for Player 2. Therefore, $\sigma'^{T,T'}$ is a $(\epsilon_1 + \epsilon_2)$-equilibrium. □

Proof of Theorem 2

Proof. After setting T and $v'^{\sigma}(I)$ for every information set I, assume T' iterations were played according to CFR, where on iteration T^*, $\forall I$ $\forall a$ we used substitute regret $R'^{T,T^*}(I,a)$.

We begin with some definitions. Define $v_i^{\sigma}(h) = \pi_{-i}^{\sigma}(h) \sum_{z \in Z} \big(\pi^{\sigma}(h,z) u_i(z)\big)$. Define $D(I)$ to be the information sets of player i reachable from I (including I). Define $\sigma|_{D(I) \to \sigma'}$ to be a strategy profile equal to σ except in the information sets in $D(I)$ where it is equal to σ'. Define $succ_i^{\sigma}(I'|I,a)$ to be the probability that I' is the next information set of player i visited given that the action a was just selected in information set I, and σ is the current strategy. Define $Succ(I,a)$ to be the set of all possible next information sets of player $P(I)$ visited given that action $a \in A(I)$ was just selected in information set I. Define $Succ(I) = \cup_{a \in A(I)} Succ(I,a)$. The *substitute full counterfactual regret* when warm starting from strategy σ and where $i = P(I)$ is

$$R_{full}'^{T,T'}(I) = \max_{\sigma' \in \Sigma_i} \Big(T\big(v^{\sigma|_{D(I) \to \sigma'}}(I) - v_i'^{\sigma}(I)\big) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}|_{D(I) \to \sigma'}}(I) - v^{\sigma^{t'}}(I)\big)\Big) \tag{18}$$

We now prove recursively that

$$R_{full}'^{T,T'}(I) \leq \sum_{I' \in D(I)} \sqrt{\pi_{-i}^{\sigma'^{T,T'}}(I')} \Delta(I') \sqrt{|A(I')|} \sqrt{T+T'} \tag{19}$$

We define the *level* of an information set as follows. Any information set I such that $Succ(I) = \emptyset$ is level 1. Let ℓ be the maximum level of any $I' \in Succ(I)$. The level of I is $\ell + 1$.

First, consider an information set I of level 1. Then there are no Player i information sets following I, so

$$R_{i,full}'^{T,T'}(I) = \max_{a \in A(I)} \Big(T\big(v^{\sigma}(I,a) - v'^{\sigma}(I)\big) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}}(I,a) - v^{\sigma^{t'}}(I)\big)\Big)$$

Since there are no further player i actions in this case, so $v^{\sigma}(I,a) = v'^{\sigma}(I,a)$. Therefore, $R_{full}'^{T,T'}(I) = R'^{T,T'}(I)$. By Lemma 2, (19) holds.

Now assume that (19) holds for all I' where the level of I' is at most ℓ. We prove (19) holds for all I with level $\ell + 1$ where $i = P(I)$.

From Lemma 2, we know that

$$T v'^{\sigma}(I) + \sum_{t'=1}^{T'} v^{\sigma^t}(I) \geq \max_{a \in A(I)} \Big(T v'^{\sigma}(I,a) + \sum_{t'=1}^{T'} v^{\sigma^t}(I,a) - \sqrt{\pi_{-i}^{\sigma'^{T,T'}}(I)} \Delta(I) \sqrt{|A(I)|} \sqrt{T+T'} \Big) \tag{20}$$

$$T v'^{\sigma}(I) + \sum_{t'=1}^{T'} v^{\sigma^t}(I) \geq \max_{a \in A(I)} \Big(T \sum_{h \in I} \sum_{h' \in Succ_i(h \cdot a)} v_i'^{\sigma}(h') + \sum_{t'=1}^{T'} v^{\sigma^t}(I,a) - \sqrt{\pi_{-i}^{\sigma'^{T,T'}}(I)} \Delta(I) \sqrt{|A(I)|} \sqrt{T+T'} \Big) \tag{21}$$

We also know that for any σ,

$$\max_{\sigma' \in \Sigma_i} v^{\sigma|_{D(I) \to \sigma'}}(I) = \max_{a \in A(I)} \max_{\sigma_i' \in \Sigma_i} \sum_{h \in I} \Big(\sum_{z \in Z : z \in Succ_i(h \cdot a)} v_i^{\sigma}(z) + \sum_{h' \notin Z : h' \in Succ_i(h \cdot a)} v^{\sigma|_{D(I(h')) \to \sigma'}}(h') \Big) \tag{22}$$

Since for any $z \in Z$, $v_i^{\sigma}(z) = v_i'^{\sigma}(z)$, so combining (21)

and (22) we get

$$\max_{\sigma' \in \Sigma_i} \Big(T\big(v^{\sigma|_{D(I)\to\sigma'}}(I) - v_i'^{\sigma}(I)\big) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}|_{D(I)\to\sigma'}}(I) - v^{\sigma^{t'}}(I)\big)\Big) \le$$
$$\max_{\sigma_i' \in \Sigma_i} \Big(T \sum_{h \in I} \sum_{h' \notin Z : h' \in Succ_i(h\cdot a)} \big(v^{\sigma|_{D(I(h'))\to\sigma'}}(h') - v_i'^{\sigma}(h')\big) + \sum_{t'=1}^{T'} \sum_{h \in I} \sum_{h' \notin Z : h' \in Succ_i(h\cdot a)} \big(v^{\sigma^{t'}|_{D(I(h'))\to\sigma'}}(h') - v_i^{\sigma^{t'}}(h')\big) + \sqrt{\pi_{-i}^{\sigma' T,T'}(I)}\Delta(I)\sqrt{|A(I)|}\sqrt{T+T'} \quad (23)$$

Since we sum over only $h' \in Succ_i(h \cdot a)$ where $h' \notin Z$, this becomes

$$\max_{\sigma' \in \Sigma_i} \Big(T\big(v^{\sigma|_{D(I)\to\sigma'}}(I) - v_i'^{\sigma}(I)\big) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}|_{D(I)\to\sigma'}}(I) - v^{\sigma^{t'}}(I)\big)\Big) \le$$
$$\max_{\sigma_i' \in \Sigma_i} \Big(T \sum_{I' \in Succ(I,a)} \big(v^{\sigma|_{D(I')\to\sigma'}}(I') - v_i'^{\sigma}(I')\big) + \sum_{t'=1}^{T'} \sum_{I' \in Succ(I,a)} \big(v^{\sigma^{t'}|_{D(I')\to\sigma'}}(I') - v_i^{\sigma^{t'}}(I')\big) + \sqrt{\pi_{-i}^{\sigma' T,T'}(I)}\Delta(I)\sqrt{|A(I)|}\sqrt{T+T'} \quad (24)$$

From the recursion assumption, for any $I' \in Succ(I,a)$,

$$\max_{\sigma' \in \Sigma_i} \Big(T\big(v^{\sigma|_{D(I)\to\sigma'}}(I') - v_i'^{\sigma}(I')\big) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}|_{D(I')\to\sigma'}}(I') - v^{\sigma^{t'}}(I')\big)\Big) \le \sum_{I'' \in D(I')} \sqrt{\pi_{-i}^{\sigma' T,T'}(I'')}\Delta(I'')\sqrt{|A(I'')|}\sqrt{T+T'} \quad (25)$$

Therefore,

$$\max_{\sigma' \in \Sigma_i} \Big(T\big(v^{\sigma|_{D(I)\to\sigma'}}(I) - v_i'^{\sigma}(I)\big) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}|_{D(I)\to\sigma'}}(I) - v^{\sigma^{t'}}(I)\big)\Big) \le \sum_{I' \in Succ(I,a)} \sum_{I'' \in D(I')} \sqrt{\pi_{-i}^{\sigma' T,T'}(I'')}\Delta(I'')\sqrt{|A(I'')|}\sqrt{T+T'} + \sqrt{\pi_{-i}^{\sigma' T,T'}(I)}\Delta(I)\sqrt{|A(I)|}\sqrt{T+T'} \quad (26)$$

Since $Succ(I,a) \subseteq Succ(I)$ and since $D(I) = \cup_{I' \in Succ(I)} D(I') \cup \{I\}$, so

$$\max_{\sigma' \in \Sigma_i} \Big(T\big(v^{\sigma|_{D(I)\to\sigma'}}(I) - v_i'^{\sigma}(I)\big) + \sum_{t'=1}^{T'} \big(v^{\sigma^{t'}|_{D(I)\to\sigma'}}(I) - v^{\sigma^{t'}}(I)\big)\Big) \le \sum_{I' \in D(I)} \sqrt{\pi_{-i}^{\sigma' T,T'}(I')}\Delta(I')\sqrt{|A(I')|}\sqrt{T+T'} \quad (27)$$

Therefore, (19) holds by recursion.

Define $R_i'^{T,T'}$ according to (17). If $P(\emptyset) = i$, then (19) implies

$$R_i'^{T,T'} \le \sum_{I \in \mathcal{I}_i} \sqrt{\pi_{-i}^{\sigma' T,T'}(I)}\Delta(I)\sqrt{|A(I)|}\sqrt{T+T'} \quad (28)$$

If $P(\emptyset) \neq i$, then we could simply add a Player i information set at the beginning of the game with a single action. Therefore, (28) holds for every player i. Since $v_1'^{\sigma} + v_2'^{\sigma} \le 0$ by construction, so we can applying Lemma 3 using (28), and thereby see that Theorem 2 holds. □

Proof of Corollary 1

Proof. After T iterations of CFR, for every information set I we could clearly assign $v'^{\bar{\sigma}^T}(I) = \frac{1}{T}\sum_{t=1}^{T} v^{\sigma^t}(I)$ in order to satisfy Theorem 2, since this would set regrets to exactly what they were before. From (8) we see this choice of $v'^{\bar{\sigma}^T}(I)$ satisfies (16). We instead choose $v'^{\bar{\sigma}^T}(I) \le \frac{1}{T}\sum_{t=1}^{T} v^{\sigma^t}(I)$, where $v'^{\bar{\sigma}^T}(I)$ still satisfies (16). Since $v'^{\bar{\sigma}^T}(I) \le \frac{1}{T}\sum_{t=1}^{T} v^{\sigma^t}(I)$ for every information set I, so from (15) we know $v_i'^{\bar{\sigma}^T} \le \frac{1}{T}\sum_{t=1}^{T} u_i(\sigma^t)$. Therefore, $v_1'^{\bar{\sigma}^T} + v_2'^{\bar{\sigma}^T} \le 0$ and we can apply Theorem 2 to warm start to T iterations. □