Equivalence Analysis between Counterfactual Regret Minimization and Online Mirror Descent

Weiming Liu [1] **Huacong Jiang** [2] **Bin Li** [2] **Houqiang Li** [2]

Abstract

Follow-the-Regularized-Leader (FTRL) and Online Mirror Descent (OMD) are regret minimization algorithms for Online Convex Optimization (OCO), they are mathematically elegant but less practical in solving Extensive-Form Games (EFGs). Counterfactual Regret Minimization (CFR) is a technique for approximating Nash equilibria in EFGs. CFR and its variants have a fast convergence rate in practice, but their theoretical results are not satisfactory. In recent years, researchers have been trying to link CFRs with OCO algorithms, which may provide new theoretical results and inspire new algorithms. However, existing analysis is restricted to local decision points. In this paper, we show that CFRs with Regret Matching and Regret Matching+ are equivalent to special cases of FTRL and OMD, respectively. According to these equivalences, a new FTRL and a new OMD algorithm, which can be considered as extensions of vanilla CFR and CFR+, are derived. The experimental results show that the two variants converge faster than conventional FTRL and OMD, even faster than vanilla CFR and CFR+ in some EFGs.

1. Introduction

An Extensive-Form Game (EFG) involves multiple players and sequential decisions. In this paper, we focus on two-player zero-sum EFGs with imperfect information, for example, heads-up no-limit Texas hold'em poker (HUNL). One can approximate a Nash equilibrium in this kind of game using iterative regret minimization algorithms, e.g., Counterfactual Regret Minimization (CFR) (Zinkevich et al., 2007). CFR minimizes the *total regret* of each player by minimizing the *counterfactual regrets* in local decision points. CFRs usually use Regret Matching (RM) (Zinkevich et al., 2007) and Regret Matching+ (RM+) (Tammelin et al., 2015) for minimizing the counterfactual regrets, resulting in CFR-RM and CFR-RM+ algorithms, respectively. In recent years, many variants of CFR have been proposed (Tammelin et al., 2015; Brown & Sandholm, 2019b; Farina et al., 2021). Although CFRs can only guarantee to converge to a Nash equilibrium at a rate of $O(1/\sqrt{T})$, they usually converge much faster in practice. Because of the superior performance and the parameter-free property, CFR and its variants have been applied in multiple super-human HUNL agents (Moravčík et al., 2017; Brown & Sandholm, 2018; 2019a). However, the theoretical results for CFRs are not satisfactory. Fundamentally, CFR-RM and CFR-RM+ are specialized for EFGs, which makes them difficult to analyze.

Follow-the-Regularized-Leader (FTRL) (Abernethy et al., 2008) and Online Mirror Descent (OMD) (Beck & Teboulle, 2003) are two prominent Online Convex Optimization (OCO) algorithms (Shalev-Shwartz, 2012; Hazan, 2016). OCO algorithms are mathematically elegant and the theoretical results are promising. In (Farina et al., 2019b), the optimistic variants of FTRL and OMD have been applied to EFGs, showing a theoretical convergence rate of $O(1/T)$. However, they remain less competitive than the SOTA CFRs (Brown & Sandholm, 2019b; Farina et al., 2021) in practice. There are some other first-order methods (Hoda et al., 2010; Kroer et al., 2020) for EFGs. However, they are also inferior in performance to the SOTA CFRs.

Recently, researchers have been interested in linking CFRs with OCO (and first-order) algorithms (Waugh & Bagnell, 2015; Farina et al., 2021), which may provide some insight on the superior performance of CFRs or may help to design new regret minimization algorithms for EFGs. In (Waugh & Bagnell, 2015), the authors have proven that RM is equivalent to dual average (a form of FTRL) (Nesterov, 2009). In (Farina et al., 2021), it has been proven that the results of RM and RM+ can be recovered by FTRL and OMD, respectively. And based on these relationships, the optimistic variants of CFR and CFR+ have been proposed. However, these work only considers the connection between RM (RM+) and FTRL (OMD), which can not be extended to

[1]School of Data Science, University of Science and Technology of China, Hefei, China [2]Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. Correspondence to: Weiming Liu <weiming@mail.ustc.edu.cn>, Bin Li <binli@ustc.edu.cn>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

the equivalence between CFR-RM (CFR-RM+) and FTRL (OMD).

In this paper, we first propose a Future-Dependent FTRL (FD-FTRL) and a Future-Dependent OMD (FD-OMD), which have a special regularizer that depends on *future* decisions. Crucially, they belong to the FTRL and OMD families and are able to leverage many existing theoretical results. Then, we prove that CFR-RM and CFR-RM+ are equivalent to special cases of FD-FTRL and FD-OMD, respectively. The equivalences reveal that: 1) the cumulative counterfactual regrets in CFRs can be viewed as adaptive regularization parameters; 2) CFRs are special adaptive FTRL and OMD, which may partially explain the superior performance of CFRs over FTRL and OMD; 3) FTRL and OMD are more general than CFRs, and, contrary to previous findings (Farina et al., 2019b), they are not necessarily worse than CFRs in EFGs, as long as they are configured properly. In order to investigate whether (FD-)FTRL and (FD-)OMD can converge faster than CFRs in EFGs, two practical implementations of FD-FTRL and FD-OMD are presented, and various configurations are tested. Experimental results show that FD-FTRL and FD-OMD can recover vanilla CFR and CFR+, respectively, and they can even converge faster than vanilla CFR and CFR+ in some EFGs. In conclusion, the contributions of the paper are:

- An equivalence between CFR-RM (CFR-RM+) and FTRL (OMD) is established, which may provide opportunities for communication between CFR and OCO.
- As the bridges, FD-FTRL and FD-OMD with a future-dependent regularizer are proposed. Besides, experiments involving various configurations are performed, showing that (FD-)FTRL and (FD-)OMD can be competitive compared with CFRs.

2. Preliminaries

The process of a player in a two-player zero-sum EFG can be described as a sequential decision process (Farina et al., 2019a). A sequential decision process consists of two kinds of points: *decision points* and *observation points*. The set of decision points is denoted by $\mathcal{J}$, and the set of observation points is denoted by $\mathcal{K}$. At each decision point $j \in \mathcal{J}$, the agent has to make a *(local) decision* $\hat{\boldsymbol{x}}_j \in \Delta^{n_j}$, where Δ^{n_j} is a simplex over the *action set* A_j and $n_j = |A_j|$. The set of all actions is denoted by $\mathcal{A}$. The combination of $\hat{\boldsymbol{x}}_j$ in all decision points is called a *strategy*. Let $\hat{\boldsymbol{x}}_j[a]$ be the probability of choosing *action* $a \in A_j$. Each action leads the agent to an observation point $k \in \mathcal{K}$, denoted by $k = \rho(j, a)$. At each observation point, the agent will receive a signal $s \in S_k$. After observing the signal, the agent reaches another decision point $j' \in \mathcal{J}$, written as $j' = \rho(k, s)$. The set of decision points that are earliest reachable after choosing action a at j is denoted by $C_{j,a}$. Formally, $C_{j,a} = \{\rho(\rho(j, a), s) : s \in S_{\rho(j,a)}\}$. If $j' \in C_{j,a}$, we say that j' is a child of j and j is the parent of j'. Denote the set of all descending decision points of j (including j) by $C_{\#j} = \{j\} \cup \bigcup_{j' 2 C_{j,a}, a 2 A_j} C_{\#j'}$. We assume that the process forms a tree. In other words, $C_{j,a} \cap C_{j',a'} = \emptyset$ for any $(j, a) \neq (j', a')$. This is equivalent to the perfect-recall assumption in EFGs. We assume a sequential decision process always starts from a decision point, named the *root decision point* and denoted by o. An illustration is given in Appendix B.

2.1. Sequence-Form Strategy

A strategy can be represented in a sequence form (Von Stengel, 1996). A *sequence* is a series of (j, a) starting from the root in a sequential decision process. In sequence-form representation, a strategy is the combination of the probabilities of playing each sequence. In this paper, we will formulate the sequence-form strategy space as a treeplex (Hoda et al., 2010) and follow the construction in (Farina et al., 2019b). Formally, denote the sequence-form strategy space by $\mathcal{X}$, and denote a strategy in $\mathcal{X}$ by $\boldsymbol{x}$. $\mathcal{X}$ can be obtained recursively: at every decision point $j \in \mathcal{J}$, let $\mathcal{X}_{j,a} = \prod_{j' 2 C_{j,a}} \mathcal{X}_{j'}$ (cartesian product); let

$$\mathcal{X}_j = \big\{(\hat{\boldsymbol{x}}_j, \hat{\boldsymbol{x}}_j[a_1]\boldsymbol{x}_{j,a_1} \dots, \hat{\boldsymbol{x}}_j[a_{n_j}]\boldsymbol{x}_{j,a_{n_j}}) : \hat{\boldsymbol{x}}_j \in \Delta^{n_j}, \boldsymbol{x}_{j,a_1} \in \mathcal{X}_{j,a_1}, \dots, \boldsymbol{x}_{j,a_{n_j}} \in \mathcal{X}_{j,a_{n_j}}\big\},$$

where $\hat{\boldsymbol{x}}_j \in \Delta^{n_j}$ and $(a_1, \dots, a_{n_j}) = A_j$; let $\mathcal{X} = \mathcal{X}_o$. As we can see, each entry is corresponding to a sequence with the value representing the probability of playing the whole sequence. Crucially, $\mathcal{X}$ and all $\mathcal{X}_j$ are treeplexes, so they are convex and compact (Hoda et al., 2010). Intuitively, $\mathcal{X}_j$ is the sequence-form strategy space of the *sub-sequential decision process* that starts from decision point j.

Given a concatenated vector $\boldsymbol{z} = (\boldsymbol{z}_{j_1}, \dots, \boldsymbol{z}_{j_m}) \in \mathsf{R}^{\sum_{k=1}^{m} n_{j_k}}$, e.g., an $\boldsymbol{x} \in \mathcal{X}$, we may need to isolate a sub-vector related to decision point j only. Formally, let $\boldsymbol{z}[j]$ represent the n_j entries related to decision point j, and let $\boldsymbol{z}[j, a]$ represent the entry corresponding to (j, a). Besides, let $\boldsymbol{z}[\downarrow j]$ denote the sub-vector corresponding to the decision points in $C_{\#j}$. For any vector $\boldsymbol{z} \in \mathsf{R}^{n_j}$, e.g., a decision $\hat{\boldsymbol{x}}_j$ at j, we use $\boldsymbol{z}[a]$ to represent the entry corresponding to a. Let p_j denote the pair (j', a') such that $j \in C_{j',a'}$. Then, $\boldsymbol{x}[p_j] = \boldsymbol{x}[j', a']$ is the probability of reaching decision point j. Note that p_o is undefined. For convenience, we let $\boldsymbol{x}[p_o] = 1$. Based on the above definitions, there are simple mappings among an $\boldsymbol{x} \in \mathcal{X}$, an $\boldsymbol{x}_j \in \mathcal{X}_j$, and a decision $\hat{\boldsymbol{x}}_j \in \Delta^{n_j}$. Formally, we *redefine* that $\boldsymbol{x}_j = \boldsymbol{x}[\downarrow j]/\boldsymbol{x}[p_j]$ and $\hat{\boldsymbol{x}}_j = \boldsymbol{x}[j]/\boldsymbol{x}[p_j]$ for any $j \in \mathcal{J}$. In the rest of the paper, we use normal symbols, e.g., $\boldsymbol{x}$ and $\boldsymbol{x}_j$, to represent the variables that are related to the sequence-form space, and use symbols with hats, e.g., $\hat{\boldsymbol{x}}_j$, to represent the variables

that are related to local decision points.

2.2. Nash Equilibrium and Regret Minimization

Based on the sequence-form representation, the problem of finding a Nash equilibrium in a two-player zero-sum EFG with perfect recall can be formulated as a bilinear saddle-point problem (BSPP) (Kroer et al., 2020). A BSPP for an EFG has the form

$$\min_{\boldsymbol{x}\in X}\max_{\boldsymbol{y}\in Y}\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y}=\max_{\boldsymbol{y}\in Y}\min_{\boldsymbol{x}\in X}\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y}, \tag{1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the strategy spaces for player 1 and player 2, respectively. $\mathbf{A}$ is a matrix encoding the losses for player 1, so $\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y}$ is the expected loss for it. Without loss of generality, the main results will be represented under the viewpoint of player 1. Define the *strategy profile* as $(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}$. The *exploitability* of $(\boldsymbol{x},\boldsymbol{y})$ is defined as

$$e(\boldsymbol{x},\boldsymbol{y})=\max_{\boldsymbol{y}'\in Y}\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y}'-\min_{\boldsymbol{x}'\in X}\boldsymbol{x}'^\top\mathbf{A}\boldsymbol{y}. \tag{2}$$

A Nash equilibrium is a strategy profile $(\boldsymbol{x}^*,\boldsymbol{y}^*)$ such that $e(\boldsymbol{x}^*,\boldsymbol{y}^*)=0$. Let $\boldsymbol{l}=\mathbf{A}\boldsymbol{y}$. The expected loss for player 1 can be reformulated as $\langle\boldsymbol{l},\boldsymbol{x}\rangle$, which is *linear* in $\boldsymbol{x}$. Note that $\langle\boldsymbol{l},\boldsymbol{x}\rangle=\sum_{j\in J}\langle\boldsymbol{l}[j],\boldsymbol{x}[j]\rangle=\sum_{j\in J}\boldsymbol{x}[p_j]\langle\boldsymbol{l}[j],\hat{\boldsymbol{x}}_j\rangle$. Based on this linearity, the sequence-form representation has been used with linear programming (Koller et al., 1996), first-order methods (Hoda et al., 2010; Kroer et al., 2020), and regret minimization algorithms (Farina et al., 2019b) for approximating Nash equilibria in zero-sum EFGs.

A regret minimization algorithm observes a loss $\boldsymbol{l}^t$ at every iteration and chooses a strategy $\boldsymbol{x}^{t+1}\in\mathcal{X}$ based on the losses $\boldsymbol{l}^1,\cdots,\boldsymbol{l}^t$ and the previous strategies $\boldsymbol{x}^1,\cdots,\boldsymbol{x}^t$. The target is to minimize the *total regret*, defined as

$$R^T=\max_{\boldsymbol{x}'\in X}\sum_{t=1}^{T}\left\{\langle\boldsymbol{l}^t,\boldsymbol{x}^t\rangle-\langle\boldsymbol{l}^t,\boldsymbol{x}'\rangle\right\}. \tag{3}$$

Define the *cumulative loss* as $\boldsymbol{L}^T=\sum_{t=1}^{T}\boldsymbol{l}^t$. The framework is given in Algorithm 1. It is well known that in a two-player zero-sum game, $\epsilon(\overline{\boldsymbol{x}}^T,\overline{\boldsymbol{y}}^T)=(R_1^T+R_2^T)/T$, where R_1^T and R_2^T are the total regrets for player 1 and player 2, respectively, $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ are the average strategies. Therefore, if R_1^T and R_2^T grow sub-linearly, $(\overline{\boldsymbol{x}}^T,\overline{\boldsymbol{y}}^T)$ will converge to a Nash equilibrium as $T\to\infty$.

Algorithm 1 Regret Minimization Framework

for iteration $t=1$ to T **do**
  $\boldsymbol{l}^t\leftarrow\text{ObserveLoss}(\boldsymbol{x}^t)$.
  $\boldsymbol{x}^{t+1}\leftarrow\text{Update}(\boldsymbol{l}^1,\dots,\boldsymbol{l}^t,\boldsymbol{x}^1,\dots,\boldsymbol{x}^t)$.
end for

Algorithm 2 Localized Regret Minimization Framework

function Update($\boldsymbol{l}^1,\dots,\boldsymbol{l}^t,\boldsymbol{x}^1,\dots,\boldsymbol{x}^t$)
  for node $j\in\mathcal{J}$ in bottom-up order **do**
    $\hat{\boldsymbol{x}}_j^{t+1}\leftarrow\text{LocalUpdate}(\boldsymbol{l}^1,\dots,\boldsymbol{l}^t,\boldsymbol{x}^1,\dots,\boldsymbol{x}^t)$.
  end for
  Construct and **Return** $\boldsymbol{x}^{t+1}$.
end function

2.3. Counterfactual Regret Minimization (CFR)

CFR is a regret minimization algorithm for two-player zero-sum games. It has been proven that the total regret of each player after T iterations is bounded by $O(\sqrt{T})$ and thus the average strategies will converge to a Nash equilibrium at a rate of $O(1/\sqrt{T})$ (Zinkevich et al., 2007). In (Farina et al., 2019a), CFR has been reformulated based on the sequence-form representation. In this paper, we will follow the formulation. Given a strategy $\boldsymbol{x}^t\in\mathcal{X}$ and a loss $\boldsymbol{l}^t\in\mathbb{R}^{\sum_{j\in J}n_j}$, CFR constructs a *counterfactual loss* $\hat{\boldsymbol{l}}_j^t\in\mathbb{R}^{n_j}$ recursively for each $j\in\mathcal{J}$:

$$\hat{\boldsymbol{l}}_j^t[a]=\boldsymbol{l}^t[j,a]+\sum_{j'\in C_{j,a}}\langle\hat{\boldsymbol{l}}_{j'}^t,\hat{\boldsymbol{x}}_{j'}^t\rangle. \tag{4}$$

Note that $\langle\hat{\boldsymbol{l}}_o^t,\hat{\boldsymbol{x}}_o\rangle=\langle\boldsymbol{l}^t,\boldsymbol{x}^t\rangle$. Define the *cumulative counterfactual loss* as $\hat{\boldsymbol{L}}_j^T=\sum_{t=1}^{T}\hat{\boldsymbol{l}}_j^t$. According to (4),

$$\hat{\boldsymbol{L}}_j^t[a]=\boldsymbol{L}^t[j,a]+\sum_{j'\in C_{j,a}}\left(\sum_{k=1}^{t}\langle\hat{\boldsymbol{l}}_{j'}^k,\hat{\boldsymbol{x}}_{j'}^k\rangle\right). \tag{5}$$

Besides, the *instantaneous counterfactual regret* is defined as $\hat{\boldsymbol{r}}_j^t=\langle\hat{\boldsymbol{l}}_j^t,\hat{\boldsymbol{x}}_j^t\rangle\mathbf{1}-\hat{\boldsymbol{l}}_j^t$, where $\mathbf{1}$ is an all-ones vector. The *cumulative counterfactual regret* is defined as

$$\hat{\boldsymbol{R}}_j^t=\sum_{k=1}^{t}\hat{\boldsymbol{r}}_j^t=\left(\sum_{k=1}^{t}\langle\hat{\boldsymbol{l}}_j^k,\hat{\boldsymbol{x}}_j^k\rangle\right)\mathbf{1}-\hat{\boldsymbol{L}}_j^t. \tag{6}$$

Let $\hat{R}_j^T=\max_{a\in A_j}\hat{\boldsymbol{R}}_j^T[a]$. The point of CFR is that $R^T\le\sum_{j\in J}[\hat{R}_j^T]^+$, where $[\cdot]^+=\max\{\cdot,0\}$ (Zinkevich et al., 2007). Accordingly, CFR instantiates a local regret minimizer to minimize $\hat{R}_j^T$ for each decision point $j\in\mathcal{J}$. A loss $\boldsymbol{l}^t$ received by a CFR algorithm is processed as follows: (i) the loss is decomposed into $\{\hat{\boldsymbol{l}}_j\}_{j\in J}$; (ii) each local minimizer observes the corresponding loss $\hat{\boldsymbol{l}}_j^t$ and returns the next local decision $\hat{\boldsymbol{x}}_j^{t+1}\in\Delta^{n_j}$; (iii) constructs $\boldsymbol{x}^{t+1}$ according to the local decisions.

Many local minimizers can be used, for example, RM and RM+, resulting in CFR-RM and CFR-RM+, respectively. The updates are summarized in Table 1, which can be plugged into Algorithm 2. As we can see, instead of tracking $\hat{\boldsymbol{R}}_j^t$, RM+ tracks a special *truncated cumulative counterfactual regret* $\hat{\boldsymbol{Q}}_j^t$ at every decision point. Note

Table 1: The local updates of the algorithms, can be plugged into Algorithm 2.

Algorithm	CFR-RM	CFR-RM+
LocalUpdate	$\hat{\boldsymbol{l}}_j^t[a] \leftarrow \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \langle \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle$, $\hat{\boldsymbol{R}}_j^t \leftarrow \hat{\boldsymbol{R}}_j^{t-1} + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle \mathbf{1} - \hat{\boldsymbol{l}}_j^t$, $\hat{\boldsymbol{x}}_j^{t+1} \leftarrow [\hat{\boldsymbol{R}}_j^t]^+ / \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$. RM	$\hat{\boldsymbol{l}}_j^t[a] \leftarrow \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \langle \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle$, $\hat{\boldsymbol{Q}}_j^t \leftarrow [\hat{\boldsymbol{Q}}_j^{t-1} + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle \mathbf{1} - \hat{\boldsymbol{l}}_j^t]^+$, $\hat{\boldsymbol{x}}_j^{t+1} \leftarrow \hat{\boldsymbol{Q}}_j^t / \|\hat{\boldsymbol{Q}}_j^t\|_1$. RM+
Algorithm	(FD-)FTRL	(FD-)OMD
LocalUpdate	$\hat{\boldsymbol{L}}_j^{0:t}[a] \leftarrow \boldsymbol{L}^t[j,a] + \sum_{j' \in C_{j,a}} -\psi_{j'}^t(-\hat{\boldsymbol{L}}_{j'}^{0:t})$, $\hat{\boldsymbol{x}}_j^{t+1} \leftarrow \nabla \psi_j^t(-\hat{\boldsymbol{L}}_j^{0:t})$.	$\hat{\boldsymbol{l}}_j^{0:t}[a] \leftarrow \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \hat{l}_{j'}^{0:t}$, where $\hat{l}_{j'}^{0:t} \leftarrow \psi_{j'}^{t-1}(\nabla \psi_{j'}^{t-1}(\hat{\boldsymbol{x}}_{j'}^t)) - \psi_{j'}^t(\nabla \psi_{j'}^{t-1}(\hat{\boldsymbol{x}}_{j'}^t) - \hat{\boldsymbol{l}}_{j'}^{0:t})$, $\hat{\boldsymbol{x}}_j^{t+1} \leftarrow \nabla \psi_j^t(\nabla \psi_j^{t-1}(\hat{\boldsymbol{x}}_j^t) - \hat{\boldsymbol{l}}_j^{0:t})$.
Algorithm	FD-FTRL(R)	FD-FTRL(R)
LocalUpdate	$\hat{\boldsymbol{L}}_j^{0:t}[a] \leftarrow \boldsymbol{L}^t[j,a] + \sum_{j' \in C_{j,a}} \alpha_{j'}^t$, solve α_j^t w.r.t. (13), $\hat{\boldsymbol{R}}_j^{0:t} \leftarrow \alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^{0:t}$, $\hat{\boldsymbol{x}}_j^{t+1} \leftarrow [\hat{\boldsymbol{R}}_j^{0:t}]^+ / \|[\hat{\boldsymbol{R}}_j^{0:t}]^+\|_1$.	$\hat{\boldsymbol{l}}_j^{0:t}[a] \leftarrow \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \alpha_{j'}^t$, solve α_j^t w.r.t. (14), $\hat{\boldsymbol{Q}}_j^{0:t} \leftarrow [\hat{\boldsymbol{Q}}_j^{0:t-1} + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^{0:t}]^+$, $\hat{\boldsymbol{x}}_j^{t+1} \leftarrow \hat{\boldsymbol{Q}}_j^{0:t} / \|\hat{\boldsymbol{Q}}_j^{0:t}\|_1$.

that $\hat{\boldsymbol{R}}_j^t \le \hat{\boldsymbol{Q}}_j^t$, which means $\hat{\boldsymbol{R}}_j^t$ will also been minimized when minimizing $\hat{\boldsymbol{Q}}_j^t$. Intuitively, RM+ is more sensitive to positive instantaneous regrets and thus can minimize the cumulative regrets faster. We assume that

Assumption 2.1. $\|[\hat{\boldsymbol{R}}_j^t]^+\|_1 > 0$ and $\|\hat{\boldsymbol{Q}}_j^t\|_1 > 0, \forall j \in \mathcal{J}$ and $t > 0$.

This assumption is needed for the equivalence analysis below. It can be satisfied by initializing $\hat{\boldsymbol{R}}_j^0$ and $\hat{\boldsymbol{Q}}_j^0$ to a small value $\epsilon \mathbf{1} > 0$.[1] A discussion is given in Appendix C.

Finally, there is a useful lemma provided in (Liu et al., 2022), as shown in Lemma 2.2. The proof is given in Appendix C. As a result, the famous regret bound $R^T \le \sum_{j \in \mathcal{J}} [\hat{R}_j^T]^+$ is immediately recovered as $\sum_{t=1}^T \{\langle \boldsymbol{l}^t, \boldsymbol{x}^t \rangle - \langle \boldsymbol{l}^t, \boldsymbol{x}^0 \rangle\} = \sum_{j \in \mathcal{J}} \boldsymbol{x}^0[p_j] \left\langle \sum_{t=1}^T \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^0 \right\rangle$.

Lemma 2.2. *(Liu et al., 2022) For any $\boldsymbol{x}, \boldsymbol{x}^0 \in \mathcal{X}$ and loss $\boldsymbol{l} \in \mathbb{R}^{\sum_{j \in \mathcal{J}} n_j}$, let $\hat{\boldsymbol{r}}_j$ be the instantaneous regret at decision point j under strategy $\boldsymbol{x}$ and loss $\boldsymbol{l}$, then, $\langle \boldsymbol{l}, \boldsymbol{x} \rangle - \langle \boldsymbol{l}, \boldsymbol{x}^0 \rangle = \sum_{j \in \mathcal{J}} \boldsymbol{x}^0[p_j] \langle \hat{\boldsymbol{r}}_j, \hat{\boldsymbol{x}}_j^0 \rangle$.*

2.4. Online Convex Optimization (OCO)

The total regret can also be minimized using OCO algorithms, e.g., FTRL and OMD. There are many variants of FTRL and OMD (McMahan & Streeter, 2010; Rakhlin & Sridharan, 2013). In this paper, we follow the generalized definitions of FTRL and OMD in (Joulani et al., 2020).[2] Formally, for any $\boldsymbol{x} \in \mathcal{X}$, define the regularizer at iteration t as $q^{0:t}(\boldsymbol{x}) = \sum_{k=0}^t q^k(\boldsymbol{x})$, where $q^k : D \to \mathbb{R}, \mathcal{X} \subseteq D$. Assume $q^{0:t}(\boldsymbol{x})$ is differentiable and strictly convex on $\mathcal{X}$. FTRL updates the strategy according to $\boldsymbol{x}^1 = \operatorname{argmin}_{\boldsymbol{x} \in \mathcal{X}} q^0(\boldsymbol{x})$ and

$$\boldsymbol{x}^{t+1} = \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \left\{ \langle \boldsymbol{L}^t, \boldsymbol{x} \rangle + q^{0:t}(\boldsymbol{x}) \right\}, \tag{7}$$

where $\boldsymbol{L}^t = \sum_{k=1}^t \boldsymbol{l}^k$ is the cumulative loss. OMD updates the strategy according to $\boldsymbol{x}^1 = \operatorname{argmin}_{\boldsymbol{x} \in \mathcal{X}} q^0(\boldsymbol{x})$ and

$$\boldsymbol{x}^{t+1} = \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \left\{ \langle \boldsymbol{l}^t, \boldsymbol{x} \rangle + q^t(\boldsymbol{x}) + \mathcal{B}_{q^{0:t-1}}(\boldsymbol{x} \| \boldsymbol{x}^t) \right\}, \tag{8}$$

where $\mathcal{B}_{q^{0:t-1}}$ is the Bregman divergence of function $q^{0:t-1}$. Formally, $\mathcal{B}_{q^{0:t}}(\boldsymbol{x} \| \boldsymbol{x}^0) = q^{0:t}(\boldsymbol{x}) - q^{0:t}(\boldsymbol{x}^0) - \left\langle \nabla q^{0:t}(\boldsymbol{x}^0), \boldsymbol{x} - \boldsymbol{x}^0 \right\rangle$ for any $\boldsymbol{x}, \boldsymbol{x}^0 \in \mathcal{X}$. Relatively, $q^{0:t}$ is called the Distance-Generating Function (DGF).

To solve (7) and (8), one needs to compute the gradient $\nabla q^{0:t}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X}$ and the gradient of the convex conjugate $q^{0:t}$: $\nabla q^{0:t}(\boldsymbol{g}) = \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}} \left\{ \langle \boldsymbol{g}, \boldsymbol{x} \rangle - q^{0:t}(\boldsymbol{x}) \right\}$ for any $\boldsymbol{g} \in \mathbb{R}^{\sum_j n_j}$. Then, (7) has a solution $\hat{\boldsymbol{x}}^{t+1} = \nabla q^{0:t}(-\boldsymbol{L}^t)$; and (8) has $\boldsymbol{x}^{t+1} = \nabla q^{0:t}(\nabla q^{0:t-1}(\boldsymbol{x}^t) - \boldsymbol{l}^t)$ (Orabona, 2019). Both FTRL and OMD can achieve a regret bound of $O(\sqrt{T})$ when the regularizers and the parameters are configured properly (Joulani et al., 2020).

2.5. Dilated DGF and Localized FTRL (OMD)

To apply FTRL and OMD to EFGs, dilated DGF (Hoda et al., 2010) is proposed to serve as the regularizer. For any sequence-form strategy $\boldsymbol{x} \in \mathcal{X}$, a dilated DGF is defined as $d(\boldsymbol{x}) = \sum_{j \in \mathcal{J}} \boldsymbol{x}[p_j] \psi_j(\hat{\boldsymbol{x}}_j)$, where $\hat{\boldsymbol{x}}_j = \boldsymbol{x}[j] / \boldsymbol{x}[p_j] \in \Delta^{n_j}$ and $\psi_j : E \to \mathbb{R}, \Delta^{n_j} \subseteq E$ is a local DGF. Assume ψ_j is differentiable and strictly convex on Δ^{n_j}. Note that d is strictly convex as long as ψ_j is strictly convex at all

[1] This has been adopted in OpenSpiel (Lanctot et al., 2019).

[2] In (Joulani et al., 2020), they are named Ada-FTRL and Ada-OMD, respectively. Besides, we ignore the "proximal" regularizer.

$j \in \mathcal{J}$ (Hoda et al., 2010). Examples of dilated DGFs can be found in (Kroer et al., 2020) and (Farina et al., 2019b).

Let the regularizer $q^{0:t}$ at iteration t be a dilated DGF:

$$q^{0:t}(\boldsymbol{x}) = \sum_{j \in \mathcal{J}} \boldsymbol{x}[p_j]\psi_j^t(\hat{\boldsymbol{x}}_j), \tag{9}$$

where $\psi_j^t(\hat{\boldsymbol{x}}_j)$ is differentiable and strictly convex on Δ^{n_j}. It is known that the updates of FTRL and OMD can be decomposed into local updates (Hoda et al., 2010; Farina et al., 2019b). In this paper, we introduce notations $\hat{\boldsymbol{L}}_j^{0t}$ and $\hat{\boldsymbol{l}}_j^{0t}$ to denote the *local losses* in FTRL and OMD, respectively, as shown in Proposition 2.3 and 2.4.

Proposition 2.3. *The update of FTRL in (7) with $q^{0:t}(\boldsymbol{x})$ being a dilated DGF defined in (9) can be decomposed as $\hat{\boldsymbol{x}}_j^{t+1} = \nabla\psi_j{}^t(-\hat{\boldsymbol{L}}_j^{0t})$, $j \in \mathcal{J}$, where*

$$\hat{\boldsymbol{L}}_j^{0t}[a] = \boldsymbol{L}^t[j,a] + \sum_{j' \in C_{j,a}} -\psi_{j'}{}^t(-\hat{\boldsymbol{L}}_{j'}^{0t}). \tag{10}$$

Proposition 2.4. *The update of OMD in (8) with $q^{0:t}(\boldsymbol{x})$ being a dilated DGF defined in (9) can be decomposed as $\hat{\boldsymbol{x}}_j^{t+1} = \nabla\psi_j{}^t(\nabla\psi_j^t{}^{-1}(\hat{\boldsymbol{x}}_j^t) - \hat{\boldsymbol{l}}_j^{0t})$, where*

$$\hat{\boldsymbol{l}}_j^{0t}[a] = \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \hat{l}_{j'}^{0t}, \tag{11}$$

and $\hat{l}_{j'}^{0t} = \psi_{j'}{}^{t\ -1}(\nabla\psi_{j'}^t{}^{-1}(\hat{\boldsymbol{x}}_{j'}^t)) - \psi_{j'}{}^t(\nabla\psi_{j'}^t{}^{-1}(\hat{\boldsymbol{x}}_{j'}^t) - \hat{\boldsymbol{l}}_{j'}^{0t})$.

In the propositions, $\psi_j{}^t$ is the convex conjugate: $\psi_j{}^t(\hat{\boldsymbol{g}}) = \max_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \{\langle \hat{\boldsymbol{g}}, \hat{\boldsymbol{x}}_j\rangle - \psi_j^t(\hat{\boldsymbol{x}}_j)\}$ for any $\hat{\boldsymbol{g}} \in \mathbb{R}^{n_j}$, and $\nabla\psi_j{}^t(\hat{\boldsymbol{g}})$ is the gradient: $\nabla\psi_j{}^t(\hat{\boldsymbol{g}}) = \operatorname{argmax}_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \{\langle \hat{\boldsymbol{g}}, \hat{\boldsymbol{x}}_j\rangle - \psi_j^t(\hat{\boldsymbol{x}}_j)\}$. The propositions mainly leverage the recursive nature of the sequence-form strategy space and the recursive property of the dilated DGF. The proofs are given in Appendix D. According to these two propositions, FTRL and OMD algorithms can be formulated into localized forms, as shown in Table 1.

3. Equivalence Analysis and Its Application

By comparing the updates of CFR-RM (CFR-RM+) and FTRL (OMD) in Table 1, we can see that these two algorithms have similar recursive structures. To examine whether they are equivalent under certain settings, we first give an example of FTRL and analyze the equivalence at terminal decision points. In this paper, we only consider the case where the regularizer is a dilated Euclidean DGF with $\psi_j^t(\hat{\boldsymbol{x}}_j) = \frac{1}{2}\beta_j^t\|\hat{\boldsymbol{x}}_j\|_2^2 + C_j^t, \beta_j^t > 0, C_j^t \in \mathbb{R}$.

Example 3.1. In FTRL with a dilated Euclidean regularizer where $\psi_j^t(\hat{\boldsymbol{x}}_j) = \frac{1}{2}\beta_j^t\|\hat{\boldsymbol{x}}_j\|_2^2, \beta_j^t > 0$, we have $\hat{\boldsymbol{x}}_j^{t+1} = \nabla\psi_j{}^t(-\hat{\boldsymbol{L}}_j^{0t}) = [\alpha_j^t\mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+/\beta_j^t$ and $-\psi_j{}^t(-\hat{\boldsymbol{L}}_j^{0t}) = \alpha_j^t - \frac{1}{2}\beta_j^t\|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2$, where $\alpha_j^t \in \mathbb{R}$ satisfies $\|\hat{\boldsymbol{x}}_j^{t+1}\|_1 = 1$.

Example 3.1 shows that $\hat{\boldsymbol{x}}_j^{t+1} = [\alpha_j^t\mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+/\beta_j^t$ when $C = 0$, which is similar to the updating rule of RM ($\hat{\boldsymbol{x}}_j^{t+1} = [\hat{\boldsymbol{R}}_j^t]^+/\|[\hat{\boldsymbol{R}}_j^t]^+\|_1 = [\sum_{k=1}^t\langle\hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle\mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+/\|[\hat{\boldsymbol{R}}_j^t]^+\|_1$). In the example, α_j^t is arisen to constrain the solution to a simplex. Note that the α_j^t that fulfills $\|\hat{\boldsymbol{x}}_j^{t+1}\|_1 = 1$, i.e., $\|[\alpha_j^t\mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+\|_1 = \beta_j^t$, exists and is unique when $\beta_j^t > 0$.

Now, Let us focus at terminal decision points where $C_{j,a} = \emptyset$. At such a point j, we have both $\hat{\boldsymbol{L}}_j^{0t}$ in FTRL and $\hat{\boldsymbol{L}}_j^t$ in CFR-RM equal $\boldsymbol{L}^t[j]$. What is more, when β_j^t in Example 3.1 equals $\|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ in CFR-RM, we have $\alpha_j^t = \sum_{k=1}^t\langle\hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle$ since $\|[\alpha_j^t\mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+\|_1 = \beta_j^t$ (recall that α_j^t exists and is unique) in FTRL and $\|[\sum_{k=1}^t\langle\hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle\mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+\|_1 = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ in CFR-RM. This means that FTRL and CFR-RM have the same local decision at terminal decision point j when $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$. The same conclusion can also be found in (Waugh & Bagnell, 2015). However, this is not true at non-terminal decision points because it is *not* guaranteed that $\hat{\boldsymbol{L}}_j^{0t} = \hat{\boldsymbol{L}}_j^t$ at such a point j.

By comparing the local updates of CFR-RM and FTRL in (5) and (10), we can see that $\hat{\boldsymbol{L}}_j^{0t}$ will be equal to $\hat{\boldsymbol{L}}_j^t$ at any decision point j if $-\psi_{j'}{}^t(-\hat{\boldsymbol{L}}_{j'}^{0t}) = \sum_{k=1}^t\langle\hat{\boldsymbol{l}}_{j'}^k, \hat{\boldsymbol{x}}_{j'}^k\rangle$ at all its children. However, as shown in Example 3.1, when $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$, $-\psi_j{}^t(-\hat{\boldsymbol{L}}_j^{0t}) \neq \sum_{k=1}^t\langle\hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle$ even at terminal decision point, although we have $\alpha_j^t = \sum_{k=1}^t\langle\hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle$. But can we design an FTRL with a special regularizer such that $-\psi_j{}^t(-\hat{\boldsymbol{L}}_j^{0t}) = \alpha_j^t = \sum_{k=1}^t\langle\hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle$ at every decision point? With this question, we propose FD-FTRL and FD-OMD and then we prove that CFR-RM and CFR-RM+ are equivalent to special cases of them, respectively.

3.1. Future-Dependent FTRL (OMD): the Bridge

In this subsection, we propose FD-FTRL and FD-OMD. As shown in Definition 3.2, FD-FTRL (FD-OMD) have a regularizer, called FD regularizer, at every iteration that depends on the next iteration strategy.

Definition 3.2. FD-FTRL (FD-OMD) is an FTRL (OMD) with $q^{0:t}(\boldsymbol{x})$ being a dilated DGF defined in (9) and $\psi_j^t(\hat{\boldsymbol{x}}_j) = \frac{1}{2}\beta_j^t\|\hat{\boldsymbol{x}}_j\|_2^2 + \frac{1}{2}\beta_j^t\|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2, \beta_j^t > 0, \forall j \in \mathcal{J}$.

Remark 3.3. In FD-FTRL (FD-OMD), according to Proposition 2.3 (2.4), $\hat{\boldsymbol{x}}_j^{t+1}$ does not depend on itself. So, $\hat{\boldsymbol{x}}_j^{t+1}$ can be solved and then $\boldsymbol{x}^{t+1}$ can be constructed bottom-up.

Similar to Example 3.1, we have Example 3.4. Also, FD-OMD has $\hat{\boldsymbol{x}}_j^{t+1} = [\beta_j^t{}^{-1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t\mathbf{1} - \hat{\boldsymbol{l}}_j^{0t}]^+/\beta_j^t$ and $\hat{l}_j^{0t} = \alpha_j^t$ where $\alpha_j^t \in \mathbb{R}$ satisfies $\|\hat{\boldsymbol{x}}_j^{t+1}\|_1 = 1$.

Example 3.4. In FD-FTRL, we have $\hat{\boldsymbol{x}}_j^{t+1} = \nabla\psi_j{}^t(-\hat{\boldsymbol{L}}_j^{0t}) = [\alpha_j^t\mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+/\beta_j^t$ and $-\psi_j{}^t(-\hat{\boldsymbol{L}}_j^{0t}) = \alpha_j^t$, where $\alpha_j^t \in \mathbb{R}$ satisfies $\|\hat{\boldsymbol{x}}_j^{t+1}\|_1 = 1$.

As we can see, the only difference between Example 3.1 and Example 3.4 is that the latter one drops the $-\frac{1}{2}\beta_j^t\|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2$ term in $-{\psi_j}^t(-\hat{\boldsymbol{L}}_j^{0t})$, which is canceled out by the term in the regularizer. From a global viewpoint, FD-FTRL has an extra linear regularizer (or an extra loss) $\langle \boldsymbol{m}, \boldsymbol{x}\rangle$, where $\boldsymbol{m}[j,a] = \sum_{j^0 2 C_{j,a}} \frac{1}{2}\beta_{j^0}^t\|\hat{\boldsymbol{x}}_{j^0}^{t+1}\|_2^2$, than the FTRL in Example 3.1. Therefore, FD-FTRL has more preference for reaching decision points that have lower l_2 norm. In other words, FD-FTRL is more conservative than the FTRL in Example 3.1. Nevertheless, the FD regularizer is still a dilated Euclidean DGF, and FD-FTRL (FD-OMD) belongs to FTRL (OMD) family. Leveraging multiple analysis techniques for FTRL (OMD), the convergence of FD-FTRL (FD-OMD) is provable, as shown in Theorem 3.5.

Theorem 3.5. *The total regret R^T of FD-FTRL (FD-OMD) after T iterations is bounded by*

$$\sum_{j 2 J}\left(\beta_j^T + \sum_{t=1}^{T}\frac{\left[\|\hat{\boldsymbol{r}}_j^t\|_2^2 + \|\beta_j^t\ {}^1\hat{\boldsymbol{x}}_j^t\|_2^2 - \|\beta_j^t\hat{\boldsymbol{x}}_j^{t+1}\|_2^2\right]^+}{2\beta_j^t}\right).$$

The proof for Theorem 3.5 is given in Appendix E. The theorem is mainly based on Theorem 3 in (Joulani et al., 2020) and Lemma 2.2 in this paper. We assume

Assumption 3.6. $\|\beta_j^t\ {}^1\hat{\boldsymbol{x}}_j^t\|_2^2 \le \|\beta_j^t\hat{\boldsymbol{x}}_j^{t+1}\|_2^2$ for all $j \in \mathcal{J}$.

This assumption is easy to be satisfied after reparameterizing β_j^t, will be discussed later. If the assumption is true, we have $R^T \le \sum_{j 2 J}\left(\beta_j^T + \sum_{t=1}^T \|\hat{\boldsymbol{r}}_j^t\|_2^2/(2\beta_j^t)\right)$. In this case, there are many ways to bound the total regret of FD-FTRL (FD-OMD) by $O(\sqrt{T})$ (Note that $\|\hat{\boldsymbol{r}}_j^t\|_2^2$ is bounded). For example, we can set $\beta_j^t = \Theta(\sqrt{t})$, or $\beta_j^t = \Theta(\sqrt{T})$ if T is known. Otherwise, we can set $\beta_j^t = \sqrt{\sum_{k=1}^t \|\hat{\boldsymbol{r}}_j^k\|_2^2}$ to obtain $R^T \le 2\sum_{j 2 J}\sqrt{\sum_{k=1}^t \|\hat{\boldsymbol{r}}_j^k\|_2^2}$. The last setting is known as *adaptation*. The first adaptive FTRL was proposed in (McMahan & Streeter, 2010). Usually, adaptive regret minimization algorithms have better regret bounds.

3.2. Equivalence Theorem

In this subsection, we prove the equivalence between FD-FTRL (FD-OMD) and CFR-RM (CFR-RM+). According to Example 3.4, it is natural to conjecture that the α_j^t in FD-FTRL is equal to $\sum_{k=1}^t\langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle$ at every decision point when $\beta_j^t = \|\hat{\boldsymbol{R}}_j^t\|_1$. In other words, we may have $\hat{\boldsymbol{x}}_j^{t+1} = [\sum_{k=1}^t\langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k\rangle\mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+/\|\hat{\boldsymbol{R}}_j^t\|_1$ and $\hat{\boldsymbol{L}}_j^{0t} = \hat{\boldsymbol{L}}_j^t$, which are exactly the local updates of CFR-RM. It turns out that this conjecture can be proven recursively.

Theorem 3.7. *CFR-RM (CFR-RM+) is equivalent to a special case of FD-FTRL (FD-OMD) with $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ ($\|\hat{\boldsymbol{Q}}_j^t\|_1$), $\forall j \in \mathcal{J}, t \ge 0$.*

The proof is given in Appendix E. We name the FD-FTRL with $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ as **FD-FTRL(CFR)** and the FD-OMD with $\beta_j^t = \|\hat{\boldsymbol{Q}}_j^t\|_1$ as **FD-OMD(CFR)**. From different perspectives, the equivalences indicate that:

- The cumulative counterfactual regrets in CFR-RM (CFR-RM+) can be viewed as adaptive regularization parameters: the greater $\|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ ($\|\hat{\boldsymbol{Q}}_j^t\|_1$), the stronger the regularization. This is intuitive as the decisions should be more conservative when the past strategies have failed in controlling the regrets.
- CFR-RM (CFR-RM+) is an excellent adaptive FTRL (OMD). As mentioned before, we can guarantee $R^T \le 2\sum_j\sqrt{\sum_{k=1}^t\|\hat{\boldsymbol{r}}_j^k\|_2^2}$ in a case. However, when $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ ($\|\hat{\boldsymbol{Q}}_j^t\|_1$), CFR-RM (CFR-RM+) is recovered, and we have $R^T \le \sum_j \max_a[\hat{\boldsymbol{R}}_j^T]^+$,[3] which is an even better regret bound as $\max_a[\hat{\boldsymbol{R}}_j^T]^+ \le \sqrt{\sum_{k=1}^t\|\hat{\boldsymbol{r}}_j^k\|_2^2}$ (Zinkevich et al., 2007). Since adaptive FTRL (OMD) can adapt to the losses, they are generally faster than the non-adaptive versions. This may partially explain the superior performance of CFRs.
- FTRL and OMD can perform as well as CFRs in practice if the parameters are configured properly. While setting β_j^t to $\|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ or $\|\hat{\boldsymbol{Q}}_j^t\|_1$ is unpractical (as this requires running a CFR-RM or a CFR-RM+ in parallel), let $\beta_j^t \approx \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ or $\|\hat{\boldsymbol{Q}}_j^t\|_1$ may still result in fast algorithms. Note that this intuition is not restricted to FD-FTRL and FD-OMD, and may apply to general FTRL and OMD with Euclidean regularizers.

3.3. Practical Implementation of FD-FTRL (FD-OMD)

As we have mentioned, the total regret of FD-FTRL is bounded by $O(\sqrt{T})$ after T iterations if 1) Assumption 3.6 is fulfilled; and 2) $\beta_j^t = \Theta(\sqrt{t})$ (or $\Theta(\sqrt{T})$). However, Assumption 3.6 is non-trivial, as it depends on the future decision $\hat{\boldsymbol{x}}_j^{t+1}$ (we need to set β_j^t before we compute $\hat{\boldsymbol{x}}_j^{t+1}$). Alternatively, we introduce a new parameter $\lambda_j^t > 0$ and reparameterize β_j^t as $\sqrt{\lambda_j^t}/\|\hat{\boldsymbol{x}}_j^{t+1}\|_2$. Consequently, Assumption 3.6 can be reformulated as

Assumption 3.8. $\lambda_j^t\ {}^1 \le \lambda_j^t$ for all $j \in \mathcal{J}$.

FD-FTRL. Recall that in FD-FTRL, the next decision is computed according to $\hat{\boldsymbol{x}}_j^{t+1} = [\alpha_j^t\mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+/\beta_j^t$, where

$$\alpha_j^t \in \mathsf{R},\ s.t.\ \|[\alpha_j^t\mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+\|_1 = \beta_j^t, \tag{12}$$

and $-{\psi_j}^t(-\hat{\boldsymbol{L}}_j^{0t}) = \alpha_j^t$. Since $\beta_j^t = \sqrt{\lambda_j^t}/\|\hat{\boldsymbol{x}}_j^{t+1}\|_2$, the

[3] Can be deduced from Lemma 2.2, but not from Theorem 3.5.

constraint becomes

$$\|[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}]^+\|_2^2 = \lambda_j^t. \tag{13}$$

Note that the α_j^t in (13) exists and is unique when $\lambda_j^t > 0$, so it must equal the α_j^t in (12). Based on the above analysis, we can implement an FD-FTRL algorithm that computes the α_j^t at every decision point with respect to (13), which is then can be used to replace $-{\psi_j}^t(-\hat{\boldsymbol{L}}_j^{0t})$ and compute the next strategy $\hat{\boldsymbol{x}}_j^{t+1}$. Let $\hat{\boldsymbol{R}}_j^{0t} = \alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}$. The updates are summarized in Table 1, and the algorithm is named **FD-FTRL(R)**. According to Theorem 3.7, when $\lambda_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_2^2$, which is equivalent to $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$, FD-FTRL(R) is equivalent to CFR-RM.

FD-OMD. Recall that FD-OMD has a constraint $\|\hat{\boldsymbol{x}}_j^{t+1}\|_1 = 1$, i.e., $\|[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^{0t}]^+\|_1 = \beta_j^t$. Since $\beta_j^t = \sqrt{\lambda_j^t}/\|\hat{\boldsymbol{x}}_j^{t+1}\|_2$, the constraint is equivalent to

$$\left\| \left[\sqrt{\lambda_j^{t\ 1}} \hat{\boldsymbol{x}}_j^t / \|\hat{\boldsymbol{x}}_j^t\|_2 + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^{0t} \right]^+ \right\|_2^2 = \lambda_j^t. \tag{14}$$

Note that the α_j^t in (14) exists and is unique when $\lambda_j^t > 0$. Let $\hat{\boldsymbol{Q}}_j^{0t} = \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} = [\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^{0t}]^+$. The updates with respect to (14) is summarized in Table 1, and the algorithm is named **FD-OMD(R)**. When $\lambda_j^t = \|\hat{\boldsymbol{Q}}_j^t\|_2^2$, which is equivalent to $\beta_j^t = \|\hat{\boldsymbol{Q}}_j^t\|_1$, according to Theorem 3.7, we have FD-OMD(R) equivalent to CFR-RM+.

Notably, FD-FTRL(R) recovers an algorithm named ReCFR (Liu et al., 2022), which is inspired by a warm starting CFR algorithm (Brown & Sandholm, 2016). So, ReCFR is also a special case of FD-FTRL.

As we mentioned before, we may let $\beta_j^t \approx \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ or $\|\hat{\boldsymbol{Q}}_j^t\|_1$ to construct a fast FD-FTRL or FD-OMD. Notice that $\|[\hat{\boldsymbol{R}}_j^t]^+\|_2^2$ ($\|\hat{\boldsymbol{Q}}_j^t\|_2^2$) $\leq \sum_{k=1}^t \|\hat{\boldsymbol{r}}_j^k\|_2^2$ (Zinkevich et al., 2007; Tammelin et al., 2015), we proposed to set λ_j^t such that $\lambda_j^t \geq \eta \sum_{k=1}^t \|\hat{\boldsymbol{r}}_j^k\|_2^2$, $\eta \in (0, \infty)$.

Corollary 3.9. *If $\eta \sum_{k=1}^t \|\hat{\boldsymbol{r}}_j^k\|_2^2 \leq \lambda_j^t$ and $\lambda_j^{t\ 1} \leq \lambda_j^t$, $\forall j \in \mathcal{J}$, $t > 0$, then, the total regret of FD-FTRL(R) (FD-OMD(R)) after T iterations is $R^T \leq \sum_{j2J} \left(\sqrt{n_j} + \frac{1}{\eta}\right) \sqrt{\lambda_j^T}$.*

The proof is given in Appendix E. Clearly, the total regret of FD-FTRL(R) (FD-OMD(R)) is $O(\sqrt{T})$ if $\lambda_j^t = O(t)$ or $O(T)$, which is achievable as $\|\hat{\boldsymbol{r}}_j^k\|_2^2 = O(\|\mathbf{A}\|_{1}^2)$. Both FD-FTRL(R) and FD-OMD(R) have the same space complexity as vanilla CFR. However, FD-FTRL(R) and FD-OMD(R) need to solve a piecewise constraint at every decision point, which has a time complexity of $O(n_j^2)$. If binary search and Quicksort are used, an $O(n_j \log n_j)$ complexity can be obtained. So, FD-FTRL(R) and FD-OMD(R) are $|\mathcal{A}|$ or $\log|\mathcal{A}|$ times more expansive than CFR.

To some extent, FD-FTRL(R) and FD-OMD(R) can be considered as the extensions of CFR-RM and CFR-RM+, respectively. However, they are able to leverage theoretical results in OCO, since they belong to FTRL and OMD families. Another advantage of them over CFRs is that they do not track the counterfactual regrets. Note that, FD-FTRL(R) does not track the cumulative loss in practice, too. This is because $\boldsymbol{L}^t = \sum_k^t A\boldsymbol{y}^k = tA\bar{\boldsymbol{y}}^t$ only depends on the average strategy of the opponent, which is available during the process. The disadvantage of FD-FTRL(R) and FD-OMD(R) is that they have a set of parameters. In the next section, we will try two configurations for these parameters.

4. Experimental Investigation

To better understand the properties of FD-FTRL(R) and FD-OMD(R), we test two different methods for setting the weighting parameters λ_j^t: 1) **Linear Weighting (LW)**: $\lambda_j^t = \eta \bar{y}_j^t n_j \|\mathbf{A}\|_{1}^2 t$; 2) **Constant Weighting (CW)**[4]: $\lambda_j^t = \eta \bar{y}_j^t n_j \|\mathbf{A}\|_{1}^2 T$. In the equations, y_j^t denotes the probability of reaching j of the opponent, and $\bar{y}_j^t$ is the average probability. $\eta \in (0, \infty)$ is a global hyper-parameter. These configurations are inspired by the observations that $\|\hat{\boldsymbol{r}}_j^t\|_2^2 = O(y_j^t n_j \|\mathbf{A}\|_{1}^2)$ and $\sum_{k=1}^t \|\hat{\boldsymbol{r}}_j^k\|_2^2 = O(\bar{y}_j^t n_j \|\mathbf{A}\|_{1}^2 t)$ (Brown & Sandholm, 2016). According to Corollary 3.9, FD-FTRL(R) and FD-OMD(R) with both weighting methods have a sub-linear regret bound $O(\sqrt{T})$. For computing the average strategy, we use: 1) **Uniform Averaging (UA)**: $\bar{\boldsymbol{x}}^T = \frac{1}{T}\sum_{t=1}^T \boldsymbol{x}^t$; 2) **Linear Averaging (LA)**: $\bar{\boldsymbol{x}}^T = \frac{2}{T(T+1)}\sum_{t=1}^T t\boldsymbol{x}^t$. FD-FTRL(R) and FD-OMD(R) use **CW** and **LA** by default.

The algorithms are compared with FTRL, OMD and CFRs. For a fair comparison, the competitors also use **LA** for computing the average strategies, namely Linear CFR (LCFR) (Brown & Sandholm, 2019b), CFR+, FTRL(LA), and OMD(LA). FTRL(LA) and OMD(LA)[5] are the algorithms with **LA** and a regularizer $q^{0:t}(\boldsymbol{x}) = q^0(\boldsymbol{x}) = \sum_{j2J} \boldsymbol{x}[p_j]\frac{1}{2}\beta_j \|\hat{\boldsymbol{x}}\|_2^2$ for $t > 0$. We set β_j for each point j in FTRL(LA) and OMD(LA) according to (Farina et al., 2019b): $\beta_j = 2\sigma + 2\max_{a2A_j} \sum_{j^0 2 C_{j,a}} \beta_{j^0}$, where σ is a hyper-parameter. In FD-FTRL(R) and FTRL(LA), the losses are also weighted linearly, as in LCFR. All the algorithms use alternating updates, as in CFRs.

We run a coarse grid search for tuning the hyper-parameter in each algorithm, and the best one will be used for the comparison. The tuning results of FD-FTRL(R) and FD-OMD(R) are given in Appendix G. We conduct our experiments in eight benchmark games, including Leduc (Southey et al., 2012) and FHP (Brown et al., 2019). A description of

[4] Since $\bar{y}_j^t$ changes slowly, we can consider it constant.

[5] A primary experiment shows that the **LA** versions of FTRL and OMD are significantly faster than the **UA** versions.

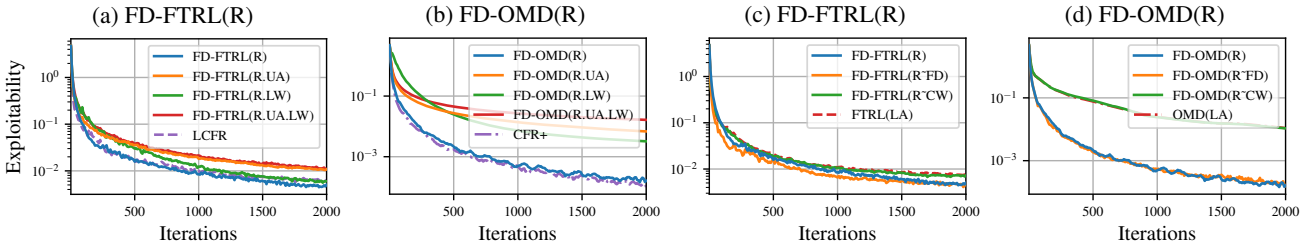

Figure 1: Exploitability curves of FD-FTRL(R) and FD-OMD(R) in different configurations in Leduc. The x-axis is the number of iterations. **(a, b)**: weighting methods and averaging methods. **(c, d)**: FD regularizer and **CW**. FD-FTRL(R˜FD) is an FD-FTRL(R) without FD regularizer, i.e., an FTRL(LA) with **CW**. FD-FTRL(R˜CW) is an FD-FTRL(R) without **CW**, i.e., an FTRL(LA) with FD regularizer. FD-OMD(R˜FD) and FD-OMD(R˜CW) are configured similarly.

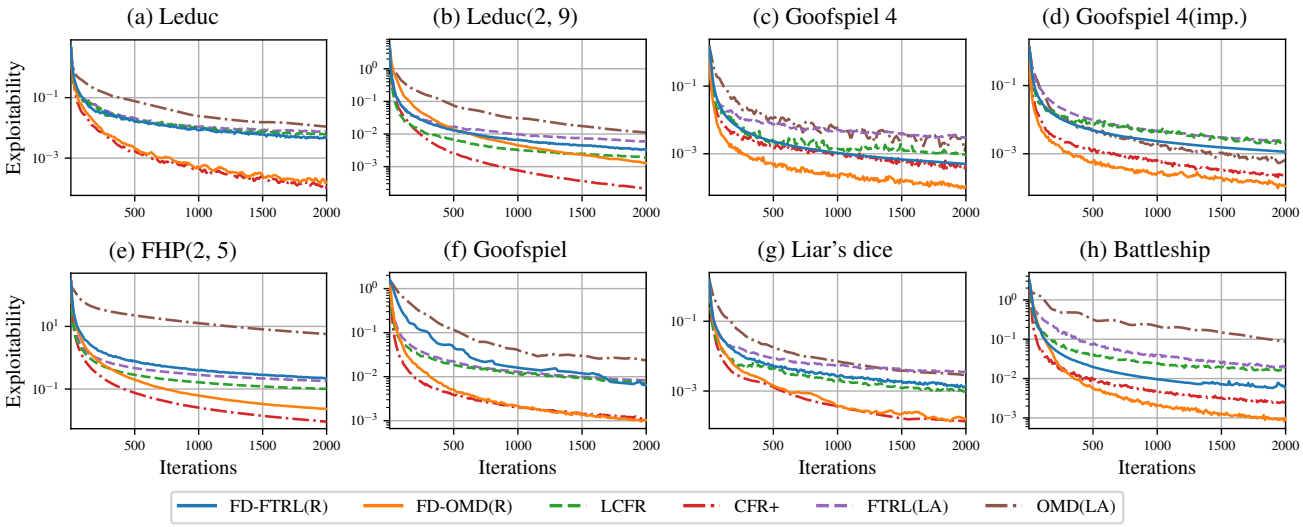

Figure 2: Exploitability curves of FD-FTRL(R), FD-OMD(R), and the competitors in eight games.

the games is given in Appendix F. The experiments use part of the code of project OpenSpiel (Lanctot et al., 2019).[6]

We first compare FD-FTRL(CFR) and FD-OMD(CFR) with vanilla CFR (i.e., CFR-RM) and CFR+ (i.e., CFR-RM+ with **LA**) in the benchmark games. The results are given in Appendix G, showing that the exploitability curves of FD-FTRL(CFR) (FD-OMD(CFR)) and vanilla CFR (CFR+) are overlapping with each other in every game. Therefore, the equivalences are also verified empirically. Then, we test FD-FTRL(R) and FD-OMD(R) in different configurations. The results in Leduc are given in Figure 1. In plot 1(a), the results show that both FD-FTRL(R.LW) and FD-FTRL(R.UA) are slower than the default FD-FTRL(R), which indicates that both **CW** and **LA** contribute to the performance of FD-FTRL(R). In plot 1(b), as we can see, the default FD-OMD(R) with **LA** and **CW** is much faster. This is not surprising, since it is also observed in CFR+ that **LA** can improve the performance significantly (Tammelin et al., 2015). However, the reason is still not clear. In plots 1(c, d), we perform an ablation study for FD-FTRL(R) and FD-OMD(R), respectively. The results show that **CW** has a significant impact on the performance, while FD regularizer has a much weaker effect. Since FD-FTRL(R˜FD) is essentially an FTRL(LA) with **CW**, the results indicate that **CW** can also apply to conventional FTRL algorithms. The results in the other games are reported in Appendix G, which are basically consistent with the results in Leduc. It is worth noting that FD-FTRL(R˜FD) is faster than FD-FTRL(R) in some benchmark games. This suggests that FD-FTRL(R) (as well as vanilla CFR and LCFR) may be too conservative in some games as it tends to keep the l_2 norms of the decisions low. However, FD-OMD(R) is always (one of) the fastest.

The results of FD-FTRL(R) and FD-OMD(R) in all the benchmark games are shown in Figure 2. As we can see, FD-FTRL(R) is tied with LCFR and FTRL(LA). However, the results show that FD-OMD(R) behaves more like a CFR instead of an OMD: it is always faster than LCFR and OMD(LA), and only slower than CFR+ in Leduc(2, 9)

[6] https://github.com/deepmind/open_spiel

and FHP(2, 5). Note that Leduc(2, 9) and FHP(2, 5) are the most stochastic among the benchmark games. So, it seems that the weighting methods (**CW** and **LW**) can not handle stochastic games well. We suspect FD-FTRL(R) and FD-OMD(R) may perform better in stochastic games when λ_j^t is set closer to $\|[\hat{\boldsymbol{R}}_j^t]^+\|_2^2$ and $\|\hat{\boldsymbol{Q}}_j^t\|_2^2$.

We also compare FD-FTRL(R) and FD-OMD(R) with PCFR and PCFR+ (Farina et al., 2021). The results are given in Appendix G, showing that FD-FTRL(R) and FD-OMD(R) are also competitive compared with them. Finally, we compare FD-FTRL(R) and FD-OMD(R) with FTRL, OMD, and vanilla CFR that use **UA** for computing the average strategies. The results are given in Appendix G, showing that both FD-FTRL(R) and FD-OMD(R) are always faster than them.

5. Conclusions and Future Work

In the paper, It is proven that CFR-RM and CFR-RM+ are equivalent to special cases of FTRL and OMD, respectively. The equivalences provide a new understanding of the counterfactual regrets in CFRs and may partially explain the superior performance of CFRs. As the bridges, FD-FTRL and FD-OMD have been proposed, and have been extensively tested in eight benchmark EFGs. The experimental results show that they are competitive compared with conventional FTRL, OMD, and CFRs. The results also suggest that FTRL and OMD are not necessarily slower than CFRs in EFGs. Therefore, more research is required in applying OCO algorithms to EFGs.

The equivalence analysis in this paper is relatively primitive, e.g., the analysis is restricted to RM, RM+, and Euclidean regularizers. Also, the analysis does not improve the theoretical results of CFR and CFR+. However, it can explain CFRs as an adaptive FTRL (OMD) and provide new ways to apply FTRL and OMD to EFGs. Besides, since there are optimistic variants of FTRL and OMD (Farina et al., 2019b) that converge at a rate of $O(1/T)$, we may also be able to develop new optimistic variants of CFRs.

In recent years, function-approximate CFRs (Waugh et al., 2015; Brown et al., 2019; Li et al., 2020; Liu et al., 2022) have been found to have problems in approximating cumulative counterfactual regrets. Combining FD-FTRL(R) or FD-OMD(R) with function approximation would not have these problems, since they do not rely on the cumulative counterfactual regrets. Furthermore, it is known that CFR-RM+ works much better with function approximation than CFR-RM (Morrill, 2016; D'Orazio, 2020). It would be interesting to see whether FD-OMD(R) is more favorable than FD-FTRL(R) in this setting.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61836011 and Grant U19B2044.

References

Abernethy, J. D., Hazan, E., and Rakhlin, A. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Conference on Learning Theory*, pp. 263–274, 2008.

Beck, A. and Teboulle, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

Brown, N. and Sandholm, T. Strategy-based warm starting for regret minimization in games. In *AAAI Conference on Artificial Intelligence*, pp. 432–438, 2016.

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

Brown, N. and Sandholm, T. Superhuman AI for multiplayer poker. *Science*, 365(6456):885–890, 2019a.

Brown, N. and Sandholm, T. Solving imperfect-information games via discounted regret minimization. In *AAAI Conference on Artificial Intelligence*, pp. 1829–1836, 2019b.

Brown, N., Lerer, A., Gross, S., and Sandholm, T. Deep counterfactual regret minimization. In *International Conference on Machine Learning*, volume 97, pp. 793–802, 2019.

D'Orazio, R. Regret minimization with function approximation in extensive-form games. Master's thesis, University of Alberta, 2020.

Farina, G., Kroer, C., Brown, N., and Sandholm, T. Stable-predictive optimistic counterfactual regret minimization. In *International Conference on Machine Learning*, volume 97, pp. 1853–1862, 2019a.

Farina, G., Kroer, C., and Sandholm, T. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. In *Advances in Neural Information Processing Systems*, pp. 5222–5232, 2019b.

Farina, G., Ling, C. K., Fang, F., and Sandholm, T. Correlation in extensive-form games: Saddle-point formulation and benchmarks. In *Advances in Neural Information Processing Systems*, pp. 9229–9239, 2019c.

Farina, G., Kroer, C., and Sandholm, T. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. In *AAAI Conference on Artificial Intelligence*, pp. 5363–5371, 2021.

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

Hoda, S., Gilpin, A., Peña, J., and Sandholm, T. Smoothing techniques for computing Nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2):494–512, 2010.

Joulani, P., György, A., and Szepesvári, C. A modular analysis of adaptive (non-)convex optimization: Optimism, composite objectives, variance reduction, and variational bounds. *Theoretical Computer Science*, 808:108–138, 2020.

Koller, D., Megiddo, N., and von Stengel, B. Efficient computation of equilibria for extensive two-person games. *Games and Economic Behavior*, 14(2):247–259, 1996.

Kroer, C., Waugh, K., Kilinç-Karzan, F., and Sandholm, T. Faster algorithms for extensive-form game solving via improved smoothing functions. *Mathematical Programming*, 179(1):385–417, 2020.

Lanctot, M., Waugh, K., Zinkevich, M., and Bowling, M. H. Monte Carlo sampling for regret minimization in extensive games. In *Advances in Neural Information Processing Systems*, pp. 1078–1086, 2009.

Lanctot, M., Lockhart, E., Lespiau, J., Zambaldi, V. F., Upadhyay, S., Pérolat, J., Srinivasan, S., Timbers, F., Tuyls, K., Omidshafiei, S., Hennes, D., Morrill, D., Muller, P., Ewalds, T., Faulkner, R., Kramár, J., Vylder, B. D., Saeta, B., Bradbury, J., Ding, D., Borgeaud, S., Lai, M., Schrittwieser, J., Anthony, T. W., Hughes, E., Danihelka, I., and Ryan-Davis, J. Openspiel: A framework for reinforcement learning in games. *CoRR*, abs/1908.09453, 2019. URL http://arxiv.org/abs/1908.09453.

Li, H., Hu, K., Zhang, S., Qi, Y., and Song, L. Double neural counterfactual regret minimization. In *International Conference on Learning Representations*, 2020.

Lisỳ, V., Lanctot, M., and Bowling, M. Online Monte Carlo counterfactual regret minimization for search in imperfect information games. In *International Conference on Autonomous Agents and Multiagent Systems*, pp. 27–36, 2015.

Liu, W., Li, B., and Togelius, J. Model-free neural counterfactual regret minimization with bootstrap learning. *IEEE Transactions on Games*, 2022.

McMahan, H. B. and Streeter, M. J. Adaptive bound optimization for online convex optimization. In *Conference on Learning Theory*, pp. 244–256, 2010.

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

Morrill, D. R. Using regret estimation to solve games compactly. Master's thesis, University of Alberta, 2016.

Nesterov, Y. E. Primal-dual subgradient methods for convex problems. *Mathematical programming*, 120(1):221–259, 2009.

Orabona, F. A modern introduction to online learning. *CoRR*, abs/1912.13213, 2019. URL http://arxiv.org/abs/1912.13213.

Rakhlin, A. and Sridharan, K. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pp. 3066–3074, 2013.

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

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

Southey, F., Bowling, M., Larson, B., Piccione, C., Burch, N., Billings, D., and Rayner, D. C. Bayes' bluff: Opponent modelling in poker. *CoRR*, abs/1207.1411, 2012. URL http://arxiv.org/abs/1207.1411.

Tammelin, O., Burch, N., Johanson, M., and Bowling, M. Solving heads-up limit texas hold'em. In *International Joint Conferences on Artificial Intelligence*, pp. 645–652, 2015.

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

Waugh, K. and Bagnell, J. A. A unified view of large-scale zero-sum equilibrium computation. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, volume WS-15-07, 2015.

Waugh, K., Morrill, D., Bagnell, J. A., and Bowling, M. H. Solving games with functional regret estimation. In *AAAI Conference on Artificial Intelligence*, pp. 2138–2145, 2015.

Zinkevich, M., Johanson, M., Bowling, M. H., and Piccione, C. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems*, pp. 1729–1736, 2007.

A. Notation Table

In this section, we list the notations that appear in the article. See table 2. We mainly follow the notations in (Farina et al., 2019b) and (Farina et al., 2021). However, as this paper discusses the equivalences between algorithms in different areas, it is unavoidable to introduce new notations. Also, some notations may not be consistent with the ones defined in the mentioned literature. For example, we use $\boldsymbol{x}_j$ to denote a sequence-form strategy in $\mathcal{X}_j$, while (Farina et al., 2019b) uses it to denote a sub-vector of $\boldsymbol{x}$ that corresponds to decision point j.

B. An illustration of Sequential Decision Process

As an illustration, consider the game of Kuhn poker. Kuhn poker consists of a three-card deck: **K**ing, **Q**ueen, and **J**ack. The sequential decision tree for the first player is shown in Figure 3. For example, we have decision set $\mathcal{J} = \{0, 1, 2, 3, 4, 5, 6\}$, and action sets $A_0 = \{start\}$, $A_1 = A_2 = A_3 = \{bet, pass\}$, and $A_4 = A_5 = A_6 = \{call, fold\}$. Moreover, we have $C_{0,start} = \{1, 2, 3\}$, $C_{1,bet} = \emptyset$, $C_{3,pass} = \{6\}$, $C_{\#0} = \mathcal{J}$, $C_{\#1} = \{1, 4\}$, and et al.

For a sequence-form strategy $\boldsymbol{x} \in \mathcal{X}$ in Kuhn poker, we have an $\boldsymbol{x}_6$ in $\mathcal{X}_6$ equals an $\hat{\boldsymbol{x}}_6 \in \Delta^2$, an $\boldsymbol{x}_3$ in $\mathcal{X}_3$ equals $(\hat{\boldsymbol{x}}_3, \hat{\boldsymbol{x}}_3[pass]\boldsymbol{x}_6) = (\hat{\boldsymbol{x}}_3, \hat{\boldsymbol{x}}_3[pass]\hat{\boldsymbol{x}}_6)$, and

$$\begin{aligned}\boldsymbol{x} &= (\hat{\boldsymbol{x}}_o, \hat{\boldsymbol{x}}_o[start]\boldsymbol{x}_1, \hat{\boldsymbol{x}}_o[start]\boldsymbol{x}_2, \hat{\boldsymbol{x}}_o[start]\boldsymbol{x}_3) \\ &= (1, \hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_1[pass]\hat{\boldsymbol{x}}_4, \hat{\boldsymbol{x}}_2, \hat{\boldsymbol{x}}_2[pass]\hat{\boldsymbol{x}}_5, \hat{\boldsymbol{x}}_3, \hat{\boldsymbol{x}}_3[pass]\hat{\boldsymbol{x}}_6)\,.\end{aligned}$$

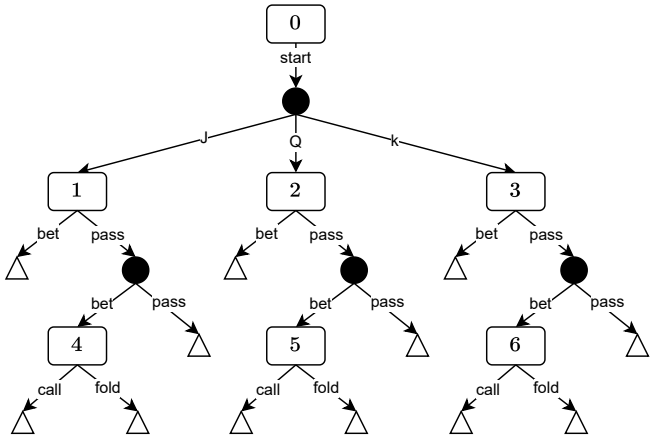

Figure 3: The sequential decision tree for the first player in the game of Kuhn poker, • denotes an observation point, △ denotes the end of the decision process. Adapted from (Farina et al., 2019b).

C. Counterfactual Regret Minimization

C.1. A Discussion on Assumption 2.1

CFRs usually initialize $\hat{\boldsymbol{R}}_j^0$ and $\hat{\boldsymbol{Q}}_j^0$ to zero for all $j \in \mathcal{J}$. In this paper, we set $\hat{\boldsymbol{R}}_j^0$ and $\hat{\boldsymbol{Q}}_j^0$ to a small value vector $\epsilon\mathbf{1} > 0$ for each $j \in \mathcal{J}$.[7] As shown in Lemma C.1 and C.2, the l_2-norm of the cumulative counterfactual regret at every decision point is monotonically increasing. Therefore, the initialization guarantees that $\|[\hat{\boldsymbol{R}}_j^t]^+\|_1 > 0$ and $\|\hat{\boldsymbol{Q}}_j^t\|_1 > 0$ for any $t > 0$. Note that, according to Lemma C.1, this initialization has a negligible effect on the convergence, as $\hat{R}_j^T \le \|[\hat{\boldsymbol{R}}_j^T]^+\|_2 \le \sqrt{\epsilon^2 n_j + \sum_{t=1}^T \|\hat{\boldsymbol{r}}_j^t\|_2^2} = O(\sqrt{T})$. The effect on CFR-RM+ is similar. Note that for simplicity, we let $\epsilon \to 0$ in the following proofs.

Lemma C.1. *In CFR-RM,* $\|[\hat{\boldsymbol{R}}_j^t\ {}^1]^+\|_2^2 \le \|[\hat{\boldsymbol{R}}_j^t]^+\|_2^2 \le \|[\hat{\boldsymbol{R}}_j^t\ {}^1]^+\|_2^2 + \|\hat{\boldsymbol{r}}_j^t\|_2^2$ *for all* $j \in \mathcal{J}$ *and* $t > 0$.

Proof. First, we prove that $\|[\hat{\boldsymbol{R}}_j^t\ {}^1]^+\|_2^2 \le \|[\hat{\boldsymbol{R}}_j^t]^+\|_2^2$. It is trivial when $\|[\hat{\boldsymbol{R}}_j^t\ {}^1]^+\|_2^2 = 0$. When $\|[\hat{\boldsymbol{R}}_j^t\ {}^1]^+\|_2^2 > 0$, we have,

[7] This setting has been adopted in OpenSpiel project (Lanctot et al., 2019): https://github.com/deepmind/open_spiel

Table 2: Notation Table

$\mathcal{J}$	the set of decision points.
j	$j \in \mathcal{J}$. A decision point.
$\mathcal{K}$	the set of observation points.
k	$k \in \mathcal{K}$. An observation point.
A_j	the action set at decision point j.
a	$a \in A_j$. An action in A_j.
$\mathcal{A}$	the set of all actions.
n_j	$n_j = \|A_j\|$, the size of the action set at decision point j.
$C_{j,a}$	the set of earliest reachable decision points after taking action a at j.
$C_{\#j}$	the set of all decision points reachable from j, including j.
o	the root decision point.
$\mathbf{1}$	an all-ones vector.
Δ^n	an n-dimensional simplex.
$\hat{\boldsymbol{x}}_j$	$\hat{\boldsymbol{x}}_j \in \Delta^{n_j}$. A decision at point j.
$\mathcal{X}$	the sequence-form strategy space of player 1.
$\mathcal{X}_j$	the sequence-form strategy space in a sub-sequential decision process starting from j.
$\mathcal{Y}$	the sequence-form strategy space of player 2.
$\boldsymbol{x}$	$\boldsymbol{x} \in \mathcal{X}$. A sequence-form strategy of player 1.
$\boldsymbol{x}_j$	$\boldsymbol{x}_j \in \mathcal{X}_j$. A sequence-form strategy of player 1 in $\mathcal{X}_j$.
$\overline{\boldsymbol{x}}$	the average strategy of player 1.
$\boldsymbol{y}$	$\boldsymbol{y} \in \mathcal{Y}$. A sequence-form strategy of player 2.
$\overline{\boldsymbol{y}}$	the average strategy of player 2.
$\boldsymbol{x}[p_j]$	the entry in $\boldsymbol{x}$ corresponding to the sequence of reaching j.
$\boldsymbol{z}[j]$	the n_j entries related to j for any $\boldsymbol{z} \in \mathsf{R}^{\sum_{k=1}^{m} n_{j_k}}$.
$\boldsymbol{z}[j,a]$	the entry corresponding to (j,a) for any vector $\boldsymbol{z} \in \mathsf{R}^{\sum_{k=1}^{m} n_{j_k}}$.
$\boldsymbol{z}[a]$	the entry corresponding to action a for any vector $\boldsymbol{z} \in \mathsf{R}^{n_j}$.
$\boldsymbol{l}$	the loss vector. $\boldsymbol{l} = \mathbf{A}\boldsymbol{y}$ in a two-player zero-sum game.
$\hat{\boldsymbol{l}}_j^t$	$\hat{\boldsymbol{l}}_j^t[a] = \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \langle \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle$. The counterfactual loss.
$\boldsymbol{L}^T$	$\boldsymbol{L}^T = \sum_{t=1}^{T} \boldsymbol{l}^t$. The cumulative loss.
$\hat{\boldsymbol{L}}_j^T$	$\hat{\boldsymbol{L}}_j^t = \sum_{t=1}^{T} \hat{\boldsymbol{l}}_j^t$. The cumulative counterfactual loss at j.
$\hat{\boldsymbol{r}}_j^t$	$\hat{\boldsymbol{r}}_j^t = \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle \mathbf{1} - \hat{\boldsymbol{l}}_j^t$. The instantaneous counterfactual regret.
$\hat{\boldsymbol{R}}_j^t$	$\hat{\boldsymbol{R}}_j^t = \sum_{k=1}^{t} \hat{\boldsymbol{r}}_j^k = \sum_{k=1}^{t} \langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k \rangle \mathbf{1} - \hat{\boldsymbol{L}}_j^t$. The cumulative counterfactual regret.
$\hat{R}_j^T$	$\hat{R}_j^T = \max_{\hat{\boldsymbol{x}}_j^0 \in \mathbb{R}^{n_j}} \sum_{t=1}^{T} \langle \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^0 \rangle = \max_{a \in A_j} \hat{\boldsymbol{R}}_j^T[a]$. The immediate counterfactual regret.
$\hat{\boldsymbol{Q}}_j^t$	$\hat{\boldsymbol{Q}}_j^t \leftarrow [\hat{\boldsymbol{Q}}_j^{t-1} + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle \mathbf{1} - \hat{\boldsymbol{l}}_j^t]^+$. The truncated cumulative counterfactual regret.
$q^{0:t}(\boldsymbol{x})$	$q^{0:t}(\boldsymbol{x}) = \sum_{k=1}^{t} q^k(\boldsymbol{x})$. The regularizer. $q^k : D \to \mathsf{R}, \mathcal{X} \subseteq D$ is a proper function.
$d(\boldsymbol{x})$	$d(\boldsymbol{x}) = \sum_{j \in \mathcal{J}} \boldsymbol{x}[p_j] \psi_j(\hat{\boldsymbol{x}}_j)$. A dilated DGF. $\psi_j : E \to \mathsf{R}, \Delta^{n_j} \subseteq E$.
$\mathcal{B}_d(\boldsymbol{x}^0 \|\| \boldsymbol{x})$	$\mathcal{B}_d(\boldsymbol{x}^0 \|\| \boldsymbol{x}) = d(\boldsymbol{x}^0) - d(\boldsymbol{x}_j) - \langle \nabla d(\boldsymbol{x}), \boldsymbol{x}^0 - \boldsymbol{x} \rangle$. The Bregman divergence of DGF d.
$\hat{\boldsymbol{L}}_j^{0t}$	$\hat{\boldsymbol{L}}_j^{0t}[a] = \boldsymbol{L}^t[j,a] + \sum_{j' \in C_{j,a}} -{\psi_{j'}}^t(-\hat{\boldsymbol{L}}_{j'}^{0t})$. The local loss of FTRL.
$\hat{\boldsymbol{l}}_j^{0t}$	$\hat{\boldsymbol{l}}_j^{0t}[a] = \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \hat{l}_{j'}^{0t}$. The local loss of OMD.
$\hat{l}_j^{0t}$	$\hat{l}_j^{0t} = {\psi_j}^{t-1}(\nabla {\psi_j}^{t-1}(\hat{\boldsymbol{x}}_j^t)) - {\psi_j}^t(\nabla {\psi_j}^{t-1}(\hat{\boldsymbol{x}}_j^t)) - \hat{\boldsymbol{l}}_j^{0t})$.
$\hat{\boldsymbol{R}}_j^{0t}$	$\hat{\boldsymbol{R}}_j^{0t} = \alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^{0t}$.
$\hat{\boldsymbol{Q}}_j^{0t}$	$\hat{\boldsymbol{Q}}_j^{0t} = \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} = [\beta_j^{t-1} \hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^{0t}]^+$.
β_j^t	$\beta_j^t > 0$. The weighting parameter for DGF ψ_j^t.
λ_j^t	$\lambda_j^t > 0$. The weighting parameter for reparameterizing β_j^t: $\beta_j^t = \sqrt{\lambda_j^t} / \|\hat{\boldsymbol{x}}_j^{t+1}\|_2$.

$$\begin{aligned}\langle[\hat{\boldsymbol{R}}_j^t]^+, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle \geq& \langle\hat{\boldsymbol{R}}_j^t, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle \\ =& \langle\hat{\boldsymbol{R}}_j^{t\ 1} + \hat{\boldsymbol{r}}_j^t, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle \\ =& \langle\hat{\boldsymbol{R}}_j^{t\ 1}, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle \\ =& \|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_2^2.\end{aligned} \tag{15}$$

The second equality is because

$$\begin{aligned}\langle\hat{\boldsymbol{r}}_j^t, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle =& \langle\hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t\rangle\|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_1 - \langle\hat{\boldsymbol{l}}_j^t, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle \\ =& \langle\hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t\rangle\|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_1 - \langle\hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t\rangle\|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_1 \\ =& 0.\end{aligned} \tag{16}$$

Secondly, according to Cauchy–Schwarz inequality, we have

$$\begin{aligned}\langle[\hat{\boldsymbol{R}}_j^t]^+, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle \leq& \|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_2\|[\hat{\boldsymbol{R}}_j^t]^+\|_2 \\ \leq& \sqrt{\langle[\hat{\boldsymbol{R}}_j^t]^+, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle}\|[\hat{\boldsymbol{R}}_j^t]^+\|_2.\end{aligned} \tag{17}$$

Since $\langle[\hat{\boldsymbol{R}}_j^t]^+, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle \geq \|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_2^2 > 0$, we have

$$\|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_2^2 \leq \|[\hat{\boldsymbol{R}}_j^t]^+\|_2^2. \tag{18}$$

Besides, we have

$$\begin{aligned}\|[\hat{\boldsymbol{R}}_j^t]^+\|_2^2 =& \|[\hat{\boldsymbol{R}}_j^{t\ 1} + \hat{\boldsymbol{r}}_j^t]^+\|_2^2 \\ \leq& \|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+ + \hat{\boldsymbol{r}}_j^t\|_2^2 \\ =& \|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_2^2 + 2\langle\hat{\boldsymbol{r}}_j^t, [\hat{\boldsymbol{R}}_j^{t\ 1}]^+\rangle + \|[\hat{\boldsymbol{r}}_j^t]^+\|_2^2 \\ =& \|[\hat{\boldsymbol{R}}_j^{t\ 1}]^+\|_2^2 + \|\hat{\boldsymbol{r}}_j^t\|_2^2.\end{aligned} \tag{19}$$

□

Lemma C.2. *In CFR-RM+, $\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 \leq \|\hat{\boldsymbol{Q}}_j^t\|_2^2 \leq \|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 + \|\hat{\boldsymbol{r}}_j^t\|_2^2$ for all $j \in \mathcal{J}$ and $t > 0$.*

Proof. First, we prove that $\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 \leq \|\hat{\boldsymbol{Q}}_j^t\|_2^2$. It is trivial when $\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 = 0$. When $\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 > 0$, we have,

$$\begin{aligned}\langle\hat{\boldsymbol{Q}}_j^t, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle \geq& \langle\hat{\boldsymbol{Q}}_j^{t\ 1} + \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle \\ =& \langle\hat{\boldsymbol{Q}}_j^{t\ 1}, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle \\ =& \|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2.\end{aligned} \tag{20}$$

The Second equality is because

$$\begin{aligned}\langle\hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle =& \langle\hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t\rangle\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_1 - \langle\hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle \\ =& \langle\hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t\rangle\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_1 - \langle\hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t\rangle\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_1 \\ =& 0.\end{aligned} \tag{21}$$

Secondly, according to Cauchy–Schwarz inequality, we have

$$\begin{aligned}\langle\hat{\boldsymbol{Q}}_j^t, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle \leq& \|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2\|\hat{\boldsymbol{Q}}_j^t\|_2 \\ \leq& \sqrt{\langle\hat{\boldsymbol{Q}}_j^t, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle}\|\hat{\boldsymbol{Q}}_j^t\|_2.\end{aligned} \tag{22}$$

So,

$$\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 \leq \|\hat{\boldsymbol{Q}}_j^t\|_2^2. \tag{23}$$

Besides, we have

$$\begin{aligned}\|\hat{\boldsymbol{Q}}_j^t\|_2^2 =&\|[\hat{\boldsymbol{Q}}_j^{t\ 1} + \hat{\boldsymbol{r}}_j^t]^+\|_2^2\\ \leq&\|\hat{\boldsymbol{Q}}_j^{t\ 1} + \hat{\boldsymbol{r}}_j^t\|_2^2\\ =&\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 + 2\langle\hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{Q}}_j^{t\ 1}\rangle + \|\hat{\boldsymbol{r}}_j^t\|_2^2\\ \leq&\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_2^2 + \|\hat{\boldsymbol{r}}_j^t\|_2^2.\end{aligned} \tag{24}$$

□

C.2. Proof for Lemma 2.2

Proof. According to the definition of counterfactual loss $\hat{\boldsymbol{l}}_j[a] = \boldsymbol{l}[j,a] + \sum_{j' \in C_{j,a}} \langle \hat{\boldsymbol{l}}_{j'}, \hat{\boldsymbol{x}}_{j'}\rangle$, it is easy to see that

$$\begin{aligned}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_j, \hat{\boldsymbol{x}}_j^0\rangle =&\boldsymbol{x}^0[p_j]\langle\boldsymbol{l}[j], \hat{\boldsymbol{x}}_j^0\rangle + \boldsymbol{x}^0[p_j]\sum_{a\in A_j}\sum_{j'\in C_{j,a}}\hat{\boldsymbol{x}}_j^0[a]\langle\hat{\boldsymbol{l}}_{j'}, \hat{\boldsymbol{x}}_{j'}\rangle\\ =&\boldsymbol{x}^0[p_j]\langle\boldsymbol{l}[j], \hat{\boldsymbol{x}}_j^0\rangle + \sum_{a\in A_j}\sum_{j'\in C_{j,a}}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_{j'}, \hat{\boldsymbol{x}}_{j'}\rangle\end{aligned} \tag{25}$$

Then, according to the definition of instantaneous counterfactual regret $\hat{\boldsymbol{r}}_j = \langle\hat{\boldsymbol{l}}_j, \hat{\boldsymbol{x}}_j\rangle\mathbf{1} - \hat{\boldsymbol{l}}_j$, the right side of the equation is

$$\begin{aligned}\sum_{j\in J}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{r}}_j, \hat{\boldsymbol{x}}_j^0\rangle =& \sum_{j\in J}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_j, \hat{\boldsymbol{x}}_j\rangle - \sum_{j\in J}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_j, \hat{\boldsymbol{x}}_j^0\rangle\\ =&\sum_{j\in J}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_j, \hat{\boldsymbol{x}}_j\rangle - \left\{\sum_{j\in J}\boldsymbol{x}^0[p_j]\langle\boldsymbol{l}[j], \hat{\boldsymbol{x}}_j^0\rangle + \sum_{j\in J}\sum_{a\in A_j}\sum_{j'\in C_{j,a}}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_{j'}, \hat{\boldsymbol{x}}_{j'}\rangle\right\}\\ =&\left\{\sum_{j\in J}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_j, \hat{\boldsymbol{x}}_j\rangle - \sum_{j\in J}\sum_{a\in A_j}\sum_{j'\in C_{j,a}}\boldsymbol{x}^0[p_j]\langle\hat{\boldsymbol{l}}_{j'}, \hat{\boldsymbol{x}}_{j'}\rangle\right\} - \sum_{j\in J}\boldsymbol{x}^0[p_j]\langle\boldsymbol{l}[j], \hat{\boldsymbol{x}}_j^0\rangle\\ =&\langle\hat{\boldsymbol{l}}_o, \hat{\boldsymbol{x}}_o\rangle - \langle\boldsymbol{l}, \boldsymbol{x}^0\rangle = \langle\boldsymbol{l}, \boldsymbol{x}\rangle - \langle\boldsymbol{l}, \boldsymbol{x}^0\rangle.\end{aligned} \tag{26}$$

So both sides of the equation are equal. The last equality is because $\langle\hat{\boldsymbol{l}}_o, \hat{\boldsymbol{x}}_o\rangle = \langle\boldsymbol{l}, \boldsymbol{x}\rangle$, according to the definition of counterfactual losses.

□

D. Online Convex Optimization and Distance-Generating Function

D.1. Recursive Definition of Dilated DGF

Given a sequence-form space $\mathcal{X}$, a dilated DGF is defined as

$$d(\boldsymbol{x}) = \sum_{j\in J}\boldsymbol{x}[p_j]\psi_j\left(\hat{\boldsymbol{x}}_j\right). \tag{27}$$

A dilated DGF can also be defined recursively. Specifically, define

$$\begin{aligned}d(\boldsymbol{x}) =&d_o(\boldsymbol{x}_o),\\ d_j(\boldsymbol{x}_j) =&\psi_j(\hat{\boldsymbol{x}}_j) + \sum_{a\in A_j, j'\in C_{j,a}}\hat{\boldsymbol{x}}_j[a]d_{j'}\left(\boldsymbol{x}_{j'}\right).\end{aligned} \tag{28}$$

D.2. Recursive Definition of Bregman Divergence

Let's begin by computing the gradient of $d_j(\boldsymbol{x}_j)$ for any $j \in \mathcal{J}$.

Lemma D.1. *Given a DGF $d_j(\boldsymbol{x}_j) = \psi_j(\hat{\boldsymbol{x}}_j) + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] d_{j^0}(\boldsymbol{x}_{j^0})$, then the partial derivative of $d_j(\boldsymbol{x}_j)$ with respect to $\boldsymbol{x}_j[j]$ is*

$$\frac{\partial d_j}{\partial \boldsymbol{x}_j[j]}(\boldsymbol{x}_j) = \nabla \psi_j(\hat{\boldsymbol{x}}_j) + \left(\sum_{j^0 \in C_{j,a}} \left(d_{j^0}(\boldsymbol{x}_{j^0}) - \langle \nabla d_{j^0}(\boldsymbol{x}_{j^0}), \boldsymbol{x}_{j^0} \rangle \right) \right)_{a \in A_j}, \tag{29}$$

and for any $a \in A_j$, $j^0 \in C_{j,a}$,

$$\frac{\partial d_j}{\partial \boldsymbol{x}_j[\downarrow j^0]}(\boldsymbol{x}_j) = \nabla d_{j^0}(\boldsymbol{x}_{j^0}). \tag{30}$$

Proof. Let's first see partial derivative of $d_j(\boldsymbol{x}_j)$ with respect to $\boldsymbol{x}_j[j,a]$:

$$\frac{\partial d_j}{\partial \boldsymbol{x}_j[j,a]}(\boldsymbol{x}_j) = \frac{\partial \psi_j}{\partial \boldsymbol{x}_j[j,a]}(\boldsymbol{x}_j[j]) + \sum_{j^0 \in C_{j,a}} \partial \left(\boldsymbol{x}_j[j,a] d_{j^0}\left(\frac{\boldsymbol{x}_j[\downarrow j^0]}{\boldsymbol{x}_j[j,a]} \right) \right) / \partial \boldsymbol{x}_j[j,a]. \tag{31}$$

Since

$$\partial \left(\boldsymbol{x}_j[j,a] d_{j^0}\left(\frac{\boldsymbol{x}_j[\downarrow j^0]}{\boldsymbol{x}_j[j,a]} \right) \right) / \partial \boldsymbol{x}_j[j,a] = d_{j^0}\left(\frac{\boldsymbol{x}_j[\downarrow j^0]}{\boldsymbol{x}_j[j,a]} \right) - \left\langle \nabla d_{j^0}\left(\frac{\boldsymbol{x}_j[\downarrow j^0]}{\boldsymbol{x}_j[j,a]} \right), \frac{\boldsymbol{x}_j[\downarrow j^0]}{\boldsymbol{x}_j[j,a]} \right\rangle, \tag{32}$$

we have

$$\frac{\partial d_j}{\partial \boldsymbol{x}_j[j]}(\boldsymbol{x}_j) = \nabla \psi_j(\hat{\boldsymbol{x}}_j) + \left(\sum_{j^0 \in C_{j,a}} \left(d_{j^0}(\boldsymbol{x}_{j^0}) - \langle \nabla d_{j^0}(\boldsymbol{x}_{j^0}), \boldsymbol{x}_{j^0} \rangle \right) \right)_{a \in A_j}. \tag{33}$$

For $\partial d_j(\boldsymbol{x}_j)/\partial \boldsymbol{x}_j[\downarrow j^0]$, we have,

$$\frac{\partial d_j}{\partial \boldsymbol{x}_j[\downarrow j^0]}(\boldsymbol{x}_j) = \partial \left(\boldsymbol{x}_j[j,a] d_{j^0}\left(\frac{\boldsymbol{x}_j[\downarrow j^0]}{\boldsymbol{x}_j[j,a]} \right) \right) / \partial \boldsymbol{x}_j[\downarrow j^0] = \nabla d_{j^0}\left(\frac{\boldsymbol{x}_j[\downarrow j^0]}{\boldsymbol{x}_j[j,a]} \right) = \nabla d_{j^0}(\boldsymbol{x}_{j^0}). \tag{34}$$

□

Based on the above lemma, we can construct the Bregman divergence recursively for any dilated DGF.

Lemma D.2. *A Bregman divergence $\mathcal{B}_d(\boldsymbol{x} \| \boldsymbol{x}^0)$ constructed from a dilated DGF $d(\boldsymbol{x})$, which is differentiable and strictly convex on a sequence-form space $\mathcal{X}$, can be constructed recursively as follows:*

$$\begin{aligned} \mathcal{B}_d(\boldsymbol{x} \| \boldsymbol{x}^0) =& \mathcal{B}_{d_o}(\boldsymbol{x}_o \| \boldsymbol{x}_o^0), \\ \mathcal{B}_{d_j}(\boldsymbol{x}_j \| \boldsymbol{x}_j^0) =& \mathcal{B}_{\psi_j}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^0) + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] \mathcal{B}_{d_{j^0}}(\boldsymbol{x}_{j^0} \| \boldsymbol{x}_{j^0}^0), \end{aligned} \tag{35}$$

where $\mathcal{B}_{\psi_j}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^0)$ is the Bregman divergence constructed from $\psi_j(\hat{\boldsymbol{x}}_j)$, which is differentiable and strictly convex on Δ^{n_j}.

Proof. According to Lemma D.1, we have

$$\begin{aligned} & \left\langle \nabla d_j(\boldsymbol{x}_j^0), \boldsymbol{x}_j - \boldsymbol{x}_j^0 \right\rangle \\ =& \left\langle \nabla \psi_j(\hat{\boldsymbol{x}}_j^0), \hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_j^0 \right\rangle + \sum_{a \in A_j, j^0 \in C_{j,a}} \left(\hat{\boldsymbol{x}}_j[a] - \hat{\boldsymbol{x}}_j^0[a] \right) \left(d_{j^0}\left(\boldsymbol{x}_{j^0}^0 \right) - \left\langle \nabla d_{j^0}\left(\boldsymbol{x}_{j^0}^0 \right), \boldsymbol{x}_{j^0}^0 \right\rangle \right) \\ & + \sum_{a \in A_j, j^0 \in C_{j,a}} \left\langle \nabla d_{j^0}\left(\boldsymbol{x}_{j^0}^0 \right), \hat{\boldsymbol{x}}_j[a] \boldsymbol{x}_{j^0} - \hat{\boldsymbol{x}}_j^0[a] \boldsymbol{x}_{j^0}^0 \right\rangle \\ =& \left\langle \nabla \psi_j(\hat{\boldsymbol{x}}_j^0), \hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_j^0 \right\rangle + \sum_{a \in A_j, j^0 \in C_{j,a}} \left(\hat{\boldsymbol{x}}_j[a] d_{j^0}\left(\boldsymbol{x}_{j^0}^0 \right) - \hat{\boldsymbol{x}}_j^0[a] d_{j^0}\left(\boldsymbol{x}_{j^0}^0 \right) \right) \\ & + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] \left\langle \nabla d_{j^0}\left(\boldsymbol{x}_{j^0}^0 \right), \boldsymbol{x}_{j^0} - \boldsymbol{x}_{j^0}^0 \right\rangle \end{aligned} \tag{36}$$

So, the Bregman divergence defined on the dilated DGF is

$$\begin{aligned}
&\mathcal{B}_{d_j}(\boldsymbol{x}_j\|\boldsymbol{x}_j^0) = d_j(\boldsymbol{x}_j) - d_j(\boldsymbol{x}_j^0) - \left\langle \nabla d_j(\boldsymbol{x}_j^0), \boldsymbol{x}_j - \boldsymbol{x}_j^0 \right\rangle \\
=&\psi_j(\boldsymbol{x}_j[j]) - \psi_j(\boldsymbol{x}_j^0[j]) + \sum_{a \in A_j, j^0 \in C_{j,a}} \left(\hat{\boldsymbol{x}}_j[a] d_{j^0}\left(\boldsymbol{x}_{j^0}\right) - \hat{\boldsymbol{x}}_j^0[a] d_{j^0}\left(\boldsymbol{x}_{j^0}^0\right)\right) - \left\langle \nabla d_j(\boldsymbol{x}_j^0), \boldsymbol{x}_j - \boldsymbol{x}_j^0 \right\rangle \\
=&\psi_j(\boldsymbol{x}_j[j]) - \psi_j(\boldsymbol{x}_j^0[j]) - \left\langle \nabla \psi_j(\hat{\boldsymbol{x}}_j^0), \hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_j^0 \right\rangle \\
&\quad + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] \left(d_{j^0}\left(\boldsymbol{x}_{j^0}\right) - d_{j^0}\left(\boldsymbol{x}_{j^0}^0\right) - \left\langle \nabla d_{j^0}\left(\boldsymbol{x}_{j^0}^0\right), \boldsymbol{x}_{j^0} - \boldsymbol{x}_{j^0}^0 \right\rangle\right) \\
=&\mathcal{B}_{\psi_j}(\hat{\boldsymbol{x}}_j\|\hat{\boldsymbol{x}}_j^0) + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] \mathcal{B}_{d_{j^0}}(\boldsymbol{x}_{j^0}\|\boldsymbol{x}_{j^0}^0).
\end{aligned} \tag{37}$$

□

D.3. Proof for Proposition 2.3

Proof. When $q^{0:t}(\boldsymbol{x}) = \sum_{j \in J} \boldsymbol{x}[p_j]\psi_j^t(\hat{\boldsymbol{x}}_j)$ is a dilated DGF, according to the recursive definition of dilate DGF, we let

$$q_j^{0:t}(\boldsymbol{x}_j) = \sum_{j^0 \in C_{\# j}} \boldsymbol{x}_j[p_j]\psi_{j^0}^t(\hat{\boldsymbol{x}}_{j^0}). \tag{38}$$

Define the target function for FTRL as

$$F^t(\boldsymbol{x}) = \langle \boldsymbol{L}^t, \boldsymbol{x} \rangle + q^{0:t}(\boldsymbol{x}). \tag{39}$$

Then, we have $\boldsymbol{x}^{t+1} = \operatorname{argmin}_{\boldsymbol{x} \in X} F^t(\boldsymbol{x})$. Furthermore, define

$$F_j^t(\boldsymbol{x}_j) = \langle \boldsymbol{L}^t[\downarrow j], \boldsymbol{x}_j \rangle + q_j^{0:t}(\boldsymbol{x}_j). \tag{40}$$

Note that $F^t(\boldsymbol{x}) = F_o^t(\boldsymbol{x}_o)$ and $q^{0:t\ \ 1}(\boldsymbol{x}) = q_o^{0:t\ \ 1}(\boldsymbol{x}_o)$. Since both the sequence-form space and the dilated DGF are defined recursively, we have

$$\begin{aligned}
F_j^t(\boldsymbol{x}_j) =& \langle \boldsymbol{L}^t[\downarrow j], \boldsymbol{x}_j \rangle + q_j^{0:t}(\boldsymbol{x}_j) \\
=& \left\{ \langle \boldsymbol{L}^t[j], \boldsymbol{x}_j[j] \rangle + \sum_{a \in A_j, j^0 \in C_{j,a}} \langle [\boldsymbol{L}^t[\downarrow j^0], \boldsymbol{x}_j[\downarrow j^0] \rangle \right\} + \left\{ \psi_j^t(\hat{\boldsymbol{x}}_j) + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] q_{j^0}^{0:t}\left(\boldsymbol{x}_{j^0}\right) \right\}.
\end{aligned} \tag{41}$$

Since $\boldsymbol{x}_j[\downarrow j^0]/\boldsymbol{x}_j[j,a] = \boldsymbol{x}_{j^0}$ and $\boldsymbol{x}_j[j,a] = \hat{\boldsymbol{x}}_j[a]$, we have

$$\begin{aligned}
F_j^t(\boldsymbol{x}_j) =& \langle \boldsymbol{L}^t[j], \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] \left(\langle \boldsymbol{L}^t[\downarrow j^0], \boldsymbol{x}_{j^0} \rangle + q_{j^0}^{0:t}\left(\boldsymbol{x}_{j^0}\right) \right) \\
=& \langle \boldsymbol{L}^t[j], \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) + \sum_{a \in A_j, j^0 \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] F_{j^0}^t(\boldsymbol{x}_{j^0}).
\end{aligned} \tag{42}$$

Let $\tilde{\boldsymbol{g}}_j^t = \left(\sum_{j^0 \in C_{j,a}} F_{j^0}^t(\boldsymbol{x}_{j^0})\right)_{a \in A_j}$, then

$$F_j^t(\boldsymbol{x}_j) = \langle \boldsymbol{L}^t[j] + \tilde{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j), \tag{43}$$

$$\begin{aligned}
\min_{\boldsymbol{x}_j \in X_j} F_j^t(\boldsymbol{x}_j) &= \min_{\boldsymbol{x}_j \in X_j} \left\{ \langle \boldsymbol{L}^t[j] + \tilde{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) \right\} \\
&= \min_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle \boldsymbol{L}^t[j] + \min_{\boldsymbol{x}_j} \tilde{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) \right\} \\
&= \min_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle \boldsymbol{L}^t[j] + \hat{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) \right\} \\
&= -\psi_j^{\ t}(-\boldsymbol{L}^t[j] - \hat{\boldsymbol{g}}_j^t).
\end{aligned} \tag{44}$$

where $\hat{\boldsymbol{g}}_j^t[a] = \sum_{j' \in C_{j,a}} \min_{\hat{\boldsymbol{x}}_{j'} \in \mathcal{X}_{j'}} F_{j'}^t(\boldsymbol{x}_{j'})$, and

$$\psi_j{}^t(-\boldsymbol{L}^t[j] - \hat{\boldsymbol{g}}_j^t) = \max_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle -\boldsymbol{L}^t[j] - \hat{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle - \psi_j^t(\hat{\boldsymbol{x}}_j) \right\}. \tag{45}$$

More importantly, we have

$$\hat{\boldsymbol{x}}_j^{t+1} = \operatorname*{argmin}_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle \boldsymbol{L}^t[j] + \hat{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) \right\} = \nabla \psi_j{}^t(-\boldsymbol{L}^t[j] - \hat{\boldsymbol{g}}_j^t). \tag{46}$$

This means we can compute the decision at point j locally. The above equation holds for all $j \in \mathcal{J}$, so we have $\hat{\boldsymbol{g}}_j^t[a] = \sum_{j' \in C_{j,a}} \min_{\hat{\boldsymbol{x}}_{j'} \in \mathcal{X}_{j'}} F_{j'}^t(\boldsymbol{x}_{j'}) = \sum_{j' \in C_{j,a}} -\psi_{j'}{}^t(-\boldsymbol{L}^t[j'] - \hat{\boldsymbol{g}}_{j'}^t)$. In conclusion, at every decision point $j \in \mathcal{J}$, we have

$$\hat{\boldsymbol{x}}_j^{t+1} = \operatorname*{argmin}_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \left\langle \hat{\boldsymbol{L}}_j^{\emptyset t}, \hat{\boldsymbol{x}}_j \right\rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) \right\} = \nabla \psi_j{}^t(-\hat{\boldsymbol{L}}_j^{\emptyset t}), \tag{47}$$

where

$$\hat{\boldsymbol{L}}_j^{\emptyset t}[a] = \boldsymbol{L}^t[j, a] + \sum_{j' \in C_{j,a}} -\psi_{j'}{}^t(-\hat{\boldsymbol{L}}_{j'}^{\emptyset t}). \tag{48}$$

□

D.4. Proof for Proposition 2.4

Proof. As in the proof for Proposition 2.4, we let

$$q_j^{\emptyset:t}(\boldsymbol{x}_j) = d_j(\boldsymbol{x}_j) = \sum_{j' \in C_{\#j}} \boldsymbol{x}_j[p_j] \psi_{j'}^t(\hat{\boldsymbol{x}}_{j'}). \tag{49}$$

Define the target function for OMD as

$$G^t(\boldsymbol{x}) = \langle \boldsymbol{l}^t, \boldsymbol{x} \rangle + q^t(\boldsymbol{x}) + \mathcal{B}_{q^{\emptyset:t}}(\boldsymbol{x} \| \boldsymbol{x}^t). \tag{50}$$

Then, we have $\boldsymbol{x}^{t+1} = \operatorname{argmin}_{\boldsymbol{x} \in \mathcal{X}} G^t(\boldsymbol{x})$ for OMD. Furthermore, define

$$G_j^t(\boldsymbol{x}_j) = \langle \boldsymbol{l}^t[\downarrow j], \boldsymbol{x}_j \rangle + q_j^t(\boldsymbol{x}_j) + \mathcal{B}_{q_j^{\emptyset:t}}(\boldsymbol{x}_j \| \boldsymbol{x}_j^t). \tag{51}$$

Note that $G^t(\boldsymbol{x}) = G_o^t(\boldsymbol{x}_o)$. Since both the sequence-form space and the Bregman divergence are defined recursively, we have

$$\begin{aligned}
&G_j^t(\boldsymbol{x}_j) = \langle \boldsymbol{l}^t[\downarrow j], \boldsymbol{x}_j \rangle + q_j^t(\boldsymbol{x}_j) + \mathcal{B}_{q_j^{\emptyset:t}}(\boldsymbol{x}_j \| \boldsymbol{x}_j^t) \\
=&\left\{ \langle \boldsymbol{l}^t[j], \boldsymbol{x}_j[j] \rangle + \sum_{a \in A_j, j' \in C_{j,a}} \langle [\boldsymbol{l}^t[\downarrow j'], \boldsymbol{x}_j[\downarrow j'] \rangle \right\} \\
&\quad + \left\{ \psi_j^t(\hat{\boldsymbol{x}}_j) - \psi_j^t{}^{\mathbf{1}}(\hat{\boldsymbol{x}}_j) + \mathcal{B}_{\psi_j^t}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^t) + \sum_{a \in A_j, j' \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] \left(q_{j'}^t(\boldsymbol{x}_{j'}) + \mathcal{B}_{q_{j'}^{\emptyset:t}}(\boldsymbol{x}_{j'} \| \boldsymbol{x}_{j'}^t) \right) \right\} \\
=&\langle \boldsymbol{l}^t[j], \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) - \psi_j^t{}^{\mathbf{1}}(\hat{\boldsymbol{x}}_j) + \mathcal{B}_{\psi_j^t}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^t) \\
&\quad + \sum_{a \in A_j, j' \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] \left(\langle \boldsymbol{l}^t[\downarrow j'], \boldsymbol{x}_{j'} \rangle + q_{j'}^t(\boldsymbol{x}_{j'}) + \mathcal{B}_{q_{j'}^{\emptyset:t}}(\boldsymbol{x}_{j'} \| \boldsymbol{x}_{j'}^t) \right) \\
=&\langle \boldsymbol{l}^t[j], \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) - \psi_j^t{}^{\mathbf{1}}(\hat{\boldsymbol{x}}_j) + \mathcal{B}_{\psi_j^t}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^t) + \sum_{a \in A_j, j' \in C_{j,a}} \hat{\boldsymbol{x}}_j[a] G_{j'}^t(\boldsymbol{x}_{j'}).
\end{aligned} \tag{52}$$

Let $\tilde{\boldsymbol{g}}_j^t = \left(\sum_{j' \in C_{j,a}} G_{j'}^t(\boldsymbol{x}_{j'}) \right)_{a \in A_j}$, then

$$G_j^t(\boldsymbol{x}_j) = \langle \boldsymbol{l}^t[j] + \tilde{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) - \psi_j^t{}^{\mathbf{1}}(\hat{\boldsymbol{x}}_j) + \mathcal{B}_{\psi_j^t}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^t), \tag{53}$$

$$\begin{aligned}\min_{\boldsymbol{x}_j \in \mathcal{X}_j} G_j^t(\boldsymbol{x}_j) &= \min_{\boldsymbol{x}_j \in \mathcal{X}_j} \left\{ \langle \boldsymbol{l}^t[j] + \tilde{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) - {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j) + \mathcal{B}_{{\psi_j^t}^{*1}}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^t) \right\} \\ &= \min_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle \boldsymbol{l}^t[j] + \min_{\boldsymbol{x}_j} \tilde{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) - {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j) + \mathcal{B}_{{\psi_j^t}^{*1}}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^t) \right\} \\ &= \min_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle \boldsymbol{l}^t[j] + \hat{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) - {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t) - \langle \nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t), \hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_j^t \rangle \right\} \\ &= {\psi_j^t}^{*1}(\nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t)) - {\psi_j^t}^{*1}(\nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t) - \boldsymbol{l}^t[j] - \hat{\boldsymbol{g}}_j^t). \end{aligned} \tag{54}$$

where $\hat{\boldsymbol{g}}_j^t[a] = \sum_{j' \in C_{j,a}} \min_{\boldsymbol{x}_{j'} \in \mathcal{X}_{j'}} G_{j'}^t(\boldsymbol{x}_{j'})$, and

$$\begin{aligned} {\psi_j^t}^{*1}(\nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t)) &\overset{\text{Fenchel-Young Inequality}}{=\!=\!=} \langle \nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t), \hat{\boldsymbol{x}}_j^t \rangle - {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t), \\ {\psi_j^t}^{*1}(\nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t) - \boldsymbol{l}^t[j] - \hat{\boldsymbol{g}}_j^t) &= \max_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle \nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t) - \boldsymbol{l}^t[j] - \hat{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle - \psi_j^t(\hat{\boldsymbol{x}}_j) \right\}. \end{aligned} \tag{55}$$

Importantly, we can compute the decision at point j locally:

$$\begin{aligned} \hat{\boldsymbol{x}}_j^{t+1} &= \operatorname*{argmin}_{\hat{\boldsymbol{x}}_j \in \mathbb{R}^{n_j}} \left\{ \langle \boldsymbol{l}^t[j] + \hat{\boldsymbol{g}}_j^t, \hat{\boldsymbol{x}}_j \rangle + \psi_j^t(\hat{\boldsymbol{x}}_j) - {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j) + \mathcal{B}_{{\psi_j^t}^{*1}}(\hat{\boldsymbol{x}}_j \| \hat{\boldsymbol{x}}_j^t) \right\} \\ &= \nabla {\psi_j^t}^{*1}(\nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t) - \boldsymbol{l}^t[j] - \hat{\boldsymbol{g}}_j^t). \end{aligned} \tag{56}$$

The above equation holds for all $j \in \mathcal{J}$. In conclusion,

$$\hat{\boldsymbol{x}}_j^{t+1} = \nabla {\psi_j^t}^{*1}(\nabla {\psi_j^t}^{*1}(\hat{\boldsymbol{x}}_j^t) - \hat{\boldsymbol{l}}_j^{0t}). \tag{57}$$

where

$$\hat{\boldsymbol{l}}_j^{0t}[a] = \boldsymbol{l}^t[j,a] + \sum_{j' \in C_{j,a}} \left\{ {\psi_{j'}^t}^{*1}(\nabla {\psi_{j'}^t}^{*1}(\hat{\boldsymbol{x}}_{j'}^t)) - {\psi_{j'}^t}(\nabla {\psi_{j'}^t}^{*1}(\hat{\boldsymbol{x}}_{j'}^t) - \hat{\boldsymbol{l}}_{j'}^{0t}) \right\}. \tag{58}$$

□

E. Equivalence Analysis and Its Application

E.1. Proof for Theorem 3.5

Proof. According to Lemma 2 and Theorem 3 in (Joulani et al., 2020), for both FD-FTRL and FD-OMD, we have

$$\begin{aligned} \sum_{t=1}^{T} \left(\langle \boldsymbol{l}^t, \boldsymbol{x}^t \rangle - \langle \boldsymbol{l}^t, \boldsymbol{x}^0 \rangle \right) &\le \sum_{t=0}^{T} \left(q^t(\boldsymbol{x}^0) - q^t(\boldsymbol{x}^{t+1}) \right) + \sum_{t=1}^{T} \left(\langle \boldsymbol{l}^t, \boldsymbol{x}^t - \boldsymbol{x}^{t+1} \rangle - \mathcal{B}_{{q^{0:t}}^{*1}}(\boldsymbol{x}^{t+1} \| \boldsymbol{x}^t) \right) \\ &= q^{0:T}(\boldsymbol{x}^0) - q^0(\boldsymbol{x}^1) + \sum_{t=1}^{T} \left(-q^t(\boldsymbol{x}^{t+1}) + \langle \boldsymbol{l}^t, \boldsymbol{x}^t - \boldsymbol{x}^{t+1} \rangle - \mathcal{B}_{{q^{0:t}}^{*1}}(\boldsymbol{x}^{t+1} \| \boldsymbol{x}^t) \right). \end{aligned} \tag{59}$$

where ${q^{0:t}}^{*1}(\boldsymbol{x}) = \sum_{j \in \mathcal{J}} \boldsymbol{x}[p_j] \beta_j^{t\ 1} \frac{1}{2}(\|\hat{\boldsymbol{x}}_j\|_2^2 + \|\hat{\boldsymbol{x}}_j^t\|_2^2)$ and $\mathcal{B}_{{q^{0:t}}^{*1}}(\boldsymbol{x}^{t+1} \| \boldsymbol{x}^t) = \sum_{j \in \mathcal{J}} \boldsymbol{x}^{t+1}[p_j] \beta_j^{t\ 1} \frac{1}{2} \|\hat{\boldsymbol{x}}_j^{t+1} - \hat{\boldsymbol{x}}_j^t\|_2^2$. Note that Assumptions 1, 2, 3, 5, and 8 in (Joulani et al., 2020) have already been fulfilled.

Firstly, for the first term $q^{0:T}(\boldsymbol{x}^0) - q^0(\boldsymbol{x}^1)$ in the above equation, according to the definition of the regularizer, we have

$$q^{0:T}(\boldsymbol{x}^0) - q^0(\boldsymbol{x}^1) \le q^{0:T}(\boldsymbol{x}^0) = \sum_{j \in \mathcal{J}} \boldsymbol{x}^0[p_j] \beta_j^T \left(\frac{1}{2} \|\hat{\boldsymbol{x}}_j^0\|_2^2 + \frac{1}{2} \|\hat{\boldsymbol{x}}_j^{T+1}\|_2^2 \right) \le \sum_{j \in \mathcal{J}} \boldsymbol{x}^0[p_j] \beta_j^T. \tag{60}$$

Then, for $-q^t(\boldsymbol{x}^{t+1})$, we have

$$\begin{aligned} &-q^t(\boldsymbol{x}^{t+1}) = {q^{0:t}}^{*1}(\boldsymbol{x}^{t+1}) - q^{0:t}(\boldsymbol{x}^{t+1}) \\ &= \sum_{j \in \mathcal{J}} \boldsymbol{x}^{t+1}[p_j] \beta_j^{t\ 1} \left(\frac{1}{2} \|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2 + \frac{1}{2} \|\hat{\boldsymbol{x}}_j^t\|_2^2 \right) - \sum_{j \in \mathcal{J}} \boldsymbol{x}^{t+1}[p_j] \beta_j^t \left(\frac{1}{2} \|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2 + \frac{1}{2} \|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2 \right) \\ &= \sum_{j \in \mathcal{J}} \boldsymbol{x}^{t+1}[p_j] \left(\frac{1}{2} \beta_j^{t\ 1} \|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2 + \frac{1}{2} \beta_j^{t\ 1} \|\hat{\boldsymbol{x}}_j^t\|_2^2 - \beta_j^t \|\hat{\boldsymbol{x}}_j^{t+1}\|_2^2 \right). \end{aligned} \tag{61}$$

For the term $\langle \boldsymbol{l}^t, \boldsymbol{x}^t - \boldsymbol{x}^{t+1} \rangle$, according to Lemma 2.2 in the paper,

$$\langle \boldsymbol{l}^t, \boldsymbol{x}^t - \boldsymbol{x}^{t+1} \rangle = \sum_{j \in J} \boldsymbol{x}^{t+1}[p_j] \langle \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^{t+1} \rangle. \tag{62}$$

Besides, according to the definition of Bregman divergence, we have

$$\begin{aligned} -\mathcal{B}_{q^{0:t-1}}(\boldsymbol{x}^{t+1} \| \boldsymbol{x}^t) &= -\sum_{j \in J} \boldsymbol{x}^{t+1}[p_j] \beta_j^{t-1} \frac{1}{2} \| \hat{\boldsymbol{x}}_j^{t+1} - \hat{\boldsymbol{x}}_j^t \|_2^2 \\ &= \sum_{j \in J} \boldsymbol{x}^{t+1}[p_j] \left(-\frac{1}{2} \beta_j^{t-1} \| \hat{\boldsymbol{x}}_j^{t+1} \|_2^2 - \frac{1}{2} \beta_j^{t-1} \| \hat{\boldsymbol{x}}_j^t \|_2^2 + \beta_j^{t-1} \langle \hat{\boldsymbol{x}}_j^{t+1}, \hat{\boldsymbol{x}}_j^t \rangle \right). \end{aligned} \tag{63}$$

Therefore, for the second term in the first equation, we have

$$\begin{aligned} & -q^t(\boldsymbol{x}^{t+1}) + \langle \boldsymbol{l}^t, \boldsymbol{x}^t - \boldsymbol{x}^{t+1} \rangle - \mathcal{B}_{q^{0:t-1}}(\boldsymbol{x}^{t+1} \| \boldsymbol{x}^t) \\ =& \sum_{j \in J} \boldsymbol{x}^{t+1}[p_j] \left(\beta_j^{t-1} \langle \hat{\boldsymbol{x}}_j^{t+1}, \hat{\boldsymbol{x}}_j^t \rangle - \beta_j^t \| \hat{\boldsymbol{x}}_j^{t+1} \|_2^2 + \langle \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^{t+1} \rangle \right) \\ =& \sum_{j \in J} \boldsymbol{x}^{t+1}[p_j] \langle \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t - \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} + \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^{t+1} \rangle. \end{aligned} \tag{64}$$

According to the Fenchel-Young inequality, we have

$$\begin{aligned} \langle \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t - \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} + \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^{t+1} \rangle \leq & \frac{\| \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t + \hat{\boldsymbol{r}}_j^t \|_2^2}{2\beta_j^t} - \frac{\| \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} \|_2^2}{2\beta_j^t} = \frac{\| \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t \|_2^2 + \| \hat{\boldsymbol{r}}_j^t \|_2^2}{2\beta_j^t} - \frac{\| \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} \|_2^2}{2\beta_j^t} \\ = & \frac{\| \hat{\boldsymbol{r}}_j^t \|_2^2 + \| \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t \|_2^2 - \| \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} \|_2^2}{2\beta_j^t}. \end{aligned} \tag{65}$$

The first equality is because $\langle \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle = \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle - \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle = 0$. Combining the above equations, we have

$$\begin{aligned} \sum_{t=1}^T \left(\langle \boldsymbol{l}^t, \boldsymbol{x}^t \rangle - \langle \boldsymbol{l}^t, \boldsymbol{x}^0 \rangle \right) \leq & q^{0:T}(\boldsymbol{x}^0) - q^0(\boldsymbol{x}^1) + \sum_{t=1}^T \left(-q^t(\boldsymbol{x}^{t+1}) + \langle \boldsymbol{l}^t, \boldsymbol{x}^t - \boldsymbol{x}^{t+1} \rangle - \mathcal{B}_{q^{0:t-1}}(\boldsymbol{x}^{t+1} \| \boldsymbol{x}^t) \right) \\ \leq & \sum_{j \in J} \boldsymbol{x}^0[p_j] \beta_j^T + \sum_{j \in J} \sum_{t=1}^T \boldsymbol{x}^{t+1}[p_j] \langle \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t - \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} + \hat{\boldsymbol{r}}_j^t, \hat{\boldsymbol{x}}_j^{t+1} \rangle \\ \leq & \sum_{j \in J} \left(\boldsymbol{x}^0[p_j] \beta_j^T + \sum_{t=1}^T \boldsymbol{x}^{t+1}[p_j] \frac{\| \hat{\boldsymbol{r}}_j^t \|_2^2 + \| \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t \|_2^2 - \| \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} \|_2^2}{2\beta_j^t} \right) \\ \leq & \sum_{j \in J} \left(\beta_j^T + \sum_{t=1}^T \frac{\left[\| \hat{\boldsymbol{r}}_j^t \|_2^2 + \| \beta_j^{t-1} \hat{\boldsymbol{x}}_j^t \|_2^2 - \| \beta_j^t \hat{\boldsymbol{x}}_j^{t+1} \|_2^2 \right]^+}{2\beta_j^t} \right). \end{aligned} \tag{66}$$

□

E.2. Proof for Theorem 3.7

Proof. We prove the equivalence by recursively proving that the local loss $\hat{\boldsymbol{L}}_j^{0t}$ and the local decision $\hat{\boldsymbol{x}}_j^{t+1}$ in FD-FTRL equal $\hat{\boldsymbol{L}}_j^t$ and $[\hat{\boldsymbol{R}}_j^t]^+ / \| [\hat{\boldsymbol{R}}_j^t]^+ \|_1$ in CFR-RM, respectively, i.e., the following equations hold at all decision points:

$$\hat{\boldsymbol{L}}_j^{0t} = \hat{\boldsymbol{L}}_j^t, \tag{67}$$

$$\hat{\boldsymbol{x}}_j^{t+1} = \frac{[\hat{\boldsymbol{R}}_j^t]^+}{\| [\hat{\boldsymbol{R}}_j^t]^+ \|_1}. \tag{68}$$

Let the depth of a decision point $j \in \mathcal{J}$ be the maximum length of the sequence to the end of the game, denoted by D_j. First, at any decision point $j \in \mathcal{J}$ that has $D_j = 1$, we have $\hat{\boldsymbol{L}}_j^{0t} = \boldsymbol{L}^t[j] = \hat{\boldsymbol{L}}_j^t$. So (67) holds. Besides, according to Proposition 2.3 in the paper, we have

$$\hat{\boldsymbol{x}}_j^{t+1} = \nabla {\psi_j}^t(-\hat{\boldsymbol{L}}_j^{0t}) = \nabla {\psi_j}^t(-\hat{\boldsymbol{L}}_j^t) = \frac{[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+}{\beta_j^t}, \tag{69}$$

where $\alpha_j^t \in \mathsf{R}$ satisfies $\|[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+\|_1 = \beta_j^t$. Since function $f(\alpha_j^t) \mapsto \|[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+\|_1$ is monotone ascending and convex in the range of $[\min_a \hat{\boldsymbol{L}}_j^t[a], \infty)$ with the minimum equals zero, α_j^t exists and is unique. Since $\hat{\boldsymbol{R}}_j^t = \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k \rangle \mathbf{1} - \hat{\boldsymbol{L}}_j^t$, when $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$, we can conclude that the unique solution of α_j^t in the equation is $\sum_{k=1}^t \langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k \rangle$. So

$$\hat{\boldsymbol{x}}_j^{t+1} = \frac{[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+}{\beta_j^t} = \frac{[\sum_{k=1}^t \langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k \rangle \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+}{\beta_j^t} = \frac{[\hat{\boldsymbol{R}}_j^t]^+}{\|[\hat{\boldsymbol{R}}_j^t]^+\|_1}, \tag{70}$$

So (68) holds at every decision point $j \in \mathcal{J}$ that has $D_j = 1$.

Now, assume that (68) and (67) are satisfied at decision points whose depth is less than or equal to k. For a decision point $j \in \mathcal{J}$ that has $D_j = k + 1$, we have

$$\hat{\boldsymbol{L}}_j^{0t}[a] = \boldsymbol{L}^t[j, a] + \sum_{j^0 2 C_{j,a}} -{\psi_{j^0}}^t(-\hat{\boldsymbol{L}}_{j^0}^{0t}), \tag{71}$$

where ${\psi_{j^0}}^t$ is the convex conjugate:

$$-{\psi_{j^0}}^t(-\hat{\boldsymbol{L}}_{j^0}^{0t}) = \langle \hat{\boldsymbol{L}}_{j^0}^{0t}, \hat{\boldsymbol{x}}_{j^0}^{t+1} \rangle + \psi_{j^0}^t(\hat{\boldsymbol{x}}_{j^0}^{t+1}). \tag{72}$$

Since (70) holds at decision point j^0 whose depth is less than or equal to k, we have $\beta_{j^0}^t \hat{\boldsymbol{x}}_{j^0}^{t+1} = [\hat{\boldsymbol{L}}_{j^0}^t - \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_{j^0}^k, \hat{\boldsymbol{x}}_{j^0}^k \rangle]^+$, and

$$\begin{aligned}
-{\psi_{j^0}}^t(-\hat{\boldsymbol{L}}_{j^0}^{0t}) &= \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_{j^0}^k, \hat{\boldsymbol{x}}_{j^0}^k \rangle + \left\langle \hat{\boldsymbol{L}}_{j^0}^t - \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_{j^0}^k, \hat{\boldsymbol{x}}_{j^0}^k \rangle, \hat{\boldsymbol{x}}_{j^0}^{t+1} \right\rangle + \psi_{j^0}^t(\hat{\boldsymbol{x}}_{j^0}^{t+1}) \\
&= \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_{j^0}^k, \hat{\boldsymbol{x}}_{j^0}^k \rangle - \beta_{j^0}^t \|\hat{\boldsymbol{x}}_{j^0}^{t+1}\|_2^2 + \frac{1}{2}\beta_{j^0}^t \|\hat{\boldsymbol{x}}_{j^0}^{t+1}\|_2^2 + \frac{1}{2}\beta_{j^0}^t \|\hat{\boldsymbol{x}}_{j^0}^{t+1}\|_2^2 \\
&= \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_{j^0}^k, \hat{\boldsymbol{x}}_{j^0}^k \rangle.
\end{aligned} \tag{73}$$

Therefore,

$$\hat{\boldsymbol{L}}_j^{0t}[a] = \boldsymbol{L}^t[j, a] + \sum_{j^0 2 C_{j,a}} \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_{j^0}^k, \hat{\boldsymbol{x}}_{j^0}^k \rangle, \tag{74}$$

which is equal to the cumulative counterfactual loss $\hat{\boldsymbol{L}}_j^t[a]$ in CFR-RM, i.e., (68) holds. Again, we have

$$\hat{\boldsymbol{x}}_j^{t+1} = \nabla {\psi_j}^t(-\hat{\boldsymbol{L}}_j^{0t}) = \nabla {\psi_j}^t(-\hat{\boldsymbol{L}}_j^t) = \frac{[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+}{\beta_j^t}, \tag{75}$$

where α_j^t satisfies $\|[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+\|_1 = \beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$. Since $\hat{\boldsymbol{R}}_j^t = \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k \rangle \mathbf{1} - \hat{\boldsymbol{L}}_j^t$, we have $\alpha_j^t = \sum_{k=1}^t \langle \hat{\boldsymbol{l}}_j^k, \hat{\boldsymbol{x}}_j^k \rangle$, and,

$$\hat{\boldsymbol{x}}_j^{t+1} = \frac{[\alpha_j^t \mathbf{1} - \hat{\boldsymbol{L}}_j^t]^+}{\beta_j^t} = \frac{[\hat{\boldsymbol{R}}_j^t]^+}{\|[\hat{\boldsymbol{R}}_j^t]^+\|_1}, \tag{76}$$

So, (67) holds. Since the local decisions between FD-FTRL and CFR-RM at all decision points are equal, we can conclude that CFR-RM is equivalent to a special case of FD-FTRL with $\beta_j^t = \|[\hat{\boldsymbol{R}}_j^t]^+\|_1$ at all decision points.

Now, we prove the equivalence between FD-OMD and CFR-RM+. We prove the equivalence by recursively proving that the following equations hold at all decision points:

$$\hat{\boldsymbol{l}}_j^{0t} = \hat{\boldsymbol{l}}_j^t, \tag{77}$$

$$\hat{\boldsymbol{x}}_j^{t+1} = \frac{\hat{\boldsymbol{Q}}_j^t}{\|\hat{\boldsymbol{Q}}_j^t\|_1}. \tag{78}$$

First, at decision point $j \in \mathcal{J}$ that has $D_j = 1$, we have $\hat{\boldsymbol{l}}_j^{0t} = \boldsymbol{l}^t[j] = \hat{\boldsymbol{l}}_j^t$. So, (77) holds. Then, according to Proposition 2.4, we have

$$\hat{\boldsymbol{x}}_j^{t+1} = \nabla {\psi_j}^t(\nabla \psi_j^{t\ 1}(\hat{\boldsymbol{x}}_j^t) - \hat{\boldsymbol{l}}_j^{0t}) = {\psi_j}^t(\nabla \psi_j^{t\ 1}(\hat{\boldsymbol{x}}_j^t) - \hat{\boldsymbol{l}}_j^t) = \frac{[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^t]^+}{\beta_j^t}, \tag{79}$$

where α_j^t fulfills the constraint $\|\hat{\boldsymbol{x}}_j^{t+1}\|_1 = 1$, i.e., $\|[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^t]^+\|_1 = \beta_j^t$. Note that α_j^t exists and is unique. Recall that, in CFR-RM+,

$$\hat{\boldsymbol{Q}}_j^t = [\hat{\boldsymbol{Q}}_j^{t\ 1} + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle - \hat{\boldsymbol{l}}_j^t]^+ = [\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_1 \hat{\boldsymbol{x}}_j^t + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle - \hat{\boldsymbol{l}}_j^t]^+. \tag{80}$$

So, when $\beta_j^{t\ 1} = \|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_1$ and $\beta_j^t = \|\hat{\boldsymbol{Q}}_j^t\|_1$, we can conclude that $\alpha_j^t = \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle$ and

$$\hat{\boldsymbol{x}}_j^{t+1} = \frac{[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^t]^+}{\beta_j^t} = \frac{[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle - \hat{\boldsymbol{l}}_j^t]^+}{\beta_j^t} = \frac{\hat{\boldsymbol{Q}}_j^t}{\|\hat{\boldsymbol{Q}}_j^t\|_1}. \tag{81}$$

So, (78) holds at every decision point $j \in \mathcal{J}$ that has $D_j = 1$.

Now, assume that (77) and (78) are satisfied at decision points whose depth is less than or equal to k. For a decision point $j \in \mathcal{J}$ that has $D_j = k + 1$, we have

$$\hat{\boldsymbol{l}}_j^{0t}[a] = \boldsymbol{l}^t[j, a] + \sum_{j' \sqsupset C_{j,a}} \hat{l}_{j'}^{0t}, \tag{82}$$

where

$$\begin{aligned}
\hat{l}_{j'}^{0t} =& {\psi_{j'}}^t(\nabla \psi_{j'}^{t\ 1}(\hat{\boldsymbol{x}}_{j'}^t)) - {\psi_{j'}}^t(\nabla \psi_{j'}^{t\ 1}(\hat{\boldsymbol{x}}_{j'}^t) - \hat{\boldsymbol{l}}_{j'}^{0t}) \\
=& \left(\langle \beta_{j'}^{t\ 1}\hat{\boldsymbol{x}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle - \frac{1}{2}\beta_{j'}^{t\ 1}\|\hat{\boldsymbol{x}}_{j'}^t\|_2^2 - \frac{1}{2}\beta_{j'}^{t\ 1}\|\hat{\boldsymbol{x}}_{j'}^t\|_2^2\right) - \left(\langle \beta_{j'}^{t\ 1}\hat{\boldsymbol{x}}_{j'}^t - \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^{t+1} \rangle - \frac{1}{2}\beta_{j'}^t\|\hat{\boldsymbol{x}}_{j'}^{t+1}\|_2^2 - \frac{1}{2}\beta_{j'}^t\|\hat{\boldsymbol{x}}_{j'}^{t+1}\|_2^2\right) \\
=& \langle \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle + \langle \hat{\boldsymbol{l}}_{j'}^t - \beta_{j'}^{t\ 1}\hat{\boldsymbol{x}}_{j'}^t - \langle \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle, \hat{\boldsymbol{x}}_{j'}^{t+1} \rangle + \beta_{j'}^t\|\hat{\boldsymbol{x}}_{j'}^{t+1}\|_2^2 \\
=& \langle \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle.
\end{aligned} \tag{83}$$

Therefore,

$$\hat{\boldsymbol{l}}_j^{0t}[a] = \boldsymbol{l}^t[j, a] + \sum_{j' \sqsupset C_{j,a}} \langle \hat{\boldsymbol{l}}_{j'}^t, \hat{\boldsymbol{x}}_{j'}^t \rangle, \tag{84}$$

which is equal to the counterfactual loss $\hat{\boldsymbol{l}}_j^t[a]$ in CFR-RM+, i.e., (77) holds. Again, we have

$$\hat{\boldsymbol{x}}_j^{t+1} = \nabla {\psi_j}^t(\nabla \psi_j^{t\ 1}(\hat{\boldsymbol{x}}_j^t) - \hat{\boldsymbol{l}}_j^{0t}) = \nabla {\psi_j}^t(\nabla \psi_j^{t\ 1}(\hat{\boldsymbol{x}}_j^t) - \hat{\boldsymbol{l}}_j^t) = \frac{[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^t]^+}{\beta_j^t}, \tag{85}$$

where α_j^t fulfills $\|[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t \mathbf{1} - \hat{\boldsymbol{l}}_j^t]^+\|_1 = \beta_j^t = \|\hat{\boldsymbol{Q}}_j^t\|_1$. Since $\hat{\boldsymbol{Q}}_j^t = [\hat{\boldsymbol{Q}}_j^{t\ 1} + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle - \hat{\boldsymbol{l}}_j^t]^+ = [\|\hat{\boldsymbol{Q}}_j^{t\ 1}\|_1 \hat{\boldsymbol{x}}_j^t + \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle - \hat{\boldsymbol{l}}_j^t]^+$, we have $\alpha_j^t = \langle \hat{\boldsymbol{l}}_j^t, \hat{\boldsymbol{x}}_j^t \rangle$ and

$$\hat{\boldsymbol{x}}_j^{t+1} = \frac{[\beta_j^{t\ 1}\hat{\boldsymbol{x}}_j^t + \alpha_j^t - \hat{\boldsymbol{l}}_j^t]^+}{\beta_j^t} = \frac{\hat{\boldsymbol{Q}}_j^t}{\|\hat{\boldsymbol{Q}}_j^t\|_1}, \tag{86}$$

So, (78) holds. Since the local decisions between FD-OMD and CFR-RM+ at all decision points are equal, we can conclude that CFR-RM+ is equivalent to a special case of FD-OMD with $\beta_j^t = \|\hat{\boldsymbol{Q}}_j^t\|_1$ at all decision points. □

E.3. Proof for Corollary 3.9

We first present a lemma from (Orabona, 2019). For completeness, the proof is also quoted.

Lemma E.1. *(Orabona, 2019) Let $a_0 \geq 0$ and $f : [0, +\infty) \to [0, +\infty)$ a non-increasing function. Then*

$$\sum_{t=1}^{T} a_t f\left(a_0 + \sum_{k=1}^{t} a_k\right) \leq \int_{a_0}^{\sum_{t=0}^{T} a_t} f(x)dx. \tag{87}$$

Proof. (Orabona, 2019). Denote by $s_t = \sum_{k=0}^{t} a_k$.

$$a_t f\left(a_0 + \sum_{k=1}^{t} a_k\right) = a_t f(s_t) \leq \int_{s_{t-1}}^{s_t} f(x)dx. \tag{88}$$

Summing over $t = 1, \ldots, T$, we have the stated bound. □

Proof. According to Theorem 3.5, the total regret is

$$\begin{aligned} R^T &\leq \sum_{j \in J} \left(\beta_j^T + \sum_{t=1}^{T} \frac{\left[\|\hat{\boldsymbol{r}}_j^t\|_2^2 + \lambda_j^{t-1} - \lambda_j^t\right]^+}{2\beta_j^t}\right) \\ &\leq \sum_{j \in J} \left(\beta_j^T + \sum_{t=1}^{T} \frac{\|\hat{\boldsymbol{r}}_j^t\|_2^2}{2\beta_j^t}\right) \\ &\leq \sum_{j \in J} \sqrt{n_j \lambda_j^T} + \sum_{j \in J} \sum_{t=1}^{T} \frac{\|\hat{\boldsymbol{r}}_j^t\|_2^2}{2\sqrt{\lambda_j^t}}. \end{aligned} \tag{89}$$

The first and the last inequality are because $\beta_j^t = \sqrt{\lambda_j^t}/\|\hat{\boldsymbol{x}}_j^{t+1}\|_2$. Since $\eta \sum_{k=1}^{t} \|\hat{\boldsymbol{r}}_j^k\|_2^2 \leq \lambda_j^t$, we have

$$\sum_{t=1}^{T} \frac{\|\hat{\boldsymbol{r}}_j^t\|_2^2}{2\sqrt{\lambda_j^t}} \leq \frac{1}{2\sqrt{\eta}} \sum_{t=1}^{T} \frac{\|\hat{\boldsymbol{r}}_j^t\|_2^2}{\sqrt{\sum_{k=1}^{t} \|\hat{\boldsymbol{r}}_j^k\|_2^2}} \leq \frac{1}{\sqrt{\eta}} \sqrt{\sum_{k=1}^{T} \|\hat{\boldsymbol{r}}_j^k\|_2^2} \leq \frac{1}{\eta} \sqrt{\lambda_j^T}. \tag{90}$$

The second inequality is because of Lemma E.1. So,

$$R^T \leq \sum_{j \in J} \left(\sqrt{n_j \lambda_j^T} + \frac{1}{\eta} \sqrt{\lambda_j^T}\right) = \sum_{j \in J} \left(\sqrt{n_j} + \frac{1}{\eta}\right) \sqrt{\lambda_j^T}. \tag{91}$$

□

F. Benchmark Games

In this section, we describe the benchmark games used in the experiments.

F.1. Description of the Games

- **Leduc** (Southey et al., 2012) is a two-player zero-sum EFG. It can be considered as a simplified Heads-up Limit Texas Hold'em (HULH).[8] In Leduc, there are two suits cards, each suit comprises three ranks, and two rounds of betting are allowed. At the beginning of the game, each player places an ante of one chip in the pot and is dealt with one card, which is only visible to itself. In the first round of betting, player 1 has to choose an action between *Call* and *Raise*. Taking the action Call means that the player will place or has placed the same chips as the opponent and leaves the choice to the opponent. Taking the action Raise means that the player will place more chips than the opponent to the pot.

[8] https://en.wikipedia.org/wiki/Texas_hold_%27em

It is only two raises allowed in the first round of betting. Sometimes when a player bets fewer chips than his opponent and is asked to take an action, he can choose to *Fold*. If he does so, the game is over, and the player loses all chips. The first round ends if one of the players has chosen Fold or if both players agree to end. If no player folds, a public card is revealed to both of the players and then the second round of betting takes place, with the same dynamic as the first round. After the two rounds of betting, if one of the players has a pair with the public card, that player wins the pot. Otherwise, the player with a higher private card wins. In the first round, the player taking action Raise should place 2 (named raise size) more chips to the pot than the opponent. In the second round, the raise size is 4.

- **Leduc(2, 9)** has the same rules as Leduc, except that it has two suits of cards with each suit comprises nine ranks.
- **FHP(2, 5)** (Brown et al., 2019) is a simplified HULH, and it has two suits of cards with each suit comprises five ranks. At the beginning of the game, each player places an ante of 50 chips in the pot and is dealt with two cards. It has the same dynamic in each round of betting as Leduc. However, it allows three raises and a raise size of 100 in each round of betting, and the first player to act in the second round is player 2. After the first round of betting, three public cards are revealed to the players. After two rounds of betting, the rank of the 5 cards (2 private cards + 3 public cards) of each player is evaluated, and the player with the higher rank wins the pot. We use the standard hand evaluating method[9] used in HULH.
- **Goofspiel** (Ross, 1971) is popular benchmark EFG. The game has three suits of cards with each suit comprises five ranks. At the beginning of the game. Each player is dealt with one suit of private cards. Another suit of cards is served as the prize and kept face down on the desk. At each round of the game, the topmost prize card is revealed. Then the players are asked to select a card from their own cards and reveal the cards simultaneously. The player who has a higher card wins the prize card. If the ranks are equal, the prize card is split. After all prize cards are revealed, the score of each player is the sum of the rank of the prize cards it won. Then, the payoff for the player with the higher score is $+1$, and the payoff for the opponent is -1. If the scores are the same, the payoffs for the players are $(0.5, 0.5)$.
- **Goofspiel 4** is the same as Goofspiel, except that Goofspiel 4 has three suits of cards with each suit comprises four ranks.
- **Goofspiel 4 (imp)** Goofspiel 4 (imp), i.e., Goofspiel 4 with imperfect information, is a variant of Goofspiel 4, which is first proposed in (Lanctot et al., 2009). In each round, the cards that the players selected are not revealed to each other. Instead, a coordinator will see the cards and determine which player wins the prize card.
- **Liar's dice** (Lisý et al., 2015) is another popular benchmark EFG. At the beginning of the game. Each player secretly rolls a dice. Then the first player claims the outcome of their roll, in the form of **Q**uantity-**V**alue, e.g., a claim of 1-2 means that the player believes that there is *one* dice with a face value of *2*. The second player can make a higher claim: either to claim a higher Q with any V or to claim the same Q with a higher V. Otherwise, he can call his opponent a "liar". When the player calls a liar, the dice are revealed. If the opponent's claim is false, the player that called a liar on the opponent wins the game; otherwise, the player loses the game.
- **Battleship** (Farina et al., 2019c) is a classic board game. At the beginning of the game, each player secretly places a set of ships on separate grids without overlapping. In this setting, the size of the ship is 1×2 and the value is 4. The size of the grid for each player is 3×2. After all the ships are placed. The players take turns to fire the opponent's ship. Ships that have been hit at all their cells are considered sunk. The game ends if a player's all ships are all sunk or each of the players has completed 3 shots. Finally, the payoff of the player is calculated as the sum of the values of the opponent's ships that were sunk, minus the sum of the values of the player's ships that were lost.

F.2. Measuring the Sizes of the Games

To better measure the sizes of the games, another model known as the *history tree* is used to describe the games. For a two-player imperfect information EFG, we have two players $P = \{1, 2\}$. A special player, the *chance* player c, is introduced to take action for random events in the game. A history $h \in H$ (i.e., a full information state) is represented as a string of actions that taken by all the players and the chance player. The histories of an EFG form a history tree by nature. Given a non-terminal history $h \in H$, $A(h) \subseteq \mathcal{A}$ gives the set of legal actions, and $P(h) \in P \cup \{c\}$ gives the player that should act in the history. If action $a \in A(h)$ leads from h to h^0, then $h^0 = ha$. Denote the set of histories that are reached immediately after taking any action $a \in A(h)$ for h as $C_h = \{h^0 | h^0 = ha, a \in A(h)\}$. In an imperfect information game, histories of

[9] https://en.wikipedia.org/wiki/List_of_poker_hands

player $p \in P$ that are indistinguishable are collected into an *information set* (infoset) $I \in \mathcal{I}_p$. According to the definition of decision points, we know that infosets are actually decision points. So, $\mathcal{J}_p = \mathcal{I}_p$ for any player $p \in P$.

With the above definitions, we can measure the sizes of the games in many dimensions. In Table 3, we give some data about the games. In the table, #Histories measures the number of histories of the game. Depth measures the depth of the history tree of the game, i.e., the maximum length of the histories. #Leaves measures the number of leaves of the history tree. The size of decisions measures the maximum number of histories that belong to the same decision point (infoset). The action factor, denoted by m, measures the contribution of the players to the number of leaves. We define m recursively: if history h belongs to $p \in P$, then $m_h = \sum_{h' 2 C_h} m_{h'}$; otherwise, $m_h = \max_{h' 2 C_h} m_{h'}$. Then, the action factor m is defined as $m = m_o$. The stochastic factor, denoted by s, measures the contribution of the chance player to the number of leaves. The stochastic factor is defined as follows: if history h belongs to $p \in P$, $s_h = \max_{h' 2 C_h} s_{h'}$; otherwise, $s_h = \sum_{h' 2 C_h} s_{h'}$. Then, the stochastic factor s is defined as $s = s_o$. Note that $m \times s \approx$ #Leaves.

Table 3: Sizes of the games

Game	#Decision points	#Histories	#Leaves	Depth	size of decisions	Action factor	Stochastic factor
Leduc	936	3780	5520	8	5	49	120
Leduc(2, 9)	9288	1.5×10^5	2.2×10^5	8	17	49	4896
Goof. 4	7304	1.1×10^4	1.4×10^4	6	4	576	24
Goof. 4 (imp)	3608	1.1×10^4	1.4×10^4	6	14	576	24
FHP(2, 5)	1.4×10^5	1.4×10^6	2.3×10^6	12	9	98	2.5×10^4
Goofspiel	9.1×10^5	1.4×10^6	1.7×10^6	8	5	1.4×10^4	36
Liar's dice	2.5×10^4	1.5×10^5	1.5×10^5	13	6	4095	24
Battleship	4.1×10^5	2.3×10^6	7.0×10^6	10	22	7.0×10^6	1

G. Additional Results

G.1. Full Results in the Benchmark Games

In Figure 4, we compare FD-FTRL(CFR), FD-OMD(CFR) with CFRs in eight games. As we can see, FD-FTRL(CFR) recovers vanilla CFR and FD-OMD(CFR) recovers CFR+. So, the equivalences are verified empirically. Note that both FD-OMD(CFR) and CFR+ use **LA** for computing the average strategies.

In Figure 5, we compare FD-FTRL(R) in different configurations in eight games. As we have stated in the paper, the default FD-FTRL(R) (with **LA** and **CW**) has the fastest convergence rate. In Figure 6, we compare FD-OMD(R) in different configurations in eight games. Similar to the conclusion for FD-FTRL(R), the default FD-OMD(R) is the fastest, it is even faster than CFR+ in four games. In Figure 7 and 8, we perform an ablation study for FD-FTRL(R) and FD-OMD(R), respectively. As we have stated in the paper, **CW** has a significant impact on performance. Sometimes, FD-FTRL(R˜FD) is faster than the default FD-FTRL(R). However, FD-OMD(R) is always faster than the other variants.

In Figure 9, we compare FD-FTRL(R) and FD-OMD(R) with Predicted CFR (PCFR) and Predicted CFR+ (PCFR+) (Farina et al., 2021). We implement PCFR with **UA** and implement PCFR+ with **LA**. As we can see, FD-OMD(R) is still competitive compared with PCFR+.

In Figure 10, the results of conventional FTRL, OMD, and vanilla CFR that use **UA** to compute the average strategies are given. As we can see, both FD-FTRL(R) and FD-OMD(R) are always faster than them.

G.2. Hyper-Parameter Tuning

For each game, We perform a coarse hyper-parameter (η) tuning for FD-FTRL(R) and FD-OMD(R). For most of the games, we choose η in $\{0.1, 0.01, 10^{-3}, 10^{-4}, 10^{-5}\}$. In Figure 11 (12), the results of FD-FTRL(R) (FD-OMD(R)) with different hyper-parameters are given. As we can see, FD-OMD(R) is more sensitive to the hyper-parameter than FD-FTRL(R). In (Liu et al., 2022), the authors have proposed a method for adapting the hyper-parameter in ReCFR, which may be available for adapting the hyper-parameters in FD-FTRL(R) and FD-OMD(R), too.

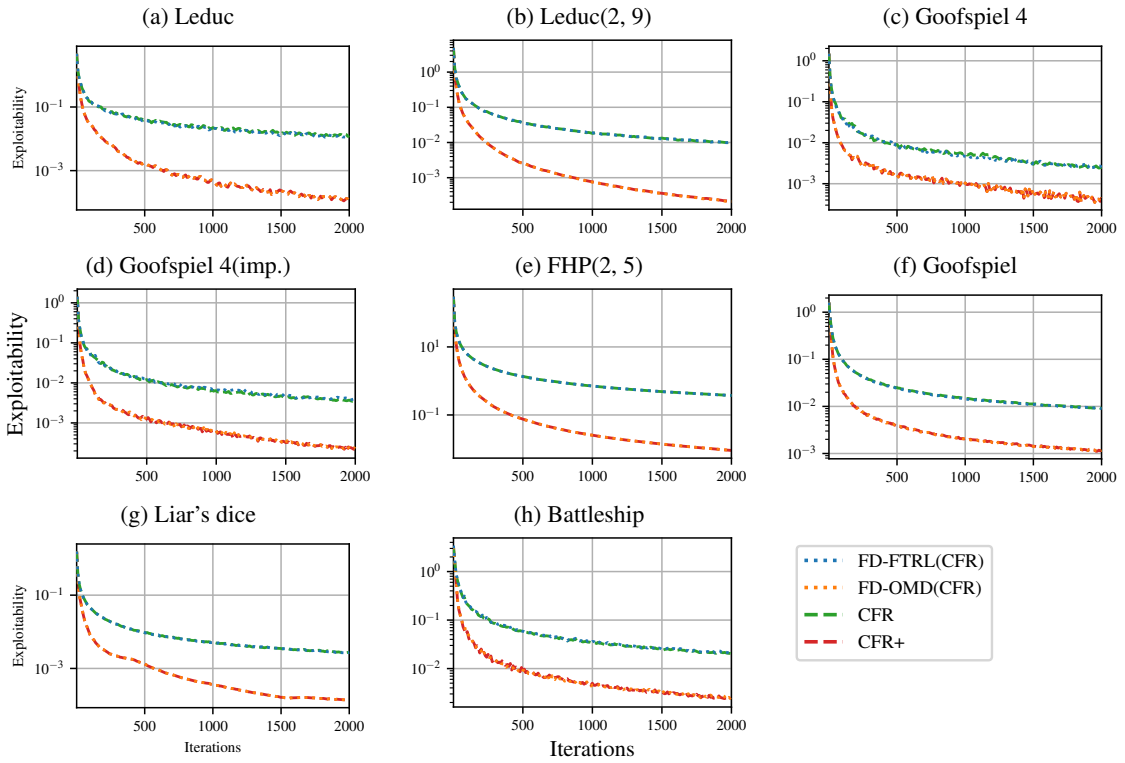

Figure 4: Exploitability curves of FD-FTRL(CFR), FD-OMD(CFR), and CFRs in eight games.

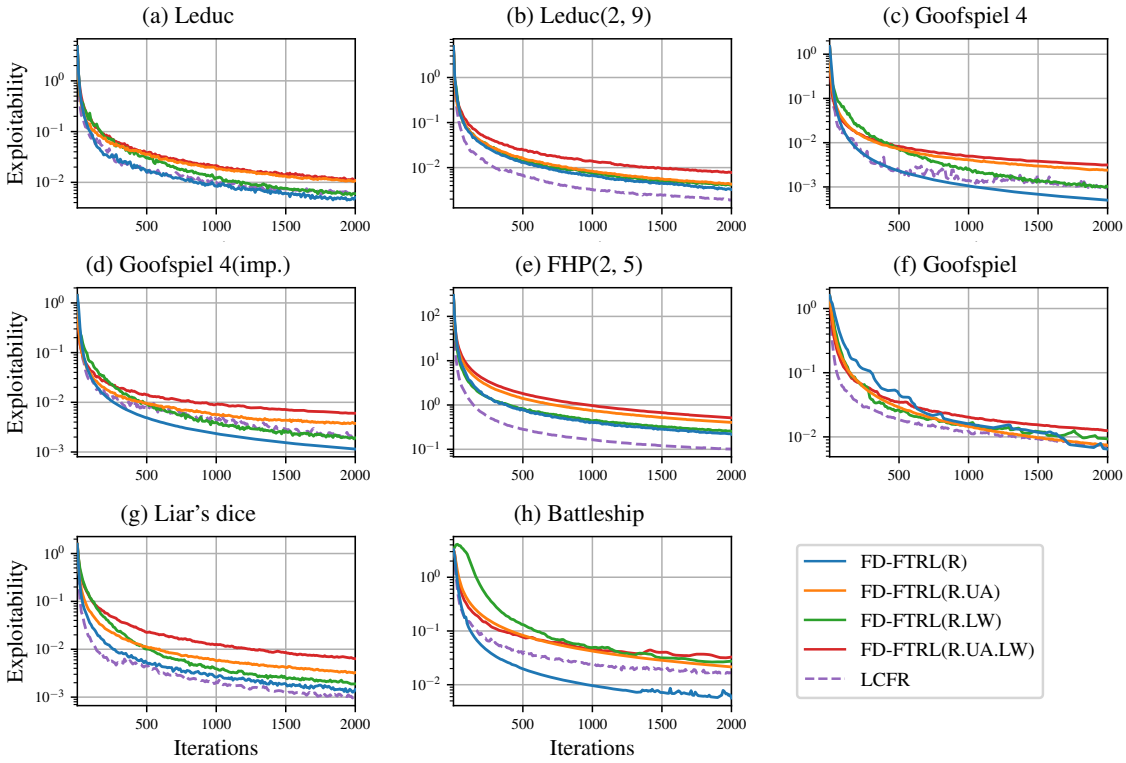

Figure 5: Exploitability curves of FD-FTRL(R) in different configurations in eight games.

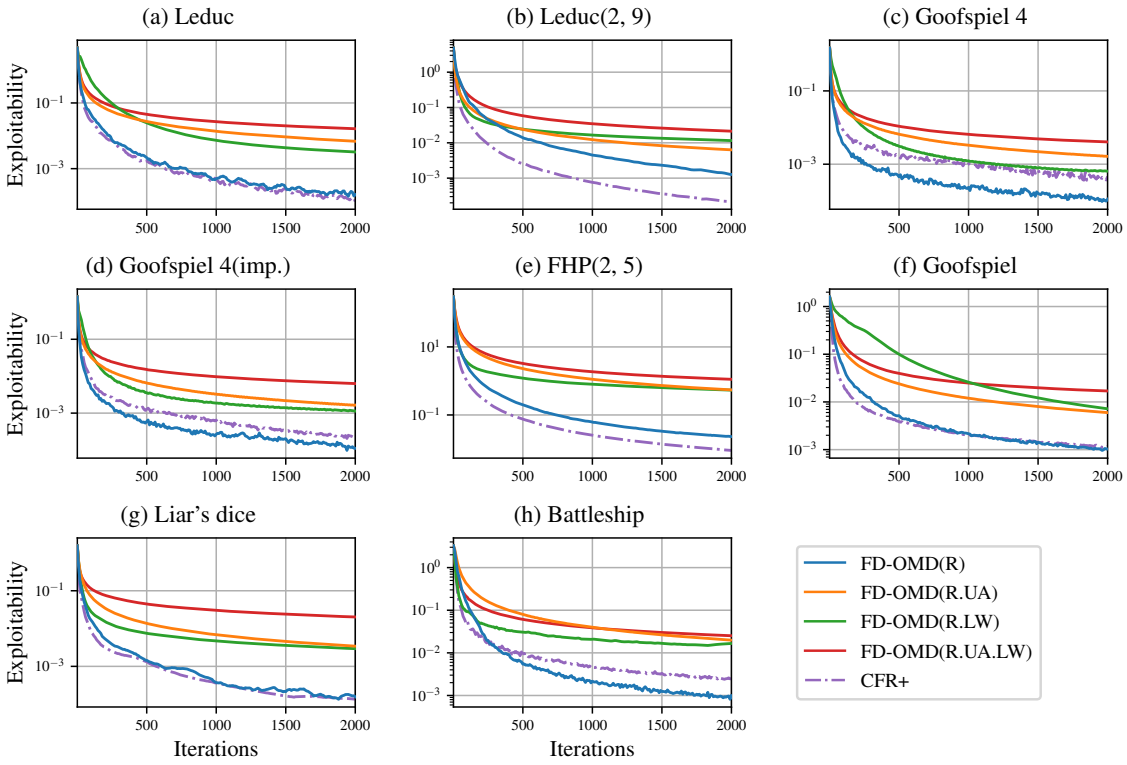

Figure 6: Exploitability curves of FD-OMD(R) in different configurations in eight games.

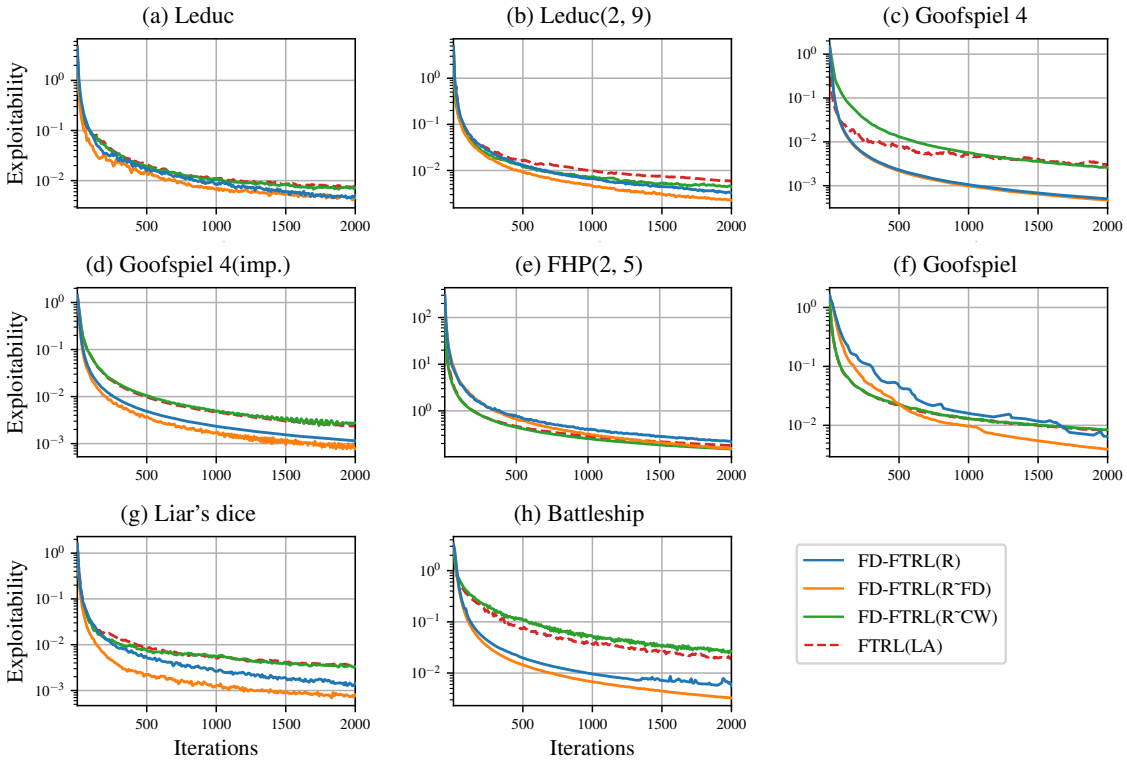

Figure 7: Exploitability curves of FD-FTRL(R) with or without certain components in eight games.

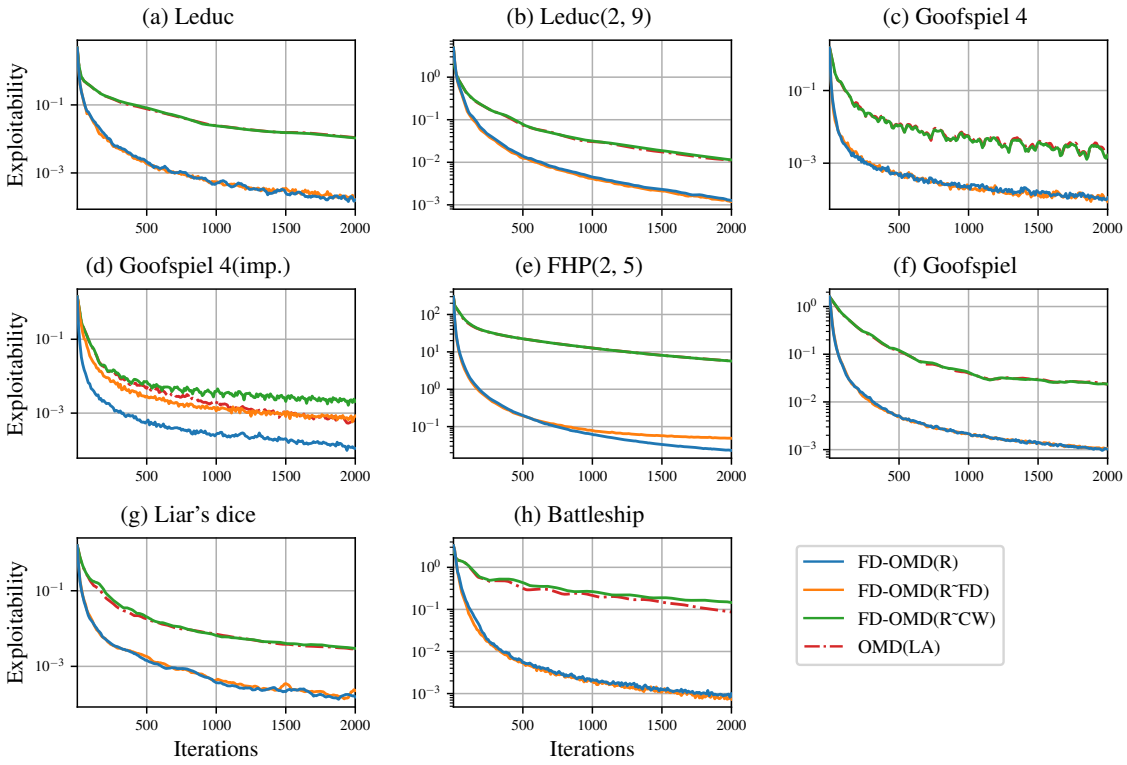

Figure 8: Exploitability curves of FD-OMD(R) with or without certain components in eight games.

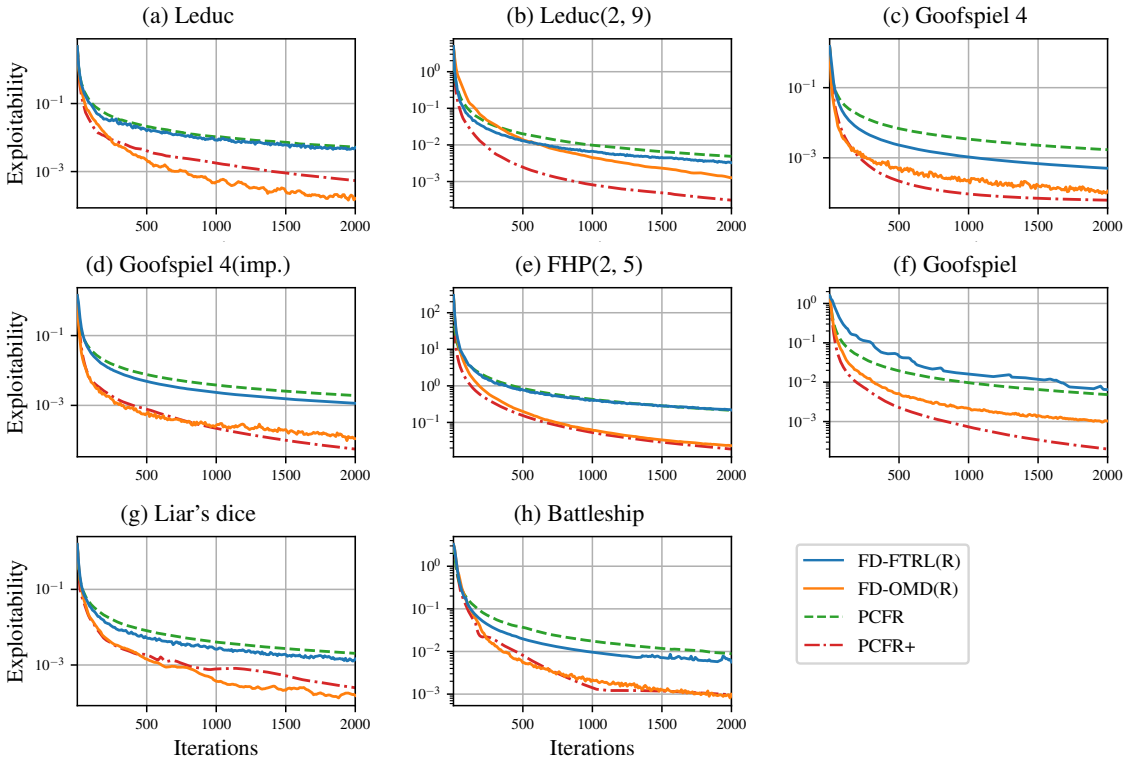

Figure 9: Exploitability curves of FD-FTRL(R), FD-OMD(R), PCFR, and PCFR+ in eight games.

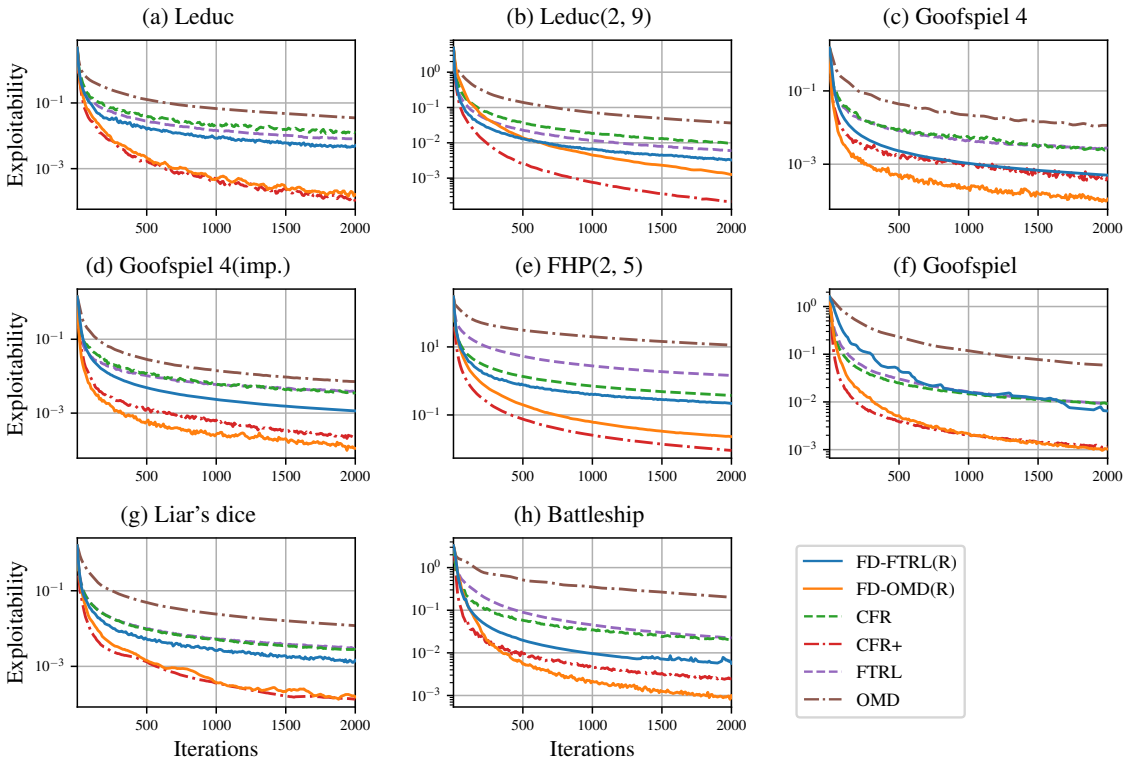

Figure 10: Exploitability curves of FD-FTRL(R), FD-OMD(R), FTRL, OMD, and CFRs in eight games.

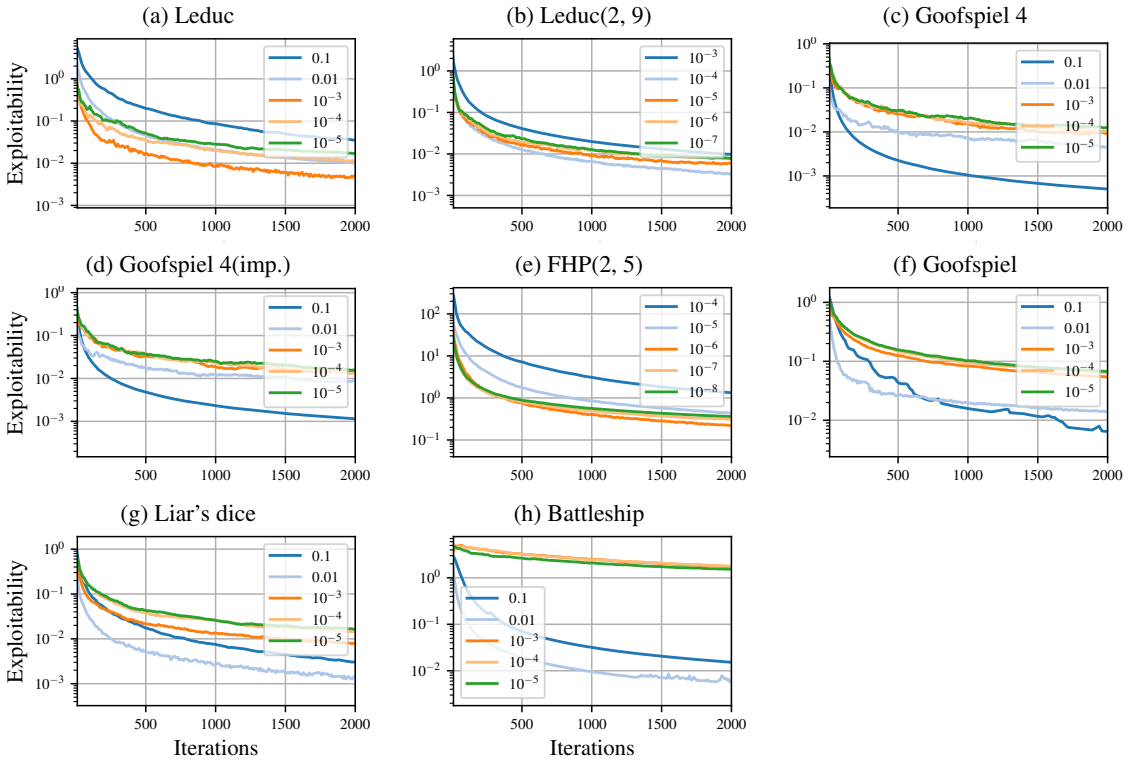

Figure 11: FD-FTRL(R) with different hyper-parameters in eight games.

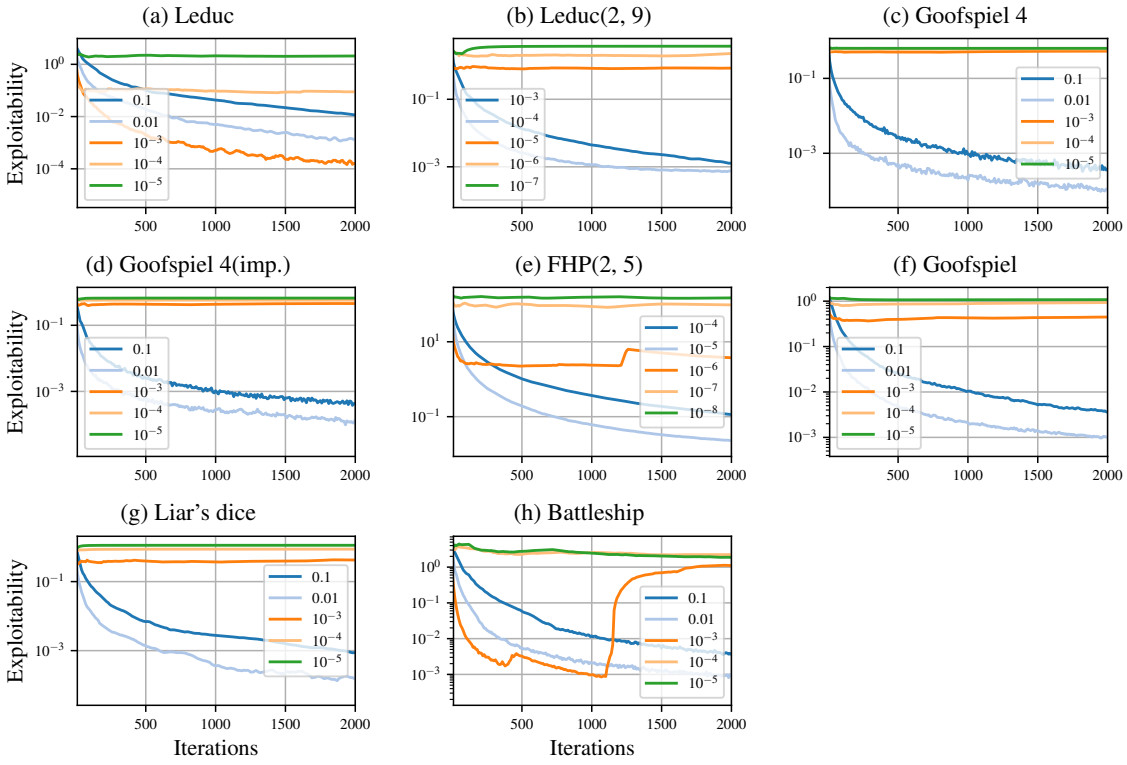

Figure 12: FD-OMD(R) with different hyper-parameters in eight games.