Modern Algorithmic Game Theory

Martin Schmid
Department of Applied Mathematics
Charles University in Prague

December 10, 2025

Counterfactual Regret Minimization

Regret Minimization In Extensive-Form Games

- We have already seen how we can apply regret minimization to normal-form games
- The resulting self-play algorithm had a particularly simple form and minimizing the external regret guaranteed the convergence to a Nash equilibrium in two-player zero-sum games
- We will now extend regret minimization to extensive-form games in the form of **counterfactual regret minimization**
- The core idea is to **decompose** the full regret into individual **per-information-state** regrets that can be minimized independently
- We will then show that minimizing these additive regrets also minimizes the full external regret of the game
- CFR was introduced in the paper *Regret Minimization in Games with Incomplete Information* in 2007

Counterfactual Regret Minimization

- We will first recall the definitions of counterfactual reach probability and counterfactual state and state-action value functions
- These notions will allow for a particularly easy and intuitive definition of counterfactual regrets and the whole CFR algorithm

Counterfactual Reach Probability

- Given a strategy profile π, we define the counterfactual reach probability of history $h \in \mathcal{H}$ as

$$P^{\pi}_{-i}(h) = \prod_{j \in \mathcal{N} \cup \{c\} \setminus \{i\}} P^{\pi}_{j}(h) = \prod_{h'a \sqsubseteq h :\, p(h') \neq i} \pi(h', a)$$

- It is the probability of reaching history h when player i **attempts** to reach that particular history and all other players stick to their strategies.
- In other words, at every history h' that is a prefix of history h, where player i has to act, they place all of their probability mass on the particular action a that is on the path to h, i.e. $\pi_i(h', a) = 1$, where $h'a \sqsubseteq h$ and $p(h') = i$.

Counterfactual Value Functions

- The **counterfactual state-action** value $q^{\pi}_{i,c}(s_i, a)$ of an information state $s_i \in \mathcal{S}_i$ and action $a \in \mathcal{A}_i(s_i)$ is defined as

$$q^{\pi}_{i,c}(s_i, a) = \sum_{h \in \mathcal{H}(s_i)} P^{\pi}_{-i}(h) q^{\pi}_i(h, a)$$

- The **counterfactual state** value $v^{\pi}_{i,c}(s_i)$ of an information state $s_i \in \mathcal{S}_i$ is defined as

$$v^{\pi}_{i,c}(s_i) = \sum_{a \in \mathcal{A}_i(s_i)} \pi_i(s_i, a) q^{\pi}_{i,c}(s_i, a)$$

Counterfactual Regret

- Given a sequence of strategy profiles $\pi^1, \ldots, \pi^t$, we can use the counterfactual state and state-action values to define the counterfactual regrets as follows:

$$R_i^t(s, a) = \sum_{k=1}^{t} \left(q_{i,c}^{\pi^k}(s, a) - v_{i,c}^{\pi^k}(s) \right)$$

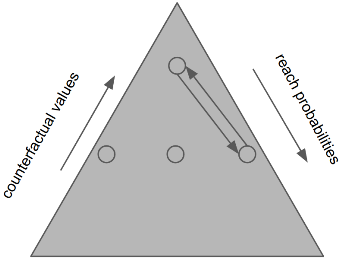
- Defining $R_i^t(s) = \max_a R_i^t(s, a)$ and $R_i^t(s)^+ = \max(R_i^t(s), 0)$, it can be shown that the overall regret R_i^t is **upper bounded** by the sum of positive counterfactual regrets $R_i^t(s)^+$, i.e. $R_i^t \leq \sum_{s \in \mathcal{S}_i} R_i^t(s)^+$
- This means that we can minimize counterfactual regrets in each information state independently and we will also minimize the overall regret

Convergence

- If a Hannan consistent regret minimizer (such as Regret matching) is used in each information state, the average strategy converges to a Nash Equilibrium
- This result relies on the Folk Theorem and the decomposition of regret shown previously
- Note that, in order to converge correctly, the average behavioral strategy $\bar{\pi}$ must be properly weighted by the player's reach probability $P_i^{\pi}(s)$

CFR Implementation

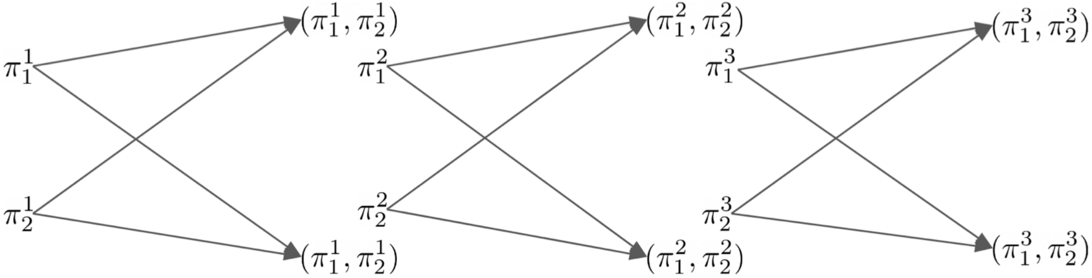
- CFR is typically implemented as a recursive traversal of the game tree
- In the forward pass, we pass the current reach probabilities down the tree
- In the backward pass, we compute the counterfactual values at the leaves and pass them up to compute regrets

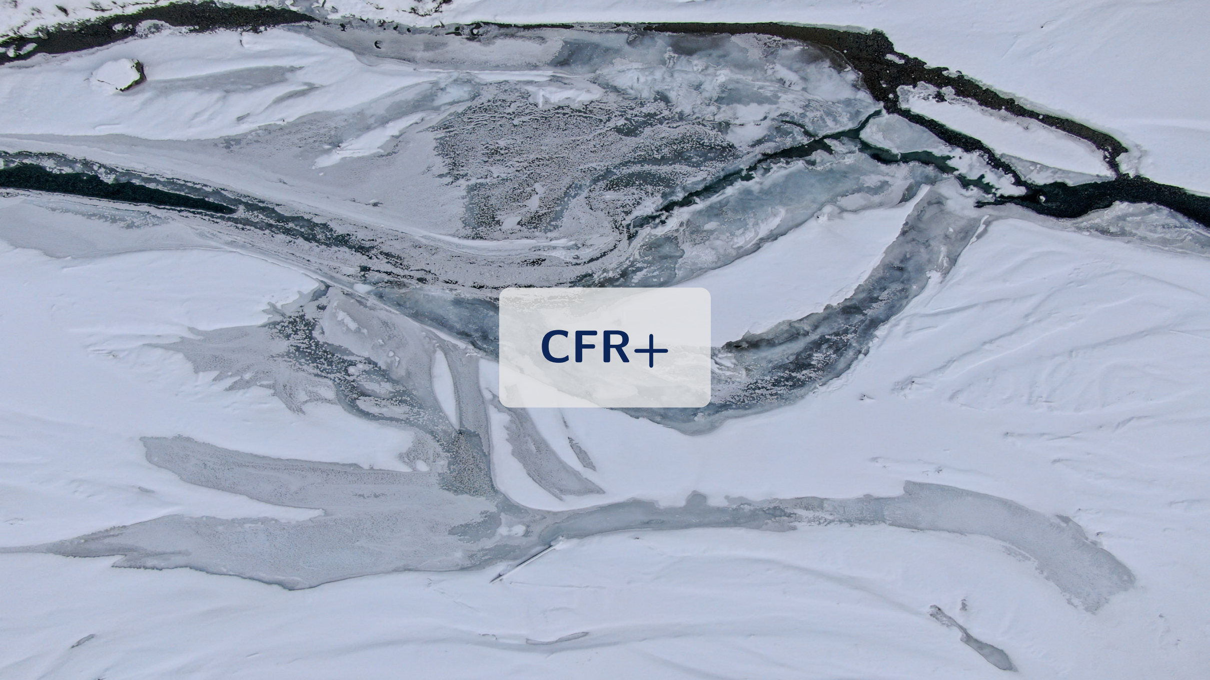


function $\textsc{ComputeValues}(s_{pub} \in \mathcal{S}_{pub},\ d_1 \in \Delta\mathcal{S}_1(s_{pub}),\ d_2 \in \Delta\mathcal{S}_2(s_{pub}))$

CFR Implementation

- CFR employs so called **simultaneous** updates, where both players update their policies at the same time and the regret computation is **symmetrical**




CFR+

CFR+

- CFR+ is an extension of the classical CFR algorithm with an impressive empirical performance
- The algorithm was proposed in *Solving Large Imperfect Information Games Using CFR+* published in 2014
- It modifies the vanilla CFR algorithm in three ways:
    1. **Regret Matching Plus**
    2. **Alternating updates**
    3. **Linear strategy averaging**
- It drastically improved the state-of-the-art performance and has been used to solve Limit Texas Hold'em Poker
- Limit Texas Hold'em Poker is **the largest imperfect information that has been solved to this day!**

Regret Matching Plus (RM+)

- **Standard Regret Matching (RM):**
  - Accumulates regret normally: $R^t(a) = R^{t-1}(a) + r^t(a)$
  - Plays proportional to positive regret: $\pi^{t+1}(a) \propto R^t(a)^+$
  - *Issue:* Negative regrets can accumulate indefinitely

- **Regret Matching Plus (RM+):**
  - **Floors** the cumulative regret at zero **at every iteration**
  - Update rule: $R^t(a) = \max(0, R^{t-1}(a) + r^t(a))$
  - Strategy is proportional to the stored values: $\pi^{t+1}(a) \propto R^t(a)$

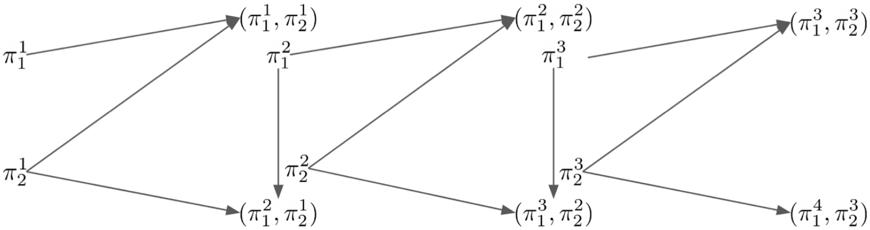
- Like RM, RM+ guarantees convergence to a Nash Equilibrium with an error bound of $\mathcal{O}(1/\sqrt{T})$.

Why RM+ Works

- Suppose an action a is bad for 100 turns; it accumulates a large negative regret (e.g., $R^t(a) = -100$)
- If the opponent changes their strategy and a suddenly becomes the **best** action, standard RM will wait until $R^t(a)$ becomes positive again
- This causes a significant **lag** in adaptation; the agent keeps ignoring the optimal action until the negative debt is cleared
- Because RM+ resets negative values to 0 immediately, it has no "memory" of how bad an action was, only that it was bad
- If a previously bad action becomes optimal, RM+ can start playing it **immediately** in the next iteration
- This makes the algorithm highly reactive to changes in the opponent's strategy

Alternating Updates

- CFR+ uses **alternating updates**:
  - Player 1 updates their strategy to π_1^{t+1}
  - Player 2 immediately observes π_1^{t+1} and updates their strategy to π_2^{t+1} using this new information
- This seemingly small change significantly speeds up the propagation of information through the game tree



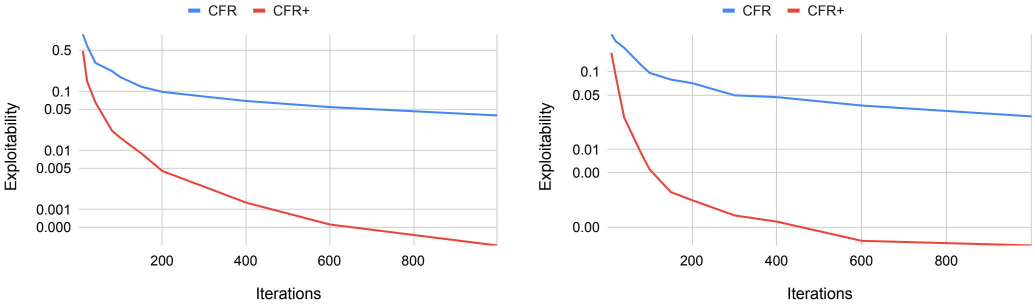
Linear Strategy Averaging

- In standard CFR, the average strategy is a uniform average over all iterations (weighted only by reach probabilities)
- In CFR+, we assume that later iterations are closer to the equilibrium and should be weighted more heavily
- CFR+ uses **linear weighting**, where the strategy at iteration t is weighted by t

$$\bar{\pi}^t \propto \sum_{k=1}^{t} k \cdot \pi^k$$

- This allows the average strategy to converge much faster to a Nash equilibrium compared to uniform averaging

Empirical Performance



(a) Leduc Poker (b) A Small Graph Chase Game

Figure: Convergence of CFR and CFR+

Week 9 Homework

You can find more detailed descriptions of homework tasks in the GitHub repository.

1. Counterfactual Regret Minimization (CFR)
2. CFR+