CSCE 631 — Algorithmic Game Theory

Lecture 13: Counterfactual Regret Minimization (CFR)

Alan Kuhnle

Summer 2026

Texas A&M University

Today's Plan

Motivation

Counterfactual Values

Counterfactual Regret

The CFR Algorithm

Convergence

Alternating Updates and Practical Considerations

Summary

Motivation

Why CFR?

- Sequence-form LP solves two-player zero-sum games in polynomial time . . . but the LP can still be enormous
- Texas Hold'em: $\sim 10^{14}$ information sets
- Even with abstraction, LP solvers hit memory limits
- Need: an **iterative** algorithm that:
  - Converges to a Nash equilibrium
  - Has low per-iteration cost
  - Can run on games with millions of information sets

CFR (Zinkevich et al., 2007)

Counterfactual Regret Minimization decomposes the game-level regret into *local* regrets at each information set, then applies regret matching independently at each one. Converges to NE in two-player zero-sum games.

Regret Minimization Recap

Recall from online learning (Lecture 5–6):

- A player repeatedly chooses action a^t from set A
- Observes loss (or utility) after each round
- **External regret** after T rounds:

$$R^T = \max_{a \in A} \sum_{t=1}^{T} u(a, s^t) - \sum_{t=1}^{T} u(a^t, s^t)$$

- **Regret matching** (Hart and Mas-Colell, 2000) achieves $R^T = O(\sqrt{T})$
- If both players in a zero-sum game use regret-minimizing strategies, the average strategy profile converges to a NE

Challenge: In extensive-form games, the action space is the set of *strategies* — exponentially large!

Counterfactual Values

Reach Probabilities

Fix a strategy profile $\sigma = (\sigma_1, \sigma_2)$ and a node h in the game tree.

Definition

The **reach probability** of node h decomposes as:

$$\pi^\sigma(h) = \pi_1^\sigma(h) \cdot \pi_2^\sigma(h) \cdot \pi_c(h)$$

where $\pi_i^\sigma(h)$ is the product of player i's action probabilities on the path to h, and $\pi_c(h)$ is the product of chance probabilities.

- **Counterfactual reach:** the probability of reaching h if player i "tries to reach h" (plays to reach it):

$$\pi_{-i}^\sigma(h) = \pi_c(h) \cdot \prod_{j \neq i} \pi_j^\sigma(h)$$

- This is the reach probability excluding player i's contribution

Counterfactual Value

Definition

The **counterfactual value** of an information set I for player i under strategy profile σ is:

$$v_i^\sigma(I) = \sum_{h \in I} \pi_{-i}^\sigma(h) \sum_{z \in Z_h} \pi^\sigma(h, z)\, u_i(z)$$

where Z_h is the set of terminal nodes reachable from h, and $\pi^\sigma(h, z)$ is the probability of reaching z from h under σ.

- Intuition: expected payoff at I *weighted by the probability that player i's opponents and chance bring the game to I*
- Player i's own contribution to reaching I is excluded
- This "counterfactual" weighting is what enables the decomposition

Counterfactual Value of an Action

Definition

The counterfactual value of action a at information set I:

$$v_i^\sigma(I, a) = \sum_{h \in I} \pi_{-i}^\sigma(h) \sum_{z \in Z_{h \cdot a}} \pi^\sigma(h \cdot a, z)\, u_i(z)$$

where $h \cdot a$ denotes the child node reached by taking action a at h.

We have the natural decomposition:

$$v_i^\sigma(I) = \sum_{a \in A(I)} \sigma_i(a \mid I) \cdot v_i^\sigma(I, a)$$

The counterfactual value at I is the weighted average of action values.

Counterfactual Regret

Instantaneous Counterfactual Regret

Definition

At iteration t, the **instantaneous counterfactual regret** of not playing action a at information set I is:

$$r_i^t(I, a) = v_i^{\sigma^t}(I, a) - v_i^{\sigma^t}(I)$$

- Measures: how much better action a would have been compared to the current strategy at I (in counterfactual terms)
- Positive $r_i^t(I, a)$: we "regret" not playing a more
- Negative $r_i^t(I, a)$: we are glad we did not play a more

Cumulative Counterfactual Regret

Definition

The **cumulative counterfactual regret** for action a at I after T iterations is:

$$R_i^T(I,a) = \sum_{t=1}^{T} r_i^t(I,a) = \sum_{t=1}^{T} \left[v_i^{\sigma^t}(I,a) - v_i^{\sigma^t}(I) \right]$$

The positive part: $R_i^{T,+}(I,a) = \max(R_i^T(I,a),\ 0)$.

Key Property

The overall regret of player i in the game is bounded by the sum of local positive regrets:

$$R_i^T \leq \sum_{I \in \mathcal{I}_i} \max_{a \in A(I)} R_i^{T,+}(I,a)$$

The CFR Algorithm

CFR Decomposition Theorem

Theorem (Zinkevich et al., 2007)

For a two-player game with perfect recall, the overall external regret of player i satisfies:

$$R_i^T \leq \sum_{I \in \mathcal{I}_i} R_i^{T,+}(I)$$

where $R_i^{T,+}(I) = \max_{a \in A(I)} R_i^{T,+}(I, a)$ is the maximum positive counterfactual regret at information set I.

Consequence: if we can minimize regret *independently* at each information set, we minimize overall regret.

- At each I: use regret matching over $|A(I)|$ actions (a small local problem)
- The game-level regret is the sum of $|\mathcal{I}_i|$ local regrets, each $O(\sqrt{T})$
- Overall: $R_i^T = O(|\mathcal{I}_i| \cdot \sqrt{T})$

Regret Matching at Each Information Set

At each iteration t, for each information set $I \in \mathcal{I}_i$:

1. Compute cumulative regrets $R_i^{t-1}(I, a)$ for each $a \in A(I)$
2. Set the strategy proportional to positive regrets:

$$\sigma_i^t(a \mid I) = \begin{cases} \dfrac{R_i^{t-1,+}(I, a)}{\sum_{a'} R_i^{t-1,+}(I, a')} & \text{if } \sum_{a'} R_i^{t-1,+}(I, a') > 0 \\ \dfrac{1}{|A(I)|} & \text{otherwise (uniform)} \end{cases}$$

3. After computing counterfactual values under σ^t, update:

$$R_i^t(I, a) \leftarrow R_i^{t-1}(I, a) + r_i^t(I, a)$$

CFR Pseudocode

Algorithm: Vanilla CFR

Input: Extensive-form game Γ, number of iterations T

Output: Approximate Nash equilibrium $\bar{\sigma}$

Initialize: $R_i^0(I, a) \leftarrow 0$, $S_i(I, a) \leftarrow 0$ for all i, I, a

for $t = 1, \ldots, T$ **do**

Compute σ^t via regret matching at each info set

for each player i **do**

Traverse game tree, compute $v_i^{\sigma^t}(I, a)$ for all I, a

Update $R_i^t(I, a) \leftarrow R_i^{t-1}(I, a) + v_i^{\sigma^t}(I, a) - v_i^{\sigma^t}(I)$

Accumulate: $S_i(I, a) \leftarrow S_i(I, a) + \pi_i^{\sigma^t}(I) \cdot \sigma_i^t(a \mid I)$

end for

end for

Return: $\bar{\sigma}_i(a \mid I) = S_i(I, a) / \sum_{a'} S_i(I, a')$

Important Detail: Average Strategy

- The *current* strategy σ^t does **not** converge (it oscillates)
- The **average strategy** $\bar{\sigma}^T$ converges to a NE:

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

- The weighting by $\pi_i^{\sigma^t}(I)$ accounts for how often player i actually reaches I under their own strategy
- In practice, we maintain running sums $S_i(I, a)$ and normalize at the end

Convergence

Convergence to Nash Equilibrium

Theorem

In a two-player zero-sum game, if both players use CFR for T iterations, the average strategy profile $\bar{\sigma}^T = (\bar{\sigma}_1^T, \bar{\sigma}_2^T)$ is a 2ε-Nash equilibrium where:

$$\varepsilon = \frac{\Delta_i \sqrt{|\mathcal{A}_i|}}{T} \cdot \sum_{I \in \mathcal{I}_i} 1 = O\left(\frac{|\mathcal{I}| \cdot \Delta \cdot \sqrt{|A_{\max}|}}{\sqrt{T}} \right)$$

with Δ_i the range of payoffs and $|A_{\max}|$ the maximum number of actions at any info set.

- **Rate:** $O(1/\sqrt{T})$ convergence to NE (exploitability)
- To achieve ε-NE: need $T = O(|\mathcal{I}|^2/\varepsilon^2)$ iterations
- Each iteration: one full traversal of the game tree

Why Does It Work? Intuition

1. **Regret matching** at each info set guarantees local regret $O(\sqrt{T})$
2. **CFR decomposition:** overall regret $\leq$ sum of local regrets
3. **Regret bound $\Rightarrow$ NE:** in zero-sum games, if both players have average regret $\leq \varepsilon$, the average strategies form a 2ε-NE

Critical insight: the counterfactual weighting ensures that regrets at different information sets compose correctly, even though the local problems are coupled through the tree structure.

Non-Zero-Sum Games

CFR still converges to a *correlated equilibrium* (not NE in general) for non-zero-sum games. For NE in general games, additional techniques are needed.

Alternating Updates and Practical Considerations

Alternating Updates

Vanilla CFR: both players update simultaneously each iteration.

Alternating CFR:

- On odd iterations: only Player 1 updates regrets (Player 2 fixed)
- On even iterations: only Player 2 updates regrets (Player 1 fixed)
- Reduces computation per iteration by $\sim 50\%$
- Same convergence rate $O(1/\sqrt{T})$
- In practice, alternating updates converge *faster* per unit of computation

Why Alternating is Better

When only one player updates, the other's strategy is fixed, so the counterfactual values are deterministic (no moving target). This reduces variance in the regret estimates.

CFR+: Regret Matching+

CFR+ (Tammelin, 2014): a simple modification with dramatic speedups.

- Replace cumulative regret update:

$$R_i^t(I, a) \leftarrow \max\big(R_i^{t-1}(I, a) + r_i^t(I, a),\ 0\big)$$

  i.e., clamp negative regrets to zero immediately
- Use weighted averaging: $\bar{\sigma}^T$ weights iteration t by t (later iterations contribute more)
- **Empirically:** converges $\sim 10\times$ faster than vanilla CFR
- **Theoretically:** same $O(1/\sqrt{T})$ worst-case guarantee, but much better in practice

CFR+ was used to solve heads-up limit Texas Hold'em (Bowling et al., Science 2015).

Vanilla CFR: Complexity per Iteration

- Each iteration requires a **full traversal** of the game tree
- **Time per iteration:** $O(|V|)$ (visit every node once)
- **Space:** $O(\sum_I |A(I)|)$ to store cumulative regrets and strategy sums
- For T iterations: total time $O(T \cdot |V|)$

	Sequence-Form LP	**Vanilla CFR**
Time	$O(\|V\|^{3.5})$ (interior point)	$O(T \cdot \|V\|)$
Space	$O(\|V\|^2)$ (LP matrix)	$O(\|V\|)$
Exact?	Yes	ε-approximate

Tradeoff: CFR uses far less memory; LP gives exact solution. For large games, memory is the binding constraint $\Rightarrow$ CFR wins.

Practical Impact

Milestones enabled by CFR and its variants:

- **2015:** Heads-up limit Hold'em essentially solved (Bowling et al., CFR+ with 10^{12} info sets)
- **2017:** Libratus defeats top human pros in heads-up no-limit Hold'em (Brown and Sandholm, subgame solving + CFR)
- **2019:** Pluribus: first superhuman multiplayer poker AI (Brown and Sandholm, MCCFR + depth-limited search)

Beyond Poker

CFR-based methods apply to any sequential imperfect-information game: security games, negotiation, medical decision-making, adversarial planning.

Summary

Summary

- **CFR** decomposes overall regret into counterfactual regrets at each information set
- **Counterfactual value:** expected payoff weighted by opponents' and chance's reach probability
- **CFR decomposition theorem:** overall regret $\leq$ sum of local positive regrets
- **Algorithm:** regret matching at each info set; full tree traversal per iteration
- **Convergence:** $O(1/\sqrt{T})$ to NE in two-player zero-sum games; the *average* strategy converges (not the current strategy)
- **Alternating updates** halve computation; **CFR+** gives $\sim 10\times$ practical speedup
- **Complexity:** $O(|V|)$ time and space per iteration

Next time: Monte Carlo Tree Search and sampling-based CFR variants for even larger games.