Lecture 14—Learning in Extensive Form Games

Iosif Sakos

November 11, 2025

Abstract

These notes study learning in extensive form games (EFGs) through the counterfactual regret minimization (CFR) algorithm. We define reach probabilities and counterfactual utility at information sets. Building on these notions, we introduce counterfactual regret and present the CFR update rule, in which behavioral strategies are adjusted via regret matching at each information set. We then state the CFR decomposition theorem, which shows that controlling local counterfactual regret at every information set is sufficient to bound a player's regret in the game and implies convergence to approximate Nash equilibria (NE) in EFGs with perfect recall [1].

Contents

1 Counterfactual regret minimization

Learning in extensive form games (EFGs) often employs the counterfactual regret minimization (CFR) algorithm [1]. CFR is an iterative method for computing approximate Nash equilibria (NE) in EFGs with imperfect information. The key idea of CFR is to minimize a player's regret for not choosing alternative actions, where each action is evaluated based on its counterfactual utility at the information sets where it is available.

1.1 Preliminaries

Summary of EFG foundations. Recall that an EFG is a game tree. The internal nodes, or histories, of the game tree are decision points for an acting player, and the outgoing arcs represent the possible actions the acting player can take at each history. Information sets group together histories that are *indistinguishable* to the player due to imperfect information. A terminal history corresponds to *a complete sequence of actions* leading to the end of the game, with associated payoffs for each player. Although EFGs may include *chance nodes* to model stochastic events, for simplicity we focus on game trees without chance nodes. All results in these notes extend to EFGs with chance nodes by treating nature as an additional player with fixed distributions over its histories.

Formally, for an n-player EFG with perfect recall (and no chance nodes), we use $\mathcal{H}$ to denote the set of histories. For each player $i \in [\![n]\!]$, we use $\mathcal{J}_i \subset 2^{\mathcal{H}}$ to denote the set of information sets of player i, and for each information set $\mathcal{I} \in \mathcal{J}_i$ we use $\mathcal{A}_{\mathcal{I}}$ to denote the set of actions available to player i at $\mathcal{I}$. We use $\mathcal{Z} \subseteq \mathcal{H}$ to denote the set of terminal histories, and for each terminal history $z \in \mathcal{Z}$, we write $u_i(z)$ for the payoff of player i at z. To ease the notation, we assume that the action sets of each player's information sets are disjoint; that is, $\mathcal{A}_{\mathcal{I}} \cap \mathcal{A}_{\mathcal{I}'} = \emptyset$ for all $i \in [\![n]\!]$ and $\mathcal{I}, \mathcal{I}' \in \mathcal{J}_i$ with $\mathcal{I} \neq \mathcal{I}'$.

Summary of EFG strategy representations. In EFGs, strategies have both normal-form representation and behavioral-form representation. In the normal-form representation, strategies correspond to mixed strategies of an induced normal-form game; thus, by Nash's theorem [2], a mixed-strategy NE always exists in EFGs. Nonetheless, the normal-form representation is often impractical due to the exponential growth of the pure-strategy space with respect to the size of the game tree.

In contrast, in the behavioral-form representation, strategies are behavioral strategies, assigning probabilities to actions at each information set. In other words, a behavioral strategy of each player is a product distribution over the player's available actions at each information set. Kuhn's theorem [3] establishes the equivalence between mixed strategies and behavioral strategies in EFGs with perfect recall, i.e., EFGs in which each player remembers their past actions. Hence, a behavioral-strategy NE also exists in such games.

For each player $i \in [\![n]\!]$, we use

$$\mathcal{B}_i = \prod_{\mathcal{I} \in \mathcal{J}_i} \Delta(\mathcal{A}_{\mathcal{I}}) \tag{1}$$

to denote the set of all behavioral strategies of player i, where $\Delta(\mathcal{A}_{\mathcal{I}})$ is the probability simplex over the action set $\mathcal{A}_{\mathcal{I}}$ at information set $\mathcal{I}$. Moreover, we use

$$\mathcal{B} = \prod_{i=1}^{n} \mathcal{B}_i \tag{2}$$

to denote the set of all behavioral-strategy profiles. Finally, for each player $i \in [\![n]\!]$, a behavioral strategy $b_i \in \mathcal{B}_i$, and action $\alpha \in \mathcal{A}_{\mathcal{I}}$ with $\mathcal{I} \in \mathcal{J}_i$, we write $b_i(\alpha)$ to denote the probability assigned by b_i to action α at information set $\mathcal{I}$. This is well-defined because we assume that the action sets of each player's information sets are disjoint.

1.2 Action sequences and perfect recall

For each EFG and each history $h \in \mathcal{H}$ in the game tree of the EFG, there exists a unique sequence of actions taken by the players to reach h from the root of the game tree (since the game tree is a tree). For each

history $h \in \mathcal{H}$ and each player $i \in [\![n]\!]$, we use

$$q_h^i = (q_{h,1}^i, \ldots, q_{h,m_h^i}^i), \tag{3}$$

to denote this *unique* sequence of actions taken by player i to reach the history h, where m_h^i denotes the length of this sequence. Furthermore, we use

$$p_h^i = (p_{h,1}^i, \ldots, p_{h,m_h^i}^i) \tag{4}$$

to denote the induced sequence of histories, where $p_{h,k}^i$ denotes the history at which action $q_{h,k}^i$ was taken. Finally, we use

$$\mathcal{J}_{i,h} = \{\mathcal{I} \in \mathcal{J}_i \mid \exists k \in [\![m_h^i]\!] \text{ such that } p_{h,k}^i \in \mathcal{I}\} \tag{5}$$

to denote the set of information sets of player i that are visited in the history sequence p_h^i.

As an example, consider the game tree in Figure 1.

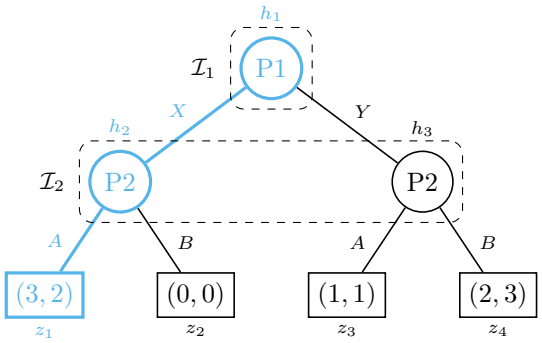

Figure 1: A simple game tree. The path from the root history h_1 to the leftmost terminal history z_1 is highlighted in sky blue.

Consider the leftmost terminal history z_1 with payoffs $(3, 2)$, which is highlighted in sky blue. The action sequence of Player 1 (P1) to reach z_1 is $q_{z_1}^1 = (X)$, with induced history sequence $p_{z_1}^1 = (h_1)$, where h_1 is the root history, while the action sequence of Player 2 (P2) to reach z_1 is $q_{z_1}^2 = (A)$, with induced history sequence $p_{z_1}^2 = (h_2)$. Finally, the set of information sets visited by Player 1 in $p_{z_1}^1$ is $\mathcal{J}_{1,z_1} = \{\mathcal{I}_1\}$, while the set of information sets visited by Player 2 in $p_{z_1}^2$ is $\mathcal{J}_{2,z_1} = \{\mathcal{I}_2\}$.

Perfect recall. Formally, an EFG has perfect recall if, for each player $i \in [\![n]\!]$, each information set $\mathcal{I} \in \mathcal{J}_i$, and each pair of histories h, $h' \in \mathcal{I}$, we have $q_h^i = q_{h'}^i$ [3]. For example, the game tree in Figure 1 has perfect recall. Indeed, Player 1 has a single information set $\mathcal{I}_1 = \{h_1\}$, so the condition is trivially satisfied for Player 1, while Player 2 has a single information set $\mathcal{I}_2 = \{h_2, h_3\}$, and at both histories h_2 and h_3, Player 2 has taken no previous actions, so their past action sequences coincide.

In an EFG with perfect recall, for each player $i \in [\![n]\!]$, each history $h \in \mathcal{H}$, and each information set $\mathcal{I} \in \mathcal{J}_{i,h}$, we can *unambiguously* define the *unique* history

$$p_{h,\mathcal{I}}^i \in \mathcal{I} \tag{6}$$

visited by player i in the history sequence p_h^i. Indeed, $p_{h,\mathcal{I}}^i$ is well-defined due to the following proposition.

Proposition 1. *For an EFG with perfect recall, the history sequence p_h^i used by each player i to reach a history h cannot contain two different histories from the same information set; i.e., if $p_{h,k}^i$, $p_{h,k'}^i \in \mathcal{I}$ for some $\mathcal{I} \in \mathcal{J}_i$, then $k = k'$.*

Proof. Suppose that, for some player $i \in [\![n]\!]$, some history $h \in \mathcal{H}$, and some k, $k' \in [\![m_h^i]\!]$, there exists $\mathcal{I} \in \mathcal{J}_i$ such that $p_{h,k}^i$, $p_{h,k'}^i \in \mathcal{I}$ and $k < k'$. Thus, by the definition of perfect recall, we have that the action sequences to reach $p_{h,k}^i$ and $p_{h,k'}^i$ are the same. However, since p_h^i is induced by the *unique* action sequence q_h^i to reach history h, and $k < k'$, it follows that the player's action sequence to reach $p_{h,k}^i$ is a subsequence of the action sequence to reach $p_{h,k'}^i$, which contradicts the definition of perfect recall. □

As an example, consider the game tree in Figure 1 and the sequence of histories $p^2_{z_1} = (h_2)$ for Player 2 to reach the leftmost terminal history z_1. Since $h_2 \in \mathcal{I}_2$, we have $p^2_{z_1,\mathcal{I}_2} = h_2$.

1.3 Counterfactual utilities and counterfactual regret

Before describing the CFR algorithm, we first introduce the notions of counterfactual utilities and counterfactual regret, which are fundamental to the operation of CFR.

Reach probabilities. The reach probability quantifies the likelihood of reaching a specific history when the players follow a given behavioral-strategy profile. Formally, for a player $i \in [\![n]\!]$, a behavioral-strategy profile $b \in \mathcal{B}$, and a history $h \in \mathcal{H}$, we define the reach probability of player i reaching h under b as

$$\pi^b_i(h) = \prod_{k=1}^{m^i_h} b_i\big(q^i_{h,k}\big). \tag{7}$$

Furthermore, we define the reach probability of all players other than player i reaching h under b as

$$\pi^b_{-i}(h) = \prod_{j \neq i} \pi^b_j(h) = \prod_{j \neq i} \prod_{k=1}^{m^j_h} b_j\big(q^j_{h,k}\big). \tag{8}$$

Observe that the probability $\pi^b(h)$ of reaching a history $h \in \mathcal{H}$ under a behavioral-strategy profile $b \in \mathcal{B}$ can be decomposed as the product of the reach probabilities for any player $i \in [\![n]\!]$ and the reach probabilities of all players other than player i reaching h; i.e.,

$$\pi^b(h) = \prod_{j=1}^{n} \prod_{k=1}^{m^j_h} b_j\big(q^j_{h,k}\big) = \pi^b_i(h) \cdot \pi^b_{-i}(h). \tag{9}$$

Then, for each history $h \in \mathcal{H}$ and each history $h' \in \mathcal{H}$ in the path from the root of the game tree to h, i.e., h' is in p^i_h for some player $i \in [\![n]\!]$, we can define the conditional reach probability of reaching h from h' under a behavioral-strategy profile $b \in \mathcal{B}$ as

$$\pi^b(h \,|\, h') = \begin{cases} \frac{\pi^b(h)}{\pi^b(h')}, & \text{if } \pi^b(h') > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

As an example, consider the game tree with three levels of decision nodes shown in Figure 2.

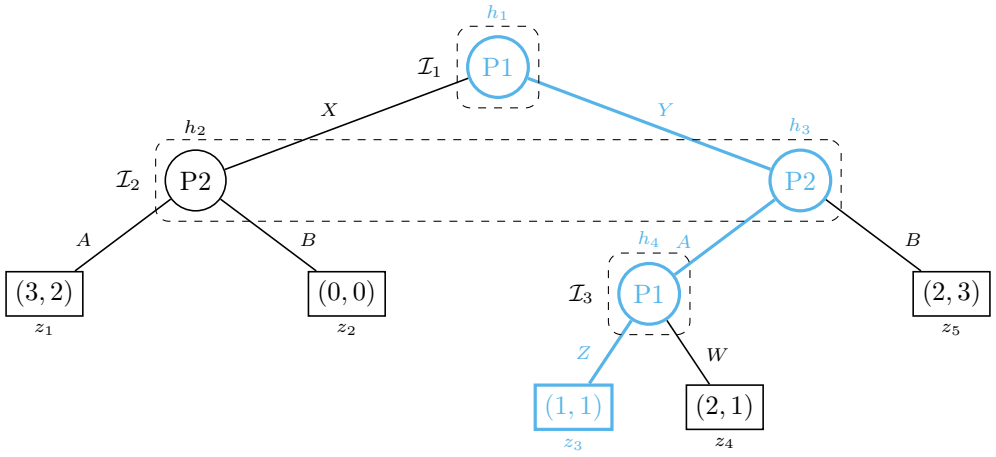

Figure 2: A game tree with three levels of decision nodes. The path from the root history h_1 to the terminal history z_3 is highlighted in sky blue.

Consider the behavioral-strategy profile b where Player 1 chooses action X at information set $\mathcal{I}_1$ with probability 0.6 and action Y with probability 0.4; Player 2 chooses action A at information set $\mathcal{I}_2$ with probability 0.7 and action B with probability 0.3; and, finally, at information set $\mathcal{I}_3$, Player 1 chooses action Z with probability 0.5 and action W with probability 0.5. Then, the reach probability of Player 1 reaching history h_4 under b is $\pi_1^b(h_4) = b_1(Y) = 0.4$, while the reach probability of Player 2 reaching h_4 under b is $\pi_2^b(h_4) = b_2(A) = 0.7$. Thus, the reach probability of both players to reach h_4 under b is $\pi^b(h_4) = \pi_1^b(h_4) \cdot \pi_2^b(h_4) = 0.4 \cdot 0.7 = 0.28$. Similarly, the reach probability of Player 1 reaching terminal history z_3 under b is $\pi_1^b(z_3) = b_1(Y) \cdot b_1(Z) = 0.4 \cdot 0.5 = 0.2$, while the reach probability of Player 2 reaching z_3 under b is $\pi_2^b(z_3) = b_2(A) = 0.7$. Thus, the reach probability of both players to reach z_3 under b is $\pi^b(z_3) = \pi_1^b(z_3) \cdot \pi_2^b(z_3) = 0.2 \cdot 0.7 = 0.14$. Finally, the conditional reach probability of reaching terminal history z_3 from h_4 under b is $\pi^b(z_3 \,|\, h_4) = \frac{\pi^b(z_3)}{\pi^b(h_4)} = \frac{0.14}{0.28} = 0.5$.

Counterfactual utilities. The counterfactual utility of a player $i \in [\![n]\!]$ at an information set $\mathcal{I} \in \mathcal{J}_i$ under a behavioral-strategy profile $b \in \mathcal{B}$ measures the expected payoff for player i under b, conditional on some history in $\mathcal{I}$ being reached, where each terminal history is weighted by the reach probability under b of all other players reaching that history. Formally, for a player $i \in [\![n]\!]$, a behavioral-strategy profile $b \in \mathcal{B}$, and an information set $\mathcal{I} \in \mathcal{J}_i$, we define the counterfactual utility of player i at $\mathcal{I}$ under b as

$$v_i^b(\mathcal{I}) = \sum_{z \in \mathcal{Z} \,:\, \mathcal{I} \in \mathcal{J}_{i,z}} \pi_{-i}^b(p_{z,\mathcal{I}}^i) \cdot \pi^b(z \,|\, p_{z,\mathcal{I}}^i) \cdot u_i(z). \tag{11}$$

Observe that the subset $\{z \in \mathcal{Z} \mid \mathcal{I} \in \mathcal{J}_{i,z}\}$ of terminal histories is exactly the set of terminal histories that can be reached from a history in the information set $\mathcal{I}$. Then, the sum in (11) simply computes the expected payoff for player i by reaching those terminal histories, but weights each terminal history by the likelihood of all other players reaching the history in $\mathcal{I}$ from which that terminal history is reachable. In the sequel, we use counterfactual utilities to evaluate the performance of actions at information sets. This induces a notion of regret we call *counterfactual regret*.

Consider again the game tree in Figure 2 and the behavioral-strategy profile b defined above. Player 1 plays X and Y at $\mathcal{I}_1$ with probabilities 0.6 and 0.4, respectively, and Z and W at $\mathcal{I}_3$ with probabilities 0.5 and 0.5, respectively, while Player 2 plays A and B at $\mathcal{I}_2$ with probabilities 0.7 and 0.3, respectively. For Player 1 at information set $\mathcal{I}_3$, the only terminal histories that pass through $\mathcal{I}_3$ are z_3 and z_4, and in both cases the history visited at $\mathcal{I}_3$ is $p_{z,\mathcal{I}_3}^1 = h_4$. Thus, $\pi_2^b(p_{z,\mathcal{I}_3}^1) = b_2(A) = 0.7$ for $z \in \{z_3, z_4\}$. Moreover, $\pi^b(z_3 \,|\, h_4) = b_1(Z) = 0.5$ and $\pi^b(z_4 \,|\, h_4) = b_1(W) = 0.5$. Finally, the payoffs to Player 1 at these terminals are $u_1(z_3) = 1$ and $u_1(z_4) = 2$. Thus, the counterfactual utility of Player 1 at $\mathcal{I}_3$ under b is

$$v_1^b(\mathcal{I}_3) = \sum_{z \in \{z_3, z_4\}} \pi_2^b(h_4) \cdot \pi^b(z \,|\, h_4) \cdot u_1(z) = 0.7 \cdot (0.5 \cdot 1 + 0.5 \cdot 2) = 1.05. \tag{12}$$

Counterfactual regret. Given a behavioral-strategy profile $b \in \mathcal{B}$ and an action $\alpha \in \mathcal{A}_\mathcal{I}$ at an information set $\mathcal{I} \in \mathcal{J}_i$ for some player $i \in [\![n]\!]$, we define $b^{\mathcal{I} \to \alpha}$ as the behavioral-strategy profile that mirrors b except that player i deterministically chooses α at $\mathcal{I}$; i.e.,

$$b_j^{\mathcal{I} \to \alpha}(\beta) = \begin{cases} 1, & \text{if } j = i \text{ and } \beta = \alpha; \\ 0, & \text{if } j = i \text{ and } \beta \in \mathcal{A}_\mathcal{I} \setminus \{\alpha\}; \\ b_j(\beta), & \text{otherwise.} \end{cases} \tag{13}$$

Then, given a time horizon $T > 0$ and a sequence of behavioral-strategy profiles $(b_1, \ldots, b_T)$ with $b_t \in \mathcal{B}$ for all $t \in [\![T]\!]$, we define the counterfactual regret for player $i \in [\![n]\!]$ from not having chosen action $\alpha \in \mathcal{A}_\mathcal{I}$ at information set $\mathcal{I} \in \mathcal{J}_i$ up to time T as

$$R_i^T(\mathcal{I}, \alpha) = \sum_{t=1}^{T} \big(v_i^{b_i^{\mathcal{I} \to \alpha}}(\mathcal{I}) - v_i^{b_t}(\mathcal{I}) \big). \tag{14}$$

In other words, the counterfactual regret quantifies the local regret at each information set $\mathcal{I}$ that player i experiences from not choosing action α at $\mathcal{I}$, with respect to the counterfactual utility at $\mathcal{I}$.

1.4 The CFR algorithm.

The CFR algorithm iteratively minimizes the counterfactual regret at each information set for all players by updating their behavioral strategies based on the accumulated counterfactual regret up to that iteration via regret matching [4]. In particular, at each iteration $t \in [\![T]\!]$, for each player $i \in [\![n]\!]$ and each information set $\mathcal{I} \in \mathcal{J}_i$, the behavioral strategy $b_i^t \in \mathcal{B}_i$ of player i is updated as

$$b_i^t(\alpha) = \begin{cases} \frac{\left(R_i^{t-1}(\mathcal{I},\alpha)\right)_+}{\sum_{\beta \in \mathcal{A}_\mathcal{I}} \left(R_i^{t-1}(\mathcal{I},\beta)\right)_+}, & \text{if } \sum_{\beta \in \mathcal{A}_\mathcal{I}} \left(R_i^{t-1}(\mathcal{I},\beta)\right)_+ > 0; \\ \frac{1}{|\mathcal{A}_\mathcal{I}|}, & \text{otherwise,} \end{cases} \tag{15}$$

where $(x)_+ = \max\{0, x\}$ for all $x \in \mathbb{R}$. In other words, at each iteration, the probability of choosing each action at each information set is proportional to the positive part of the accumulated counterfactual regret from not having chosen that action up to the previous iteration.

Under standard regularity assumptions, the regret-matching updates guarantee that the cumulative counterfactual regret at each information set grows at most on the order of $\mathcal{O}\left(\sqrt{T}\right)$ [4]; however, this bound concerns local counterfactual regret rather than the standard regret of the game, and the CFR decomposition theorem below provides the connection between the two.

Theorem 2 (CFR decomposition [1])**.** *For an n-player EFG with perfect recall, let $(b_1, \ldots, b_T)$ be a sequence of behavioral-strategy profiles generated by the CFR algorithm in (15) over T iterations. Then, for each player $i \in [\![n]\!]$, the regret R_i^T of player i after T iterations is bounded above by the sum of the positive parts of the counterfactual regrets at each of their information sets; i.e.,*

$$R_i^T \leq \sum_{\mathcal{I} \in \mathcal{J}_i} \max_{\alpha \in \mathcal{A}_\mathcal{I}} \left(R_i^T(\mathcal{I}, \alpha)\right)_+. \tag{16}$$

Thus, in an EFG with perfect recall, CFR treats each information set as a local regret-minimization problem based on counterfactual utilities, and the decomposition theorem above shows that controlling counterfactual regret at every information set is sufficient to control a player's overall regret in the original game.

A List of abbreviations

CFR	**C**ounter**f**actual **R**egret **M**inimization 1, 3, 5, 7
EFG	**E**xtensive **F**orm **G**ame 1, 3, 4, 7
NE	**N**ash **E**quilibrium 1, 3

B Index

References

[1] Martin A. Zinkevich et al. "Regret Minimization in Games with Incomplete Information." In: *21st Annual Conference on Neural Information Processing Systems 2007*. Neural Information Processing Systems (Dec. 3–6, 2007). Advances in Neural Information Processing Systems 20. Neural Information Processing Systems Foundation, 2007, pp. 1729–1736. ISBN: 978-1-6056-0352-0.

[2] John Forbes Nash Jr. "Non-Cooperative Games." In: *Annals of Mathematics* 54.2 (Sept. 1951), pp. 286–295. ISSN: 0003-486X. DOI: 10.2307/1969529.

[3] Harold William Kuhn. "Extensive Games and the Problem of Information." In: *Contributions to the Theory of Games*. Ed. by Harold William Kuhn and Albert William Tucker. Vol. 2. 2 vols. Annals of Mathematics Studies 24. Princeton, New Jersey, United States: Princeton University Press, 1953, pp. 193–216.

[4] Sergiu Hart and Andreu Mas-Colell. "A Simple Adaptive Procedure Leading to Correlated Equilibrium." In: *Econometrica* 68.5 (2000), pp. 1127–1150. ISSN: 1468-0262. DOI: 10.1111/1468-0262.00153.