INVARIANT LAST-ITERATE CONVERGENCE FOR OPTIMISTIC COUNTERFACTUAL REGRET MINIMIZATION IN SOLVING PERTURBED EXTENSIVE FORM GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Counterfactual regret minimization (CFR) represents a family of online algorithms for efficiently learning the Nash equilibrium (NE) in imperfect-information extensive-form games (EFGs). Among them, optimistic/predictive CFR methods incorporate optimism from online optimization to accelerate convergence toward NEs, and often exhibit superior theoretical and empirical performance. However, the last-iterate convergence properties of optimistic CFR variants remain relatively underexplored. In this work, we take a step toward closing this gap by establishing asymptotic last-iterate convergence guarantees for two representative optimistic CFR algorithms, PCFR$^+$ and PDCFR$^+$, in perturbed EFGs. Specifically, we first establish last-iterate convergence of PCFR$^+$ in perturbed regularized EFGs. This result holds for arbitrary step sizes and arbitrary strategy initializations at each decision node, and we therefore refer to it as *invariant* last-iterate convergence. Building on this invariance property, we derive a tighter average-regret bound for PCFR$^+$ in EFGs than the existing bound. We then extend our analysis to PDCFR$^+$, which discounts regrets accumulated in earlier iterations. The invariant last-iterate convergence guarantee continues to hold for any non-decreasing and bounded sequence of discounting factors and arbitrary strategy initializations. Furthermore, we characterize how the last-iterate strategies generated by these two optimistic CFR algorithms approximate NEs of the original EFG. In particular, we show that the approximation error vanishes as the perturbation and regularization coefficients become sufficiently small. Experiments demonstrate that the two algorithms outperform the most recent baseline on most medium-scale EFGs while achieving performance comparable to discounted CFR on large-scale poker games.

1 INTRODUCTION

Extensive-form games (EFGs) are a foundational framework in game theory used to model sequential decision making of multiple agents with imperfect information, and have been widely applied to solve real-world multi-agent learning problems, such as auctions Shubik (1971), cyber security Kakkad et al. (2019), medical treatment Sandholm (2015) and Poker games Bowling et al. (2015). The goal in solving EFGs is to find a Nash Equilibrium (NE), where no agent can increase her utility by unilaterally changing her strategy.

Efficient approaches for solving EFGs include first-order methods (FOM) Hoda et al. (2010); Kroer et al. (2020), and counterfactual regret minimization (CFR) Zinkevich et al. (2007); Tammelin et al. (2015). The latter is often preferred due to its flexibility in combination with acceleration techniques (like sampling Lanctot et al. (2009) and function approximation Brown et al. (2019)) and its superhuman practical performance Moravčík et al. (2017); Brown & Sandholm (2018). Among all CFR variants, the optimistic/predictive CFR algorithms (e.g. PCFR$^+$ in Farina et al. (2021)) are considered representative methods, as they leverage the optimism technique from online optimization Chiang et al. (2012); Rakhlin & Sridharan (2013a;b); Syrgkanis et al. (2015) to accelerate convergence to the NEs of EFG and attain exceptional empirical performance. Conventionally, all CFR-type algorithms work by minimizing the individual regret of each player, where the average strategy over all iterations forms an approximate NE, resulting in the so-called *average-iterate* convergence in solving EFGs Farina et al. (2019a); Brown & Sandholm (2019).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

However, relying on strategies averaged over all iterations can introduce several complications. Maintaining an average strategy incurs additional computational and storage overhead and, when combined with function approximation, may also introduce extra approximation error. For example, Deep CFR Brown et al. (2019) requires a separate neural network to approximate the average strategy. These limitations motivate the study of whether the strategy produced at the current iteration itself converges to the Nash equilibrium (NE) set of an EFG, a property commonly referred to as *last-iterate convergence* Wei et al. (2021). Nevertheless, (Lee et al., 2021; Cai et al., 2025) show, empirically or theoretically, that vanilla CFR Zinkevich et al. (2007) and its optimistic variant PCFR$^+$ may fail to exhibit last-iterate convergence even in normal-form games (NFGs), which form a strict subclass of EFGs. Therefore, establishing last-iterate convergence guarantees for commonly used CFR algorithms appears challenging without additional structural conditions or algorithmic techniques.

The reward transformation (RT) framework provides one such approach for establishing last-iterate convergence of CFR-type algorithms in EFGs Pérolat et al. (2021; 2022); Liu et al. (2023); Bernasconi et al. (2024). The framework modifies the utility functions of the original game by introducing *regularization* and *perturbation*, thereby relating equilibrium computation in the original EFG to learning equilibria in perturbed regularized EFGs. Within this framework, Liu et al. (2023) show that regularized CFR, equipped with an adaptive online mirror descent (OMD) algorithm Nemirovskij & Yudin (1983) as the local regret minimizer, achieves asymptotic last-iterate convergence to the unique NE of the perturbed regularized EFG. More recently, building on an OMD-based formulation of CFR$^+$ Tammelin (2014), a non-optimistic CFR variant, Meng et al. (2025a) establish asymptotic last-iterate convergence for reward-transformation-based CFR$^+$ in perturbed regularized EFGs. Despite this progress, the last-iterate convergence of optimistic CFR algorithms, such as PCFR$^+$, under the RT framework remains poorly understood. Moreover, it remains unclear how the perturbation and regularization coefficients affect the approximation of NEs of the original EFG.

To address the above issues, this work establishes last-iterate convergence for two representative optimistic CFR algorithms, PCFR$^+$ and its discounted variant PDCFR$^+$, under the RT framework, and characterizes how the perturbation and regularization coefficients affect the approximation to the NEs of original EFG by the last-iterate strategy. The main contributions are summarized as follows:

- We establish the asymptotic last-iterate convergence of PCFR$^+$ in perturbed regularized EFGs. Our analysis subsumes the last-iterate convergence of CFR$^+$ as a special case and holds for arbitrary step sizes and arbitrary strategy initializations at every decision node. We therefore refer to this property as *invariant* last-iterate convergence. Building on this invariance property, we derive a novel average-regret bound for PCFR$^+$ in perturbed regularized EFGs. The same bound continues to hold in the original unperturbed and unregularized EFG, yielding a sharper average-iterate convergence guarantee to the NE set of the original EFG than that established in Farina et al. (2021).
- We then incorporate the discounting technique of Brown & Sandholm (2019) into the optimistic CFR framework to downweight regrets accumulated in earlier iterations, and establish both last-iterate and average-iterate convergence for the resulting PDCFR$^+$ algorithm in perturbed regularized EFGs. Its invariant last-iterate convergence continues to hold for any non-decreasing and upper-bounded sequence of discounting factors, as well as for arbitrary strategy initializations.
- We next show that the pointwise convergence, in terms of the Euclidean distance, of the unique NE in perturbed regularized EFG to the NE set of the original EFG generally requires strong assumptions (e.g. unique NE of original EFG). Rather than relying on such assumptions, we quantify how well the last-iterate strategies generated by the two optimistic CFR algorithms approximate an NE of the original EFG in terms of the duality gap. Concretely, we demonstrate that this approximation error vanishes as the perturbation and regularization coefficients become sufficiently small, as well as the last-iterate strategy being close to the unique NE of perturbed regularized EFG.
- Experiments on four EFGs validate the invariant last-iterate convergence of PCFR$^+$ and PDCFR$^+$. Furthermore, experiments on eight small- to medium-scale EFGs and two large-scale EFGs demonstrate that these two reward-transformation-based optimistic CFR variants achieve competitive or superior last-iterate convergence performance compared with existing baselines.

2 PRELIMINARIES

For any integer d, denote the set $[d] = \{0, 1, 2, \cdots, d\}$ for conciseness, Δ^d as the $(d-1)$-dimension simplex, $\|\cdot\|$ as a certain norm in Euclidean space $\mathbb{R}^d$. We use upper boldface letter $\boldsymbol{\mathcal{X}}$, lower boldface letter $\boldsymbol{x}$ and non-bold letter x to denote certain space, vector and scalar respectively.

2.1 IMPERFECT-INFORMATION EXTENSIVE-FORM GAMES

Extensive-form games (EFGs). An imperfect-information EFG can be characterized by a tree-based formalism $G = \{\mathcal{N}, \mathcal{H}, P, A, \mathcal{I}, \{u_i\}_{i\in\mathcal{N}}\}$ (Osborne & Rubinstein, 1994, p. 200). Concretely, $\mathcal{N}$ is the set of players, $\mathcal{H}$ the set of all possible histories/nodes. For each node $h \in \mathcal{H}$, $P(h) \in \mathcal{N}$ represents the player taking action at node h, and $A(h)$ denotes the actions available at node h. For any player $i \in \mathcal{N}$, the set of nodes h that player i can not distinguish their correct position in the tree $\mathcal{H}$ (according to the disclosed information) is called an **information set** (infoset). All infosets belonging to player i constitute a partition $\mathcal{I}_i$ of $\mathcal{H}$, and $\mathcal{I} = \cup_{i\in\mathcal{N}}\mathcal{I}_i$. For any infoset $I \in \mathcal{I}_i$, since any two histories $h, h' \in I$ are indistinguishable to player i, we have $A(h) = A(h')$ and denote $A(I) \triangleq A(h)$ for simplicity. We use $C_i(I, a) \subseteq \mathcal{I}_i$ to denote the set of infosets that is encountered immediately after player i choosing action $a \in A(I)$. Denote $A_{max} \triangleq \max_{I\in\mathcal{I}} |A(I)|$ and $C_{max} \triangleq \max_{i\in\mathcal{N}, I\in\mathcal{I}_i, a\in A(I)} |C_i(I, a)|$. When player i reaches the terminal node $z \in \mathcal{Z} \subseteq \mathcal{H}$, it receives the utility $u_i(z) \in [-1,\ 1]$. In two-player zero-sum EFGs, $u_0(z) = -u_1(z), \forall z \in \mathcal{Z}$. Let H be the maximum number of actions taken along any path from the root to a leaf node.

Sequence-Form strategy and Behavioral strategy. The sequence-form strategy space is originally proposed as a *treeplex* by Hoda et al. (2010). For any player $i \in \mathcal{N}$, the sequence is an infoset-action pair $(I, a) \in \Sigma_i \triangleq \{(I, a) : I \in \mathcal{I}_i, a \in A(I)\}$. For any sequence-form strategy $\boldsymbol{x}_i \in \boldsymbol{\mathcal{X}}_i$, it is indexed by all sequences, and $\boldsymbol{x}_i(I, a)$ represents the probability that player i reaches the sequence (I, a) from the root node r of tree. We assume that $\forall i \in \mathcal{N}$ and $\boldsymbol{x}_i \in \boldsymbol{\mathcal{X}}_i$, $\|\boldsymbol{x}_i\|_1 \le D$. The sequence-form strategy $\boldsymbol{x}_i$ can be considered as the **global strategy** over the whole tree $\mathcal{H}$, and the following behavioral strategy σ_i can be regarded as the **local strategy** defined on each infoset $I \in \mathcal{I}_i$. Concretely, the behavioral strategy $\sigma_i(I) = [\sigma_i(I, a)|a \in A(I)] \in \Delta^{|A(I)|}$ at infoset I is a probability vector, where $\sigma_i(I, a)$ is the probability of player i taking action $a \in A(I)$. If all players follow the strategy profile $\sigma = \{\sigma_i\}_{i\in\mathcal{N}}$, then the reaching probability to infoset I is denoted by $\pi^\sigma(I)$. The contribution of i to this probability is $\pi_i^\sigma(I)$ and $\pi_{-i}^\sigma(I)$ is the contribution of other players. The overall value of a strategy profile σ to player i is the expected payoff over all terminal nodes, $u_i(\sigma) = \sum_{h\in\mathcal{Z}} u_i(h)\pi^\sigma(h)$. It is not difficult to see that there exists a one-one mapping between the sequence-form strategy $\boldsymbol{x}_i \in \boldsymbol{\mathcal{X}}_i$ and the behavioral strategy $\sigma_i \in \times_{I\in\mathcal{I}_i}\Delta^{|A(I)|}$. More precisely, for any $i \in \mathcal{N}, I \in \mathcal{I}_i, a \in A(I)$, we have $\boldsymbol{x}_i(I, a) = \pi_i^\sigma(I)\sigma_i(I, a)$.

Nash equilibrium (NE). A NE is a behavioral strategy profile σ^* where no player can increase her utility $u_i(\sigma^*)$ by unilaterally changing her strategy σ_i^*. Under the sequence-form strategy framework (Von Stengel, 1996; Hoda et al., 2010), learning a NE of two-player zero-sum EFGs can be reformulated as a bilinear saddle point problem: $\min_{\boldsymbol{x}_0\in\boldsymbol{\mathcal{X}}_0}\max_{\boldsymbol{x}_1\in\boldsymbol{\mathcal{X}}_1} \boldsymbol{x}_0^\top \boldsymbol{A}\boldsymbol{x}_1$, where $\boldsymbol{A} \in [-1, 1]^{|\Sigma_0|\times|\Sigma_1|}$ is the payoff matrix, $\boldsymbol{\mathcal{X}}_0$ and $\boldsymbol{\mathcal{X}}_1$ are the sequence-form strategy spaces for player 0 and player 1 respectively. Denote $\boldsymbol{\mathcal{X}} = \times_{i\in\mathcal{N}}\boldsymbol{\mathcal{X}}_i$, and $\boldsymbol{\mathcal{X}}^*$ as the set of NEs, and define the loss $\boldsymbol{\ell}_0^\sigma = \boldsymbol{A}\boldsymbol{x}_1$ and $\boldsymbol{\ell}_1^\sigma = -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_0$ for players 0/1 respectively. Following previous works Farina et al. (2023); Meng et al. (2025a), we assume that there exist game-dependent constants $L > 0$ and $P > 0$, such that $\forall \boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{\mathcal{X}}$, $\|\boldsymbol{\ell}^\sigma - \boldsymbol{\ell}^{\sigma'}\|_1 \le L\|\boldsymbol{x} - \boldsymbol{x}'\|_1$ and $\|\boldsymbol{\ell}^\sigma\|_1 \le P$, where $\boldsymbol{\ell}^\sigma = [\boldsymbol{\ell}_i^\sigma | i \in \mathcal{N}]$. The duality gap/exploitability $\mathrm{Gap}(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) = \max_{\boldsymbol{x}_1\in\boldsymbol{\mathcal{X}}_1} \bar{\boldsymbol{x}}_0^\top \boldsymbol{A}\boldsymbol{x}_1 - \min_{\boldsymbol{x}_0\in\boldsymbol{\mathcal{X}}_0} \boldsymbol{x}_0^\top \boldsymbol{A}\bar{\boldsymbol{x}}_1$ measures the distance between strategy $(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1)$ from NE.

2.2 FINDING NASH EQUILIBRIUMS VIA ONLINE LEARNING

Online Learning over Sequence-Form Strategies. To find the NE $\boldsymbol{x}_i^* \in \boldsymbol{\mathcal{X}}^*$, player i chooses to output a series of sequence-form strategies $\{\boldsymbol{x}_i^t\}_{t\ge1}$ by running an online learning algorithm Shalev-Shwartz et al. (2012). Under such framework, Player i's cumulative regret is defined as $R_i^T \triangleq \max_{\boldsymbol{x}_i\in\boldsymbol{\mathcal{X}}_i} \sum_{t=1}^T \langle \boldsymbol{\ell}_i^t, \boldsymbol{x}_i^t - \boldsymbol{x}_i\rangle$, where $\boldsymbol{\ell}_i^t = \boldsymbol{\ell}_i^{\sigma^t}$ is the loss suffered at iteration t after choosing strategy $\boldsymbol{x}_i^t$. It is well known that if all players use online algorithms to minimize the cumulative regret with an upper bound $R_i^T \le \epsilon_i$, then duality gap of the average strategy profile $\sum_{t=1}^T \boldsymbol{x}^t/T$ is of order $(\epsilon_1 + \epsilon_2)/T$. Such typical algorithms include excess gap technique Hoda et al. (2010) and online mirror descent Farina et al. (2019c), all of which need to be ran with a dilated regularization function over the sequence-form strategy space $\boldsymbol{\mathcal{X}}_i$.

Counterfactual Regret Minimization (CFR) over Behavioral Strategies. Instead of directly minimizing the cumulative regret R_i^T of player i, CFR algorithm (Zinkevich et al., 2007; Farina et al., 2019b) chooses to decompose the global regret R_i^T into the sum of local counterfactual regret

$R_i^T(I)$ at each infoset I. Minimizing the counterfactual regrets $R_i^T(I)$ has been demonstrated more effectively in practice, and has led to several superhuman agents in EFGs (Bowling et al., 2015; Moravčík et al., 2017; Brown & Sandholm, 2018). Formally, given any strategy profile σ and loss $\boldsymbol{\ell}_i^\sigma$ for player i, the **counterfactual value** $\boldsymbol{v}_i^\sigma(I,a)$ of action $a \in A(I)$ and the **instantaneous counterfactual regret** $\boldsymbol{m}_i^\sigma(I)$ at infoset $I \in \mathcal{I}_i$ are defined as follow:

$$\boldsymbol{v}_i^\sigma(I,a) = -\boldsymbol{\ell}_i^\sigma(I,a) + \sum_{I' \in C_i(I,a)} \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I') \rangle, \qquad \boldsymbol{m}_i^\sigma(I) = \boldsymbol{v}_i^\sigma(I) - \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I) \rangle \mathbf{1}, \quad (1)$$

where $\boldsymbol{v}_i^\sigma(I) = (\boldsymbol{v}_i^\sigma(I,a))_{a \in A(I)}$, and σ_i represents the behavioral strategy associated with $\boldsymbol{x}_i$. Denote $R_i^T(I) = \sum_{t=1}^T \boldsymbol{m}_i^{\sigma^t}(I)$ as the **cumulative counterfactual regret** at I, (Liu et al., 2022, Lemma 2.2) obtains the following equivalence relation between $R_i^T(I)$ and cumulative regret R_i^T:

$$R_i^T = \max_{\sigma'} \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle R_i^T(I), \sigma_i'(I) \rangle = \max_{\sigma'} \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \sum_{t=1}^T \langle \boldsymbol{v}_i^{\sigma^t}(I), \sigma_i' - \sigma_i^t \rangle \quad (2)$$

Thus, CFR method need to choose an online algorithm to output strategies $\{\sigma_i^t\}_{t \geq 1}$ and minimize the local regret $\sum_{t=1}^T \langle \boldsymbol{v}_i^{\sigma^t}(I), \sigma_i' - \sigma_i^t \rangle$ in the right-hand side of Eq. (2) for any reference strategy σ_i'.

Regret Matching$^+$ as Local Regret Minimizer in CFR. When setting the local regret minimizer as regret matching$^+$ (RM$^+$) algorithm, we get CFR$^+$ Tammelin (2014), the classical CFR variant that always converges faster in practice than the original CFR Zinkevich et al. (2007). By applying Blackwell approachability Blackwell (1956); Abernethy et al. (2011), Farina et al. (2021) shows that RM$^+$ is equivalent to running online mirror descent (OMD) algorithm over the convex cone of the simplex $\Delta^{|A(I)|}$, i.e. $\mathbb{R}_{\geq 0}^{|A(I)|}$. Concretely, at each iteration t and infoset $I \in \mathcal{I}_i$, the update rule of OMD algorithm is $\boldsymbol{\theta}_I^{t+1} \in \arg\min_{\boldsymbol{\theta}_I \in \mathbb{R}_{\geq 0}^{|A(I)|}} \langle -\boldsymbol{m}_i^t(I), \boldsymbol{\theta}_I \rangle + \frac{1}{\eta} D_\psi(\boldsymbol{\theta}_I || \boldsymbol{\theta}_I^t)$, where $\eta > 0$ is the step size, $\boldsymbol{m}_i^t(I) = \boldsymbol{m}_i^{\sigma^t}(I)$ is the instantaneous counterfactual regret, and $D_\psi(\boldsymbol{u}||\boldsymbol{v}) = \psi(\boldsymbol{u}) - \psi(\boldsymbol{v}) - \langle \nabla\psi(\boldsymbol{v}), \boldsymbol{u} - \boldsymbol{v} \rangle$ is the Bregman divergence generated by function $\psi(\cdot) = \frac{1}{2}\|\cdot\|_2^2$. Then the output vector is normalized to get the behavioral strategy $\sigma_i^{t+1}(I) = \boldsymbol{\theta}_I^{t+1} / \langle \boldsymbol{\theta}_I^{t+1}, \mathbf{1} \rangle$. Farina et al. (2021) further show that $\sigma_i^{t+1}(I)$ remains unchanged for any step size $\eta > 0$, which is referred to as *infoset step size invariance* and is proved as a key property for the practical success of CFR$^+$ algorithm Chakrabarti et al. (2024).

3 Optimization Objective

Perturbed Extensive-Form Games (Perturbed EFGs). Such game is proposed to ensure that each player reaches every infoset with at least a positive probability $\gamma \in (0,1)$. Concretely, for any infoset $I \in \mathcal{I}_i$, the behavioral strategy space over the action set $A(I)$ in a γ-perturbed EFG is a γ-perturbed simplex $\Delta_\gamma^{|A(I)|}$, where for any $\hat{\sigma}_i(I) \in \Delta_\gamma^{|A(I)|}$, we have $\gamma \leq \hat{\sigma}_i(I,a) \leq 1$ for any action $a \in A(I)$. The sequence-form strategy space for player i is denoted as $\boldsymbol{\mathcal{X}}_i^\gamma$. The joint strategy space is $\boldsymbol{\mathcal{X}}^\gamma = \times_{i \in \mathcal{N}} \boldsymbol{\mathcal{X}}_i^\gamma \subseteq \boldsymbol{\mathcal{X}}$. When $\gamma \to 0$, the set of *limit points* of sequence $\{\boldsymbol{x}^{*,\gamma}\}_{\gamma > 0}$ of NEs is the set of extensive-form perfect equilibria (EFPE) Selten (1975), which is a refinement of NE.

Perturbed Regularized Extensive-Form Games. Recent works Pérolat et al. (2021; 2022); Liu et al. (2023); Abe et al. (2024); Meng et al. (2025a) adopt the Reward Transformation (RT) framework to demonstrate the last-iterate convergence of algorithms in solving EFGs. Concretely, these works transform the original EFG into a sequence of perturbed regularized EFGs by adding perturbation and regularization terms, thus the reward function (i.e. the loss $\boldsymbol{\ell}^\sigma$) will be changed. $\forall \hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}' \in \boldsymbol{\mathcal{X}}^\gamma$, the introduction of perturbation ensures the smoothness of counterfactual value $\|\boldsymbol{v}_i^{\hat{\sigma}}(I) - \boldsymbol{v}_i^{\hat{\sigma}'}(I)\|_1 \leq O(\|\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}'\|_1)$ Liu et al. (2023), and the regularization term guarantees the strongly monotonicity of loss function $O(\langle \boldsymbol{\ell}^{\hat{\sigma}} - \boldsymbol{\ell}^{\hat{\sigma}'}, \hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}' \rangle) \geq \|\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}'\|_2^2$ and the uniqueness of NE Sokota et al. (2023). Therefore, it suffices to establish the last-iterate convergence to the NEs of perturbed regularized EFGs, and then ensure the sequence of NEs of these perturbed regularized EFGs converges to the set of NEs of original EFG. Formally, the γ-perturbed μ-regularized EFG is defined as below:

$$\min_{\hat{\boldsymbol{x}}_0 \in \boldsymbol{\mathcal{X}}_0^\gamma} \max_{\hat{\boldsymbol{x}}_1 \in \boldsymbol{\mathcal{X}}_1^\gamma} \hat{\boldsymbol{x}}_0^\top \boldsymbol{A} \hat{\boldsymbol{x}}_1 + \mu D_\psi(\hat{\boldsymbol{x}}_0, \boldsymbol{r}_0) - \mu D_\psi(\hat{\boldsymbol{x}}_1, \boldsymbol{r}_1) \triangleq f_{\gamma,\mu}(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1), \quad (3)$$

where $\gamma > 0$ and $\mu > 0$ are constants, and $\boldsymbol{r} = [\boldsymbol{r}_0; \boldsymbol{r}_1] \in \boldsymbol{\mathcal{X}}$ is the reference sequence-form strategy profile. The unique NE of this perturbed regularized EFG is denoted by $\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ or $\hat{\sigma}^{*,\gamma,\mu}$. The

transformed reward/loss is then defined as follow:

$$\boldsymbol{\ell}_0^{\hat{\sigma}} = \boldsymbol{A}\hat{\boldsymbol{x}}_1 + \mu\nabla\psi(\hat{\boldsymbol{x}}_0) - \mu\nabla\psi(\boldsymbol{r}_0), \quad \boldsymbol{\ell}_1^{\hat{\sigma}} = -\boldsymbol{A}^\top\hat{\boldsymbol{x}}_0 + \mu\nabla\psi(\hat{\boldsymbol{x}}_1) - \mu\nabla\psi(\boldsymbol{r}_1). \tag{4}$$

$\boldsymbol{v}_i^{\hat{\sigma}}(I)$ and $\boldsymbol{m}_i^{\hat{\sigma}}(I)$ can be defined as in Eq. (1) by replacing the original loss $\boldsymbol{\ell}_i^{\sigma}$ with $\boldsymbol{\ell}_i^{\hat{\sigma}}$ in Eq. (4). During the online learning process, algorithm generates a sequence of perturbed sequence-form strategy profiles $\{\hat{\boldsymbol{x}}^t\}_{t\geq 1}$ (associated with the perturbed behavioral strategy profiles $\{\hat{\sigma}^t\}_{t\geq 1}$), and the corresponding sequence of transformed losses is denoted as $\{\hat{\boldsymbol{\ell}}^t = \boldsymbol{\ell}^{\hat{\sigma}^t}\}_{t\geq 1}$ based on Eq. (4).

4 Convergence of Optimistic CFR in Solving Regularized EFG

In this section, we give last-iterate convergence analysis for two classical predictive/optimistic versions of CFR$^+$ algorithm (i.e. PCFR$^+$ Farina et al. (2021) and PDCFR$^+$ Xu et al. (2024)) in solving regularized EFG in Eq. (3). Besides, we use the last-iterate convergence analysis to derive improved average-regret bounds for these two algorithms. The main technical novelties of this work and the detailed conceptual distinctions between it and the most closely related works (Meng et al., 2025a; Liu et al., 2023) are discussed in Remarks C.1 and C.2.

4.1 Invariant Last-/Average-Iterate Convergence for Predictive CFR$^+$

Within the Blackwell approachability framework Farina et al. (2021); Chakrabarti et al. (2024), PCFR$^+$ algorithm is equivalent to running optimistic online mirror descent (OOMD) algorithm Rakhlin & Sridharan (2013a;b) over the convex cone of simplex and then normalizing the output vector. Formally, at iteration t and infoset $I \in \mathcal{I}_i$, the update rule of PCFR$^+$ with step size $\eta > 0$ for learning an NE of the perturbed regularized EFG in Eq. (3) is

$$\boldsymbol{\theta}_I^{t+1} \in \underset{\boldsymbol{\theta}_I \in \mathbb{R}_{\geq 0}^{|A(I)|}}{\arg\min}\ \langle -\hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}_I\rangle + \frac{1}{\eta} D_\psi(\boldsymbol{\theta}_I \| \hat{\boldsymbol{\theta}}_I^t), \quad \hat{\boldsymbol{\theta}}_I^{t+1} \in \underset{\boldsymbol{\theta}_I \in \mathbb{R}_{\geq 0}^{|A(I)|}}{\arg\min}\ \langle -\hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I\rangle + \frac{1}{\eta} D_\psi(\boldsymbol{\theta}_I \| \hat{\boldsymbol{\theta}}_I^t). \tag{5}$$

where the instantaneous counterfactual regret is denoted as $\hat{\boldsymbol{m}}_i^t(I) = \boldsymbol{m}_i^{\hat{\sigma}^t}(I)$ for conciseness, the distance generating function $\psi(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{\theta}\|_2^2$. The output unnormalized strategy of the first update step in Eq. (5) has a closed-form solution $\boldsymbol{\theta}^{t+1} = [\hat{\boldsymbol{\theta}}^t + \eta\hat{\boldsymbol{m}}_i^t(I)]^+$. Then, the output behavioral strategy and the perturbed behavioral strategy for solving Eq. (3) are computed as follow

$$\sigma_i^{t+1}(I) = \boldsymbol{\theta}_I^{t+1}/\langle \boldsymbol{\theta}_I^{t+1}, \mathbf{1}\rangle, \qquad \hat{\sigma}_i^{t+1}(I) = (1-\alpha_I)\sigma_i^{t+1}(I) + \gamma\mathbf{1}, \tag{6}$$

where $\alpha_I = \gamma|A(I)|$. The corresponding perturbed sequence-form strategy can be determined as $\hat{\boldsymbol{x}}_i^{t+1}(I) = \pi_i^{\hat{\sigma}^{t+1}}(I)\hat{\sigma}_i^{t+1}(I)$. Besides, it is not difficult to see that the update rule of CFR$^+$ algorithm is a special case of PCFR$^+$ by setting $\hat{\boldsymbol{m}}_i^t(I) = 0$ in the first update of Eq. (5). We then give the asymptotic last-iterate convergence of PCFR$^+$ algorithm as follow, which involves the last-iterate convergence for reward transformation-based CFR$^+$ algorithm Meng et al. (2025a) as a special case.

Theorem 4.1 (**Last-Iterate Convergence of Predictive CFR$^+$**). *If all players use PCFR$^+$ with any $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$ and $\eta > 0$, then the strategy profiles $\{\hat{\boldsymbol{x}}^t\}_{t\geq 1}$ converges to the unique NE of perturbed regularized EFGs in Eq. (3) with any $\gamma > 0$ and $\mu > 0$, i.e. $\lim_{T\to\infty}\|\hat{\boldsymbol{x}}^T - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2 = 0$.*

In the perturbed regularized EFGs, Theorem 4.1 provides the first asymptotic last-iterate convergence for optimistic CFR algorithms that holds for any step size $\eta > 0$ and any initialization (which we call the ***invariant last-iterate convergence***, hereafter). Previous asymptotic last-iterate convergence with any step size $\eta > 0$ holds for the non-optimistic CFR variant, CFR$^+$ Meng et al. (2025a). Another asymptotic last-iterate convergence in perturbed regularized EFGs within the CFR framework is (Liu et al., 2023, Thm 5.3) for Reg-CFR algorithm, which uses an optimistic OMD algorithm as the local regret minimizer with adaptive step sizes. However, Reg-CFR (Liu et al., 2023, Eq (5.2)) directly solves the constrained optimization problem over the γ-perturbed simplex, which requires additional projection to the γ-perturbed simplex and is less efficient than PCFR$^+$ that has a closed-form behavioral strategy $\hat{\sigma}_i(I)$ in Eq. (6).

Based on the invariant last-iterate convergence in Theorem 4.1 that holds for any initialization $\hat{\boldsymbol{\theta}}^0$ and any step size $\eta > 0$, we further establish the upper bound on cumulative regret R_i^T and the average-iterate convergence of PCFR$^+$ for perturbed regularized EFGs with any $\gamma \geq 0$ and $\mu \geq 0$.

Theorem 4.2 (**Average-Iterate Convergence of Predictive CFR$^+$**). *Assume that player i follow the update rule of PCFR$^+$ to solve the perturbed regularized EFGs in Eq. (3) with any $\gamma \geq 0$ and $\mu \geq 0$. Then, for any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$ and any step size $\eta > 0$, the per-player individual regret is upper bounded as: $R_i^T \leq \sum_{I \in \mathcal{I}_i} 2\sqrt{\|\hat{\boldsymbol{\theta}}_I^0\|_2^2/\eta^2 + \frac{1}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2}$.*

The above regret bound holds not only for the strictly perturbed regularized EFG with $\gamma > 0, \mu > 0$, but also holds for PCFR$^+$ algorithm in solving the original non-perturbed EFG with $\gamma = \mu = 0$. When $\hat{\boldsymbol{\theta}}_I^0 = \mathbf{0}$, it recovers the regret bound for PCFR$^+$ in (Farina et al., 2021, Thm 3). To the best of our knowledge, this is the first time to derive regret bound for the celebrated PCFR$^+$ algorithm without using its complicated Blackwell approachability analysis (Farina et al., 2021, Prop 2). We then give the following remark to show the key differences between these two regret bounds.

Remark 4.1. *(Farina et al., 2021, Eq. (24)) gives the regret bound for PCFR$^+$ at an inforset with any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$ and step size $\eta > 0$ as: $\frac{1+\|\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta} + \eta\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2$, which is of order $O(1 + \sum_t\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2)$ and is larger than our bound $O(\sqrt{1 + \sum_t\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2})$ in Thm 4.2. Even setting the optimal η in (Farina et al., 2021, Eq. (24)), its lowest value $2\sqrt{(1 + \|\hat{\boldsymbol{\theta}}_I^0\|_2^2)\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2}$ is larger than ours when $\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 \geq 1/\eta^2$, validating the improvement of our average-iterate convergence analysis. More comparisons between regret bounds of different CFR variants are listed in Table 1.*

4.2 Invariant Last-/Average-Iterate Convergence for Discounted PCFR$^+$

In this section, we integrate the discounting technique Brown & Sandholm (2019) into PCFR$^+$ method to solve perturbed regularized EFG in Eq. (3). Such technique discounts players' regrets from earlier iterations, and has shown promising advantage in solving large-scale EFGs (like heads-up no-limit Texas hold'em). Formally, the update rule of PDCFR$^+$ for learning an NE of the perturbed regularized EFGs in Eq. (3) at iteration t and infoset $I \in \mathcal{I}_i$ is (we write λ_I^t as λ^t for simplicity):

$$\boldsymbol{\theta}_I^{t+1} = [\lambda^{t+1}\hat{\boldsymbol{\theta}}_I^t + \hat{\boldsymbol{m}}_i^t(I)]^+, \quad \hat{\boldsymbol{\theta}}_I^{t+1} = [\lambda^{t+1}\hat{\boldsymbol{\theta}}_I^t + \hat{\boldsymbol{m}}_i^{t+1}(I)]^+, \tag{7}$$

where $\{\lambda^t\}_{t\geq 0}$ is the non-decreasing positive sequence with an upper bound λ (i.e. $0 < \lambda^t \leq \lambda^{t+1} \leq \lambda$), and the output behavioral strategy $\hat{\sigma}_i^{t+1}(I)$ is computed as in Eq. (6). Due to the linearity of the ReLU function $[\cdot]^+ = \max(\cdot, 0)$ and the ℓ_1-normalizing operator, the update rule of PDCFR$^+$ in Eq. (7) is equivalent to running optimistic OMD algorithm in Eq. (5) with non-increasing step sizes $\{\eta^t\}_{t\geq 0}$ (i.e. $\lambda^t = 1/\eta^t$). The non-decreasing positive sequence $\{\lambda^t\}_{t\geq 0}$ is widely used in solving EFGs: in DCFR$^+$, $\lambda^t = \frac{t^\alpha}{1+t^\alpha} \leq 1$ (with $\alpha > 0$) and is called *discounting factor*; in Reg-CFR Liu et al. (2023) and AdOGD Zhang et al. (2026), the adaptive coefficient $\lambda_I^t = (\sum_{s=1}^{t-1}\|\hat{\boldsymbol{m}}_i^s(I) - \hat{\boldsymbol{m}}_i^{s-1}(I)\|^2 + \delta)^{1/2}$ with $\delta > 0$.

Then, we give the asymptotic last-iterate convergence for PDCFR$^+$ in solving perturbed EFG. Such asymptotic last-iterate convergence also holds for DCFR$^+$ Li et al. (2024), and holds for RT-based AdOGD algorithms Zhang et al. (2026) if its adaptive step size at any inforset is upper bounded.

Theorem 4.3 (**Last-Iterate Convergence of PDCFR$^+$**). *If all players use PDCFR$^+$ with any $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$ and non-decreasing upper-bounded sequence $\{\lambda^t\}_{t\geq 1}$, then the strategy $\hat{\boldsymbol{x}}^t$ converges to the NE of perturbed regularized EFG in Eq. (3) for any $\gamma > 0, \mu > 0$, i.e. $\lim_{T\to\infty}\|\hat{\boldsymbol{x}}^T - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2 = 0$.*

In Theorem 4.3, the requirement for a non-decreasing sequence λ^t (which corresponds to a non-increasing step size η^t in Eq. (5)) is more general than the constant step size $\eta > 0$ used in Theorem 4.1, and thus applies to a broader class of problems. According to its proof in Lemma G.2, it is the linearity of ReLU function in OOMD algorithm with ℓ_2-regularization function (which does not exist in optimistic MWU algorithm) that makes the key contribution to the invariant last-iterate convergence of PDCF$^+$ for *any* non-decreasing sequence of discounting factors $\{\lambda^t\}_{t\geq 1}$.

Furthermore, we use the invariance property analysis of last-iterate convergence to establish the average-iterate convergence of PDCFR$^+$ for perturbed regularized EFGs. Specifically, we prove the convergence of weighted-average strategy profile $\bar{\boldsymbol{x}}^{\boldsymbol{w}} = \sum_{t=1}^T w^t\boldsymbol{x}^t / \sum_{t=1}^T w^t$ as in Xu et al. (2024). When $w^t = 1$, the weighted-average strategy reduces to the uniform average strategy $\sum_{t=1}^T \boldsymbol{x}^t/T$.

Theorem 4.4 (**Average-Iterate Convergence of PDCFR$^+$**). *Assume that player i follow the update rule of PDCFR$^+$ to solve the perturbed regularized EFG in Eq. (3) with any $\gamma \geq 0, \mu \geq 0$. If $\{w^t\lambda^t\}_{t\geq 0}$ is a positive non-increasing sequence, then for any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$, we have*

$$R_i^{T,\boldsymbol{w}} \triangleq \sum_{t=0}^{T-1} w^{t+1}\langle \hat{\boldsymbol{\ell}}_i^{t+1}, \hat{\boldsymbol{x}}_i^{t+1} - \hat{\boldsymbol{x}}_i\rangle \leq 2\sum_{I\in\mathcal{I}_i}\sqrt{w^1\lambda^1\Big[w^1\lambda^1\|\hat{\boldsymbol{\theta}}_I^0\|_2^2 + \sum_{t=0}^{T-1}\frac{w^{t+1}}{2\lambda^{t+1}}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big]}.$$

Analogous to Remark 4.1, the above bound reduces to the result (Xu et al., 2024) for original EFGs under initialization $\hat{\boldsymbol{\theta}}_I^0 = \mathbf{0}$, and remains tighter for arbitrary initialization $\hat{\boldsymbol{\theta}}_I^0$.

5 Convergence to Approximate Nash Equilibrium of Original EFG

In this section, we first investigate how the unique NE of perturbed regularized EFG converge to the NEs of original EFG. Then, we show how the last-iterate strategy of reward-transformation based PCFR$^+$/PDCFR$^+$ in Algorithm 1 converge to the set of NEs of original EFG in terms of duality gap.

5.1 Convergence of the Unique Nash Equilibrium of Perturbed Regularized EFG

We first explore whether the NE of the perturbed regularized EFG in Eq. (3) will *converge pointwise* to an exact NE of the original EFG, and what property such a limit point (if exists) possesses. Due to the strongly-convex strongly-concave property of perturbed regularized EFG with any $\gamma > 0, \mu > 0$, such game has only one NE $\hat{\boldsymbol{x}}^{*,\gamma,\mu}$. The following proposition shows that if the double limit $\lim_{(\gamma,\mu)\to(0,0)}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ exists, then it is equal to the iterated limit $\lim_{\mu\to 0}\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$.

Proposition 5.1. *For the γ-perturbed μ-regularized EFG in Eq. (3), if the double limit $\lim_{(\gamma,\mu)\to(0,0)}\hat{\boldsymbol{x}}^{*,\gamma,\mu} = \hat{\boldsymbol{x}}^*$ exists, then the iterated limit $\lim_{\mu\to 0}\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ exists and $\lim_{\mu\to 0}\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu} = \hat{\boldsymbol{x}}^*$.*

When the original EFG has the unique NE, the double limit $\lim_{(\gamma,\mu)\to(0,0)}\hat{\boldsymbol{x}}^{*,\gamma,\mu} = \hat{\boldsymbol{x}}^*$ exists (see more details in Lemma I.1). The EFG example that has iterated limit but no double limit can be found in Prop.4.1 in Bernasconi et al. (2024). Therefore, Proposition 5.1 indicates that the iterated limit $\lim_{\mu\to 0}\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ is a more general notion of equilibrium than the double limit, as it is well-defined under broader conditions. Consequently, we postulate the existence of the iterated limit and derive the conditions for tuning discrete perturbation/regularization parameters that guarantee the pointwise convergence of the perturbed regularized EFG's NEs to the NEs of original EFG.

Proposition 5.2. *For any EFG with perfect recall, if the iterated limit $\lim_{\mu\to 0}\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ exists, then there exist a function $g : \mathbb{R}^+\backslash\{0\} \to \mathbb{R}^+\backslash\{0\}$ such that, for any sequence $\{(\gamma_k,\mu_k)\}_{k\geq 1}$ satisfying $\lim_{k\to\infty}\mu_k = 0$, $\lim_{k\to\infty}\gamma_k = 0$ and $\gamma_k < g(\mu_k)$, the limit $\lim_{k\to\infty}\hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}$ exists and is a Nash Equilibirum of the original EFG.*

Notice that another related work (Bernasconi et al., 2024, Thm 4.3) assumes the existence of the iterated limit $\lim_{\gamma\to 0}\lim_{\mu\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$. This assumption is more restrictive than ours, because it implicitly requires the existence of both inner limit $\lim_{\mu\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ and outer limit $\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma}$. In contrast, for the iterated limit $\lim_{\mu\to 0}\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ in Proposition 5.2, the existence of the inner limit $\lim_{\gamma\to 0}\hat{\boldsymbol{x}}^{*,\gamma,\mu}$ is guaranteed by the strong convexity-strong concavity property of the regularized EFG (see Lemma I.1). Thus, we only need to assume the existence of outer limit $\lim_{\mu\to 0}\hat{\boldsymbol{x}}^{*,\mu}$.

5.2 Approximation to the NE of Original EFG with Small Regularization

Section 5.1 has shown that the convergence of $\hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}$ to the NE of original EFG requires strong assumptions. In this section, we choose to study the convergence of RTPCFR$^+$/RTPDCFR$^+$ in Algorithm 1 to the approximate NE of original EFG in terms of duality gap. The next proposition gives an upper bound on the duality gap in terms of the distance to the NE of perturbed EFG.

Proposition 5.3. *Define $|\Sigma| = \max\{|\Sigma_0|, |\Sigma_1|\}$ Let $\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) \in \boldsymbol{\mathcal{X}}_0 \times \boldsymbol{\mathcal{X}}_1$ be such that $\|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \bar{\boldsymbol{x}}\|_2 \leq \nu$ for some $\nu \geq 0$, then we have $\mathrm{Gap}(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) \leq 2D\nu|\Sigma| + 6\mu D^2$.*

When setting $\bar{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{T_k+1}$ as the last-iterate output of Algorithm 1 in the k-th phase, the last-iterate convergence $\lim_{T_k\to\infty}\|\hat{\boldsymbol{x}}^{T_k+1} - \hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}\|_2 = 0$ in Section 4 guarantees that, for any given $\epsilon > 0$, when the regularization factor $\mu < \epsilon/(12D^2)$, and the iterate number T_k is large enough such that $\|\hat{\boldsymbol{x}}^{T_k+1} - \hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}\|_2 \leq \epsilon/(4D|\Sigma|)$, then $\hat{\boldsymbol{x}}^{T_k+1}$ forms an ϵ-NE of the original EFG. Next, we give another bound on the duality gap in terms of the duality gap in the perturbed regularized EFG.

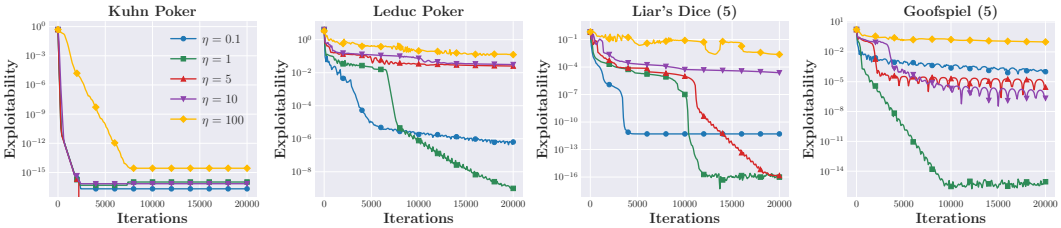

Figure 1: Invariant last-iterate convergence of RTPCFR$^+$ w.r.t. different step sizes

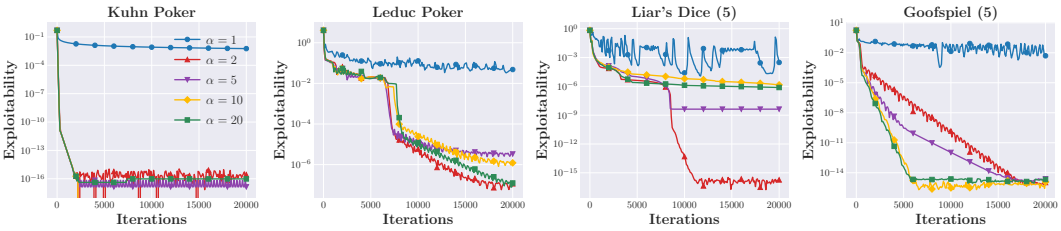

Figure 2: Invariant last-iterate convergence of RTPDCFR$^+$ w.r.t. different discounting factors.

Proposition 5.4. *Let $F(\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x}_1, -\boldsymbol{A}^\top \boldsymbol{x}_0)$. The approximate NE $\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) \in \boldsymbol{\mathcal{X}}_0^\gamma \times \boldsymbol{\mathcal{X}}_1^\gamma$ of a γ-perturbed EFG is an approximate of the NE of original EFG in terms of duality gap:*

$$\max_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} \langle F(\bar{\boldsymbol{x}}), \bar{\boldsymbol{x}} - \boldsymbol{x} \rangle \leq \max_{\hat{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}^\gamma} \langle F(\bar{\boldsymbol{x}}), \bar{\boldsymbol{x}} - \hat{\boldsymbol{x}} \rangle + 2\gamma P H A_{\max}$$

The above proposition characterizes the convergence of the last-iterate strategy $\hat{\boldsymbol{x}}^{T_k+1}$ or the average-iterate strategy $\sum_{t=1}^{T_k} \hat{\boldsymbol{x}}^t / T$ in the k-th phase to the NE of original EFG. Furthermore, the upper bound of order $O(\gamma P H A_{\max})$ in Proposition 5.4 is tighter than the bound of $O(\gamma P^2)$ in (Liu et al., 2023, Lemma F.4) since $P \gg H A_{\max}$. This improvement stems from the efficient computation of the perturbed behavioral strategy $\hat{\sigma}_i$ (see Eq. (6)), instead of solving a γ-perturbed constrained optimization problem in (Liu et al., 2023, Eq (5.2)). Combining the above two propositions obtains the following result.

Theorem 5.1. *Denote* $\mathrm{Gap}_\gamma(\hat{\boldsymbol{x}}_0^{T_k+1}, \hat{\boldsymbol{x}}_1^{T_k+1}) = \max_{\hat{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}^\gamma} \langle F(\hat{\boldsymbol{x}}^{T_k+1}), \hat{\boldsymbol{x}}^{T_k+1} - \hat{\boldsymbol{x}} \rangle$. *Then, the output of RTPCFR+/RTPDCFR+ algorithm (i.e. Algorithm 1) in the k-th phase is an approximate of the NE of original EFG in terms of duality gap, i.e.,* $\mathrm{Gap}(\hat{\boldsymbol{x}}_0^{T_k+1}, \hat{\boldsymbol{x}}_1^{T_k+1})$

$$\leq \min\{2D|\Sigma| \|\hat{\boldsymbol{x}}^{T_k+1} - \hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}\|_2 + 6\mu_k D^2, \quad \mathrm{Gap}_{\gamma_k}(\hat{\boldsymbol{x}}_0^{T_k+1}, \hat{\boldsymbol{x}}_1^{T_k+1}) + 2\gamma_k P H A_{\max}\}$$

Theorem 5.1 characterizes the convergence rate of the last-iterate strategy $\hat{\boldsymbol{x}}^{T_k+1}$ (generated by Algorithm 1 in the k-th phase) to the NEs of original EFG. Roughly, the approximation error is bounded by $O(\|\hat{\boldsymbol{x}}^{T_k+1} - \hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}\|_2 + \min\{\mu_k, \gamma_k\})$ in terms of the duality gap. By employing small perturbation and regularization parameters γ_k and μ_k, our Algorithm 1 yields an approximate NE of the original EFG after a sufficient number of iterations in the k-th phase.

6 Experiments

Experimental Settings. We conduct experiments on eight small-to-medium-sized instances of five EFG benchmarks: Kuhn Poker, Leduc Poker, Goofspiel, Liar's Dice, and Battleship (see Table 2 for details). **Source code is provided in supplementary material**. We compare our methods with three groups of baselines: (1) Traditional CFR algorithms like CFR$^+$ Tammelin et al. (2015), DCFR Brown & Sandholm (2019), and PCFR$^+$ Farina et al. (2021); (2) Recent algorithms that update the sequence-form strategy over the treeplex and enjoy last-iterate convergence, like OMWU and OGDA Lee et al. (2021). (3) Recent CFR variants that are proposed within the reward transformation framework and enjoy last-iterate convergence guarantees, like Reg-CFR Liu et al. (2023) and RTCFR$^+$ Meng et al. (2025a). All algorithms are built on the open-source LiteEFG code Liu et al. (2024). For RTPCFR$^+$, we set the initial values $\gamma = 1\mathrm{e}{-10}$, $\mu = 1\mathrm{e}{-4}$, the number of iterations $T_k = 100$ to update $\boldsymbol{r}$. For RTPDCFR$^+$, $\alpha = 1.4, \mu = 0.0005, T_k = 1000$. We set hyperparameters of other baselines as in their original papers (see App. L). All algorithms are ran for 20,000 ($N = 20000/T_k$) iterations. More experiments, including the last-iterate results of RTPDCFR$^+$ (Fig. 5), the comparison between last-iterate and average-iterate convergence of CFR variants (Fig. 6), are deferred to Appendix L.

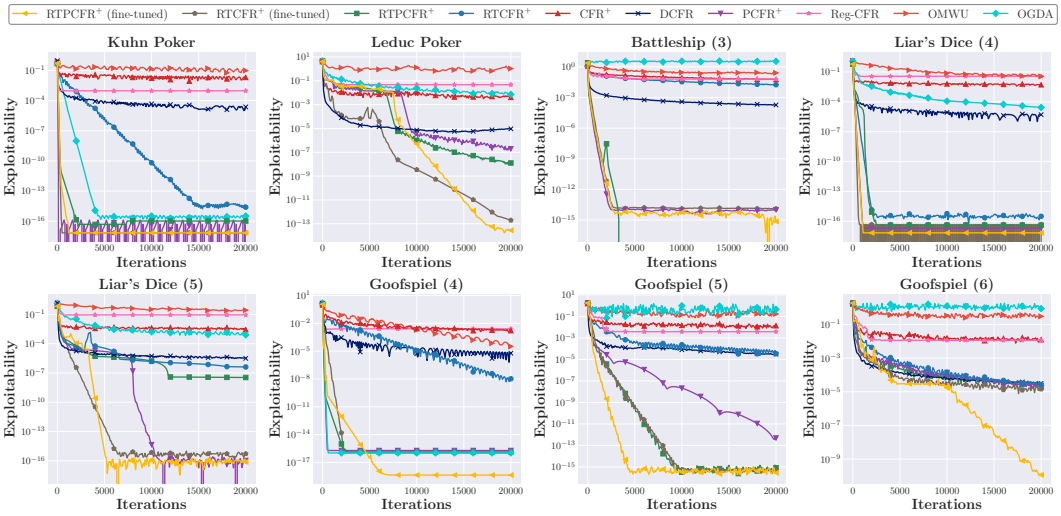

Figure 3: Last-iterate convergence of different algorithms in small-to-medium-sized EFGs.

Invariant convergence. Experiments in Figs 1-2 show that: **(1)** RTPCFR$^+$ and RTPDCFR$^+$ achieves last-iterate convergence across a wide range of step sizes η and discounting factors α. **(2)** However, extreme parameter choices, such as a large step size $\eta = 100$ or a small discounting factor $\alpha = 1$, may result in slow convergence.

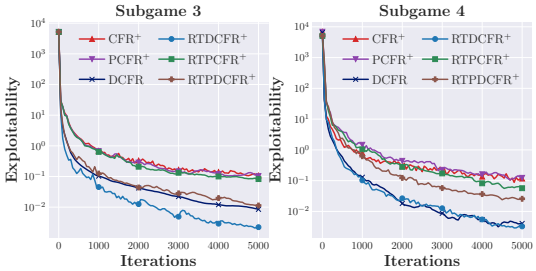

Figure 4: Last-iterate convergence of different CFR algorithms in large-size EFGs.

Last-Iterate convergence in small-to-medium-sized EFGs. In Fig. 3, we can conclude that: **(1)** RTPCFR$^+$ achieves faster convergence than the latest RT-based method RTCFR$^+$ in most benchmarks except Leduc Poker. In some EFGs like Battleship, Goofspiel, and Goofspiel (5), RTPCFR$^+$ outperforms RTCFR$^+$ by over five orders of magnitude in speed. **(2)** RTPCFR$^+$ achieves performance comparable to PCFR$^+$. Specifically, it exhibits faster last-iterate convergence in Leduc Poker, Battleship, Goofspiel (5), while matching PCFR$^+$ in Liar's Dice (4) and Goofspiel (4). **(3)** RTPCFR$^+$ (fine-tuned) also converges faster than the state-of-the-art RTCFR$^+$ (fine-tuned) in all experiments, and such improvement is up to five orders of magnitude in Goofspiel (6), demonstrating the strengths of optimism technique in RT framework.

Last-Iterate convergence in Large-size EFGs. The empirical last-iterate convergence of different CFR algorithms on two large-size EFG benchmarks, HUNL Subgame (3) and Subgame (4) Brown & Sandholm (2019), are shown in Fig. 4. We can find that: **(1)** The previous state-of-the-art DCFR Brown & Sandholm (2019) still achieves better faster convergence rates and lower exploitabilities than the proposed RTPCFR$^+$. **(2)** RTPDCFR$^+$ achieves comparable result with DCFR in Subgame (3), but performs worse in Subgame (4). **(3)** Nevertheless, the proposed RTDCFR$^+$ (with $\alpha = 1.3, \mu = 0.01, T_k = 1000$), the non-predictive version of RTPDCFR$^+$, yields lower exploitabilities than DCFR in both Subgame (3) and Subgame (4).

7 Conclusions

We establish invariant last-iterate convergence for predictive and discounted CFR variants under the reward transformation framework. For PCFR$^+$, the convergence holds for arbitrary step sizes and strategy initializations, while for PDCFR$^+$ it further holds for any non-decreasing and upper-bounded sequence of discounting factors. Our analysis also yields improved average-regret bounds for optimistic CFR methods. Moreover, we characterize how the perturbation and regularization coefficients affect the approximation of NEs of the original EFG in terms of the duality gap. Experiments validate our theoretical results and demonstrate the practical benefits of incorporating optimism and discounting into the reward transformation framework. An important but challenging direction for future work is to establish explicit finite-time last-iterate convergence rates for regularization-based CFR variants and develop adaptive schemes for selecting the perturbation and regularization coefficients.

AI USE STATEMENT

In this work, we used generative AI tools for [language polishing]. We have not used generative AI tools for [help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, assist with translation, interpret results], and [generate synthetic data sets,clean and reformat dataset, support qualitative and thematic data analysis] are not applicable to this work.

ETHICS STATEMENT

Intended as a theoretical contribution to the study of games in extensive form, this work may have wider societal ramifications; nevertheless, we do not think any particular one warrants explicit attention here.

REPRODUCIBILITY STATEMENT

For the comparison methods, the corresponding outcomes are obtained by executing the implementation made publicly available at `https://github.com/menglinjian/NeurIPS-2025-RTCFR`. The source code is also provided in the supplementary material.

REFERENCES

Kenshi Abe, Kaito Ariu, Mitsuki Sakamoto, and Atsushi Iwasaki. Adaptively perturbed mirror descent for learning in games. In *International Conference on Machine Learning (ICML)*, 2024.

Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *Annual Conference on Learning Theory (COLT)*, pp. 27–46, 2011.

Yu Bai, Chi Jin, Song Mei, Ziang Song, and Tiancheng Yu. Efficient phi-regret minimization in extensive-form games via online mirror descent. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

James P. Bailey and Georgios Piliouras. Multiplicative weights update in zero-sum games. In *ACM Conference on Economics and Computation (EC)*, pp. 321–338, 2018.

Martino Bernasconi, Alberto Marchesi, and Francesco Trovò. Learning extensive-form perfect equilibria in two-player zero-sum sequential games. In *International Conference on Artificial Intelligence and Statistics (AISTAT)*, pp. 2152–2160, 2024.

David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 1956.

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

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1829–1836, 2019.

Noam Brown, Adam Lerer, Sam Gross, and Tuomas Sandholm. Deep counterfactual regret minimization. In *International Conference on Machine Learning (ICML)*, pp. 793–802, 2019.

Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Weiqiang Zheng. Last-iterate convergence properties of regret-matching algorithms in games. In *The Thirteenth International Conference on Learning Representations, (ICLR)*, 2025.

Darshan Chakrabarti, Julien Grand-Clément, and Christian Kroer. Extensive-form game solving via Blackwell approachability on treeplexes. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Annual Conference on Learning Theory (COLT)*, pp. 6.1–6.20, 2012.

Constantinos Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *Innovations in Theoretical Computer Science Conference (ITCS)*, pp. 27:1–27:18, 2019.

Zhiyuan Fan, Christian Kroer, and Gabriele Farina. On the optimality of dilated entropy and lower bounds for online learning in extensive-form games. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. Stable-predictive optimistic counterfactual regret minimization. In *International Conference on Machine Learning (ICML)*, pp. 1853–1862, 2019a.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Online convex optimization for sequential decision processes and extensive-form games. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1917–1925, 2019b.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 5221–5231, 2019c.

Gabriele Farina, Chun Kai Ling, Fei Fang, and Tuomas Sandholm. Correlation in extensive-form games: Saddle-point formulation and benchmarks. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 9229–9239, 2019d.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. In *International Conference on Machine Learning (ICML)*, pp. 6337–6357, 2022.

Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, and Haipeng Luo. Regret matching+: (in)stability and fast convergence in games. In *annual Conference on Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 61546–61572, 2023.

Zachary Feinstein. Continuity and sensitivity analysis of parameterized nash games. *Economic Theory Bulletin*, 10:2196–1093, 2022.

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

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

Samid Hoda, Andrew Gilpin, Javier Pena, and Tuomas Sandholm. Smoothing techniques for computing nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2): 494–512, 2010.

Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to Nash equilibrium. In *Conference on Learning Theory (COLT)*, pp. 2388–2422, 2021.

Vishruti Kakkad, Hitarth Shah, Reema Patel, and Nishant Doshi. A comparative study of applications of game theory in cyber security and cloud computing. *Procedia Computer Science*, 155:680–685, 2019.

Christian Kroer, Kevin Waugh, Fatma Kılınç-Karzan, and Tuomas Sandholm. Faster algorithms for extensive-form game solving via improved smoothing functions. *Mathematical Programming*, 179 (1):385–417, 2020.

Harold W. Kuhn. A simplified two-person poker. In *Contributions to the Theory of Games*, volume 1, pp. 97–103. Princeton University Press, 1950.

Marc Lanctot, Kevin Waugh, Martin Zinkevich, and Michael Bowling. Monte carlo sampling for regret minimization in extensive games. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 1078–1086, 2009.

Chung-Wei Lee, Christian Kroer, and Haipeng Luo. Last-iterate convergence in extensive-form games. In *Annual Conference on Neural Information Processing (NeurIPS)*, pp. 14293–14305, 2021.

Qi Lei, Sai Ganesh Nagarajan, Ioannis Panageas, et al. Last iterate convergence in no-regret learning: constrained min-max optimization for convex-concave landscapes. In *International Conference on Artificial Intelligence and Statistics (AISTAT)*, pp. 1441–1449, 2021.

Kai Li, Hang Xu, Haobo Fu, Qiang Fu, and Junliang Xing. Automatically designing counterfactual regret minimization algorithms for solving imperfect-information games. *Artif. Intell.*, 337:104232, 2024.

Viliam Lisý, Marc Lanctot, and Michael H. Bowling. Online monte carlo counterfactual regret minimization for search in imperfect information games. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 27–36, 2015.

Mingyang Liu, Asuman E. Ozdaglar, Tiancheng Yu, and Kaiqing Zhang. The power of regularization in solving extensive-form games. In *International Conference on Learning Representations (ICLR)*, 2023.

Mingyang Liu, Gabriele Farina, and Asuman Ozdaglar. LiteEFG: An efficient python library for solving extensive-form games. *arXiv preprint arXiv:2407.20351*, 2024.

Weiming Liu, Huacong Jiang, Bin Li, and Houqiang Li. Equivalence analysis between counterfactual regret minimization and online mirror descent. In *International Conference on Machine Learning (ICML)*, pp. 13717–13745, 2022.

Linjian Meng, Tianpei Yang, Youzhi Zhang, Zhenxing Ge, Shangdong Yang, Tianyu Ding, Wenbin Li, Bo An, and Yang Gao. Efficient last-iterate convergence in solving extensive-form games. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025a.

Linjian Meng, Youzhi Zhang, Zhenxing Ge, Tianyu Ding, Shangdong Yang, Zheng Xu, Wenbin Li, and Yang Gao. Last-iterate convergence of smooth Regret Matching+ variants in learning Nash equilibria, 2025b.

Linjian Meng, Youzhi Zhang, Shangdong Yang, Wenbin Li, Tianyu Ding, and Yang Gao. A faster parameter-free regret matching algorithm. In *International Conference on Learning Representations (ICLR)*, 2026.

Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations (ICLR)*, 2018a.

Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2703–271, 2018b.

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

Arkadij Semenovič Nemirovskij and David Borisovich Yudin. Problem complexity and method efficiency in optimization. *Journal of the Operational Research Society*, 1983.

Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, 1994.

Julien Pérolat, Rémi Munos, Jean-Baptiste Lespiau, Shayegan Omidshafiei, Mark Rowland, Pedro A. Ortega, Neil Burch, Thomas W. Anthony, David Balduzzi, Bart De Vylder, Georgios Piliouras, Marc Lanctot, and Karl Tuyls. From Poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *International Conference on Machine Learning (ICML)*, pp. 8525–8535, 2021.

Julien Pérolat, Bart De Vylder, Daniel Hennes, Eugene Tarassov, Florian Strub, Vincent de Boer, Paul Muller, Jerome T Connor, Neil Burch, Thomas Anthony, et al. Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science*, 378(6623):990–996, 2022.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory (COLT)*, pp. 993–1019, 2013a.

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 3066–3074, 2013b.

Sheldon M. Ross. Goofspiel—the game of pure strategy. *Journal of Applied Probability*, 8(3): 621–625, 1971.

Tuomas Sandholm. Steering evolution strategically: Computational game theory and opponent exploitation for treatment planning, drug design, and synthetic biology. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 4057–4061, 2015.

Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 1975.

Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Martin Shubik. The dollar auction game: a paradox in noncooperative behavior and escalation. *Journal of Conflict Resolution*, 15(1):109–111, 1971.

Samuel Sokota, Ryan D'Orazio, J. Zico Kolter, Nicolas Loizou, Marc Lanctot, Ioannis Mitliagkas, Noam Brown, and Christian Kroer. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. In *International Conference on Learning Representations (ICLR)*, 2023.

Finnegan Southey, Michael Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and Chris Rayner. Bayes' bluff: Opponent modelling in poker. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 550–558, 2005.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 2989–2997, 2015.

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

Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit texas hold'em. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 645–652, 2015.

Bernhard Von Stengel. Efficient computation of behavior strategies. *Games and Economic Behavior*, 14(2):220–246, 1996.

Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *International Conference on Learning Representations (ICLR)*, 2021.

Hang Xu, Kai Li, Bingyun Liu, Haobo Fu, Qiang Fu, Junliang Xing, and Jian Cheng. Minimizing weighted counterfactual regret with optimistic online mirror descent. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 5272–5280, 2024.

Brian Hu Zhang, Ioannis Anagnostides, and Tuomas Sandholm. Scale-invariant regret matching and online learning with optimal convergence: Bridging theory and practice in zero-sum games. In *ACM Conference on Economics and Computation (EC)*, 2026.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 905–912, 2007.

A Related Work

Last-/Average-Iterate Convergence in Perturbed Regularized EFGs. To the best of our knowledge, Pérolat et al. (2021) is the first work to employ reward transformation for solving imperfect-information games (i.e., EFGs) under continuous-time dynamics. In particular, (Pérolat et al., 2021, Thms. 5.2 & 6.1) establish linear last-iterate convergence in regularized EFGs and asymptotic last-iterate convergence in the original, unregularized EFGs for their regularized follow-the-regularized-leader algorithm. Moving to the discrete-time setting, (Liu et al., 2023, Thms. 4.1–4.2) establish the first linear last-iterate convergence result for their dilated online mirror descent (OMD) algorithm in regularized EFGs, while guaranteeing sublinear convergence to the set of Nash equilibria (NEs) in the original EFGs. Subsequently, (Bernasconi et al., 2024, Thm. 4.3) rigorously show that, for the sequence of NEs of perturbed regularized EFGs to converge pointwise to the NE set of the original EFG, the regularization must vanish faster than the perturbation. Under the CFR framework, (Liu et al., 2023, Eq. (5.2)) employ dual-stabilized OMD as the local regret minimizer and establish asymptotic last-iterate convergence together with the optimal $O(1/T)$ average-iterate convergence to the unique NE of the perturbed regularized EFG. More recently, (Meng et al., 2025a, Eq. (3)) apply regret matching$^+$ (Tammelin et al., 2015) at each local decision point to generate an action strategy (i.e., a probability vector), followed by an affine transformation that maps the resulting strategy into the feasible set induced by the perturbation. For the resulting RTCFR$^+$ algorithm, (Meng et al., 2025a, Thms. 4.1 & F.1) establish asymptotic last-iterate convergence and a sublinear $O(1/\sqrt{T})$ average-iterate convergence rate to the NE of the perturbed regularized EFG. Our work follows the line of Meng et al. (2025a) and develops invariant last-iterate convergence guarantees, together with average-iterate convergence results, for predictive and discounted variants of CFR under the reward-transformation (RT) framework.

Last-Iterate Convergence in EFGs. We first describe the last-iterate convergence results in solving normal-form games (NFGs) / matrix games, which is a strict subset of EFGs. It is first discovered in (two-player) zero-sum matrix games that the vanilla online gradient descent (OGD) and multiplicative weight update (MWU) fail to converge last iterate (Mertikopoulos et al., 2018b, Thm 4.3) with continuous-time feedback, even using shrinking step sizes (Bailey & Piliouras, 2018, Thm 3.5). In contrast, their optimistic versions, optimistic gradient descent ascent (OGDA) and optimist multiplicative weight update (OMWU) Daskalakis & Panageas (2019); Mertikopoulos et al. (2018a); Lei et al. (2021), are proved to enjoy asymptotic last-iterate convergence. (Wei et al., 2021, Thm 3) further reveals that OMWU achieves linear last-iterate convergence in bilinear games over the simplex with the same unique NE assumption as in Daskalakis & Panageas (2019), and OGDA yields such linear convergence even without the uniqueness assumption (Wei et al., 2021, Thm 8). (Hsieh et al., 2021, Thm 7) shows the asymptotic last-iterate convergence of dual-stabilized optimistic OMD (DS-OOMD) with adaptive step size. (Cai et al., 2025, Thm 1) proves the last-iterate convergence of the famous regret matching$^+$ (RM$^+$) algorithm under the unique strict NE assumption, and their experimental counterexample shows the non-convergence of RM$^+$ under the unique non-strict NE condition. Besides, Cai et al. (2025); Meng et al. (2025b) further prove the last-iterate convergence results for the smooth versions of RM$^+$, such as smooth predictive RM$^+$ in (Cai et al., 2025, Thm 5) and (Meng et al., 2025b, Thm 5.3). Extending the theoretical results (or algorithms) from NFG to EFG while maintaining efficient implementation is nontrivial, because directly doing so comes at the cost of an exponential blowup of the strategy space. (Lee et al., 2021, Thm 4) for the first time demonstrates that optimistic OMD (OOMD) that runs over the sequence-form strategy space Von Stengel (1996) of EFG achieve asymptotic last-iterate convergence. With additional unique NE assumption, vanilla OMWU and OGDA (both of which are special cases of OOMD by setting the Bregman divergence as KL-divergence and Euclidean distance, respectively) yield sublinear and linear last-convergence rates (Lee et al., 2021, Thm 6), respectively. Besides, running OOMD with a dilated entropy function is much more efficient (i.e having a closed-form solution at each decision point) and still maintains a linear last-iterate convergence rate, e.g. see (Lee et al., 2021, Thm 7) and (Farina et al., 2022, Thm 5.5).

Counterfactual Regret Minimization (CFR) Algorithms in EFGs . In extensive-form game, the core idea of CFR method is to decompose the (global) regret of each player as the summation of (local) counterfactual regrets at all information sets of the game tree Zinkevich et al. (2007); Farina et al. (2019b). The next step is to utilize a no-regret online algorithm Hazan et al. (2016); Shalev-Shwartz et al. (2012) to minimize the counterfactual regret at each information set. The first CFR algorithm Zinkevich et al. (2007) uses regret matching (RM) Hart & Mas-Colell (2000) as the

local regret minimizer, which is a parameter-free algorithm that outputs a distribution over actions in an infoset in proportion to the positive cumulative counterfactual regret on those actions. Subsequent works along this line mainly focus on developing new RM-type regret minimizer, such as RM$^+$ Tammelin et al. (2015), discounted RM Brown & Sandholm (2019), predictive RM$^+$ Farina et al. (2021), smooth predictive RM$^+$ (Farina et al., 2023, Alg 2), conceptual RM$^+$ (Farina et al., 2023, Alg 3), predictive and discounted RM$^+$ Xu et al. (2024), monotone increasing smooth predictive RM$^+$ (Meng et al., 2026, Alg 1), leading to CFR$^+$, DCFR, PCFR$^+$, smooth PCFR$^+$, conceptual smooth CFR$^+$, PDCFR$^+$ and MI-smooth PCFR$^+$ algorithms, respectively. Most of these RM-based CFR algorithms enjoys an average-iterate convergence rate $O(1/\sqrt{T})$ in two-player zero-sum EFGs, except that smooth PCFR$^+$ (Farina et al., 2023, Thm 4.2) and conceptual smooth CFR$^+$ (Farina et al., 2023, Thm 5.5) achieving fast average-iterate convergence rate $O(1/T)$ (although conceptual smooth CFR$^+$, also called clairvoyant CFR$^+$, requires additional $O(T \log T)$ gradient computations to achieve this rate). Recent works (Zhang et al., 2026; Meng et al., 2026) also develop scale-invariant regret matching$^+$ algorithms with fast average-iterate convergence guarantee $O(1/T)$ in normal-form games. Besides RM-type CFR algorithms, there also exist other CFR variants that take optimistic OMD as local regret minimizer, such as Farina et al. (2019a) choosing OMWU and (Liu et al., 2023, Eq (5.2)) selecting dual-stabilized OOMD as local regret minimizers, and both of them achieve $O(1/T^{3/4})$ average-iterate convergence rate. We list the average-iterate convergence rates of different CFR variants in Table 1, for the reader's reference.

Other Average-Iterate Convergence Results in EFGs. In contrast to CFR algorithms that use no-regret learning algorithm to minimize local counterfactual regrets in decision points and to output behavioral strategy, another group directly runs online algorithm over the whole treeplex to output sequence-form strategy and to minimize player's individual regret. Farina et al. (2019c) directly runs optimistic OMD over treeplex and obtains $O(1/T)$ average-iterate convergence. Farina et al. (2022) develop a kernelized OMWU algorithm that can be implemented efficiently with a per-iteration complexity linear in the number of sequences (i.e. information-action pairs) of the player, and can achieve $O(\log^4(T)/T)$ per-player regret in multiplayer general-sum game. (Bai et al., 2022, Thm 8) proves that running kernelized OMWU is equivalent to running OMD with dilated KL-divergence over the whole treeplex. (Fan et al., 2024, Thm 5.4) proposes new treeplex primal-dual norms and originally designs an optimistic OMD algorithm that is implemented (in theory but not in practice) over the terminal observation points in treeplex with per-player regret $O(1/\sqrt{T})$.

B Pseudocode for Optimistic CFR$^+$ in Solving Perturbed Regularized EFG

Algorithm 1: Reward-Transformation based PCFR$^+$/PDCFR$^+$ (i.e. RTPCFR$^+$/RTPDCFR$^+$) in solving Perturbed Regularized EFGs

Input: T_1, μ_1, γ_1, $\boldsymbol{r}_1$, α (α is only for PDCFR$^+$), descreasing factors $\kappa = 0.3$ & $\tau = 0.5$
Output: $\hat{\boldsymbol{x}}^{T_K+1}$

$\hat{\boldsymbol{\theta}}_I^1 \leftarrow \mathbf{0}, \eta \leftarrow 1, \forall I \in \mathcal{I}$;
for $k = 1$ **to** K **do**
  for $i \in \mathcal{N} = \{0, 1\}$ **do**
    Build the perturbed regularized EFGs in Eq. (3) via μ_k, γ_k, and $\boldsymbol{r}_k$;
    for $t = 1$ **to** T_k **do**
      Option I (PCFR$^+$): Update $\boldsymbol{\theta}_I^{t+1} = [\hat{\boldsymbol{\theta}}_I^t + \hat{\boldsymbol{m}}_i^t(I)]^+, \quad \hat{\boldsymbol{\theta}}_I^{t+1} = [\hat{\boldsymbol{\theta}}_I^t + \hat{\boldsymbol{m}}_i^{t+1}(I)]^+$;
      Option II (PDCFR$^+$): Update $\boldsymbol{\theta}_I^{t+1} = [\frac{t^\alpha}{1+t^\alpha}\hat{\boldsymbol{\theta}}_I^t + \hat{\boldsymbol{m}}_i^t(I)]^+, \hat{\boldsymbol{\theta}}_I^{t+1} = [\frac{t^\alpha}{1+t^\alpha}\hat{\boldsymbol{\theta}}_I^t + \hat{\boldsymbol{m}}_i^{t+1}(I)]^+$;
      Output probability vector: $\sigma_i^{t+1}(I) = \boldsymbol{\theta}_I^{t+1} / \langle \boldsymbol{\theta}_I^{t+1}, \mathbf{1} \rangle$;
      Output perturbed behavioral strategy: $\hat{\sigma}_i^{t+1}(I) = (1 - \alpha_I)\sigma_i^{t+1}(I) + \gamma \mathbf{1}$;
      Update perturbed sequence-form strategy: $\hat{\boldsymbol{x}}_i^{t+1}(I) = \pi_i^{\hat{\sigma}^{t+1}}(I)\hat{\sigma}_i^{t+1}(I), \forall I \in \mathcal{I}_i$;
    end
    Update initialized cumulative counterfactual regrets: $\hat{\boldsymbol{\theta}}_I^1 \leftarrow \hat{\boldsymbol{\theta}}_I^{T_k+1}, \forall I \in \mathcal{I}_i$;
  end
  Update hyperparameters: $\mu_{k+1} = \mu_k * \kappa, \gamma_{k+1} = \gamma_k * \tau, \boldsymbol{r}_{k+1} = \hat{\boldsymbol{x}}^{T_k+1}$ ▷ Decreasing Perturbation/Regularization Coefficients;
end

Table 1: Comparisons of per-player individual regret (i.e. average-iterate convergence rate) of different CFR variants for two-player zero-sum extensive-form games.

Algorithm	**Local Regret Minimizer**	**Per-Player Individual Regret**
CFR (Zinkevich et al., 2007, NeurIPS)	RM	$O(1/\sqrt{T})$
CFR$^+$ (Tammelin et al., 2015, IJCAI)	RM+	$O(1/\sqrt{T})$
DCFR (Brown & Sandholm, 2019, AAAI)	Discounted RM$^+$	$O(1/\sqrt{T})$
PCFR$^+$ (Farina et al., 2021, AAAI)	Predictive RM$^+$	$O\left(\sum_{I\in\mathcal{I}_i}\sqrt{\sum_{t=1}^{T}\|\hat{\boldsymbol{m}}_i^t(I)-\hat{\boldsymbol{m}}_i^{t-1}(I)\|_2^2}/T\right)$
Opt FTRL (Farina et al., 2019a, ICML)	Stable-Predictive OptMWU	$O(1/T^{3/4})$
Reg-CFR (Liu et al., 2023, ICLR)	DS-OptOMD	$O(1/T^{3/4})$
Conceptual Smooth CFR+ (Farina et al., 2023, NeurIPS)	Conceptual RM$^+$	$O(1/T)$ Cost: $O(T\log T)$ gradient computations
PDCFR+ (Xu et al., 2024, IJCAI)	Predictive Discounted RM+	$O\left(\sum_{I\in\mathcal{I}_i}\sqrt{\sum_{t=1}^{T}\|\hat{\boldsymbol{m}}_i^t(I)-\hat{\boldsymbol{m}}_i^{t-1}(I)\|_2^2}/T\right)$
Invariant-PCFR+ our Theorem 4.2	Predictive RM$^+$	$O\left(\sum_{I\in\mathcal{I}_i}\sqrt{\frac{\|\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta^2}+\sum_{t=1}^{T}\|\hat{\boldsymbol{m}}_i^t(I)-\hat{\boldsymbol{m}}_i^{t-1}(I)\|_2^2}/T\right)$

C Technical Novelties of Our Work and Detailed Comparisons with the most Related Works

Remark C.1 (Five Technical Novelties in our Convergence Analysis for Optimistic CFR in Solving Perturbed EFG (e.g. including Conceptual Distinctions between our Results and the most Related Work Meng et al. (2025a))). *We explain five main technical novelties in our convergence analysis.*

1. ***For the asymptotic last-iterate convergence for PCFR$^+$/PDCFR$^+$:*** *instead of directly applying the first-order optimality condition to the first OMD update step in Eq. (5) as previous work (Meng et al., 2025a, Sect D), we choose to apply the first-order optimality condition to the second OMD update step in Eq. (5), and then use first-order optimality condition of the first OMD update step in Eq. (5) to bound the extra term $-\eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\rangle$, finally leading to a telescoping sum.*
2. ***For the invariant last-iterate convergence of PDCFR$^+$:*** *for any non-increasing step sizes $\{\eta^t\}_{t\geq 0}$, we construct a novel sequence of step sizes $\eta^{t+1\prime} = \frac{\eta^{t'}}{\eta^t}\eta^{t+1}$ with $\eta^{0'} \leq \frac{\mu}{3|\mathcal{I}|\kappa^2}$, and prove the non-increasing property as well as the existence of the limit point of $\{\eta^{t+1\prime}\}_{t\geq 0}$ by applying monotone convergence theorem (see Lemma G.1). The above construction is critical to prove $\hat{\boldsymbol{\theta}}_I^{t'} = \frac{\eta^{t'}}{\eta^t}\hat{\boldsymbol{\theta}}_I^t$ by mathematical induction in Lemma G.2, finally resulting in the invariant last-iterate convergence for PDCFR$^+$.*
3. ***For the improved average-iterate regret bound for PCFR$^+$/PDCFR$^+$:*** *We first use Eq. (2) in Lemma D.6 to connect the individual-player regret and the summation of counterfactual regrets over all inforsets, hence avoiding the complicated Blackwell approachability analysis as in (Farina et al., 2021, Prop 2). Then, we leverage the invariance property of the last-iterate convergence (w.r.t. any step size and any strategy initialization) of PCFR$^+$/PDCFR$^+$ in Claims E.1 & G.1 to obtain improved dependence on the initialization strategy.*
4. ***For the pointwise convergence to the original NE:*** *By employing previous result in Feinstein (2022), we first show that the limit set of the sequence of NE of perturbed regularized EFG is actually a subset of the set of NEs of original game. Thus, the pointwise convergence of discrete perturbed regularized EFG to the NE of original game actually requires strong assumptions. Such analysis is absent from RTCFR$^+$ (Meng et al., 2025a), but point out that we need to use other convergence notion (like duality gap) instead of the pointwise convergence that is used in the last-iterate convergence in perturbed EFG.*
5. ***For the convergence to the approximate NE of original EFG:*** *We use the explicit form of the behavioral strategy in Eq. (6) of RTPCFR$^+$/RTPDCFR$^+$ in Algorithm 1 to obtain an improved dependence on P (i.e. the number of all information sets in 2-player zero-sum EFG) in Proposition 5.4 in duality gap, which finally results in an upper bound $O(\|\hat{\boldsymbol{x}}^{T_k+1} - \hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}\|_2 + \min\{\mu_k, \gamma_k\})$ on the duality gap, characterizing how the perturbation/regularization coefficients affect the last-iterate convergence of our Algorithm 1. Such guarantee is absent from (Meng et al., 2025a).* □

Remark C.2 (Technical Differences between the Last-Iterate Convergence for Optimistic Methods in (Liu et al., 2023) and Ours). *Given that Liu et al. (2023) consists of two parts: (1) the explicit last-iterate convergence of global optimistic mirror descent in (Liu et al., 2023, Thms 4.1-4.2) to the NE of original EFG, and (2) the asymptotic last-iterate convergence of adaptive mirror descent-based CFR variant in (Liu et al., 2023, Thms 5.3) to the NE of perturbed EFG, the explicit comparison between our work and Liu et al. (2023) is detailed in two aspects:*

1. *When compared with the global optimistic mirror descent in (Liu et al., 2023, Sect 4): It is hard to directly apply the optimistic operator-splitting arguments of (Liu et al., 2023) to the CFR decomposition to obtain the last-iterate convergence of RTPCFR$^+$ algorithm, because the global optimistic mirror descent updates the treeplex sequence-form strategy in each iteration. However, in PCFR$^+$ (and its variants) the regret minimizer in each information set operate independently and cannot synchronize with other regret minimizers. We then choose to decompose the global counterfactual regret into individual ones, use first-order optimality condition twice of predictive regret matching+ on each information set to bound the individual regret, and leverage the smoothness of instantaneous regret (which is guaranteed by the introduction of perturbation technique that was not used in global mirror descent in (Liu et al., 2023)), finally leading to*

our Eq. (32) $\frac{\eta\mu}{3}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \le O(1)$*, which implies the asymptotic convergence of regret-matching-based optimistic CFR variants.*

2. *When compared with the adaptive OMD-based CFR variant Reg-CFR in (Liu et al., 2023, Sect 5), our conceptual contributions are two-fold:*

   - *Reg-CFR chooses mirror descent (with adaptive step size) as local regret minimizer and solve a constrained minimization problem on perturbed strategy space (see (Liu et al., 2023, Eq.(5.2))). Instead, our PCFR+ uses predictive regret matching+ as local regret minimizer and chooses a more efficient update rule*

   $$\hat{\sigma}_i^{t+1}(I) = (1-\alpha_I)\sigma_i^{t+1}(I) + \gamma\mathbf{1},$$

   which finally leads to a sharper bound $O(\gamma P)$ on duality gap in Proposition 6.4 than that of $O(\gamma P^2)$ in (Liu et al., 2023, Lemma F.4).
   - *When obtaining the last-iterate convergence for Reg-CFR, (Liu et al., 2023) uses the smoothness of instantaneous regret $\hat{m}_i^t$ and the property of adaptive step size, while our work uses the smoothness of instantaneous regret $\hat{m}_i^{*,\gamma}$ and show the invariant last-iterate convergence w.r.t. any step size.*

D Auxiliary Results

Lemma D.1 (Lemma 4 in Farina et al. (2021))**.** *Let $\mathcal{D}\subseteq\mathbb{R}^d$ be a convex and closed set. Consider the update rule of online mirror descent (OMD) algorithm with step size $\eta>0$ and convex distance generating function ψ: $\boldsymbol{\theta}^*\in\arg\min_{\boldsymbol{\theta}\in\mathcal{D}}\left\{\langle\boldsymbol{g},\boldsymbol{\theta}\rangle+\frac{1}{\eta}D_\psi(\boldsymbol{\theta}||\boldsymbol{\theta}^0)\right\}$, we have the following result for any $\boldsymbol{\theta}\in\mathcal{D}$:*

$$\eta\langle\boldsymbol{g},\boldsymbol{\theta}^*-\boldsymbol{\theta}\rangle \le D_\psi(\boldsymbol{\theta}||\boldsymbol{\theta}^0) - D_\psi(\boldsymbol{\theta}||\boldsymbol{\theta}^*) - D_\psi(\boldsymbol{\theta}^*||\boldsymbol{\theta}^0) \tag{8}$$

Lemma D.2 (Adapted from Lemma D.4 in Sokota et al. (2023))**.** *For any $\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$, $\mu\ge 0$, and $\gamma\ge 0$,*

$$\sum_{i\in\mathcal{N}}\langle\boldsymbol{\ell}_i^{\boldsymbol{x}},\boldsymbol{x}_i-\boldsymbol{x}_i^{*,\mu,\gamma,\boldsymbol{r}}\rangle \ge \sum_{i\in\mathcal{N}}\langle\boldsymbol{\ell}_i^{\boldsymbol{x}}-\boldsymbol{\ell}_i^{\boldsymbol{x}^{*,\mu,\gamma,\boldsymbol{r}}},\boldsymbol{x}_i-\boldsymbol{x}_i^{*,\mu,\gamma,\boldsymbol{r}}\rangle \ge \mu\|\boldsymbol{x}-\boldsymbol{x}^{*,\mu,\gamma,\boldsymbol{r}}\|_2^2, \tag{9}$$

where $\boldsymbol{\ell}_0^{\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}_1+\mu\nabla\psi(\boldsymbol{x}_0)-\mu\nabla\psi(\boldsymbol{r}_0)$ and $\boldsymbol{\ell}_1^{\boldsymbol{x}}=-\boldsymbol{A}^T\boldsymbol{x}_0+\mu\nabla\psi(\boldsymbol{x}_1)-\mu\nabla\psi(\boldsymbol{r}_1)$.

Lemma D.3 (Lemmas D.2-D.3 in Meng et al. (2025a))**.** *For any $\boldsymbol{x},\boldsymbol{x}'\in\boldsymbol{\mathcal{X}}$, $i\in\mathcal{N}$, $I\in\mathcal{I}_i$, $\mu\ge 0$, and $\gamma\ge 0$,*

$$\|\hat{\boldsymbol{v}}_i^{\sigma}(I)\|_2 \le \|\hat{\boldsymbol{v}}_i^{\sigma}(I)\|_1 \le P+2\mu D \tag{10}$$

$$\|\hat{\boldsymbol{v}}_i^{\sigma}(I)-\hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2 \le (L+\mu)\|\boldsymbol{x}-\boldsymbol{x}'\|_1 + (P+2\mu D)\|\sigma_i-\sigma_i'\|_1, \tag{11}$$

where $\hat{\boldsymbol{v}}_i^{\sigma}(I)=[\hat{\boldsymbol{v}}_i^{\sigma}(I,a)|a\in A(I)]$, $\hat{\boldsymbol{v}}_i^{\sigma}(I,a)=-\hat{\boldsymbol{\ell}}_i^{\boldsymbol{x}}+\sum_{I'\in C_i(I,a)}\langle\hat{\boldsymbol{v}}_i^{\sigma}(I'),\sigma_i(I')\rangle$ with $\hat{\boldsymbol{\ell}}_0^{\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}_1+\mu\nabla\psi(\boldsymbol{x}_0)-\mu\nabla\psi(\boldsymbol{r}_0)$ and $\hat{\boldsymbol{\ell}}_1^{\boldsymbol{x}}=-\boldsymbol{A}^T\boldsymbol{x}_0+\mu\nabla\psi(\boldsymbol{x}_1)-\mu\nabla\psi(\boldsymbol{r}_1)$, as well as σ and σ' are the behavioral strategy profiles associated with $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively.

Lemma D.4 (Lemma D.4 in Meng et al. (2025a))**.** *For any $\hat{\boldsymbol{x}},\hat{\boldsymbol{x}}'\in\boldsymbol{\mathcal{X}}^\gamma$ with $\gamma>0$, $i\in\mathcal{N}$, $I\in\mathcal{I}_i$, and $\mu\ge 0$,*

$$\|\hat{\sigma}_i-\hat{\sigma}_i'\|_1 \le \frac{A_{max}C_{max}+1}{\gamma^H}\|\hat{\boldsymbol{x}}_i-\hat{\boldsymbol{x}}_i'\|_1, \tag{12}$$

where $\hat{\sigma}$ and $\hat{\sigma}'$ are the behavioral strategy profiles associated with $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$, respectively.

Lemma D.5 (Lemma 4.4 in Meng et al. (2025a))**.** *$\forall i\in\mathcal{N}$, $I\in\mathcal{I}_i$, and $\boldsymbol{\theta}_I\in\mathbb{R}_{\ge 0}^{|A(I)|}$, $\langle-\hat{\boldsymbol{m}}_i^{*,\mu,\gamma,\boldsymbol{r}}(I),\boldsymbol{\theta}_I\rangle\ge 0$.*

The following lemma establishes the equivalence relation between individual regret $\langle\boldsymbol{\ell}_i,\boldsymbol{x}_i-\boldsymbol{x}_i'\rangle$ and weighted summation of counterfactual regrets $\{\boldsymbol{m}_i^{\sigma}(I)\}_{I\in\mathcal{I}_i}$. Such result is very essential and we give the whole proof details for the completeness of this work.

Lemma D.6 (Lemma 2.2 in Liu et al. (2022)). *For any $\boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{\mathcal{X}}$, $\boldsymbol{\ell} \in \mathbb{R}^{|\boldsymbol{\mathcal{X}}|}$, $i \in \mathcal{N}$, $\mu \geq 0$, and $\gamma \geq 0$, define*

$$\begin{aligned}
&\boldsymbol{m}_i^\sigma(I) = \boldsymbol{v}_i^\sigma(I) - \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \mathbf{1}, \\
&\boldsymbol{v}_i^\sigma(I,a) = -\boldsymbol{\ell}_i^\sigma(I,a) + \sum_{I' \in C_i(I,a)} \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I')\rangle, \\
&\boldsymbol{\ell}_0^\sigma = \boldsymbol{A}\boldsymbol{x}_1^\sigma + \mu\nabla\psi(\boldsymbol{x}_0^\sigma) - \mu\nabla\psi(\boldsymbol{r}_0), \quad \boldsymbol{\ell}_1^\sigma = -\boldsymbol{A}^\top \boldsymbol{x}_0^\sigma + \mu\nabla\psi(\boldsymbol{x}_1^\sigma) - \mu\nabla\psi(\boldsymbol{r}_1).
\end{aligned} \tag{13}$$

Then, we have the following result for any player $i \in \mathcal{N}$:

$$\langle \boldsymbol{\ell}_i, \boldsymbol{x}_i - \boldsymbol{x}_i' \rangle = \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{m}_i^\sigma(I), \sigma_i'(I)\rangle \tag{14}$$

Proof. Plug the definition of instantaneous counterfactual regret $\boldsymbol{m}_i^\sigma(I)$ into the right-hand side of the above equality,

$$\begin{aligned}
&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{m}_i^\sigma(I), \sigma_i'(I)\rangle \\
=&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I) - \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \mathbf{1}, \sigma_i'(I)\rangle \\
=&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i'(I) - \sigma_i(I)\rangle \\
\overset{(i)}{=}&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \Big\langle -\boldsymbol{\ell}_i^\sigma(I) + \Big(\sum_{I' \in C_i(I,a)} \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I')\rangle \Big)_{a \in A(I)}, \sigma_i'(I) \Big\rangle - \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \\
=&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle + \sum_{I \in \mathcal{I}_i} \sum_{a \in A(I)} \sum_{I' \in C_i(I,a)} \pi_i^{\sigma'}(I) \sigma_i'(I,a) \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I')\rangle - \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \\
\overset{(ii)}{=}&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle + \sum_{I \in \mathcal{I}_i} \sum_{a \in A(I)} \sum_{I' \in C_i(I,a)} \pi_i^{\sigma'}(I') \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I')\rangle - \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \\
\overset{(iii)}{=}&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle + \sum_{I \in \mathcal{I}_i \setminus \{r\}} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle - \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \\
=&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle - \sum_{I \in \{r\}} \pi_i^{\sigma'}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \\
\overset{(iv)}{=}&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle - \sum_{I \in \{r\}} \pi_i^{\sigma}(I) \langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle \\
\overset{(v)}{=}&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle - \sum_{I \in \{r\}} \Big[\pi_i^\sigma(I) \langle -\boldsymbol{\ell}_i^\sigma(I) + \Big(\sum_{I' \in C_i(I,a)} \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I')\rangle \Big)_{a \in A(I)}, \sigma_i(I)\rangle \Big] \\
=&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle - \sum_{I \in \{r\}} \Big[\pi_i^\sigma(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i(I)\rangle + \sum_{a \in A(I)} \sum_{I' \in C_i(I,a)} \pi_i^\sigma(I) \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I')\rangle \sigma_i(I,a) \Big] \\
\overset{(vi)}{=}&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle - \sum_{I \in \{r\}} \Big[\pi_i^\sigma(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i(I)\rangle + \sum_{a \in A(I)} \sum_{I' \in C_i(I,a)} \pi_i^\sigma(I') \langle \boldsymbol{v}_i^\sigma(I'), \sigma_i(I')\rangle \Big] \\
\overset{(vii)}{=}&\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i'(I)\rangle - \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma}(I) \langle -\boldsymbol{\ell}_i^\sigma(I), \sigma_i(I)\rangle \\
\overset{(viii)}{=}&\sum_{I \in \mathcal{I}_i} \langle -\boldsymbol{\ell}_i^\sigma(I), \boldsymbol{x}_i^{\sigma'}(I)\rangle - \sum_{I \in \mathcal{I}_i} \langle -\boldsymbol{\ell}_i^\sigma(I), \boldsymbol{x}_i^{\sigma}(I)\rangle \\
=&\langle \boldsymbol{\ell}_i^\sigma, \boldsymbol{x}_i^\sigma - \boldsymbol{x}_i^{\sigma\prime} \rangle
\end{aligned} \tag{15}$$

where the equality (i) and (v) hold due to the definition of counterfactual value vector $\boldsymbol{v}_i^\sigma(I)$, (ii) and (vi) due to the fact that $\pi_i^\sigma(I)\sigma_i(I,a) = \pi_i^\sigma(I')$ for any $I' \in C_i(I,a)$ and any σ, (iii) due

to Lemma D.7 by setting $f(\sigma_i(I), \pi_i^\sigma(I)) = \pi_i^\sigma(I)\langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle$; (iv) holds since in the root decision node $r \in \mathcal{I}_i$ we have $\pi_i^\sigma(r) = \pi_i^{\sigma'}(r) = 1$, (vii) due to the recursion of expression $\pi_i^\sigma(I)\langle \boldsymbol{v}_i^\sigma(I), \sigma_i(I)\rangle$, $(viii)$ due to the fact that the equality $\pi_i^\sigma(I)\sigma_i(I) = \boldsymbol{x}_i^\sigma(I)$ holds for any σ and information set I. □

Applying Eq. (15) in the proof for Lemma D.6 obtains the following lemma.

Corollary D.1 (Lemma 4.3 in Meng et al. (2025a)). *For any $\boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{\mathcal{X}}$, $\boldsymbol{\ell} \in \mathbb{R}^{|\mathcal{X}|}$, $i \in \mathcal{N}$, $\mu \geq 0$, and $\gamma \geq 0$,*

$$\langle \boldsymbol{\ell}_i, \boldsymbol{x}_i - \boldsymbol{x}_i' \rangle = \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma'}(I)\langle -\boldsymbol{v}_i^\sigma(I, a), \sigma_i(I) - \sigma_i'(I)\rangle. \tag{16}$$

Lemma D.7. *For any behavioral strategy σ, any player $i \in \mathcal{N}$, any information set $I \in \mathcal{I}_i$, define any measurable function f over the product space between the simplex space $\Delta^{|A(I)|}$ and unit internal $[0, 1]$ as $f : \Delta^{|A(I)|} \times [0, 1] \to \mathbb{R}$ as $f(\sigma_i(I), \pi_i^\sigma(I))$. Then we have*

$$\sum_{I \in \mathcal{I}_i} \sum_{a \in A(I)} \sum_{I' \in C_i(I,a)} f(\sigma_i(I'), \pi_i^\sigma(I')) = \sum_{I \in \mathcal{I}_i \setminus \{r\}} f(\sigma_i(I), \pi_i^\sigma(I)) \tag{17}$$

Proof. For any information set $I' \in \mathcal{I}_i \setminus \{r\}$, there exists only one information set $I \in \mathcal{I}_i$ and only one action $a \in A(I)$, such that $I' \in C_i(I, a)$, due to the perfect recall property of the extensive-form game. Besides, for any information set I, and any action $a \in A(I)$, information set $I \notin C_i(I, a)$; and the root information set r does not belong to any $C_i(I, a)$. Therefore, for any information set $I'(I' \notin \{r\})$, the term $f(\sigma_i(I'), \pi_i^\sigma(I'))$ in the left-hand side of Eq. (17) appears only once, thus the equality in Eq. (17) holds. The proof is finished. □

E Proof for the Asymptotic Last-Iterate Convergence of PCFR+ in Solving Perturbed Regularized EFG

Lemma E.1. *Define the instantaneous regret associated with any behavioral strategy profile $\sigma = \{\sigma_1, \sigma_2\}$:*

$$\begin{aligned}
&\hat{\boldsymbol{m}}_i^\sigma(I) = \hat{\boldsymbol{v}}_i^\sigma(I) - \langle \hat{\boldsymbol{v}}_i^\sigma(I), \sigma_i(I)\rangle \mathbf{1}, \\
&\hat{\boldsymbol{v}}_i^\sigma(I, a) = -\hat{\boldsymbol{\ell}}_i^\sigma(I, a) + \sum_{I' \in C_i(I,a)} \langle \hat{\boldsymbol{v}}_i^\sigma(I'), \hat{\sigma}_i(I)\rangle, \\
&\hat{\boldsymbol{\ell}}_0^\sigma = \boldsymbol{A}\hat{\boldsymbol{x}}_1^\sigma + \mu\nabla\psi(\hat{\boldsymbol{x}}_0^\sigma) - \mu\nabla\psi(\boldsymbol{r}_0), \quad \hat{\boldsymbol{\ell}}_1^\sigma = -\boldsymbol{A}^\top\hat{\boldsymbol{x}}_0^\sigma + \mu\nabla\psi(\hat{\boldsymbol{x}}_1^\sigma) - \mu\nabla\psi(\boldsymbol{r}_1).
\end{aligned} \tag{18}$$

Let $\kappa = \Big[\big(A_{max} + 1\big)(L + \mu) + \big(2A_{max} + 1\big)\frac{(P+2\mu D)}{1-\gamma A_{max}}\frac{A_{max}C_{max}+1}{\gamma^H}\Big]\sqrt{|\mathcal{I}|A_{max}}$ Then for any strategy profile σ, σ', any player $i \in \mathcal{N}$, any information set $I \in \mathcal{I}_i$, we have:

$$\|\hat{\boldsymbol{m}}_i^\sigma(I) - \hat{\boldsymbol{m}}_i^{\sigma'}(I)\|_2 \leq \kappa\|\hat{\boldsymbol{x}}^\sigma - \hat{\boldsymbol{x}}^{\sigma'}\|_2 \tag{19}$$

Proof.

$$\begin{aligned}
&\|\hat{\boldsymbol{m}}_i^\sigma(I) - \hat{\boldsymbol{m}}_i^{\sigma'}(I)\|_2 = \|\hat{\boldsymbol{v}}_i^\sigma(I) - \langle \hat{\boldsymbol{v}}_i^\sigma(I), \sigma_i(I)\rangle\mathbf{1} - \hat{\boldsymbol{v}}_i^{\sigma'}(I) + \langle \hat{\boldsymbol{v}}_i^{\sigma'}(I), \sigma_i'(I)\rangle\mathbf{1}\|_2 \\
&\overset{(i)}{\leq} \|\hat{\boldsymbol{v}}_i^\sigma(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2 + \big|\langle \hat{\boldsymbol{v}}_i^\sigma(I), \sigma_i(I)\rangle - \langle \hat{\boldsymbol{v}}_i^{\sigma'}(I), \sigma_i'(I)\rangle\big|\|\mathbf{1}\|_2 \\
&\overset{(ii)}{\leq} \|\hat{\boldsymbol{v}}_i^\sigma(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2 + \big|\langle \hat{\boldsymbol{v}}_i^\sigma(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I), \sigma_i(I)\rangle + \langle \hat{\boldsymbol{v}}_i^{\sigma'}(I), \sigma_i(I) - \sigma_i'(I)\rangle\big||A(I)| \\
&\overset{(iii)}{\leq} \|\hat{\boldsymbol{v}}_i^\sigma(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2 + |A(I)|\Big(\|\hat{\boldsymbol{v}}_i^\sigma(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2\|\sigma_i(I)\|_2 + \|\hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2\|\sigma_i(I) - \sigma_i'(I)\|_2\Big) \\
&\overset{(iv)}{\leq} \|\hat{\boldsymbol{v}}_i^\sigma(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2 + |A(I)|\Big(\|\hat{\boldsymbol{v}}_i^\sigma(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2 + (P + 2\mu D)\|\sigma_i(I) - \sigma_i'(I)\|_2\Big)
\end{aligned} \tag{20}$$

$$
\begin{aligned}
&= \big(|A(I)|+1\big)\|\hat{\boldsymbol{v}}_i^{\sigma}(I) - \hat{\boldsymbol{v}}_i^{\sigma'}(I)\|_2 + |A(I)|(P+2\mu D)\|\sigma_i(I) - \sigma_i'(I)\|_2 \\
&\overset{(v)}{\leq} \big(|A(I)|+1\big)\Big[(L+\mu)\|\hat{\boldsymbol{x}}^{\sigma} - \hat{\boldsymbol{x}}^{\sigma'}\|_1 + (P+2\mu D)\|\sigma_i - \sigma_i'\|_1\Big] + |A(I)|(P+2\mu D)\|\sigma_i(I) - \sigma_i'(I)\|_2 \\
&= \big(|A(I)|+1\big)\Big[(L+\mu)\|\hat{\boldsymbol{x}}^{\sigma} - \hat{\boldsymbol{x}}^{\sigma'}\|_1 + \frac{(P+2\mu D)}{1-\alpha_I}\|\hat{\sigma}_i - \hat{\sigma}_i'\|_1\Big] + \frac{|A(I)|(P+2\mu D)}{1-\alpha_I}\|\hat{\sigma}_i(I) - \hat{\sigma}_i'(I)\|_2 \\
&\leq \big(|A(I)|+1\big)\Big[(L+\mu)\|\hat{\boldsymbol{x}}^{\sigma} - \hat{\boldsymbol{x}}^{\sigma'}\|_1 + \frac{(P+2\mu D)}{1-\alpha_I}\|\hat{\sigma}_i - \hat{\sigma}_i'\|_1\Big] + \frac{|A(I)|(P+2\mu D)}{1-\alpha_I}\|\hat{\sigma}_i - \hat{\sigma}_i'\|_1 \\
&= \big(|A(I)|+1\big)(L+\mu)\|\hat{\boldsymbol{x}}^{\sigma} - \hat{\boldsymbol{x}}^{\sigma'}\|_1 + \big(2|A(I)|+1\big)\frac{(P+2\mu D)}{1-\alpha_I}\|\hat{\sigma}_i - \hat{\sigma}_i'\|_1 \\
&\overset{(vi)}{\leq} \big(|A(I)|+1\big)(L+\mu)\|\hat{\boldsymbol{x}}^{\sigma} - \hat{\boldsymbol{x}}^{\sigma'}\|_1 + \big(2|A(I)|+1\big)\frac{(P+2\mu D)}{1-\alpha_I}\frac{A_{max}C_{max}+1}{\gamma^H}\|\hat{\boldsymbol{x}}_i^{\sigma} - \hat{\boldsymbol{x}}_i^{\sigma'}\|_1 \\
&\leq \Big[\big(|A(I)|+1\big)(L+\mu) + \big(2|A(I)|+1\big)\frac{(P+2\mu D)}{1-\alpha_I}\frac{A_{max}C_{max}+1}{\gamma^H}\Big]\|\hat{\boldsymbol{x}}^{\sigma} - \hat{\boldsymbol{x}}^{\sigma'}\|_1 \\
&\leq \kappa\|\hat{\boldsymbol{x}}^{\sigma} - \hat{\boldsymbol{x}}^{\sigma'}\|_2
\end{aligned}
$$

where (i) holds due to the triangle inequality, (ii) due to the fact $\|\mathbf{1}\|_2 \leq \|\mathbf{1}\|_1 \leq |A(I)|$, (iii) due to Cauchy-Schwartz inequality $\langle \boldsymbol{a}, \boldsymbol{b}\rangle \leq \|\boldsymbol{a}\|_2\|\boldsymbol{b}\|_2$, (iv) due to the boundedness of counterfactual value $\hat{\boldsymbol{v}}_i^{\sigma'}(I)$ in Eq. (10) and the boundedness of ℓ_1-norm of a simplex vector $\sigma_i(I)$, (v) due to Lemma D.3, (vi) due to Lemma D.4. The last inequality holds since for any d-dimensional vector $\boldsymbol{x}$, we have $\|\boldsymbol{x}\|_1 \leq \sqrt{d}\|\boldsymbol{x}\|_2$. □

Proof of Theorem 4.1. The update rule of optimistic OMD is as follow:

$$
\boldsymbol{\theta}_I^{t+1} \in \arg\min_{\boldsymbol{\theta}_I \in \mathbb{R}_{\geq 0}^{|A(I)|}} \left\{ \langle -\hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}_I\rangle + \frac{1}{\eta} D_\psi(\boldsymbol{\theta}_I || \hat{\boldsymbol{\theta}}_I^t) \right\}, \tag{21}
$$

$$
\hat{\boldsymbol{\theta}}_I^{t+1} \in \arg\min_{\boldsymbol{\theta}_I \in \mathbb{R}_{\geq 0}^{|A(I)|}} \left\{ \langle -\hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I\rangle + \frac{1}{\eta} D_\psi(\boldsymbol{\theta}_I || \hat{\boldsymbol{\theta}}_I^t) \right\}. \tag{22}
$$

We apply Lemma D.1 for the update rule of optimistic OMD in Eq. (22), obtaining that for any $\boldsymbol{\theta} \in \mathbb{R}_{\geq 0}^{|A(I)|}$:

$$
\eta\langle -\hat{\boldsymbol{m}}_i^{t+1}(I), \hat{\boldsymbol{\theta}}_I^{t+1} - \boldsymbol{\theta}\rangle \leq D_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\boldsymbol{\theta}||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t). \tag{23}
$$

Setting $\boldsymbol{\theta} = \sigma_i^{*,\gamma}(I)$ in the above inequality, we have

$$
\eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i^{*,\gamma}(I) - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle \leq D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t). \tag{24}
$$

Adding $\eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), -\boldsymbol{\theta}_I^{t+1}\rangle$ into both sides of Eq. (24), and rearranging the terms, we have

$$
\begin{aligned}
&\eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i^{*,\gamma}(I) - \boldsymbol{\theta}_I^{t+1}\rangle \\
\leq& D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle \\
=& \underbrace{D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^{t+1})}_{①} \underbrace{- D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle}_{②} \\
&\underbrace{- \eta\langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle}_{③} \underbrace{- \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\rangle}_{④}.
\end{aligned} \tag{25}
$$

It is not difficult to see that, when taking a summation of the right-hand-side of Eq. (25) over $t \in [T]$, both term ① and term ③ can lead to the telescoping sum. Therefore, it remains to bound terms ② and ④ respectively.

(I) Bounding ② in Eq. (25)

$$\begin{aligned}
&- D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle \\
\leq &- D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) + \eta\Big[\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2 \cdot \|\hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1}\|_2\Big] \\
\leq &- D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) + \eta\Big[\frac{\eta\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2}{2} + \frac{\|\hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1}\|_2^2}{2\eta}\Big] \\
= &\frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2,
\end{aligned} \tag{26}$$

where the first and the second inequalities in Eq. (26) are due to Cauchy-Schwartz and mean-value inequality $ab \leq \frac{\eta a^2}{2} + \frac{b^2}{2\eta}$ for any $\eta > 0$, respectively.

(II) Bounding ④ in Eq. (25)

$$\begin{aligned}
&- \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\rangle \\
= &- \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\rangle - \eta\langle \hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\rangle \\
\overset{(i)}{\leq} &\eta\Big[\frac{\eta}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 + \frac{1}{\eta}\|\boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\|_2^2\Big] + \Big[D_\psi(\hat{\boldsymbol{\theta}}_I^t||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\hat{\boldsymbol{\theta}}_I^t||\boldsymbol{\theta}_I^{t+1}) - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t)\Big] \\
= &\eta\Big[\frac{\eta}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 + \frac{1}{\eta}\|\boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\|_2^2\Big] + \Big[-\|\boldsymbol{\theta}^{t+1} - \hat{\boldsymbol{\theta}}_I^t\|_2^2\Big] \\
= &\frac{\eta^2}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2,
\end{aligned} \tag{27}$$

where in the inequality (i) of Eq. (27), the first part holds due to Cauchy-Schwartz/mean-value inequality (cf. Eq. (26)), and the second part due to the application of Lemma D.1 to the OMD update in Eq. (21).

Combining the results in Eq. (26) and Eq. (27), and plugging them into Eq. (25), we have

$$\begin{aligned}
&\eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i^{*,\gamma}(I) - \boldsymbol{\theta}_I^{t+1}\rangle - \frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 - \frac{\eta^2}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 \\
\leq &D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - \eta\langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle.
\end{aligned} \tag{28}$$

Notice that $\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I^{t+1}\rangle = \langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i^{t+1}\rangle\|\boldsymbol{\theta}_I^{t+1}\|_1 = \langle \hat{\boldsymbol{v}}_i^{t+1}(I) - \langle \hat{\boldsymbol{v}}_i^{t+1}(I), \sigma_i^{t+1}(I)\rangle\mathbf{1}, \sigma_i^{t+1}\rangle\|\boldsymbol{\theta}_I^{t+1}\|_1 = 0$, then the first term in the left-hand side of Eq. (28) can be simplified as

$$\begin{aligned}
\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i^{*,\gamma}(I) - \boldsymbol{\theta}_I^{t+1}\rangle &= \langle \hat{\boldsymbol{v}}_i^{t+1}(I) - \langle \hat{\boldsymbol{v}}_i^{t+1}(I), \sigma_i^{t+1}(I)\rangle\mathbf{1}, \sigma_i^{*,\gamma}(I)\rangle \\
&= \langle \hat{\boldsymbol{v}}_i^{t+1}(I), \sigma_i^{*,\gamma}(I) - \sigma_i^{t+1}(I)\rangle \\
&= \frac{1}{1-\alpha_I}\langle \hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{*,\gamma}(I) - \hat{\sigma}_i^{t+1}(I)\rangle
\end{aligned} \tag{29}$$

Then, let $\beta_I = \pi_i^{\hat{\sigma}^{*,\gamma,\mu}}(I)$, $\zeta_I = \beta_I(1-\alpha_I)$ (obviously $\zeta_I < 1$), multiplying both sides in Eq. (28) by ζ_I, we have

$$\begin{aligned}
&\eta\beta_I\langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I) - \hat{\sigma}_i^{*,\gamma}(I)\rangle - \frac{\eta^2\zeta_I}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 - \frac{\eta^2\zeta_I}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 \\
\leq &\zeta_I\Big[D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - \eta\langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle\Big].
\end{aligned} \tag{30}$$

Taking summation of both sides in Eq. (30) over $t = 0, 1, \cdots, T-1$, $i \in \mathcal{N}$, $I \in \mathcal{I}_i$, we have

$$\begin{aligned}
&\sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i}\zeta_I\Big[D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^0) - \eta\langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^0\rangle\Big] \\
\overset{(i)}{\geq} &\sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i}\zeta_I\Big[D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^0) - D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^T) - \eta\langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^0 - \hat{\boldsymbol{\theta}}_I^T\rangle\Big]
\end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{t=0}^{T-1}\sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i}\Big[\eta\beta_I\langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I)-\hat{\sigma}_i^{*,\gamma}(I)\rangle - \frac{\eta^2\zeta_I}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 - \frac{\eta^2\zeta_I}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{t}(I)\|_2^2\Big]\\
&\geq \sum_{t=0}^{T-1}\sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i}\Big[\eta\beta_I\langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I)-\hat{\sigma}_i^{*,\gamma}(I)\rangle - \frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 - \frac{\eta^2}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{t}(I)\|_2^2\Big]\\
&\overset{(ii)}{=} \sum_{t=0}^{T-1}\sum_{i\in\mathcal{N}}\Big[\eta\langle \hat{\boldsymbol{\ell}}_i^{t+1}, \hat{\boldsymbol{x}}_i^{t+1}-\hat{\boldsymbol{x}}_i^{*,\gamma,\mu}\rangle - \sum_{I\in\mathcal{I}_i}\frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 - \sum_{I\in\mathcal{I}_i}\frac{\eta^2}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{t}(I)\|_2^2\Big]\\
&\overset{(iii)}{\geq} \sum_{t=0}^{T-1}\sum_{i\in\mathcal{N}}\Big[\eta\mu\|\hat{\boldsymbol{x}}_i^{t+1}-\hat{\boldsymbol{x}}_i^{*,\gamma,\mu}\|_2^2 - \sum_{I\in\mathcal{I}_i}\frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 - \sum_{I\in\mathcal{I}_i}\frac{\eta^2}{4}\|\hat{\boldsymbol{m}}_i^{t+1}(I)-\hat{\boldsymbol{m}}_i^{t}(I)\|_2^2\Big]\\
&\overset{(iv)}{\geq} \sum_{t=0}^{T-1}\sum_{i\in\mathcal{N}}\Big[\eta\mu\|\hat{\boldsymbol{x}}_i^{t+1}-\hat{\boldsymbol{x}}_i^{*,\gamma,\mu}\|_2^2 - \sum_{I\in\mathcal{I}_i}\frac{\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - \sum_{I\in\mathcal{I}_i}\frac{\eta^2\kappa^2}{4}\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{t}\|_2^2\Big]\\
&= \sum_{t=0}^{T-1}\Big[\eta\mu\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}|\frac{\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}|\frac{\eta^2\kappa^2}{4}\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{t}\|_2^2\Big]\\
&\overset{(v)}{\geq} \sum_{t=0}^{T-1}\Big[\eta\mu\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}|\frac{\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}|\frac{\eta^2\kappa^2}{2}(\|\hat{\boldsymbol{x}}^{*,\gamma,\mu}-\hat{\boldsymbol{x}}^{t}\|_2^2 + \|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2)\Big]\\
&= \sum_{t=0}^{T-1}\Big[\eta\mu\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}|\eta^2\kappa^2\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}|\frac{\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{*,\gamma,\mu}-\hat{\boldsymbol{x}}^{t}\|_2^2\Big]\\
&= \sum_{t=0}^{T-1}\Big[(\frac{2\eta\mu}{3} - |\mathcal{I}|\eta^2\kappa^2)\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2\Big] + \sum_{t=0}^{T-1}\Big[(\frac{\eta\mu}{3})\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}|\frac{\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{*,\gamma,\mu}-\hat{\boldsymbol{x}}^{t}\|_2^2\Big]\\
&= \sum_{t=0}^{T-1}\Big[(\frac{2\eta\mu}{3} - |\mathcal{I}|\eta^2\kappa^2)\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2\Big] + \sum_{t=1}^{T-1}\Big[(\frac{\eta\mu}{3} - \frac{|\mathcal{I}|\eta^2\kappa^2}{2})\|\hat{\boldsymbol{x}}^{t}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2\Big]\\
&\quad + \frac{\eta\mu}{3}\|\hat{\boldsymbol{x}}^{T}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - \frac{|\mathcal{I}|\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{*,\gamma,\mu}-\hat{\boldsymbol{x}}^{0}\|_2^2. \qquad (31)
\end{aligned}
$$

where (i) due to Lemma D.5 and the non-negativity of Bregman divergence, (ii) due to Corollary D.1, (iii) due to the strong monotonicity in Lemma D.2, (iv) due to the smoothness of instantaneous regret in Lemma E.1,(v) due to the mean-value inequality $\|\boldsymbol{a}+\boldsymbol{b}\|_2^2 \leq 2\|\boldsymbol{a}\|_2^2 + 2\|\boldsymbol{b}\|_2^2$. Rearranging the terms in Eq. (31), we have

$$
\begin{aligned}
&\sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i}\zeta_I\Big[D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^0) - \eta\langle\hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^0\rangle\Big] + \frac{|\mathcal{I}|\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{*,\gamma,\mu}-\hat{\boldsymbol{x}}^{0}\|_2^2 \qquad (32)\\
&\geq \sum_{t=0}^{T-1}\Big[(\frac{2\eta\mu}{3} - |\mathcal{I}|\eta^2\kappa^2)\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2\Big] + \sum_{t=1}^{T-1}\Big[(\frac{\eta\mu}{3} - \frac{|\mathcal{I}|\eta^2\kappa^2}{2})\|\hat{\boldsymbol{x}}^{t}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2\Big] + \frac{\eta\mu}{3}\|\hat{\boldsymbol{x}}^{T}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2
\end{aligned}
$$

Setting $\frac{\eta\mu}{3} - \frac{|\mathcal{I}|\eta^2\kappa^2}{2} \geq \frac{\eta\mu}{6}$, i.e., $0 < \eta \leq \frac{\mu}{3|\mathcal{I}|\kappa^2}$, Eq. (32) leads to for any $T > 0$:

$$
\frac{\eta\mu}{3}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{x}}^{t+1}-\hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \leq \sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i}\zeta_I\Big[D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^0) - \eta\langle\hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^0\rangle\Big] + \frac{|\mathcal{I}|\eta^2\kappa^2}{2}\|\hat{\boldsymbol{x}}^{*,\gamma,\mu}-\hat{\boldsymbol{x}}^{0}\|_2^2 = O(1).
$$

Therefore, the sequence $\{\hat{\boldsymbol{x}}^t\}_{t\geq 1}$ converges to $\hat{\boldsymbol{x}}^{*,\gamma,\mu}$, showing the asymptotic last-iterate convergence of RTPCFR+ algorithm.

We next prove the following claim, which, combining with the aforementioned last-iterate convergence of optimistic OMD with step size $\eta \leq \mu/(3|\mathcal{I}|\kappa^2)$, finishes the whole proof for Theorem 4.1.

Claim E.1. *For any information set $I \in \mathcal{I}_i$, any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$, any step size $\eta > 0$, the output behavioral strategy $\hat{\sigma}_i^t$ of the optimistic OMD algorithm converges asymptotically to the Nash equilibrium $\hat{\sigma}^{*,\gamma,\mu}$.*

It suffices to show that there exists an initialization $\hat{\boldsymbol{\theta}}_I^{0'}$ and step size $\eta' = \mu/(3|\mathcal{I}|\kappa^2)$, such that the output strategy of optimistic OMD with $\hat{\boldsymbol{\theta}}_I^{0'}$ and $\eta' = \mu/(3|\mathcal{I}|\kappa^2)$ (whose output sequence strategy $\{\hat{\boldsymbol{x}}^t\}_{t\geq 1}$ converges to $\hat{\boldsymbol{x}}^{*,\gamma,\mu}$) is the same as that of OOMD with any fixed initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$ and any step size $\eta > 0$.

Actually, $\forall$ fixed initialization $\hat{\boldsymbol{\theta}}_I^0$, let $\hat{\boldsymbol{\theta}}_I^{0'} = \frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^0$. We demonstrate by induction the following relationship between unnormalized strategy $\{\hat{\boldsymbol{\theta}}_I^{t'}\}_{t\geq 0}$ and $\{\hat{\boldsymbol{\theta}}_I^t\}_{t\geq 0}$ generated by running OOMD with $(\hat{\boldsymbol{\theta}}_I^{0'}, \eta')$ and $(\hat{\boldsymbol{\theta}}_I^0, \eta)$ respectively: $\hat{\boldsymbol{\theta}}_I^{t'} = \frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^t$, for $t \geq 0$. The base case when $t = 0$ holds obviously due to the construction of initialization $\hat{\boldsymbol{\theta}}_I^{0'} = \frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^0$. Assume that claim holds when for iteration $t \geq 0$, then for the iteration $t+1$, we have

$$\hat{\boldsymbol{\theta}}_I^{t+1'} = [\hat{\boldsymbol{\theta}}_I^{t'} + \eta'\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = [\frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^t + \eta'\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = \frac{\eta'}{\eta}[\hat{\boldsymbol{\theta}}_I^t + \eta\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = \frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^{t+1},$$

where the above first inequality holds due to the update rule of OOMD, the second equality due to the inductive hypothesis. Therefore, $\hat{\boldsymbol{\theta}}_I^{t'} = \frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^t$, for $t \geq 0$, and so we further have that the output behavioral strategy of OOMD is

$$\begin{aligned}
\boldsymbol{\theta}_I^{t+1'} &= [\hat{\boldsymbol{\theta}}_I^{t'} + \eta'\hat{\boldsymbol{m}}_i^t(I)]^+ = [\frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^t + \eta'\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = \frac{\eta'}{\eta}[\hat{\boldsymbol{\theta}}_I^t + \eta\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = \frac{\eta'}{\eta}\boldsymbol{\theta}_I^{t+1} \\
\sigma_i^{t+1'} &= \frac{\boldsymbol{\theta}_I^{t+1'}}{\langle \boldsymbol{\theta}_I^{t+1'}, \mathbf{1}\rangle} = \frac{\frac{\eta'}{\eta}\boldsymbol{\theta}_I^{t+1}}{\frac{\eta'}{\eta}\langle \boldsymbol{\theta}_I^{t+1}, \mathbf{1}\rangle} = \sigma_i^{t+1},
\end{aligned}$$

which finishes the proof of Claim E.1. □

F Proof for the Average-Iterate Convergence of PCFR+ in Solving Perturbed Regularized EFG

Theorem F.1 (Equivalent Restatement of Theorem 4.2). *Assuming all players follow the update rule of PCFR$^+$ with any $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$ and $\eta > 0$, then the uniformly average strategy profile $\bar{\boldsymbol{x}}^T = \sum_{t=1}^T \boldsymbol{x}^t/T$ converges with a rate $O(1/\sqrt{T})$ to the set of NEs of the perturbed regularized EFGs in (3) for any $\gamma \geq 0$ and $\mu \geq 0$.*

Proof. We aim to give an upper bound for the i-th player's cumulative regret $R_i^T = \max_{\hat{\boldsymbol{x}}_i \in \boldsymbol{\mathcal{X}}_i^\gamma} \sum_{t=1}^T \langle \hat{\boldsymbol{\ell}}_i^t, \hat{\boldsymbol{x}}_i^t - \hat{\boldsymbol{x}}_i\rangle$. We start from Eq. (23) and set $\boldsymbol{\theta} = \sigma_i(I)$, obtaining that:

$$\begin{aligned}
&\eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i(I) - \boldsymbol{\theta}_I^{t+1}\rangle \\
\leq& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle \\
=& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - \eta\langle \hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle - \eta\langle \hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle \\
\leq& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) \\
&+ \eta\Big[\frac{\eta}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 + \frac{\|\boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\|_2^2}{2\eta}\Big] + \Big[D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\boldsymbol{\theta}_I^{t+1}) - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t)\Big] \\
=& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) + \frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t),
\end{aligned} \tag{33}$$

where the second inequality holds due to Cauchy-Schwartz/mean-value inequality (in the first square brackets) and Lemma D.1 (in the second square brackets). By using $\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I^{t+1}\rangle = \langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i^{t+1}(I)\rangle\|\boldsymbol{\theta}_I^{t+1}\|_1 = 0$, we can further simplify the left-hand side of Eq. (33) as

$$\begin{aligned}
&\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i(I) - \boldsymbol{\theta}_I^{t+1}\rangle = \langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i(I)\rangle = \langle \hat{\boldsymbol{v}}_i^{t+1}(I) - \langle \hat{\boldsymbol{v}}_i^{t+1}(I), \sigma_i^{t+1}(I)\rangle\mathbf{1}, \sigma_i(I)\rangle \\
=& \langle \hat{\boldsymbol{v}}_i^{t+1}(I), \sigma_i(I) - \sigma_i^{t+1}(I)\rangle = \frac{1}{1-\alpha_I}\langle \hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i(I) - \hat{\sigma}_i^{t+1}(I)\rangle.
\end{aligned} \tag{34}$$

Then, multiplying both sides in Eq. (33) with $\zeta_I = (1-\alpha_I)\pi_i^{\hat{\sigma}}(I)$, we have

$$\begin{aligned}&\pi_i^{\hat{\sigma}}(I)\langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I) - \hat{\sigma}_i(I)\rangle \\ \leq &\frac{\zeta_I}{\eta}\Big[D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) + \frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big]\end{aligned} \tag{35}$$

Taking summation of both sides in Eq. (35) over $t = 0, 1, \cdots, T-1$, $I \in \mathcal{I}_i$, and applying Lemma D.1, we have

$$\begin{aligned}&\sum_{t=0}^{T-1}\langle \hat{\boldsymbol{\ell}}_i^{t+1}, \hat{\boldsymbol{x}}_i^{t+1} - \hat{\boldsymbol{x}}_i\rangle \\ \leq &\sum_{t=0}^{T-1}\sum_{I\in\mathcal{I}_i}\frac{\zeta_I}{\eta}\Big[D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) + \frac{\eta^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\ \leq &\sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^0) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^T)}{\eta} + \frac{\eta}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big].\end{aligned} \tag{36}$$

According to the proof for Claim E.1, we know that, as long as we set another initialization $\hat{\boldsymbol{\theta}}_I^{0'} = \frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^0$, then for any step size $\eta' > 0$, the output sequence strategy $\{\hat{\boldsymbol{x}}_i^{t'}\}_{t\geq 1}$ of OOMD algorithm (with initialization $\hat{\boldsymbol{\theta}}_I^{0'}$ and step size η') is the same as that $\{\hat{\boldsymbol{x}}_i^t\}_{t\geq 1}$ of OOMD algorithm (with initialization $\hat{\boldsymbol{\theta}}_I^0$ and step size η). Applying such result into Eq. (36), we have for any $\eta' > 0$:

$$\begin{aligned}&\sum_{t=0}^{T-1}\langle \hat{\boldsymbol{\ell}}_i^{t+1}, \hat{\boldsymbol{x}}_i^{t+1} - \hat{\boldsymbol{x}}_i\rangle \\ \leq &\sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{0'}) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{T'})}{\eta'} + \frac{\eta'}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\ = &\sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{D_\psi(\sigma_i(I)||\frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^0) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{T'})}{\eta'} + \frac{\eta'}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\ \leq &\sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{\|\sigma_i(I) - \frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^0\|_2^2}{2\eta'} + \frac{\eta'}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\ \leq &\sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{\|\sigma_i(I)\|_2^2 + \|\frac{\eta'}{\eta}\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta'} + \frac{\eta'}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\ \leq &\sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{1}{\eta'} + \frac{\eta'\|\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta^2} + \frac{\eta'}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\ \leq &\sum_{I\in\mathcal{I}_i}2\zeta_I\sqrt{\frac{\|\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta^2} + \frac{1}{2}\sum_{t=0}^{T-1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2}\end{aligned}$$

where the third inequality holds due to the mean-value inequality $\|\boldsymbol{a}+\boldsymbol{b}\|^2 \leq 2\|\boldsymbol{a}\|^2 + 2\|\boldsymbol{b}\|^2$, the fourth inequality since for the simplex vector $\|\sigma_i(I)\|_1 = 1$, the last inequality due to the basic inequality $\frac{a}{\eta} + b\eta \leq 2\sqrt{ab}$ for any $\eta > 0$. □

G Proof for the Asymptotic Last-Iterate Convergence of PDCFR+ in Solving Perturbed Regularized EFG

Proof of Theorem 4.3. we first consider non-increasing sequence of step size $\{\eta^t\}_{t\geq 0}$, i.e. $\lim_{t\to\infty}\eta^t = \eta, \eta \in \big(0, \frac{\mu}{3|\mathcal{I}|\kappa^2}\big)$. The update rule of OOMD with bounded non-increasing learning

rate $\{\eta^t\}_{t\geq 0}$ is as follow:

$$\boldsymbol{\theta}_I^{t+1} \in \arg\min_{\boldsymbol{\theta}_I \in \mathbb{R}_{\geq 0}^{|A(I)|}} \left\{ \langle -\hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}_I \rangle + \frac{1}{\eta^{t+1}} D_\psi(\boldsymbol{\theta}_I || \hat{\boldsymbol{\theta}}_I^t) \right\}, \tag{37}$$

$$\hat{\boldsymbol{\theta}}_I^{t+1} \in \arg\min_{\boldsymbol{\theta}_I \in \mathbb{R}_{\geq 0}^{|A(I)|}} \left\{ \langle -\hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I \rangle + \frac{1}{\eta^{t+1}} D_\psi(\boldsymbol{\theta}_I || \hat{\boldsymbol{\theta}}_I^t) \right\}. \tag{38}$$

By replacing the constant step size η with η^{t+1} in Eqs. (24)-(30), we can obtain

$$\begin{aligned}
&\eta^{t+1}\beta_I \langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I) - \hat{\sigma}_i^{*,\gamma}(I) \rangle - \frac{(\eta^{t+1})^2 \zeta_I}{2} \|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 - \frac{(\eta^{t+1})^2 \zeta_I}{4} \|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 \\
&\leq \zeta_I \Big[D_\psi(\sigma_i^{*,\gamma}(I) || \hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i^{*,\gamma}(I) || \hat{\boldsymbol{\theta}}_I^{t+1}) - \eta^{t+1} \langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1} \rangle \Big].
\end{aligned} \tag{39}$$

Then, taking summation of both sides in Eq. (39) over $t = 0, 1, \cdots, T-1$, $i \in \mathcal{N}$, $I \in \mathcal{I}_i$, we have

$$\begin{aligned}
&\sum_{t=0}^{T-1} \sum_{i \in \mathcal{N}} \sum_{I \in \mathcal{I}_i} \Big[\eta^{t+1}\beta_I \langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I) - \hat{\sigma}_i^{*,\gamma}(I) \rangle - \frac{(\eta^{t+1})^2 \zeta_I}{2} \|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 \\
&\quad - \frac{(\eta^{t+1})^2 \zeta_I}{4} \|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 \Big] \\
&\leq \sum_{i \in \mathcal{N}} \sum_{I \in \mathcal{I}_i} \zeta_I \Big[D_\psi(\sigma_i^{*,\gamma}(I) || \hat{\boldsymbol{\theta}}_I^0) - D_\psi(\sigma_i^{*,\gamma}(I) || \hat{\boldsymbol{\theta}}_I^T) - \langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \sum_{t=0}^{T-1} \eta^{t+1} \big(\hat{\boldsymbol{\theta}}_I^t - \hat{\boldsymbol{\theta}}_I^{t+1} \big) \rangle \Big] \\
&= \sum_{i \in \mathcal{N}} \sum_{I \in \mathcal{I}_i} \zeta_I \Big[D_\psi(\sigma_i^{*,\gamma}(I) || \hat{\boldsymbol{\theta}}_I^0) - D_\psi(\sigma_i^{*,\gamma}(I) || \hat{\boldsymbol{\theta}}_I^T) - \langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \eta^1 \hat{\boldsymbol{\theta}}_I^0 - \eta^T \hat{\boldsymbol{\theta}}_I^T + \sum_{t=1}^{T-1} \big(\eta^{t+1} - \eta^t \big) \hat{\boldsymbol{\theta}}_I^t \rangle \Big] \\
&\leq \sum_{i \in \mathcal{N}} \sum_{I \in \mathcal{I}_i} \zeta_I \Big[D_\psi(\sigma_i^{*,\gamma}(I) || \hat{\boldsymbol{\theta}}_I^0) - \langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \eta^1 \hat{\boldsymbol{\theta}}_I^0 \rangle \Big],
\end{aligned} \tag{40}$$

where the last inequality holds due to the non-negativity of Bregman divergence, and the non-increasing property of sequence $\{\eta^t\}_{t\geq 0}$ (i.e. $\eta^{t+1} - \eta^t \leq 0$) and Lemma D.5. The left-hand side of Eq. (40) can be further lower bounded as in Eq. (31):

$$\begin{aligned}
&\sum_{t=0}^{T-1} \sum_{i \in \mathcal{N}} \sum_{I \in \mathcal{I}_i} \Big[\eta^{t+1}\beta_I \langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I) - \hat{\sigma}_i^{*,\gamma}(I) \rangle - \frac{(\eta^{t+1})^2 \zeta_I}{2} \|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^{*,\gamma}(I)\|_2^2 \\
&\quad - \frac{(\eta^{t+1})^2 \zeta_I}{4} \|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 \Big] \\
&\geq \sum_{t=0}^{T-1} \Big[(\frac{2\eta^{t+1}\mu}{3} - |\mathcal{I}|(\eta^{t+1})^2\kappa^2) \|\hat{\boldsymbol{x}}^{t+1} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big] + \sum_{t=0}^{T-1} \Big[(\frac{\eta^{t+1}\mu}{3}) \|\hat{\boldsymbol{x}}^{t+1} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - |\mathcal{I}| \frac{(\eta^{t+1})^2\kappa^2}{2} \|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^t\|_2^2 \Big] \\
&= \sum_{t=0}^{T-1} \Big[(\frac{2\eta^{t+1}\mu}{3} - |\mathcal{I}|(\eta^{t+1})^2\kappa^2) \|\hat{\boldsymbol{x}}^{t+1} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big] + \sum_{t=1}^{T-1} \Big[(\frac{\eta^t\mu}{3} - \frac{|\mathcal{I}|(\eta^{t+1})^2\kappa^2}{2}) \|\hat{\boldsymbol{x}}^t - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big] \\
&\quad + \frac{\eta^T\mu}{3} \|\hat{\boldsymbol{x}}^T - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 - \frac{|\mathcal{I}|(\eta^1)^2\kappa^2}{2} \|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^0\|_2^2 \\
&\geq \sum_{t=0}^{T-1} \Big[(\frac{2\eta^{t+1}\mu}{3} - |\mathcal{I}|(\eta^{t+1})^2\kappa^2) \|\hat{\boldsymbol{x}}^{t+1} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big] \\
&\quad + \sum_{t=1}^{T-1} \Big[(\frac{\eta^{t+1}\mu}{3} - \frac{|\mathcal{I}|(\eta^{t+1})^2\kappa^2}{2}) \|\hat{\boldsymbol{x}}^t - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big] - \frac{|\mathcal{I}|(\eta^1)^2\kappa^2}{2} \|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^0\|_2^2,
\end{aligned} \tag{41}$$

where the last inequality holds due to the fact that $\eta^t \geq \eta^{t+1}$. Combining the upper bound in Eq. (40) and lower bound in Eq. (41), and setting $\eta^{t+1} \in (\eta, \frac{\mu}{3|\mathcal{I}|\kappa^2}]$ we have

$$
\begin{aligned}
O(1) &= \sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i} \zeta_I \Big[D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^0) - \langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \eta^1 \hat{\boldsymbol{\theta}}_I^0 \rangle \Big] + \frac{|\mathcal{I}|(\eta^1)^2\kappa^2}{2} \|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^0\|_2^2 \\
&\geq \sum_{t=0}^{T-1} \Big[(\frac{2\eta^{t+1}\mu}{3} - |\mathcal{I}|(\eta^{t+1})^2\kappa^2) \|\hat{\boldsymbol{x}}^{t+1} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big] + \sum_{t=1}^{T-1} \Big[(\frac{\eta^{t+1}\mu}{3} - \frac{|\mathcal{I}|(\eta^{t+1})^2\kappa^2}{2}) \|\hat{\boldsymbol{x}}^{t} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big] \\
&\geq \sum_{t=0}^{T-1} \Big[(\frac{\eta\mu}{6}) \|\hat{\boldsymbol{x}}^{t+1} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \Big].
\end{aligned}
$$

Therefore, when running (weighted) OOMD algorithm in Eqs. (37)-(38), with non-increasing step sizes $\{\eta^t\}_{t\geq 0}$ that satisfies $\eta^t \in (\eta, \frac{\mu}{3|\mathcal{I}|\kappa^2}]$, the output sequence $\{\hat{\boldsymbol{x}}^t\}_{t\geq 1}$ satisfies the inequality for any $T \geq 1$:

$$
\sum_{t=1}^{T} \|\hat{\boldsymbol{x}}^t - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2^2 \leq \frac{6}{\eta\mu} \Big[\sum_{i\in\mathcal{N}}\sum_{I\in\mathcal{I}_i} \zeta_I \Big[D_\psi(\sigma_i^{*,\gamma}(I)||\hat{\boldsymbol{\theta}}_I^0) - \langle \hat{\boldsymbol{m}}_i^{*,\gamma}(I), \eta^1 \hat{\boldsymbol{\theta}}_I^0 \rangle \Big] + \frac{|\mathcal{I}|(\eta^1)^2\kappa^2}{2} \|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^0\|_2^2 \Big] = O(1).
$$

Therefore, $\{\hat{\boldsymbol{x}}^t\}_{t\geq 1}$ converges to the Nash equilibrium $\hat{\boldsymbol{x}}^{*,\gamma,\mu}$, completing the proof of the first part. □

We next prove the following claim, which, combining with the aforementioned last-iterate convergence of (weighted) optimistic OMD with step size $\eta^t \in (\eta, \frac{\mu}{3|\mathcal{I}|\kappa^2}]$, finishes the whole proof for Theorem 4.3.

Claim G.1. *For any information set $I \in \mathcal{I}_i$, any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$, any non-increasing step sizes $\{\eta^t\}_{t\geq 0}$ with $\lim_{t\to\infty} \eta^t = \eta > 0$, the output behavior strategy $\hat{\sigma}_i^t$ of the weighted OOMD algorithm converges asymptotically to the Nash equilibrium $\hat{\sigma}^{*,\gamma,\mu}$.*

For any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$, non-increasing step sizes $\{\eta^t\}_{t\geq 0}$, we construct another initialization $\hat{\boldsymbol{\theta}}_I^{0'} = \frac{\eta^{0'}}{\eta^0}\hat{\boldsymbol{\theta}}_I^0$ and sequence of step size $\{\eta^{t'}\}_{t\geq 0}$ which satisfies $\eta^{t+1\prime} = \frac{\eta^{t'}}{\eta^t}\eta^{t+1}$ and $\eta^{0'} \leq \frac{\mu}{3|\mathcal{I}|\kappa^2}$. We then prove the following lemma.

Lemma G.1. *For any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$, any non-increasing step sizes $\{\eta^t\}_{t\geq 0}$ with $\lim_{t\to\infty} \eta^t = \eta > 0$, we have that: the constructed sequence of step sizes $\{\eta^{t'}\}_{t\geq 0}$ is non-increasing and $\lim_{t\to\infty} \eta^{t'} = \eta^{0'}\frac{\eta}{\eta^0} \in (0, \frac{\mu}{3|\mathcal{I}|\kappa^2}]$.*

Proof. Due to the non-increasing property of $\{\eta^t\}_{t\geq 0}$, we have $\eta^{t+1\prime} = \frac{\eta^{t'}}{\eta^t}\eta^{t+1} \leq \eta^{t'}$ and so $\{\eta^{t'}\}_{t\geq 0}$ is non-increasing and non-negative. From the monotone convergence theorem, the limit of sequence $\{\eta^{t'}\}_{t\geq 0}$ exists, and can be computed as:

$$
\lim_{t\to\infty} \eta^{t+1\prime} = \lim_{t\to\infty} \frac{\eta^{t+1\prime}}{\eta^{t'}} \frac{\eta^{t'}}{\eta^{t-1\prime}} \cdots \frac{\eta^{1\prime}}{\eta^{0\prime}} \eta^{0\prime} = \lim_{t\to\infty} \frac{\eta^{t+1}}{\eta^t} \frac{\eta^t}{\eta^{t-1}} \cdots \frac{\eta^1}{\eta^0} \eta^{0\prime} = \lim_{t\to\infty} \eta^{t+1} \frac{\eta^{0\prime}}{\eta^0} = \frac{\eta^{0\prime}}{\eta^0}\eta.
$$

□

From Lemma G.1, the constructed non-increasing step sizes $\{\eta^{t'}\}_{t\geq 0}$ satisfy $\eta^{t'} \in (\eta^{0'}\frac{\eta}{\eta^0}, \frac{\mu}{3|\mathcal{I}|\kappa^2}]$, so that the output strategy $\hat{\sigma}_i^t$ of weighted optimistic OMD with $\hat{\boldsymbol{\theta}}_I^{0'}$ and $\{\eta^{t'}\}_{t\geq 0}$ converges to Nash equilibrium $\hat{\sigma}_i^{*,\gamma}$. We next give the following lemma, which completes the whole proof for Claim G.1.

Lemma G.2. *The output behavioral strategy of weighted OOMD algorithm with any (fixed) initialization $\hat{\boldsymbol{\theta}}_I^0$ and step sizes $\{\eta^t\}_{t\geq 0}$, is the same as that of weighted OOMD algorithm with the constructed initialization $\hat{\boldsymbol{\theta}}_I^{0'}$ and step size $\{\eta^{t'}\}_{t\geq 0}$*

Proof. We first demonstrate by induction that, for any iteration $t \geq 0$: $\hat{\boldsymbol{\theta}}_I^{t'} = \frac{\eta^{t'}}{\eta^t}\hat{\boldsymbol{\theta}}_I^t$.

The base case of $t = 0$ holds trivially due to the construction of initialization $\hat{\boldsymbol{\theta}}_I^{0'} = \frac{\eta^{0'}}{\eta^0}\hat{\boldsymbol{\theta}}_I^0$. Assume the proposition holds for the iteration t. Consider the iteration $t+1$ of weighted OOMD algorithm:

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{t+1'} =&[\hat{\boldsymbol{\theta}}^{t'} + \eta^{t+1'}\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = [\frac{\eta^{t'}}{\eta^t}\hat{\boldsymbol{\theta}}_I^t + \eta^{t+1'}\hat{\boldsymbol{m}}_i^{t+1}(I)]^+\\ =&[\frac{\eta^{t'}}{\eta^t}\hat{\boldsymbol{\theta}}_I^t + \frac{\eta^{t'}}{\eta^t}\eta^{t+1}\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = \frac{\eta^{t'}}{\eta^t}[\hat{\boldsymbol{\theta}}_I^t + \eta^{t+1}\hat{\boldsymbol{m}}_i^{t+1}(I)]^+\\ =&\frac{\eta^{t+1'}}{\eta^{t+1}}[\hat{\boldsymbol{\theta}}_I^t + \eta^{t+1}\hat{\boldsymbol{m}}_i^{t+1}(I)]^+ = \frac{\eta^{t+1'}}{\eta^{t+1}}\hat{\boldsymbol{\theta}}^{t+1},\end{aligned}$$

where the second equality holds due to the inductive hypothesis for iteration t, the third and the fifth equality due to the construction of step size $\eta^{t+1'}$. Therefore, for any $t \geq 0$: $\hat{\boldsymbol{\theta}}_I^{t'} = \frac{\eta^{t'}}{\eta^t}\hat{\boldsymbol{\theta}}_I^t$.

Then, for the output unnormalized strategy $\boldsymbol{\theta}_I^{t+1'}$ of weighted OOMD algorithm, we have

$$\boldsymbol{\theta}_I^{t+1'} = [\hat{\boldsymbol{\theta}}^{t'} + \eta^{t+1'}\hat{\boldsymbol{m}}_i^t(I)]^+ = [\frac{\eta^{t'}}{\eta^t}\hat{\boldsymbol{\theta}}^t + \frac{\eta^{t'}}{\eta^t}\eta^{t+1}\hat{\boldsymbol{m}}_i^t(I)]^+ = \frac{\eta^{t'}}{\eta^t}[\hat{\boldsymbol{\theta}}^t + \eta^{t+1}\hat{\boldsymbol{m}}_i^t(I)]^+ = \frac{\eta^{t'}}{\eta^t}\boldsymbol{\theta}_I^{t+1}.$$

Finally, the behavioral strategy of the weighted OOMD algorithm with constructed initialization and step size is computed as follow, and is the same as the original behavioral strategy:

$$\sigma_i^{t+1'} = \frac{\boldsymbol{\theta}_I^{t+1'}}{\langle \boldsymbol{\theta}_I^{t+1'}, \mathbf{1}\rangle} = \frac{\frac{\eta^{t'}}{\eta^t}\boldsymbol{\theta}_I^{t+1}}{\frac{\eta^{t'}}{\eta^t}\langle \boldsymbol{\theta}_I^{t+1}, \mathbf{1}\rangle} = \sigma_i^{t+1}. \qquad \square$$

H Proof for the Average-Iterate Convergence of PDCFR+ in Solving Perturbed Regularized EFG

Theorem H.1 (Equivalent Restatement of Theorem 4.4). *Assuming all players follow the update rule of PDCFR$^+$ with any initialization $\hat{\boldsymbol{\theta}}_I^0 \in \mathbb{R}_{\geq 0}^{|A(I)|}$ and any sequence of positive step sizes $\{\eta^t\}_{t\geq 0}$. If $\{\frac{w^t}{\eta^t}\}_{t\geq 0}$ is a positive non-increasing sequence, then the weighted average strategy profile $\bar{\boldsymbol{x}}_{\boldsymbol{w}}^T = \frac{\sum_{t=1}^T w^t \boldsymbol{x}^t}{\sum_{t=1}^T w^t}$ converges with a rate of $O(\frac{|\mathcal{I}_i|\sqrt{\sum_{t=0}^{T-1} w^{t+1}\eta^{t+1}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2}}{\sum_{t=0}^{T-1} w^{t+1}})$ to the set of NEs of the perturbed regularized EFGs in (3) for any $\gamma \geq 0$ and $\mu \geq 0$.*

Proof. We aim to give an upper bound for the i-th player's weighted cumulative regret $R_i^T = \max_{\hat{\boldsymbol{x}}_i \in \boldsymbol{\mathcal{X}}_i^\gamma} \sum_{t=1}^T w^t \langle \hat{\boldsymbol{\ell}}_i^t, \hat{\boldsymbol{x}}_i^t - \hat{\boldsymbol{x}}_i\rangle$. Similar to the proof for Theorem F.1, by replacing η in Eq. (23) with step size η^{t+1} and setting $\boldsymbol{\theta} = \sigma_i(I)$, we have

$$\begin{aligned}&\eta^{t+1}\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \sigma_i(I) - \boldsymbol{\theta}_I^{t+1}\rangle\\ \leq& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - \eta^{t+1}\langle \hat{\boldsymbol{m}}_i^{t+1}(I), \boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle\\ =& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t)\\ &- \eta^{t+1}\langle \hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle - \eta^{t+1}\langle \hat{\boldsymbol{m}}_i^t(I), \boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\rangle\\ \leq& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t)\\ &+ \eta^{t+1}\Big[\frac{\eta^{t+1}}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 + \frac{\|\boldsymbol{\theta}_I^{t+1} - \hat{\boldsymbol{\theta}}_I^{t+1}\|_2^2}{2\eta^{t+1}}\Big] + \Big[D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\hat{\boldsymbol{\theta}}_I^{t+1}||\boldsymbol{\theta}_I^{t+1}) - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t)\Big]\\ =& D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) + \frac{(\eta^{t+1})^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t), \qquad (42)\end{aligned}$$

Multiplying both sides in Eq. (42) with $w^{t+1}\zeta_I = w^{t+1}(1-\alpha_I)\pi_i^{\hat{\sigma}}(I)$, and applying Lemma D.1 to simplify the left-hand side of Eq. (42) we have

$$w^{t+1}\pi_i^{\hat{\sigma}}(I)\langle -\hat{\boldsymbol{v}}_i^{t+1}(I), \hat{\sigma}_i^{t+1}(I) - \hat{\sigma}_i(I)\rangle \qquad (43)$$

$$\leq \frac{w^{t+1}\zeta_I}{\eta^{t+1}}\Big[D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) + \frac{(\eta^{t+1})^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2 - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t)\Big],$$

Then, taking summation of both sides in Eq. (44) over $t = [T-1]$, $I \in \mathcal{I}_i$, and applying Lemma D.1, we have

$$\begin{aligned}
&\sum_{t=0}^{T-1} w^{t+1}\langle \hat{\boldsymbol{\ell}}_i^{t+1}, \hat{\boldsymbol{x}}_i^{t+1} - \hat{\boldsymbol{x}}_i\rangle \\
&\leq \sum_{t=0}^{T-1}\sum_{I\in\mathcal{I}_i}\frac{\zeta_I w^{t+1}}{\eta^{t+1}}\Big[D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1}) - D_\psi(\boldsymbol{\theta}_I^{t+1}||\hat{\boldsymbol{\theta}}_I^t) + \frac{(\eta^{t+1})^2}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\
&\leq \sum_{I\in\mathcal{I}_i}\zeta_I\Big[\sum_{t=0}^{T-1}\frac{w^{t+1}}{\eta^{t+1}}\Big(D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) - D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{t+1})\Big) + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1})}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\
&= \sum_{I\in\mathcal{I}_i}\zeta_I\Big[\sum_{t=1}^{T-1}\Big(\frac{w^{t+1}}{\eta^{t+1}} - \frac{w^t}{\eta^t}\Big)D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^t) + \frac{w^1}{\eta^1}D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^0) - \frac{w^T}{\eta^T}D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^T) \\
&\qquad + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1})}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \qquad (44) \\
&\leq \sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{w^1}{\eta^1}D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^0) + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1})}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big]. \qquad (45)
\end{aligned}$$

where the third inequality holds due to the non-increasing property of sequence $\{\frac{w^t}{\eta^t}\}_{t\geq 0}$ and the non-negativity of Bregman divergence $D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^T)$. Then, according to the analysis in Lemma G.2, for weighted OOMD algorithm running with any (fixed) initialization $\hat{\boldsymbol{\theta}}_I^0$ and step sizes $\{\eta^t\}_{t\geq 0}$, another weighted OOMD algorithm running with the initialization $\hat{\boldsymbol{\theta}}_I^{0'} = \frac{\eta^{0'}}{\eta^0}\hat{\boldsymbol{\theta}}_I^0$ (for any $\eta^{0'} > 0$) and the sequence of step sizes $\{\eta^{t+1'} = \frac{\eta^{t'}}{\eta^t}\eta^{t+1}\}_{t\geq 0}$ outputs the same sequence-form strategy $\hat{\boldsymbol{x}}_i^{t+1}$. Then, replacing η^{t+1} and $\hat{\boldsymbol{\theta}}_I^0$ in Eq. (45) with $\eta^{t+1'}$ and $\hat{\boldsymbol{\theta}}_I^{0'}$, we obtain that for any $\eta^{0'} > 0$:

$$\begin{aligned}
&\sum_{t=0}^{T-1} w^{t+1}\langle \hat{\boldsymbol{\ell}}_i^{t+1}, \hat{\boldsymbol{x}}_i^{t+1} - \hat{\boldsymbol{x}}_i\rangle \\
&\leq \sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{w^1}{\eta^{1'}}D_\psi(\sigma_i(I)||\hat{\boldsymbol{\theta}}_I^{0'}) + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1'})}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\
&\leq \sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{w^1(\|\sigma_i(I)\|_2^2 + \|\hat{\boldsymbol{\theta}}_I^{0'}\|_2^2)}{\eta^{1'}} + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1'})}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\
&= \sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{w^1\|\sigma_i(I)\|_2^2}{\eta^{1'}} + \frac{w^1\|\hat{\boldsymbol{\theta}}_I^{0'}\|_2^2}{\eta^{1'}} + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1'})}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\
&= \sum_{I\in\mathcal{I}_i}\zeta_I\Big[\frac{\eta^0 w^1\|\sigma_i(I)\|_2^2}{\eta^1\eta^{0'}} + \frac{w^1\eta^{0'}\|\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta^1\eta^0} + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1}\eta^{0'})}{2\eta^0}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big] \\
&\leq \sum_{I\in\mathcal{I}_i}\zeta_I 2\sqrt{\frac{\eta^0 w^1}{\eta^1}\Big[\frac{w^1\|\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta^1\eta^0} + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1})}{2\eta^0}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big]} \\
&= \sum_{I\in\mathcal{I}_i}\zeta_I 2\sqrt{\frac{w^1}{\eta^1}\Big[\frac{w^1\|\hat{\boldsymbol{\theta}}_I^0\|_2^2}{\eta^1} + \sum_{t=0}^{T-1}\frac{(w^{t+1}\eta^{t+1})}{2}\|\hat{\boldsymbol{m}}_i^{t+1}(I) - \hat{\boldsymbol{m}}_i^t(I)\|_2^2\Big]} \qquad (46)
\end{aligned}$$

where the second inequality holds due to the mean-value inequality $\|\boldsymbol{a}+\boldsymbol{b}\|_2^2 \le 2\|\boldsymbol{a}\|_2^2 + 2\|\boldsymbol{b}\|_2^2$, the second equality holds due to the fact that $\hat{\boldsymbol{\theta}}_I^{0'} = \frac{\eta^{0'}}{\eta^0}\hat{\boldsymbol{\theta}}_I^0$ and $\frac{\eta^{t+1'}}{\eta^{t+1}} = \frac{\eta^{t'}}{\eta^t} = \frac{\eta^{0'}}{\eta^0}$, the last inequality due to $\|\hat{\boldsymbol{\theta}}_I^0\|_2 \le \|\hat{\boldsymbol{\theta}}_I^0\|_1 = 1$ and the basic inequality $\frac{a}{\eta} + b\eta \le 2\sqrt{ab}$ for any $\eta > 0$. □

I Proofs for the Convergence of NE of Perturbed Regularized EFG

For any γ-perturbed μ-regularized EFG, let $h(\gamma,\mu) = \arg\min_{\hat{\boldsymbol{x}}_0 \in \boldsymbol{\mathcal{X}}_0^\gamma} \max_{\hat{\boldsymbol{x}}_1 \in \boldsymbol{\mathcal{X}}_1^\gamma} \hat{\boldsymbol{x}}_0^\top \boldsymbol{A}\hat{\boldsymbol{x}}_1 + \mu D_\psi(\hat{\boldsymbol{x}}_0, \boldsymbol{r}_0) - \mu D_\psi(\hat{\boldsymbol{x}}_1, \boldsymbol{r}_1)$ denote the set of NEs. The following lemma is a direct application of (Feinstein, 2022, Corollary 3.6) to the perturbed regularized EFG.

Lemma I.1. *For the perturbed regularized EFG in Eq. (3), let $\{(\gamma_k,\mu_k)\}_{k\ge1} \to (0,0)$ be a convergent positive sequence of parameters. Then,*

1. *The limit superior of the sequence of sets $\limsup_{k\to\infty} h(\gamma_k,\mu_k) \subseteq h(0,0)$ is a subset of the set of NEs of original EFG.*
2. *Besides, if the original EFG has the unique Nash Equilibrium, then for **any** convergent sequence $\{(\gamma_k,\mu_k)\}_{k\ge1} \to (0,0)$, we have $\lim_{k\to\infty} h(\gamma_k,\mu_k) = h(0,0)$.*

Proof of Proposition 5.1. ① $\lim\limits_{(\gamma,\mu)\to(0,0)} \hat{\boldsymbol{x}}^{*,\gamma,\mu} = \hat{\boldsymbol{x}}^* \Leftrightarrow \forall\epsilon > 0$, $\exists\delta_1 > 0$, when $|\gamma| < \delta_1$ and $|\mu| < \delta_1$, we have $\|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^*\|_2 < \epsilon/2$.

② Due to the strongly-convex strongly-concave property, the unperturbed regularized EFG $\min\max\limits_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}} f_\mu(\boldsymbol{x}_0,\boldsymbol{x}_1)$ has only one solution $\hat{\boldsymbol{x}}^{*,\mu}$. Then, according to Lemma I.1, the limit $\lim\limits_{\gamma\to0} \hat{\boldsymbol{x}}^{*,\gamma,\mu} = \hat{\boldsymbol{x}}^{*,\mu}$ exists. That is, $\forall\epsilon > 0$, $\exists\ \delta_2 > 0$, when $|\gamma| < \delta_2$, we have $\|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^{*,\mu}\|_2 < \epsilon/2$.

Combining the results in ① and ②, let $\delta = \min\{\delta_1,\delta_2\}$, then for any $\epsilon > 0$, when $|\mu| < \delta$ (and $|\gamma| < \delta$) , by applying triangle inequality we have

$$\|\hat{\boldsymbol{x}}^{*,\mu} - \hat{\boldsymbol{x}}^*\|_2 \le \|\hat{\boldsymbol{x}}^{*,\mu} - \hat{\boldsymbol{x}}^{*,\gamma,\mu}\|_2 + \|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^*\|_2 < \epsilon/2 + \epsilon/2 = \epsilon.$$

Therefore, we have $\lim\limits_{\mu\to0}\lim\limits_{\gamma\to0} \hat{\boldsymbol{x}}^{*,\gamma,\mu} = \lim\limits_{\mu\to0} \hat{\boldsymbol{x}}^{*,\mu} = \hat{\boldsymbol{x}}^*$. □

Proof of Proposition 5.2. According to Lemma I.1, the iterated limit $\lim\limits_{\mu\to0}\lim\limits_{\gamma\to0} \hat{\boldsymbol{x}}^{*,\gamma,\mu}$ is a Nash Equilibrium of the original EFG.

① $\lim_{\gamma\to0} \hat{\boldsymbol{x}}^{*,\gamma,\mu} = \hat{\boldsymbol{x}}^{*,\mu} \Leftrightarrow \forall\epsilon > 0$, there exists a real number $R(\epsilon,\mu) > 0$ when $0 < \gamma < R(\epsilon,\mu)$, we have $\|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^{*,\mu}\|_2 < \epsilon$.

In the above analysis, let $\epsilon = \mu$, and define function $g(\mu) = R(\mu,\mu) > 0$, then if $0 < \gamma < g(\mu)$, we have $\|\hat{\boldsymbol{x}}^{*,\gamma,\mu} - \hat{\boldsymbol{x}}^{*,\mu}\|_2 < \mu$. Further, notice that the sequence $\{(\gamma_k,\mu_k)\}_{k\ge1}$ satisfies $\gamma_k < g(\mu_k)$ and $\lim_{k\to\infty}\mu_k = 0$. Therefore, we have for any $\epsilon > 0$, there exists integer $K_1 \in \mathbb{N}^+$, when $k > K_1$, we have $\mu_k < \epsilon/2$ and so $\|\hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k} - \hat{\boldsymbol{x}}^{*,\mu_k}\|_2 < \epsilon/2$.

② Recall the assumption $\lim\limits_{\mu\to0} \hat{\boldsymbol{x}}^{*,\mu} = \lim\limits_{\mu\to0}\lim\limits_{\gamma\to0} \hat{\boldsymbol{x}}^{*,\gamma,\mu} = \hat{\boldsymbol{x}}^*$, that is, $\forall\epsilon > 0$, there exists $H(\epsilon) > 0$, when $0 < \mu < H(\epsilon)$, we have $\|\hat{\boldsymbol{x}}^{*,\mu} - \hat{\boldsymbol{x}}^*\|_2 < \epsilon$. Combining the convergent property $\lim_{k\to\infty}\mu_k = 0$, we can conclude that, for any $\epsilon > 0$, there exists $K_2 \in \mathbb{N}^+$, when $k > K_2$, $0 < \mu_k < H(\epsilon)$, thus we have $\|\hat{\boldsymbol{x}}^{*,\mu_k} - \hat{\boldsymbol{x}}^*\|_2 < \epsilon/2$.

Combining the results in ① and ②, for any $\epsilon > 0$, when $k > K = \max\{K_1,K_2\}$, we have

$$\|\hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k} - \hat{\boldsymbol{x}}^*\|_2 \le \|\hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k} - \hat{\boldsymbol{x}}^{*,\mu_k}\|_2 + \|\hat{\boldsymbol{x}}^{*,\mu_k} - \hat{\boldsymbol{x}}^*\|_2 < \epsilon.$$

In conclusion, the limit $\lim_{k\to\infty} \hat{\boldsymbol{x}}^{*,\gamma_k,\mu_k}$ exists. □

J Proofs for the Convergence of Algorithm 1 to the Approximate Nash Equilibrium of Original EFG

Proposition J.1. *Let $\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) \in \boldsymbol{\mathcal{X}}_0 \times \boldsymbol{\mathcal{X}}_1$ be such that $\|\boldsymbol{x}^{*,\gamma,\mu} - \bar{\boldsymbol{x}}\|_2 \le \nu$ for some $\nu \ge 0$, then we have* $\mathrm{Gap}(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) \le 2D\nu\sqrt{|\Sigma_0| \cdot |\Sigma_1|} + 6\mu D^2$.

Proof. Recall that $f_{\gamma,\mu}(\boldsymbol{x}_0, \boldsymbol{x}_1) = \boldsymbol{x}_0^\top \boldsymbol{A}\boldsymbol{x}_1 + \mu D_\psi(\boldsymbol{x}_0||\boldsymbol{r}_0) - \mu D_\psi(\boldsymbol{x}_1||\boldsymbol{r}_1)$. Then

$$
\begin{aligned}
&f_{\gamma,\mu}(\boldsymbol{x}_0, \bar{\boldsymbol{x}}_1) - f_{\gamma,\mu}(\bar{\boldsymbol{x}}_0, \boldsymbol{x}_1)\\
=&\boldsymbol{x}_0^\top \boldsymbol{A}\bar{\boldsymbol{x}}_1 + \mu D_\psi(\boldsymbol{x}_0||\boldsymbol{r}_0) - \mu D_\psi(\bar{\boldsymbol{x}}_1||\boldsymbol{r}_1) - \big[\bar{\boldsymbol{x}}_0^\top \boldsymbol{A}\boldsymbol{x}_1 + \mu D_\psi(\bar{\boldsymbol{x}}_0||\boldsymbol{r}_0) - \mu D_\psi(\boldsymbol{x}_1||\boldsymbol{r}_1)\big]\\
=&\boldsymbol{x}_0^\top \boldsymbol{A}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_1^{*,\gamma,\mu}) - (\bar{\boldsymbol{x}}_0 - \bar{\boldsymbol{x}}_0^{*,\gamma,\mu})^\top \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{x}_0^\top \boldsymbol{A}\boldsymbol{x}_1^{*,\gamma,\mu} - (\boldsymbol{x}_0^{*,\gamma,\mu})^\top \boldsymbol{A}\boldsymbol{x}_1\\
&+ \mu\big[D_\psi(\boldsymbol{x}_0||\boldsymbol{r}_0) - D_\psi(\bar{\boldsymbol{x}}_0||\boldsymbol{r}_0)\big] - \mu[D_\psi(\bar{\boldsymbol{x}}_1||\boldsymbol{r}_1) - D_\psi(\boldsymbol{x}_1||\boldsymbol{r}_1)]\\
&+ \mu D_\psi(\boldsymbol{x}_0||\boldsymbol{r}_0) - \mu D_\psi(\boldsymbol{x}_1^{*,\gamma,\mu}||\boldsymbol{r}_1) - \mu D_\psi(\boldsymbol{x}_0^{*,\gamma,\mu}||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_1||\boldsymbol{r}_1)\\
&- \mu D_\psi(\boldsymbol{x}_0||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_1^{*,\gamma,\mu}||\boldsymbol{r}_1) + \mu D_\psi(\boldsymbol{x}_0^{*,\gamma,\mu}||\boldsymbol{r}_0) - \mu D_\psi(\boldsymbol{x}_1||\boldsymbol{r}_1)\\
=&\boldsymbol{x}_0^\top \boldsymbol{A}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_1^{*,\gamma,\mu}) - (\bar{\boldsymbol{x}}_0 - \bar{\boldsymbol{x}}_0^{*,\gamma,\mu})^\top \boldsymbol{A}\boldsymbol{x}_1 + f_{\gamma,\mu}(\boldsymbol{x}_0, \boldsymbol{x}_1^{*,\gamma,\mu}) - f_{\gamma,\mu}(\boldsymbol{x}_0^{*,\gamma,\mu}, \boldsymbol{x}_1)\\
&- \mu D_\psi(\bar{\boldsymbol{x}}_0||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_0^{*,\gamma,\mu}||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_1^{*,\gamma,\mu}||\boldsymbol{r}_1) - \mu D_\psi(\bar{\boldsymbol{x}}_1||\boldsymbol{r}_1)\\
\overset{(i)}{\le}&\underbrace{\boldsymbol{x}_0^\top \boldsymbol{A}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_1^{*,\gamma,\mu}) - (\bar{\boldsymbol{x}}_0 - \bar{\boldsymbol{x}}_0^{*,\gamma,\mu})^\top \boldsymbol{A}\boldsymbol{x}_1}_{①}\\
&\underbrace{-\mu D_\psi(\bar{\boldsymbol{x}}_0||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_0^{*,\gamma,\mu}||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_1^{*,\gamma,\mu}||\boldsymbol{r}_1) - \mu D_\psi(\bar{\boldsymbol{x}}_1||\boldsymbol{r}_1)}_{②},
\end{aligned}
$$

where Inequality (i) holds due to the property of Nash equilibrium $\boldsymbol{x}^{*,\gamma,\mu}$ of perturbed regularized EFG. In remains to bound terms ①-②.

(1) Bounding $① = \boldsymbol{x}_0^\top \boldsymbol{A}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_1^{*,\gamma,\mu}) - (\bar{\boldsymbol{x}}_0 - \bar{\boldsymbol{x}}_0^{*,\gamma,\mu})^\top \boldsymbol{A}\boldsymbol{x}_1$.

$$
① \le \nu(\|\boldsymbol{A}^\top \boldsymbol{x}_0\|_2 + \|\boldsymbol{A}\boldsymbol{x}_1\|_2) \le \nu\|\boldsymbol{A}\|_2(\|\boldsymbol{x}_0\|_2 + \|\boldsymbol{x}_1\|_2) \le 2D\nu\sqrt{|\Sigma_0| \cdot |\Sigma_1|},
$$

where the last inequality holds because $\|\boldsymbol{A}\|_2 \le \|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j} \boldsymbol{A}_{i,j}^2} \le \sqrt{\sum_{i,j} 1} = \sqrt{|\Sigma_0| \cdot |\Sigma_1|}$.
(2) Bounding $② = -\mu D_\psi(\bar{\boldsymbol{x}}_0||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_0^{*,\gamma,\mu}||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_1^{*,\gamma,\mu}||\boldsymbol{r}_1) - \mu D_\psi(\bar{\boldsymbol{x}}_1||\boldsymbol{r}_1)$. Notice that for any $\boldsymbol{x}, \boldsymbol{y}$, we have $D_\psi(\boldsymbol{x}||\boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2 \le \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 \le 2D^2$. Then,

$$
② \le \mu D_\psi(\boldsymbol{x}_0^{*,\gamma,\mu}||\boldsymbol{r}_0) + \mu D_\psi(\boldsymbol{x}_1^{*,\gamma,\mu}||\boldsymbol{r}_1) \le 4\mu D^2.
$$

Combining the upper bounds for ① and ② , we have

$$
f_{\gamma,\mu}(\boldsymbol{x}_0, \bar{\boldsymbol{x}}_1) - f_{\gamma,\mu}(\bar{\boldsymbol{x}}_0, \boldsymbol{x}_1) \le 2D\nu\sqrt{|\Sigma_0| \cdot |\Sigma_1|} + 2\mu D^2. \tag{47}
$$

Besides, it is not difficult to see that for any $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1) \in \boldsymbol{\mathcal{X}}$,

$$
\begin{aligned}
&f(\boldsymbol{x}_0, \bar{\boldsymbol{x}}_1) - f(\bar{\boldsymbol{x}}_0, \boldsymbol{x}_1) - \big[f_{\gamma,\mu}(\boldsymbol{x}_0, \bar{\boldsymbol{x}}_1) - f_{\gamma,\mu}(\bar{\boldsymbol{x}}_0, \boldsymbol{x}_1)\big]\\
=&-\mu D_\psi(\boldsymbol{x}_0||\boldsymbol{r}_0) + \mu D_\psi(\bar{\boldsymbol{x}}_1||\boldsymbol{r}_1) + \big[\mu D_\psi(\bar{\boldsymbol{x}}_0||\boldsymbol{r}_0) - \mu D_\psi(\boldsymbol{x}_1||\boldsymbol{r}_1)\big] \le 4\mu D^2.
\end{aligned} \tag{48}
$$

Combing Eq. (47) and Eq.(48), we have the duality gap for the original EFG can be bounded as

$$
\mathrm{Gap}(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) = \min_{\boldsymbol{x}_0} \max_{\boldsymbol{x}_1} f(\boldsymbol{x}_0, \bar{\boldsymbol{x}}_1) - f(\bar{\boldsymbol{x}}_0, \boldsymbol{x}_1) \le 2D\nu\sqrt{|\Sigma_0| \cdot |\Sigma_1|} + 6\mu D^2. \qquad \square
$$

Proposition J.2. *The approximate Nash Equilibrium $\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) \in \boldsymbol{\mathcal{X}}_0^\gamma \times \boldsymbol{\mathcal{X}}_1^\gamma$ of a γ-perturbed EFG is an approximate of the NE of original EFG in terms of duality gap:*

$$
\max_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} \langle F(\bar{\boldsymbol{x}}), \bar{\boldsymbol{x}} - \boldsymbol{x}\rangle \le \max_{\hat{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}^\gamma} \langle F(\bar{\boldsymbol{x}}), \bar{\boldsymbol{x}} - \hat{\boldsymbol{x}}\rangle + 2\gamma PHA_{\max}
$$

Proof. Define $F(\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x}_1, -\boldsymbol{A}^\top \boldsymbol{x}_0)$. Let $\boldsymbol{x}_1 = \arg\max_{\tilde{\boldsymbol{x}}_1 \in \boldsymbol{\mathcal{X}}_1} \bar{\boldsymbol{x}}_0^\top \boldsymbol{A}\tilde{\boldsymbol{x}}_1$, where $\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1)$ is an approximate Nash Equilibrium of a γ-perturbed EFG, we have for any $\hat{\boldsymbol{x}}_1 \in \boldsymbol{\mathcal{X}}_1^\gamma$:

$$
\bar{\boldsymbol{x}}_0^\top \boldsymbol{A}\boldsymbol{x}_1 = \bar{\boldsymbol{x}}_0^\top \boldsymbol{A}\hat{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_0^\top \boldsymbol{A}(\boldsymbol{x}_1 - \hat{\boldsymbol{x}}_1) \le \max_{\hat{\boldsymbol{x}}_1 \in \boldsymbol{\mathcal{X}}_1^\gamma} \big[\bar{\boldsymbol{x}}_0^\top \boldsymbol{A}\hat{\boldsymbol{x}}_1 + \|\boldsymbol{A}^\top \bar{\boldsymbol{x}}_0\|_1 \cdot \|\boldsymbol{x}_1 - \hat{\boldsymbol{x}}_1\|_\infty\big].
$$

Table 2: Sizes of different extensive-form games, where #Histories is the total number of histories in the game tree, #Infosets the number of all information sets, #Terminal histories the number of leaf nodes, #Depth the depth of game tree, and #Max size of infosets is the largest number of histories contained at an infoset.

Game	#Histories	#Infosets	#Terminal histories	Depth	Max size of infosets
Kuhn Poker	58	12	30	6	2
Leduc Poker	9,457	936	5,520	12	5
Liar's Dice (4)	8,181	1,024	4,080	12	4
Liar's Dice (5)	51,181	5,120	25,575	14	5
Goofspiel (4)	1,077	162	576	7	14
Goofspiel (5)	26,931	2,124	14,400	9	46
Goofspiel (6)	969,523	34,482	518,400	11	230
Battleship (3)	732,607	81,027	552,132	9	7
HUNL Subgame (3)	398,112,843	69,184	261,126,360	10	1,980
HUNL Subgame (4)	244,005,483	43,240	158,388,120	8	1,980

Similarly, Let $\boldsymbol{x}_0 = \arg\min_{\tilde{\boldsymbol{x}}_0 \in \boldsymbol{\mathcal{X}}_0} \tilde{\boldsymbol{x}}_0^\top \boldsymbol{A} \bar{\boldsymbol{x}}_1$, we have for any $\hat{\boldsymbol{x}}_0 \in \boldsymbol{\mathcal{X}}_0^\gamma$:

$$\boldsymbol{x}_0^\top \boldsymbol{A} \bar{\boldsymbol{x}}_1 = \hat{\boldsymbol{x}}_0^\top \boldsymbol{A} \bar{\boldsymbol{x}}_1 + (\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)^\top \boldsymbol{A} \bar{x}_1 \geq \min_{\hat{\boldsymbol{x}}_0 \in \boldsymbol{\mathcal{X}}_0^\gamma} \left[\hat{\boldsymbol{x}}_0^\top \boldsymbol{A} \bar{\boldsymbol{x}}_1 + (\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)^\top \boldsymbol{A} \bar{x}_1 \right].$$

Combining the above two inequalities, we have

$$\begin{aligned}
&\mathrm{Gap}(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_1) = \max_{\tilde{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}} \langle F(\bar{\boldsymbol{x}}), \tilde{\boldsymbol{x}} \rangle \\
&= \max_{\tilde{\boldsymbol{x}}_1 \in \boldsymbol{\mathcal{X}}_1} \bar{\boldsymbol{x}}_0^\top \boldsymbol{A} \tilde{\boldsymbol{x}}_1 - \min_{\tilde{\boldsymbol{x}}_0 \in \boldsymbol{\mathcal{X}}_0} \tilde{\boldsymbol{x}}_0^\top \boldsymbol{A} \bar{\boldsymbol{x}}_1 \\
&\leq \max_{\hat{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}^\gamma} \left[\bar{\boldsymbol{x}}_0^\top \boldsymbol{A} \hat{\boldsymbol{x}}_1 - \hat{\boldsymbol{x}}_0^\top \boldsymbol{A} \bar{\boldsymbol{x}}_1 + \|\boldsymbol{A}^\top \bar{\boldsymbol{x}}_0\|_1 \cdot \|\boldsymbol{x}_1 - \hat{\boldsymbol{x}}_1\|_\infty - (\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)^\top \boldsymbol{A} \bar{\boldsymbol{x}}_1 \right] \\
&\leq \max_{\hat{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}^\gamma} \left[\langle F(\bar{\boldsymbol{x}}), \hat{\boldsymbol{x}} \rangle + \|\boldsymbol{A}^\top \bar{\boldsymbol{x}}_0\|_1 \cdot \|\boldsymbol{x}_1 - \hat{\boldsymbol{x}}_1\|_\infty + \|\boldsymbol{A} \bar{\boldsymbol{x}}_1\|_1 \cdot \|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0\|_\infty \right] \\
&\leq \max_{\hat{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}^\gamma} \left[\langle F(\bar{\boldsymbol{x}}), \hat{\boldsymbol{x}} \rangle + P \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_\infty \right], \qquad (49)
\end{aligned}$$

where the last inequality holds due to the P-boundedness of global loss $\boldsymbol{\ell}^\sigma$. It remains to upper bound $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_\infty$. Recall that for any information set $I \in \mathcal{I}_i$, the quality $\boldsymbol{x}_i(I) = \pi_i^\sigma(I)\sigma_i(I)$ holds for any sequence-form strategy $\boldsymbol{x}$ and the corresponding behavioral strategy σ, and the perturbed behavioral strategy $\hat{\sigma}_i(I) = (1 - \gamma|A(I)|)\sigma_i(I) + \gamma \mathbf{1}$, then we have

$$\begin{aligned}
&\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_\infty = \max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\pi_i^\sigma(I)\sigma_i(I) - \pi_i^{\hat{\sigma}}(I)\hat{\sigma}_i(I)\|_\infty \\
&= \max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\sigma_i(I)(\pi_i^{\hat{\sigma}}(I) - \pi_i^\sigma(I)) + \left[\gamma \mathbf{1} - \gamma |A(I)| \sigma_i(I)\right] \pi_i^{\hat{\sigma}(I)}\|_\infty \\
&\leq \underbrace{\max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\sigma_i(I)(\pi_i^{\hat{\sigma}}(I) - \pi_i^\sigma(I))\|_\infty}_{②} + \underbrace{\max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\left[\gamma \mathbf{1} - \gamma |A(I)| \sigma_i(I)\right] \pi_i^{\hat{\sigma}(I)}\|_\infty}_{①}, \qquad (50)
\end{aligned}$$

(1) Bounding $① = \max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\left[\gamma \mathbf{1} - \gamma |A(I)| \sigma_i(I)\right] \pi_i^{\hat{\sigma}(I)}\|_\infty$**.**

$$\begin{aligned}
① &= \gamma \max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\left[\mathbf{1} - |A(I)| \sigma_i(I)\right] \pi_i^{\hat{\sigma}(I)}\|_\infty \\
&\leq \gamma \max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\mathbf{1} - |A(I)| \sigma_i(I)\|_\infty \leq \gamma A_{\max}.
\end{aligned}$$

(2) Bounding $② = \max_{i \in \mathcal{N}, I \in \mathcal{I}_i} \|\sigma_i(I)(\pi_i^{\hat{\sigma}}(I) - \pi_i^\sigma(I))\|_\infty$**.**

For any information set $I \in \mathcal{I}_i$, let $\mathrm{Path}(I) = \{(I_1, a_1), (I_2, a_2), \ldots, (I_{H(I)}, a_{H(I)})\}$ denote the information-action pairs located in the path from the root of the game tree to the inforset I, where $H(I) \leq H$ is the depth of inforset I in the tree. Then, for any behavioral strategy σ, the reaching probability $\pi_i^\sigma(I) = \prod_{(I', a') \in \mathrm{Path}(I)} \sigma_i(I', a')$. Therefore, we have

$$|\pi_i^{\hat{\sigma}}(I) - \pi_i^\sigma(I)|$$

$$
\begin{aligned}
=&|\prod_{(I',a')\in\mathrm{Path}(I)}\hat{\sigma}_i(I',a')-\prod_{(I',a')\in\mathrm{Path}(I)}\sigma_i(I',a')|\\
\leq&|\prod_{(I',a')\in\mathrm{Path}(I)}\hat{\sigma}_i(I',a')-\Big(\prod_{\substack{(I',a')\in\mathrm{Path}(I)\\(I',a')\neq(I_1,a_1)}}\hat{\sigma}_i(I',a')\Big)\sigma_i(I_1,a_1)|\\
&+|\Big(\prod_{\substack{(I',a')\in\mathrm{Path}(I)\\(I',a')\neq(I_1,a_1)}}\hat{\sigma}_i(I',a')\Big)\sigma_i(I_1,a_1)-\prod_{(I',a')\in\mathrm{Path}(I)}\sigma_i(I',a')|\\
\leq&\gamma A_{\max}+|\prod_{\substack{(I',a')\in\mathrm{Path}(I)\\(I',a')\neq(I_1,a_1)}}\hat{\sigma}_i(I',a')-\prod_{\substack{(I',a')\in\mathrm{Path}(I)\\(I',a')\neq(I_1,a_1)}}\sigma_i(I',a')|\\
\leq&\gamma H A_{\max}.
\end{aligned}
$$

where the first inequality holds due to the triangle inequality; the second inequality due to the fact that the gap between original behavioral strategy σ_i and perturbed behavioral strategy $\hat{\sigma}_i$ can be bounded as $|\hat{\sigma}_i(I',a')-\sigma_i(I',a')|=|\gamma(1-|A(I')|\sigma_i(I'))|\leq\gamma A_{\max}$ and the fact that $\sigma_i(I_1,a_1)\in[0,1]$; the last inequality holds due to the induction. Then we have ② $=\max_{i\in\mathcal{N},I\in\mathcal{I}_i}\|\sigma_i(I)(\pi_i^{\hat{\sigma}}(I)-\pi_i^{\sigma}(I))\|_\infty\leq\max_{i\in\mathcal{N},I\in\mathcal{I}_i}|\pi_i^{\hat{\sigma}}(I)-\pi_i^{\sigma}(I)|\leq\gamma H A_{\max}$.

Plugging the results of **(1)**-**(2)** into Eq.(50), we obtain $\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_\infty\leq 2\gamma H A_{\max}$. Plugging this upper bound into Eq.(49) finishes the whole proof. □

K Details of Extensive-Form Games (EFGs)

Kuhn Poker Kuhn (1950) is a simplified two-player zero-sum poker game with a three-card deck and one chance to bet for each player. **Leduc Poker** Southey et al. (2005) is a larger game with a 6-card deck and two betting rounds. In **Liar's Dice (x)** (x=4, 5) Lisý et al. (2015), each player gets an x-sided dice, which they will roll at the beginning of each hand and take turn placing bets on the outcome. **Goofspiel (x)** (x=4, 5) Ross (1971) is a card game where each player has x cards and their goal is to score points by bidding simultaneously in x rounds. In **Battleship (3)** Farina et al. (2019d), each player conceals a 1×2 ship on her own $2\times x$ grids and then alternates turns launching three attacks against the opponent's vessel. **HUNL Subgame (x)** (x=3, 4) refers to a heads-up no-limit Texas hold'em (HUNL) subgame constructed by the superhuman poker agent Libratus Brown & Sandholm (2019). The sizes of different Extensive-Form Games (EFGs) are shown in Table 2. More details of the rule of these EFGs can be found in (Xu et al., 2024, Appendix B).

L Additional Experiments

Table 3: Hyperparameters used in RTPCFR$^+$(fine-tuned).

	Kuhn Poker	Leduc Poker	Battleship (3)	Liar's Dice (4)
μ	0.005	0.001	0.00005	0.001
T_u	50	200	10	10
	Liar's Dice (5)	Goofspiel (4)	Goofspiel (5)	Goofspiel (6)
μ	0.0005	0.0001	0.005	0.005
T_u	50	200	100	500

Computer Resources. Our implementations are based on Python. We run all algorithms on a platform with 4 NVIDIA RTX 6000 GPUs and 2 AMD EPYC 7543 Processors. Each GPU has 48G memory, and each CPU has 64 cores. The CUDA version is 12.1, the Python version 3.7.16, the matplotlib version 3.5.3. The source code for reproducing all experimental results is provided in the supplementary material.

Hyperparameter Setting of Baselines in EFGs. For OMWU and OGDA, we set the step size η to 0.5 and 0.1, respectively. For Reg-CFR, we use the same hyperparameters as in their original

Table 4: Hyperparameters used in RTPDCFR$^+$(fine-tuned).

	Kuhn Poker	Leduc Poker	Battleship (3)	Liar's Dice (4)
μ	0.1	0.0005	0.005	0.00005
T_u	10	100	10	10
α	1.3	2.5	2.8	1.5
	Liar's Dice (5)	Goofspiel (4)	Goofspiel (5)	Goofspiel (6)
μ	0.0001	0.01	0.01	0.005
T_u	50	10	50	500
α	2.0	3.0	2.8	3.0

paper (Liu et al., 2023, Appendix B). For RTCFR$^+$ Meng et al. (2025a), we set the initial values of γ and μ as $1\mathrm{e}{-10}$ and $1\mathrm{e}{-3}$, respectively, and set the number of iterations T_u required to update γ and $\boldsymbol{r}$ to 100. For our RTPDCFR$^+$, the hyperparameter α in discounting factor $\frac{t^\alpha}{1+t^\alpha}$ is set as $\alpha = 2.8$, the initial values $\gamma = 1\mathrm{e}{-10}$, $\mu = 5\mathrm{e}-5$, and the number of iterations is set as $T_u = 100$ to update $\boldsymbol{r}$. The optimal hyperparameters used in RTPCFR$^+$ and RTPDCFR$^+$ (i.e. corresponding to RTPCFR$^+$ (fine-tuned) and RTPDCFR$^+$ (fine-tuned) respectively) over eight small-to-medium-size EFG benchmarks are shown in Tables 3-4.

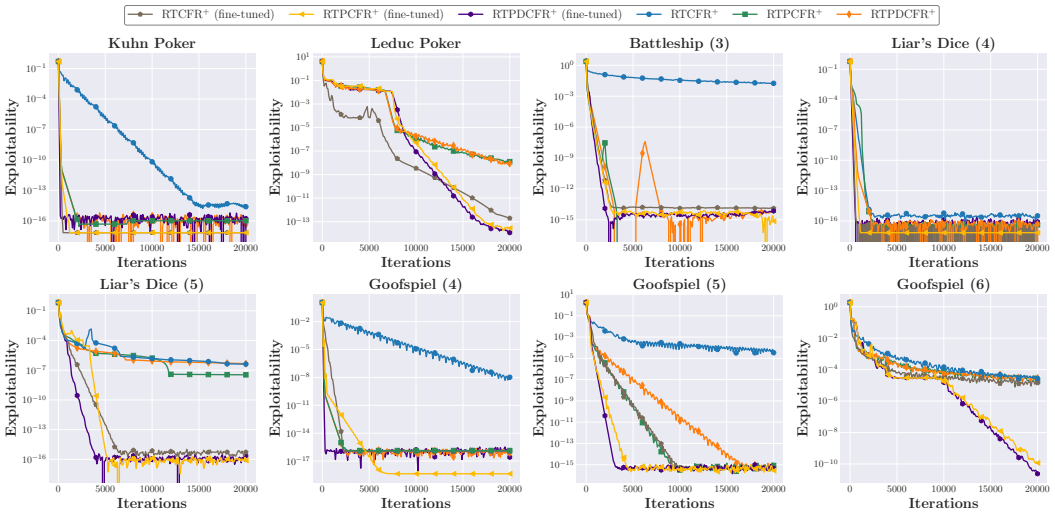

Figure 5: Last-iterate convergence of different reward transformation-based CFR methods in EFGs.

Empirical Last-Iterate Convergence Rates of different RT-based CFR Algorithms. We compare the last-iterate convergence of three reward transformation-based algorithms: RTCFR$^+$ Meng et al. (2025a), RTPCFR$^+$, RTPDCFR$^+$, as well as their fine-tuned versions in Figure 5. We can see that: **(1)** Our proposed RTPCFR$^+$ and RTPDCFR$^+$ (without fine-tuning) can achieve faster convergence than the latest RTPCFR$^+$ (without fine-tuning) algorithm in most EFG benchmarks (expect Leduc Poker), demonstrating the benefits of optimism and discounting technique in solving EFGs. **(2)** RTPCFR$^+$ (fine-tuned) and RTPDCFR$^+$ (fine-tuned) also outperform RTCFR$^+$ (fine-tuned) in most EFGs. Specifically, RTPDCFR$^+$ (fine-tuned) yields faster convergence rates and lower exploitabilities in all EFG benchmarks. RTPCFR$^+$ (fine-tuned) enjoys faster convergence rates and lower exploitabilities in 8 EFG benchmarks. In particular, RTPCFR$^+$ (fine-tuned) obtains much lower exploitability than RTCFR$^+$ (fine-tuned) in Goofspiel (4). **(3)** For the comparison between RTPCFR$^+$ and RTPDCFR$^+$, RTPDCFR$^+$ can achieve lower exploitabilities in medium-size EFG benchmarks, like Goofspiel (6), revealing the potential of discounting technique in solving large-scale EFGs. RTPDCFR$^+$ (fine-tuned) also obtains lower exploitabilities in these two EFGs than the fine-tuned version of RTPCFR$^+$. However, in Goofspiel (4), RTPCFR$^+$ (fine-tuned) attains a much lower exploitability than RTPDCFR$^+$ (fine-tuned). This indicates that, RTPCFR$^+$ seemingly has an edge in solving small-sized EFGs.

Comparison with Average-Iterate Convergence of CFR algorithms. We compare the last/average-iterate convergence of RTPCFR$^+$/RTPDCFR$^+$ with the linear-average-iterate convergence of tradi-

tional CFR algorithms. The experimental results are shown in Figure 6. We can observe that: **(1)** In most EFGs except Goofspiel (6), the last-iterate strategies of RTPCFR$^+$/RTPDCFR$^+$ always obtain lower exploitability than their linear-average strategies. **(2)** In most EFGs, the linear-average strategies of RTPCFR$^+$/RTPDCFR$^+$ obtains truly similar performance as the linear-average strategy of PCFR$^+$. **(3)** In most EFG benchmarks except Goofspiel (6), RTPCFR$^+$/RTPDCFR$^+$ achieves significantly faster convergence rates to the NEs of original EFGs, than the average-iterate convergence rate of the state-of-the-art PCFR$^+$ algorithm. Concretely, in Kuhn Poker, Battleship (3), Liar's Dice (4), Goofspiel (4), and Goofspiel (5) these five EFG settings, our RTPCFR$^+$ achieves at least five orders of magnitude faster convergence rate than the average-iterate convergence rate of PCFR$^+$. But in Goofspiel (6), PCFR$^+$ converges faster than the last-iterate strategy of RTPCFR$^+$/RTPDCFR$^+$.

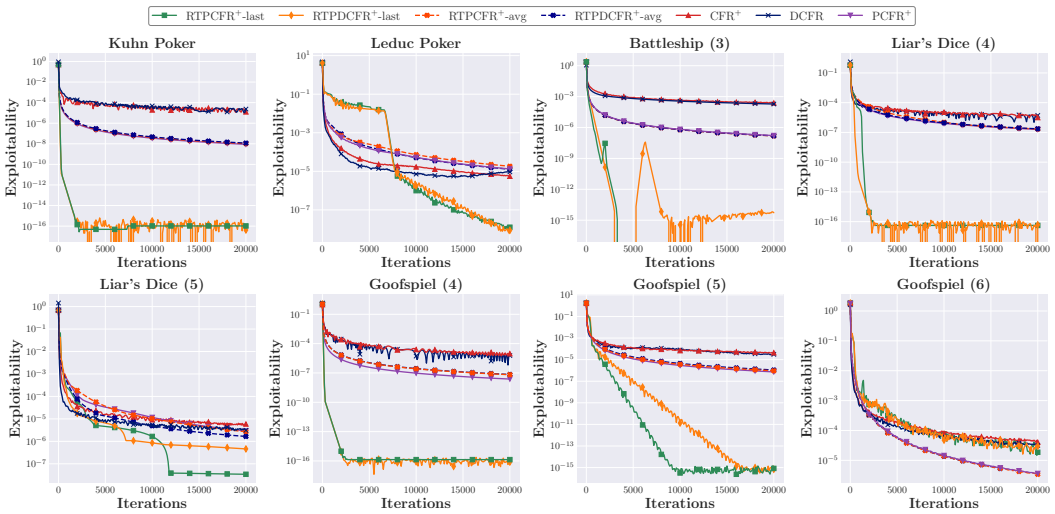

Figure 6: The comparison between last-iterate convergence of RTPCFR+/RTPDCFR+ and linear-average-iterate convergence of CFR variants in EFGs.