Proceedings of Machine Learning Research vol 291:1–47, 2025 38th Annual Conference on Learning Theory

Corrupted Learning Dynamics in Games

Taira Tsuchiya TSUCHIYA@MIST.I.U-TOKYO.AC.JP
The University of Tokyo and RIKEN

Shinji Ito SHINJI@MIST.I.U-TOKYO.AC.JP
The University of Tokyo and RIKEN

Haipeng Luo HAIPENGL@USC.EDU
University of Southern California

Editors: Nika Haghtalab and Ankur Moitra

Abstract

Learning in games refers to scenarios where multiple players interact in a shared environment, each aiming to minimize their regret. It is well known that an equilibrium can be computed at a fast rate of $O(1/T)$ when all players follow the optimistic follow-the-regularized-leader (OFTRL). However, this acceleration is limited to the *honest regime*, in which all players fully adhere to a prescribed algorithm—a situation that may not be realistic in practice. To address this issue, we present *corrupted learning dynamics* that adaptively find an equilibrium at a rate that depends on the extent to which each player deviates from the strategy suggested by the prescribed algorithm. First, in two-player zero-sum corrupted games, we provide learning dynamics for which the external regret of the x-player (and similarly for the y-player) is roughly bounded by $O(\log(m_x m_y) + \sqrt{\widehat{C}_y} + \widehat{C}_x)$, where m_x and m_y denote the number of actions of the x- and y-players, respectively, and $\widehat{C}_x$ and $\widehat{C}_y$ represent their cumulative deviations. We then extend our approach to multiplayer general-sum corrupted games, providing learning dynamics for which the swap regret of player i is bounded by $O(\log T + \sqrt{\sum_k \widehat{C}_k \log T} + \widehat{C}_i)$ ignoring dependence on the number of players and actions, where $\widehat{C}_i$ is the cumulative deviation of player i from the prescribed algorithm. Our learning dynamics are agnostic to the levels of corruption. A key technical contribution is a new analysis that ensures the stability of a stationary distribution of a Markov chain under a new adaptive learning rate, thereby allowing us to achieve the desired bound in the corrupted regime while matching the best existing bound in the honest regime. Notably, our framework can be extended to address not only corruption in strategies but also corruption in the observed expected utilities, and we provide several matching lower bounds.

Keywords: learning in games, optimistic follow-the-regularized-leader, external/swap regret minimization, Nash/correlated equilibrium

1. Introduction

Learning in games refers to settings where multiple players interact in a shared environment, each aiming to minimize their regret by iteratively adapting their strategies based on repeated interactions (Freund and Schapire, 1999; Hart and Mas-Colell, 2000). Each player can minimize their own regret using the framework of *online learning*, which allows us to minimize the cumulative loss based on sequentially obtained past observations (Cesa-Bianchi and Lugosi, 2006). It is well known that an approximate equilibrium can be obtained when all players employ no-regret algorithms. For instance, in two-player zero-sum games, if each player employs an online learning algorithm with regret Reg^T, then an $O(\mathsf{Reg}^T/T)$-approximate Nash equilibrium is attained after T rounds. Since many

Table 1: Comparison of individual (external) regret upper bounds of the x-player in two-player zero-sum games with a payoff matrix $A \in [-1,1]^{m_x \times m_y}$ after T rounds. The variables $\widehat{C}_x, \widehat{C}_y \in [0, 2T]$ are the amount of corruption in strategies for the x- and y-players, respectively, $g^{(t)} \in [-1,1]^{m_x}$ is a utility vector for the x-player, and $P_\infty^T(g) = \sum_{t=1}^T \|g^{(t)} - g^{(t-1)}\|_\infty^2$.

References	Honest	Corrupted (no corruption in observed utilities)
Rakhlin and Sridharan (2013b)	$\log(m_x m_y T)$	$\sqrt{P_\infty^T(g)}\log(m_x T) + \widehat{C}_x$
Kangarshahi et al. (2018)	$\log(m_x m_y T)$	$\sqrt{T \log m_x} + \widehat{C}_x$
Syrgkanis et al. (2015)	$\log(m_x m_y)$	$\log(m_x m_y) + \sqrt{T \log m_x} + \widehat{C}_x$
This work (Theorem 10)	$\sqrt{\log(m_x m_y) \log m_x}$	$\min\left\{\sqrt{(\log(m_x m_y) + \widehat{C}_x + \widehat{C}_y)\log m_x}, \sqrt{P_\infty^T(g) \log m_x}\right\} + \widehat{C}_x$

online learning algorithms achieve regret bounds of $\mathsf{Reg}^T = O(\sqrt{T})$, this implies that an $O(1/\sqrt{T})$-approximate equilibrium is obtained after T rounds. In online learning, the regret bound of $\mathsf{Reg}^T = O(\sqrt{T})$ is generally unimprovable even for stochastic losses (Cesa-Bianchi and Lugosi, 2006).

However, in the context of learning in games, the observations of each player are determined by the strategies of their opponents. By leveraging this property, it has been shown that the $O(\sqrt{T})$ individual regret upper bounds, and the corresponding $O(1/\sqrt{T})$ convergence rates, can be significantly improved, a phenomenon first observed by Daskalakis et al. (2011). A prominent approach to achieving fast rates is the use of optimistic prediction, which involves predicting the next observation based on past observations. These include optimistic follow-the-regularized-leader (OFTRL) and optimistic online mirror descent (Rakhlin and Sridharan, 2013a,b; Syrgkanis et al., 2015). For example, in two-player zero-sum games, if both players employ a specific OFTRL algorithm with a constant learning rate, each player can achieve a regret bound of $O(\log(m_x m_y))$, independent of T, where m_x and m_y denote the number of actions of the x- and y-players, respectively (Syrgkanis et al., 2015).

However, such improvements can only be achieved in the *honest regime*, where all players fully adhere to the prescribed algorithm. A player may deviate from the output of the prescribed algorithm, for example, to address their own constraints or to amplify the regret of other players. When an opponent's sequence of actions deviates entirely from the algorithm's output—a scenario often referred to as the adversarial regime in the literature—approaches that rely on the assumption of perfect adherence to the prescribed algorithm may fail to achieve sublinear regret.

Several approaches are known to address such adversarial opponents. For example, in two-player zero-sum games, by appropriately switching algorithms, it is possible to simultaneously achieve regret bounds of $O(\log(m_x m_y))$ in the honest regime and $\widetilde{O}(\sqrt{T})$ against adversarial opponents (Syrgkanis et al., 2015). However, such theoretical guarantees in the adversarial scenario can be very pessimistic. Even when a player's strategies deviate only slightly from the output of the prescribed algorithm, we can only ensure the $\widetilde{O}(\sqrt{T})$ bound. For example, an opposing player might initially fail to follow the prescribed algorithm (in which case the algorithm in Syrgkanis et al. 2015 switches to one designed to handle dishonest players), but later revert to honest behavior. Ideally, it is desirable to design learning dynamics that achieve regret bounds that smoothly bridge the $\widetilde{O}(1)$ bound in the honest regime and the $\widetilde{O}(\sqrt{T})$ worst-case bound.

Our contributions To achieve this goal, we present *corrupted learning dynamics* that yield regret upper bounds which adapt to the degree of deviation from the prescribed algorithm. We introduce the *corrupted regime* for n-player games, characterized by corruption levels $\{\widehat{C}_i\}_{i \in [n]}$. In the corrupted

Table 2: Comparison of individual swap regret upper bounds of player i in multiplayer general-sum games with n-players and m-actions after T rounds. The variable $\widehat{C}_i \in [0, 2T]$ is the cumulative amount of corruption in strategies for player i, and $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$.

References	Honest	Corrupted (no corruption in observed utilities)
Chen and Peng (2020)	$\sqrt{n}(m \log m)^{3/4} T^{1/4}$	$\sqrt{mT \log m} + \widehat{C}_i$
Anagnostides et al. (2022a)	$nm^4 \log m \log^4 T$	N/A
Anagnostides et al. (2022b)	$nm^{5/2} \log T$	$nm^{5/2} \log T + \sqrt{Tm \log m} + \widehat{C}_i$
This work (Theorem 13)	$nm^{5/2} \log T$	$nm^{5/2} \log T + \min\left\{\sqrt{\widehat{S}(nm^2 + m^{5/2}) \log T}, m\sqrt{T \log T}\right\} + \widehat{C}_i$

regime, each player $i \in [n]$ accumulates a deviation of $\widehat{C}_i \geq 0$ from the prescribed algorithm. When $\widehat{C}_i = 0$ for all i, this regime corresponds to the honest regime. See Section 3 for details.

We propose learning dynamics that can effectively handle the corrupted regime for both two-player zero-sum games and multiplayer general-sum games. The proposed learning dynamics also achieve regret bounds that are comparable to or even better than the best existing dynamics in the honest regime. Our dynamics builds on OFTRL with adaptive learning rates. For each problem setup, we provide the following regret bounds and corresponding convergence rates to an equilibrium.

Two-player zero-sum games We start with two-player zero-sum games and provide learning dynamics with the following regret guarantees:

Theorem 1 (Informal version of Theorem 10) *In two-player zero-sum games, there exists $(\widehat{C}_x, \widehat{C}_y)$-agnostic learning dynamics such that the external regret of the x-player is bounded by $\sqrt{\log(m_x) \log(m_x m_y)}$ in the honest regime and by $\min\left\{\sqrt{\log(m_x)\left(\log(m_x m_y) + \widehat{C}_x + \widehat{C}_y\right)}, \sqrt{T \log m_x}\right\} + \widehat{C}_x$ in the corrupted regime, where $\widehat{C}_x$ and $\widehat{C}_y$ are the cumulative amount of corruption in strategies for the x- and y-players, respectively. The external regret of the y-player is bounded by a similar quantity.*

This is the first external regret upper bound for the corrupted regime. Note that in the literature, the *adversarial regime for the x-player* typically refers to the case where $\widehat{C}_x = 0$ and $\widehat{C}_y$ is arbitrary, whereas here we provide a bound for the more general setting with $\widehat{C}_x \geq 0$. This regret guarantee incentivizes players to follow the prescribed dynamics: any deviation by an opponent from the algorithm's output incurs only a square-root penalty, whereas a deviation by a player from the prescribed algorithm incurs a linear penalty. A comparison with existing bounds is provided in Table 1. From the regret bound in Theorem 1 and the relationship between no-external-regret learning in online linear optimization and an (approximate) Nash equilibrium in Theorem 4 (see next section), it follows that the average play after T rounds is an $\widetilde{O}((\widehat{C}_x + \widehat{C}_y)/T)$-approximate Nash equilibrium in the corrupted regime, where we ignore the dependence on m_x and m_y.

Multiplayer general-sum games Building on the analysis for two-player zero-sum games, we propose new corrupted learning dynamics for multiplayer general-sum games. In multiplayer general-sum games, to obtain a *correlated equilibrium* (Aumann, 1974; Foster and Vohra, 1997; Hart and Mas-Colell, 2000), we focus on minimizing the *swap regret* to get the following guarantees:

Theorem 2 (Informal version of Theorem 13) *In multiplayer general-sum games with n-players and m-actions, there exists $\{\widehat{C}_i\}_{i \in [n]}$-agnostic learning dynamics such that the swap regret of each player i is at most $nm^{5/2} \log T$ in the honest regime and $nm^{5/2} \log T + \min\Big\{\sqrt{\widehat{S}(nm^2 + m^{5/2}) \log T}$,*

$m\sqrt{T \log T}\Big\} + \widehat{C}_i$ *in the corrupted regime, where* $\widehat{C}_i$ *is the cumulative amount of corruption in strategies of player* i *and* $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$.

This is the first swap regret upper bound for the corrupted regime. Note again that in the literature, the adversarial regime for player i typically refers to the corrupted regime where $\widehat{C}_i = 0$ and $\widehat{C}_j$ is arbitrary for all $j \neq i$, whereas here we provide upper bounds for the more general setting with $\widehat{C}_i \geq 0$. Compared to the best swap regret bound by Anagnostides et al. (2022b), our algorithm achieves the same bound in the honest regime, a new adaptive bound in the corrupted regime in terms of $\widehat{S}$, and a worst-case $\sqrt{T}$-type bound that is $\sqrt{m}$ time worse than theirs (which is due to the use of the adaptive learning rate). A comparison with existing swap regret bounds is summarized in Table 2. Our algorithm is a variant of the algorithm by Anagnostides et al. (2022b), who use OFTRL with a constant learning rate (Syrgkanis et al., 2015) and the reduction of swap regret minimization to external regret minimization (Blum and Mansour, 2007). From the swap regret upper bound in Theorem 2 and the relationship between no-swap-regret learning in online linear optimization and a correlated equilibrium in Theorem 6 (see next section), it follows that when the above algorithm is used by all players, the time-averaged history of joint play after T rounds is an $O\big((\log T + \sqrt{\widehat{S} \log T} + \max_{k \in [n]} \widehat{C}_k)/T\big)$-approximate correlated equilibrium in the corrupted regime, where we ignore the dependence on n and m.

Achieving the swap regret bound in the corrupted regime requires a new analysis of OFTRL and the stability of stationary distribution of Markov chains. The existing analysis in Anagnostides et al. (2022b) relies heavily on the fact that OFTRL uses a constant learning rate, and thus we cannot directly employ it when we use an adaptive learning rate. By setting the learning rate sufficiently small and applying an analysis similar to that of two-player zero-sum games to prove Theorem 1, as well as the analysis in Wei and Luo (2018), we can show a swap regret bound of $O(nm^8 \log T)$ in the honest regime. However, the dependence on m of this upper bound is significantly worse than the $O(nm^{5/2} \log T)$ bound by Anagnostides et al. (2022b).

To address this issue, we leverage the fact that swap regret minimization can be achieved via multiple no-external-regret algorithms (Blum and Mansour, 2007). Specifically, we construct a transition probability matrix from the outputs of the external regret minimizer for each action and adopt the stationary distribution of the resulting Markov chain as the final strategy. We consider analyzing the stability of the transition probability matrix exploiting this property (Lemma 30), which leads to the significant improvement in the dependence on the number of actions m.

Our analysis, like previous work, leverages the RVU (Regret bounded by Variation in Utilities) bound for OFTRL, which was shown to be useful to obtain fast rates by Syrgkanis et al. (2015). In multiplayer general-sum games, we require an RVU bound for OFTRL with the log-barrier regularizer, which is a special case of a self-concordant barrier. However, RVU bounds for self-concordant barriers are only known for constant learning rates (Anagnostides et al., 2022b). To the best of our knowledge, we are the first to derive RVU bounds for OFTRL with a self-concordant barrier and an adaptive learning rate, which may be of independent interest (see Lemmas 23 and 24).

Our problem setting and analysis can be extended to account for corruption not only in the strategies but also in the observed expected utilities. This setup plays a crucial role in practical applications. For a detailed definition of the problem setting and an explanation of scenarios in which such corruption occurs, see Section 3. Finally, at the end of this paper, we provide several matching lower bounds (Theorem 15) that depend on the degree of corruption in strategies and utilities, offering a partial characterization of corrupted learning (with a complete characterization left for future work).

It is worth mentioning that our dynamics is *strongly uncoupled* (Hart and Mas-Colell, 2000; Daskalakis et al., 2011), meaning that each player requires no prior information about the game and determines their strategies solely based on past observations gathered through game interactions. Naturally, players are unaware of other players' utilities, and there is no communication between players.

2. Preliminaries

Notation and conventions For a natural number $n \in \mathbb{N}$, we use $[n] = \{1, \ldots, n\}$. Let $\mathbf{0}$ and $\mathbf{1}$ denote the zero vector and one vector with all entries equal to 0 and 1, respectively. Given a vector x, we use $x(a)$ to denote its a-th element and use $\|x\|_p$ to denote its ℓ_p-norm for $p \in [1, \infty]$. For a matrix A, we use $A(k, \cdot)$ to denote its k-th row. We use $\Delta(\mathcal{K})$ to denote the set of all probability distributions over $\mathcal{K}$, and use $\Delta_d = \{x \in [0,1]^d \colon \|x\|_1 = 1\}$ to denote the $(d-1)$-dimensional probability simplex. Let $D_\psi(x, y)$ denote the Bregman divergence between x and y induced by a differentiable convex function ψ, that is, $D_\psi(x, y) = \psi(x) - \psi(y) - \langle \nabla \psi(y), x - y \rangle$. We use $\|h\|_{x,f} = \sqrt{h^\top \nabla^2 f(x) h}$ and $\|h\|_{*,x,f} = \sqrt{h^\top (\nabla^2 f(x))^{-1} h}$ to denote the local norm and its dual norm of a vector h at a point x with respect to a convex function f, respectively. To simplify the notation, we use $f \lesssim g$ to denote $f = O(g)$. For a sequence $z = (z^{(1)}, \ldots, z^{(T)})$ and $q \in [1, \infty]$, we define the (squared) path-length values in terms of the ℓ_q-norm as $P_q^T(z) = \sum_{t=2}^T \|z^{(t)} - z^{(t-1)}\|_q^2$. Throughout this paper, we frequently use i as the player index, a as the action index, d or m as the dimension of a feasible set.

2.1. Online linear optimization and external/swap regret

Online linear optimization, a fundamental area of online learning, serves as a key framework for learning in games. In online linear optimization, a learner is given a convex set $\mathcal{K} \subseteq \mathbb{R}^d$ before the game starts. Then at each round $t = 1, \ldots, T$, the learner first selects a point $x^{(t)} \in \mathcal{K}$ based on past observations, and the environment then determines a loss vector $\ell^{(t)} \in \mathbb{R}^d$ without knowledge of $x^{(t)}$. The learner then suffers a loss of $\langle x^{(t)}, \ell^{(t)} \rangle \in \mathbb{R}$ and observes the loss vector $\ell^{(t)}$. The most common objective for the learner is to minimize the external regret Reg^T, which is defined as $\mathsf{Reg}^T = \max_{u \in \mathcal{K}} \mathsf{Reg}^T(u)$ for $\mathsf{Reg}^T(u) = \sum_{t=1}^T \langle x^{(t)} - u, \ell^{(t)} \rangle$. We will see below that minimizing this external regret relates to the problem of finding an approximate Nash equilibrium.

Another important objective of the learner is to minimize the swap regret. Here, we consider only the case where the feasible set is the probability simplex. Let $\mathcal{M}_d = \{M \in [0,1]^{d \times d} \colon M(k, \cdot) \in \Delta_d \text{ for } k \in [d]\}$ be the set of all $d \times d$ row stochastic matrices, which we also refer to as transition probability matrices. Then, the swap regret is defined as $\mathsf{SwapReg}^T = \max_{M \in \mathcal{M}_d} \mathsf{SwapReg}^T(M)$ for $\mathsf{SwapReg}^T(M) = \sum_{t=1}^T \langle x^{(t)}, \ell^{(t)} - M \ell^{(t)} \rangle$. We will see below how minimizing the swap regret is related to the problem of finding a correlated equilibrium. It should be noted that the swap regret is nonnegative, a property leveraged to obtain a desirable regret in multiplayer general-sum games.

2.2. Two-player zero-sum games

Here we introduce two-player zero-sum games with a payoff matrix $A \in [-1, 1]^{m_x \times m_y}$, where m_x and m_y are the number of actions of the x- and y-players, respectively. The procedure of this game is as follows: at each round $t = 1, \ldots, T$, the x-player chooses a *mixed strategy* (or strategy for short) $x^{(t)} \in \Delta_{m_x}$ and the y-player chooses $y^{(t)} \in \Delta_{m_y}$. Then the x-player observes a utility vector $g^{(t)} = A y^{(t)}$ and gains a reward of $\langle x^{(t)}, g^{(t)} \rangle$, and the y-player observes a loss vector $\ell^{(t)} =$

$A^\top x^{(t)}$ and incurs a loss of $\langle y^{(t)}, \ell^{(t)}\rangle$. The regret of the x-player, $\mathsf{Reg}_{x,g}^T$ is given by $\mathsf{Reg}_{x,g}^T = \max_{x^* \in \Delta_{m_x}} \mathsf{Reg}_{x,g}^T(x^*)$ for $\mathsf{Reg}_{x,g}^T(x^*) = \sum_{t=1}^T \langle x^* - x^{(t)}, Ay^{(t)}\rangle = \sum_{t=1}^T \langle x^* - x^{(t)}, g^{(t)}\rangle$, and the regret of the y-player is given by $\mathsf{Reg}_{y,\ell}^T = \max_{y^* \in \Delta_{m_y}} \mathsf{Reg}_{y,\ell}^T(y^*)$ for $\mathsf{Reg}_{y,\ell}^T(y^*) = \sum_{t=1}^T \langle y^{(t)} - y^*, A^\top x^{(t)}\rangle = \sum_{t=1}^T \langle y^{(t)} - y^*, \ell^{(t)}\rangle$. With this framework established, we now introduce the concept of a Nash equilibrium.

Definition 3 (Nash equilibrium) *In a two-player zero-sum game with a payoff matrix $A \in [-1,1]^{m_x \times m_y}$, a pair of probability distributions $\sigma = (x^*, y^*)$ over action sets $[m_x]$ and $[m_y]$ is an ε-*approximate Nash equilibrium *for $\varepsilon \geq 0$ if for any distributions $x \in \Delta_{m_x}$ and $y \in \Delta_{m_y}$, $x^\top Ay^* - \varepsilon \leq {x^*}^\top Ay^* \leq {x^*}^\top Ay + \varepsilon$. The distribution σ is a* Nash equilibrium *if σ is a 0-approximate Nash equilibrium.*

The following lemma is well known in the literature and provides a connection between no-external-regret learning dynamics and learning in two-player zero-sum games.

Theorem 4 (Freund and Schapire 1999) *In two-player zero-sum games, suppose that the regrets of the x- and y-players are $\mathsf{Reg}_{x,g}^T$ and $\mathsf{Reg}_{y,\ell}^T$, respectively. Then, the product distribution of the average play $(\frac{1}{T}\sum_{t=1}^T x^{(t)}, \frac{1}{T}\sum_{t=1}^T y^{(t)})$ is a $((\mathsf{Reg}_{x,g}^T + \mathsf{Reg}_{y,\ell}^T)/T)$-approximate Nash equilibrium.*

2.3. Multiplayer general-sum games

We next introduce multiplayer general-sum games. Let $n \geq 2$ be the number of players and $[n] = \{1, \ldots, n\}$ be the set of players. Each player $i \in [n]$ has an action set $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a utility function $u_i \colon \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1,1]$. The procedure of this game is as follows: at each round $t = 1, \ldots, T$, each player $i \in [n]$ chooses a mixed strategy $x_i^{(t)} \in \Delta_{m_i}$ and observes a *expected utility* $u_i^{(t)} \in [-1,1]^{m_i}$. Here, the a_i-th element of $u_i^{(t)}$ is defined as $u_i^{(t)}(a_i) = \mathbb{E}_{a_{-i} \sim x_{-i}^{(t)}}[u_i(a_i, a_{-i})]$, which is the expected reward when player i selects a_i and the other players choose their actions following a distribution $x_{-i}^{(t)} = (x_1^{(t)}, \ldots, x_{i-1}^{(t)}, x_{i+1}^{(t)}, \ldots, x_n^{(t)})$. The swap regret of each player $i \in [n]$ is given by $\mathsf{SwapReg}_{x_i,u_i}^T = \max_{M \in \mathcal{M}_{m_i}} \mathsf{SwapReg}_{x_i,u_i}^T(M)$ for $\mathsf{SwapReg}_{x_i,u_i}^T(M) = \sum_{t=1}^T \langle x_i^{(t)}, Mu_i^{(t)} - u_i^{(t)}\rangle$, where we recall that $\mathcal{M}_m = \{M \in [0,1]^{m \times m} \colon M(k, \cdot) \in \Delta_m \text{ for } k \in [m]\}$ is the set of all $m \times m$ row stochastic matrices and note that here we consider the utility vector $u_i^{(t)}$ instead of a loss vector. Note also that two-player zero-sum games are special cases of multiplayer general-sum games, and hereafter, corresponding notations may be used without clarification. With this framework established, we can now introduce the concept of a correlated equilibrium.

Definition 5 (Correlated equilibrium, Aumann 1974) *A probability distribution σ over action sets $\times_{i=1}^n \mathcal{A}_i$ is an ε-*approximate correlated equilibrium *for $\varepsilon \geq 0$ if for any player $i \in [n]$ and any (swap) function $\phi_i \colon \mathcal{A}_i \to \mathcal{A}_i$ that swap action a_i with $\phi_i(a_i)$, $\mathbb{E}_{a \sim \sigma}[u_i(a)] \geq \mathbb{E}_{a \sim \sigma}[u_i(\phi_i(a_i), a_{-i})] - \varepsilon$. The distribution σ is a* correlated equilibrium *if σ is a 0-approximate correlated equilibrium.*

The following lemma is folklore and provides a connection between no-swap-regret learning dynamics and learning in multiplayer general-sum games.

Theorem 6 (Foster and Vohra 1997) *In multiplayer general-sum games, suppose that the swap regret of each player i is $\mathsf{SwapReg}_{x_i,u_i}^T$ and let $\sigma^{(t)} = \otimes_{i \in [n]} x_i^{(t)} \in \Delta(\times_{i=1}^n \mathcal{A}_i)$ given by $\sigma^{(t)}(a_1, \ldots, a_n) = \prod_{i \in [n]} x_i^{(t)}(a_i)$ for each $a_i \in \mathcal{A}_i$ be the joint distribution at round t. Then, its time-averaged distribution $\sigma = \frac{1}{T}\sum_{t=1}^T \sigma^{(t)}$ is a $(\max_{i \in [n]} \mathsf{SwapReg}_{x_i,u_i}^T / T)$-approximate correlated equilibrium.*

2.4. Optimistic follow-the-regularized-leader

Here we present the framework of OFTRL for online linear optimization introduced in Section 2.1, which we employ in both two-player zero-sum games and multiplayer general-sum games. At each round $t \in [T]$, OFTRL selects a point $x^{(t)} \in \mathcal{K} \subseteq \mathbb{R}^d$ such that $x^{(t)} \in \arg\min_{x\in\mathcal{K}}\{\langle x, m^{(t)} + \sum_{s=1}^{t-1} \ell^{(s)}\rangle + \psi^{(t)}(x)\}$ for a convex regularizer $\psi^{(t)}$ over $\mathcal{K}$ and an optimistic prediction $m^{(t)} \in \mathbb{R}^d$ of the true loss vector $\ell^{(t)}$. The optimistic prediction $m^{(t)}$ serves as the predicted value of $\ell^{(t)}$ based on information available up to round $t-1$, and the closer this prediction is to the actual loss vector $\ell^{(t)}$, the smaller the regret. Intuitively, this property can be used to achieve fast convergence in learning in games because the loss vectors are determined by the strategies chosen by the other players. If all players follow a no-regret algorithm, their strategies can be predicted to some extent, enabling us to speed up convergence to an equilibrium. For this reason, $m^{(t)} = \ell^{(t-1)}$ is often used as the optimistic prediction. By employing OFTRL, we can obtain an RVU (Regret bounded by Variation in Utilities) bound, which was observed to be useful for learning in games by Syrgkanis et al. (2015). See Lemma 17 and Lemma 24 in Appendix B for the RVU bounds for OFTRL with the negative Shannon entropy regularizer and with the log-barrier regularizer, respectively.

3. Honest and Corrupted Regimes

This section introduces the honest and corrupted regimes, focusing on multiplayer general-sum games (with two-player zero-sum games being a special case, so all definitions apply). Let $\widehat{x}_i^{(t)} \in \Delta_{m_i}$ denote the strategy suggested by the prescribed algorithm of player i at round t, and let $x_i^{(t)} \in \Delta_{m_i}$ denote the strategy actually chosen by player i at round t. A game is said to be in an *honest regime* if all players fully adhere to the prescribed algorithm at every round; that is, $x_i^{(t)} = \widehat{x}_i^{(t)}$ for all $i \in [n]$ and $t \in [T]$.

Corrupted regime In practice, the assumption that all players fully adhere to the prescribed algorithm may not always hold; it is desirable for the individual regret to be bounded based on the degree of deviation from the prescribed algorithm. Furthermore, not only the strategies but also the observed utilities may be subject to corruption for various reasons. Motivated by the potential corruption in both strategies and utilities, we define the following corrupted regime:

Definition 7 (Corrupted regime with corruption level $\{(\widehat{C}_i, \widetilde{C}_i)\}_{i\in[n]}$) *A game is said to be in a* corrupted regime with corruption level $\{(\widehat{C}_i, \widetilde{C}_i)\}_{i\in[n]}$ *if the following two conditions hold:*

(i) the strategies $\{x_i^{(1)}, \dots, x_i^{(T)}\}$ chosen by each player $i \in [n]$ deviate from outputs of the prescribed algorithm $\{\widehat{x}_i^{(1)}, \dots, \widehat{x}_i^{(T)}\}$ by at most $\widehat{C}_i$, that is, $\sum_{t=1}^T \|x_i^{(t)} - \widehat{x}_i^{(t)}\|_1 \le \widehat{C}_i$ for all $i \in [n]$;

(ii) the utilities $\{\widetilde{u}_i^{(1)}, \dots, \widetilde{u}_i^{(T)}\}$ observed by each player $i \in [n]$ deviate from the expected utility $\{u_i^{(1)}, \dots, u_i^{(T)}\}$ by at most $\widetilde{C}_i$, that is, $\sum_{t=1}^T \|u_i^{(t)} - \widetilde{u}_i^{(t)}\|_\infty \le \widetilde{C}_i$ for all $i \in [n]$.

Define $C_i = 2\widehat{C}_i + 2\widetilde{C}_i$. Then the corrupted regime with $C_i = 0$ for all i corresponds to the honest regime, and the corrupted regime with arbitrary $\{C_j\}_{j\neq i}$ is the adversarial regime for player i. In the following sections, we analyze the external and swap regrets in the corrupted regime for two-player zero-sum and multiplayer general-sum games. The corrupted procedure is summarized as follows:

At each round $t = 1, \dots, T$:

1. A prescribed algorithm suggests a strategy $\widehat{x}_i^{(t)} \in \Delta_{m_i}$ for each player $i \in [n]$;

2. Each player $i \in [n]$ selects a strategy $x_i^{(t)} \leftarrow \widehat{x}_i^{(t)} + \widehat{c}_i^{(t)}$;
3. Each player i observes a corrupted utility vector $\widetilde{u}_i^{(t)} \leftarrow u_i^{(t)} + \widetilde{c}_i^{(t)}$;
4. Each player i gains a reward of $\langle x_i^{(t)}, u_i^{(t)} \rangle$ in Setting (I) and $\langle x_i^{(t)}, \widetilde{u}_i^{(t)} \rangle$ in Setting (II).

Here, $\widehat{c}_i^{(t)}$ and $\widetilde{c}_i^{(t)}$ are corruption vectors for strategies and utility, respectively, such that $\sum_{t=1}^T \|\widehat{c}_i^{(t)}\|_1 = \sum_{t=1}^T \|x_i^{(t)} - \widehat{x}_i^{(t)}\|_1 \leq \widehat{C}_i$, $\sum_{t=1}^T \|\widetilde{c}_i^{(t)}\|_\infty = \sum_{t=1}^T \|u_i^{(t)} - \widetilde{u}_i^{(t)}\|_\infty \leq \widetilde{C}_i$, and $C_i = 2\widehat{C}_i + 2\widetilde{C}_i$. Note that the corrupted feedback setting differs from the noisy feedback setting, where we observe feedback $u_i^{(t)} + \xi_i^{(t)}$ for a zero-mean noise $\xi_i^{(t)}$ (Cohen et al., 2017; Hsieh et al., 2022; Abe et al., 2023).

Remark 8 *As described above, there are two possible settings for the rewards obtained by each player: Setting (I), where the reward is given by $\langle x_i^{(t)}, u_i^{(t)} \rangle$, and Setting (II), where the reward is given by $\langle x_i^{(t)}, \widetilde{u}_i^{(t)} \rangle$. Setting (I) is natural in scenarios where communication channels or other external factors introduce noise before information from other players $j \neq i$ reaches player i. In contrast, Setting (II) is more appropriate when the true utility function u_i (or the payoff matrix A) is time-varying. As we will see below, the definition of regret should differ between these two settings.*

In the corrupted regime, the following four types of external regret can be defined: $\mathsf{Reg}_{x_i,u_i}^T(x^*) = \sum_{t=1}^T \langle x^* - x_i^{(t)}, u_i^{(t)} \rangle$, $\mathsf{Reg}_{\widehat{x}_i,u_i}^T(x^*) = \sum_{t=1}^T \langle x^* - \widehat{x}_i^{(t)}, u_i^{(t)} \rangle$, $\mathsf{Reg}_{x_i,\widetilde{u}_i}^T(x^*) = \sum_{t=1}^T \langle x^* - x_i^{(t)}, \widetilde{u}_i^{(t)} \rangle$, and $\mathsf{Reg}_{\widehat{x}_i,\widetilde{u}_i}^T(x^*) = \sum_{t=1}^T \langle x^* - \widehat{x}_i^{(t)}, \widetilde{u}_i^{(t)} \rangle$. As discussed in Remark 8, depending on the cause of the corruption in the utility vector $u_i^{(t)}$, it is natural to consider either $\mathsf{Reg}_{x_i,u_i}^T(x^*)$ or $\mathsf{Reg}_{x_i,\widetilde{u}_i}^T(x^*)$ as the evaluation metric for the player. Specifically, in Setting (I), it is natural to use $\mathsf{Reg}_{x_i,u_i}^T(x^*)$, while in Setting (II), it is natural to use $\mathsf{Reg}_{x_i,\widetilde{u}_i}^T(x^*)$. For these external regrets, the following inequalities hold:

Proposition 9 *For any $i \in [n]$ and $x^* \in \Delta_{m_i}$, $|\mathsf{Reg}_{x_i,u_i}^T(x^*) - \mathsf{Reg}_{\widehat{x}_i,u_i}^T(x^*)| \leq \widehat{C}_i$, $|\mathsf{Reg}_{x_i,\widetilde{u}_i}^T(x^*) - \mathsf{Reg}_{\widehat{x}_i,\widetilde{u}_i}^T(x^*)| \leq \widehat{C}_i$, $|\mathsf{Reg}_{x_i,u_i}^T(x^*) - \mathsf{Reg}_{x_i,\widetilde{u}_i}^T(x^*)| \leq 2\widetilde{C}_i$, and $|\mathsf{Reg}_{\widehat{x}_i,u_i}^T(x^*) - \mathsf{Reg}_{\widehat{x}_i,\widetilde{u}_i}^T(x^*)| \leq 2\widetilde{C}_i$.*

The proof can be found in Appendix C. Following the above four regrets, define $\mathsf{SwapReg}_{x_i,u_i}^T(M) = \sum_{t=1}^T \langle x_i^{(t)}, Mu_i^{(t)} - u_i^{(t)} \rangle$, $\mathsf{SwapReg}_{\widehat{x}_i,u_i}^T(M) = \sum_{t=1}^T \langle \widehat{x}_i^{(t)}, Mu_i^{(t)} - u_i^{(t)} \rangle$, $\mathsf{SwapReg}_{x_i,\widetilde{u}_i}^T(M) = \sum_{t=1}^T \langle x_i^{(t)}, M\widetilde{u}_i^{(t)} - \widetilde{u}_i^{(t)} \rangle$, and $\mathsf{SwapReg}_{\widehat{x}_i,\widetilde{u}_i}^T(M) = \sum_{t=1}^T \langle \widehat{x}_i^{(t)}, M\widetilde{u}_i^{(t)} - \widetilde{u}_i^{(t)} \rangle$. For these swap regrets, a similar proposition as Proposition 9 holds (see Appendix C for the statement and proof).

4. Corrupted Learning Dynamics in Two-Player Zero-Sum Games

This section investigates learning dynamics of two-player zero-sum games in the corrupted regime (see Appendix D.1 for the procedure). We will upper bound the external regrets of the x- and y-players.

4.1. Learning dynamics

We use OFTRL with the negative Shannon entropy regularizer as the prescribed algorithm. Let $\widetilde{g}^{(t)} = g^{(t)} + \widetilde{c}_x^{(t)}$ for $g^{(t)} = Ay^{(t)}$ and $\widetilde{\ell}^{(t)} = \ell^{(t)} + \widetilde{c}_y^{(t)}$ for $\ell^{(t)} = A^\top x^{(t)}$. Then, $\{\widehat{x}^{(t)}\}_{t=1}^T$ and

$\{\widehat{y}^{(t)}\}_{t=1}^T$ are given by

$$\widehat{x}^{(t)} = \mathop{\arg\max}_{x\in\Delta_{m_x}} \left\{ \left\langle x, \widetilde{g}^{(t-1)} + \sum_{s=1}^{t-1} \widetilde{g}^{(s)} \right\rangle + \frac{H(x)}{\eta_x^{(t)}} \right\}, \ \eta_x^{(t)} = \sqrt{\frac{\log_+(m_x)/2}{\log_+(m_x) + \sum_{s=1}^{t-1} \|\widetilde{g}^{(s)} - \widetilde{g}^{(s-1)}\|_\infty^2}},$$
$$\widehat{y}^{(t)} = \mathop{\arg\min}_{y\in\Delta_{m_y}} \left\{ \left\langle y, \widetilde{\ell}^{(t-1)} + \sum_{s=1}^{t-1} \widetilde{\ell}^{(s)} \right\rangle - \frac{H(y)}{\eta_y^{(t)}} \right\}, \ \eta_y^{(t)} = \sqrt{\frac{\log_+(m_y)/2}{\log_+(m_y) + \sum_{s=1}^{t-1} \|\widetilde{\ell}^{(s)} - \widetilde{\ell}^{(s-1)}\|_\infty^2}},$$

where $H(x) = \sum_k x(k) \log(1/x(k))$ is the Shannon entropy, $\log_+(z) = \max\{\log z, 4\}$, and we let $g^{(0)} = \ell^{(0)} = \mathbf{0}$. It is worth noting that the learning rates are adjusted to satisfy $\eta_x^{(t)} \leq 1/\sqrt{2}$ and $\eta_y^{(t)} \leq 1/\sqrt{2}$. OFTRL with the negative Shannon entropy regularizer corresponds to optimistic Hedge, and can be written in closed form as $\widehat{x}^{(t)}(i) \propto \exp\big(\eta_x^{(t)}\big(\widetilde{g}^{(t-1)}(i) + \sum_{s=1}^{t-1} \widetilde{g}^{(s)}(i)\big)\big)$ for $i \in [m_x]$ and $\widehat{y}^{(t)}(i) \propto \exp\big(-\eta_y^{(t)}\big(\widetilde{\ell}^{(t-1)}(i) + \sum_{s=1}^{t-1} \widetilde{\ell}^{(s)}(i)\big)\big)$ for $i \in [m_y]$.

4.2. External regret bounds and analysis

The learning dynamics determined by the above algorithm guarantee the following bounds.

Theorem 10 (External regret upper bounds) *Suppose that the x- and y-players use the algorithms in Section 4.1 to obtain strategies $\{\widehat{x}^{(t)}\}_{t=1}^T$ and $\{\widehat{y}^{(t)}\}_{t=1}^T$, respectively. Then, in the corrupted regime, it holds that*

$$\mathsf{Reg}_{x,g}^T \lesssim \min\left\{ \sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{\big(P_\infty^T(\widetilde{g}) + \log m_x\big) \log m_x} \right\} + C_x\,, \quad (1)$$
$$\mathsf{Reg}_{x,\widetilde{g}}^T \lesssim \min\left\{ \sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{\big(P_\infty^T(\widetilde{g}) + \log m_x\big) \log m_x} \right\} + \widehat{C}_x\,, \quad (2)$$

where $P_\infty^T(\widetilde{g}) = \sum_{t=2}^T \|\widetilde{g}^{(t)} - \widetilde{g}^{(t-1)}\|_\infty^2$ is the (squared) path-length values of $\widetilde{g}^{(t)}$ in terms of the ℓ_∞-norm. For the upper bounds on $\mathsf{Reg}_{y,\ell}^T$ and $\mathsf{Reg}_{y,\widetilde{\ell}}^T$, the upper bounds replacing x with y and $\widetilde{g}$ with $\widetilde{\ell}$ in (1) and (2) holds, respectively. (We include these bounds in Appendix D for completeness.)

A comparison with existing external regret upper bounds is provided in Table 1. The above bounds are the first individual external regret bounds for the corrupted regime, and they are adaptive to the corruption levels C_x and C_y. Note that by setting $C_x = C_y = 0$ in the above bounds, we obtain $\mathsf{Reg}_{x,g}^T \lesssim \sqrt{\log(m_x m_y) \log m_x}$ and $\mathsf{SwapReg}_{x,\widetilde{g}}^T \lesssim \sqrt{\log(m_x m_y) \log m_x}$ in the honest regime. It is worth noting that the regret guarantees incentivize players to follow the prescribed dynamics: any deviation by an opponent from the algorithm's output incurs only a square-root penalty, whereas a deviation by a player from the prescribed algorithm incurs a linear penalty. The bound $O(\sqrt{\log(m_x m_y) \log m_x})$ in the honest regime is slightly better than the classical bound of $O(\log(m_x m_y))$. The x-player (and similarly for the x-player) simultaneously achieves a path-length bound of $O(\sqrt{P_\infty^T(g) \log m_x})$ when $C_x = 0$ against any sequence of strategies by the y-player, which is $(\log(m_x T)/\sqrt{\log m_x})$-times better than the bound of Rakhlin and Sridharan (2013b) and also improves upon Syrgkanis et al. (2015) as our bound is adaptive to the path-length of the y-player's strategies. Finally, note that the upper bound on $\mathsf{Reg}_{x,g}^T$ depends linearly on the magnitude of $\widehat{C}_x$, whereas the upper bound on $\mathsf{Reg}_{x,\widetilde{g}}^T$ depends on $\widehat{C}_x$ via a square-root term.

It is worth noting that, in Theorem 10, we assume that both the x- and y-players follow the proposed algorithm described in Section 4.1. However, even if the opponent employs optimistic Hedge with a constant learning rate (Syrgkanis et al., 2015), a player following our algorithm can still achieve regret bounds similar to those in Theorem 10, in both the honest and corrupted regimes. This provides an incentive for players to use our algorithm. A similar observation also holds for the algorithm for multiplayer general-sum games in Section 5. A more detailed statement and proof can be found in Appendix D.5.

Note also that $C_y = \sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1$ in the regret upper bounds in (1) and (2) can be replaced with $C'_y = \sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1^2 \le 2C_y$, though we do not present this refinement in the main text for the sake of notational simplicity. This refinement can be easily verified by fixing the final inequality in (17) within the proof of Lemma 11. A similar refinement is also possible in the analysis of swap regret for multiplayer general-sum games in Section 5.

We provide the proof of Theorem 10 below. We begin by providing the following lemma, which follows from Lemma 17 and the definitions of corruption levels.

Lemma 11 *Suppose that the x- and y-players use the algorithms in Section 4.1 to obtain strategies $\{\widehat{x}^{(t)}\}_{t=1}^T$ and $\{\widehat{y}^{(t)}\}_{t=1}^T$, respectively. Then,*

$$\mathsf{Reg}^T_{\widehat{x},\widetilde{g}} \le 2\sqrt{2\log_+(m_x)\big(\log_+(m_x) + 8(\widehat{C}_y + \widetilde{C}_x) + 4P_1^T(\widehat{y})\big)} - \frac{1}{\sqrt{8}} P_1^T(\widehat{x}) + 2\,,$$
$$\mathsf{Reg}^T_{\widehat{y},\widetilde{\ell}} \le 2\sqrt{2\log_+(m_y)\big(\log_+(m_y) + 8(\widehat{C}_x + \widetilde{C}_y) + 4P_1^T(\widehat{x})\big)} - \frac{1}{\sqrt{8}} P_1^T(\widehat{y}) + 2\,,$$

where we recall that $P_1^T(z) = \sum_{t=1}^T \|z^{(t)} - z^{(t-1)}\|_1^2$ for $z = (z^{(1)}, \ldots, z^{(T)})$.

The proof of Lemma 11 can be found in Appendix D. Now we are ready to prove Theorem 10.

Proof [Proof sketch of Theorem 10] Here we provide the regret upper bounds on $\mathsf{Reg}^T_{x,g}$ and $\mathsf{Reg}^T_{y,\ell}$, and the rest of the proof can be found in Appendix D. From Lemma 11 and Proposition 9, we have

$$\mathsf{Reg}^T_{x,g} \le C_x + \sqrt{8\log_+(m_x)\big(1 + \log_+(m_x) + 8(\widehat{C}_y + \widetilde{C}_x) + 4P_1^T(\widehat{y})\big)} - \frac{1}{\sqrt{8}} P_1^T(\widehat{x}) + 2\,, \quad (3)$$
$$\mathsf{Reg}^T_{y,\ell} \le C_y + \sqrt{8\log_+(m_y)\big(1 + \log_+(m_y) + 8(\widehat{C}_x + \widetilde{C}_y) + 4P_1^T(\widehat{x})\big)} - \frac{1}{\sqrt{8}} P_1^T(\widehat{y}) + 2\,. \quad (4)$$

Summing up the above two bounds, we can upper bound the social regret $\mathsf{Reg}^T_{x,g} + \mathsf{Reg}^T_{y,\ell}$ by

$$\begin{aligned}
&\sqrt{8\log_+(m_x)\Big(1 + \log_+(m_x) + 8(\widehat{C}_y + \widetilde{C}_x) + 4P_1^T(\widehat{y})\Big)} - \frac{1}{\sqrt{8}} P_1^T(\widehat{x}) + C_x + 4 \\
&\quad + \sqrt{8\log_+(m_y)\Big(1 + \log_+(m_y) + 8(\widehat{C}_x + \widetilde{C}_y) + 4P_1^T(\widehat{x})\Big)} - \frac{1}{\sqrt{8}} P_1^T(\widehat{y}) + C_y \\
&= O\left(\sqrt{(\widehat{C}_y + \widetilde{C}_x)\log m_x} + \sqrt{(\widehat{C}_x + \widetilde{C}_y)\log m_y} + \log(m_x m_y) + C_x + C_y\right) - \frac{1}{4\sqrt{2}}\big(P_1^T(\widehat{x}) + P_1^T(\widehat{y})\big)\,,
\end{aligned}$$

where we used $b\sqrt{z} - az \le b^2/(4a)$ for $a > 0$, $b \ge 0$ and $z \ge 0$. Combining this with $\mathsf{Reg}^T_{x,g} + \mathsf{Reg}^T_{y,\ell} \ge 0$, which holds from the definition of the Nash equilibrium (see Lemma 26), we have

$$\begin{aligned}
P_1^T(\widehat{x}) + P_1^T(\widehat{y}) &\lesssim \sqrt{(\widehat{C}_y + \widetilde{C}_x)\log m_x} + \sqrt{(\widehat{C}_x + \widetilde{C}_y)\log m_y} + \log(m_x m_y) + C_x + C_y \\
&\lesssim \log(m_x m_y) + C_x + C_y\,,
\end{aligned} \quad (5)$$

Algorithm 1: No-swap-regret algorithm of player i in multiplayer general-sum games

for $t = 1, 2, \dots, T$ **do**
- For each expert $a \in \mathcal{A}_i$, compute $y_{i,a}^{(t)} \in \Delta_{m_i}$ by OFTRL in (6);
- Let $Q_i^{(t)} \in [0,1]^{m_i \times m_i}$ be a matrix whose a-th row is $y_{i,a}^{(t)}$, that is $Q_i^{(t)}(a, \cdot) = (y_{i,a}^{(t)})^\top$;
- Let $\widehat{x}_i^{(t)}$ be a stationary distribution of Markov chain induced by $Q_i^{(t)}$, $(Q_i^{(t)})^\top \widehat{x}_i^{(t)} = \widehat{x}_i^{(t)}$;
- Play $\widehat{x}_i^{(t)} = x_i^{(t)} + \widehat{c}_i^{(t)} \in \Delta_{m_i}$ for some corruption $\widehat{c}_i^{(t)}$;
- Observe corrupted utility $\widetilde{u}_i^{(t)} = u_i^{(t)} + \widetilde{c}_i^{(t)} \in [0,1]^{m_i}$ for $u_i^{(t)}(a_i) = \mathbb{E}_{a_{-i} \sim x_{-i}^{(t)}}[u_i(a_i, a_{-i})]$;
- For each $a \in \mathcal{A}_i$, let $\widetilde{u}_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a) \widetilde{u}_i^{(t)} \in [0,1]^{m_i}$;

where the last line follows from the AM–GM inequality and the definitions of C_x and C_y. Finally, plugging (5) in (3) and (4) gives the desired bounds on $\mathsf{Reg}_{x,g}^T$ and $\mathsf{Reg}_{y,\ell}^T$. The bound $\mathsf{Reg}_{x,g}^T \lesssim \sqrt{(P_\infty^T(\widetilde{g}) + \log m_x) \log m_x} + C_x$ follows from (16) in the proof of Lemma 11. ∎

5. Corrupted Learning Dynamics in Multiplayer General-Sum Games

This section presents learning dynamics for multiplayer general-sum games in the corrupted regime. Our algorithm for computing $\{\widehat{x}_i^{(t)}\}_{t=1}^T$ is a variant of the algorithm by Anagnostides et al. (2022b).

Reducing swap regret minimization to external regret minimization We begin by introducing a well-known result for swap regret minimization due to Blum and Mansour (2007), taking an algorithm for player i for example. They developed a method to reduce swap regret minimization to m_i instances of external regret minimization. Specifically, they consider the following procedure. First, for each $a \in \mathcal{A}_i$, let $u_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a) u_i^{(t)} \in [-\widehat{x}_i^{(t)}(a), \widehat{x}_i^{(t)}(a)]^{m_i}$ be the utility vector for the external regret minimizer $a \in \mathcal{A}_i$, which we call expert a, at round t. Then, let $y_{i,a}^{(t)} \in \Delta_{m_i}$ be the output of expert a. Then, the external regret for expert a is given by $\mathsf{Reg}_{i,a}^T = \max_{y \in \Delta_{m_i}} \sum_{t=1}^T \langle y - y_{i,a}^{(t)}, u_{i,a}^{(t)} \rangle$. Using $\{y_{i,a}^{(t)}\}_{a \in \mathcal{A}_i}$, we construct a transition probability matrix $Q_i^{(t)} \in [0,1]^{m_i \times m_i}$ whose a-th row is $y_{i,a}^{(t)}$ for each $a \in \mathcal{A}_i$, *i.e.*, $Q_i^{(t)}(a, \cdot) = (y_{i,a}^{(t)})^\top$. Let $\widehat{x}_i^{(t)}$ be a stationary distribution of the Markov chain associated with $Q_i^{(t)}$. That is $\widehat{x}_i^{(t)}(j) = \sum_{k=1}^{m_i} Q_i^{(t)}(k, j) \widehat{x}_i^{(t)}(k)$, which is equivalent to $(Q_i^{(t)})^\top \widehat{x}_i^{(t)} = \widehat{x}_i^{(t)}$, where $\widehat{x}_i^{(t)}$ is a column vector. Finally, we use this $\widehat{x}_i^{(t)}$ as a final suggested strategy for player i.[1]

Blum and Mansour (2007) showed that, with the above procedure, the swap regret is equal to the sum of the external regret of external regret minimizer $a \in \mathcal{A}_i$. In our corrupted scenario, in which there can be corruption in strategies and observed utilities, their result can be restated as follows:

Lemma 12 *Define* $\widetilde{u}_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a) \widetilde{u}_i^{(t)}$ *and* $\widetilde{\mathsf{Reg}}_{i,a}^T(y^*) = \sum_{t=1}^T \langle y^* - y_{i,a}^{(t)}, \widetilde{u}_{i,a}^{(t)} \rangle$. *Suppose that* $\widehat{x}_i^{(t)} = (Q_i^{(t)})^\top \widehat{x}_i^{(t)}$, *where* $Q_i^{(t)}(a, \cdot) = y_{i,a}^{(t)}$ *for each* $a \in \mathcal{A}_i$ *and recall that* $\mathsf{SwapReg}_{\widehat{x}_i, \widetilde{u}_i}^T(M) = \sum_{t=1}^T \langle \widehat{x}_i^{(t)}, M \widetilde{u}_i^{(t)} - \widetilde{u}_i^{(t)} \rangle$. *Then, it holds that* $\mathsf{SwapReg}_{\widehat{x}_i, \widetilde{u}_i}^T(M) = \sum_{a \in \mathcal{A}_i} \widetilde{\mathsf{Reg}}_{i,a}^T(M(a, \cdot))$.

We include the proof of this lemma for completeness in Appendix E.

1. Blum and Mansour (2007) does not consider the corrupted regime and thus $\widehat{x}_i^{(t)} = x_i^{(t)}$, $\widetilde{u}_i^{(t)} = u_i^{(t)}$, and $\widetilde{u}_{i,a}^{(t)} = u_{i,a}^{(t)}$.

No-external-regret algorithm From Lemma 12, it is sufficient to construct a no-external-regret algorithm that aims to minimize the external regret $\widetilde{\mathsf{Reg}}_{i,a}^T(M(a,\cdot))$. For this, we employ OFTRL with the log-barrier regularizer and an adaptive learning rate: we compute $y_{i,a}^{(t)} \in \Delta_{m_i}$ by

$$y_{i,a}^{(t)} = \operatorname*{arg\,max}_{y \in \Delta_{m_i}} \left\{ \left\langle y, \widetilde{u}_{i,a}^{(t-1)} + \sum_{s=1}^{t-1} \widetilde{u}_{i,a}^{(s)} \right\rangle - \frac{\phi(y)}{\eta_{i,a}^{(t)}} \right\}, \ \eta_{i,a}^{(t)} = \min \left\{ \sqrt{\frac{m_i \log T/8}{4 + \sum_{s=1}^{t-1} \|\widetilde{u}_{i,a}^{(s)} - \widetilde{u}_{i,a}^{(s-1)}\|_\infty^2}}, \eta_{i,\max} \right\} \tag{6}$$

for each $a \in \mathcal{A}_i$ and $\eta_{i,\max} = \frac{1}{256n\sqrt{m_i}}$, where $\eta_{i,a}^{(t)}$ is the learning rate for expert $a \in \mathcal{A}_i$ of player i at round t, $\phi(x) = -\sum_k \log(x(k))$ is the logarithmic barrier function, and recall $\widetilde{u}_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a)\widetilde{u}_i^{(t)}$. Here, we let $\widetilde{u}_{i,a}^{(0)} = \mathbf{0}$ for simplicity. We use the *expert-wise* adaptive learning rate $\eta_{i,a}^{(t)}$ for each player $i \in [n]$, while Anagnostides et al. (2022b) uses a constant common learning rate, $\eta_{i,a}^{(t)} = \eta_i$. The algorithm for each player $i \in [n]$ is summarized in Algorithm 1.

Let $m = \max_{i \in [n]} m_i$ denote the maximum number of actions. We define $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$, $\widetilde{S} = \sum_{i \in [n]} \widetilde{C}_i$, and $S = \sum_{i \in [n]} C_i$, where $\widehat{C}_i$, and $\widetilde{C}_i$ are defined in Definition 7, and $C_i = 2\widehat{C}_i + \widetilde{C}_i$. Then, Algorithm 1 achieves the following individual swap regret upper bounds.

Theorem 13 (Swap regret upper bounds) *In the corrupted regime, Algorithm 1 achieves*

$$\mathsf{SwapReg}_{x_i,u_i}^T \lesssim nm^{5/2} \log T + \min\left\{ m\sqrt{\left(\widehat{S}(n+\sqrt{m}) + \widetilde{S}\sqrt{m}\right)\log T}, m_i\sqrt{T \log T} \right\} + C_i \,.$$

$$\mathsf{SwapReg}_{x_i,\widetilde{u}_i}^T \lesssim nm^{5/2} \log T + \min\left\{ m\sqrt{\left(\widehat{S}(n+\sqrt{m}) + \widetilde{C}_i\right)\log T} + (\widetilde{S}nm^6 \log T)^{\frac{1}{4}}, m_i\sqrt{T \log T} \right\} + \widehat{C}_i \,.$$

The proof can be found in Appendix E, which provides slightly better upper bounds. A comparison against existing bounds can be found in Table 2. Note that setting $C_i = 0$ for all i in the above bounds yields $\mathsf{SwapReg}_{x_i,u_i}^T \lesssim nm^{5/2} \log T$ and $\mathsf{SwapReg}_{x_i,\widetilde{u}_i}^T \lesssim nm^{5/2} \log T$ in the honest regime. Compared to the best swap regret bounds by Anagnostides et al. (2022b), our algorithm achieves the same bound in the honest regime, a new adaptive bound in the corrupted regime in terms of $\widehat{S}$ and $\widetilde{S}$, and a worst-case bound that is $\sqrt{m}$-times worse than their bound of $O(nm^{5/2} \log T + \sqrt{Tm \log m})$. It is worth noting that the bound on $\mathsf{SwapReg}_{x_i,u_i}^T$ deteriorates linearly with respect to $\{\widetilde{C}_j\}_{j \neq i}$, whereas the bound on $\mathsf{SwapReg}_{x_i,\widetilde{u}_i}^T$ is affected by $\widetilde{C}_i$ only through a square root dependence and by $\{\widetilde{C}_j\}_{j \neq i}$ through a *fourth root* dependence. Interestingly, the fourth root dependence on $\widetilde{C}_j$ (for $j \neq i$) is better than the squared root on $\widetilde{C}_y$ in the upper bound of $\mathsf{Reg}_{x,\widetilde{g}}^T$. This difference arises because, whereas $\mathsf{Reg}_{x,\widetilde{g}}^T + \mathsf{Reg}_{y,\widetilde{\ell}}^T \not\geq 0$ (see Appendix D.4), we always have $\mathsf{SwapReg}_{x_i,\widetilde{u}_i}^T \geq 0$.

Remark 14 *Using expert-wise learning rates $\eta_{i,a}^{(t)}$ as in (6), is crucial for proving an upper bound of $\mathsf{SwapReg}_{x_i,\widetilde{u}_i}^T \lesssim \sqrt{\widetilde{C}_i} + (\sum_{j \neq i} \widetilde{C}_j)^{1/4}$, which has a sublinear dependence on $\{\widetilde{C}_i\}_{i \in [n]}$. If one modifies the algorithm of Anagnostides et al. (2022b) by using an adaptive learning rate, one might choose the same learning rate for all $a \in \mathcal{A}_i$ as $\eta_i^{(t)} \simeq \min\left\{ m_i\sqrt{\log T/(4+\sum_{s=1}^{t-1}\|u_i^{(s)} - u_i^{(s-1)}\|_\infty^2)}, 1/(n\sqrt{m_i}) \right\}$. However, despite our attempts, using this learning rate causes $\mathsf{SwapReg}_{\widehat{x}_i,\widetilde{u}_i}^T$ to depend linearly on $\widetilde{C}_i$, which fails to ensure the desired robustness against corruption in utilities.*

The key analysis to prove Theorem 13 lies in analyzing the stability of stationary distributions of the Markov chains determined by OFTRL with the adaptive learning rate in (6), which defines the output $y_{i,a}^{(t)}$ of each $a \in \mathcal{A}_i$. In contrast to the analysis by Anagnostides et al. (2022b), which uses a constant learning rate, we use the adaptive learning rate. By setting the learning rate sufficiently small and applying an analysis similar to that for two-player zero-sum games in Section 4, as well as the analysis in Wei and Luo (2018), we can show a swap regret bound of $O(nm^8 \log T)$ in the honest regime. However, the dependence on m is significantly worse than the $O(nm^{5/2} \log T)$ bound in Anagnostides et al. (2022b). To address this issue, we leverage the property of reducing swap regret minimization to external regret minimization (Lemma 12) in the stability analysis: exploiting this property ensures the stability of transition probability matrices without making the learning rate too small (Lemma 30), thereby allowing us to establish a swap regret bound of $O(nm^{5/2} \log T)$ in the honest regime while maintaining robustness in the corrupted regime.

6. Lower Bounds for Corrupted Games

This section provides several matching lower bounds of Theorems 10 and 13. All the lower bounds are constructed for two-player zero-sum games and proofs can be found in Appendix F.

Theorem 15 (Lower bounds in the corrupted regime) *For any learning dynamics,*

(i) there exists a game in the corrupted regime with $\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x$ *such that* $\mathsf{Reg}_{x,\widetilde{g}}^T = \mathsf{Reg}_{\widehat{x},\widetilde{g}}^T = \Omega\big(\sqrt{\widetilde{C}_x \log m_x}\big)$*, and there exists a game in the corrupted regime with* $\sum_{t=1}^T \|\ell^{(t)} - \widetilde{\ell}^{(t)}\|_\infty \le \widetilde{C}_y$ *such that* $\mathsf{Reg}_{y,\widetilde{\ell}}^T = \mathsf{Reg}_{\widehat{y},\widetilde{\ell}}^T = \Omega\big(\sqrt{\widetilde{C}_y \log m_y}\big)$*;*

(ii) there exists a game in the corrupted regime with $\sum_{t=1}^T \|x^{(t)} - \widehat{x}^{(t)}\|_1 \le \widehat{C}_x$ *such that* $\mathsf{Reg}_{x,g}^T = \mathsf{Reg}_{x,\widetilde{g}}^T = \Omega(\widehat{C}_x)$*, and there exists a game in the corrupted regime with* $\sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1 \le \widehat{C}_y$ *such that* $\mathsf{Reg}_{y,\ell}^T = \mathsf{Reg}_{y,\widetilde{\ell}}^T = \Omega(\widehat{C}_y)$*;*

(iii) there exists a game in the corrupted regime with $\sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1 \le \widehat{C}_y$ *such that* $\max\big\{\mathsf{Reg}_{\widehat{x},g}^T, \mathsf{Reg}_{\widehat{y},\ell}^T\big\} = \Omega\big(\sqrt{\widehat{C}_y}\big)$*, and there exists a game in the corrupted regime with* $\sum_{t=1}^T \|x^{(t)} - \widehat{x}^{(t)}\|_1 \le \widehat{C}_x$ *such that* $\max\big\{\mathsf{Reg}_{\widehat{x},g}^T, \mathsf{Reg}_{\widehat{y},\ell}^T\big\} = \Omega\big(\sqrt{\widehat{C}_x}\big)$.

Note that the lower bounds in (iii), unlike those in (ii), are for regrets $\mathsf{Reg}_{\widehat{x},g}^T$ and $\mathsf{Reg}_{\widehat{y},\ell}^T$, which are defined for the pre-corruption strategies $\{\widehat{x}^{(t)}\}_{t=1}^T$ and $\{\widehat{y}^{(t)}\}_{t=1}^T$. The lower bound in (iii) is stated in a form similar to the lower bounds of Syrgkanis et al. (2015, Theorem 24) and Chen and Peng (2020, Theorem 4.2). However, while their lower bounds focus only on a specific algorithm (the Hedge algorithm), our lower bounds hold for any learning dynamics. Moreover, their lower bounds do not address the corrupted regime. The detailed results of their lower bounds can be found in Appendix A.

The lower bound in (i) is proven by considering the payoff matrix $A = 0$ and reducing the problem to the finite-time lower bound for online linear optimization over the probability simplex (Lemma 35). Specifically, we consider a scenario in which corruption occurs only in the first $t = 1, \ldots, \widetilde{C}_x/2$ rounds, where the corrupted reward vectors $\widetilde{g}^{(t)}$ are designed to achieve the lower bound, and no corruption occurs in any other rounds.

The lower bound in (ii) (specifically, the first statement) can be proven by considering a payoff matrix with an action with a lower reward, and designing corrupted strategies that select the action. In particular, we consider the payoff matrix A whose first $m_x - 1$ rows consist entirely of 1s and whose last row consists entirely of 0s. Then, by considering corrupted strategies $x^{(t)} = e_{m_x}$ for rounds $t = 1, \ldots, \widehat{C}_x/2$ and $x^{(t)} = \widehat{x}^{(t)}$ for $t > \widehat{C}_x/2$, the desired bound can be proven.

For the lower bound in (iii), we consider $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$, in which the optimal action for y-player is action 3. We first observe that the first statement of (iii) is equivalent to the existence of a constant $\kappa > 0$ and a corrupted game satisfying $\sum_{t=1}^{T} \|y^{(t)} - \widehat{y}^{(t)}\|_1 \le \widehat{C}_y$ such that $\mathsf{Reg}_{\widehat{x},g}^{T} < \kappa\sqrt{\widehat{C}_y}$ implies $\mathsf{Reg}_{\widehat{y},\ell}^{T} \ge \kappa\sqrt{\widehat{C}_y}$. We consider a corruption to the strategies where, for the initial rounds $t = 1, \ldots, \widehat{C}_y/2$, the strategy $y^{(t)}$ is chosen so that $\check{g}^{(t)} = Ay^{(t)}$, where $\check{g}^{(t)}$ is the gradient vector achieving the lower bound for online linear optimization over the simplex. Then, the regret in the first $\widehat{C}_y/2$ rounds is lower bounded by $\frac{1}{2}\sqrt{\widehat{C}_y}$. For the remaining rounds, from the construction of the matrix A, in order for the y-player to satisfy $\mathsf{Reg}_{\widehat{y},\ell}^{T} < \kappa\sqrt{\widehat{C}_y}$, the y-player can select actions 1 and 2 at most $\kappa\sqrt{\widehat{C}_y}$ times. Using this, the regret after round $t = \widehat{C}_y/2 + 1$ can be lower bounded by $-\kappa\sqrt{\widehat{C}_y}$, and by choosing $\kappa = 1/4$, we obtain the desired lower bound.

7. Conclusion and Future Work

In this paper, we establish a new framework of the corrupted regime in learning in normal-form games, where players may deviate from the strategies given by a prescribed algorithm and the observed utilities are also subject to corruption. We provide an almost complete characterization of learning dynamics in the corrupted regime. Specifically, we design algorithms that achieve strong regret guarantees in the corrupted regime for both two-player zero-sum games and multiplayer general-sum games, while matching the best known regret bounds in the honest regime. Furthermore, we investigate regret lower bounds in the corrupted regime for two-player zero-sum games and show that our regret upper bounds are optimal for a wide range of corruption types. Our regret analysis reveals that the player's regret grows at the rate of the square root of the cumulative deviation of the opponent, whereas the player's own deviation contributes linearly.

There are many important directions for future work. A natural next step is to investigate whether our analysis can be extended beyond normal-form games to more general settings such as extensive-form games (Zinkevich et al., 2007) and Markov games (Bai and Jin, 2020). Another promising direction is to explore the partial information settings, such as semi-bandit feedback (Rakhlin and Sridharan, 2013b; Wei and Luo, 2018) and bandit feedback (O'Donoghue et al., 2021), and to study regret notions beyond external or swap regret, for example Φ-regret (Greenwald and Jafari, 2003; Stoltz and Lugosi, 2007; Bai et al., 2022). It would also be interesting to examine the relationship between the (adversarially) corrupted observed utilities studied in this paper and utilities with the zero-mean noise (Cohen et al., 2017).

From a more technical perspective, we highlight two directions for future work. The first direction is to investigate the optimality of corruption of opponents' observed utilities. In Section 6, we showed that our regret upper bounds are minimax optimal for a wide range of corruption types. However, in two-player zero-sum games, the regret upper bound for the x-player depends on the square root of $\widetilde{C}_y$, the cumulative corruptions of y-player's observed utilities, and the optimality of this dependence remains an open question. The second direction is to close the $\sqrt{m}$ gap between the swap-regret upper bounds for multiplayer general-sum games in the adversarial setting, specifically between the bound established by Anagnostides et al. (2022b) and our own bound, as discussed shortly after Theorem 13.

Acknowledgments

TT was supported by JSPS KAKENHI Grant Number JP24K23852. SI was supported by JSPS KAKENHI Grant Number JP25K03184. HL was supported by NSF award IIS-1943607.

References

Kenshi Abe, Kaito Ariu, Mitsuki Sakamoto, Kentaro Toyoshima, and Atsushi Iwasaki. Last-iterate convergence with full and noisy feedback in two-player zero-sum games. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pages 7999–8028. PMLR, 2023. 8

Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, page 736–749. Association for Computing Machinery, 2022a. 3, 19

Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $O(\log T)$ swap regret in multiplayer games. In *Advances in Neural Information Processing Systems*, volume 35, pages 3292–3304. Curran Associates, Inc., 2022b. 3, 4, 11, 12, 13, 14, 19, 26, 30, 37

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On the convergence of no-regret learning dynamics in time-varying games. In *Advances in Neural Information Processing Systems*, volume 36, pages 16367–16405. Curran Associates, Inc., 2023. 20

Kimon Antonakopoulos, Veronica Belmega, and Panayotis Mertikopoulos. An adaptive mirror-prox method for variational inequalities with singular operators. In *Advances in Neural Information Processing Systems*, volume 32, pages 8455–8465. Curran Associates, Inc., 2019. 20

Kimon Antonakopoulos, Veronica Belmega, and Panayotis Mertikopoulos. Adaptive extra-gradient methods for min-max optimization and games. In *International Conference on Learning Representations*, 2021. 20

Robert J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, 1974. 3, 6

Yu Bai and Chi Jin. Provable self-play algorithms for competitive reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 551–560. PMLR, 2020. 14

Yu Bai, Chi Jin, Song Mei, Ziang Song, and Tiancheng Yu. Efficient phi-regret minimization in extensive-form games via online mirror descent. In *Advances in Neural Information Processing Systems*, volume 35, pages 22313–22325. Curran Associates, Inc., 2022. 14

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(47):1307–1324, 2007. 4, 11, 19

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006. 1, 2

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Advances in Neural Information Processing Systems*, volume 33, pages 18990–18999. Curran Associates, Inc., 2020. 3, 13, 19, 20

Johanne Cohen, Amélie Heliou, and Panayotis Mertikopoulos. Learning with bandit feedback in potential games. In *Advances in Neural Information Processing Systems*, volume 30, pages 6369–6378. Curran Associates, Inc., 2017. 8, 14

Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. From external to swap regret 2.0: An efficient reduction for large action spaces. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, page 1216–1222. Association for Computing Machinery, 2024. 20

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, page 235–254. Society for Industrial and Applied Mathematics, 2011. 2, 5

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. In *Advances in Neural Information Processing Systems*, volume 34, pages 27604–27616. Curran Associates, Inc., 2021. 19

Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning dynamics for general convex games. In *Advances in Neural Information Processing Systems*, volume 35, pages 39076–39089. Curran Associates, Inc., 2022. 19

Yi Feng, Hu Fu, Qun Hu, Ping Li, Ioannis Panageas, bo peng, and Xiao Wang. On the last-iterate convergence in time-varying zero-sum games: Extra gradient succeeds where optimism fails. In *Advances in Neural Information Processing Systems*, volume 36, pages 21933–21944. Curran Associates, Inc., 2023. 20

Dean P. Foster and Rakesh V. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1):40–55, 1997. 3, 6

Dylan J Foster, Zhiyuan Li, Thodoris Lykouris, Karthik Sridharan, and Eva Tardos. Learning in games: Robustness of fast convergence. In *Advances in Neural Information Processing Systems*, volume 29, pages 4734–4742. Curran Associates, Inc., 2016. 19

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997. 19

Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1):79–103, 1999. 1, 6

Kaito Fujii. Bayes correlated equilibria and no-regret dynamics. *arXiv preprint arXiv:2304.05005*, 2023. 20

Amy Greenwald and Amir Jafari. A general class of no-regret learning algorithms and game-theoretic equilibria. In *Learning Theory and Kernel Machines*, pages 2–12. Springer Berlin Heidelberg, 2003. 14

Keegan Harris, Ioannis Anagnostides, Gabriele Farina, Mikhail Khodak, Steven Wu, and Tuomas Sandholm. Meta-learning in games. In *The Eleventh International Conference on Learning Representations*, 2023. 20

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000. 1, 3, 5

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016. 44

Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to Nash equilibrium. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134, pages 2388–2422. PMLR, 2021. 20

Yu-Guan Hsieh, Kimon Antonakopoulos, Volkan Cevher, and Panayotis Mertikopoulos. No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation. In *Advances in Neural Information Processing Systems*, volume 35, pages 6544–6556. Curran Associates, Inc., 2022. 8

Ehsan Asadi Kangarshahi, Ya-Ping Hsieh, Mehmet Fatih Sahin, and Volkan Cevher. Let's be honest: An optimal no-regret framework for zero-sum games. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 2488–2496. PMLR, 2018. 2

Tianyi Lin, Zhengyuan Zhou, Panayotis Mertikopoulos, and Michael Jordan. Finite-time last-iterate convergence for multi-agent learning in games. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 6161–6171. PMLR, 2020. 20

Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994. 19

Arkadi S Nemirovski and Michael J Todd. Interior-point methods for optimization. *Acta Numerica*, 17:191–234, 2008. 23

Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994. 23, 24

Brendan O'Donoghue, Tor Lattimore, and Ian Osband. Matrix games with bandit feedback. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161, pages 279–289. PMLR, 2021. 14

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019. 21, 44

Francesco Orabona and Dávid Pál. Optimal non-asymptotic lower bound on the minimax regret of learning with expert advice. *arXiv preprint arXiv:1511.02176*, 2015. 43

Binghui Peng and Aviad Rubinstein. Fast swap regret minimization and applications to approximate correlated equilibria. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, page 1223–1234. Association for Computing Machinery, 2024. 20

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30, pages 993–1019, 2013a. 2, 19

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, volume 26, pages 3066–3074. Curran Associates, Inc., 2013b. 2, 9, 14, 19, 20

Gilles Stoltz and Gábor Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, 59(1):187–208, 2007. 14

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, volume 28, pages 2989–2997. Curran Associates, Inc., 2015. 2, 4, 7, 9, 10, 13, 19, 20, 30

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1263–1291. PMLR, 2018. 4, 13, 14, 19

Yu-Hu Yan, Peng Zhao, and Zhi-Hua Zhou. Fast rates in time-varying strongly monotone games. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 39138–39164. PMLR, 2023. 20

Mengxiao Zhang, Peng Zhao, Haipeng Luo, and Zhi-Hua Zhou. No-regret learning in time-varying zero-sum games. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pages 26772–26808. PMLR, 2022. 20

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems*, volume 20, pages 1729–1736. Curran Associates, Inc., 2007. 14

Appendix A. Additional Related Work

This section discusses additional related work that could not be included above.

External regret minimization for computing Nash equilibrium In two-player zero-sum games, it is known that when each player uses an (external) regret minimization algorithm, the average iterate of the chosen strategies converges to a Nash equilibrium (see Theorem 4). For example, when each player employs FTRL with the negative Shannon entropy corresponding to Hedge (Littlestone and Warmuth, 1994; Freund and Schapire, 1997), which has an upper regret bound of $\widetilde{O}(\sqrt{T})$, an approximate Nash equilibrium is obtained at a rate of $\widetilde{O}(1/\sqrt{T})$. A significant improvement in this convergence rate can be achieved by employing the optimistic frameworks such as OFTRL and the optimistic online mirror descent (Rakhlin and Sridharan, 2013a,b; Syrgkanis et al., 2015). The best-known upper bound for the individual regret in the honest regime is $O(\log(m_x m_y))$, achieved with OFTRL with the negative Shannon entropy regularizer and a constant learning rate (Syrgkanis et al., 2015). Importantly, the authors ensure robustness against adversarial opponents by monitoring the cumulative variation of the actions chosen by the player and switching to an algorithm with a learning rate of $\Theta(1/\sqrt{T})$ once it exceeds a certain threshold. Since the work of Syrgkanis et al. (2015), the optimistic prediction has been widely used as a tool for achieving an $o(\sqrt{T})$ regret in learning in games (Foster et al., 2016; Wei and Luo, 2018; Chen and Peng, 2020; Anagnostides et al., 2022b).

Although not the focus of this paper, it is known that in multiplayer general-sum games, if each player uses a no-external-regret algorithm, their time-averaged distribution of joint play converges to a coarse correlated equilibrium (CCE), which is a less favorable notion of equilibrium compared to the correlated equilibrium. Syrgkanis et al. (2015) showed that external regret minimization with OFTRL using the negative Shannon entropy achieves an individual regret of $\widetilde{O}(T^{1/4})$ in the honest regime. This result was later improved to $\widetilde{O}(T^{1/6})$ for two-player games by Chen and Peng (2020). Subsequently, significant advances in algorithms and analytical techniques have shown that (poly)logarithmic external regret can be achieved (Daskalakis et al., 2021; Farina et al., 2022). Although this paper does not address it, we believe that a similar extension of the work by Farina et al. (2022) to the corrupted regime is possible.

Swap regret minimization for computing correlated equilibrium In multiplayer general-sum games, the correlated equilibrium, which is considered a more desirable solution concept compared to the coarse correlated equilibrium, can be obtained when each player employs a no-swap-regret algorithm (see Theorem 6). As discussed in Section 5, to minimize the swap regret, we prepare a set of base experts for each action and determine the player's strategies using a Markov chain defined by these outputs (Blum and Mansour, 2007). The challenge in analyzing swap regret minimization lies in the need for precise analysis of the stability of the stationary distribution of the Markov chain. Notably, Chen and Peng (2020) analyzed this stability using the Markov chain tree theorem and demonstrated that individual swap regret can be bounded by $O(T^{1/4})$ by employing OFTRL with the negative Shannon entropy. Later, this swap regret bound was significantly improved by Anagnostides et al. (2022a,b). In particular, Anagnostides et al. (2022b) demonstrated that using OFTRL with the log-barrier regularizer and a constant learning rate makes the stability of the Markov chain stronger, leading to the first $O(\log T)$ individual swap upper bound. To ensure the adversarial robustness of the algorithm, they also considered a procedure that switches to an algorithm with a learning rate of $\Theta(1/\sqrt{T})$ whenever the cumulative variation of the observed utilities exceeds $\log T$, which is similar to Syrgkanis et al. (2015). Although not addressed in this study, there are lines of research that consider variations of the swap regret in Bayesian games (Fujii, 2023) and that focus on dependencies on the number of actions m instead of the number of rounds T when analyzing the swap regret (Dagan et al., 2024; Peng and Rubinstein, 2024).

Adaptive learning rate The adaptive learning rate is a critical technical element in our corrupted learning dynamics. The use of adaptive learning rate in learning in games is not new. Adaptive learning rate has been used in two-player zero-sum matrix games (Rakhlin and Sridharan, 2013b), cocoercive games (Lin et al., 2020), variational inequalities (Antonakopoulos et al., 2019, 2021), and variationally stable games (Hsieh et al., 2021). Of these, Rakhlin and Sridharan (2013b); Hsieh et al. (2021) are the most relevant to our study. Some of our results partially subsume theirs as a special case. However, unlike their work, the use of adaptive learning rate for the corrupted regime has not been studied before, especially in terms of the swap regret.

Time-varying games Zhang et al. (2022) introduced the concept of time-varying games in two-player zero-sum settings, demonstrating that regret upper bounds can be achieved depending on the variation of the payoff matrix. Following their framework, time-varying games have gained increasing interest (Harris et al., 2023; Yan et al., 2023; Feng et al., 2023; Anagnostides et al., 2023). Our setting, where corrupted expected utilities are observed, can be seen as a generalization of the time-varying setting, where the underlying payoff matrix A or utility function u_i are time-varying. (When considering such variations, recall from Remark 8 that defining regret in terms of $\mathsf{Reg}^T_{x_i,\widetilde{u}_i}$ or $\mathsf{SwapReg}^T_{x_i,\widetilde{u}_i}$ is the natural choice.) However, a key distinction is that our primary motivation lies in corruption in strategies. Moreover, we consider a more general corruption model for utility than Zhang et al. (2022) and its subsequent works, making a direct comparison with their regret upper bounds impossible.

Lower bounds In learning in games, while there has been extensive research on regret upper bounds, there has been relatively little work on regret lower bounds. Initially, Syrgkanis et al. (2015, Theorem 24) considered a setting where the x-player uses (vanilla) Hedge (recall that this corresponds to FTRL with the negative Shannon entropy) with any learning rate, while the y-player follows a (pure) best response (*i.e.,* minimizing the expected utility of the current round based on the x-player's choice). They showed that for Hedge with any learning rate, there exists a two-player zero-sum game where the x-player must suffer a $\sqrt{T}$ regret. Later, Chen and Peng (2020, Theorem 4.2) extended this result to a setting where both the x- and y-players use Hedge. They demonstrated that for Hedge with any learning rate, there exists a two-player general-sum game where at least one player suffers a $\sqrt{T}$ regret. Note that these regret lower bounds do not apply to the corrupted regime, and these results rely on regret upper bounds that depend on the update rule of Hedge. In contrast, our lower bounds for corrupted games in Theorem 15 hold for any learning dynamics.

Appendix B. Regret Analysis of Optimistic Follow-the-Regularized-Leader

In this section, we analyze the regret of the Optimistic Follow-the-Regularized-Leader (OFTRL) for online linear optimization. First, we provide a general analysis of OFTRL. Then, we present RVU bounds for OFTRL with the negative Shannon entropy and the log-barrier regularizer, which are used respectively in two-player zero-sum games in Section 4 and multi-player general-sum games in Section 5. In the context of learning in games, it is common to consider a *utility* vector $u^{(t)}$ instead of a loss vector $\ell^{(t)}$. Even in such cases, our statements for online linear optimization provided in this paper can be applied by letting $\ell^{(t)} = -u^{(t)}$.

B.1. Common analysis

The following lemma provides a regret bound for OFTRL that holds for a general regularizer $\psi^{(t)}$.

Lemma 16 (Regret bound for OFTRL) *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a nonempty closed convex set. Let $x^{(t)} \in \arg\min_{x\in\mathcal{K}}\{\langle x, m^{(t)} + \sum_{s=1}^{t-1} \ell^{(s)}\rangle + \psi^{(t)}(x)\}$ be the output of OFTRL (with linearized losses) at round t. Then, for any $x^* \in \mathcal{K}$,*

$$\sum_{t=1}^{T}\langle x^{(t)} - x^*, \ell^{(t)}\rangle \leq \psi^{(T+1)}(x^*) - \psi^{(1)}(x^{(1)}) + \sum_{t=1}^{T}\Big(\psi^{(t)}(x^{(t+1)}) - \psi^{(t+1)}(x^{(t+1)})\Big)$$
$$+ \sum_{t=1}^{T}\Big(\langle x^{(t)} - x^{(t+1)}, \ell^{(t)} - m^{(t)}\rangle - D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)})\Big) + \langle x^* - x^{(T+1)}, m^{(T+1)}\rangle\,. \quad (7)$$

This lemma follows from a standard analysis of OFTRL and can be seen as a variant of Orabona (2019, Theorem 7.36). We include the proof for completeness.

Proof Define $\widetilde{\psi}^{(t)}(x) = \psi^{(t)}(x) + \langle x, m^{(t)}\rangle$, and let

$$F^{(t)}(x) = \left\langle x, m^{(t)} + \sum_{s=1}^{t-1}\ell^{(s)}\right\rangle + \psi^{(t)}(x) = \left\langle x, \sum_{s=1}^{t-1}\ell^{(s)}\right\rangle + \widetilde{\psi}^{(t)}(x)$$

be the objective function of OFTRL at round t. From $-\sum_{t=1}^{T}\langle x^*, \ell^{(t)}\rangle = \widetilde{\psi}^{(T+1)}(x^*) - F^{(T+1)}(x^*)$, we have

$$\begin{aligned}
&\sum_{t=1}^{T}\langle x^{(t)} - x^*, \ell^{(t)}\rangle \\
&= \widetilde{\psi}^{(T+1)}(x^*) - F^{(T+1)}(x^*) + \big(-F^{(1)}(x^{(1)}) + F^{(1)}(x^{(1)})\big) + \big(-F^{(T+1)}(x^{(T+1)}) + F^{(T+1)}(x^{(T+1)})\big) \\
&\quad + \sum_{t=1}^{T}\langle x^{(t)}, \ell^{(t)}\rangle \\
&= \widetilde{\psi}^{(T+1)}(x^*) - F^{(T+1)}(x^*) - F^{(1)}(x^{(1)}) + \sum_{t=1}^{T}\big(F^{(t)}(x^{(t)}) - F^{(t+1)}(x^{(t+1)})\big) + F^{(T+1)}(x^{(T+1)}) \\
&\quad + \sum_{t=1}^{T}\langle x^{(t)}, \ell^{(t)}\rangle \\
&\leq \widetilde{\psi}^{(T+1)}(x^*) - \widetilde{\psi}^{(1)}(x^{(1)}) + \sum_{t=1}^{T}\big(F^{(t)}(x^{(t)}) - F^{(t+1)}(x^{(t+1)}) + \langle x^{(t)}, \ell^{(t)}\rangle\big)\,,
\end{aligned} \quad (8)$$

where in the second inequality we considered the telescoping sum and in the last line we used the fact that $x^{(T+1)}$ is the minimizer of $F^{(T+1)}$. The last term in the last inequality is bounded as

$$\begin{aligned}
&F^{(t)}(x^{(t)}) - F^{(t+1)}(x^{(t+1)}) + \langle x^{(t)}, \ell^{(t)}\rangle \\
&= F^{(t)}(x^{(t)}) - F^{(t)}(x^{(t+1)}) + \langle x^{(t+1)}, m^{(t)} - m^{(t+1)}\rangle \\
&\quad + \psi^{(t)}(x^{(t+1)}) - \psi^{(t+1)}(x^{(t+1)}) + \langle x^{(t)} - x^{(t+1)}, \ell^{(t)}\rangle \\
&\leq -D_{F^{(t)}}(x^{(t+1)}, x^{(t)}) + \psi^{(t)}(x^{(t+1)}) - \psi^{(t+1)}(x^{(t+1)}) \\
&\quad + \langle x^{(t+1)}, m^{(t)} - m^{(t+1)}\rangle + \langle x^{(t)} - x^{(t+1)}, \ell^{(t)}\rangle\,.
\end{aligned} \quad (9)$$

Here in the last inequality we used

$$D_{F^{(t)}}(x^{(t+1)}, x^{(t)}) = F^{(t)}(x^{(t+1)}) - F^{(t)}(x^{(t)}) - \langle x^{(t+1)} - x^{(t)}, \nabla F^{(t)}(x^{(t)}) \rangle \leq F^{(t)}(x^{(t+1)}) - F^{(t)}(x^{(t)}) ,$$

where the inequality follows from the first-order optimality condition at $x^{(t)}$, that is, $\langle x - x^{(t)}, \nabla F^{(t)}(x^{(t)}) \rangle \geq 0$ for any $x \in \mathcal{K}$. Combining (8), (9), and the fact that $D_{F^{(t)}}(x^{(t+1)}, x^{(t)}) = D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)})$, we obtain

$$\begin{aligned}
&\sum_{t=1}^{T} \langle x^{(t)} - x^*, \ell^{(t)} \rangle \\
&\leq \widetilde{\psi}^{(T+1)}(x^*) - \widetilde{\psi}^{(1)}(x^{(1)}) + \sum_{t=1}^{T} \Big(\psi^{(t)}(x^{(t+1)}) - \psi^{(t+1)}(x^{(t+1)}) - D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)}) \Big) \\
&\qquad + \sum_{t=1}^{T} \langle x^{(t+1)}, m^{(t)} - m^{(t+1)} \rangle + \sum_{t=1}^{T} \langle x^{(t)} - x^{(t+1)}, \ell^{(t)} \rangle \\
&= \psi^{(T+1)}(x^*) - \psi^{(1)}(x^{(1)}) + \sum_{t=1}^{T} \Big(\psi^{(t)}(x^{(t+1)}) - \psi^{(t+1)}(x^{(t+1)}) - D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)}) \Big) \\
&\qquad + \sum_{t=1}^{T} \langle x^{(t+1)} - x^{(t)}, m^{(t)} \rangle + \langle x^* - x^{(T+1)}, m^{(T+1)} \rangle + \sum_{t=1}^{T} \langle x^{(t)} - x^{(t+1)}, \ell^{(t)} \rangle ,
\end{aligned}$$

where the last line follows from the telescoping $\sum_{t=1}^{T} \langle x^{(t+1)}, m^{(t)} - m^{(t+1)} \rangle = \sum_{t=1}^{T} \langle x^{(t+1)} - x^{(t)}, m^{(t)} \rangle + \langle x^{(1)}, m^{(1)} \rangle - \langle x^{(T+1)}, m^{(T+1)} \rangle$. This completes the proof. ■

B.2. OFTRL with negative Shannon entropy regularizer

Lemma 16 immediately yields the well-known RVU bound for OFTRL with the negative Shannon entropy, which we use in our analysis for two-player zero-sum games in Section 4.

Lemma 17 (RVU bound for OFTRL with negative Shannon entropy regularizer) *Let $\psi^{(t)}(x) = -\frac{1}{\eta^{(t)}} H(x)$ for $H(x) = \sum_{k=1}^{d} x(k) \log(1/x(k))$ be the negative Shannon entropy regularizer with nonincreasing learning rate $\eta^{(t)}$ and $x^{(t)} \in \arg\min_{x \in \Delta_d} \{ \langle x, m^{(t)} + \sum_{s=1}^{t-1} \ell^{(s)} \rangle + \psi^{(t)}(x) \}$ be the output of OFTRL at round t. Then, for any $x^* \in \Delta_d$,*

$$\sum_{t=1}^{T} \langle x^{(t)} - x^*, \ell^{(t)} \rangle \leq \frac{\log d}{\eta^{(T+1)}} + \sum_{t=1}^{T} \eta^{(t)} \| \ell^{(t)} - m^{(t)} \|_\infty^2 - \sum_{t=1}^{T} \frac{1}{4\eta^{(t)}} \| x^{(t)} - x^{(t+1)} \|_1^2 + 2 \| m^{(T+1)} \|_\infty .$$

Proof We will upper bound the RHS of (7) in Lemma 16. Since $H(x) \leq \log d$ for all $x \in \Delta_d$, we have

$$\begin{aligned}
&\psi^{(T+1)}(x^*) - \psi^{(1)}(x^{(1)}) + \sum_{t=1}^{T} \Big(\psi^{(t)}(x^{(t+1)}) - \psi^{(t+1)}(x^{(t+1)}) \Big) \\
&\leq \frac{\log d}{\eta^{(1)}} + \sum_{t=1}^{T} \left(\frac{1}{\eta^{(t+1)}} - \frac{1}{\eta^{(t)}} \right) \log d = \frac{\log d}{\eta^{(T+1)}} .
\end{aligned}$$

Since $\psi^{(t)}$ is $(1/\eta^{(t)})$-strongly convex w.r.t. $\|\cdot\|_1$, we also have

$$
\begin{aligned}
&\langle x^{(t)} - x^{(t+1)}, \ell^{(t)} - m^{(t)} \rangle - D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)}) \\
&\le \|x^{(t)} - x^{(t+1)}\|_1 \|\ell^{(t)} - m^{(t)}\|_\infty - \frac{1}{2\eta^{(t)}} \|x^{(t)} - x^{(t+1)}\|_1^2 \\
&= \|x^{(t)} - x^{(t+1)}\|_1 \|\ell^{(t)} - m^{(t)}\|_\infty - \frac{1}{4\eta^{(t)}} \|x^{(t)} - x^{(t+1)}\|_1^2 - \frac{1}{4\eta^{(t)}} \|x^{(t)} - x^{(t+1)}\|_1^2 \\
&\le \eta^{(t)} \|\ell^{(t)} - m^{(t)}\|_\infty^2 - \frac{1}{4\eta^{(t)}} \|x^{(t)} - x^{(t+1)}\|_1^2 ,
\end{aligned}
$$

where the first inequality follows from Hölder's inequality and the $(1/\eta^{(t)})$-strong convexity of $\psi^{(t)}$ with respect to $\|\cdot\|_1$, and the last inequality follows from $b\sqrt{z} - az \le b^2/(4a)$ for $a > 0, b \ge 0$, and $z \ge 0$. Combining Lemma 16 with the last two inequalities gives the desired bound. ■

B.3. OFTRL with self-concordant barrier and adaptive learning rate

Here, we present the RVU bound for OFTRL with a self-concordant barrier, a class of functions that includes the log-barrier regularizer. We first introduce the concepts and key properties of self-concordant functions and self-concordant barriers.

Self-concordant function and self-concordant barrier We first define the self-concordant function and self-concordant barrier (see *e.g.,* Nesterov and Nemirovskii 1994; Nemirovski and Todd 2008 for detailed background). Recall that $\|h\|_{x,f} = \sqrt{h^\top \nabla^2 f(x) h}$ and $\|h\|_{*,x,f} = \sqrt{h^\top (\nabla^2 f(x))^{-1} h}$.

Definition 18 *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a closed convex set. Then a function $f \colon \mathrm{int}(\mathcal{K}) \to \mathbb{R}$ is* self-concordant *if f is three times continuously differentiable convex function with $f(x_k) \to \infty$ if $x_k \to x_\infty \in \partial\mathcal{K}$, and*

$$|D^3 f(x)[h,h,h]| \le 2\big(D^2 f(x)[h,h]\big)^{3/2}$$

for all $x \in \mathrm{int}(\mathcal{K})$ and $h \in \mathbb{R}^d$.

Definition 19 *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a closed convex set and $\vartheta \ge 1$. Then a function g is* ϑ-self-concordant barrier *for $\mathcal{K}$ if g is self-concordant and*

$$|Df(x)[h]| \le (\vartheta D^2 f(x)[h,h])^{1/2}$$

for all $x \in \mathrm{int}(\mathcal{K})$ and $h \in \mathbb{R}^d$.

We use the fact that the log-barrier function $\phi(x) = -\sum_{k=1}^d \log(x(k))$ is a d-self-concordant barrier over the positive orthant. For self-concordant functions, we use the following lemma.

Lemma 20 (Nesterov and Nemirovskii 1994, Theorem 2.1.1) *Let $\mathcal{S}$ be an open convex subset of a finite-dimensional real vector space. Let f be a self-concordant function on $\mathcal{S}$. Then, for any $y \in \mathcal{S}$ such that $\|x - y\|_{y,f} < 1$,*

$$(1 - \|x-y\|_{y,f})^2 \nabla^2 f(y) \preceq \nabla^2 f(x) \preceq \frac{1}{(1 - \|x-y\|_{y,f})^2} \nabla^2 f(y) .$$

We will use the following properties of the self-concordant barrier.

Lemma 21 (Nesterov and Nemirovskii 1994, Proposition 2.3.2) *Let g be a ϑ-self-concordant barrier for $\mathcal{K}$. Then, for any $x, y \in \operatorname{int}(\mathcal{K})$,*

$$g(y) - g(x) \le \vartheta \log\left(\frac{1}{1 - \pi(y; x)}\right),$$

where $\pi(y, x) = \inf\{s \ge 0 \colon x + s^{-1}(y - x) \in \mathcal{K}\}$ is the Minkowski function on $\mathcal{K}$ with pole at x.

The following lemma will be used in the proof of Theorem 13.

Lemma 22 *Let f be a ϑ-self-concordant barrier on $\mathcal{K}$. Then, for any $x \in \operatorname{int}(\mathcal{K})$ such that $\nabla^2 f(x)$ is invertible, it holds that $\|\nabla f(x)\|_{*,x,f}^2 \le \vartheta$.*

Proof This follows from the definition of the self-concordant barrier in Theorem 19 with $h = (\nabla^2 f(x))^{-1} f(x)$. ■

OFTRL with self-concordant barrier Using these properties, we can establish an upper bound on the regret of OFTRL when employing a self-concordant barrier regularizer and an adaptive learning rate. To our knowledge, few thorough analyses derive the negative term when using an adaptive learning rate in OFTRL with self-concordant barriers.

Lemma 23 (RVU bound for OFTRL with self-concordant barrier and adaptive learning rate) *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a nonempty closed convex set with a diameter $D = \max_{x,y\in\mathcal{K}}\|x - y\|$. Let ϕ be a ϑ-self-concordant barrier for $\mathcal{K}$ and $\psi^{(t)}(x) = \frac{1}{\eta^{(t)}}\phi(x)$ be a regularizer with nonincreasing and nonnegative learning rate $\{\eta^{(t)}\}_{t=1}^T$. For this $\psi^{(t)}$, consider the OFTRL update $x^{(t)} \in \arg\min_{x\in\mathcal{K}}\{\langle x, m^{(t)} + \sum_{s=1}^{t-1}\ell^{(s)}\rangle + \psi^{(t)}(x)\}$. Suppose that the sequence of iterates $\{x^{(t)}\}_{t=1}^T$ satisfies $\|x^{(t+1)} - x^{(t)}\|_{x^{(t)},\phi} \le 1/2$ for all $t \in [T]$. Then, for any $x^* \in \mathcal{K}$,*

$$\sum_{t=1}^T \langle x^{(t)} - x^*, \ell^{(t)}\rangle \le \frac{\vartheta \log T}{\eta^{(T+1)}} + \sum_{t=1}^T 4\eta^{(t)}\|\ell^{(t)} - m^{(t)}\|_{*,x^{(t)},\phi}^2 - \sum_{t=1}^T \frac{1}{16\eta^{(t)}}\|x^{(t+1)} - x^{(t)}\|_{x^{(t)},\phi}^2 + 3DL,$$

where $L = \max\{\max_{t\in[T]}\|\ell^{(t)}\|_*, \max_{t\in[T+1]}\|m^{(t)}\|_*\}$.

Proof We will upper bound the RHS of (7) in Lemma 16. We consider OFTRL with the regularizer $\bar{\psi}^{(t)}(\cdot) = \frac{1}{\eta^{(t)}}\bar{\phi}(\cdot)$ for $\bar{\phi}(x) = \phi(x) - \min_{x'\in\Delta_d}\phi(x') \ge 0$, where we note that the output of this OFTRL at each round t is same as $x^{(t)}$. Fix $x^* \in \mathcal{K}$ and define $v = (1 - 1/T)x^* + (1/T)x^{(1)} \in \mathcal{K}$, which satisfies $(1 - 1/T)(x^* - x^{(1)}) = v - x^{(1)}$. Then, from Lemma 16 with the fact that $\bar{\psi}^{(t)}$ is nonnegative and nondecreasing, we have

$$\begin{aligned}
\sum_{t=1}^T \langle x^{(t)} - x^*, \ell^{(t)}\rangle &= \sum_{t=1}^T \langle x^{(t)} - v, \ell^{(t)}\rangle + \sum_{t=1}^T \langle v - x^*, \ell^{(t)}\rangle \\
&\le \frac{\bar{\phi}(v)}{\eta^{(T+1)}} + \sum_{t=1}^T \Big(\langle x^{(t)} - x^{(t+1)}, \ell^{(t)} - m^{(t)}\rangle - D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)})\Big) + 3DL, \qquad (10)
\end{aligned}$$

where we used $\langle v - x^*, \ell^{(t)} \rangle \le \|v - x^*\| \|\ell^{(t)}\|_* = \frac{1}{T}\|x^{(1)} - x^*\| \|\ell^{(t)}\|_* \le DL/T$.

We first upper bound the first term in (10). From Lemma 21, $\bar{\phi}(v)$ in the first term in (10) is upper bounded by

$$\bar{\phi}(v) = \bar{\phi}(v) - \bar{\phi}(x^{(1)}) \le \vartheta \log\left(\frac{1}{1 - \pi(v; x^{(1)})}\right) \le \vartheta \log T\,, \tag{11}$$

where the equality follows from $\bar{\phi}(x^{(1)}) = 0$ and the last inequality follows from $\pi(v; x^{(1)}) \le 1 - 1/T$ since $x^{(1)} + (1 - 1/T)^{-1}(v - x^{(1)}) = x^{(1)} + (x^* - x^{(1)}) = x^{(1)} \in \mathcal{K}$.

We next upper bound the second term in (10). From Taylor's theorem, there exists a point $\xi^{(t)} = \gamma x^{(t+1)} + (1-\gamma)x^{(t)}$ with some $\gamma \in [0,1]$ such that $D_\phi(x^{(t+1)}, x^{(t)}) \ge \frac{1}{2}\|x^{(t+1)} - x^{(t)}\|^2_{\xi^{(t)},\phi}$. Hence from Lemma 20,

$$\begin{aligned}
D_\phi(x^{(t+1)}, x^{(t)}) &\ge \frac{1}{2}\|x^{(t+1)} - x^{(t)}\|^2_{\xi^{(t)},\phi} \\
&\ge \frac{1}{2}\left(1 - \|\xi^{(t)} - x^{(t)}\|_{x^{(t)},\phi}\right)^2 \|x^{(t+1)} - x^{(t)}\|^2_{x^{(t)},\phi} \\
&= \frac{1}{2}\left(1 - \gamma\|x^{(t+1)} - x^{(t)}\|_{x^{(t)},\phi}\right)^2 \|x^{(t+1)} - x^{(t)}\|^2_{x^{(t)},\phi} \\
&\ge \frac{1}{8}\|x^{(t+1)} - x^{(t)}\|^2_{x^{(t)},\phi}\,,
\end{aligned}$$

where the last inequality follows from the assumption that $\|x^{(t+1)} - x^{(t)}\|_{x^{(t)},\phi} \le 1/2$. Hence, $D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)}) = \frac{1}{\eta^{(t)}} D_\phi(x^{(t+1)}, x^{(t)}) \ge \frac{1}{8\eta^{(t)}}\|x^{(t+1)} - x^{(t)}\|^2_{x^{(t)},\phi}$. Therefore, using this inequality, we have

$$\begin{aligned}
&\langle x^{(t)} - x^{(t+1)}, \ell^{(t)} - m^{(t)} \rangle - D_{\psi^{(t)}}(x^{(t+1)}, x^{(t)}) \\
&\le \|x^{(t)} - y^*\|_{x^{(t)},\phi} \cdot \|\ell^{(t)} - m^{(t)}\|_{*,x^{(t)},\phi} - \frac{1}{8\eta^{(t)}}\|x^{(t+1)} - x^{(t)}\|^2_{x^{(t)},\phi} \\
&\le 4\eta^{(t)}\|\ell^{(t)} - m^{(t)}\|^2_{*,x^{(t)},\phi} - \frac{1}{16\eta^{(t)}}\|x^{(t+1)} - x^{(t)}\|^2_{x^{(t)},\phi}\,,
\end{aligned} \tag{12}$$

where the first inequality follows from Hölder's inequality and $b\sqrt{z} - az \le b^2/(4a)$ for $a > 0, b \ge 0$, and $z \ge 0$. Combining (10) with (11) and (12) completes the proof. ■

OFTRL with log-barrier regularizer Now we are ready to prove the RVU bound for OFTRL with the log-barrier regularizer and adaptive learning rate, which serves as a foundation for our analysis of multi-player general-sum games in Section 5.

Lemma 24 (RVU bound for OFTRL with log-barrier regularizer and adaptive learning rate) *Let $\psi^{(t)}(x) = \frac{1}{\eta^{(t)}}\phi(x)$ for $\phi(x) = -\sum_{k=1}^d \log(x(k))$ be the logarithmic barrier regularizer with nonincreasing learning rate $\eta^{(t)}$ and $x^{(t)} \in \arg\min_{x \in \Delta_d}\{\langle x, m^{(t)} + \sum_{s=1}^{t-1}\ell^{(s)}\rangle + \psi^{(t)}(x)\}$ be the output of OFTRL at round t. Suppose that $\|x^{(t+1)} - x^{(t)}\|_{x^{(t)},\phi} \le 1/2$. Then, for any $x^* \in \Delta_d$,*

$$\sum_{t=1}^T \langle x^{(t)} - x^*, \ell^{(t)} \rangle \le \frac{d \log T}{\eta^{(T+1)}} + \sum_{t=1}^T 4\eta^{(t)}\|\ell^{(t)} - m^{(t)}\|^2_{*,x^{(t)},\phi} - \sum_{t=1}^T \frac{1}{16\eta^{(t)}}\|x^{(t+1)} - x^{(t)}\|^2_{x^{(t)},\phi} + 6L\,, \tag{13}$$

where $L = \max\{\max_{t\in[T]}\|\ell^{(t)}\|_*, \max_{t\in[T+1]}\|m^{(t)}\|_*\}$.

Proof Combining Lemma 23 with the fact that ϕ is a d-self-concordant barrier and $\max_{x,y\in\Delta_d}\|x-y\|_1 = 2$ yields the desired bound. ∎

For a constant learning rate, the RVU bound for OFTRL has been studied (Anagnostides et al., 2022b). However, to our knowledge, no RVU bounds have been established for the case with an adaptive learning rate, which is of independent interest. One can observe that the key to proving Lemma 24 (and more generally Lemma 23) is the stability assumption for $x^{(t)}$, namely, $\|x^{(t+1)} - x^{(t)}\|_{x^{(t)},\phi} \le 1/2$. This assumption can be verified in swap regret minimization by analyzing the stability of a Markov chain, as detailed in Section 5 and Appendix E.

Appendix C. Deferred Details from Section 3

Here we provide the omitted details from Section 3. We first provide the proof of Proposition 9.
Proof [Proof of Proposition 9] From the triangle inequality and the Cauchy–Schwarz inequality, we have

$$\left|\mathsf{Reg}^T_{x_i,u_i}(x^*) - \mathsf{Reg}^T_{\widehat{x}_i,u_i}(x^*)\right| = \left|\sum_{t=1}^T \langle \widehat{x}_i^{(t)} - x_i^{(t)}, u_i^{(t)}\rangle\right| \le \sum_{t=1}^T \|\widehat{x}_i^{(t)} - x_i^{(t)}\|_1 \|u_i^{(t)}\|_\infty \le \widehat{C}_i\,.$$

Similarly, we also have

$$\left|\mathsf{Reg}^T_{x_i,u_i}(x^*) - \mathsf{Reg}^T_{x_i,\widetilde{u}_i}(x^*)\right| = \left|\sum_{t=1}^T \langle x^* - x_i^{(t)}, u_i^{(t)} - \widetilde{u}_i^{(t)}\rangle\right| \le \sum_{t=1}^T \|x^* - x_i^{(t)}\|_1 \|u_i^{(t)} - \widetilde{u}_i^{(t)}\|_\infty \le 2\widetilde{C}_i\,.$$

The other inequalities can be proven in the same manner. ∎

A similar proposition to Proposition 9 also holds for the four types of swap regret in Section 3.

Proposition 25 *For any $i \in [n]$ and $M \in \mathcal{M}_{m_i}$, it holds that $|\mathsf{SwapReg}^T_{x_i,u_i}(M) - \mathsf{SwapReg}^T_{\widehat{x}_i,u_i}(M)| \le 2\widehat{C}_i$, $|\mathsf{SwapReg}^T_{x_i,\widetilde{u}_i}(M) - \mathsf{SwapReg}^T_{\widehat{x}_i,\widetilde{u}_i}(M)| \le 2\widehat{C}_i$, $|\mathsf{SwapReg}^T_{x_i,u_i}(M) - \mathsf{SwapReg}^T_{x_i,\widetilde{u}_i}(M)| \le 2\widetilde{C}_i$, and $|\mathsf{SwapReg}^T_{\widehat{x}_i,u_i}(v) - \mathsf{SwapReg}^T_{\widehat{x}_i,\widetilde{u}_i}(M)| \le 2\widetilde{C}_i$.*

Proof [Proof of Proposition 25] From the triangle inequality and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\left|\mathsf{SwapReg}^T_{x_i,u_i}(M) - \mathsf{SwapReg}^T_{\widehat{x}_i,u_i}(M)\right| &= \left|\sum_{t=1}^T \langle \widehat{x}_i^{(t)} - x_i^{(t)}, M u_i^{(t)} - u_i^{(t)}\rangle\right| \\ &\le \sum_{t=1}^T \|\widehat{x}_i^{(t)} - x_i^{(t)}\|_1 \|M u_i^{(t)} - u_i^{(t)}\|_\infty \le 2\widehat{C}_i\,.\end{aligned}$$

Similarly, we also have

$$\begin{aligned}&\left|\mathsf{SwapReg}^T_{x_i,u_i}(M) - \mathsf{SwapReg}^T_{x_i,\widetilde{u}_i}(M)\right| = \left|\sum_{t=1}^T \left\langle x_i^{(t)}, M(u_i^{(t)} - \widetilde{u}_i^{(t)}) - (u_i^{(t)} - \widetilde{u}_i^{(t)})\right\rangle\right| \\ &\le \sum_{t=1}^T \|x_i^{(t)}\|_1 \left(\|M(u_i^{(t)} - \widetilde{u}_i^{(t)})\|_\infty + \|u_i^{(t)} - \widetilde{u}_i^{(t)}\|_\infty\right) \le 2\sum_{t=1}^T \|u_i^{(t)} - \widetilde{u}_i^{(t)}\|_\infty \le 2\widetilde{C}_i\,,\end{aligned}$$

where we used $\|Mx\|_\infty = \max_{k\in[m]}|\langle M(k,\cdot), x\rangle| \le \max_{k\in[m]}\|M(k,\cdot)\|_1\|x\|_\infty \le \|x\|_\infty$ for a row stochastic matrix $M \in \mathcal{M}_m$ and a vector $x \in \mathbb{R}^m$. The other inequalities can be proven in the same manner. ■

Appendix D. Deferred Proofs for Two-Player Zero-Sum Games from Section 4

This section provides the details and deferred proofs from Section 4.

D.1. Corrupted procedure in two-player zero-sum games

For clarity, we summarize the corrupted learning procedure for a two-player zero-sum game with payoff matrix A. This procedure corresponds to the one used for multiplayer general-sum games described in Section 3.

At each round $t = 1, \dots, T$:

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. The x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and the y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
3. The x-player observes a corrupted expected reward vector $\widetilde{g}^{(t)} = g^{(t)} + \widetilde{c}_x^{(t)}$ for $g^{(t)} = Ay^{(t)}$ and the y-player observes a corrupted expected loss vector $\widetilde{\ell}^{(t)} = \ell^{(t)} + \widetilde{c}_y^{(t)}$ for $\ell^{(t)} = A^\top x^{(t)}$;
4. The x-player gains a payoff of $\langle x^{(t)}, g^{(t)}\rangle$ in Setting (I) and $\langle x^{(t)}, \widetilde{g}^{(t)}\rangle$ in Setting (II), and the y-player incurs a loss of $\langle y^{(t)}, \ell^{(t)}\rangle$ in Setting (I) and $\langle y^{(t)}, \widetilde{\ell}^{(t)}\rangle$ in Setting (II);

Here, $\widehat{c}_x^{(t)}$ and $\widetilde{c}_x^{(t)}$ are corruption vectors for strategies and utility of the x-player at round t, respectively, such that $\sum_{t=1}^T \|\widehat{c}_x^{(t)}\|_1 = \sum_{t=1}^T \|x^{(t)} - \widehat{x}^{(t)}\|_1 \le \widehat{C}_x$, $\sum_{t=1}^T \|\widetilde{c}_x^{(t)}\|_\infty = \sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x$, and $C_x = \widehat{C}_x + 2\widetilde{C}_x$. Similarly, $\widehat{c}_y^{(t)}$ and $\widetilde{c}_y^{(t)}$ are corruption levels of strategies and utility of the y-player, respectively, such that $\sum_{t=1}^T \|\widehat{c}_y^{(t)}\|_1 = \sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1 \le \widehat{C}_y$, $\sum_{t=1}^T \|\widetilde{c}_y^{(t)}\|_\infty = \sum_{t=1}^T \|\ell^{(t)} - \widetilde{\ell}^{(t)}\|_\infty \le \widetilde{C}_y$, and $C_y = \widehat{C}_y + 2\widetilde{C}_y$.

D.2. Preliminary lemmas

Lemma 26 *Let (x^*, y^*) be a Nash equilibrium. Then,* $\mathrm{Reg}_{x,g}^T(x^*) + \mathrm{Reg}_{y,\ell}^T(y^*) \ge 0$.

Proof We have

$$\mathrm{Reg}_{x,g}^T(x^*) + \mathrm{Reg}_{y,\ell}^T(y^*) = \sum_{t=1}^T \langle x^* - x^{(t)}, Ay^{(t)}\rangle + \sum_{t=1}^T \langle y^{(t)} - y^*, Ax^{(t)}\rangle$$
$$= \sum_{t=1}^T \langle x^*, Ay^{(t)}\rangle - \sum_{t=1}^T \langle y^*, Ax^{(t)}\rangle = \sum_{t=1}^T \langle x^*, Ay^{(t)} - Ay^*\rangle - \sum_{t=1}^T \langle y^*, Ax^{(t)} - Ax^*\rangle \ge 0,$$

where the last inequality follows from the definition of the Nash equilibrium. ■

D.3. Proof of Lemma 11

Proof [Proof of Lemma 11] From Lemma 17, the regret of the x-player is bounded by

$$\mathsf{Reg}^T_{\widehat{x},\widetilde{g}} \leq \frac{\log m_x}{\eta_x^{(T+1)}} + \sum_{t=1}^T \eta_x^{(t)} \|\widetilde{g}^{(t)} - \widetilde{g}^{(t-1)}\|_\infty^2 - \sum_{t=1}^T \frac{1}{4\eta_x^{(t)}} \|\widehat{x}^{(t+1)} - \widehat{x}^{(t)}\|_1^2 + 2\,. \tag{14}$$

From the definition of $\eta_x^{(t)}$, the third term in (14) is upper bounded by

$$\sum_{t=1}^T \eta_x^{(t)} \|\widetilde{g}^{(t)} - \widetilde{g}^{(t-1)}\|_\infty^2 \leq \sqrt{\frac{\log_+(m_x)}{2}} \sum_{t=1}^T \frac{\|\widetilde{g}^{(t)} - \widetilde{g}^{(t-1)}\|_\infty^2}{\sqrt{4 + P_\infty^{t-1}(\widetilde{g})}} \leq \sqrt{2P_\infty^T(\widetilde{g}) \log_+(m_x)}\,, \tag{15}$$

where the first inequality follows from $\log_+(m_x) \geq 4$ and the last inequality from $\sum_{t=1}^T z_t / \sqrt{4 + \sum_{s=1}^{t-1} z_s} \leq 2\sqrt{\sum_{t=1}^T z_t}$ for $z_t \in [0, 4]$. Hence, from (14), (15), and the definition of $\eta_x^{(T+1)}$, we obtain

$$\mathsf{Reg}^T_{\widehat{x},\widetilde{g}} \leq 2\sqrt{2\log_+(m_x)\big(\log_+(m_x) + P_\infty^T(\widetilde{g})\big)} - \sum_{t=1}^T \frac{1}{4\eta_x^{(t)}} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + 2\,. \tag{16}$$

Now we have $P_\infty^T(\widetilde{g}) = \sum_{t=1}^T \|\widetilde{g}^{(t)} - \widetilde{g}^{(t-1)}\|_\infty^2 \leq 2\sum_{t=1}^T \|g^{(t)} - g^{(t-1)}\|_\infty^2 + 8\widetilde{C}_x$ and

$$\begin{aligned}
\sum_{t=1}^T \|g^{(t)} - g^{(t-1)}\|_\infty^2 &= \sum_{t=1}^T \|A(y^{(t)} - y^{(t-1)})\|_\infty^2 \leq \sum_{t=1}^T \|y^{(t)} - y^{(t-1)}\|_1^2 \\
&\leq \sum_{t=1}^T \Big(2\|y^{(t)} - \widehat{y}^{(t)}\|_1^2 + 4\|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 + 2\|\widehat{y}^{(t-1)} - y^{(t-1)}\|_1^2\Big) \\
&\leq 4\sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1^2 + 4\sum_{t=1}^T \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \leq 8\widehat{C}_y + 4P_1^T(\widehat{y})\,.
\end{aligned} \tag{17}$$

Hence, plugging the last two inequalities in (16) and using $\eta_x^{(t)} \leq 1/\sqrt{2}$ for all $t \in [T]$, we obtain

$$\mathsf{Reg}^T_{\widehat{x},\widetilde{g}} \leq 2\sqrt{2\log_+(m_x)\Big(\log_+(m_x) + 8(\widehat{C}_y + \widetilde{C}_x) + 4P_1^T(\widehat{y})\Big)} - \frac{1}{\sqrt{8}} P_1^T(\widehat{x}) + 2\,.$$

We next consider the regret of the y-player. Using Lemma 17 and following the similar analysis as above, we can upper bound the regret of the y-player as

$$\mathsf{Reg}^T_{\widehat{y},\widetilde{\ell}} \leq 2\sqrt{2\log_+(m_y)\left(\log_+(m_y) + \sum_{t=1}^T \|\widetilde{\ell}^{(t)} - \widetilde{\ell}^{(t-1)}\|_\infty^2\right)} - \sum_{t=2}^T \frac{1}{4\eta_y^{(t)}} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 + 2\,. \tag{18}$$

Now we have $\sum_{t=1}^{T}\|\widetilde{\ell}^{(t)}-\widetilde{\ell}^{(t-1)}\|_\infty^2 \le 2\sum_{t=1}^{T}\|\ell^{(t)}-\ell^{(t-1)}\|_\infty^2 + 8\widetilde{C}_y$ and

$$\begin{aligned}
\sum_{t=1}^{T}\|\ell^{(t)}-\ell^{(t-1)}\|_\infty^2 &\le \sum_{t=1}^{T}\|x^{(t)}-x^{(t-1)}\|_1^2 \\
&\le \sum_{t=1}^{T}\Big(2\|x^{(t)}-\widehat{x}^{(t)}\|_1^2 + 4\|\widehat{x}^{(t)}-\widehat{x}^{(t-1)}\|_1^2 + 2\|\widehat{x}^{(t-1)}-x^{(t-1)}\|_1^2\Big) \\
&\le 4\sum_{t=1}^{T}\|x^{(t)}-\widehat{x}^{(t)}\|_1^2 + 4\sum_{t=1}^{T}\|\widehat{x}^{(t)}-\widehat{x}^{(t-1)}\|_1^2 \le 8\widehat{C}_x + 4P_1^T(\widehat{x})\,.
\end{aligned}$$

Combining (18) with the last two inequalities and using $\eta_y^{(t)} \le 1/\sqrt{2}$, we obtain

$$\mathsf{Reg}_{\widehat{y},\widetilde{\ell}}^T \le 2\sqrt{2\log_+(m_y)\Big(\log_+(m_y)+8(\widehat{C}_x+\widetilde{C}_y)+4P_1^T(\widehat{x})\Big)} - \frac{1}{\sqrt{8}}P_1^T(\widehat{y}) + 2\,,$$

which completes the proof. ■

D.4. Remaining proof of Theorem 10

The following bounds are omitted bounds in Theorem 10 for the y-player:

$$\begin{aligned}
\mathsf{Reg}_{y,\ell}^T &\lesssim \min\Big\{\sqrt{(\log(m_x m_y)+C_x+C_y)\log m_y},\sqrt{\big(P_\infty^T(\widetilde{\ell})+\log m_y\big)\log m_y}\Big\} + C_y\,,\\
\mathsf{Reg}_{y,\widetilde{\ell}}^T &\lesssim \min\Big\{\sqrt{(\log(m_x m_y)+C_x+C_y)\log m_y},\sqrt{\big(P_\infty^T(\widetilde{\ell})+\log m_y\big)\log m_y}\Big\} + \widehat{C}_y\,.
\end{aligned}$$

Below, we include the omitted proofs of Theorem 10:

Proof [Remaining proof of Theorem 10] Here we provide the remaining proof to prove the upper bounds on $\mathsf{Reg}_{x,\widetilde{g}}^T$ and $\mathsf{Reg}_{y,\widetilde{\ell}}^T$. From Lemma 11 and Proposition 9, we have

$$\mathsf{Reg}_{x,\widetilde{g}}^T \le \widehat{C}_x + \sqrt{8\log_+(m_x)\Big(1+\log_+(m_x)+8(\widehat{C}_y+\widetilde{C}_x)+4P_1^T(\widehat{y})\Big)} - \frac{1}{\sqrt{8}}P_1^T(\widehat{x}) + 2\,, \tag{19}$$

$$\mathsf{Reg}_{y,\widetilde{\ell}}^T \le \widehat{C}_y + \sqrt{8\log_+(m_y)\Big(1+\log_+(m_y)+8(\widehat{C}_x+\widetilde{C}_y)+P_1^T(\widehat{x})\Big)} - \frac{1}{\sqrt{8}}P_1^T(\widehat{y}) + 2\,. \tag{20}$$

Summing up the last two upper bounds, we have

$$\begin{aligned}
&\mathsf{Reg}_{x,\widetilde{g}}^T + \mathsf{Reg}_{y,\widetilde{\ell}}^T \\
&\le \sqrt{8\log_+(m_x)\Big(1+\log_+(m_x)+8(\widehat{C}_y+\widetilde{C}_x)+4P_1^T(\widehat{y})\Big)} - \frac{1}{\sqrt{8}}P_1^T(\widehat{x}) + \widehat{C}_x + 4 \\
&\quad + \sqrt{8\log_+(m_y)\Big(1+\log_+(m_y)+8(\widehat{C}_x+\widetilde{C}_y)+4P_1^T(\widehat{x})\Big)} - \frac{1}{\sqrt{8}}P_1^T(\widehat{y}) + \widehat{C}_y \\
&= O\Big(\sqrt{(\widehat{C}_y+\widetilde{C}_x)\log m_x}+\sqrt{(\widehat{C}_x+\widetilde{C}_y)\log m_y}+\log(m_x m_y)+\widehat{C}_x+\widehat{C}_y\Big) - \frac{1}{4\sqrt{2}}\big(P_1^T(\widehat{x})+P_1^T(\widehat{y})\big)\,,
\end{aligned} \tag{21}$$

where we used $b\sqrt{z} - az \le b^2/(4a)$ for $a > 0$, $b \ge 0$, and $z \ge 0$. Now from Proposition 9 and Lemma 26, we have

$$\mathsf{Reg}^T_{x,\widetilde{g}} + \mathsf{Reg}^T_{x,\widetilde{\ell}} \ge \mathsf{Reg}^T_{x,g} + \mathsf{Reg}^T_{y,\ell} - 2(\widetilde{C}_x + \widetilde{C}_y) \ge -2(\widetilde{C}_x + \widetilde{C}_y)\,.$$

Combining this inequality and (21), we have

$$\begin{aligned} P_1^T(\widehat{x}) + P_1^T(\widehat{y}) &\lesssim \sqrt{(\widehat{C}_y + \widetilde{C}_x)\log m_x} + \sqrt{(\widehat{C}_x + \widetilde{C}_y)\log m_y} + \log(m_x m_y) + C_x + C_y \\ &\lesssim \log(m_x m_y) + C_x + C_y\,, \end{aligned} \tag{22}$$

where the last line follows from the AM–GM inequality and the definitions of C_x and C_y. Finally, plugging (22) in (19) and (20) gives the desired bounds on $\mathsf{Reg}^T_{x,\widetilde{g}}$ and $\mathsf{Reg}^T_{y,\widetilde{\ell}}$. The upper bound of $\mathsf{Reg}^T_{x,\widetilde{g}} \lesssim \sqrt{(P_\infty^T(\widetilde{g}) + \log m_x)\log m_x} + C_x$ follows from (16) in the proof of Lemma 11. ■

D.5. Analysis when the opponent follows optimistic Hedge with a constant learning rate

The proposed learning dynamics in Sections 4.1 and 5, remain effective even when the opponent follows OFTRL with a constant learning rate. Specifically, in two-player zero-sum games, if the opponent follows the algorithm of Syrgkanis et al. (2015), and in multiplayer general-sum games, if the opponents follow the algorithm of Anagnostides et al. (2022b), then a player employing our algorithm can still achieve regret bounds comparable to those established in Theorems 10 and 13 respectively, in both the honest and corrupted regimes. This provides an incentive for players to employ our learning dynamics. Here, we focus on demonstrating this result for two-player zero-sum games.

Theorem 27 *Suppose that the x-player uses the algorithm in Section 4.1 to obtain strategies $\{\widehat{x}^{(t)}\}_{t=1}^T$ and the y-player uses optimistic Hedge with a fixed learning rate $\eta_y > 0$ to obtain $\{\widehat{y}^{(t)}\}_{t=1}^T$. Then, the regret of the x-player is upper bounded as*

$$\begin{aligned} \mathsf{Reg}^T_{x,g} \lesssim &\sqrt{\log_+(m_x)\big(\log(m_x m_y) + \max\{1,\eta_y^2\}\log_+(m_x) + \widehat{C}_y + \widetilde{C}_x + \max\{\eta_y,\eta_y^2\}(C_x + C_y)\big)} \\ &+ \eta_y^2 \log_+(m_x)\mathbb{1}\Big[\eta_y \ge 1/(32\sqrt{2})\Big] + C_x\,. \end{aligned}$$

When $\eta_y < 1/(32\sqrt{2})$, the regret of the y-player is upper bounded by

$$\mathsf{Reg}^T_{y,\ell} \le C_y + \frac{\log m_y}{\eta_y} + O(\eta_y(\widehat{C}_x + \widetilde{C}_y)) + \frac{1}{\frac{1}{4\sqrt{2}} - 8\eta_y} O\big(\log(m_x m_y) + \eta_y^2 \log m_x + \eta_y(C_x + C_y)\big)\,.$$

This result shows that, regardless of the choice of constant learning rate in the optimistic Hedge algorithm employed by the y-player, the x-player can attain the same regret upper bound as in Theorem 10. On the other hand, the regret of the y-player is linearly affected by the corruption level C_y by the opponent, *i.e.,* the x-player. It is also worth noting that, as long as both η_y and $1/(4\sqrt{2}) - 8\eta_y$ are positive absolute constants, a convergence rate to a Nash equilibrium similar to that in Theorem 10 can still be guaranteed.

Proof From Lemma 11 and Proposition 9, the regret of the x-player is bounded as

$$\mathsf{Reg}_{x,g}^T \leq C_x + \sqrt{8\log_+(m_x)\big(1+\log_+(m_x) + 8(\widehat{C}_y + \widetilde{C}_x) + 4P_1^T(\widehat{y})\big)} - \frac{1}{\sqrt{8}}P_1^T(\widehat{x}) + 2\,, \quad (23)$$

We next consider the regret of the y-player. From Lemma 17, the regret of the y-player is bounded as

$$\mathsf{Reg}_{\widehat{y},\widetilde{\ell}}^T \leq \frac{\log m_x}{\eta_y} + \eta_y \sum_{t=1}^T \|\widetilde{\ell}^{(t)} - \widetilde{\ell}^{(t-1)}\|_\infty^2 - \frac{1}{4\eta_y}\sum_{t=1}^T \|\widehat{y}^{(t+1)} - \widehat{y}^{(t)}\|_1^2 + 2\,.$$

From the proof of Lemma 11, we know that $\sum_{t=1}^T \|\widetilde{\ell}^{(t)} - \widetilde{\ell}^{(t-1)}\|_\infty^2 \leq 2\sum_{t=1}^T \|\ell^{(t)} - \ell^{(t-1)}\|_\infty^2 + 8\widetilde{C}_y$ and $\sum_{t=1}^T \|\ell^{(t)} - \ell^{(t-1)}\|_\infty^2 \leq 8\widehat{C}_x + 4P_1^T(\widehat{x})$, and thus

$$\mathsf{Reg}_{\widehat{y},\widetilde{\ell}}^T \leq \frac{\log m_y}{\eta_y} + 8\eta_y\Big(2\widehat{C}_x + \widetilde{C}_y + P_1^T(\widehat{x})\Big) - \frac{1}{4\eta_y}P_1^T(\widehat{y}) + 2\,.$$

Hence, the last inequality with Proposition 9 gives

$$\mathsf{Reg}_{y,\ell}^T \leq C_y + \frac{\log m_y}{\eta_y} + 8\eta_y\Big(2\widehat{C}_x + \widetilde{C}_y + P_1^T(\widehat{x})\Big) - \frac{1}{4\eta_y}P_1^T(\widehat{y}) + 2\,, \quad (24)$$

Summing up (23) and (24), we can upper bound the social regret $\mathsf{Reg}_{x,g}^T + \mathsf{Reg}_{y,\ell}^T$ by

$$\begin{aligned}
&\sqrt{8\log_+(m_x)\Big(1+\log_+(m_x) + 8(\widehat{C}_y + \widetilde{C}_x) + 4P_1^T(\widehat{y})\Big)} - \frac{1}{\sqrt{8}}P_1^T(\widehat{x}) + C_x + 4 \\
&\quad + \frac{\log m_y}{\eta_y} + 8\eta_y\Big(2\widehat{C}_x + \widetilde{C}_y + P_1^T(\widehat{x})\Big) - \frac{1}{4\eta_y}P_1^T(\widehat{y}) + C_y \\
&= O\bigg(\sqrt{(\widehat{C}_y + \widetilde{C}_x)\log m_x} + \eta_y \log m_x + \frac{\log m_y}{\eta_y} + \max\{1,\eta_y\}(C_x + C_y)\bigg) \\
&\qquad + \bigg(8\eta_y - \frac{1}{4\sqrt{2}}\bigg)P_1^T(\widehat{x}) - \frac{1}{4\eta_y}P_1^T(\widehat{y})\,,
\end{aligned}$$

where we used $b\sqrt{z} - az \leq b^2/(4a)$ for $a > 0$, $b \geq 0$ and $z \geq 0$. Combining this with $\mathsf{Reg}_{x,g}^T + \mathsf{Reg}_{y,\ell}^T \geq 0$ in Lemma 26, we have

$$\begin{aligned}
\frac{1}{4\eta_y}P_1^T(\widehat{y}) &= O\bigg(\sqrt{(\widehat{C}_y + \widetilde{C}_x)\log m_x} + \eta_y \log m_x + \frac{\log m_y}{\eta_y} + \max\{1,\eta_y\}(C_x + C_y)\bigg) \\
&\qquad + \bigg(8\eta_y - \frac{1}{4\sqrt{2}}\bigg)P_1^T(\widehat{x}) \\
&= O\bigg(\frac{\log(m_x m_y)}{\eta_y} + \eta_y \log m_x + \max\{1,\eta_y\}(C_x + C_y)\bigg) + \bigg(8\eta_y - \frac{1}{4\sqrt{2}}\bigg)P_1^T(\widehat{x})\,,
\end{aligned} \quad (25)$$

where the last line follows from the AM–GM inequality and the definitions of C_x and C_y. When $\eta_y \geq 1/(32\sqrt{2})$, plugging (25) in (23) gives

$$
\begin{aligned}
\mathsf{Reg}_{x,g}^T &\lesssim \sqrt{\log_+(m_x)\big(\log_+(m_x) + \widehat{C}_y + \widetilde{C}_x + \log(m_x m_y) + \eta_y^2 \log m_x + \max\{\eta_y, \eta_y^2\}(C_x + C_y)\big)} \\
&\quad + \sqrt{\eta_y^2 \log_+(m_x) P_1^T(\widehat{x})} - \frac{1}{\sqrt{8}} P_1^T(\widehat{x}) + C_x \\
&\lesssim \sqrt{\log_+(m_x)\big(\log_+(m_x) + \widehat{C}_y + \widetilde{C}_x + \log(m_x m_y) + \eta_y^2 \log m_x + \max\{\eta_y, \eta_y^2\}(C_x + C_y)\big)} \\
&\quad + \eta_y^2 \log_+(m_x) + C_x\,, \qquad (26)
\end{aligned}
$$

where we used $b\sqrt{z} - az \leq b^2/(4a)$ for $a > 0$, $b \geq 0$ and $z \geq 0$. When $\eta_y < 1/(32\sqrt{2})$, we upper bound the last term in (25) by zero and repeating the similar analysis as in (26) gives the desired regret upper bound of the x-player.

Further, when $\eta_y < 1/(32\sqrt{2})$, from (25) we have

$$
\left(\frac{1}{4\sqrt{2}} - 8\eta_y\right) P_1^T(\widehat{x}) = O\!\left(\frac{\log(m_x m_y)}{\eta_y} + \eta_y \log m_x + \max\{1, \eta_y\}(C_x + C_y)\right),
$$

and thus, combining this with (24) gives

$$
\begin{aligned}
\mathsf{Reg}_{y,\ell}^T &\leq C_y + \frac{\log m_y}{\eta_y} + 8\eta_y \left(2\widehat{C}_x + \widetilde{C}_y\right) \\
&\quad + \frac{1}{\frac{1}{4\sqrt{2}} - 8\eta_y} O\big(\log(m_x m_y) + \eta_y^2 \log m_x + \eta_y (C_x + C_y)\big) + 2\,,
\end{aligned}
$$

which is the desired regret upper bound of the y-player. ■

Appendix E. Deferred Proofs for Multiplayer General-Sum Games from Section 5

E.1. Proof of Lemma 12

Proof The swap regret $\mathsf{SwapReg}_{\widehat{x}_i, \widetilde{u}_i}^T(M)$ can be decomposed as

$$
\begin{aligned}
\mathsf{SwapReg}_{\widehat{x}_i, \widetilde{u}_i}^T(M) &= \sum_{t=1}^T \langle \widehat{x}_i^{(t)}, M \widetilde{u}_i^{(t)} - \widetilde{u}_i^{(t)} \rangle \\
&= \sum_{t=1}^T \left\langle M^\top \widehat{x}_i^{(t)} - (Q_i^{(t)})^\top \widehat{x}_i^{(t)}, \widetilde{u}_i^{(t)} \right\rangle + \sum_{t=1}^T \left\langle (Q_i^{(t)})^\top \widehat{x}_i^{(t)} - \widehat{x}_i^{(t)}, \widetilde{u}_i^{(t)} \right\rangle \\
&= \sum_{t=1}^T \left\langle (M - Q_i^{(t)})^\top \widehat{x}_i^{(t)}, \widetilde{u}_i^{(t)} \right\rangle \qquad \text{(since } (Q_i^{(t)})^\top \widehat{x}_i^{(t)} = \widehat{x}_i^{(t)}\text{)} \\
&= \sum_{t=1}^T \operatorname{tr}\left(\left(M - Q_i^{(t)}\right) \widetilde{u}_i^{(t)} (\widehat{x}_i^{(t)})^\top\right) \\
&= \sum_{a \in \mathcal{A}_i} \sum_{t=1}^T \left\langle M(a, \cdot) - Q_i^{(t)}(a, \cdot), \widehat{x}_i^{(t)}(a) \widetilde{u}_i^{(t)} \right\rangle
\end{aligned}
$$

$$= \sum_{a \in \mathcal{A}_i} \sum_{t=1}^{T} \left\langle M(a, \cdot) - y_{i,a}^{(t)}, \widetilde{u}_{i,a}^{(t)} \right\rangle \qquad (\widetilde{u}_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a)\widetilde{u}_i^{(t)})$$
$$= \sum_{a \in \mathcal{A}_i} \widetilde{\mathsf{Reg}}_{i,a}^{T}(M(a, \cdot)) \,,$$

which completes the proof. ■

E.2. Preliminary lemmas for the proof of Theorem 13

Here we prepare several lemmas to prove Theorem 13. We begin with the following lemma, which is useful to evaluate the stability of the Markov chain in the proof of Lemma 30.

Lemma 28 *Let $\mathcal{K}$ be a bounded nonempty convex set and $y \in \mathcal{K}$. Let $\delta > 0$ and f be a real-valued strictly convex function over $\mathcal{K}$ and $x^* = \arg\min_{x' \in \mathcal{K}} f(x')$ be the unique minimizer of f. Suppose that for any $z \in \mathcal{K}$ such that $\|z - y\| = \delta$ for a norm $\|\cdot\|$, it holds that $f(z) \geq f(y)$. Then, $\|x^* - y\| < \delta$.*

Proof We begin by proving that for any $z \in \mathbb{R}^d$ such that $\|z - y\| \geq \delta$, it holds that $f(z) \geq f(y)$. This follows from the convexity of f. To formally prove this, we take $z \in \mathcal{K}$ satisfying $\|z - y\| > \delta$ arbitrarily and define $w(\gamma) = (1-\gamma)y + \gamma z = y + \gamma(z - y)$ for $\gamma \in [0, 1]$. Define $\gamma^\circ = \delta / \|z - y\|$, which satisfies $\|w(\gamma^\circ) - y\| = \gamma^\circ \|z - y\| = \delta$. Then, we have

$$f(y) \leq f(w(\gamma^\circ)) \leq (1 - \gamma^\circ) f(y) + \gamma^\circ f(z) \,, \tag{27}$$

where the first inequality follows from $\|w(\gamma^\circ) - y\| = \delta$ and the assumption that for any $z \in \mathcal{K}$ satisfying $\|z - y\| = \delta$, $f(z) \geq f(y)$, and the second inequality from the convexity of f. Rearranging the terms in (27) gives $f(y) \leq f(z)$.

Therefore, we have $\{z \in \mathcal{K} \colon \|z - y\| \geq \delta\} \subseteq \{z \in \mathcal{K} \colon f(z) \geq f(y)\}$ and thus from the assumption that x^* is the minimizer of the strictly convex function f,

$$x^* \in \{z \in \mathcal{K} \colon f(z) < f(y)\} \subseteq \{z \in \mathcal{K} \colon \|z - y\| < \delta\} \,,$$

which implies $\|x^* - y\| < \delta$. This completes the proof. ■

The following lemma upper bounds the increase of the reciprocal of the learning rate.

Lemma 29 *The learning rate $\{\eta_{i,a}^{(t)}\}_{t=1}^T$ in (6) satisfies*

$$\frac{1}{\eta_{i,a}^{(t)}} - \frac{1}{\eta_{i,a}^{(t-1)}} \leq \frac{\sqrt{8}(\widehat{x}_i^{(t-1)}(a) + \widehat{x}_i^{(t-2)}(a))}{\sqrt{m_i \log T}} \,.$$

Proof From the definition of the learning rate $\eta_{i,a}^{(t)}$, we have

$$\frac{1}{\eta_{i,a}^{(t)}} - \frac{1}{\eta_{i,a}^{(t-1)}} = \frac{1}{\min\left\{\sqrt{\frac{m_i \log T/8}{4 + \sum_{s=1}^{t-1} \|\widetilde{u}_{i,a}^{(s)} - \widetilde{u}_{i,a}^{(s-1)}\|_\infty^2}}, \frac{1}{256 n \sqrt{m_i}}\right\}} - \frac{1}{\min\left\{\sqrt{\frac{m_i \log T/8}{4 + \sum_{s=1}^{t-2} \|\widetilde{u}_{i,a}^{(s)} - \widetilde{u}_{i,a}^{(s-1)}\|_\infty^2}}, \frac{1}{256 n \sqrt{m_i}}\right\}}$$

$$
\begin{aligned}
&\leq \sqrt{\frac{4 + \sum_{s=1}^{t-1} \|\widetilde{u}_{i,a}^{(s)} - \widetilde{u}_{i,a}^{(s-1)}\|_\infty^2}{m_i \log T/8}} - \sqrt{\frac{4 + \sum_{s=1}^{t-2} \|\widetilde{u}_{i,a}^{(s)} - \widetilde{u}_{i,a}^{(s-1)}\|_\infty^2}{m_i \log T/8}} \\
&\leq \sqrt{\frac{8}{m_i \log T}} \sqrt{\|\widetilde{u}_i^{(t-1)} - \widetilde{u}_i^{(t-2)}\|_\infty^2} \\
&\leq \frac{\sqrt{8}(\|\widetilde{u}_{i,a}^{(t-1)}\|_\infty + \|\widetilde{u}_{i,a}^{(t-2)}\|_\infty)}{\sqrt{m_i \log T}} \leq \frac{\sqrt{8}(\widehat{x}_i^{(t-1)}(a) + \widehat{x}_i^{(t-2)}(a))}{\sqrt{m_i \log T}} ,
\end{aligned}
$$

where in the first inequality we used the fact that $z \mapsto 1/\min\{a, z\} - 1/\min\{b, z\}$ is nondecreasing in z when $a \leq b$, in the second inequality we used the subadditivity of $z \mapsto \sqrt{z}$ for $z \geq 0$, and in the last inequality we used $\widetilde{u}_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a)\widetilde{u}_i^{(t)}$ and $\|\widetilde{u}_i^{(t)}\|_\infty \leq 1$. ■

From Lemmas 22, 28 and 29, we can prove the following key lemma, which guarantees the stability of the Markov chain under the adaptive learning rate $\eta_{i,a}^{(t)}$ in (6). In this lemma and its proof, we ignore the player index $i \in [n]$ for notational simplicity; for example, we abbreviate $\widehat{x}_i^{(t)}$ as $\widehat{x}^{(t)}$, $y_{i,a}^{(t)}$ as $y_a^{(t)}$, $u_{i,a}^{(t)}$ as $u_a^{(t)}$, $u_i^{(t)}$ as $u^{(t)}$, and $\eta_{i,a}^{(t)}$ as $\eta_a^{(t)}$, where we recall that we use a for the index for actions and i for the index of players.

Lemma 30 *Suppose that $T \geq 3$ and consider the following OFTRL update in (6):*

$$
y_a^{(t)} = \operatorname*{arg\,max}_{y \in \Delta_m} \left\{ -F_a^{(t)}(y) \right\} , \quad F_a^{(t)}(y) \coloneqq - \left(\left\langle y, \widetilde{u}_a^{(t-1)} + \sum_{s=1}^{t-1} \widetilde{u}_a^{(s)} \right\rangle - \frac{1}{\eta_a^{(t)}} \phi(y) \right) .
$$

Then,

$$
\sum_{a \in \mathcal{A}} \|y_a^{(t+1)} - y_a^{(t)}\|_{y_a^{(t)}, F_a^{(t+1)}} = \frac{1}{\sqrt{\eta_a^{(t+1)}}} \sum_{a \in \mathcal{A}} \|y_a^{(t+1)} - y_a^{(t)}\|_{y_a^{(t)}, \phi} \leq \frac{1}{2} . \tag{28}
$$

To ensure the sufficient condition for the stability of the stationary distribution, $\sum_{a \in \mathcal{A}} \mu_a^{(t)} \leq 1/2$ in Lemma 32, under our learning rate in (6), we will see in Lemma 31 that it suffcies to show $\sum_{a \in \mathcal{A}} \|y_a^{(t)} - y_a^{(t-1)}\|_{y_a^{(t-1)}, F_a^{(t)}} \leq 1/2$, which is the claim of Lemma 30. To prove Lemma 30, we will analyze the stability of the outputs $y_1^{(t)}, \ldots, y_m^{(t)}$ of the $m(= |\mathcal{A}|)$ experts simultaneously, which is used for swap regret minimization. Without this new analysis, we need to choose a rather smaller learning rate to ensure the stability of the Markov chain, resulting in a swap regret bound of $O(nm^8 \log T)$ in the honest regime.

Proof [Proof of Lemma 30] We recall that $m = |\mathcal{A}|$ and use $\mathcal{M}_m = (\Delta_m)^m = \Delta_m \times \cdots \times \Delta_m$ to denote the Cartesian product of m probability simplices.[2] Define a strictly convex function $G^{(t+1)} \colon \mathcal{M}_m \to \mathbb{R}$ by

$$
G^{(t+1)}(\boldsymbol{w}) = G^{(t+1)}(w_1, \ldots, w_m) = \sum_{a \in \mathcal{A}} F_a^{(t+1)}(w_a) .
$$

Note that for any $\boldsymbol{h} = (h_1, \ldots, h_m) \in \mathcal{M}_m$, the local norm $\|\boldsymbol{h}\|_{\boldsymbol{y}^{(t)}, G^{(t+1)}}$ is given by

$$
\|\boldsymbol{h}\|_{\boldsymbol{y}^{(t)}, G^{(t+1)}} = \sqrt{\boldsymbol{h}^\top \operatorname{diag}\left(\{\nabla^2 F_a^{(t+1)}(y_a^{(t)})\}_{a \in \mathcal{A}} \right) \boldsymbol{h}}
$$

2. We use $\mathcal{M}_m$ to denote the Cartesian product of m probability simplices since this is equivalent to the set of all row stochastic matrices.

$$= \sqrt{\sum_{a \in \mathcal{A}} h_a^\top \nabla^2 F_a^{(t+1)}(y_a^{(t)}) h_a} = \sqrt{\sum_{a \in \mathcal{A}} \|h_a\|^2_{y_a^{(t)}, F_a^{(t+1)}}} \,. \tag{29}$$

Now from the fact that $y_a^{(t+1)}$ is the minimizer of the strongly convex function $F_a^{(t+1)}$ for each $a \in \mathcal{A}$, the point $\boldsymbol{y}^{(t)} := (y_1^{(t)}, \dots, y_m^{(t)}) \in \mathcal{M}_m$ is the unique minimizer of $G^{(t+1)}$. Hence from Lemma 28, to prove the claim of the lemma, it suffices to prove that for any $\boldsymbol{z} = (z_1, \dots, z_m) \in \mathcal{M}_m$ satisfying $\|\boldsymbol{z} - \boldsymbol{y}^{(t)}\|_{\boldsymbol{y}^{(t)}, G^{(t+1)}} = 1/(2\sqrt{m})$, it holds that $G^{(t+1)}(\boldsymbol{z}) \geq G^{(t+1)}(\boldsymbol{y}^{(t)})$. In fact, if this is proven, then Lemma 28 implies $\|\boldsymbol{y}^{(t+1)} - \boldsymbol{y}^{(t)}\|_{\boldsymbol{y}^{(t)}, G^{(t+1)}} \leq 1/(2\sqrt{m})$, and thus the LHS of (28) is upper bounded by

$$\sum_{a \in \mathcal{A}} \|y_a^{(t+1)} - y_a^{(t)}\|_{y_a^{(t)}, F_a^{(t+1)}} \leq \sqrt{m \sum_{a \in \mathcal{A}} \|y_a^{(t+1)} - y_a^{(t)}\|^2_{y_a^{(t)}, F_a^{(t+1)}}} = \sqrt{m} \|\boldsymbol{y}^{(t+1)} - \boldsymbol{y}^{(t)}\|_{y^{(t)}, G^{(t+1)}} \leq \frac{1}{2} \,,$$

where the first inequality follows from the Cauchy–Schwarz inequality and the equality follows from (29). This is the claim of Lemma 30.

In the following, we will prove that for any $\boldsymbol{z} = (z_1, \dots, z_m) \in \mathcal{M}_m$ satisfying $\|\boldsymbol{z} - \boldsymbol{y}^{(t)}\|_{\boldsymbol{y}^{(t)}, G^{(t+1)}} = 1/(2\sqrt{m})$, it holds that $G^{(t+1)}(\boldsymbol{z}) \geq G^{(t+1)}(\boldsymbol{y}^{(t)})$. Let $h_a = z_a - y_a^{(t)} \in \mathbb{R}^m$ for each $a \in \mathcal{A}$. Let us fix $a \in \mathcal{A}$ and we will lower bound $F_a^{(t+1)}(z_a)$. From Taylor's theorem, there exists a point $\xi_a^{(t)} = \gamma z_a + (1-\gamma) y_a^{(t)}$ for some $\gamma \in [0, 1]$ such that

$$F_a^{(t+1)}(z_a) = F_a^{(t+1)}(y_a^{(t)}) + \langle \nabla F_a^{(t+1)}(y_a^{(t)}), h_a \rangle + \frac{1}{2} h_a^\top \nabla^2 F_a^{(t+1)}(\xi_a^{(t)}) h_a \,. \tag{30}$$

We will lower bound the second term in the RHS of (30) below. From the first-order optimality condition at $y_a^{(t)}$, we have

$$\begin{aligned}
&\nabla F_a^{(t+1)}(y_a^{(t)}) = -\widetilde{u}_a^{(t)} - \sum_{s=1}^{t} \widetilde{u}_a^{(s)} + \frac{1}{\eta_a^{(t+1)}} \nabla \phi(y_a^{(t)}) \\
&= \left(-\widetilde{u}_a^{(t-1)} - \sum_{s=1}^{t-1} \widetilde{u}_a^{(s)} + \frac{1}{\eta_a^{(t)}} \nabla \phi(y_a^{(t)}) \right) - 2\widetilde{u}_a^{(t)} + \widetilde{u}_a^{(t-1)} + \left(\frac{1}{\eta_a^{(t+1)}} - \frac{1}{\eta_a^{(t)}} \right) \nabla \phi(y_a^{(t)}) \\
&= -2\widetilde{u}_a^{(t)} + \widetilde{u}_a^{(t-1)} + \left(\frac{1}{\eta_a^{(t+1)}} - \frac{1}{\eta_a^{(t)}} \right) \nabla \phi(y_a^{(t)}) \,,
\end{aligned} \tag{31}$$

where the last equality follows from $\nabla F_a^{(t)}(y_a^{(t)}) = -\widetilde{u}_a^{(t-1)} - \sum_{s=1}^{t-1} \widetilde{u}_a^{(s)} + (1/\eta_a^{(t)}) \nabla \phi(y_a^{(t)}) = 0$. Hence, the second term in the RHS of (30) is lower bounded as

$$\begin{aligned}
\langle \nabla F_a^{(t+1)}(y_a^{(t)}), h_a \rangle &\geq - \left\| -2\widetilde{u}_a^{(t)} + \widetilde{u}_a^{(t-1)} + \left(\frac{1}{\eta_a^{(t+1)}} - \frac{1}{\eta_a^{(t)}} \right) \nabla \phi(y_a^{(t)}) \right\|_{*, y_a^{(t)}, \phi} \|h_a\|_{y_a^{(t)}, \phi} \\
&\qquad \text{(by (31) and Hölder)} \\
&\geq - \left(\|-2\widetilde{u}_a^{(t)} + \widetilde{u}_a^{(t-1)}\|_{*, y_a^{(t)}, \phi} + \left(\frac{1}{\eta_a^{(t+1)}} - \frac{1}{\eta_a^{(t)}} \right) \|\nabla \phi(y_a^{(t)})\|_{*, y_a^{(t)}, \phi} \right) \cdot \frac{1}{2} \sqrt{\frac{\eta_a^{(t+1)}}{m}}
\end{aligned}$$

$$\geq -\frac{1}{2}\left(\|-2\widetilde{u}_a^{(t)}+\widetilde{u}_a^{(t-1)}\|_\infty+\frac{\sqrt{8}(\widehat{x}^{(t-1)}(a)+\widehat{x}^{(t-2)}(a))}{\sqrt{m\log T}}\|\nabla\phi(y_a^{(t)})\|_{*,y_a^{(t)},\phi}\right)\sqrt{\frac{\eta_a^{(t+1)}}{m}}$$

$$\text{(} \|w\|_{*,y_a^{(t-1)},\phi}\leq\|w\|_\infty \text{ for any } w\in\mathbb{R}^m \text{ and Lemma 29)}$$

$$\geq -\frac{1}{2}\left(2\widehat{x}^{(t)}(a)+\widehat{x}^{(t-1)}(a)+\frac{\sqrt{8}(\widehat{x}^{(t-1)}(a)+\widehat{x}^{(t-2)}(a))}{\sqrt{\log T}}\right)\sqrt{\frac{\eta_a^{(t+1)}}{m}}\,. \tag{32}$$

Here, the second inequality follows from the triangle inequality and

$$\|h_a\|_{y_a^{(t)},\phi}=\|z_a-y_a^{(t)}\|_{y_a^{(t)},\phi}=\sqrt{\eta_a^{(t+1)}}\|z_a-y_a^{(t)}\|_{y_a^{(t)},F_a^{(t+1)}}$$
$$\leq\sqrt{\eta_a^{(t+1)}}\|\boldsymbol{z}-\boldsymbol{y}^{(t)}\|_{\boldsymbol{y}_a^{(t)},G^{(t+1)}}=\frac{1}{2}\sqrt{\frac{\eta_a^{(t+1)}}{m}}\,,$$

and the fourth inequality from $\widetilde{u}_a^{(t)}=\widehat{x}^{(t)}(a)\widetilde{u}^{(t)}$, $\|\widetilde{u}^{(t)}\|_\infty\leq 1$ and $\|\nabla\phi(y_a^{(t)})\|_{*,y_a^{(t)},\phi}\leq\sqrt{m}$, which holds since ϕ is m-self-concordant barrier and Lemma 22. Combining (30) with (32) gives

$$F_a^{(t+1)}(z_a)$$
$$\geq F_a^{(t+1)}(y_a^{(t)})-\frac{1}{2}\left(2\widehat{x}^{(t)}(a)+\widehat{x}^{(t-1)}(a)+\frac{\sqrt{8}\big(\widehat{x}^{(t-1)}(a)+\widehat{x}^{(t-2)}(a)\big)}{\sqrt{\log T}}\right)\sqrt{\frac{\eta_a^{(t+1)}}{m}}+\frac{1}{2}\|h_a\|^2_{\xi_a^{(t)},F_a^{(t+1)}}\,. \tag{33}$$

We next consider the last term in the RHS of (33). We have $\|z_a-y_a^{(t)}\|_{y_a^{(t)},\phi}=\sqrt{\eta_a^{(t+1)}}\|z_a-y_a^{(t)}\|_{y_a^{(t)},F_a^{(t+1)}}\leq\sqrt{\eta_a^{(t+1)}}\|\boldsymbol{z}-\boldsymbol{y}^{(t)}\|_{\boldsymbol{y}^{(t)},G^{(t+1)}}=\frac{1}{2}\sqrt{\eta_a^{(t+1)}/m}\leq\frac{1}{32}m^{-3/4}$, where the last inequality follows from $\eta_a^{(t+1)}\leq 1/(256n\sqrt{m})$. Combining this with a property of self-concordant barriers in Lemma 20, we have

$$\|h_a\|^2_{\xi_a^{(t)},\phi}\geq\left(1-\|y_a^{(t)}-\xi_a^{(t)}\|_{y_a^{(t)},\phi}\right)^2\|h_a\|^2_{y_a^{(t)},\phi}=\left(1-\gamma\|z_a-y_a^{(t)}\|_{y_a^{(t)},\phi}\right)^2\|h_a\|^2_{y_a^{(t)},\phi}\geq\frac{15}{16}\|h_a\|^2_{y_a^{(t)},\phi}\,. \tag{34}$$

Using this inequality, we can lower bound the last term in the RHS of (33) as

$$\frac{1}{2}\sum_{a\in\mathcal{A}}\|h_a\|^2_{\xi_a^{(t)},F_a^{(t+1)}}=\frac{1}{2\eta_a^{(t+1)}}\sum_{a\in\mathcal{A}}\|h_a\|^2_{\xi_a^{(t)},\phi}$$
$$\geq\frac{15}{32}\frac{1}{\eta_a^{(t+1)}}\sum_{a\in\mathcal{A}}\|h_a\|^2_{y_a^{(t)},\phi}=\frac{15}{32}\sum_{a\in\mathcal{A}}\|h_a\|^2_{y_a^{(t)},F_a^{(t+1)}}=\frac{15}{32}\|\boldsymbol{h}\|^2_{\boldsymbol{y}^{(t)},G^{(t+1)}}=\frac{15}{128m}\,, \tag{35}$$

where the first inequality follows from (34), the third equality from (29), and the last equality from $\|\boldsymbol{h}\|_{\boldsymbol{y}^{(t)},G^{(t+1)}}=1/(2\sqrt{m})$.

Therefore, summing up the inequality (33) over $a\in\mathcal{A}$ and using (35), we obtain

$$G^{(t+1)}(\boldsymbol{z})=\sum_{a\in\mathcal{A}}F_a^{(t+1)}(z_a)\geq\sum_{a\in\mathcal{A}}F_a^{(t+1)}(y_a^{(t)})-\frac{1}{2}\left(\frac{3}{\sqrt{m}}+\sqrt{\frac{8}{\log T}}\right)\sqrt{\eta_a^{(t+1)}}+\frac{15}{128}$$

$$\geq \sum_{a\in\mathcal{A}} F_a^{(t+1)}(y_a^{(t)}) = G^{(t+1)}(\boldsymbol{y}^{(t)}) \,,$$

where in the first inequality we used the fact that $\widehat{x}^{(t)}, \widehat{x}^{(t-1)}, \widehat{x}^{(t-2)} \in \Delta_m$ are elements in the probability simplex and in the last inequality we used $T \geq 3$ and $\eta_a^{(t+1)} \leq 1/256$. This completes the proof. ■

Finally we will use the key lemma, Lemma 30, to prove Lemma 33, which relates the stability of the outputs of m_i-experts $\{y_{i,a}^{(t)}\}_{a\in\mathcal{A}_i}$ and the stability of the output $\widehat{x}_i^{(t)}$ of the Markov chain defined by the transition matrix $Q_i^{(t)}$. To prove this relation, we define

$$\mu_a^{(t)} = \max_{b\in\mathcal{A}} \left| 1 - \frac{y_a^{(t)}(b)}{y_a^{(t-1)}(b)} \right| ,$$

in which the player index $i \in [n]$ is again ignored for simplicity and we will also ignore the index i for lemmas involving $\mu_a^{(t)}$. Note that $\mu_a^{(t)}$ is a lower bound of $\|y_a^{(t)} - y_a^{(t-1)}\|_{y_a^{(t-1)},\phi}$, which appears in the LHS of (28) in Lemma 30, since

$$\mu_a^{(t)} = \max_{b\in\mathcal{A}} \left| 1 - \frac{y_a^{(t)}(b)}{y_a^{(t-1)}(b)} \right| \leq \sqrt{\sum_{b\in\mathcal{A}} \left(1 - \frac{y_a^{(t)}(b)}{y_a^{(t-1)}(b)} \right)^2} = \|y_a^{(t)} - y_a^{(t-1)}\|_{y_a^{(t-1)},\phi} \,. \tag{36}$$

The following lemma is a direct consequence of Lemma 30.

Lemma 31 *Under the same assumptions as Lemma 30,* $\sum_{a\in\mathcal{A}} \mu_a^{(t)} \leq \sum_{a\in\mathcal{A}} \|y_a^{(t)} - y_a^{(t-1)}\|_{y_a^{(t-1)},\phi} \leq 1/2$.

Thanks to this lemma, we can use the RVU bound for the log-barrier regularizer in Lemma 24 when proving Theorem 13.

Proof The claim directly follows from Lemma 30. In fact, from (36) and Lemma 30, we have

$$\sum_{a\in\mathcal{A}} \mu_a^{(t)} \leq \sum_{a\in\mathcal{A}} \|y_a^{(t)} - y_a^{(t-1)}\|_{y_a^{(t-1)},\phi} \leq \frac{\sqrt{\eta_a^{(t)}}}{2} \leq \frac{1}{32\sqrt{2}n^{1/2}m^{1/4}} \leq \frac{1}{2} \,,$$

where we used $\eta_a^{(t)} \leq 1/(256 n\sqrt{m})$. ■

The following lemma, which is proven in Anagnostides et al. (2022b) based on the Markov chain tree theorem, relates the stability of $\widehat{x}^{(t)}$ and $\mu_a^{(t)}$.

Lemma 32 (Anagnostides et al. 2022b, Eq. (26) in the proof of Lemma 4.2) *Suppose that* $\sum_{a\in\mathcal{A}} \mu_a^{(t)} \leq 1/2$. *Then,*

$$\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1 \leq 8 \sum_{a\in\mathcal{A}} \mu_a^{(t)} \,.$$

Finally from Lemmas 31 and 32, we can prove the following lemma relating the stability of the output of m_i-experts and the stability of the Markov chain defined by $Q_i^{(t)}$.

Lemma 33 *We assume the conditions of Lemma 30. Then, it holds that*

$$\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 \le 64|\mathcal{A}| \sum_{a\in\mathcal{A}} \|y_a^{(t)} - y_a^{(t-1)}\|^2_{y_a^{(t-1)},\phi} .$$

This lemma will be used in the proof of Theorem 13 to upper bound the negative term in the RVU bound in (13).

Proof From Lemma 32 combined with $\sum_{a\in\mathcal{A}} \mu_a^{(t)} \le 1/2$ in Lemma 31, we have $\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1 \le 8\sum_{a\in\mathcal{A}} \mu_a^{(t)}$, which implies

$$\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 \le 64 \left(\sum_{a\in\mathcal{A}} \mu_a^{(t)}\right)^2 \le 64|\mathcal{A}| \sum_{a\in\mathcal{A}} \left(\mu_a^{(t)}\right)^2 \le 64|\mathcal{A}| \sum_{a\in\mathcal{A}} \|y_a^{(t)} - y_a^{(t-1)}\|^2_{y_a^{(t-1)},\phi} ,$$

where the second inequality follows from the Cauchy–Schwarz inequality and the last inequality from (36). This completes the proof. ■

E.3. Upper bound on $\mathsf{SwapReg}^T_{\widehat{x}_i,\widetilde{u}_i}$

Now from the preliminary lemmas provided in Appendix E.2, we are ready to prove the following lemma, which will lead to Theorem 13.

Lemma 34 *Algorithm 1 achieves*

$$\mathsf{SwapReg}^T_{\widehat{x}_i,\widetilde{u}_i} \le 512 n m_i^{5/2} \log T + 16 m_i \sqrt{\left(2n\sum_{j\ne i} P_1^T(\widehat{x}_j) + 4n \sum_{j\ne i} \widehat{C}_j + \widetilde{C}_i\right)\log T} - \frac{2n}{\sqrt{m_i}} P_1^T(\widehat{x}_i) ,$$

$$\mathsf{SwapReg}^T_{\widehat{x}_i,\widetilde{u}_i} \le 512 m_i^{5/2} \log T + 64 m_i \sqrt{T \log T} .$$

Proof [Proof of Lemma 34] From the definition of $\{y_{i,a}^{(t)}\}_{t=1}^T$ and Lemma 24 with $\sum_{a\in\mathcal{A}} \|y_{i,a}^{(t+1)} - y_{i,a}^{(t)}\|_{y_{i,a}^{(t)},\phi} \le 1/2$ in Lemma 31, for each $a \in \mathcal{A}_i$ we have

$$\begin{aligned}
\widetilde{\mathsf{Reg}}^T_{i,a} &= \max_{y\in\Delta(\mathcal{A}_i)} \sum_{t=1}^T \langle y - y_{i,a}^{(t)}, \widetilde{u}_{i,a}^{(t)}\rangle \\
&\le \frac{m_i \log T}{\eta_{i,a}^{(T+1)}} + 4\sum_{t=1}^T \eta_{i,a}^{(t)} \|\widetilde{u}_{i,a}^{(t)} - \widetilde{u}_{i,a}^{(t-1)}\|^2_{*,y_{i,a}^{(t)},\phi} - \sum_{t=1}^T \frac{1}{16\eta_{i,a}^{(t)}} \|y_{i,a}^{(t+1)} - y_{i,a}^{(t)}\|^2_{y_{i,a}^{(t)},\phi} + 6 \\
&\le \frac{m_i \log T}{\eta_{i,\max}} + \sqrt{32 m_i \left(4 + \sum_{t=1}^T \|\widetilde{u}_{i,a}^{(t)} - \widetilde{u}_{i,a}^{(t-1)}\|^2_{*,y_{i,a}^{(t)},\phi}\right)\log T} - \sum_{t=1}^T \frac{1}{16\eta_{i,a}^{(t)}} \|y_{i,a}^{(t+1)} - y_{i,a}^{(t)}\|^2_{y_{i,a}^{(t)},\phi} + 6 ,
\end{aligned} \tag{37}$$

where we used the inequality $\sum_{t=1}^T z_t / \sqrt{4 + \sum_{s=1}^{t-1} z_s} \le 2\sqrt{\sum_{t=1}^T z_t}$ for $z_t \in [0,4]$ for all $t \in [T]$. Hence, using Lemma 12 and (37), we have

$$\mathsf{SwapReg}^T_{\widehat{x}_i,\widetilde{u}_i} = \sum_{a\in\mathcal{A}_i} \widetilde{\mathsf{Reg}}^T_{i,a}$$

$$\le \frac{m_i^2 \log T}{\eta_{i,\max}} + \sum_{a \in \mathcal{A}_i} \sqrt{32 m_i \left(4 + \sum_{t=1}^T \| \widetilde{u}_{i,a}^{(t)} - \widetilde{u}_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2\right) \log T}$$

$$- \sum_{t=1}^T \sum_{a \in \mathcal{A}_i} \frac{1}{16 \eta_{i,a}^{(t)}} \| y_{i,a}^{(t+1)} - y_{i,a}^{(t)} \|_{y_{i,a}^{(t)},\phi}^2 + 6 m_i \,. \tag{38}$$

Now we have $\| \widetilde{u}_{i,a}^{(t)} - \widetilde{u}_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 \le 2 \| u_{i,a}^{(t)} - u_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 + 4 \widehat{x}_i^{(t)}(a) \| \widetilde{c}_i^{(t)} \|_{*,y_{i,a}^{(t)},\phi}^2 + 4 \widehat{x}_i^{(t-1)}(a) \| \widetilde{c}_i^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2$ $\le 2 \| u_{i,a}^{(t)} - u_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 + 4 \widehat{x}_i^{(t)}(a) \| \widetilde{c}_i^{(t)} \|_\infty^2 + 4 \widehat{x}_i^{(t-1)}(a) \| \widetilde{c}_i^{(t-1)} \|_\infty^2$, and thus the summation in the second term of (38) is upper bounded by

$$\sum_{a \in \mathcal{A}_i} \| \widetilde{u}_{i,a}^{(t)} - \widetilde{u}_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 \le 2 \sum_{a \in \mathcal{A}_i} \| u_{i,a}^{(t)} - u_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 + 8 \widetilde{C}_i \,. \tag{39}$$

The second term in (39) is further bounded as

$$\begin{aligned}
&\sum_{a \in \mathcal{A}_i} \| u_{i,a}^{(t)} - u_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 = \sum_{a \in \mathcal{A}_i} \| u_i^{(t)} \widehat{x}_i^{(t)}(a) - u_i^{(t-1)} \widehat{x}_i^{(t-1)}(a) \|_{*,y_{i,a}^{(t)},\phi}^2 \\
&\le 2 \sum_{a \in \mathcal{A}_i} \| u_i^{(t)} \widehat{x}_i^{(t)}(a) - u_i^{(t-1)} \widehat{x}_i^{(t)}(a) \|_{*,y_{i,a}^{(t)},\phi}^2 + 2 \sum_{a \in \mathcal{A}_i} \| u_i^{(t-1)} \widehat{x}_i^{(t)}(a) - u_i^{(t-1)} \widehat{x}_i^{(t-1)}(a) \|_{*,y_{i,a}^{(t)},\phi}^2 \\
&= 2 \sum_{a \in \mathcal{A}_i} (\widehat{x}_i^{(t)}(a))^2 \| u_i^{(t)} - u_i^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 + 2 \sum_{a \in \mathcal{A}_i} (\widehat{x}_i^{(t)}(a) - \widehat{x}_i^{(t-1)}(a))^2 \| u_i^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2 \\
&\le 2 \| u_i^{(t)} - u_i^{(t-1)} \|_\infty^2 + 2 \sum_{a \in \mathcal{A}_i} (\widehat{x}_i^{(t)}(a) - \widehat{x}_i^{(t-1)}(a))^2 \\
&= 2 \| u_i^{(t)} - u_i^{(t-1)} \|_\infty^2 + 2 \| \widehat{x}_i^{(t)} - \widehat{x}_i^{(t-1)} \|_2^2 \,.
\end{aligned} \tag{40}$$

Hence, from (39) and (40), the sum of the second and third terms in (38) is upper bounded by

$$\sqrt{32 m_i^2 \sum_{a \in \mathcal{A}_i} \left(4 + \sum_{t=1}^T \| u_{i,a}^{(t)} - u_{i,a}^{(t-1)} \|_{*,y_{i,a}^{(t)},\phi}^2\right) \log T} - \frac{1}{16 \eta_{i,\max}} \sum_{t=1}^T \sum_{a \in \mathcal{A}_i} \| y_{i,a}^{(t+1)} - y_{i,a}^{(t)} \|_{y_{i,a}^{(t)},\phi}^2$$

(Cauchy–Schwarz and $\eta_{i,a}^{(t)} \le \eta_{i,\max}$)

$$\le \sqrt{128 m_i^2 \left(m_i + P_\infty^T(u_i) + P_2^T(\widehat{x}_i) + 4 \widetilde{C}_i\right) \log T} - \frac{1}{64 m_i \eta_{i,\max}} P_1^T(\widehat{x}_i)$$

(Lemma 33, (39), and (40))

$$\le 8\sqrt{2} m_i^{3/2} \log T + 8\sqrt{2} m_i \sqrt{(P_\infty^T(u_i) + 4 \widetilde{C}_i) \log T} + 8\sqrt{2} m_i \sqrt{P_1^T(\widehat{x}_i) \log T} - \frac{1}{64 m_i \eta_{i,\max}} P_1^T(\widehat{x}_i)$$

(subadditivity of $z \mapsto \sqrt{z}$ and $\|\cdot\|_2 \le \|\cdot\|_1$)

$$\le 8\sqrt{2} m_i^{3/2} \log T + 8\sqrt{2} m_i \sqrt{(P_\infty^T(u_i) + 4 \widetilde{C}_i) \log T} + 8192 m_i^3 \eta_{i,\max} \log T - \frac{1}{128 m_i \eta_{i,\max}} P_1^T(\widehat{x}_i) \,,$$

where the last line follows from the inequality $b\sqrt{z} - az \le b^2/(4a)$ for any $a > 0$, $b \ge 0$ and $z \ge 0$.

Therefore, combining (38) with the last inequality, we obtain

$$\mathsf{SwapReg}_{\widehat{x}_i, \widetilde{u}_i}^T \le \frac{m_i^2 \log T}{\eta_{i,\max}} + 8\sqrt{2} m_i^{3/2} \log T + 8\sqrt{2} m_i \sqrt{(P_\infty^T(u_i) + 4 \widetilde{C}_i) \log T}$$

$$+ 8192 m_i^2 m \eta_{i,\max} \log T - \frac{1}{128 m_i \eta_{i,\max}} P_1^T(\widehat{x}_i) + 6 m_i \tag{41}$$

$$\le 512 m_i^{5/2} \log T + 64 m_i \sqrt{T \log T}\,, \tag{42}$$

where the last inequality follows from $\eta_{i,\max} \le 1/(256 n \sqrt{m_i})$. The last inequality (42) is the second upper bound on $\mathsf{SwapReg}^T_{\widehat{x}_i, \widetilde{u}_i}$ in Lemma 34.

Now we will upper bound $P_\infty^T(u_i) = \sum_{t=1}^T \|u_i^{(t)} - u_i^{(t-1)}\|_\infty^2$ in (41) to prove the first upper bound on $\mathsf{SwapReg}^T_{\widehat{x}_i, \widetilde{u}_i}$ in Lemma 34. Let $\mathcal{A}_{-i} = \times_{j \neq i} \mathcal{A}_j$. Then,

$$\begin{aligned}
\|u_i^{(t)} - u_i^{(t-1)}\|_\infty &= \max_{a_i \in \mathcal{A}_i} \left| \sum_{a_{-i} \in \mathcal{A}_{-i}} u_i(a_i, a_{-i}) \prod_{j \neq i} x_j^{(t)}(a_j) - \sum_{a_{-i} \in \mathcal{A}_{-i}} u_i(a_i, a_{-i}) \prod_{j \neq i} x_j^{(t-1)}(a_j) \right| \\
&\le \sum_{a_{-i} \in \mathcal{A}_{-i}} \left| \prod_{j \neq i} x_j^{(t)}(a_j) - \prod_{j \neq i} x_j^{(t-1)}(a_j) \right| \le \sum_{j \neq i} \|x_j^{(t)} - x_j^{(t-1)}\|_1\,,
\end{aligned} \tag{43}$$

where the first inequality follows from $u_i^{(t)}(a_i, a_{-i}) \in [-1, 1]$ and the last inequality follows from the fact that the total variation of two product distributions is bounded by the sum of the total variations of each marginal distribution. Now, for any $j \in [n]$, we have

$$\begin{aligned}
\sum_{t=1}^T \|x_j^{(t)} - x_j^{(t-1)}\|_1^2 &\le \sum_{t=1}^T \left(2\|x_j^{(t)} - \widehat{x}_j^{(t)}\|_1^2 + 4\|\widehat{x}_j^{(t)} - \widehat{x}_j^{(t-1)}\|_1^2 + 2\|x_j^{(t-1)} - \widehat{x}_j^{(t-1)}\|_1^2 \right) \\
&\le \sum_{t=1}^T \left(4\|x_j^{(t)} - \widehat{x}_j^{(t)}\|_1^2 + 4\|\widehat{x}_j^{(t)} - \widehat{x}_j^{(t-1)}\|_1^2 \right) \\
&\le 8\widehat{C}_j + \sum_{t=1}^T \|\widehat{x}_j^{(t)} - \widehat{x}_j^{(t-1)}\|_1^2 = 8\widehat{C}_j + P_1^T(\widehat{x}_j)\,,
\end{aligned} \tag{44}$$

where we define $x_j^{(0)} = \widehat{x}_j^{(0)} = (1/m)\mathbf{1}$ for simplicity. Hence, from (43), the Cauchy–Schwarz inequality, and (44), we have

$$\begin{aligned}
P_\infty^T(u_i) &= \sum_{t=1}^T \|u_i^{(t)} - u_i^{(t-1)}\|_\infty^2 \le \sum_{t=1}^T \left(\sum_{j \neq i} \|x_j^{(t)} - x_j^{(t-1)}\|_1 \right)^2 \le (n-1) \sum_{j \neq i} \sum_{t=1}^T \|x_j^{(t)} - x_j^{(t-1)}\|_1^2 \\
&\le 8(n-1) \sum_{j \neq i} \widehat{C}_j + 4(n-1) \sum_{j \neq i} \sum_{t=1}^T \|\widehat{x}_j^{(t)} - \widehat{x}_j^{(t-1)}\|_1^2 = 8(n-1) \sum_{j \neq i} \widehat{C}_j + 4(n-1) \sum_{j \neq i} P_1^T(\widehat{x}_j)\,.
\end{aligned} \tag{45}$$

Finally, combining (41) with (45), we obtain

$$\begin{aligned}
&\mathsf{SwapReg}^T_{\widehat{x}_i, \widetilde{u}_i} \\
&\le \frac{m_i^2 \log T}{\eta_{i,\max}} + 8\sqrt{2} m_i^{3/2} \log T + 8\sqrt{2} m_i \sqrt{\left(4(n-1) \sum_{j \neq i} P_1^T(\widehat{x}_j) + 8(n-1) \sum_{j \neq i} \widehat{C}_j + 4\widetilde{C}_i \right) \log T}
\end{aligned}$$

$$+ 8192 m_i^2 m \eta_{i,\max} \log T - \frac{1}{128 m_i \eta_{i,\max}} P_1^T(\widehat{x}_i) + 6 m_i$$
$$\le 512 n m_i^{5/2} \log T + 16\sqrt{2} m_i \sqrt{\left(n \sum_{j \ne i} P_1^T(\widehat{x}_j) + 2n \sum_{j \ne i} \widehat{C}_j + \widetilde{C}_i \right) \log T} - \frac{2n}{\sqrt{m_i}} P_1^T(\widehat{x}_i) \,,$$

where we used $\eta_{i,\max} \le 1/(256 n \sqrt{m_i})$. This is the desired second upper bound of $\mathsf{SwapReg}^T_{\widehat{x}_i, \widetilde{u}_i}$ in Lemma 34. ∎

E.4. Proof of Theorem 13

Now, from the preliminary lemmas provided in Appendix E.2 and Lemma 34, we are ready to prove Theorem 13.

Proof [Proof of Theorem 13] We first prove the upper bound on $\mathsf{SwapReg}^T_{x_i,u_i}$. From Lemma 34 and Proposition 25, we have

$$\mathsf{SwapReg}^T_{x_i,u_i} \le 512 n m_i^{5/2} \log T + 16\sqrt{2} m_i \sqrt{\left(n \sum_{j \ne i} P_1^T(\widehat{x}_j) + 2n \widehat{S}_{-i} + \widetilde{C}_i \right) \log T} + C_i - \frac{2n}{\sqrt{m_i}} P_1^T(\widehat{x}_i) \,, \tag{46}$$

where we recall that $\widehat{S}_{-i} = \sum_{j \ne i} \widehat{C}_j$ and $C_i = 2\widehat{C}_i + 2\widetilde{C}_i$. Taking the summation of the inequality (46) over the set of players $[n]$ and using the same argument as in (47), we can upper bound the sum of the swap regret, $\sum_{i \in [n]} \mathsf{SwapReg}^T_{x_i,u_i}$, as follows:

$$\begin{aligned}
&\sum_{i \in [n]} \mathsf{SwapReg}^T_{x_i,u_i} - \sum_{i \in [n]} \left(512 n m_i^{5/2} \log T + C_i \right) \\
&\le 16\sqrt{2} \sum_{i \in [n]} m_i \sqrt{\left(n \sum_{j \ne i} P_1^T(\widehat{x}_j) + 2n \widehat{S}_{-i} + \widetilde{C}_i \right) \log T} - \sum_{i \in [n]} \frac{2n}{\sqrt{m_i}} P_1^T(\widehat{x}_i) \\
&\le 16\sqrt{2} \sqrt{\sum_{i \in [n]} m_i^2 \log T} \sqrt{n \sum_{i \in [n]} \sum_{j \ne i} P_1^T(\widehat{x}_j) + 2n^2 \widehat{S}_{-i} + \sum_{i \in [n]} \widetilde{C}_i} - \frac{2n}{\sqrt{m}} \sum_{i \in [n]} P_1^T(\widehat{x}_i) \\
&\hspace{28em} \text{(Cauchy–Schwarz)} \\
&\le 16\sqrt{2} \sqrt{\sum_{i \in [n]} m_i^2 \log T} \sqrt{n^2 \sum_{i \in [n]} P_1^T(\widehat{x}_i) + 2n^2 \widehat{S}_{-i} + \widetilde{S}} - \frac{2n}{\sqrt{m}} \sum_{i \in [n]} P_1^T(\widehat{x}_i) \\
&\le 512 n \sqrt{m} \sum_{i \in [n]} m_i^2 \log T + 16 \sqrt{2 \sum_{i \in [n]} m_i^2 \left(n^2 \widehat{S}_{-i} + 2\widetilde{S} \right) \log T} - \frac{n}{\sqrt{m}} \sum_{i \in [n]} P_1^T(\widehat{x}_i) \,, \qquad (47)
\end{aligned}$$

where we let $\widehat{S}_{-i} = \sum_{j \ne i} \widehat{C}_j$ and recall that $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$, $\widetilde{S} = \sum_{i \in [n]} \widetilde{C}_i$, and $m = \max_{i \in [n]} m_i$. Here, in the last inequality we used the subadditivity of $z \mapsto \sqrt{z}$ for $z \ge 0$ and the inequality $b\sqrt{z} - az \le b^2/(4a)$ for $a > 0$, $b \ge 0$ and $z \ge 0$ with $z = \sum_{i \in [n]} P_1^T(\widehat{x}_i)$. Combining (47) with the fact that $\sum_{i\in[n]} \mathsf{SwapReg}^T_{x_i,u_i} \geq 0$, we obtain

$$n \sum_{i\in[n]} P_1^T(\widehat{x}_i) \leq 768nm \sum_{i\in[n]} m_i^2 \log T + 16\sqrt{\sum_{i\in[n]} m_i^2 \left(n^2 \widehat{S}_{-i} + \widetilde{S}\right)\log T} + \sqrt{m}S\,,$$

where we recall $S = \sum_{i\in[n]} C_i$. Plugging the last inequality in (46), we have

$$\begin{aligned}
&\mathsf{SwapReg}^T_{x_i,u_i} - 512nm_i^{5/2}\log T \\
&\leq m_i\sqrt{\left(768nm\sum_{i\in[n]} m_i^2\log T + 16\sqrt{\sum_{i\in[n]} m_i^2\left(n^2\widehat{S}_{-i}+\widetilde{S}\right)\log T} + \sqrt{m}S + 2n\widehat{S}_{-i} + \widetilde{C}_i\right)\log T + C_i} \\
&\lesssim m_i\sqrt{\left(nm\sum_{i\in[n]} m_i^2\log T + n\widehat{S} + \sqrt{m}S\right)\log T + C_i}\,,
\end{aligned}$$

where in the last line we used the AM–GM inequality and $\widehat{S}_{-i} \leq \widehat{S}$. To simplify the last inequality, from $m_i \leq m$ we get

$$\begin{aligned}
\mathsf{SwapReg}^T_{x_i,u_i} &\lesssim m\sqrt{\left(n^2 m^3 \log T + \left(n+\sqrt{m}\right)\widehat{S} + \sqrt{m}\widetilde{S}\right)\log T + C_i} \\
&\lesssim nm^{5/2}\log T + m\sqrt{\left(\widehat{S}\left(n+\sqrt{m}\right) + \widetilde{S}\sqrt{m}\right)\log T + C_i}\,.
\end{aligned}$$

Finally, from Lemma 34 and Proposition 25, for any opponents we have

$$\mathsf{SwapReg}^T_{x_i,u_i} \lesssim nm_i^{5/2}\log T + m_i\sqrt{T\log T + C_i}\,.$$

Taking the minimum of the last two upper bounds on $\mathsf{SwapReg}^T_{x_i,u_i}$ gives the desired upper bound on $\mathsf{SwapReg}^T_{x_i,u_i}$.

We next prove the upper bound on $\mathsf{SwapReg}^T_{x_i,\widetilde{u}_i}$. Since it holds that $\mathsf{SwapReg}^T_{x_i,\widetilde{u}_i} \geq 0$ by the definition of the swap regret, we can apply a similar argument as in the case of deriving the regret upper bound for $\mathsf{SwapReg}^T_{x_i,u_i}$. From Lemma 34 and Proposition 25, we have

$$\mathsf{SwapReg}^T_{x_i,\widetilde{u}_i} \leq 512nm_i^{5/2}\log T + 16\sqrt{2}m_i\sqrt{\left(n\sum_{j\neq i} P_1^T(\widehat{x}_j) + 2n\widehat{S}_{-i} + \widetilde{C}_i\right)\log T + \widehat{C}_i - \frac{2n}{\sqrt{m_i}} P_1^T(\widehat{x}_i)}\,. \tag{48}$$

Taking the summation of the inequality (48) over the set of players $[n]$, we can upper bound the sum of the swap regret, $\sum_{i\in[n]} \mathsf{SwapReg}^T_{x_i,\widetilde{u}_i}$, by

$$\begin{aligned}
&\sum_{i\in[n]} \mathsf{SwapReg}^T_{x_i,\widetilde{u}_i} - \sum_{i\in[n]}\left(512nm_i^{5/2}\log T + \widehat{C}_i\right) \\
&\leq 512n\sqrt{m}\sum_{i\in[n]} m_i^2\log T + 16\sqrt{2\sum_{i\in[n]} m_i^2\left(n^2\widehat{S}_{-i} + 2\widetilde{S}\right)\log T} - \frac{n}{\sqrt{m}}\sum_{i\in[n]} P_1^T(\widehat{x}_i)\,,
\end{aligned} \tag{49}$$

where we recall that $\widehat{S}_{-i} = \sum_{j \neq i} \widehat{C}_j$, $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$, and $m = \max_{i \in [n]} m_i$. Combining (49) with the fact that $\sum_{i \in [n]} \mathsf{SwapReg}^T_{x_i, \widetilde{u}_i} \geq 0$, we obtain

$$n \sum_{i \in [n]} P_1^T(\widehat{x}_i) \leq 768nm \sum_{i \in [n]} m_i^2 \log T + 16 \sqrt{\sum_{i \in [n]} m_i^2 \left(n^2 \widehat{S}_{-i} + \widetilde{S}\right) \log T} + \sqrt{m} \widehat{S} \,.$$

Plugging the last inequality in (48), we have

$$\begin{aligned}
&\mathsf{SwapReg}^T_{x_i, \widetilde{u}_i} - 512nm_i^{5/2} \log T \\
&\leq 16\sqrt{2} m_i \sqrt{\left(768nm \sum_{i \in [n]} m_i^2 \log T + 16 \sqrt{\sum_{i \in [n]} m_i^2 \left(n^2 \widehat{S}_{-i} + \widetilde{S}\right) \log T} + \sqrt{m}\widehat{S} + 2n\widehat{S}_{-i} + \widetilde{C}_i\right) \log T + \widehat{C}_i} \\
&\lesssim m_i \sqrt{\left(nm \sum_{i \in [n]} m_i^2 \log T + (n + \sqrt{m})\widehat{S} + \widetilde{C}_i\right) \log T} + m_i \left(\sum_{i \in [n]} m_i^2 \widetilde{S} \log T\right)^{1/4} + \widehat{C}_i \,,
\end{aligned}$$

where the last line follows from the AM–GM inequality. To simplify this inequality, from $m_i \leq m$ we get

$$\begin{aligned}
\mathsf{SwapReg}^T_{x_i, \widetilde{u}_i} &\lesssim m \sqrt{\left(n^2 m^3 \log T + (n + \sqrt{m})\widehat{S}\right) \log T} + m \left(nm^2 \widetilde{S} \log T\right)^{1/4} + \widehat{C}_i \\
&\lesssim nm^{5/2} \log T + m \sqrt{\left(\widehat{S}(n + \sqrt{m}) + \widetilde{C}_i\right) \log T} + (\widetilde{S} nm^6 \log T)^{1/4} + \widehat{C}_i \,.
\end{aligned}$$

Finally, from Lemma 34 and Proposition 25, for any opponents we have

$$\mathsf{SwapReg}^T_{x_i, \widetilde{u}_i} \lesssim nm_i^{5/2} \log T + m_i \sqrt{T \log T} + \widehat{C}_i \,.$$

Taking the minimum of the last two upper bounds on $\mathsf{SwapReg}^T_{x_i, \widetilde{u}_i}$ completes the proof. ■

Appendix F. Deferred Proofs of Lower Bounds from Section 6

This section provides the proof of Theorem 15.

F.1. Lower bounds for online linear optimization over the simplex

Here we provide bounds for online linear optimization over the simplex, which will be used in the proof of Theorem 15.

Lemma 35 (Orabona and Pál 2015, Theorem 8) *Consider online linear optimization for $T \geq 7$ rounds over the $(d-1)$-dimensional probability simplex Δ_m for $d \in [2, \exp(T/3)]$. Then, there exists a sequence of loss vectors $\ell^{(1)}, \ldots, \ell^{(T)} \in [-1, 1]^d$ with $\|\ell^{(t)}\|_\infty \leq 1$ and $u \in \Delta_d$ such that the regret of any algorithm that selects $x^{(t)} \in \Delta_d$ at each round $t = 1, \ldots, T$ is lower bounded by*

$$\mathsf{Reg}^T(u) = \sum_{t=1}^T \langle x^{(t)} - u, \ell^{(t)} \rangle \geq 0.09\sqrt{T \log d} - 2\sqrt{T} \,.$$

Lemma 36 *Consider online linear optimization for $T \geq 1$ rounds over the $(d-1)$-dimensional probability simplex Δ_d. Then, there exists a sequence of loss vectors $\ell^{(1)}, \dots, \ell^{(T)} \in [-1,1]^d$ with $\|\ell^{(t)}\|_\infty \leq 1$ and $u \in \Delta_d$ such that the regret of any algorithm that selects $x^{(t)} \in \Delta_d$ at each round $t = 1, \dots, T$ is lower bounded by*

$$\mathsf{Reg}^T(u) = \sum_{t=1}^T \langle x^{(t)} - u, \ell^{(t)} \rangle \geq \sqrt{\frac{T}{2}} \,.$$

This is a very minor variant of the well-known lower bound in online linear optimization (see *e.g.,* Hazan 2016; Orabona 2019). We include the proof for completeness.

Proof Let $\sigma^{(1)}, \dots, \sigma^{(T)}$ be i.i.d. Rademacher random variables, *i.e.,* $\mathbb{P}(\sigma^{(t)} = 1) = \mathbb{P}(\sigma^{(t)} = -1) = 1/2$. Then define $z = e_1 - e_2$ and $\bar{\ell}^{(t)} = \sigma^{(t)} z$ for each $t \in [T]$. Note that this $\bar{\ell}^{(t)}$ satisfies $\|\bar{\ell}^{(t)}\|_\infty = \|z\|_\infty = 1$. Then, we have

$$\begin{aligned}
\sup_{\ell^{(1)},\dots,\ell^{(T)}} \max_{u \in \Delta_d} \mathsf{Reg}^T(u) &\geq \mathbb{E}_{\bar{\ell}^{(1)},\dots,\bar{\ell}^{(T)}} \left[\sum_{t=1}^T \langle x^{(t)}, \bar{\ell}^{(t)} \rangle - \min_{u \in \Delta_d} \sum_{t=1}^T \langle u, \bar{\ell}^{(t)} \rangle \right] \\
&= \mathbb{E}_{\sigma^{(1)},\dots,\sigma^{(T)}} \left[\sum_{t=1}^T \sigma^{(t)} \langle x^{(t)}, z \rangle - \min_{u \in \Delta_d} \sum_{t=1}^T \sigma^{(t)} \langle u, z \rangle \right] \\
&= \mathbb{E}_{\sigma^{(1)},\dots,\sigma^{(T)}} \left[\max_{u \in \Delta_d} \sum_{t=1}^T \sigma^{(t)} \langle u, z \rangle \right] \geq \mathbb{E}_{\sigma^{(1)},\dots,\sigma^{(T)}} \left[\max_{u \in \{e_1, e_2\}} \sum_{t=1}^T \sigma^{(t)} \langle u, z \rangle \right],
\end{aligned}$$

where the last equality follows from the fact that $\sigma^{(t)}$ and $-\sigma^{(t)}$ follow the same distribution. From $\max\{a,b\} \geq (a+b)/2 + |a-b|/2$ for $a, b \in \mathbb{R}$, this is further lower bounded by

$$\begin{aligned}
&\mathbb{E}_{\sigma^{(1)},\dots,\sigma^{(T)}} \left[\frac{1}{2} \sum_{t=1}^T \sigma^{(t)} \langle e_1 + e_2, z \rangle + \frac{1}{2} \left| \sum_{t=1}^T \sigma^{(t)} \langle e_1 - e_2, z \rangle \right| \right] \\
&= \frac{1}{2} \mathbb{E}_{\sigma^{(1)},\dots,\sigma^{(T)}} \left[\left| \sum_{t=1}^T \sigma^{(t)} \langle e_1 - e_2, z \rangle \right| \right] = \mathbb{E}_{\sigma^{(1)},\dots,\sigma^{(T)}} \left[\left| \sum_{t=1}^T \sigma^{(t)} \right| \right] \geq \sqrt{\frac{T}{2}} \,,
\end{aligned}$$

where the first equality follows from $\mathbb{E}[\sigma_t] = 0$, the last equality from $\langle e_1 - e_2, z \rangle = \|e_1 - e_2\|_2^2 = 2$, and the last inequality from the Khintchine's inequality. This completes the proof. ∎

F.2. Proof of Theorem 15

Proof [Proof of Theorem 15 (i)] We will prove $\mathsf{Reg}_{x,\widetilde{g}}^T = \Omega(\sqrt{\widetilde{C}_x \log m_x})$. To simplify the analysis, we focus on the case where $\widetilde{C}_x/2$ is an integer. We will consider a two-player zero-sum game where a payoff matrix is $A = 0$, and for rounds $t = 1, \dots, \widetilde{C}_x/2$, the expected reward vectors are corrupted so that $\sum_{t=1}^{\widetilde{C}_x/2} \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \leq \widetilde{C}_x$, while no corruption occurs beyond this; that is, $g^{(t)} = \widetilde{g}^{(t)}$ for $t = \widetilde{C}_x/2 + 1, \dots, T$, and $x^{(t)} = \widehat{x}^{(t)}$, $y^{(t)} = \widehat{y}^{(t)}$, and $\widetilde{\ell}^{(t)} = \ell^{(t)}$ for $t = 1, \dots, T$. Note that in this case we have $\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \leq \widetilde{C}_x$.

From Lemma 35, there exists a sequence of $\{\widetilde{g}^{(t)}\}_{t=1}^{\widetilde{C}_x/2}$ such that

$$\max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{\widetilde{C}_x/2} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle \geq 0.09\sqrt{(\widetilde{C}_x/2) \log m_x} - 2\sqrt{\widetilde{C}_x/2} \geq 0.06\sqrt{\widetilde{C}_x \log m_x} - \sqrt{2\widetilde{C}_x}\,.$$

For this $\{\widetilde{g}^{(t)}\}_{t=1}^{\widetilde{C}_x/2}$, since $\widetilde{g}^{(t)} = Ay^{(t)} + \widetilde{c}_x^{(t)}$, $A = 0$, and $\widetilde{c}_x^{(t)} = 0$ for all $t \geq \widetilde{C}_x/2 + 1$, we can lower bound $\mathsf{Reg}_{x,\widetilde{g}}^T$ as follows:

$$\begin{aligned}
\mathsf{Reg}_{x,\widetilde{g}}^T &= \max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{T} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle = \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^{\widetilde{C}_x/2} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle + \sum_{t=\widetilde{C}_x/2+1}^{T} \langle x^* - x^{(t)}, Ay^{(t)} \rangle \right\} \\
&= \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^{\widetilde{C}_x/2} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} \geq 0.06\sqrt{\widetilde{C}_x \log m_x} - \sqrt{2\widetilde{C}_x}\,,
\end{aligned}$$

which is the desired lower bound on $\mathsf{Reg}_{x,\widetilde{g}}^T$. The lower bound for $\mathsf{Reg}_{y,\widetilde{\ell}}^T$ can be proven in a similar manner. ■

Proof [Proof of Theorem 15 (ii)] We will prove $\mathsf{Reg}_{x,g}^T = \Omega(\widehat{C}_x)$. To simplify the discussion, we consider the case where $\widehat{C}_x/2$ is an integer. We will consider a two-player zero-sum game in the corrupted regime with the following property. Consider the payoff matrix A such that all elements in the first $m_x - 1$ rows are 1, and all elements in the m_x-th row are 0. The strategies is corrupted so that $x^{(t)} = \widehat{x}^{(t)} + \widehat{c}_x^{(t)} = e_{m_x}$ for each round $t = 1, \ldots, \widehat{C}_x/2$, and no corruption occurs beyond this. Then, the corruption level in the strategies of the x-player is upper bounded by $\sum_{t=1}^{T} \|\widehat{c}_x^{(t)}\|_1 = \sum_{t=1}^{\widehat{C}_x/2} \|e_{m_x} - \widehat{x}^{(t)}\|_1 \leq \widehat{C}_x$. From the construction of the payoff matrix A, for each $t = 1, \ldots, \widehat{C}_x/2$, we have $Ay^{(t)} = \mathbf{1} - e_{m_x}$ and $\langle x^{(t)}, Ay^{(t)} \rangle = 0$ since $A^\top x^{(t)} = 0$.

From this construction of the corrupted game, for any $x^* \in \Delta_{m_x}$, we have

$$\sum_{t=1}^{\widehat{C}_x/2} \langle x^* - x^{(t)}, g^{(t)} \rangle = \sum_{t=1}^{\widehat{C}_x/2} \langle x^*, Ay^{(t)} \rangle = \sum_{t=1}^{\widehat{C}_x/2} \langle x^*, \mathbf{1} - e_{m_x} \rangle = \frac{\widehat{C}_x}{2}(1 - x^*(m_x))\,, \tag{50}$$

where we used $\langle x^{(t)}, Ay^{(t)} \rangle = 0$, $Ay^{(t)} = \mathbf{1} - e_{m_x}$, and $x^* \in \Delta_{m_x}$. Therefore,

$$\begin{aligned}
\mathsf{Reg}_{x,g}^T &= \max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{T} \langle x^* - x^{(t)}, g^{(t)} \rangle \\
&= \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^{\widehat{C}_x/2} \langle x^* - x^{(t)}, g^{(t)} \rangle + \sum_{t=\widehat{C}_x/2+1}^{T} \langle x^* - x^{(t)}, Ay^{(t)} \rangle \right\} \\
&= \max_{x^* \in \Delta_{m_x}} \left\{ \frac{\widehat{C}_x}{2}(1 - x^*(m_x)) + \sum_{t=\widehat{C}_x/2+1}^{T} \langle x^* - x^{(t)}, \mathbf{1} - e_{m_x} \rangle \right\} \qquad \text{(by (50))}
\end{aligned}$$

$$= \max_{x^* \in \Delta_{m_x}} \left\{ \frac{\widehat{C}_x}{2}(1 - x^*(m_x)) + \sum_{t=\widehat{C}_x/2+1}^{T} (x^{(t)}(m_x) - x^*(m_x)) \right\} \geq \frac{\widehat{C}_x}{2} ,$$

where the last inequality follows by choosing x^* with $x^*(m_x) = 0$. This completes the proof of $\mathsf{Reg}^T_{x,g} = \Omega(\widehat{C}_x)$. The lower bound of $\mathsf{Reg}^T_{y,\ell} \geq \widehat{C}_y/2$ can be proven in a similar manner. ■

Proof [Proof of Theorem 15 (iii)] Here, we show that there exists a corrupted game with $\sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1 \leq \widehat{C}_y$ such that $\max\{\mathsf{Reg}^T_{\widehat{x},g}, \mathsf{Reg}^T_{\widehat{y},\ell}\} = \Omega\big(\sqrt{\widehat{C}_y}\big)$. Let $\kappa = 1/4$. Then, it suffices to prove that there exists a two-player zero-sum game with $\sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1 \leq \widehat{C}_y$ such that $\mathsf{Reg}^T_{\widehat{y},\ell} < \kappa\sqrt{\widehat{C}_y}$ implies $\mathsf{Reg}^T_{\widehat{x},g} \geq \kappa\sqrt{\widehat{C}_y}$. To simplify the discussion, we consider only the case where $\widehat{C}_y/2$ is an integer.

We will consider a two-player zero-sum game in the corrupted regime with the following property. Consider the payoff matrix $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$.

The suggested strategies $\widehat{y}^{(t)}$ is corrupted to $y^{(t)}$ for rounds $t = 1, \ldots, \widehat{C}_y/2$, and no corruption occurs beyond this. Then, the corruption level in the strategies of the y-player is upper bounded by $\sum_{t=1}^T \|\widehat{c}_y^{(t)}\|_1 = \sum_{t=1}^{\widehat{C}_y/2} \|\widehat{c}_y^{(t)}\|_1 \leq \widehat{C}_y$. In particular, we take the strategies $\widehat{y}^{(1)}, \ldots, \widehat{y}^{(\widehat{C}_y/2)}$ of the y-player as follows. From Lemma 36, there exists a sequence of reward vectors $\{\breve{g}^{(t)}\}_{t=1}^{\widehat{C}_y/2}$ such that $\|\breve{g}^{(t)}\|_\infty \leq 1$ and

$$\max_{x \in \Delta_{m_x}} \sum_{t=1}^{\widehat{C}_y/2} \langle x - \widehat{x}^{(t)}, \breve{g}^{(t)} \rangle \geq \frac{1}{2}\sqrt{\widehat{C}_y}. \tag{51}$$

Then for this $\{\breve{g}^{(t)}\}_{t=1}^{\widehat{C}_y/2}$, we take $y^{(t)}$ satisfying $\breve{g}^{(t)} = Ay^{(t)}$. This is indeed possible since for any $g \in [-1,1]^2$ there exists $y \in \Delta_{m_y}$ such that $g = Ay = \begin{pmatrix} y(1) - y(3) \\ y(2) - y(3) \end{pmatrix} = \begin{pmatrix} y(1) \\ y(2) \end{pmatrix} - y(3)\,\mathbf{1}$.

Finally, we are ready to prove $\mathsf{Reg}^T_{\widehat{x},g} \geq \kappa\sqrt{\widehat{C}_y}$. Now, from $\mathsf{Reg}^T_{\widehat{y},\ell} < \kappa\sqrt{\widehat{C}_y}$ and the choice of the payoff matrix A, we have

$$\kappa\sqrt{\widehat{C}_y} > \mathsf{Reg}^T_{\widehat{y},\ell} \geq \sum_{t=1}^T (1 - \widehat{y}^{(t)}(3)) \geq \sum_{t=\widehat{C}_y/2+1}^{T} (1 - \widehat{y}^{(t)}(3)) = \sum_{t=\widehat{C}_y/2+1}^{T} (\widehat{y}^{(t)}(1) + \widehat{y}^{(t)}(2)) . \tag{52}$$

We also observe that the instantaneous regret against any $x \in \Delta_{m_x}$ is lower bounded by

$$\langle x - \widehat{x}^{(t)}, Ay^{(t)} \rangle = (x(1) - \widehat{x}^{(t)}(1))y^{(t)}(1) + (x(2) - \widehat{x}^{(t)}(2))y^{(t)}(2) \geq -(y^{(t)}(1) + y^{(t)}(2)) , \tag{53}$$

where we used $Ay^{(t)} = \begin{pmatrix} y^{(t)}(1) \\ y^{(t)}(2) \end{pmatrix} - y^{(t)}(3)\,\mathbf{1}$. By combining (52) and (53), for any $x \in \Delta_{m_x}$ the regret compared to any point x after round $t = \widehat{C}_y/2 + 1$ is lower bounded by

$$\sum_{t=\widehat{C}_y/2+1}^{T} \langle x - \widehat{x}^{(t)}, Ay^{(t)} \rangle \geq - \sum_{t=\widehat{C}_y/2+1}^{T} (y^{(t)}(1) + y^{(t)}(2)) \geq -\kappa\sqrt{\widehat{C}_y} . \tag{54}$$

Therefore, from (51) and (54), we obtain

$$\mathsf{Reg}_{\widehat{x},g}^{T} \geq \frac{1}{2}\sqrt{\widehat{C}_y} - \kappa\sqrt{\widehat{C}_y} = \left(\frac{1}{2} - \kappa\right)\sqrt{\widehat{C}_y} = \kappa\sqrt{\widehat{C}_y}\,,$$

which completes the proof. ■