Faster Game Solving via Asymmetry of Step Sizes

Linjian Meng,[1] Tianpei Yang,[1*] Youzhi Zhang,[2*] Zhenxing Ge,[1] Yang Gao[1]

[1] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[2] Centre for Artificial Intelligence and Robotics, Hong Kong Institute of Science & Innovation, CAS
menglinjian@smail.nju.edu.cn, tianpei.yang@nju.edu.cn, youzhi.zhang@cair-cas.org.hk,
zhenxingge@smail.nju.edu.cn, gaoy@nju.edu.cn

Abstract

Counterfactual Regret Minimization (CFR) algorithms are widely used to compute a Nash equilibrium (NE) in two-player zero-sum imperfect-information extensive-form games (IIGs). Among them, Predictive CFR$^+$ (PCFR$^+$) is particularly powerful, achieving an exceptionally fast empirical convergence rate via the prediction in many games. However, the empirical convergence rate of PCFR$^+$ would significantly degrade if the prediction is inaccurate, leading to unstable performance on certain IIGs. To enhance the robustness of PCFR$^+$, we propose Asymmetric PCFR$^+$ (APCFR$^+$), which employs an adaptive asymmetry of step sizes between the updates of implicit and explicit accumulated counterfactual regrets to mitigate the impact of the prediction inaccuracy on convergence. We present a theoretical analysis demonstrating why APCFR$^+$ can enhance the robustness. To the best of our knowledge, we are the first to propose the asymmetry of step sizes, a simple yet novel technique that effectively improves the robustness of PCFR$^+$. Then, to reduce the difficulty of implementing APCFR$^+$ caused by the adaptive asymmetry, we propose a simplified version of APCFR$^+$ called Simple APCFR$^+$ (SAPCFR$^+$), which uses a fixed asymmetry of step sizes to enable only a single-line modification compared to original PCFR$^+$. Experimental results on five standard IIG benchmarks and two heads-up no-limit Texas Hold'em (HUNL) Subagems show that (i) both APCFR$^+$ and SAPCFR$^+$ outperform PCFR$^+$ in most of the tested games, (ii) SAPCFR$^+$ achieves a comparable empirical convergence rate with APCFR$^+$, and (iii) our approach can be generalized to improve other CFR algorithms, *e.g.*, Discount CFR (DCFR).

Code —
https://github.com/menglinjian/AAAI-2026-APCFRPlus

1 Introduction

Imperfect-information extensive-form games (IIGs) are foundational models to capture interactions among multiple agents in sequential settings with hidden information. IIGs are widely used to simulate real-world scenarios such as medical treatment (Sandholm 2015), security games (Lisý, Davis, and Bowling 2016), cybersecurity (Chen et al. 2017), and recreational games (Brown and Sandholm 2018, 2019b).

*Corresponding Authors.

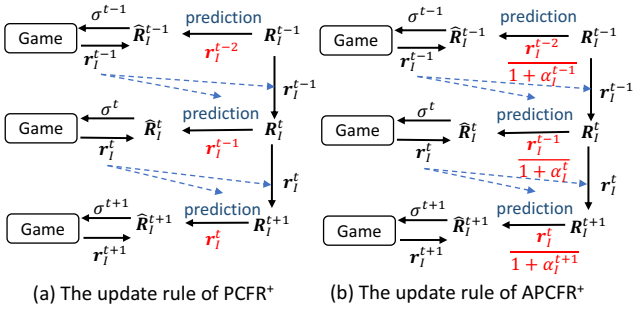

Figure 1: Comparison between PCFR$^+$ and APCFR$^+$, with differences highlighted in red. Note that the notation t in α_I^t denotes iteration t, rather than an exponent.

To address IIGs, a primary goal is to compute a Nash equilibrium (NE), where no player can unilaterally improve its payoff by deviating from the equilibrium.

As with much of the literature on solving IIGs, we focus on computing an NE in two-player zero-sum IIGs. The most widely used method for computing an NE in these IIGs is Counterfactual Regret Minimization (CFR) (Zinkevich et al. 2007; Lanctot et al. 2009; Tammelin 2014; Brown and Sandholm 2019a; Farina, Kroer, and Sandholm 2021, 2019; Liu et al. 2021, 2023; Meng et al. 2023; Farina et al. 2023; Xu et al. 2022, 2024b,a; Zhang, McAleer, and Sandholm 2024), as evidenced by their success in superhuman game AIs (Bowling et al. 2015; Moravčík et al. 2017; Brown and Sandholm 2018, 2019b; Pérolat et al. 2022). The key insight of CFR algorithms is to decompose the total regret over the game into a sum of counterfactual regrets associated within information sets (infosets) and employ a local regret minimizer to minimize counterfactual regrets within each infoset.

Many technologies have been proposed to improve the empirical convergence rate of CFR algorithms. For example, Counterfactual Regret Minimization$^+$ (CFR$^+$) (Tammelin 2014) replaces the local regret minimizer—Regret Matching (RM) (Hart and Mas-Colell 2000; Gordon 2006)—used in vanilla CFR with Regret Matching$^+$ (RM$^+$). CFR$^+$ improves the empirical convergence rate by ensuring that the accumulated counterfactual regrets remain non-negative. Then, Farina, Kroer, and Sandholm (2021) introduce Predictive CFR$^+$ (PCFR$^+$), an improved variant of CFR$^+$.

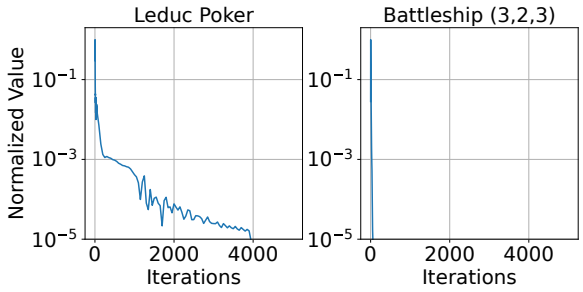

Figure 2: Dynamics of inaccuracy in PCFR$^+$ between predicted and observed instantaneous counterfactual regrets in Leduc Poker and Battleship (3,2,3). This inaccuracy is related to the theoretical convergence rate of PCFR$^+$. The values on the y-axis are normalized to the range [0, 1], which is displayed on a logarithmic scale.

PCFR$^+$ significantly outperforms other CFR algorithms including CFR$^+$ in many IIGs by using the prediction. Specifically, PCFR$^+$ maintains two types of accumulated counterfactual regrets: the implicit and the explicit. As shown in Figure 1, at each iteration t, PCFR$^+$ uses the prediction and the observed instantaneous counterfactual regret $\boldsymbol{r}_I^t$ to derive the new explicit accumulated counterfactual regret $\hat{\boldsymbol{R}}_I^t$ and the new implicit counterfactual regret $\boldsymbol{R}_I^{t+1}$, respectively. If the prediction aligns with the observed instantaneous counterfactual regret $\boldsymbol{r}_I^t$, the theoretical convergence rate of PCFR$^+$ can be improved from $O(1/\sqrt{T})$ of CFR$^+$ to $O(1/T)$ (Farina, Kroer, and Sandholm 2021). However, PCFR$^+$ sets the instantaneous counterfactual regret $\boldsymbol{r}_I^{t-1}$ observed at iteration $t-1$ as the prediction at iteration t. This operation may cause inaccurate prediction on certain IIGs, which harms the empirical convergence rate of PCFR$^+$. As noted by Farina, Kroer, and Sandholm (2021), PCFR$^+$ underperforms other CFR algorithms in Leduc Poker (Game [**O**] in Farina, Kroer, and Sandholm (2021)), yet significantly surpasses them in Battleship (3,2,3) (Game [**R**] in Farina, Kroer, and Sandholm (2021)). This aligns with the results in Figure 2: the gap between predicted and observed instantaneous counterfactual regret decreases slowly in Leduc Poker but diminishes rapidly in Battleship (3,2,3), validating our hypothesis.

To enhance the robustness of PCFR$^+$, we propose a novel variant of PCFR$^+$, termed Asymmetric PCFR$^+$ (APCFR$^+$). Similar to PCFR$^+$, APCFR$^+$ leverages the prediction to improve the convergence rate, but it mitigates the impact of the prediction inaccuracy on convergence. Specifically, as illustrated in Figure 1, APCFR$^+$ utilizes an adaptive asymmetry mechanism for step sizes between implicit and explicit accumulated counterfactual regret updates, which dynamically reduces the step size when updating via the prediction. We prove that when the step size for updating the explicit accumulated counterfactual regret via the prediction at iteration t is set to $1/(1+\alpha^t)$ for APCFR$^+$, where $\alpha^t \geq 0$ is a constant, the effect of the prediction inaccuracy on the convergence rate for APCFR$^+$ is reduced by a factor of $1+\alpha^t$ compared to PCFR$^+$. Therefore, APCFR$^+$ mitigates the impact of the prediction inaccuracy on the convergence rate. Then, through the theoretical analysis of APCFR$^+$, we propose an automatic learning mechanism for α^t, eliminating the need for fine-tuning parameters across different games. To the best of our knowledge, we are the first to propose the asymmetry of step sizes updating implicit and explicit accumulated counterfactual regrets.

To simplify the implementation of APCFR$^+$ caused by the automatic learning approach of α^t, we introduce a simplified version of APCFR$^+$, called Simple APCFR$^+$ (SAPCFR$^+$). Specifically, by analyzing the upper bounds of different terms within the theoretical guarantee of APCFR$^+$ (detailed at the beginning of Section 4.2), SAPCFR$^+$ sets α^t to 2, ensuring SAPCFR$^+$ requires only a single-line modification to the original PCFR$^+$ code.

We conduct extensive experimental evaluations of APCFR$^+$ and SAPCFR$^+$ across five standard IIG benchmarks as well as two heads-up no-limit Texas Hold'em (HUNL) Subgames generated by the top poker agent, Libratus (Brown and Sandholm 2018). The experiments demonstrate that APCFR$^+$ and SAPCFR$^+$ outperforms PCFR$^+$ in nearly all tested games and achieve an empirical convergence rate comparable to that of PCFR$^+$ in the remaining games. Moreover, we observe that SAPCFR$^+$ achieves comparable empirical convergence rate with APCFR$^+$. Finally, we can observe that our approach can be generalized to improve other CFR algorithms, *e.g.*, Discount CFR (DCFR) (Brown and Sandholm 2019a).

2 Related Work

We consider CFR algorithms (Zinkevich et al. 2007; Tammelin 2014; Brown and Sandholm 2019a; Farina, Kroer, and Sandholm 2021, 2019; Liu et al. 2021; Pérolat et al. 2021; Liu et al. 2023; Meng et al. 2023; Farina et al. 2023; Xu et al. 2022, 2024b,a; Zhang, McAleer, and Sandholm 2024), the most widely used method for learning an NE in two-player zero-sum IIGs, as evidenced by their success in superhuman game AIs (Bowling et al. 2015; Moravčík et al. 2017; Brown and Sandholm 2018, 2019b; Pérolat et al. 2022).

The key insight of CFR algorithms is the decomposition of the regret over the game into the sum of counterfactual regrets associated with infosets. The vanilla CFR algorithm, introduced by Zinkevich et al. (2007), employs RM (Hart and Mas-Colell 2000) as the local regret minimizer. To improve the empirical convergence rate of CFR, it is common to design more effective local regret minimizers, as the selection of the local regret minimizers has a significant impact on the overall performance of the CFR algorithm. Examples include RM$^+$ (Bowling et al. 2015), Discounted RM (DRM) (Brown and Sandholm 2019a), and PRM$^+$ (Farina, Kroer, and Sandholm 2021), which correspond to CFR$^+$, DCFR, and PCFR$^+$, respectively. PCFR$^+$ can demonstrate an extremely faster empirical convergence rate than other CFR variants. However, as shown in its original paper, PCFR$^+$ is outperformed by CFR$^+$ and DCFR even on standard IIG benchmarks like Leduc Poker.

To improve the robustness of PCFR$^+$, Farina et al. (2023) propose Stable PCFR$^+$ and Smooth PCFR$^+$. These algorithms improve the robustness by addressing the instability, *i.e.*, rapid strategy fluctuations across iterations, via ensuring

the lower bound of the 1-norm of accumulated counterfactual regrets exceeds a positive constant. However, these algorithms never outperform PCFR$^+$ in terms of the empirical convergence rate even though they achieve a faster theoretical convergence rate than PCFR$^+$, as demonstrated in our experiments. APCFR$^+$ does not focus on addressing the instability, but instead aims to mitigate the impact of the prediction inaccuracy on the convergence to improve the robustness. In our experiments, APCFR$^+$ consistently outperforms Stable PCFR$^+$ and Smooth PCFR$^+$ in all tested games.

3 Preliminaries

Imperfect-information Extensive-form games (IIGs). To model tree-form sequential decision-making problems with hidden information, a common used model is IIG (Osborne et al. 2004). An IIG can be formulated as $G = \{\mathcal{N}, \mathcal{H}, P, A, \mathcal{I}, \{u_i\}\}$. Here, $\mathcal{N} = \{0, 1\}$ is the set of players. "Nature" is also considered a player c (representing chance) and chooses actions with a fixed known probability distribution. $\mathcal{H}$ is the set of all possible histories. For each history $h \in \mathcal{H}$, the function $P(h)$ represents the player acting at history h, and $A(h)$ denotes the actions available at history h. To account for private information, the histories for each player i are partitioned into a collection $\mathcal{I}_i$, referred to as information sets (infosets). For any infoset $I \in \mathcal{I}_i$, histories $h, h' \in I$ are indistinguishable to player i. The notation $\mathcal{I}$ denotes $\mathcal{I} = \{\mathcal{I}_i | i \in \mathcal{N}\}$. Thus, we have $P(I) = P(h)$, $A(I) = A(h), \forall h \in I$. The set of leaf nodes is denoted by $\mathcal{Z}$. For each leaf node z, there is a pair $(u_0(z), u_1(z)) \in [-1,\ 1]$ which denotes the payoffs for the min player (player 0) and the max player (player 1) respectively. In two-player zero-sum IIGs, $u_0(z) = -u_1(z), \forall z \in \mathcal{Z}$.

Behavioral strategy. This strategy σ_i is defined on each infoset. For any infoset $I \in \mathcal{I}_i$, the probability for an action $a \in A(I)$ is denoted by $\sigma_i(I, a)$. We use $\sigma_i(I) = [\sigma_i(I, a)|a \in A(I)] \in \Delta^{|A(I)|}$ to denote the strategy at infoset I, where $\Delta^{|A(I)|}$ is a $(|A(I)| - 1)$-dimension simplex. If every player follows the strategy profile $\sigma = [\sigma_0; \sigma_1]$ and reaches infoset I, the reaching probability is denoted by $\pi^\sigma(I)$. The probability contribution of player i is $\pi_i^\sigma(I)$, while for players other than i, denoted as $-i$, the contribution is $\pi_{-i}^\sigma(I)$. In IIGs, $u_i(\sigma) = u_i(\sigma_i, \sigma_{-i}) = \sum_{z \in \mathcal{Z}} u_i(z)\pi^\sigma(z)$.

Nash equilibrium (NE). NE denotes a rational behavior where no player can benefit by unilaterally deviating from the equilibrium. For any player, her strategy is the best-response to the strategies of others. Formally, for any NE strategy profile σ^* and $i \in \mathcal{N}$, it holds that $u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(\sigma_i, \sigma_{-i}^*)$ for all σ. A widely used metric to measure the distance from the given strategy profile $\boldsymbol{x}$ to NE is the exploitability, which is defined as $\epsilon(\sigma) = \sum_{i \in \mathcal{N}} \max_{\sigma_i'}(u_i(\sigma_i', \sigma_{-i}) - u_i(\sigma_i, \sigma_{-i}))/|\mathcal{N}|$.

Computing an NE via regret minimization algorithms. To compute an NE in IIGs, a common used method is regret minimization algorithms (Rakhlin and Sridharan 2013a,b; Hazan et al. 2016; Joulani, György, and Szepesvári 2017). For any sequence of strategies $\sigma_i^1, \cdots, \sigma_i^T$ of player i, player i's regret is $R_i^T = \max_{\sigma_i} \sum_{t=1}^T u_i(\sigma_i, \sigma_{-i}^t) - \sum_{t=1}^T u_i(\sigma_i^t, \sigma_{-i}^t)$. Regret minimization algorithms are algorithms ensuring R_i^T grows sublinearly. If each player follows a regret minimization algorithm, then their average strategy converges to the set of the NE in two-player zero-sum IIGs. Formally, assume the regret of each player i is R_i^T, then it holds that

$$\epsilon(\bar{\sigma}) = \epsilon(\bar{\sigma}_0, \bar{\sigma}_1) \leq \sum_{i \in \mathcal{N}} R_i^T/(|\mathcal{N}|T),$$

where $\bar{\sigma}_i(I) = \sum_{t=1}^T \pi_i^{\sigma^t}(I)\sigma_i^t(I) / \sum_{t=1}^T \pi_i^{\sigma^t}(I)$.

Counterfactual Regret Minimization (CFR) framework. This framework (Zinkevich et al. 2007; Farina, Kroer, and Sandholm 2019; Liu et al. 2021) is designed to compute an NE of two-player zero-sum IIGs. Instead of directly minimizing the global regret R_i^T, it decomposes the regret to each infoset and independently minimizes the local regret within each infoset. Let σ^t be the strategy profile at iteration t. This framework computes the counterfactual value at infoset I for action a as $v^{\sigma^t}(I, a) = \sum_{h \in I} \sum_{z \in \mathcal{Z}_{ha}} \pi_{-i}^{\sigma^t}(h)\pi^{\sigma^t}(ha, z)u_i(z)$, where $\pi^{\sigma^t}(ha, z)$ denotes the probability from ha to z if all players play according to σ^t and $\mathcal{Z}_{ha}$ is the set of the leaf nodes that are reachable after choosing action a at history h. For any infoset I, the counterfactual regret is $R^T(I) = \max_{a \in A(I)} \sum_{t=1}^T v^{\sigma^t}(I, a) - \sum_{t=1}^T \sum_{a' \in A(I)} \sigma_i^t(I, a')v^{\sigma^t}(I, a')$. The regret over the game $R_i^T = \max_{\sigma_i} \sum_{t=1}^T u_i(\sigma_i, \sigma_{-i}^t) - \sum_{t=1}^T u_i(\sigma_i^t, \sigma_{-i}^t)$ is less than the sum of the counterfactual regrets within infosets: $R_i^T \leq \sum_{I \in \mathcal{I}_i} R^T(I)$. So any regret minimization algorithms can be used as the local regret minimizer to minimize the regret $R^T(I)$ over each infoset to minimize the regret R_i^T.

Predictive Counterfactual Regret Minimization$^+$ (PCFR$^+$). PCFR$^+$ (Farina, Kroer, and Sandholm 2021) is a powerful CFR algorithm, which significantly outperforms other CFR algorithm in many IIGs. PCFR$^+$ employs Predictive RM$^+$ (PRM$^+$) (Farina, Kroer, and Sandholm 2021) as its local regret minimizer, with its key insight is to use the prediction. Specifically, as shown Figure 1, at each iteration t, PCFR$^+$ maintains implicit and explicit accumulated counterfactual regrets: $\boldsymbol{R}_I^t$ and $\hat{\boldsymbol{R}}_I^{t-1}$. Firstly, PCFR$^+$ makes a prediction and uses this prediction to derive new explicit accumulated counterfactual regrets $\hat{\boldsymbol{R}}_I^t$ from $\boldsymbol{R}_I^t$. Then, PCFR$^+$ observes the instantaneous counterfactual regret $\boldsymbol{r}_I^t$ by following the strategy σ^t defined by $\hat{\boldsymbol{R}}_I^t$. Lastly, $\boldsymbol{r}_I^t$ is subsequently used to derive $\boldsymbol{R}_I^{t+1}$ from $\boldsymbol{R}_I^t$. If the prediction aligns with the observed instantaneous counterfactual regret $\boldsymbol{r}_I^t$, Farina, Kroer, and Sandholm (2021) show that the theoretical convergence of PCFR$^+$ can be improved from $O(1/\sqrt{T})$ of CFR$^+$ to $O(1/T)$. As tested in Farina, Kroer, and Sandholm (2021), using the instantaneous counterfactual regret $\boldsymbol{r}_I^{t-1}$ observed at the previous iteration $t-1$ as the prediction is both simple and effective. Therefore, in practice, PCFR$^+$ uses $\boldsymbol{r}_I^{t-1}$ as the prediction at iteration t. Formally, at each iteration t and for each infoset $I \in \mathcal{I}$, PCFR$^+$ updates its strategy according to

$$\hat{\boldsymbol{R}}_I^t = [\boldsymbol{R}_I^t + \boldsymbol{r}_I^{t-1}]^+, \quad \boldsymbol{R}_I^{t+1} = [\boldsymbol{R}_I^t + \boldsymbol{r}_I^t]^+,$$

$$\sigma_i^t(I) = \frac{[\hat{\boldsymbol{R}}_I^t]^+}{\|[\hat{\boldsymbol{R}}_I^t]^+\|_1} = \frac{\hat{\boldsymbol{R}}_I^t}{\|\hat{\boldsymbol{R}}_I^t\|_1},$$

where $i = P(I)$, $\boldsymbol{R}_I^1 = \mathbf{0}$, and the forth equality comes from the fact that $\hat{\boldsymbol{R}}_I^t \geq \mathbf{0}$.

4 Methodology

PCFR$^+$ leverages the prediction to accelerate the empirical convergence rate. However, when the prediction is inaccurate, its empirical convergence rate may decrease significantly, leading to unstable performance on certain IIGs. To enhance the robustness of PCFR$^+$, we propose Asymmetric PCFR$^+$ (APCFR$^+$), which mitigates the impact of the prediction inaccuracy on the convergence rate via the adaptive asymmetry of step sizes. We then provide a theoretical analysis for APCFR$^+$ to demonstrate the reason why it enhances the robustness. To simplify the implementation of APCFR$^+$ due to the adaptive asymmetry, we propose Simple APCFR$^+$ (SAPCFR$^+$), using a constant asymmetry to guarantee that it can be implemented with a single-line modification compared to PCFR$^+$.

4.1 Asymmetric PCFR$^+$ (APCFR$^+$)

To mitigate the impact of the prediction inaccuracy on convergence of PCFR$^+$, APCFR$^+$ adaptively reduces the step size when updating via the prediction, i.e., when updating the explicitly accumulated counterfactual regret. In other words, APCFR$^+$ exploits the adaptive asymmetry of step sizes between the updates of the implicit and explicit ones. Formally, at iteration t and infoset I, the update rule of APCFR$^+$ is

$$\hat{\boldsymbol{R}}_I^t = [\boldsymbol{R}_I^t + \frac{1}{1+\alpha_I^t}\boldsymbol{r}_I^{t-1}]^+, \quad \boldsymbol{R}_I^{t+1} = [\boldsymbol{R}_I^t + \boldsymbol{r}_I^t]^+,$$
$$\sigma_i^t(I) = \frac{[\hat{\boldsymbol{R}}_I^t]^+}{\|[\hat{\boldsymbol{R}}_I^t]^+\|_1} = \frac{\hat{\boldsymbol{R}}_I^t}{\|\hat{\boldsymbol{R}}_I^t\|_1},$$

where $i = P(I)$, $\boldsymbol{R}_I^1 = \mathbf{0}$, and $\boldsymbol{r}_I^0 = \mathbf{0}$. The comparison between the update rules of PCFR$^+$ and APCFR$^+$ has been shown in Figure 1. In the rest of this subsection, we first present the regret upper bound for APCFR$^+$ with respect to any α_I^t, as stated in Theorem 4.1. According to the discussion about Theorem 4.1, we show why APCFR$^+$ can enhance the robustness of PCFR$^+$ by mitigating the impact of the prediction inaccuracy on the convergence rate. Lastly, we discuss how to automatically learn α_I^t from the regret bound shown in Theorem 4.1.

Theorem 4.1. *[Proof is in Appendix A[1]]. Assuming that T iterations of APCFR$^+$ with any $\alpha_I^t \geq 0$ are conducted, the counterfactual regret at any infoset $I \in \mathcal{I}$ is bound by*

$$R^T(I) \leq \sqrt{\sum_{t=1}^{T}\left(\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2\right)}.$$

Why the asymmetry mechanism is effective. To assess the effectiveness of the asymmetry mechanism for step sizes in decreasing the regret upper bound (improving the convergence rate), we show the upper bound of $\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$ is four times than that of $\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$. Firstly, we introduce Lemma 4.2.

[1]https://arxiv.org/abs/2503.12770

Lemma 4.2. *[Adapted from Lemma 11 of Wei et al. (2021)]. Assume that T iterations of APCFR$^+$ with any $\alpha_I^t \geq 0$ are conducted. Then for any infoset $I \in \mathcal{I}$ and $t \geq 1$, we have*

$$\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2 \leq \|\boldsymbol{r}_I^t\|_2^2.$$

Assume that for any infoset $I \in \mathcal{I}$ and $t \geq 1$, $\|\boldsymbol{r}_I^t\|_2^2 \leq E$. Then, from Lemma 4.2, we have

$$\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2 \leq E. \tag{1}$$

Similarly, for $\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$, we have

$$\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2 \leq 4E. \tag{2}$$

In experiments, we also analyze the values of two terms $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$ and $\sum_{t=1}^T \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$, for both PCFR$^+$ and our algorithms (Figures 6 and 7). Among all algorithms, we observe that the value of $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$ is at least three times than that of $\sum_{t=1}^T \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$. This indicates that introducing the term $\sum_{t=1}^T \alpha_I^t \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$ and modifying the term $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$ to $\sum_{t=1}^T \frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t}$, can reduce the regret upper bound. Furthermore, compared to PCFR$^+$, both of these two terms are smaller in our algorithms, further decreasing the regret upper bound. Then, we evaluate the values of $\sum_{t=1}^T (\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2)$, for both PCFR$^+$ and our algorithms (Figures 8 and 9). In all games, the value of $\sum_{t=1}^T (\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2)$ is consistently smaller in our algorithms than in PCFR$^+$. See more details in Appendix D.

An alternative regret upper bound of APCFR$^+$. Notably, Theorem 4.1 does not conflict the upper regret bound of CFR$^+$ (where $\alpha_I^t \to \infty$), as it provides a larger upper regret bound than the original CFR$^+$ upper bound. By altering the proof method, we get $R^T(I) \leq \sqrt{\sum_{t=1}^T \|\boldsymbol{r}_I^t - \frac{1}{1+\alpha_I^t}\boldsymbol{r}_I^{t-1}\|_2^2}$, as shown in Theorem B.1 (detailed in Appendix B). By setting $\alpha_I^t \to \infty$, the original bound of CFR$^+$ can be recovered. Additionally, for PCFR$^+$ (where $\alpha_I^t \to 0$), the bound in Theorem 4.1 is identical to the one in its original version (the result in Theorem 3 of the original PCFR$^+$ version can be easily improved to the bound presented in Theorem 4.1). The reason why we employ Theorem 4.1 in the main text rather than Theorem B.1 is that the regret bound in Theorem B.1 is typically larger than that in Theorem 4.1, as demonstrated in Appendix D (Figures 8, 9, 10, and 11).

Automatic learning approach for α_I^t. To eliminate the fine-tuning of α_I^t, we propose an automatic learning approach for α_I^t. From Theorem 4.1, we have

$$R^T(I) \leq \sqrt{\sum_{t=1}^{T}\left(\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2\right)}$$
$$\leq \sqrt{\sum_{t=1}^{T}\left(\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{\alpha_I^t} + \alpha_I^t\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2\right)}.$$

To minimize the right-hand side of the last inequality, we can set $\alpha_I^t = \sqrt{\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2}}$. However, this is not feasible, as we

need α_I^t to compute $\boldsymbol{r}_I^t$. Therefore, we adopt an alternative approach:

$$\alpha_I^t = \min\left(\sqrt{\frac{\sum_{\tau=1}^{t-1}\|\boldsymbol{r}_I^\tau - \boldsymbol{r}_I^{\tau-1}\|_2^2}{\sum_{\tau=1}^{t-1}\|\boldsymbol{R}_I^{\tau+1} - \boldsymbol{R}_I^\tau\|_2^2}}, \alpha_{max}\right). \quad (3)$$

Note that the parameter in α_{max} in Eq. (3) is included solely to ensure that the bound in Theorem 4.1 remains finite. In this paper, we directly set it as 5 to reduce the cost of hyperparameter tuning. In practice, we rarely observed α_I^t reaching 5 (Figures 4 and 5).

4.2 Simple APCFR$^+$ (SAPCFR$^+$)

To simplify the implementation of APCFR$^+$ caused by the automatic learning approach of α^t, we introduce SAPCFR$^+$, which is implemented with a single-line modification to the PCFR$^+$ code. Specifically, SAPCFR$^+$ sets $\alpha_I^t = 2$. The key insight of setting $\alpha_I^t = 2$ lies in the fact that the upper bound of $\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$ is only a quarter of the upper bound of $\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$, as shown in Eq. (1), and Eq. (2).

Specifically, combining Theorem 4.1, Eq. (1), and Eq. (2), in the worst case, we obtain

$$\begin{aligned} R^T(I) &\le \sqrt{\sum_{t=1}^{T}\left(\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2\right)} \\ &\le \sqrt{\sum_{t=1}^{T}\left(\frac{4E}{1+\alpha_I^t} + \alpha_I^t E\right)}. \end{aligned}$$

It is evident that when $\alpha_I^t = 0$, i.e., for PCFR$^+$, the worst-case counterfactual regret upper bound is

$$R^T(I) \le \sqrt{\sum_{t=1}^{T} 4E}.$$

From the facts that (i) 2 minimizes $\sum_{t=1}^{T}(4E/\alpha_I^t + \alpha_I^t E)$ for any positive E and (ii) $\sum_{t=1}^{T}(4E/(1+\alpha_I^t) + \alpha_I^t E) \le \sum_{t=1}^{T}(4E/\alpha_I^t + \alpha_I^t E)$, we can set $\alpha_I^t = 2$, which implies the counterfactual regret is bound by

$$R^T(I) \le \sqrt{\sum_{t=1}^{T}\left(\frac{4E}{1+2} + 2E\right)} = \sqrt{\sum_{t=1}^{T}\frac{10E}{3}}.$$

Clearly, setting $\alpha_I^t = 2$ results in a lower regret upper bound than PCFR$^+$. Therefore, for SAPCFR$^+$, we set $\alpha_I^t = 2$ for all $t \ge 1$. Formally, at each iteration t, SAPCFR$^+$ updates its strategy at each infoset $I \in \mathcal{I}$ according to the following update rule:

$$\hat{\boldsymbol{R}}_I^t = [\boldsymbol{R}_I^t + \frac{1}{3}\boldsymbol{r}_I^{t-1}]^+, \quad \boldsymbol{R}_I^{t+1} = [\boldsymbol{R}_I^t + \boldsymbol{r}_I^t]^+,$$

$$\sigma_i^t(I) = \frac{[\hat{\boldsymbol{R}}_I^t]^+}{\|[\hat{\boldsymbol{R}}_I^t]^+\|_1} = \frac{\hat{\boldsymbol{R}}_I^t}{\|\hat{\boldsymbol{R}}_I^t\|_1},$$

where $i = P(I)$, $\boldsymbol{R}_I^1 = \boldsymbol{0}$, and $\boldsymbol{r}_I^0 = \boldsymbol{0}$.

5 Experiments

Configurations. We now evaluate the empirical convergence rates of APCFR$^+$ and SAPCFR$^+$ by comparing them to PCFR$^+$, Stable PCFR$^+$, Smooth PCFR$^+$, Reg-CFR (Liu et al. 2023), and Clairvoyant CFR (Farina et al. 2023). Stable PCFR$^+$ and Smooth PCFR$^+$ are advanced PCFR$^+$ variants. Reg-CFR and Clairvoyant CFR achieve theoretical convergence rates of $O(1/T^{\frac{3}{4}})$ and $O(1/T)$, respectively, while that of other algorithms is $O(1/\sqrt{T})$. Following the settings in PCFR$^+$, we employ alternating updates for both APCFR$^+$ and SAPCFR$^+$. For Stable PCFR$^+$, Smooth PCFR$^+$, and Reg-CFR, we apply alternating updates, as described in their original paper or open-source code. Clairvoyant CFR does not utilize alternating updates, in accordance with its original design. For all algorithms, we utilize quadratic averaging. For all compared algorithms, we adopt the hyperparameters as suggested in their respective original versions. Details on the size of the tested games are in Appendix D (Table 3). The experiments are conducted on a machine equipped with a Xeon(R) Gold 6444Y CPU and 256 GB of memory.

Empirical convergence rates in standard IIG benchmarks. We now present the empirical convergence rates across five standard IIG benchmarks, *e.g.*, Kuhn Poker, Leduc Poker, Goofspiel Poker, Liar's Dice, and Battleship. These games are implemented using OpenSpiel (Lanctot et al. 2019). The algorithm implementations are based on LiteEFG (Liu, Farina, and Ozdaglar 2024), as LiteEFG provides approximately 100 times speedup compared to the default implementation in OpenSpiel. The results are in Figure 3. For most of the tested games, except for Battleship (3,2,3) and Goofspiel (4), APCFR$^+$ and SAPCFR$^+$, significantly outperform all baselines. Even in Battleship (3,2,3) and Goofspiel (4), APCFR$^+$ and SAPCFR$^+$ outperform all algorithms except PCFR$^+$. Remarkably, they exhibit performance comparable to PCFR$^+$, reaching similar levels of exploitability after 5000 iterations. Based on the experimental results in Appendix D (Figures 6 and 7), we observe that in the games where our algorithms perform similar to PCFR$^+$, such as Battleship (3,2,3) and Goofspiel (4), PCFR$^+$ also exhibits a rapid decrease in the inaccuracy between the predicted and observed instantaneous counterfactual regrets (detailed discussions are in Appendix D). Furthermore, the performance gap between APCFR$^+$ and SAPCFR$^+$ is relatively small. Specifically, APCFR$^+$ only outperforms SAPCFR$^+$ in Leduc Poker, Battleship (3,2,3), and Liar's Dice (5). This small performance gap means that in practical applications, SAPCFR$^+$ can be directly used due to its ease of implementation and a faster empirical convergence rate compared to PCFR$^+$. Regarding Stable PCFR$^+$ and Smooth PCFR$^+$, we find that they significantly underperform PCFR$^+$. Reg-CFR and Clairvoyant CFR significantly underperform relative to other algorithms.

Empirical convergence rates in HUNL Subgames. To assess the performance of APCFR$^+$ and SAPCFR$^+$ in addressing real-world games, we also conduct evaluations in HUNL Subgames, which are considerably larger than standard IIG benchmarks. Despite the presence of code related to HUNL Subgames in Openspiel, we have not successfully executed it. Therefore, we utilize HUNL Subgames imple-

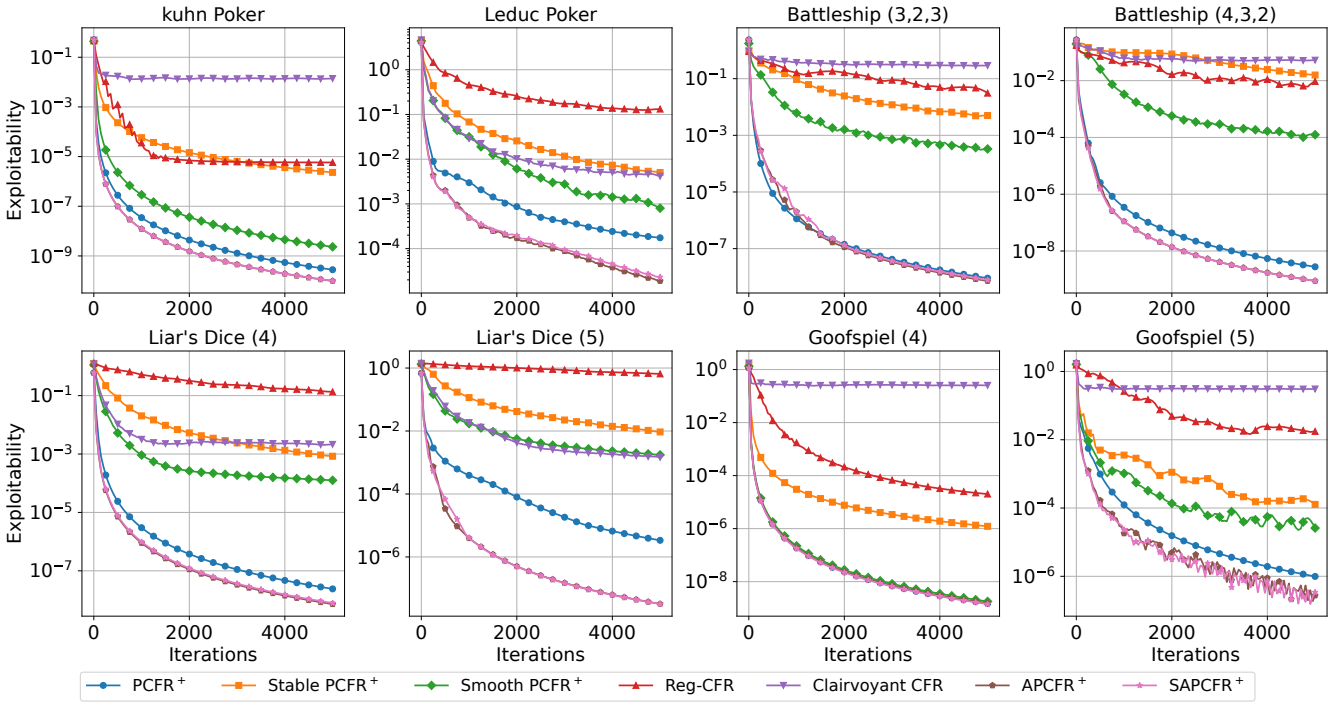

Figure 3: Empirical convergence rates of the tested algorithms in standard commonly used IIG benchmarks. In all plots, the x-axis is the number of iterations, and the y-axis is exploitability, displayed on a logarithmic scale. Liar's Dice (x) represents that every player is given a die with x sides. Goofspiel (x) denotes that each player is dealt x cards. Battleship (x, y, z) implies the size of the grid is $x \times y$, and the number of shots is z.

mented by Poker RL (Steinberger 2019). More precisely, our code is based on the code from Xu et al. (2024b). The code in Xu et al. (2024b) supports only Subgame 3 and Subgame 4, so we conduct experiments solely on these two HUNL Subgames. We do not compare Reg-CFR and Clairvoyant CFR in HUNL Subgames, as they perform significantly worse than other CFR algorithms, even in standard IIG benchmarks. The results are shown in Table 1: APCFR+ and SAPCFR+ consistently outperform all baselines in both subgames.

Running times. To validate the efficiency of APCFR$^+$ and SAPCFR$^+$, we compare their running time with that of PCFR$^+$ under the same number of iterations (*i.e.*, 5000). The experimental results are in Appendix D (Table 4). The running time of APCFR$^+$ is slightly higher compared to PCFR$^+$, primarily due to the additional α_I^t learning process in APCFR$^+$. However, the running time of SAPCFR$^+$ is nearly identical to that of PCFR$^+$, as the only difference between their implementations is a single line of code, which does not alter the computational complexity. Notably, the computational complexity remains exactly the same, even with no change in the constant factors.

Dynamics of α_I^t in APCFR$^+$. To study the behavior of α_I^t, we analyze its dynamics, as shown in Appendix D (Figures 4 and 5). We observe that α_I^t experiences a rapid increase during the initial phase but ceases to grow after approximately 100 iterations. This might be due to that the values of $\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$ and $\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$ are significantly larger in the initial phase than at later stages (as also observed in Figures 6 and 7). More details are in Appendix D.

Dynamics of $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$, $\sum_{t=1}^T \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$, $\sum_{t=1}^T (\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2)$, and $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \frac{\boldsymbol{r}_I^{t-1}}{1+\alpha_I^t}\|_2^2$. To evaluate the regret bound presented in our theoretical analysis, we examine the dynamics of these terms, as demonstrated in Appendix D. Specifically, the dynamics of $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$ and $\sum_{t=1}^T \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$ are presented in Figures 6 and 7. Similarly, Figures 8 and 9 present the dynamics of $\sum_{t=1}^T \left(\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2\right)$. Additionally, Figures 10 and 11 depict the dynamics of $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \frac{\boldsymbol{r}_I^{t-1}}{1+\alpha_I^t}\|_2^2$. This experimental results show that $\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2$ are larger than $\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$, which confirms that APCFR$^+$ effectively reduces the impact of $\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$ by increasing the weights on $\|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2$. For the first three terms, their values are smaller in APCFR$^+$ and SAPCFR$^+$ compared to PCFR$^+$, implying a lower regret bound in Theorem 4.1. However, the value of $\sum_{t=1}^T \|\boldsymbol{r}_I^t - \frac{\boldsymbol{r}_I^{t-1}}{1+\alpha_I^t}\|_2^2$ significantly exceeds that of $\sum_{t=1}^T (\frac{\|\boldsymbol{r}_I^t - \boldsymbol{r}_I^{t-1}\|_2^2}{1+\alpha_I^t} + \alpha_I^t \|\boldsymbol{R}_I^{t+1} - \boldsymbol{R}_I^t\|_2^2)$, indicating that the regret bound in Theo-

	PCFR$^+$	Stable PCFR$^+$	Smooth PCFR$^+$	APCFR$^+$	SAPCFR$^+$
Subgame 3	1.44e-3	1.41e-3 (-2.1%)	1.42e-3 (-1.4%)	1.02e-3 (-29.2%)	9.44e-4 (-34.4%)
Subgame 4	1.04e-3	9.77e-4 (-5.3%)	1.02e-3 (-1.9%)	7.53e-4 (-27.6%)	7.83e-4 (-24.7%)

Table 1: Final exploitability for the tested algorithms in HUNL Subgames. Values in red indicate percentages relative to PCFR$^+$.

	Leduc Poker (5)	Leduc Poker (9)	Leduc Poker (13)
DCFR	2.79e-5	1.27e-5	1.09e-5
PCFR$^+$	2.69e-5	5.21e-5	3.15e-5
APCFR$^+$	4.80e-6 (-82.1%)	4.03e-5 (-22.6%)	1.45e-5 (-54.0%)
SAPCFR$^+$	3.49e-6 (-87.0%)	4.07e-5 (-21.9%)	1.42e-5 (-55.0%)
DCFR$^+$	1.15e-5 (-58.8%)	6.41e-6 (-49.5%)	8.56e-6 (-21.6%)
APDCFR$^+$	3.69e-6 (-86.3%, -86.7%)	3.42e-6 (-93.4%, -73.1%)	3.02e-6 (-90.4%, -72.3%)

Table 2: The final exploitability for DCFR, PCFR$^+$, APCFR$^+$, SAPCFR$^+$, DCFR$^+$, and APDCFR$^+$ in Leduc Poker variants. Values in red indicate percentages of PCFR$^+$ variants relative to PCFR$^+$, and values in blue indicate percentages of DCFR variants relative to DCFR. Notably, APDCFR$^+$ can serve as a variant of both PCFR$^+$ and DCFR.

rem B.1 is extremely higher than that in Theorem 4.1. Thus, we use Theorem 4.1 in the main text instead of Theorem B.1.

Empirical convergence rates of APCFR$^+$ with an alternative learning approach for α^t. We also experiment with a different learning approach for α^t, other than the one in Eq. (3), *e.g.*, $\alpha_I^t = \min\left(\sqrt{\frac{\max_{\tau\in[t-1]}\|\boldsymbol{r}_I^\tau - \boldsymbol{r}_I^{\tau-1}\|_2^2}{\max_{\tau\in[t-1]}\|\boldsymbol{R}_I^{\tau+1} - \boldsymbol{R}_I^{\tau}\|_2^2}}, \alpha_{max}\right)$, where we also set $\alpha_{max} = 5$ as did in Eq. (3) to reduce the cost of hyperparameter tuning. The results in Appendix D (Figure 12 and Table 5) indicate that this approach performs similarly to the one in Eq. (3).

Comparison with other classical CFR algorithms and the generalization of our approach. In addition to the CFR algorithms that have already been compared, we also compare APCFR$^+$ and SAPCFR$^+$ with the classic CFR algorithms: CFR, CFR$^+$, and DCFR. Initially, we conducted experiments using standard IIG benchmarks and HUNL Subgames (Figure 13 and Table 6), where APCFR$^+$ and SAPCFR$^+$ consistently outperformed CFR and CFR$^+$ across all games. However, in poker games like Leduc Poker and HUNL Subgames, APCFR$^+$ and SAPCFR$^+$ did not surpass DCFR. Notably, our algorithms and DCFR are not mutually exclusive and can be combined effectively. The core innovation of our algorithms—the asymmetry of step sizes—can be integrated with DCFR, which involves discounting prior iterations when calculating accumulated regrets. Therefore, we propose APDCFR$^+$ by combining APCFR$^+$ with DCFR (details of APDCFR$^+$ are in Appendix C). In addition to CFR, CFR$^+$, DCFR, APCFR$^+$, and SAPCFR$^+$, we also compare APDCFR$^+$ with DCFR$^+$ (Xu et al. 2024b), which is an advanced variant of DCFR. Experimental results, detailed in Appendix D (Table 6), demonstrate that APDCFR$^+$ achieves a substantially faster empirical convergence rate compared to the other evaluated algorithms.

Additionally, to further evaluate the performance of DCFR, PCFR$^+$, APCFR$^+$, SAPCFR$^+$, DCFR$^+$, and APDCFR$^+$ in poker games, we conduct tests on various Leduc Poker variants that are used in the original PCFR$^+$ paper (Farina, Kroer, and Sandholm 2021). Specifically, we test on Leduc Poker with ranks of 5, 9, or 13. We denote these Leduc Poker variants as Leduc Poker (x), where x represents the number of ranks, noting that the original Leduc Poker has 3 ranks. The results, in Table 2, demonstrate that APCFR$^+$ and SAPCFR$^+$ consistently outperform PCFR$^+$ across all Leduc Poker variants. Notably, the degree to which APCFR$^+$ and SAPCFR$^+$ surpass PCFR$^+$ does not depend on the size of the game. Specifically, the smallest improvement of APCFR$^+$ and SAPCFR$^+$ over PCFR$^+$ occurs in Leduc Poker (9), where the reduction in exploitability is less than half of the reduction observed in Leduc Poker (13). Moreover, the results indicate that DCFR does not consistently outperform PCFR$^+$. For instance, in Leduc Poker (5), the performance of DCFR is inferior to that of PCFR$^+$. More importantly, APDCFR$^+$ consistently outperforms all other algorithms across each Leduc Poker variant tested, except in Leduc Poker (5), where it slightly underperforms compared to SAPCFR$^+$.

6 Conclusions

We propose a novel variant of PCFR$^+$ called APCFR$^+$, which employs the adaptive asymmetry of step sizes in the updates of implicit and explicit accumulated counterfactual regrets to improve the robustness of PCFR$^+$. We also introduce SAPCFR$^+$, requiring only a single line modification to PCFR$^+$. Experimental results validate that APCFR$^+$ and SAPCFR$^+$ exhibit a faster empirical convergence rate than PCFR$^+$. To our knowledge, we are the first to propose the asymmetry of step sizes in the updates of implicit and explicit accumulated counterfactual regrets, a simple yet novel technique that effectively improves the robustness of PCFR$^+$. Moreover, the techniques used in other CFR$^+$ algorithms are compatible with our algorithm, which shows the generalization of our approach. For example, for DCFR, by using our approach, we propose APDCFR$^+$, which significantly outperforms DCFR in poker games. Future work involves designing more effective α^t learning approaches to further enhance the empirical convergence rate.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under Grants 62192783 and 62506157, the Jiangsu Science and Technology Major Project BG2024031, the Fundamental Research Funds for the Central Universities (14380128), the Collaborative Innovation Center of Novel Software Technology and Industrialization, and the InnoHK funding.

References

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. 2018. Superhuman AI for Heads-Up No-Limit Poker: Libratus Beats Top Professionals. *Science*, 359(6374): 418–424.

Brown, N.; and Sandholm, T. 2019a. Solving Imperfect-Information Games via Discounted Regret Minimization. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, 1829–1836.

Brown, N.; and Sandholm, T. 2019b. Superhuman AI for Multiplayer Poker. *Science*, 365(6456): 885–890.

Chen, X.; Han, Z.; Zhang, H.; Xue, G.; Xiao, Y.; and Bennis, M. 2017. Wireless Resource Scheduling in Virtualized Radio Access Networks Using Stochastic Learning. *IEEE Transactions on Mobile Computing*, 17(4): 961–974.

Farina, G.; Grand-Clément, J.; Kroer, C.; Lee, C.-W.; and Luo, H. 2023. Regret Matching+: (In)Stability and Fast Convergence in Games. In *Proceedings of the 37th Conference on Neural Information Processing Systems*.

Farina, G.; Kroer, C.; and Sandholm, T. 2019. Online Convex Optimization for Sequential Decision Processes and Extensive-Form Games. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, 1917–1925.

Farina, G.; Kroer, C.; and Sandholm, T. 2021. Faster Game Solving via Predictive Blackwell Approachability: Connecting Regret Matching and Mirror Descent. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, 5363–5371.

Gordon, G. J. 2006. No-regret Algorithms for Online Convex Programs. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, 489–496. MIT Press.

Hart, S.; and Mas-Colell, A. 2000. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5): 1127–1150.

Hazan, E.; et al. 2016. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4): 157–325.

Joulani, P.; György, A.; and Szepesvári, C. 2017. A Modular Analysis of Adaptive (Non-)Convex Optimization: Optimism, Composite Objectives, Variance Reduction, and Variational Bounds. In *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, 681–720.

Lanctot, M.; Lockhart, E.; Lespiau, J.-B.; Zambaldi, V.; Upadhyay, S.; Pérolat, J.; Srinivasan, S.; Timbers, F.; Tuyls, K.; Omidshafiei, S.; et al. 2019. OpenSpiel: A Framework for Reinforcement Learning in Games.

Lanctot, M.; Waugh, K.; Zinkevich, M.; and Bowling, M. 2009. Monte Carlo Sampling for Regret Minimization in Extensive Games. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, 1078–1086.

Lisý, V.; Davis, T.; and Bowling, M. 2016. Counterfactual Regret Minimization in Sequential Security Games. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 544–550.

Liu, M.; Farina, G.; and Ozdaglar, A. 2024. LiteEFG: An Efficient Python Library for Solving Extensive-form Games. *arXiv preprint arXiv:2407.20351*.

Liu, M.; Ozdaglar, A. E.; Yu, T.; and Zhang, K. 2023. The Power of Regularization in Solving Extensive-Form Games. In *Proceedings of the 12th International Conference on Learning Representations*.

Liu, W.; Jiang, H.; Li, B.; and Li, H. 2021. Equivalence Analysis between Counterfactual Regret Minimization and Online Mirror Descent. *arXiv preprint arXiv:2110.04961*.

Meng, L.; Zhang, Y.; Ge, Z.; Yang, S.; Ding, T.; Li, W.; Yang, T.; An, B.; and Gao, Y. 2023. Efficient Last-iterate Convergence Algorithms in Solving Games. arXiv:2308.11256.

Moravčík, M.; Schmid, M.; Burch, N.; Lisý, V.; Morrill, D.; Bard, N.; Davis, T.; Waugh, K.; Johanson, M.; and Bowling, M. 2017. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337): 508–513.

Nemirovskij, A. S.; and Yudin, D. B. 1983. Problem complexity and method efficiency in optimization.

Osborne, M. J.; et al. 2004. *An introduction to game theory*, volume 3. Oxford university press New York.

Pérolat, J.; De Vylder, B.; Hennes, D.; Tarassov, E.; Strub, F.; de Boer, V.; Muller, P.; Connor, J. T.; Burch, N.; Anthony, T.; et al. 2022. Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science*, 378(6623): 990–996.

Pérolat, J.; Munos, R.; Lespiau, J.; Omidshafiei, S.; Rowland, M.; Ortega, P. A.; Burch, N.; Anthony, T. W.; Balduzzi, D.; Vylder, B. D.; Piliouras, G.; Lanctot, M.; and Tuyls, K. 2021. From Poincaré Recurrence to Convergence in Imperfect Information Games: Finding Equilibrium via Regularization. In *Proceedings of the 38th International Conference on Machine Learning*, 8525–8535.

Rakhlin, A.; and Sridharan, K. 2013a. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference on Learning Theory*, 993–1019.

Rakhlin, A.; and Sridharan, K. 2013b. Optimization, learning, and games with predictable sequences. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, 3066–3074.

Sandholm, T. 2015. Steering Evolution Strategically: Computational Game Theory and Opponent Exploitation for Treatment Planning, Drug Design, and Synthetic Biology. In

Proceedings of the 29th AAAI Conference on Artificial Intelligence, 4057–4061.

Steinberger, E. 2019. PokerRL. https://github.com/TinkeringCode/PokerRL.

Tammelin, O. 2014. Solving large imperfect information games using CFR+. *arXiv preprint arXiv:1407.5042*.

Tammelin, O.; Burch, N.; Johanson, M.; and Bowling, M. 2015. Solving heads-up limit Texas Hold'em. In *Proceedings of the 24th International Conference on Artificial Intelligence*, 645–652.

Wei, C.; Lee, C.; Zhang, M.; and Luo, H. 2021. Linear Last-iterate Convergence in Constrained Saddle-point Optimization. In *Proceedings of the 9th International Conference on Learning Representations*.

Xu, H.; Li, K.; Fu, H.; Fu, Q.; and Xing, J. 2022. AutoCFR: learning to design counterfactual regret minimization algorithms. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, volume 36, 5244–5251.

Xu, H.; Li, K.; Fu, H.; Fu, Q.; Xing, J.; and Cheng, J. 2024a. Dynamic discounted counterfactual regret minimization. In *Proceedings of the 12th International Conference on Learning Representations*.

Xu, H.; Li, K.; Liu, B.; Fu, H.; Fu, Q.; Xing, J.; and Cheng, J. 2024b. Minimizing weighted counterfactual regret with optimistic online mirror descent. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 5272–5280.

Zhang, N.; McAleer, S.; and Sandholm, T. 2024. Faster Game Solving via Hyperparameter Schedules. *arXiv preprint arXiv:2404.09097*.

Zinkevich, M.; Johanson, M.; Bowling, M.; and Piccione, C. 2007. Regret Minimization in Games with Incomplete Information. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, 1729–1736.