Tight Regret Upper and Lower Bounds for Optimistic Hedge in Two-Player Zero-Sum Games

Taira Tsuchiya
The University of Tokyo & RIKEN

Abstract

In two-player zero-sum games, the learning dynamic based on optimistic Hedge achieves one of the best-known regret upper bounds among strongly-uncoupled learning dynamics. With an appropriately chosen learning rate, the social and individual regrets can be bounded by $O(\log(mn))$ in terms of the numbers of actions m and n of the two players. This study investigates the optimality of the dependence on m and n in the regret of optimistic Hedge. To this end, we begin by refining existing regret analysis and show that, in the strongly-uncoupled setting where the opponent's number of actions is known, both the social and individual regret bounds can be improved to $O(\sqrt{\log m \log n})$. In this analysis, we express the regret upper bound as an optimization problem with respect to the learning rates and the coefficients of certain negative terms, enabling refined analysis of the leading constants. We then show that the existing social regret bound as well as these new social and individual regret upper bounds cannot be further improved for optimistic Hedge by providing algorithm-dependent individual regret lower bounds. Importantly, these social regret upper and lower bounds match exactly including the constant factor in the leading term. Finally, building on these results, we improve the last-iterate convergence rate and the dynamic regret of a learning dynamic based on optimistic Hedge, and complement these bounds with algorithm-dependent dynamic regret lower bounds that match the improved bounds.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

1 INTRODUCTION

Learning in games is a central problem in both game theory and machine learning, and it is well known that players can learn an equilibrium by employing online learning algorithms (Freund and Schapire, 1999; Hart and Mas-Colell, 2000; Cesa-Bianchi and Lugosi, 2006). Such equilibrium learning has led to the development of AI systems that surpass human performance (Bowling et al., 2015; Moravčík et al., 2017; Perolat et al., 2022; FAIR et al., 2022). Furthermore, its effectiveness has also been demonstrated recently for the alignment of large language models (LLMs) (Munos et al., 2024; Swamy et al., 2024).

The advantage of equilibrium learning based on online learning is that it can be realized through *uncoupled* learning dynamics (Hart and Mas-Colell, 2003). In particular, in two-player zero-sum games, it can be achieved by *strongly-uncoupled* learning dynamics (Daskalakis et al., 2011), namely dynamics in which each player learns solely from the rewards they have observed in the past, without knowing the opponent's strategies, observations, or even the number of actions. Such dynamics, which attain an approximate equilibrium as a consequence of maximizing cumulative reward based only on their own observations, are consistent with realistic behavioral models (Hart and Mas-Colell, 2003), and many recent learning dynamics based on online learning exhibit this property (*e.g.,* Syrgkanis et al., 2015; Anagnostides et al., 2022b).

The most representative online learning algorithm used by each player in learning in games is Hedge (Littlestone and Warmuth, 1994; Freund and Schapire, 1997), also known as multiplicative weights update (MWU) or exponential weights. The Hedge algorithm selects actions using weights exponentially scaled by past cumulative rewards, and guarantees a worst-case (external) regret of $O(\sqrt{T \log m})$, where T is the number of rounds and m is the number of actions.

In learning in games, this worst-case regret upper bound can be significantly improved if each player employs specific online learning algorithms. In particular, in two-player zero-sum games, one of the most powerful algorithms is

Table 1: Regret upper and lower bounds of learning dynamics based on optimistic Hedge in two-player zero-sum games. The lower bounds are algorithm-dependent and correspond to the learning rates used in the upper bounds. We write $M = \log m$ and $N = \log n$. The individual regret upper bounds are for the case where the focus is solely on minimizing the regret of the x-player. The upper bounds when minimizing the maximum of the individual regrets are provided in Theorems 9 and 10. The learning rates corresponding to each regret bound are summarized in Table 2 in Appendix G.

	Upper bound	Lower bound
▷ Strongly-uncoupled learning dynamics		
Social regret	$2(M+N)+1$ (Rakhlin and Sridharan, 2013)	$2(M+N)-o(1)$ (**This work, Theorem 11**)
Individual regret	$M+N+1/2$ (Rakhlin and Sridharan, 2013), Eq. (7)	$M-o(1)$ (**This work, Theorem 11**)
Dynamic regret	$(2(M+N)+1)\log T$ (Cai et al., 2025)	$M\log(T+1)-o(1)$ (**This work, Theorem 15**)
▷ Cardinality-aware strongly-uncoupled learning dynamics		
Social regret	$2\sqrt{M(N+1/2)}+2\sqrt{N(M+1/2)}$ (**This work, Theorem 5**)	$\simeq 4\sqrt{MN}-o(1)$ (**This work, Theorem 11**)
Individual regret	$2\sqrt{M(N+1/2)}$ (**This work, Eq. (6)**)	$\simeq \sqrt{MN}-o(1)$ (**This work, Theorem 11**)
Dynamic regret	$2(\sqrt{M(N+1/2)}+\sqrt{N(M+1/2)})\log T$ (**This work, Theorem 14**)	$\simeq \sqrt{MN}\log T-o(1)$ (**This work, Theorem 15**)

optimistic Hedge (Rakhlin and Sridharan, 2013; Syrgkanis et al., 2015), which selects actions according to weights that are exponentially scaled by the cumulative past rewards plus an additional copy of the most recently observed reward. When the learning rates of optimistic Hedge are set to an absolute constant, the social regret, *i.e.,* the sum of the regrets of all players, can be bounded by $O(\log(mn))$, where m and n denote the numbers of actions of the x- and y-players, respectively. This implies convergence to a Nash equilibrium at the rate of $O(\log(mn)/T)$, which is optimal up to the $\log(mn)$ factor (Daskalakis et al., 2011).

While standard no-regret online learning algorithms such as optimistic Hedge typically guarantee only average-iterate (time-averaged) convergence, very recent work shows that a learning dynamic that outputs the time-averaged strategy of optimistic Hedge can also guarantee anytime last-iterate convergence (Cai et al., 2025). Specifically, this approach achieves $O(\log(mn)/t)$ last-iterate convergence at every round t, which implies that each player's dynamic regret is bounded by $O(\log(mn)\log T)$.

As we have seen, optimistic Hedge is one of the best algorithms for learning dynamics in two-player zero-sum games. However, even with this learning dynamic, there remains a gap of an $O(\log(mn))$ factor compared with the existing lower bound on the convergence rate. This raises the fundamental question: what are the optimal upper bounds for the social and individual regrets when using uncoupled learning dynamics? Despite its fundamental nature, this question has not yet been investigated. Additional related work is provided in Appendix A.

Contributions of this paper As a first step toward addressing this open question, this study investigates the following question: in learning two-player zero-sum games, how optimal are optimistic-Hedge-based learning dynamics and their analysis in terms of dependence on the numbers of actions m and n and on the leading constants? To answer this question, we make the following contributions.

First, in Section 3, we refine the regret analysis of optimistic Hedge so that we can compare it precisely with the lower bounds we derive. The existing upper bounds are somewhat ad hoc: the analysis pays little attention to the magnitude of the leading constants, and although exploiting a certain negative term that appears in the upper bound of optimistic Hedge is crucial, it has not been treated with sufficient care. We therefore conduct a careful analysis to investigate how much we can improve the leading constants of the regret bounds and their dependence on m and n.

In the analysis of optimistic Hedge, it is important not only to tune the learning rate but also to leverage the negative term. We observe that this negative term plays two distinct roles, and introduce a new parameter that captures the tradeoff between them. We then formulate the regret upper bound as an optimization problem, which can be transformed into a convex optimization problem via a change of variables. This formulation elucidates the tradeoff between the x- and y-players' individual regrets, which allows us to make a precise comparison with the individual regret lower bound derived next.

Using this optimization perspective, we show that, in strongly-uncoupled learning dynamics where each player is additionally allowed to know the opponent's number of actions, one can achieve social and individual regret bounds of $O(\sqrt{\log m \log n})$. This improvement is particularly effective in games where $\log m$ and $\log n$ are highly imbalanced. In particular, this occurs when the number of actions of a player is exponentially large: for example, network interdiction, where the set of source–sink paths is exponential (Washburn and Wood, 1995); extensive-form games whose normal-form strategy space is exponential (Koller et al., 1996); zero-sum games with submodular structure (Wilder, 2018); maximum flow (Rakhlin and Sridharan, 2013); and asymmetric combinatorial–continuous zero-sum games (Li et al., 2025). Through numerical experiments, we confirm that being cardinality-aware indeed leads to empirical improvements in both the social

regret and the maximum of the individual regrets. The detailed experimental setup and results are mainly provided in Appendix G. The source code for reproducing figures is available at https://github.com/tsuchhiii/games-lower-bounds.

Next, in Section 4, to investigate the optimality of the refined regret upper bounds, we derive regret lower bounds for the learning dynamic based on optimistic Hedge. We show that when each player uses optimistic Hedge with learning rates $\eta, \eta' > 0$, their individual regrets are lower bounded by $\log(m)/\eta - o(1)$ and $\log(n)/\eta' - o(1)$, respectively, where $o(1)$ denotes a term that vanishes as $T \to \infty$. These regret lower bounds imply that the social regret of optimistic Hedge matches the existing best social regret bounds *including the leading constants*. For the individual regret, the upper and lower bounds match in many cases up to constant factors. To our knowledge, our work is the first to derive regret lower bounds for optimistic Hedge and to investigate their dependence on the numbers of actions m and n.

As the third contribution, in Section 5, we extend the above refinement and analysis of the (external) regret upper and lower bounds to dynamic regret. First, we present an improved result on the last-iterate convergence rate (and the corresponding dynamic regret bound) for learning dynamics based on optimistic Hedge that enjoy the last-iterate convergence property. We then provide an improvement of the dynamic regret itself, together with an algorithm-dependent dynamic regret lower bound that matches these results. From these derived lower bounds, we can see that the approach of Cai et al. (2025) cannot be further improved with respect to m, n, and T. These contributions are summarized in Table 1.

It is worth noting that Anagnostides et al. (2024) prove conditional, dimension-dependent lower bounds for Hedge and optimistic Hedge in multiplayer general-sum games under the Exponential Time Hypothesis for PPAD. By contrast, we prove a direct, unconditional, information-theoretic lower bound for optimistic Hedge. Our proof is based on an explicit game instance and a closed-form analysis of the dynamics, rather than via an indirect complexity-theoretic reduction. In this sense, our lower bound may be viewed as a concrete partial realization of a direction suggested in their paper. A more detailed comparison is given in Appendix A.

2 PRELIMINARIES

This section provides some preliminaries. For $n \in \mathbb{N}$, we denote $[n] = \{1, \dots, n\}$. We use $\mathbf{0}$ and $\mathbf{1}$ to denote the all-zero and all-one vectors, respectively. For a vector x, we write $x(i)$ for its i-th coordinate, and $\|x\|_p$ for its ℓ_p-norm, where $p \in [1, \infty]$.

2.1 Learning in Two-Player Zero-Sum Games

Setup Learning in a two-player zero-sum game is characterized by a payoff matrix $A \in [-1, 1]^{m \times n}$, where m and n denote the number of actions of the x- and y-players, respectively. The procedure of this game is as follows: at each round $t = 1, \dots, T$, the x-player selects a strategy $x_t \in \Delta_m$ and the y-player selects $y_t \in \Delta_n$. Then, the x-player observes an expected gain vector $g_t = Ay_t$ and the y-player observes an expected loss vector $\ell_t = A^\top x_t$. Finally, the x-player gains a payoff of $\langle x_t, g_t \rangle$ and the y-player incurs a loss of $\langle y_t, \ell_t \rangle$.

The goal of each player is to minimize the (external) regret given by $\mathsf{Reg}_T^x = \max_{x^* \in \Delta_m} \mathsf{Reg}_T^x(x^*)$ and $\mathsf{Reg}_T^y = \max_{y^* \in \Delta_n} \mathsf{Reg}_T^y(y^*)$ for $\mathsf{Reg}_T^x(x^*) = \sum_{t=1}^T \langle x^* - x_t, Ay_t \rangle = \sum_{t=1}^T \langle x^* - x_t, g_t \rangle$ and $\mathsf{Reg}_T^y(y^*) = \sum_{t=1}^T \langle y_t - y^*, A^\top x_t \rangle = \sum_{t=1}^T \langle y_t - y^*, \ell_t \rangle$. Note that, in the external regret, one compares against the strategy that maximizes the cumulative gain or the strategy that minimizes the cumulative loss. The social regret is defined as the sum of the regrets of both players, that is, $\mathsf{SocialReg}_T = \mathsf{Reg}_T^x + \mathsf{Reg}_T^y$. In addition, in this paper, we also consider the following notion of dynamic regret, which compares against the best strategy at each round: $\mathsf{DReg}_T^x = \sum_{t=1}^T \max_{x_t^* \in \Delta_m} \langle x_t^* - x_t, g_t \rangle$ and $\mathsf{DReg}_T^y = \sum_{t=1}^T \max_{y_t^* \in \Delta_n} \langle y_t - y_t^*, \ell_t \rangle$.

No-regret learning and Nash equilibrium We say that a pair of probability distributions (x^*, y^*) over action sets $[m]$ and $[n]$ is an *ε-approximate Nash equilibrium* for $\varepsilon \geq 0$ if for any distributions $x \in \Delta_m$ and $y \in \Delta_n$, $\langle x, Ay^* \rangle - \varepsilon \leq \langle x^*, Ay^* \rangle \leq \langle x^*, Ay \rangle + \varepsilon$. The pair (x^*, y^*) is a *Nash equilibrium* if it is a 0-approximate Nash equilibrium.

It is well known that an approximate Nash equilibrium is obtained as a consequence of no-regret learning dynamics:

Theorem 1 (Freund and Schapire, 1999)**.** *Let the sequences of strategies $(x_t)_{t=1}^T$ and $(y_t)_{t=1}^T$ be generated by online learning algorithms with regrets Reg_T^x and Reg_T^y, respectively. Then, the product distribution of the average strategies $(\frac{1}{T}\sum_{t=1}^T x_t, \frac{1}{T}\sum_{t=1}^T y_t)$ is a $(\mathsf{SocialReg}_T/T)$-approximate Nash equilibrium.*

Uncoupled learning dynamics In learning in games, each player often employs some form of decentralized algorithm. The most representative form of this is uncoupled learning dynamics (Hart and Mas-Colell, 2003): a learning dynamic is said to be uncoupled if each player's strategy does not depend on the other players' utility functions (*i.e.,* gain or loss functions). However, in two-player zero-sum games, the x-player's utility $\langle x, Ay \rangle$ and the y-player's utility $-\langle x, Ay \rangle$ differ only in sign, so this condition does not impose any restriction.

Algorithm 1 Optimistic Hedge for the x-player

1: **for** $t = 1, 2, \dots, T$ **do**
2: Choose a strategy $x_t \in \Delta_m$ by the optimistic Hedge algorithm in (1) or (2).
3: Observe a gain vector $g_t = Ay_t \in [-1, 1]^m$.

Daskalakis et al. (2011) introduce the notion of strongly-uncoupled dynamics: a learning dynamic is said to be strongly-uncoupled if, at each round t, the x-player determines its strategy using only $(Ay_s)_{s=1}^{t-1}$, while the y-player determines its strategy using only $(A^\top x_s)_{s=1}^{t-1}$. Note that under strongly-uncoupled learning dynamics, each player has no access to the opponent's strategies or even to the number of actions available to the opponent.

In this study, we also consider the following intermediate learning dynamics. A learning dynamic is said to be *cardinality-aware strongly-uncoupled* if, in addition to the information available under strongly-uncoupled dynamics, each player is informed of the number of actions of the opponent. Note that this definition naturally extends to multiplayer general-sum games. As we will discuss in Section 3, allowing each player to know the opponent's number of actions allows us to improve social and individual regret bounds.

2.2 Optimistic Hedge

In the optimistic Hedge algorithm, the strategies of the x- and y-players are determined as follows:

$$\begin{aligned} x_t(i) &\propto \exp\left(\eta\left(\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\right)\right) =: w_t(i)\,, \\ y_t(i) &\propto \exp\left(-\eta'\left(\sum_{s=1}^{t-1} \ell_s(i) + \ell_{t-1}(i)\right)\right) =: v_t(i)\,, \end{aligned} \tag{1}$$

where $\eta, \eta' > 0$ are learning rates of each player, and we set $w_1 = v_1 = \mathbf{1}$ and let $g_0 = \ell_0 = \mathbf{0}$ for simplicity. Note that $x_1 = \frac{1}{m}\mathbf{1}$ and $y_1 = \frac{1}{n}\mathbf{1}$. For $t \geq 2$, the update rule can be equivalently written as

$$\begin{aligned} x_t(i) &\propto w_{t-1}(i)\exp(\eta(2g_{t-1}(i) - g_{t-2}(i)))\,, \\ y_t(i) &\propto v_{t-1}(i)\exp(-\eta'(2\ell_{t-1}(i) - \ell_{t-2}(i)))\,. \end{aligned} \tag{2}$$

The algorithm for the x-player is summarized in Algorithm 1, and the algorithm for the y-player can be described analogously.

In Section 5, we provide a theoretical analysis of a learning dynamic based on optimistic Hedge that achieves $O(1/t)$ last-iterate convergence and, as a consequence, an $O(\log T)$ dynamic regret upper bound, slightly improving the bounds in Cai et al. (2025). Further details of this learning dynamic are given in Section 5.

3 REGRET UPPER BOUNDS

This section provides a refined regret analysis of optimistic Hedge to enable a precise comparison with the regret lower bounds derived in the next section.

3.1 Common Analysis

We first prepare the following lemma, which generalizes the standard upper bound of optimistic Hedge.

Lemma 2. *Suppose that the x- and y-players use optimistic Hedge (Algorithm 1) with learning rates η and η', respectively. Then, for any $c, c' > 0$,*

$$\begin{aligned} \mathsf{Reg}_T^x &\leq \frac{\log m}{\eta} + \frac{\eta}{2c}\sum_{t=1}^{T}\|g_t - g_{t-1}\|_\infty^2 - \frac{1-c}{2\eta}\sum_{t=2}^{T}\|x_t - x_{t-1}\|_1^2\,, \\ \mathsf{Reg}_T^y &\leq \frac{\log n}{\eta'} + \frac{\eta'}{2c'}\sum_{t=1}^{T}\|\ell_t - \ell_{t-1}\|_\infty^2 - \frac{1-c'}{2\eta'}\sum_{t=2}^{T}\|y_t - y_{t-1}\|_1^2\,. \end{aligned}$$

The proof is provided in Appendix B. Here, the parameters c and c' arise from an appropriate decomposition of negative terms that appear in the regret analysis. Intuitively, the larger these parameters are (within the range up to 1), the smaller the corresponding player's individual regret becomes.

For $\eta, \eta' > 0$ and $c, c' > 0$, define

$$\begin{aligned} &\Omega(\eta, \eta', c, c') = \omega(\eta, c) + \omega'(\eta', c') \\ &\omega(\eta, c) = \frac{\log m}{\eta} + \frac{\eta}{2c}\,,\ \omega'(\eta', c') = \frac{\log n}{\eta'} + \frac{\eta'}{2c'}\,. \end{aligned} \tag{3}$$

Then, by Lemma 2, we can derive the following upper bounds on the social regret and the individual regrets.

Theorem 3. *Under the assumptions of Lemma 2, for any $c, c' > 0$ satisfying $\eta\eta' \leq \min\{c'(1-c), c(1-c')\}$,*

$$\begin{aligned} \mathsf{SocialReg}_T &\leq \Omega(\eta, \eta', c, c')\,, \\ \mathsf{Reg}_T^x &\leq \omega(\eta, c) + \frac{\frac{\eta}{2c}}{\frac{1-c'}{2\eta'} - \frac{\eta}{2c}}\Omega(\eta, \eta', c, c') =: f(\eta, \eta', c, c')\,, \\ \mathsf{Reg}_T^y &\leq \omega'(\eta', c') + \frac{\frac{\eta'}{2c'}}{\frac{1-c}{2\eta} - \frac{\eta'}{2c'}}\Omega(\eta, \eta', c, c') =: g(\eta, \eta', c, c')\,. \end{aligned}$$

For simplicity, we define $f(\eta, \eta', c, c') = \infty$ when $\eta\eta' = c(1-c')$, and $g(\eta, \eta', c, c') = \infty$ when $\eta\eta' = c'(1-c)$.

Theorem 3 is similar to the standard analysis of optimistic Hedge (Rakhlin and Sridharan, 2013; Syrgkanis et al., 2015). Here, we provide only a proof sketch, while the complete proof is deferred to Appendix B.

Proof sketch of Theorem 3. We first note that when $\eta\eta' < c'(1-c)$, we have $\frac{\eta'}{2c'} - \frac{1-c}{2\eta} < 0$ and when $\eta\eta' < c(1-c')$

we have $\frac{\eta}{2c} - \frac{1-c'}{2\eta'} < 0$. Hence, summing the two inequalities in Lemma 2 and using the inequalities $\|g_t - g_{t-1}\|_\infty \le \|y_t - y_{t-1}\|_1$ and $\|\ell_t - \ell_{t-1}\|_\infty \le \|x_t - x_{t-1}\|_1$, we have $\mathsf{SocialReg}_T \le \Omega(\eta, \eta', c, c') + \left(\frac{\eta'}{2c'} - \frac{1-c}{2\eta}\right) \sum_{t=2}^T \|x_t - x_{t-1}\|_1^2 + \left(\frac{\eta}{2c} - \frac{1-c'}{2\eta'}\right) \sum_{t=2}^T \|y_t - y_{t-1}\|_1^2 \le \Omega(\eta, \eta', c, c')$, where the last inequality follows from the assumption that $\eta\eta' \le \min\{c'(1-c), c(1-c')\}$. Combining the last upper bound on $\mathsf{SocialReg}_T$ with the fact that $\mathsf{SocialReg}_T \ge 0$ gives $\sum_{t=2}^T \|y_t - y_{t-1}\|_1^2 \le \frac{1}{\frac{1-c'}{2\eta'} - \frac{\eta}{2c}} \cdot \Omega(\eta, \eta', c, c')$. As in Theorem 3, we define the right-hand side to be ∞ whenever $\frac{1-c'}{2\eta'} - \frac{\eta}{2c} = 0$. Finally, combining Lemma 2 with $\|g_t - g_{t-1}\|_\infty \le \|y_t - y_{t-1}\|_1$ and the last two inequalities, we obtain the desired upper bound on Reg_T^x. The upper bound on Reg_T^y can be proven in a similar manner. □

By Theorem 3, the optimal learning rates $\eta, \eta' > 0$ for the social and individual regrets under this analysis can be determined by solving optimization problems over the following feasible region Λ of the functions Ω, f, and g:

$$\Lambda = \left\{(\eta, \eta', c, c') \in \mathbb{R}^4_{>0} \colon \eta\eta' \le \min\{c'(1-c), c(1-c')\}\right\}$$

For example, within this framework, the optimization problem that determines the optimal parameters (η, η', c, c') for the social regret is given by $\min_{(\eta,\eta',c,c')\in\Lambda} \Omega(\eta, \eta', c, c')$. For notational convenience, we sometimes denote an element of Λ by $\lambda = (\eta, \eta', c, c')$ and write $M = \log m$ and $N = \log n$.

3.2 Social Regret Bounds

The learning rates η, η' that minimize the social regret and those that minimize the individual regrets are not necessarily the same. We first focus on the social regret.

Lemma 4. *Let $M' = M + 1/2$, $N' = N + 1/2$, and $D = \sqrt{MN} + \sqrt{M'N'}$. Then, it holds that $\min_{\lambda\in\Lambda} \Omega(\eta, \eta', c, c') = 2\sqrt{MN'} + 2\sqrt{M'N}$. The minimum is achieved by λ in the boundary of Λ such that*

$$c = c' = \frac{\sqrt{M'N'}}{D}\,,\ \eta = \frac{\sqrt{MM'}}{D}\,,\ \eta' = \frac{\sqrt{NN'}}{D}\,. \quad (4)$$

The proof can be found in Appendix B. Combining Theorem 3 with Lemma 4 yields the following bound:

Theorem 5. *Suppose that the x- and y-players use optimistic Hedge (Algorithm 1) with learning rates η and η' in (4). Then, $\mathsf{SocialReg}_T \le 2\sqrt{\log m\,(\log n + 1/2)} + 2\sqrt{\log n\,(\log m + 1/2)}$.*

An important remark is that this optimization problem focuses solely on the social regret, and with this adjustment it is not possible to upper bound the individual regrets. This is consistent with the fact that the optimal solution of Lemma 4 lies on $\eta\eta' = \min\{c'(1-c), c(1-c')\}$, in which case $f = g = \infty$.

In the setting where each player does not know the opponent's number of actions, that is, under strongly-uncoupled learning dynamics, the analysis based on Theorem 3 shows that the following social regret cannot be further improved. A rigorous argument is deferred to Appendix B.

Theorem 6. *Suppose that the x- and y-players use optimistic Hedge (Algorithm 1) with learning rates $\eta = \eta' = 1/2$. Then, $\mathsf{SocialReg}_T \le 2\log(mn) + 1$.*

Note that Theorem 6 is not a new result although no prior literature has explicitly stated this upper bound with the leading constants, and our cardinality-aware upper bound in Theorem 5 is strictly better than the one in Theorem 6. In fact, by the AM–GM inequality, we have $2\sqrt{\log m\,(\log n + 1/2)} + 2\sqrt{\log n\,(\log m + 1/2)} \le 2\log(mn) + 1$. As suggested by the nature of the AM–GM inequality, this implies that the advantage of being cardinality-aware becomes more significant as $\max\{\log m/\log n, \log n/\log m\}$ increases, which is particularly illustrated in the examples mentioned in Section 1.

Related to prior work, an asymmetric learning-rate tuning akin to our cardinality-aware refinement already appears in the maximum-flow analysis of Rakhlin and Sridharan (2013). Our contribution here is to isolate this phenomenon in two-player zero-sum matrix games, derive explicit social- and individual-regret bounds for optimistic Hedge, and complement them with algorithm-dependent lower bounds and dynamic-regret guarantees (as we will see in the following sections). Appendix F discusses how part of the upper-bound argument extends to Hölder-smooth convex–concave games.

3.3 Individual Regret Bounds

Next, we turn to individual regret. Although $\mathsf{Reg}_T^x \le \min_{\lambda\in\Lambda} f(\lambda)$ and $\mathsf{Reg}_T^y \le \min_{\lambda\in\Lambda} g(\lambda)$ hold, the learning rates attaining the minima on the right-hand sides need not coincide. Thus, we need a criterion for choosing λ.

A natural approach is to minimize $\max\{\mathsf{Reg}_T^x, \mathsf{Reg}_T^y\}$ by solving $\min_{\lambda\in\Lambda} \max\{f(\lambda), g(\lambda)\}$. On the other hand, one may also ask how much one player's regret can be reduced at the expense of the other's, which will be useful when comparing with the lower bounds derived later. To unify these viewpoints, we consider minimizing a weighted sum of f and g: for $\gamma \in [0, 1]$,

$$J_\gamma(\lambda) = \gamma\, f(\lambda) + (1-\gamma)\, g(\lambda)\,. \quad (5)$$

The problem of minimizing J_γ over λ is nonconvex, and unlike the case of social regret, it does not admit a closed-form solution. However, after an appropriate change of variables, it can be reformulated as a convex optimization problem; see Appendix B.4.4 for details.

Proposition 7. *For any $\gamma \in [0, 1]$, the minimization problem $\min_{\lambda\in\Lambda} J_\gamma(\lambda)$ can be transformed into a convex optimization problem by an appropriate change of variables.*

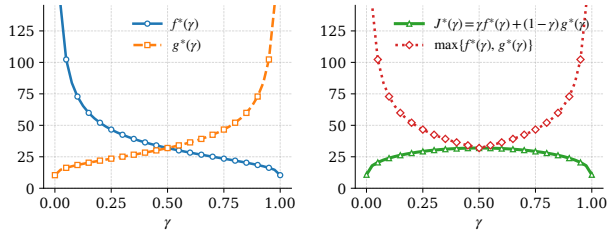

Figure 1: Tradeoff versus $\gamma \in (0,1)$ when $m = n = 10^2$: (a) $f^*(\gamma) = f(\lambda_\gamma)$ and $g^*(\gamma) = g(\lambda_\gamma)$ for $\lambda_\gamma = \arg\min_{\lambda\in\Lambda} J_\gamma(\lambda)$; (b) $J^*(\gamma) = J_\gamma(\lambda_\gamma)$ and $\max\{f^*, g^*\}$.

The optimal values of f and g for each γ, obtained by solving the resulting convex optimization problem, are shown in Figure 1. This figure illustrates the tradeoff between the individual regrets of the x- and y-players.

Extreme cases Here we discuss only the upper bound in the extreme case of this tradeoff. If we optimize solely for f, that is, if we choose η, η', c, c' to minimize only the x-player's regret, then the parameters approach $c \to 1$, $c' \to 0$, $\eta = \sqrt{M/(N+1/2)}$, and $\eta' \to 0$. In this case, the regret of the x-player is bounded by

$$\mathsf{Reg}_T^x \le f(\lambda) \to \frac{M}{\eta} + \frac{\eta}{2} + \eta N = 2\sqrt{M(N+1/2)}\,. \quad (6)$$

As we will see in Section 4, this exhibits exactly a factor of two gap compared with the corresponding lower bound.

In the case without cardinality-awareness, replacing the above η with $\eta = 1$ yields

$$\mathsf{Reg}_T^x \le f(\lambda) \to \frac{M}{\eta} + \frac{\eta}{2} + \eta N = M + N + \frac{1}{2}\,. \quad (7)$$

As we will see in Section 4, this corresponds to an additive $\log n$ factor gap compared with the corresponding lower bound.

These parameter settings correspond to the following learning dynamic: the x-player runs optimistic Hedge with a learning rate of either $\eta = \sqrt{M/(N+1/2)}$ or $\eta = 1$, while the y-player runs optimistic Hedge with a learning rate of zero (equivalent to playing the uniform strategy at every round). Although Theorem 3 cannot be applied directly to analyze the algorithm in this limiting case, the upper bounds in (6) and (7) can be obtained directly from Lemma 2 (the proof can be found in Appendix B).

Bounding max of regrets It is also important to upper bound the maximum of the two players' individual regrets. As discussed above, a closed-form expression is not available in general. Here, we derive an upper bound on the optimal solution for the case $\gamma = 1/2$, thereby controlling the maximum individual regret. By applying an appropriate change of variables (see Appendix B), we can prove the following bounds.

Lemma 8. *Let $M' = M + 1/2$, $N' = N + 1/2$, and $D = \sqrt{M'N'} + \sqrt{MN}$. Then, $\min_{\lambda\in\Lambda} \max\{f(\lambda), g(\lambda)\} \le 2\min_{\lambda\in\Lambda} J_{1/2}(\lambda) \le (20/3)(\sqrt{MN'} + \sqrt{M'N})$, where $J_{1/2}(\eta, \eta', c, c') \le$ (RHS) holds when*

$$\eta = \frac{\sqrt{MM'}}{2D}\,, \quad \eta' = \frac{\sqrt{NN'}}{2D}\,. \quad (8)$$

The learning-rate behavior under action-set cardinality asymmetry is shown in Appendix B.5, where we numerically examine how the optimal learning rates vary as the ratio between the action-set cardinalities changes.

Combining Theorem 3 with Lemma 8 immediately yields the following result.

Theorem 9. *Suppose that the x- and y-players use optimistic Hedge (Algorithm 1) with learning rates η and η' in (8). Then, $\max\{\mathsf{Reg}_T^x, \mathsf{Reg}_T^y\} \le (20/3)(\sqrt{\log m(\log n + 1/2)} + \sqrt{\log n(\log m + 1/2)})$.*

In the above analysis, an unnecessary gap arises in the first inequality of Lemma 8. By directly bounding $\max\{f(\lambda), g(\lambda)\}$ instead of bounding the minimizer of $J_{1/2}$, one can obtain some improvement.

As in the case of social regret, under strongly-uncoupled learning dynamics, the analysis based on Theorem 3 cannot achieve a bound better than the following individual regret. A rigorous argument is provided in Appendix B.

Theorem 10. *Suppose that the x- and y-players use optimistic Hedge (Algorithm 1) with learning rates $\eta = \eta' = 1/(2\sqrt{3})$ (and $c = c' = 1/2$). Then, $\max\{\mathsf{Reg}_T^x, \mathsf{Reg}_T^y\} \le 3\sqrt{3}\log(mn) + 1/\sqrt{3}$.*

Note that in the cardinality-unaware case, for both the analysis of social regret in Theorem 6 and the analysis of the maximum of the individual regrets in Theorem 10, the corresponding optimization problems $\min_\lambda \Omega(\lambda)$ and $\min_\lambda \max\{f(\lambda), g(\lambda)\}$ admit closed-form optimal solutions. This implies that, in the cardinality-unaware setting, the leading constants of the upper bounds obtained in Theorems 6 and 10 cannot be further improved as long as one relies on the commonly used regret analysis approach (Rakhlin and Sridharan, 2013; Syrgkanis et al., 2015; Anagnostides et al., 2022a,b).

4 REGRET LOWER BOUNDS

This section investigates individual regret lower bounds when each player uses optimistic Hedge in learning in two-player zero-sum games. Our main result is as follows.

Theorem 11. *Suppose that $m, n \ge 2$ and that the x- and y-players use the optimistic Hedge algorithm (Algorithm 1) with learning rates $\eta, \eta' > 0$. Then, there exists an instance of learning in two-player zero-sum games such that the re-*

gret of each player is lower bounded as

$$\mathrm{Reg}_T^x \geq \begin{cases} \dfrac{\log m}{\eta} - \dfrac{\log((m-1)(T+1))+1}{\eta(T+1)} \\ \qquad \textit{if } \eta \geq \dfrac{\log((m-1)(T+1))}{T+1}\,, \\ \dfrac{1}{\eta}\left(\log m - \eta - (m-1)\mathrm{e}^{-\eta(T+1)}\right) \\ \qquad \textit{if } \eta \in \left(0, \dfrac{\log((m-1)(T+1))}{T+1}\right], \end{cases}$$

$$\mathrm{Reg}_T^y \geq \begin{cases} \dfrac{\log n}{\eta'} - \dfrac{\log((n-1)(T+1))+1}{\eta'(T+1)} \\ \qquad \textit{if } \eta' \geq \dfrac{\log((n-1)(T+1))}{T+1}\,, \\ \dfrac{1}{\eta'}\left(\log n - \eta' - (n-1)\mathrm{e}^{-\eta'(T+1)}\right) \\ \qquad \textit{if } \eta' \in \left(0, \dfrac{\log((n-1)(T+1))}{T+1}\right]. \end{cases}$$

To the best of our knowledge, this work is the first to derive regret lower bounds for optimistic Hedge and to investigate their dependence on the numbers of actions m and n. Note that the social regret lower bound can be obtained directly from the individual regret lower bounds.

We compare the regret upper bounds in Section 3 with the lower bounds in Theorem 11. We begin with the case of cardinality-aware strongly-uncoupled learning dynamics (summarized in the lower half of Table 1). For simplicity, we focus only on the leading terms. In this case, the social regret upper bound is $4\sqrt{\log m \log n}$ for η, η' in (4), which matches the lower bound $4\sqrt{\log m \log n} - o(1)$ from Theorem 11, including the leading constant. On the other hand, for the individual regret upper bound, even if we focus solely on that of the x-player, with η, η' chosen as in (6), the individual regret asymptotically becomes $2\sqrt{\log m \log n}$. This exhibits exactly a factor-of-two gap compared with the lower bound $\sqrt{\log m \log n} - o(1)$ obtained from Theorem 11.

Next, we consider the case of strongly-uncoupled learning dynamics (summarized in the upper half of Table 1). In this setting, the social regret upper bound is $2\log(mn)$ for $\eta = \eta' = 1/2$, which matches the lower bound $2\log(mn) - o(1)$ from Theorem 11, including the leading constant. On the other hand, for the individual regret upper bound, even with η, η' chosen as in (7), it can only be improved asymptotically to $\log(mn)$. This leaves an additive gap of $\log n$ compared with $\log m - o(1)$ from Theorem 11. Closing the gaps between the upper and lower bounds for individual regret remains an important direction for future work.

Note that while we focus on the case of optimistic Hedge, which is equivalent to optimistic follow-the-regularized-leader (OFTRL) with the negative Shannon entropy, a similar argument to that in the proof of Theorem 11 can be carried out for OFTRL with other regularizers. For example, using the same instance as in Theorem 11, we can show that the regret of OFTRL with the log-barrier regularizer is roughly lower bounded by $\Omega(m \log T/\eta)$ and that the regret of OFTRL with the negative α-Tsallis entropy by $\Omega(m^{1-\alpha}/((1-\alpha)\eta))$. See Theorems 19 and 20 in Appendix E for details.

Finally, it is worth noting that Anagnostides et al. (2024) prove conditional lower bounds for Hedge and optimistic Hedge in multiplayer general-sum games under a standard complexity-theoretic assumption. By contrast, Theorem 11 gives a direct, unconditional lower bound for optimistic Hedge in two-player zero-sum games. Thus the two results are complementary rather than directly comparable.

4.1 Proof of Theorem 11

Here we provide the proof of Theorem 11. The key idea is to build a game in which action 1 keeps the same gap against every other action, regardless of the opponent's mixed strategy. Concretely, for every $i \neq 1$ and $j \neq 1$, we construct a matrix satisfying $A(1,:) - A(i,:) = \Delta_x \mathbf{1}^\top$ and $A(:,j) - A(:,1) = \Delta_y \mathbf{1}$ (see Eq. (9)). Hence, at every round, $g_t(1) - g_t(i) = \Delta_x$ and $\ell_t(j) - \ell_t(1) = \Delta_y$. The optimistic correction therefore preserves exactly the same gaps, so the weight ratios evolve as $w_t(i)/w_t(1) = \mathrm{e}^{-\eta\Delta_x t}$ and $v_t(j)/v_t(1) = \mathrm{e}^{-\eta'\Delta_y t}$. Thus both players move toward the unique equilibrium (e_1, e_1) only at an exponential rate, and summing the resulting per-round regrets yields lower bounds of order $(\log m)/\eta$ and $(\log n)/\eta'$.

In what follows, we will formally prove Theorem 11. We use the following inequality to prove the theorem:

Lemma 12. *For any $z > 0$ and $a > 0$,*

$$\frac{z}{1+z} \geq \frac{1}{a}\left[\log(1+z) - \log(1+z\mathrm{e}^{-a})\right].$$

The proof is provided in Appendix C. Now, we are ready to prove Theorem 11.

Proof of Theorem 11. We consider a game with the following payoff matrix $A \in [-1,1]^{m\times n}$:

$$A(i,j) = \begin{cases} 0 & \text{if } (i,j) = (1,1)\,, \\ \Delta_y & \text{if } i = 1, j \neq 1\,, \\ -\Delta_x & \text{if } i \neq 1, j = 1\,, \\ \Delta_y - \Delta_x & \text{otherwise} \end{cases} \tag{9}$$

for $\Delta_x, \Delta_y \in (0,1]$. Then, we can observe that

$$\begin{aligned} g_t(1) - g_t(i) &= \Delta_x \quad \text{for } i \neq 1\,, \\ \ell_t(j) - \ell_t(1) &= \Delta_y \quad \text{for } j \neq 1\,. \end{aligned}$$

We first evaluate the regret of the x-player. Since action 1 is optimal for both players, the regret of the x-player can be rewritten as

$$\mathrm{Reg}_T^x = \sum_{t=1}^T \Delta_x(1 - x_t(1))\,. \tag{10}$$

In what follows, we will evaluate $x_t(1)$. Fix arbitrary $k \in [m] \setminus \{1\}$. Then, recalling that the update rule of the optimistic Hedge algorithm can be written as (2), for any $t \geq 3$, we have

$$\begin{aligned}
&\frac{w_t(k)}{w_t(1)} \\
&= \frac{w_{t-1}(k)}{w_{t-1}(1)} \exp[\eta(2(g_{t-1}(k) - g_{t-1}(1)) - (g_{t-2}(k) - g_{t-2}(1)))] \\
&= \frac{w_{t-1}(k)}{w_{t-1}(1)} \exp(-\eta\Delta_x)\,.
\end{aligned}$$

Repeatedly applying the last equality and noting that $g_0 = \mathbf{0}$, we have

$$\begin{aligned}
\frac{w_t(k)}{w_t(1)} &= \frac{w_2(k)}{w_2(1)} \exp(-\eta\Delta_x(t-2)) \\
&= \frac{w_1(k)}{w_1(1)} \exp(-\eta\Delta_x t) = \exp(-\eta\Delta_x t)\,,
\end{aligned} \tag{11}$$

where we used $w_1(i) = 1$ for all $i \in [m]$. From (11),

$$\frac{1}{x_t(1)} = \frac{w_t(1) + \sum_{i\in[m]\setminus\{1\}} w_t(i)}{w_t(1)} = 1 + (m-1)\mathrm{e}^{-\eta\Delta_x t}\,,$$

which implies that for each $t \geq 2$ it holds that

$$x_t(1) = (1+\alpha_t)^{-1}\,,\quad \alpha_t := (m-1)\exp(-\eta\Delta_x t)\,. \tag{12}$$

For $t = 1$, since $x_1(1) = 1/m \leq 1/(1+\alpha_1)$, the same lower bound still follows.

Finally, combining (10) with (12), we can lower bound the regret of the x-player as

$$\begin{aligned}
\mathrm{Reg}_T^x &\geq \sum_{t=1}^T \Delta_x \left(1 - \frac{1}{1+\alpha_t}\right) = \Delta_x \sum_{t=1}^T \frac{\alpha_t}{1+\alpha_t} \\
&\geq \Delta_x \sum_{t=1}^T \frac{1}{\eta\Delta_x}(\log(1+\alpha_t) - \log(1+\alpha_{t+1})) \\
&\qquad \text{(Lemma 12 with } z = \alpha_t \text{ and } a = \eta\Delta_x\text{)} \\
&= \frac{1}{\eta}(\log(1+\alpha_1) - \log(1+\alpha_{T+1})) \\
&\geq \frac{1}{\eta}\left(\log m - \eta\Delta_x - (m-1)\mathrm{e}^{-\eta\Delta_x(T+1)}\right),
\end{aligned} \tag{13}$$

where in the last inequality we used $\log(1+\alpha_1) = \log(1 + (m-1)\exp(-\eta\Delta_x)) \geq \log(m\exp(-\eta\Delta_x))$ and $\log(1+z) \leq z$ for $z \geq 0$. Since the function $h\colon (0,1] \to \mathbb{R}$ given by $h(\Delta) = \eta\Delta + (m-1)\exp(-\eta\Delta(T+1))$ is minimized when $\Delta_x^* = \min\left\{1, \frac{\log((m-1)(T+1))}{\eta(T+1)}\right\}$ and its optimal value $h(\Delta_x^*)$ is

$$\begin{cases} \dfrac{\log((m-1)(T+1))+1}{T+1} & \text{if } \eta \geq \dfrac{\log((m-1)(T+1))}{T+1}\,, \\ \eta + (m-1)\mathrm{e}^{-\eta(T+1)} & \text{otherwise}\,, \end{cases}$$

choosing $\Delta_x = \Delta_x^*$ in (13) gives the desired lower bound for the x-player. Applying the same calculation on the same payoff matrix, we can obtain the desired regret lower bound for the y-player. □

Algorithm 2 Algorithm based on optimistic Hedge for the x-player with $\widetilde{O}(\log T)$ dynamic regret

1: **for** $t = 1, 2, \ldots, T$ **do**
2: Compute a strategy $\widehat{x}_t \in \Delta_m$ by the optimistic Hedge algorithm in (14).
3: Choose the time-averaged strategy x_t in (15).
4: Observe a gain vector $g_t = Ay_t \in [-1,1]^m$, where y_t is given by (15).
5: Recover $\widehat{g}_t \in [-1,1]^m$ by (16).

5 DYNAMIC REGRET BOUNDS

This section provides upper and lower bounds on dynamic regret, which are closely related to last-iterate convergence.

5.1 Dynamic Regret Upper Bounds

We describe a learning dynamic that guarantees $\widetilde{O}(1/T)$ last-iterate convergence by outputting the average of the optimistic Hedge iterates (Cai et al., 2025), which implies that each player's dynamic regret can be bounded by $\widetilde{O}(\log T)$. In this learning dynamic, each player first uses optimistic Hedge to compute $\widehat{x}_t \in \Delta_m$ and $\widehat{y}_t \in \Delta_n$ as follows:

$$\begin{aligned}
\widehat{x}_t(i) &\propto \exp\left(\eta\left(\sum_{s=1}^{t-1} \widehat{g}_s(i) + \widehat{g}_{t-1}(i)\right)\right),\ \widehat{g}_t = A\widehat{y}_t\,, \\
\widehat{y}_t(i) &\propto \exp\left(-\eta'\left(\sum_{s=1}^{t-1} \widehat{\ell}_s(i) + \widehat{\ell}_{t-1}(i)\right)\right),\ \widehat{\ell}_t = A^\top \widehat{x}_t\,,
\end{aligned} \tag{14}$$

Then, each player adopts the average of these past outputs as their strategies:

$$x_t = \frac{1}{t}\sum_{s=1}^t \widehat{x}_s\,,\quad y_t = \frac{1}{t}\sum_{s=1}^t \widehat{y}_s\,. \tag{15}$$

Their key observation is that, using the gradients defined by the actual average strategies x_t, y_t, namely $g_t = Ay_t$ and $\ell_t = A^\top x_t$, one can reconstruct the gradients $\widehat{g}_t = A\widehat{y}_t$ and $\widehat{\ell}_t = A^\top \widehat{x}_t$ (*i.e.*, the gradients that would have been obtained if optimistic Hedge outputs $\widehat{x}_t, \widehat{y}_t$ had been used directly as the strategies) as follows:

$$\widehat{g}_t = t \cdot g_t - \sum_{s=1}^{t-1} \widehat{g}_s\,,\quad \widehat{\ell}_t = t \cdot \ell_t - \sum_{s=1}^{t-1} \widehat{\ell}_s\,. \tag{16}$$

In fact, the quantities $\widehat{g}_t$ and $\widehat{\ell}_t$ can be computed from the information available up to time $t-1$ together with the gradients of the average strategies $g_t = Ay_t$ and $\ell_t = A^\top x_t$. By induction and using (14) and (15), we obtain $t \cdot g_t - \sum_{s=1}^{t-1} \widehat{g}_s = t \cdot A\left(\frac{1}{t}\sum_{s=1}^t \widehat{y}_s\right) - \sum_{s=1}^{t-1} \widehat{g}_s = \widehat{g}_t$,

which verifies the equality in (16). The algorithm for the x-player is summarized in Algorithm 2.

From the above observation, Theorem 1, and Theorem 6, we can see that this dynamic achieves the following last-iterate convergence and dynamic regret bound:

Theorem 13 (Cai et al., 2025, Theorem 3)**.** *Let $(x_t)_t$ and $(y_t)_t$ be sequences of strategies generated by Algorithm 2 with $\eta = \eta' = 1/2$. Then for any $t \geq 1$, the product distribution (x_t, y_t) is an $((2\log(mn)+1)/t)$-approximate Nash equilibrium. Consequently, the dynamic regret of each player is upper bounded as* $\max\{\mathsf{DReg}_T^x, \mathsf{DReg}_T^y\} \leq (2\log(mn)+1)(\log T + 1)$.

Under cardinality-aware strongly-uncoupled learning dynamics, the bounds of Theorem 13 can be improved as follows by using the social regret bound in Theorem 5:

Theorem 14. *Let $(x_t)_t$ and $(y_t)_t$ be sequences of strategies generated by Algorithm 2 with η, η' in (4). Then for any $t \geq 1$, the product distribution (x_t, y_t) is an $((2\sqrt{\log m\,(\log n + 1/2)} + 2\sqrt{\log n\,(\log m + 1/2)})/t)$-approximate Nash equilibrium. Consequently, the dynamic regrets are bounded as* $\max\{\mathsf{DReg}_T^x, \mathsf{DReg}_T^y\} \leq 2(\sqrt{\log m\,(\log n + 1/2)} + \sqrt{\log n\,(\log m + 1/2)})(\log T + 1)$.

Note that, by the AM–GM inequality, the convergence rate in Theorem 14 is better than that in Theorem 13.

5.2 Dynamic Regret Lower Bounds

Here we provide dynamic regret lower bounds for the learning dynamic based on optimistic Hedge in Algorithm 2.

Theorem 15. *Suppose that $m, n \geq 2$ and that the x- and y-players use Algorithm 2 with learning rates $\eta, \eta' > 0$. Let $\kappa(T) = \lfloor \sqrt{T+1} \rfloor + 1$. Then, there exists an instance of learning in two-player zero-sum games such that the dynamic regret of each player is lower bounded as*

$$\mathsf{DReg}_T^x \geq \begin{cases} \dfrac{\log(T+1)}{2\eta}\left(\log m - \dfrac{\log((m-1)\kappa(T))+1}{\kappa(T)}\right) \\ \qquad \text{if } \eta \geq \dfrac{\log((m-1)\kappa(T))}{\kappa(T)}, \\ \dfrac{\log(T+1)}{2\eta}\left(\log m - \eta - (m-1)\mathrm{e}^{-\eta\kappa(T)}\right) \\ \qquad \text{if } \eta \in \left(0, \dfrac{\log((m-1)\kappa(T))}{\kappa(T)}\right], \end{cases}$$

$$\mathsf{DReg}_T^y \geq \begin{cases} \dfrac{\log(T+1)}{2\eta'}\left(\log n - \dfrac{\log((n-1)\kappa(T))+1}{\kappa(T)}\right) \\ \qquad \text{if } \eta' \geq \dfrac{\log((n-1)\kappa(T))}{\kappa(T)}, \\ \dfrac{\log(T+1)}{2\eta'}\left(\log n - \eta' - (n-1)\mathrm{e}^{-\eta'\kappa(T)}\right) \\ \qquad \text{if } \eta' \in \left(0, \dfrac{\log((n-1)\kappa(T))}{\kappa(T)}\right]. \end{cases}$$

The proof is provided in Appendix D. A comparison with the dynamic regret upper bounds is summarized in Table 1. These results show that, for the learning dynamics presented above, the bounds are nearly optimal with respect to the numbers of actions m, n, and the number of rounds T. In the analysis of the dynamic regret lower bound, it is necessary to extract not only the logarithmic factor in the number of actions but also an additional $\log T$ factor, which worsens the constant in the lower bound by a factor of two compared with that of the external regret lower bound in Theorem 11.

6 CONCLUSION AND FUTURE WORK

In this paper, we investigated the regret upper and lower bounds of learning dynamics based on optimistic Hedge, one of the most representative dynamics for learning in two-player zero-sum games. Specifically, we first refined the regret upper bounds of optimistic Hedge. We then derived algorithm-dependent lower bounds, showing that most of these upper bounds are in fact optimal, and that the social regret is optimal even with respect to the leading constant. Finally, we extended these techniques to provide an improved upper bound and a new lower bound for dynamic regret.

This paper opens several interesting directions for future research. The first is to investigate the intermediate regimes between uncoupled learning dynamics and strongly-uncoupled learning dynamics in multiplayer general-sum games, such as cardinality-aware strongly-uncoupled learning dynamics. We showed that allowing players to know the opponent's number of actions leads to improved regret bounds. It is an interesting question whether such improvements also extend to external regret minimization (Anagnostides et al., 2022a) and swap regret minimization (Anagnostides et al., 2022b) in multiplayer general-sum games. Moreover, in our analysis of the individual and dynamic regret, the upper and lower bounds still exhibit a certain gap, and investigating whether this gap can be closed remains an important direction for future work.

A more important direction for future work is to investigate the dependence on the numbers of actions m and n for general strongly-uncoupled learning dynamics in two-player zero-sum games. A limitation of this paper is that the derived regret lower bounds are specific to the learning dynamics based on optimistic Hedge. One possible direction is to derive lower bounds separately for dynamics that possess a certain form of stability and for those that do not. For example, follow-the-leader (or fictitious play), which is an unstable algorithm, achieves $O(1)$ regret on the payoff matrix used in our lower-bound construction. However, follow-the-leader suffers linear regret against an appropriately constructed payoff matrix. Accordingly, it may be a fruitful approach to distinguish between algorithms that exhibit a certain stability property (such as Hedge) and those that lack such stability, and to analyze them separately.

Acknowledgments

The author would like to express his sincere gratitude to the anonymous reviewers for their insightful feedback and constructive suggestions. TT is supported by JSPS KAKENHI Grant Number JP24K23852.

References

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.

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.

Ioannis Anagnostides, Alkis Kalavasis, Tuomas Sandholm, and Manolis Zampetakis. On the complexity of computing sparse equilibria and lower bounds for no-regret learning in games. In *15th Innovations in Theoretical Computer Science Conference*, volume 287, pages 5:1–5:24, 2024.

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

Stephen P Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Yang Cai, Haipeng Luo, Chen-Yu Wei, and Weiqiang Zheng. From average-iterate to last-iterate convergence in games: A reduction and its applications. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

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.

Constantinos Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *10th Innovations in Theoretical Computer Science conference*, 2019.

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.

Meta Fundamental AI Research Diplomacy Team FAIR, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sasha Mitts, Adithya Renduchintala, Stephen Roller, Dirk Rowe, Weiyan Shi, Joe Spisak, Alexander Wei, David Wu, Hugh Zhang, and Markus Zijlstra. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

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.

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.

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

Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *Advances in Neural Information Processing Systems*, volume 33, pages 20766–20778. Curran Associates, Inc., 2020a.

Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman Ozdaglar. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125, pages 1758–1784. PMLR, 2020b.

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

Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93(5):1830–1836, 2003.

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.

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

Yuheng Li, Wang Panpan, and Haipeng Chen. Can reinforcement learning solve asymmetric combinatorial-

continuous zero-sum games? In *The Thirteenth International Conference on Learning Representations*, 2025.

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

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

Remi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J Mankowitz, Doina Precup, and Bilal Piot. Nash learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 36743–36768. PMLR, 2024.

Yuyuan Ouyang and Yangyang Xu. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Mathematical Programming*, 185(1):1–35, 2021.

Julien Perolat, Bart De Vylder, Daniel Hennes, Eugene Tarassov, Florian Strub, Vincent de Boer, Paul Muller, Jerome T. Connor, Neil Burch, Thomas Anthony, Stephen McAleer, Romuald Elie, Sarah H. Cen, Zhe Wang, Audrunas Gruslys, Aleksandra Malysheva, Mina Khan, Sherjil Ozair, Finbarr Timbers, Toby Pohlen, Tom Eccles, Mark Rowland, Marc Lanctot, Jean-Baptiste Lespiau, Bilal Piot, Shayegan Omidshafiei, Edward Lockhart, Laurent Sifre, Nathalie Beauguerlange, Remi Munos, David Silver, Satinder Singh, Demis Hassabis, and Karl Tuyls. Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science*, 378(6623):990–996, 2022.

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., 2013.

Gokul Swamy, Christoph Dann, Rahul Kidambi, Steven Wu, and Alekh Agarwal. A minimaximalist approach to reinforcement learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 47345–47377. PMLR, 2024.

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.

Taira Tsuchiya, Shinji Ito, and Haipeng Luo. Corrupted learning dynamics in games. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291, pages 5506–5552. PMLR, 2025.

Alan Washburn and Kevin Wood. Two-person zero-sum games for network interdiction. *Operations Research*, 43 (2):243–251, 1995.

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

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

Bryan Wilder. Equilibrium computation and robust optimization in zero sum games with submodular structure. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2018.

Taeho Yoon and Ernest K Ryu. Accelerated algorithms for smooth convex-concave minimax problems with $O(1/k^2)$ rate on squared gradient norm. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 12098–12109. PMLR, 2021.

Yuheng Zhao, Yu-Hu Yan, Kfir Yehuda Levy, and Peng Zhao. Gradient-variation online adaptivity for accelerated optimization with Hölder smoothness. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

   The problem setting is detailed in Section 2. The complexity analysis is omitted since the algorithm is the well-known optimistic Hedge. The source code for Figure 1 is included in the supplemental material.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

   The assumptions are fully provided for each proposition. The complete proofs are fully provided in Section 3 and the appendix.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] The source code for Figure 1 is included in the supplemental material.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

   This study is primarily theoretical and did not use existing assets.

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

   This study is primarily theoretical and did not use crowdsourcing or conduct research with human subjects.

Tight Regret Upper and Lower Bounds for Optimistic Hedge in Two-Player Zero-Sum Games: Supplementary Materials

A ADDITIONAL RELATED WORK

In this section, we discuss additional related work that could not be included in the main text. In two-player zero-sum games, it was first pointed out by Daskalakis et al. (2011) that a fast convergence rate of $\widetilde{O}(1/T)$ can be achieved. Subsequently, it was shown that both the optimistic Hedge algorithm and its generalized framework, optimistic follow-the-regularized-leader (OFTRL), can guarantee $\widetilde{O}(1)$ social regret, which corresponds to an $\widetilde{O}(1/T)$ convergence rate (Rakhlin and Sridharan, 2013; Syrgkanis et al., 2015). Since then, achieving fast rates via optimistic prediction has become a central approach when designing learning dynamics (*e.g.,* Foster et al. 2016; Wei and Luo 2018; Chen and Peng 2020; Anagnostides et al. 2022b). It is now known that such dynamics can guarantee $O(1)$ individual regret upper bounds in two-player zero-sum games and an $O(\log T)$ bound in multiplayer general-sum games, ignoring dependencies other than on T. It is worth noting that the optimistic Hedge algorithm does not always guarantee last-iterate convergence, but it has been shown to achieve last-iterate convergence under certain conditions (Daskalakis and Panageas, 2019; Hsieh et al., 2021; Wei et al., 2021).

While a large body of work in learning in games has focused on deriving upper bounds, lower bounds remain relatively underexplored. Several studies, including this work, investigated algorithm-dependent lower bounds. Syrgkanis et al. (2015) considered a setting where the x-player employs the vanilla Hedge algorithm with an arbitrary learning rate, while the y-player plays a (pure) best response. For such a scenario, they showed that there exists an instance of learning in two-player zero-sum games in which the x-player must suffer $\sqrt{T}$ regret. Chen and Peng (2020) demonstrated that when both the x- and y-players use the vanilla Hedge with any learning rate, there exists a two-player *general-sum* game in which at least one of the players incurs $\sqrt{T}$ regret. In contrast, our work is the first to analyze regret lower bounds for *optimistic* Hedge, to investigate their dependence on the numbers of actions m and n, and to study the dynamic regret lower bounds.

Algorithm-dependent lower bounds on convergence rates have also been investigated in the context of last-iterate convergence. For example, Golowich et al. (2020b) provided an $\Omega(1/\sqrt{T})$ lower bound on the last-iterate convergence rate for a class of algorithms that includes the extragradient method, and Golowich et al. (2020a) established a similar lower bound for the optimistic gradient algorithm. However, these results differ from ours not only in that they focus on the convergence rate of the last iterates, but also in that they assume an unconstrained setting. In a related line of research, the fundamental limits of first-order methods have also been studied (Ouyang and Xu, 2021; Yoon and Ryu, 2021). For a comprehensive overview of the literature on last-iterate convergence in the full-information (*i.e.,* gradient feedback) setting, we refer the reader to Cai et al. (2025) and the references therein.

We finally compare our lower bounds with the computational lower bounds. A complementary line of work is due to Anagnostides et al. (2024), who study computational lower bounds for no-regret learning in multiplayer general-sum games. Under the Exponential Time Hypothesis for PPAD, they show that if each player in an n-player m-action normal-form game follows MWU (Hedge) or OMWU (optimistic MWU), then there exists a game and an absolute constant $\varepsilon > 0$ such that achieving $\frac{1}{T}\max_{1\le i\le n}\mathsf{Reg}_i^T \le \varepsilon$ requires $T \ge 2^{(\log_2\log_2 m)^{1/2-o(1)}}$. Equivalently, as noted after their Corollary 3.9, if one writes a regret guarantee in the canonical form $\mathsf{Reg}_i^T \le R(m)\,T^{1-\gamma}$ for some $\gamma \in (0,1)$, then their result implies $R(m) \ge 2^{\gamma(\log_2\log_2 m)^{1/2-o(1)}}$. Thus the proven lower-bound term is dimension-dependent, while still growing strictly slower than any fixed power of $\log m$. They also obtain broader hardness results for no-regret learning in extensive-form games in a more permissive, potentially centralized model. In particular, they explicitly note that analogous hardness cannot hold in that model for two-player zero-sum games, since a Nash equilibrium can be computed and then played forever. Our result is therefore complementary: we work in the classical uncoupled online-learning model for two-player zero-sum matrix games and prove unconditional, information-theoretic lower bounds for optimistic Hedge by analyzing an explicit game instance in closed form, without passing through sparse coarse correlated equilibria. Moreover, Anagnostides et al. (2024) suggest that more elementary unconditional dimension-dependent lower bounds for MWU-type algorithms may exist under suitable parameterizations; in this sense, Theorem 11 can be viewed as a concrete partial realization of that

direction in the classical two-player zero-sum setting.

B OMITTED DETAILS FROM SECTION 3

This section provides deferred omitted details from Section 3.

B.1 Regret Analysis of Optimistic Hedge (Proof of Lemma 2)

Here we provide an analysis of optimistic Hedge. In this subsection, we use the standard notation of online linear optimization. Specifically, at each round $t = 1, \dots, T$, the player selects a point $w_t \in \mathcal{K}$ from a convex feasible set $\mathcal{K} \subseteq \mathbb{R}^d$, the environment chooses a loss vector $z_t \in \mathbb{R}^d$, and the player incurs a loss of $\langle w_t, z_t \rangle$. We use $D_\psi(v, w)$ to denote the Bregman divergence between v and w induced by a differentiable convex function ψ, that is, $D_\psi(v, w) = \psi(v) - \psi(w) - \langle \nabla\psi(w), v - w \rangle$.

The following lemma provides a regret bound for the optimistic follow-the-regularized-leader (OFTRL), which generalizes the optimistic Hedge algorithm (adapted from Tsuchiya et al. 2025, Lemma 16):

Lemma 16. *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a nonempty closed convex set. Suppose that a sequence of points $w_1, \dots, w_T \in \mathcal{K}$ is selected by OFTRL, $w_t \in \arg\min_{w \in \mathcal{K}} \{\langle w, m_t + \sum_{s=1}^{t-1} z_s \rangle + \psi_t(w)\}$, for each round $t \in [T]$. Then, for any $w^* \in \mathcal{K}$, it holds that*

$$\sum_{t=1}^{T} \langle w_t - w^*, z_t \rangle \le \psi_{T+1}(w^*) - \psi_1(w_1) + \sum_{t=1}^{T} (\psi_t(w_{t+1}) - \psi_{t+1}(w_{t+1})) + \sum_{t=1}^{T} (\langle w_t - w_{t+1}, z_t - m_t \rangle - D_{\psi_t}(w_{t+1}, w_t)) + \langle w^* - w_{T+1}, m_{T+1} \rangle . \tag{17}$$

Lemma 16 immediately yields the following regret upper bound:

Lemma 17. *Let $\psi_t(w) = -\frac{1}{\eta_t} H(w)$ for $H(w) = \sum_{i=1}^{d} w(i) \log(1/w(i))$ be the negative Shannon entropy regularizer with nonincreasing learning rate η_t and $w_t \in \arg\min_{w \in \Delta_d} \{\langle w, m_t + \sum_{s=1}^{t-1} z_s \rangle + \psi_t(w)\}$ with $m_{T+1} = \mathbf{0}$. Then, for any $w^* \in \Delta_d$ and $c_1, \dots, c_T > 0$, it holds that*

$$\sum_{t=1}^{T} \langle w_t - w^*, z_t \rangle \le \frac{\log d}{\eta_{T+1}} + \sum_{t=1}^{T} \frac{\eta_t}{2c_t} \|z_t - m_t\|_\infty^2 - \sum_{t=1}^{T} \frac{1 - c_t}{2\eta_t} \|w_t - w_{t+1}\|_1^2 .$$

This lemma generalizes Tsuchiya et al. (2025, Lemma 17). Choosing $\eta_t = \eta$, $c_t = c$ for all $t \in [T]$ and $m_{T+1} = \mathbf{0}$ yields Lemma 2.

Proof. We will upper bound the RHS of (17) in Lemma 16. From $\max_{w \in \Delta_d} H(w) \le \log d$,

$$\psi_{T+1}(w^*) - \psi_1(w_1) + \sum_{t=1}^{T} (\psi_t(w_{t+1}) - \psi_{t+1}(w_{t+1})) \le \frac{\log d}{\eta_1} + \sum_{t=1}^{T} \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \log d = \frac{\log d}{\eta_{T+1}} .$$

Fix an arbitrary $c_t > 0$. Then, we also have

$$\begin{aligned}
\langle w_t - w_{t+1}, z_t - m_t \rangle - D_{\psi_t}(w_{t+1}, w_t) &= \langle w_t - w_{t+1}, z_t - m_t \rangle - \frac{1}{\eta_t} D_{(-H)}(w_{t+1}, w_t) \\
&\le \|w_t - w_{t+1}\|_1 \|z_t - m_t\|_\infty - \frac{1}{2\eta_t} \|w_t - w_{t+1}\|_1^2 \\
&= \|w_t - w_{t+1}\|_1 \|z_t - m_t\|_\infty - \frac{c_t}{2\eta_t} \|w_t - w_{t+1}\|_1^2 - \frac{1 - c_t}{2\eta_t} \|w_t - w_{t+1}\|_1^2 \\
&\le \frac{\eta_t}{2c_t} \|z_t - m_t\|_\infty^2 - \frac{1 - c_t}{2\eta_t} \|w_t - w_{t+1}\|_1^2 ,
\end{aligned}$$

where the first inequality follows from Hölder's inequality and the fact that the function $(-H)$ is 1-strongly convex with respect to $\|\cdot\|_1$, and the last inequality follows by considering the worst-case with respect to $\|w_t - w_{t+1}\|_1$ in the first two terms. Combining Lemma 16 with the above two inequalities completes the proof. □

B.2 Proof of Theorem 3

Here we provide the proof of Theorem 3.

Proof of Theorem 3. Summing the regret upper bounds in Lemma 2 and using $\|g_1 - g_0\|_\infty \le 1$, $\|\ell_1 - \ell_0\|_\infty \le 1$, $\|g_t - g_{t-1}\|_\infty = \|A(y_t - y_{t-1})\|_\infty \le \|y_t - y_{t-1}\|_1$ and $\|\ell_t - \ell_{t-1}\|_\infty \le \|x_t - x_{t-1}\|_1$ that hold for all $t \in [T]$, we have

$$\begin{aligned}
&\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \\
&\le \frac{\log m}{\eta} + \frac{\eta}{2c} + \frac{\log n}{\eta'} + \frac{\eta'}{2c'} + \left(\frac{\eta'}{2c'} - \frac{1-c}{2\eta}\right)\sum_{t=2}^T \|x_t - x_{t-1}\|_1^2 + \left(\frac{\eta}{2c} - \frac{1-c'}{2\eta'}\right)\sum_{t=2}^T \|y_t - y_{t-1}\|_1^2 \\
&= \Omega(\eta, \eta', c, c') + \left(\frac{\eta'}{2c'} - \frac{1-c}{2\eta}\right)\sum_{t=2}^T \|x_t - x_{t-1}\|_1^2 + \left(\frac{\eta}{2c} - \frac{1-c'}{2\eta'}\right)\sum_{t=2}^T \|y_t - y_{t-1}\|_1^2 , \qquad (18)
\end{aligned}$$

where we recall

$$\Omega(\eta, \eta', c, c') = \omega(\eta, c) + \omega'(\eta', c') , \quad \omega(\eta, c) = \frac{\log m}{\eta} + \frac{\eta}{2c} , \quad \omega'(\eta', c') = \frac{\log n}{\eta'} + \frac{\eta'}{2c'} .$$

Note that when $\eta\eta' < c'(1-c)$, we have $\frac{\eta'}{2c'} - \frac{1-c}{2\eta} < 0$ and when $\eta\eta' < c(1-c')$ we have $\frac{\eta}{2c} - \frac{1-c'}{2\eta'} < 0$. Hence, when $\eta\eta' \le c'(1-c)$ and $\eta\eta' < c(1-c')$, from (18) and the fact that $\mathsf{SocialReg}_T \ge 0$, we have

$$\sum_{t=2}^T \|y_t - y_{t-1}\|_1^2 \le \frac{1}{\frac{1-c'}{2\eta'} - \frac{\eta}{2c}} \cdot \Omega(\eta, \eta', c, c') .$$

Hence, combining Lemma 2 with $\|g_t - g_{t-1}\|_\infty \le \|y_t - y_{t-1}\|_1$ and the last inequality, we obtain

$$\mathsf{Reg}_T^x \le \frac{\log m}{\eta} + \frac{\eta}{2c} + \frac{\eta}{2c}\sum_{t=2}^T \|y_t - y_{t-1}\|_1^2 \le \omega(\eta, c) + \frac{\frac{\eta}{2c}}{\frac{1-c'}{2\eta'} - \frac{\eta}{2c}} \cdot \Omega(\eta, \eta', c, c') = f(\eta, \eta', c, c') .$$

Similarly, if $\eta\eta' < c'(1-c)$ and $\eta\eta' \le c(1-c')$, from (18) and the fact that $\mathsf{SocialReg}_T \ge 0$ we have

$$\sum_{t=2}^T \|x_t - x_{t-1}\|_1^2 \le \frac{1}{\frac{1-c}{2\eta} - \frac{\eta'}{2c'}} \cdot \Omega(\eta, \eta', c, c') .$$

Hence, combining Lemma 2 with $\|\ell_t - \ell_{t-1}\|_\infty \le \|x_t - x_{t-1}\|_1$ and the last inequality, we obtain

$$\mathsf{Reg}_T^y \le \frac{\log n}{\eta'} + \frac{\eta'}{2c'} + \frac{\eta'}{2c'}\sum_{t=2}^T \|x_t - x_{t-1}\|_1^2 \le \omega'(\eta', c') + \frac{\frac{\eta'}{2c'}}{\frac{1-c}{2\eta} - \frac{\eta'}{2c'}} \cdot \Omega(\eta, \eta', c, c') = g(\eta, \eta', c, c') ,$$

This completes the proof. □

B.3 Social Regret Analysis Deferred from Section 3.2

Here we first provide the proof of Lemma 4, which will be used to obtain the tight upper bound on the social regret. We then provide the proof of Theorem 6.

B.3.1 Proof of Lemma 4 (cardinality-aware case)

Recall that Lemma 4 is for cardinality-aware strongly-uncoupled learning dynamics, and thus we can choose λ and λ' by considering the RHS of the following inequality:

$$\mathsf{SocialReg}_T \le \inf_{\lambda \in \Lambda} \Omega(\eta, \eta', c, c') , \quad \Omega(\eta, \eta', c, c') = \frac{M}{\eta} + \frac{\eta}{2c} + \frac{N}{\eta'} + \frac{\eta'}{2c'} , \qquad (19)$$

where we recall that $\Lambda = \{\lambda = (\eta, \eta', c, c') \in \mathbb{R}^4_{>0} \colon \eta\eta' \le c'(1-c),\ \eta\eta' \le c(1-c')\}$, $M = \log m$, and $N = \log n$.

Proof of Lemma 4. From the constraints that $\eta, \eta' > 0$ and $\eta\eta' \le c'(1-c)$, $\eta\eta' \le c(1-c')$, we have $c, c' \in (0,1)$. Hence, the constraints $\eta\eta' \le c'(1-c)$ and $\eta\eta' \le c(1-c')$ can be rewritten as $\frac{\eta}{c} \cdot \frac{\eta'}{c'} \le \frac{1}{c} - 1$ and $\frac{\eta}{c} \cdot \frac{\eta'}{c'} \le \frac{1}{c'} - 1$, respectively.

Now we consider the following change of variables:

$$a = \frac{\eta}{c}\,, \quad a' = \frac{\eta'}{c'}\,, \quad b = \frac{1}{c} - 1\,, \quad b' = \frac{1}{c'} - 1\,. \tag{20}$$

Note that this is a bijective transformation, and we have

$$c = \frac{1}{1+b} \in (0,1)\,, \quad c' = \frac{1}{1+b'} \in (0,1)\,, \quad \eta = ac\,, \quad \eta' = a'c'\,. \tag{21}$$

Then the constraints $\eta\eta' \le c'(1-c)$ and $\eta\eta' \le c(1-c')$ can be rewritten as $aa' \le b$ and $aa' \le b'$, respectively, and the RHS of the inequality in (19) can be rewritten as

$$\inf_{a,a',b,b' > 0\colon\, aa' \le b, aa' \le b'} \Omega(a,a',b,b')\,, \quad \Omega(a,a',b,b') = (1+b)\frac{M}{a} + \frac{a}{2} + (1+b')\frac{N}{a'} + \frac{a'}{2}\,,$$

where we abuse the notation of Ω. The rewritten function Ω is monotonically increasing with respect to b and b'. Hence the optimal choices of b and b' are $b = b' = aa'$ and in this case, we have

$$c = c' = \frac{1}{1+aa'}\,, \quad \eta\eta' = \frac{aa'}{(1+aa')^2} = c'(1-c) = c(1-c')\,,$$

and

$$\Omega(a,a',b,b') = (1+aa')\left(\frac{M}{a} + \frac{N}{a'}\right) + \frac{a}{2} + \frac{a'}{2} = \left(\frac{M}{a} + \left(N + \frac{1}{2}\right)a\right) + \left(\frac{N}{a'} + \left(M + \frac{1}{2}\right)a'\right).$$

From the AM–GM inequality, choosing

$$a = \sqrt{\frac{M}{N+\frac{1}{2}}}\,, \quad a' = \sqrt{\frac{N}{M+\frac{1}{2}}}\,,$$

gives the minimum value of Ω and its optimal value is $2\sqrt{M\left(N+\frac{1}{2}\right)} + 2\sqrt{N\left(M+\frac{1}{2}\right)}$. From (21), the optimal parameters of (19) are given by

$$c = c' = \frac{\sqrt{\left(M+\frac{1}{2}\right)\left(N+\frac{1}{2}\right)}}{\sqrt{\left(M+\frac{1}{2}\right)\left(N+\frac{1}{2}\right)} + \sqrt{MN}}\,, \; \eta = \frac{\sqrt{M\left(M+\frac{1}{2}\right)}}{\sqrt{\left(M+\frac{1}{2}\right)\left(N+\frac{1}{2}\right)} + \sqrt{MN}}\,, \; \eta' = \frac{\sqrt{N\left(N+\frac{1}{2}\right)}}{\sqrt{\left(M+\frac{1}{2}\right)\left(N+\frac{1}{2}\right)} + \sqrt{MN}}\,,$$

which are indeed elements in the feasible set Λ, and we have completed the proof of Lemma 4. □

B.3.2 Proof of Theorem 6 (cardinality-unaware case)

We next provide the proof of Theorem 6. Under strongly-uncoupled learning dynamics without cardinality-awareness, the learning rates $\eta, \eta' > 0$ cannot be chosen as functions of $M = \log m$ and $N = \log n$. Note, however, that the parameters $c, c' > 0$ are not algorithm-dependent variables and thus may depend on M and N.

Proof of Theorem 6. Define

$$\Lambda(\eta,\eta') = \{(c,c') \in \mathbb{R}^2_{>0}\colon \eta\eta' \le c'(1-c),\, \eta\eta' \le c(1-c')\}\,.$$

Then, we will show that

$$\min_{(c,c') \in \Lambda(\eta,\eta')} \left\{\frac{\eta}{2c} + \frac{\eta'}{2c'}\right\} = \frac{\eta+\eta'}{1+\sqrt{1-4\eta\eta'}}\,,$$

and the optimal $c, c' \in \Lambda(\eta,\eta')$ achieving the minimum are given by $c = c' = \frac{1+\sqrt{1-4\eta\eta'}}{2}$. To prove this, from the KKT condition, letting $\mathcal{L}(c,c',\mu_1,\mu_2) = \frac{\eta}{2c} + \frac{\eta'}{2c'} + \mu_1(\eta\eta' - c'(1-c)) + \mu_2(\eta\eta' - c(1-c'))$, we have

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial c} = -\frac{\eta}{2c^2} + \mu_1 c' - \mu_2(1-c') = 0\,, \quad \frac{\partial \mathcal{L}}{\partial c'} = -\frac{\eta'}{2c'^2} - \mu_1(1-c) + \mu_2 c = 0\,, \\ &\mu_1(\eta\eta' - c'(1-c)) = 0\,, \quad \mu_2(\eta\eta' - c(1-c')) = 0\,, \quad \mu_1, \mu_2 \ge 0\,. \end{aligned} \tag{22}$$

From the third and fourth equalities in (22), we claim that $\mu_1, \mu_2 > 0$. Indeed, if $\mu_1 = 0$, then the first equality in (22) gives $-\frac{\eta}{2c^2} - \mu_2(1-c') = 0$, which is impossible since $c \in (0,1)$ and $\mu_2 \geq 0$. Similarly, if $\mu_2 = 0$, then the second equality in (22) gives $-\frac{\eta'}{2c'^2} - \mu_1(1-c) = 0$, which is a contradiction. Hence, from $\mu_1, \mu_2 > 0$, we have $\eta\eta' = c'(1-c) = c(1-c')$. From these two equalities, we have $c = c'$. Hence, $c(1-c) = \eta\eta'$ and thus $c = c' = \frac{1+\sqrt{1-4\eta\eta'}}{2}$.

From the above observation, it suffices to choose absolute constants $\eta, \eta' > 0$ (independent of M, N) to minimize

$$\min_{(c,c') \in \Lambda(\eta,\eta')} \Omega(\eta, \eta', c, c') = \frac{M}{\eta} + \frac{N}{\eta'} + \frac{\eta + \eta'}{1 + \sqrt{1 - 4\eta\eta'}} =: F(\eta, \eta') \,.$$

A natural approach is to minimize a linear upper bound that holds for all $M, N > 0$:

$$F(\eta, \eta') \leq S(\eta, \eta')(M + N) + \frac{\eta + \eta'}{1 + \sqrt{1 - 4\eta\eta'}} \,, \quad S(\eta, \eta') = \max\left\{\frac{1}{\eta}, \frac{1}{\eta'}\right\} .$$

The slope $S(\eta, \eta')$ is minimized when $\eta = \eta'$. Since $\eta\eta' \leq \max_{c,c' \in [0,1]} \min\{c'(1-c), c(1-c')\} = 1/4$, choosing $\eta = \eta' = 1/2$ and thus $c = c' = 1/2$ are optimal choices. Therefore, with these choices of η, η', for all $M, N > 0$ we have

$$F(\eta, \eta') \leq 2(M + N) + 1 \,,$$

which leads to the desired social regret upper bound under strongly-uncoupled learning dynamics. □

B.4 Individual Regret Analysis Deferred from Section 3.3

Here we provide deferred details from Section 3.3. Recall that, as defined in Theorem 3, f and g are given by

$$\begin{aligned} f(\eta, \eta', c, c') &= \omega(\eta, c) + \frac{\frac{\eta}{2c}}{\frac{1-c'}{2\eta'} - \frac{\eta}{2c}} \Omega(\eta, \eta', c, c') = \frac{\log m}{\eta} + \frac{\eta}{2c} + \frac{\frac{\eta}{2c}}{\frac{1-c'}{2\eta'} - \frac{\eta}{2c}} \left(\frac{\log m}{\eta} + \frac{\log n}{\eta'} + \frac{\eta}{2c} + \frac{\eta'}{2c'} \right) , \\ g(\eta, \eta', c, c') &= \omega'(\eta', c') + \frac{\frac{\eta'}{2c'}}{\frac{1-c}{2\eta} - \frac{\eta'}{2c'}} \Omega(\eta, \eta', c, c') = \frac{\log n}{\eta'} + \frac{\eta'}{2c'} + \frac{\frac{\eta'}{2c'}}{\frac{1-c}{2\eta} - \frac{\eta'}{2c'}} \left(\frac{\log m}{\eta} + \frac{\log n}{\eta'} + \frac{\eta}{2c} + \frac{\eta'}{2c'} \right) , \end{aligned} \tag{23}$$

where ω, ω', and Ω are defined in (3).

B.4.1 Common Analysis

We introduce the variables $s, s' \geq 0$ so that the optimization problem for f, g in (23) can be equivalently expressed as an optimization problem over (a, a', s, s'):

$$s = \frac{b}{a} - a' = \frac{b - aa'}{a} \geq 0 \,, \quad s' = \frac{b'}{a'} - a = \frac{b' - aa'}{a'} \geq 0 \,,$$

where (a, a', b, b') are defined in (20). Then, we have

$$c = \frac{1}{1+b} = \frac{1}{1 + a(a' + s)} \,, \quad c' = \frac{1}{1 + a'(a + s')} \,, \quad \eta = ca = \frac{a}{1 + a(a' + s)} \,, \quad \eta' = \frac{a'}{1 + a'(a + s')} \,. \tag{24}$$

Using (a, a', s, s'), we can rewrite ω, ω', and Ω in (3) as (we abuse notation of ω, ω', and Ω again)

$$\begin{aligned} &\omega(a, a', s) = \frac{M}{\eta} + \frac{\eta}{2c} = \frac{M}{a} + (a' + s)M + \frac{a}{2} \,, \quad \omega'(a, a', s') = \frac{N}{\eta'} + \frac{\eta'}{2c'} = \frac{N}{a'} + (a + s')N + \frac{a'}{2} \,, \\ &\Omega(a, a', s, s') = h(a, a') + sM + s'N \,, \end{aligned}$$

where we defined

$$h(a, a') := \frac{M}{a} + a'M + \frac{N}{a'} + aN + \frac{a}{2} + \frac{a'}{2} = \frac{M}{a} + \left(N + \frac{1}{2}\right)a + \frac{N}{a'} + \left(M + \frac{1}{2}\right)a' \,. \tag{25}$$

We also have

$$\frac{1-c'}{2\eta'} - \frac{\eta}{2c} = \frac{1}{2}\left(\frac{(1-c')/c'}{\eta'/c'} - a\right) = \frac{1}{2}\left(\frac{b'}{a'} - a\right) = \frac{s'}{2} \,, \quad \frac{1-c}{2\eta} - \frac{\eta'}{2c'} = \frac{s}{2} \,,$$

and thus f and g in (23) can be rewritten as

$$\begin{aligned} f(a,a',s,s') &= \left(\frac{M}{a} + a'M + \frac{a}{2} + aN\right) + sM + \frac{a}{s'}(h(a,a') + sM)\,, \\ g(a,a',s,s') &= \left(\frac{N}{a'} + aN + \frac{a'}{2} + a'M\right) + s'N + \frac{a'}{s}(h(a,a') + s'N)\,. \end{aligned} \tag{26}$$

where we replace the arguments (η, η', c, c') of f, g with (a, a', s, s') by abuse of notation.

B.4.2 Extreme Cases

We supplement the discussion of the extreme cases, where the goal is to minimize only one player's individual regret, which was omitted in the main text. We considered the setting where the x-player uses optimistic Hedge (with a learning rate of either $\eta = \sqrt{M/(N+1/2)}$ in the cardinality-aware case or $\eta = 1$ in the cardinality-unaware case), while the y-player plays the uniform strategy at every round. There, we showed that by the asymptotic argument (that is in fact not applicable in this limiting scenario) it holds that $\mathsf{Reg}_T^x \le 2\sqrt{M(N+1/2)}$ in (6) for the cardinality-aware case and $\mathsf{Reg}_T^x \le M + N + 1/2$ in (7) for the cardinality-unaware case. These bounds (or their improved bounds) can be obtained via a direct analysis. Specifically, combining the x-player's regret upper bound in Lemma 2 with the fact that $g_t = g_{t-1} = A(\frac{1}{n}\mathbf{1})$ for all $t \ge 2$, we have

$$\begin{aligned} \mathsf{Reg}_T^x &\le \inf_{c>0}\left\{\frac{\log m}{\eta} + \frac{\eta}{2c}\sum_{t=1}^{T}\|g_t - g_{t-1}\|_\infty^2 - \frac{1-c}{2\eta}\sum_{t=2}^{T}\|x_t - x_{t-1}\|_1^2\right\} \\ &\le \frac{\log m}{\eta} + \frac{\eta}{2} \le \begin{cases} \frac{3}{2}\sqrt{M\left(N+\frac{1}{2}\right)} & \text{cardinality-aware case}\,, \\ M + \frac{1}{2} & \text{cardinality-unaware case}\,, \end{cases} \end{aligned}$$

where we chose $c = 1$ and used $\|g_1 - g_0\|_\infty \le 1$.

B.4.3 Upper Bounding the Maximum of the Individual Regrets

Here we provide the omitted details to upper bound the maximum of the individual regrets provided in Lemma 8 and Theorems 9 and 10. From the analysis in Appendix B.4.1, we can rewrite $J_\gamma(\eta, \eta', c, c')$ in (5) as

$$\begin{aligned} J_\gamma(a,a',s,s') &= \gamma\left(\frac{M}{a} + a'M + \frac{a}{2} + aN\right) + (1-\gamma)\left(\frac{N}{a'} + aN + \frac{a'}{2} + a'M\right) \\ &\quad + \gamma sM + \gamma\frac{a}{s'}(h(a,a') + sM) + (1-\gamma)s'N + (1-\gamma)\frac{a'}{s}(h(a,a') + s'N)\,, \end{aligned} \tag{27}$$

where we replaced the arguments (η, η', c, c') with (a, a', s, s') by abuse of notation.

Cardinality-aware case As discussed in the main text, in the cardinality-aware case it is difficult to compute a closed-form expression for the exact minimum of the individual regret or the minimizer that attains it. To address this, we will derive an upper bound of $\max\{f, g\}$. Specifically, we will focus on the case of $\gamma = 1/2$, which is for Lemma 8 and Theorem 9.

Proof of Lemma 8. We consider the case where $s, s' \ge 0$ are expressed as $s = \theta a'$ and $s' = \theta a$ for some $\theta > 0$. In this case, we can rewrite J_γ in (27) as

$$\begin{aligned} J_\gamma(a,a',s,s') &= \frac{1}{2}\left(\frac{M}{a} + a'M + \frac{a}{2} + aN\right) + \frac{1}{2}\left(\frac{N}{a'} + aN + \frac{a'}{2} + a'M\right) \\ &\quad + \frac{1}{2}\theta a'M + \frac{1}{2\theta}(h(a,a') + \theta a'M) + \frac{1}{2}\theta aN + \frac{1}{2\theta}(h(a,a') + \theta aN) \\ &= \frac{1}{2a}\left(M + \frac{2M}{\theta}\right) + \frac{a}{2}\left(\frac{1}{2} + 2N + \frac{2(N+4)}{\theta} + \theta N + N\right) \\ &\quad + \frac{1}{2a'}\left(N + \frac{2N}{\theta}\right) + \frac{a'}{2}\left(\frac{1}{2} + 2M + \frac{2(M+4)}{\theta} + \theta M + M\right), \end{aligned} \tag{28}$$

where h is given in (25). Then, we can evaluate $\inf_{\lambda\in\Lambda} J_{1/2}(\lambda) = \inf_{a,a',s,s'>0} J_{1/2}(a,a',s,s')$ as

$$\begin{aligned}
\inf_{\lambda\in\Lambda} J_{1/2}(\lambda) &\le \inf_{\theta\ge 0}\left\{\sqrt{M\left(1+\frac{2}{\theta}\right)\left(\left(\theta+3+\frac{2}{\theta}\right)N+\frac{8}{\theta}+\frac{1}{2}\right)}+\sqrt{N\left(1+\frac{2}{\theta}\right)\left(\left(\theta+3+\frac{2}{\theta}\right)M+\frac{8}{\theta}+\frac{1}{2}\right)}\right\}\\
&\le \sqrt{\frac{5}{3}M\cdot\left(\frac{20}{3}N+\frac{19}{6}\right)}+\sqrt{\frac{5}{3}N\cdot\left(\frac{20}{3}M+\frac{19}{6}\right)}\\
&\le \sqrt{\frac{5}{3}M\cdot\frac{20}{3}\left(N+\frac{1}{2}\right)}+\sqrt{\frac{5}{3}N\cdot\frac{20}{3}\left(M+\frac{1}{2}\right)}\\
&= \frac{10}{3}\left(\sqrt{M\left(N+\frac{1}{2}\right)}+\sqrt{N\left(M+\frac{1}{2}\right)}\right),
\end{aligned}$$

where in the first inequality, we chose

$$a=\sqrt{\frac{M+\frac{2M}{\theta}}{\frac{2}{3}+(\theta+3)N+\frac{2(N+4)}{\theta}}},\quad a'=\sqrt{\frac{N+\frac{2N}{\theta}}{\frac{2}{3}+(\theta+3)M+\frac{2(M+4)}{\theta}}},$$

(here we chose the first term in the denominator inside the square roots of a and a' to be $2/3$ instead of $1/2$, which is suboptimal but simplifies the resulting expressions) and in the second inequality we chose $\theta=3$. Therefore,

$$\inf_{\lambda\in\Lambda}\max\{f(\lambda),g(\lambda)\}\le 2\cdot\inf_{\lambda\in\Lambda}\frac{f(\lambda)+g(\lambda)}{2}=2\cdot\inf_{\lambda\in\Lambda}J_{1/2}(\lambda)\le\frac{20}{3}\left(\sqrt{M\left(N+\frac{1}{2}\right)}+\sqrt{N\left(M+\frac{1}{2}\right)}\right).$$

The corresponding a and a' achieving the above bound is

$$a=\sqrt{\frac{\frac{5}{3}M}{\frac{2}{3}+6N+\frac{2}{3}(N+4)}}=\sqrt{\frac{\frac{5}{3}M}{\frac{20}{3}N+\frac{10}{3}}}=\frac{1}{2}\sqrt{\frac{M}{N+\frac{1}{2}}},\quad a'=\frac{1}{2}\sqrt{\frac{N}{M+\frac{1}{2}}},$$

and thus corresponding $\eta,\eta'>0$ are given by

$$\begin{aligned}
\eta&=\frac{a}{1+a(a'+s)}=\frac{a}{1+a(a'+\theta a')}=\frac{\frac{1}{2}\sqrt{\frac{M}{N+1/2}}}{1+\sqrt{\frac{MN}{(M+1/2)(N+1/2)}}}=\frac{1}{2}\frac{\sqrt{M(M+1/2)}}{\sqrt{(M+1/2)(N+1/2)}+\sqrt{MN}},\\
\eta'&=\frac{1}{2}\frac{\sqrt{N(N+1/2)}}{\sqrt{(M+1/2)(N+1/2)}+\sqrt{MN}}.
\end{aligned}$$

This completes the proof. □

Cardinality-unaware case We next provide the proof of Theorem 10, which is for the cardinality-unaware case.

Proof of Theorem 10. Recall the rewritten forms of f and g in (26):

$$\begin{aligned}
f(a,a',s,s')&=\left(\frac{M}{a}+a'M+\frac{a}{2}+aN\right)+sM+\frac{a}{s'}(h(a,a')+sM),\\
g(a,a',s,s')&=\left(\frac{N}{a'}+aN+\frac{a'}{2}+a'M\right)+s'N+\frac{a'}{s}(h(a,a')+s'N),
\end{aligned}$$

where we recall $h(a,a')=\frac{M}{a}+a'M+\frac{N}{a'}+aN+\frac{a}{2}+\frac{a'}{2}$ as defined in (25).

As in the analysis of Theorem 6, we minimize the worst-case coefficients of f and g with respect to M and N. In particular, we define the coefficients of f and g with respect to M and N as follows:

$$\begin{aligned}
\kappa_M^f&=\frac{1}{a}+a'+s+\frac{a}{s'}\left(\frac{1}{a}+a'+s\right)=\left(1+\frac{a}{s'}\right)\left(\frac{1}{a}+a'+s\right),\quad \kappa_N^f=a+\frac{a}{s'}\left(\frac{1}{a'}+a\right)=\frac{a}{s'}\left(\frac{1}{a'}+a+s'\right),\\
\kappa_M^g&=a'+\frac{a'}{s}\left(\frac{1}{a}+a'\right)=\frac{a'}{s}\left(\frac{1}{a}+a'+s\right),\quad \kappa_N^g=\frac{1}{a'}+a+s'+\frac{a'}{s}\left(\frac{1}{a'}+a+s'\right)=\left(1+\frac{a'}{s}\right)\left(\frac{1}{a'}+a+s'\right).
\end{aligned}$$

Then, we will find (a, a', s, s') by considering the following optimization problem:

$$\min_{(a,a',s,s')\colon\, a>0,a'>0,s>0,s'>0} \max\left\{\kappa_M^f(a,a',s,s'),\, \kappa_N^f(a,a',s,s'),\, \kappa_M^g(a,a',s,s'),\, \kappa_N^g(a,a',s,s')\right\}, \tag{29}$$

where we note that since κ_M^f, κ_N^f and κ_M^g, κ_N^g contain s' and s in their denominators, respectively, it suffices to consider the case of $s > 0$ and $s' > 0$.

It suffices to consider the case of $a = a'$, $s = s'$ to find the optimal (a, a', s, s') in (29). This can be verified from the fact that, under the change of variables $p = \log a$, $p' = \log a'$, $q = \log s$, and $q' = \log s'$, the functions $\kappa_M^f, \kappa_N^f, \kappa_M^g, \kappa_N^g$ are convex in (p, p', q, q') (as discussed in detail in Appendix B.4.4), together with the invariance of the objective in (29) under the swaps $a \leftrightarrow a'$ and $s \leftrightarrow s'$. When $a = a'$ and $s = s'$, we have

$$\kappa_M^f = \kappa_N^g = \left(1 + \frac{a}{s}\right)\left(\frac{1}{a} + a + s\right) =: \kappa_1\,, \quad \kappa_N^f = \kappa_M^g = \frac{a}{s}\left(\frac{1}{a} + a + s\right) =: \kappa_2\,.$$

From this we have $\kappa_1 \geq \kappa_2$, and thus

$$\max\left\{\kappa_M^f, \kappa_N^f, \kappa_M^g, \kappa_N^g\right\} = \kappa_1\,.$$

The optimal value of (a, s) that minimizes $\kappa_1(a, s)$ is $(a, s) = (1/\sqrt{3}, 2/\sqrt{3})$, and at that point we have $\kappa_1(a, s) = (3/2) \cdot (\sqrt{3} + 1/\sqrt{3} + 2/\sqrt{3}) = 3\sqrt{3}$. Consequently, the optimal argument of the optimization problem in (29) is given by $(a, a', s, s') = (1/\sqrt{3}, 1/\sqrt{3}, 2/\sqrt{3}, 2/\sqrt{3})$. Therefore, from (24), the corresponding (η, η', c, c') equals $\left(\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{1}{2}, \frac{1}{2}\right)$. This completes the proof. □

B.4.4 Convex Reformulation and Numerical Learning-Rate Computation

As discussed in Lemma 8, in the cardinality-aware setting, it is difficult to obtain closed-form learning rates $\eta, \eta' > 0$ that minimize the maximum of the individual regret. In what follows, we discuss a convex reformulation of f and g (for possibly given $M = \log m$ and $N = \log n$) and numerical methods for computing $\eta, \eta' > 0$ that minimize either the maximum of the individual regrets or the convex sum of individual regrets J_γ in (5).

Recall that a function $f\colon \mathbb{R}^d \to \mathbb{R}$ with $\operatorname{dom} f = \mathbb{R}^d_{>0}$ is called a monomial function if there exists $c > 0$ and $a_i \in \mathbb{R}$ such that $f(z) = c z_1^{a_1} z_2^{a_2} \dots z_d^{a_d}$, and a function is a posynomial if it can be written as a sum of monomials (Boyd and Vandenberghe, 2004, Section 4.5). An important observation is that f and g in (26) are posynomials over $(a, a', s, s') \in \mathbb{R}^4_p$. As noted in the main text, optimizing f and g is in general nonconvex in the original variables, but by performing the change of variables described below, one can convert the problem into a convex optimization problem and solve it efficiently (see *e.g.,* Boyd and Vandenberghe 2004, Section 4.5.3) (As before, after the change of variables we will reuse the same symbols for the transformed functions by abuse of notation). We use the change of the variables given by

$$p = \log a\,, \quad p' = \log a'\,, \quad q = \log s\,, \quad q' = \log s'\,.$$

Then, we have

$$h(p, p') = M\mathrm{e}^{-p} + M\mathrm{e}^{p'} + N\mathrm{e}^{-p'} + N\mathrm{e}^{p} + \frac{1}{2}\mathrm{e}^{p} + \frac{1}{2}\mathrm{e}^{p'}\,,$$

and thus we can rewrite f and g in (26) as

$$\begin{aligned} f(p, p', q, q') = {} & M\mathrm{e}^{-p} + M\mathrm{e}^{p'} + \left(N + \frac{1}{2}\right)\mathrm{e}^{p} + M\mathrm{e}^{q} \\ & + M\mathrm{e}^{-q'} + \left(M + \frac{1}{2}\right)\mathrm{e}^{p+p'-q'} + N\mathrm{e}^{p-p'-q'} + \left(N + \frac{1}{2}\right)\mathrm{e}^{2p-q'} + M\mathrm{e}^{p+q-q'}\,, \end{aligned}$$

and

$$\begin{aligned} g(p, p', q, q') = {} & N\mathrm{e}^{-p'} + N\mathrm{e}^{p} + \left(M + \frac{1}{2}\right)\mathrm{e}^{p'} + N\mathrm{e}^{q'} \\ & + N\mathrm{e}^{-q} + \left(M + \frac{1}{2}\right)\mathrm{e}^{2p'-q} + \left(N + \frac{1}{2}\right)\mathrm{e}^{p+p'-q} + M\mathrm{e}^{p'-p-q} + N\mathrm{e}^{p'+q'-q}\,. \end{aligned}$$

Since f and g are convex in (p, p', q, q'), both their pointwise maximum and any convex combination are also convex in (p, p', q, q'). Hence their (globally optimal) solutions can be computed efficiently. In the cardinality-aware case, several

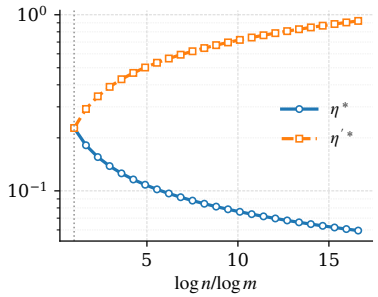

Figure 2: Optimal learning rates under cardinality asymmetry for the weighted-sum objective with $\gamma = 1/2$. We fix $m = 2$ and vary n from 2 to 10^5. The horizontal axis is $\log n/\log m$, and the vertical axis (in log scale) shows the numerically computed optimal learning rates η^* and η'^* for $J_{1/2}$.

propositions in the main text selected slightly suboptimal learning rates chosen to obtain closed-form expressions. However, if one is allowed to solve the above convex optimization, one can numerically obtain learning rates that further minimize the maximum of the individual regrets or the convex sum of individual regrets. We used this convexification to do numerical experiments.

B.5 Numerical Illustration on Optimal Learning Rates under Cardinality Asymmetry

In addition to Figure 1, which illustrates the tradeoff induced by varying γ for the fixed symmetric case $(m, n) = (10^2, 10^2)$, we provide a numerical illustration for the fixed choice $\gamma = 1/2$. Specifically, we examine how action-set cardinality asymmetry is reflected in the optimal learning rates for the weighted-sum objective $J_{1/2}$. We fix $m = 2$ and vary n from 2 to 10^5. As a measure of asymmetry, we use $\log n/\log m$ on the horizontal axis. For each pair (m, n), we numerically compute the corresponding optimal learning rates η^* and η'^* for the weighted-sum objective $J_{1/2}$, and plot them on the vertical axis. The vertical axis is shown in logarithmic scale so that the separation between η^* and η'^* can be seen more clearly even when their magnitudes differ substantially.

The result is shown in Figure 2. At the symmetric point $\log n/\log m = 1$, which corresponds to $n = m = 2$, the two optimal learning rates coincide, as expected from symmetry. As n increases, however, the gap between η^* and η'^* gradually widens. Thus, even under the fixed choice $\gamma = 1/2$, the figure numerically indicates that action-set cardinality asymmetry is reflected in an asymmetric allocation of the optimal learning rates. Because the vertical axis is on a log scale, this separation is also easy to observe on a multiplicative scale.

This observation suggests that, in the $\gamma = 1/2$ setting discussed in the main text, action-set cardinality asymmetry is reflected directly in the optimal choice of learning rates. In particular, when the two players' action-set cardinalities are highly asymmetric, a symmetric choice of learning rates is generally not numerically optimal.

C OMITTED DETAILS FROM SECTION 4

Here we provide the deferred details from Section 4.

Proof of Lemma 12. We first show that the function $f(a) = \log(1 + z\mathrm{e}^{-a})$ is convex with respect to a for $z > 0$. This can be confirmed from the fact that

$$f'(a) = \frac{-z\mathrm{e}^{-a}}{1 + z\mathrm{e}^{-a}}, \quad f''(a) = \frac{z\mathrm{e}^{-a}}{(1 + z\mathrm{e}^{-a})^2} > 0.$$

Hence, from the convexity of f, we have $f(a) - f(0) \geq f'(0) \cdot a$, which is equivalent to

$$\log(1 + z\mathrm{e}^{-a}) - \log(1 + z) \geq \frac{-z}{1 + z} \cdot a.$$

Rearranging the last inequality completes the proof. □

D OMITTED DETAILS FROM SECTION 5

This section provides the proof of Theorem 15.

Proof of Theorem 15. We consider the game with the payoff matrix A in (9). Then, noting that the optimal strategy for each player is to keep choosing only action 1 (that is to play the pure strategy e_1), and following the analysis used in the proof of Theorem 11, we can evaluate the dynamic regret as

$$\mathsf{DReg}_T^x = \sum_{t=1}^{T} \Delta_x (1 - x_t(1)) \,, \tag{30}$$

where we note that x_t denotes the time-averaged strategy of the optimistic Hedge up to round t as given by (15).

Following the analysis used in the proof of Theorem 11 again, for each $t \in [T]$ we have

$$\widehat{x}_t(1) = \frac{1}{1+\alpha_t} \,, \quad \alpha_t = (m-1)\exp(-\eta\Delta_x t) \,. \tag{31}$$

Combining (30) and (31), we can lower bound the dynamic regret of the x-player as

$$\mathsf{DReg}_T^x = \Delta_x \sum_{t=1}^{T} \left(1 - \frac{1}{t} \sum_{s=1}^{t} \frac{1}{1+\alpha_s} \right) = \Delta_x \sum_{t=1}^{T} \frac{1}{t} \sum_{s=1}^{t} \frac{\alpha_s}{1+\alpha_s} = \Delta_x \sum_{s=1}^{T} \frac{\alpha_s}{1+\alpha_s} \sum_{t=s}^{T} \frac{1}{t} \,, \tag{32}$$

where in the last equality, we exchanged the order of summation. For any $S_0 \in [T]$, the last quantity is lower bounded as

$$\begin{aligned}
\Delta_x \sum_{s=1}^{T} \frac{\alpha_s}{1+\alpha_s} \sum_{t=s}^{T} \frac{1}{t} &\geq \Delta_x \sum_{s=1}^{S_0} \frac{\alpha_s}{1+\alpha_s} \sum_{t=s}^{T} \frac{1}{t} \\
&\geq \Delta_x \sum_{s=1}^{S_0} \frac{\alpha_s}{1+\alpha_s} \log\left(\frac{T+1}{s} \right) \\
&\geq \Delta_x \log\left(\frac{T+1}{S_0} \right) \sum_{s=1}^{S_0} \frac{\alpha_s}{1+\alpha_s} \,,
\end{aligned} \tag{33}$$

where the second inequality follows from $\sum_{t=s}^{T} 1/t \geq \int_{t=s}^{T+1} (1/z)\mathrm{d}z = \log((T+1)/s)$ for any $s \in [T]$.

From Lemma 12 with $z = \alpha_s$ and $a = \eta\Delta_x$, we have

$$\frac{\alpha_s}{1+\alpha_s} \geq \frac{1}{\eta\Delta_x} (\log(1+\alpha_s) - \log(1+\alpha_{s+1})) \,,$$

and thus

$$\begin{aligned}
\sum_{s=1}^{S_0} \frac{\alpha_s}{1+\alpha_s} &\geq \frac{1}{\eta\Delta_x} (\log(1+\alpha_1) - \log(1+\alpha_{S_0+1})) \\
&\geq \frac{1}{\eta\Delta_x} (\log m - \eta\Delta_x - (m-1)\exp(-\eta\Delta_x(S_0+1))) \,,
\end{aligned} \tag{34}$$

where in the last inequality we used $\log(1+\alpha_1) = \log(1+(m-1)\exp(-\eta\Delta_x)) \geq \log(m\exp(-\eta\Delta_x))$ and $\log(1+z) \leq z$ for $z \geq 0$.

Combining (32) to (34), we can lower bound the dynamic regret of the x-player as

$$\mathsf{DReg}_T^x \geq \frac{1}{\eta} \log\left(\frac{T+1}{S_0} \right) (\log m - \eta\Delta_x - (m-1)\exp(-\eta\Delta_x(S_0+1))) \,.$$

Since $S_0 \in [T]$ is arbitrary, choosing $S_0 = \lfloor \sqrt{T+1} \rfloor$ gives

$$\mathsf{DReg}_T^x \geq \frac{1}{2\eta} \log(T+1) \big(\log m - \eta\Delta_x - (m-1)\exp(-\eta\Delta_x(\lfloor \sqrt{T+1} \rfloor + 1)) \big) \,.$$

Finally, choosing

$$\Delta_x = \min\left\{1, \frac{\log((m-1)(\lfloor\sqrt{T+1}\rfloor+1))}{\eta(\lfloor\sqrt{T+1}\rfloor+1)}\right\}$$

in the last inequality as in the proof of Theorem 11, we obtain the desired lower bound for the x-player. The dynamic regret of the y-player can be lower bounded by the same argument, and we have completed the proof. □

E LOWER BOUNDS FOR OPTIMISTIC FTRL WITH OTHER REGULARIZERS

In this section, we extend the lower-bound proof based on the game instance (9) to OFTRL with other separable regularizers. We first derive a common exact dual-gap identity, and then apply it to the log-barrier regularizer and the negative α-Tsallis entropy.

Consider OFTRL on the simplex Δ_m with a separable regularizer and a constant learning rate $\eta > 0$ given by

$$x_t \in \arg\max_{x\in\Delta_m}\left\{\left\langle x, \sum_{s=1}^{t-1} g_s + g_{t-1}\right\rangle - \frac{1}{\eta}\psi(x)\right\}, \quad \psi(x) = \sum_{i=1}^{m}\phi(x(i)), \tag{35}$$

and $g_0 = \mathbf{0}$.

We analyze the same hard instance as in (9). Recall that, for the x-player, the gain gap between action 1 and any action $k \neq 1$ is constant: $g_t(1) - g_t(k) = \Delta_x$ for all $t \geq 1$, and $k \neq 1$.

Lemma 18. *Suppose that the OFTRL iterate x_t lies in the relative interior of Δ_m for any $t \geq 1$. Then, for any $t \geq 2$, $k \neq 1$, and the game instance with a matrix of the form (9), it holds that*

$$\phi'(x_t(1)) - \phi'(x_t(k)) = \phi'(r_t) - \phi'(s_t) = \eta t \Delta_x$$

for

$$x_t(1) =: r_t, \qquad x_t(k) =: s_t = \frac{1-r_t}{m-1} \qquad \text{for } k \neq 1\,.$$

Proof. Since x_t is in the relative interior of Δ_m, the KKT conditions imply that, for any $t \geq 2$ and $i \in [m]$, there exists $\lambda_t \in \mathbb{R}$ such that

$$\phi'(x_t(i)) = \eta\left(\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\right) + \lambda_t\,.$$

Subtracting the equations for actions 1 and $k \neq 1$, we obtain

$$\phi'(x_t(1)) - \phi'(x_t(k)) = \eta\left(\sum_{s=1}^{t-1}(g_s(1) - g_s(k)) + g_{t-1}(1) - g_{t-1}(k)\right) = \eta t \Delta_x\,,$$

where in the last equality we used $g_s(1) - g_s(k) = \Delta_x$ for $s \geq 1$ and $k \neq 1$. From the symmetry of the game instance and the uniform initialization, it holds that all suboptimal actions have the same probability at each round, and thus we can write

$$x_t(1) =: r_t\,, \qquad x_t(k) =: s_t = \frac{1-r_t}{m-1} \qquad \text{for all } k \neq 1\,,$$

which completes the proof. □

Using Lemma 18, we can prove the following lower bound for OFTRL with the log-barrier regularizer.

Theorem 19. *Suppose that $m, n \geq 2$ and that the x- and y-players use OFTRL with the log-barrier regularizer $\psi(z) = -\sum_i \log z(i)$ with learning rates $\eta, \eta' > 0$, respectively. Then, there exists an instance of learning in two-player zero-sum games such that the regret of each player is lower bounded as*

$$\mathsf{Reg}_T^x \geq 1 - \frac{1}{m} + \frac{m-1}{\eta}\log\left(\frac{m+\eta(T+1)}{m+2\eta}\right)$$

and

$$\mathsf{Reg}_T^y \geq 1 - \frac{1}{n} + \frac{n-1}{\eta'}\log\left(\frac{n+\eta'(T+1)}{n+2\eta'}\right).$$

Proof. We consider the game instance in (9) with $\Delta_x = \Delta_y = 1$. From Lemma 18, for any $t \geq 2$ we have

$$\frac{m-1}{1-r_t} - \frac{1}{r_t} = \eta t\,.$$

Define

$$F(r) \coloneqq \frac{m-1}{1-r} - \frac{1}{r}\,, \qquad r \in (0,1)\,.$$

Since $F'(r) = \frac{m-1}{(1-r)^2} + \frac{1}{r^2} > 0$, the function F is strictly increasing. We also have $F(1/m) = 0$. Therefore, for $t \geq 2$, we have $F(r_t) = \eta t > 0$, and thus it holds that $r_t \geq 1/m$. Using this inequality, we obtain

$$\frac{m-1}{1-r_t} = \eta t + \frac{1}{r_t} \leq \eta t + m\,,$$

which implies

$$1 - r_t \geq \frac{m-1}{m+\eta t}$$

for all $t \geq 2$.

Therefore, using the definition of the regret and the fact that $r_1 = 1/m$ (recall $g_0 = \mathbf{0}$), and the last inequality, we obtain

$$\begin{aligned}
\mathsf{Reg}_T^x &= \sum_{t=1}^{T}(1-r_t) = 1 - \frac{1}{m} + \sum_{t=2}^{T}(1-r_t) \\
&\geq 1 - \frac{1}{m} + (m-1)\sum_{t=2}^{T}\frac{1}{m+\eta t} \\
&\geq 1 - \frac{1}{m} + \frac{m-1}{\eta}\log\left(\frac{m+\eta(T+1)}{m+2\eta}\right),
\end{aligned}$$

where in the last inequality we used

$$\sum_{t=2}^{T}\frac{1}{m+\eta t} \geq \int_2^{T+1}\frac{1}{m+\eta z}\mathrm{d}z = \frac{1}{\eta}\log\left(\frac{m+\eta(T+1)}{m+2\eta}\right).$$

The regret of the y-player can be lower bounded by the same argument, and we complete the proof. □

Using Lemma 18, we can prove the following lower bound for OFTRL with the negative Tsallis entropy.

Theorem 20. *Suppose that $m, n \geq 3$ and that the x- and y-players use OFTRL with the negative α-Tsallis entropy regularizer $\psi(z) = -\frac{1}{\alpha(1-\alpha)}\sum_i (z(i)^\alpha - z(i))$ for some $\alpha \in (0,1)$ with learning rates $\eta, \eta' > 0$, respectively. Then, there exists an instance of learning in two-player zero-sum games such that the regret of each player is lower bounded as*

$$\mathsf{Reg}_T^x \geq \frac{1}{2}\min\left\{T,\ \frac{(2(m-1))^{1-\alpha} - 2^{1-\alpha}}{(1-\alpha)\eta}\right\}$$

and

$$\mathsf{Reg}_T^y \geq \frac{1}{2}\min\left\{T,\ \frac{(2(n-1))^{1-\alpha} - 2^{1-\alpha}}{(1-\alpha)\eta'}\right\}.$$

Proof. Define

$$B_\alpha(m) \coloneqq (2(m-1))^{1-\alpha} - 2^{1-\alpha} > 0,$$

and consider the game instance in (9) with

$$\Delta_x \coloneqq \min\left\{1,\ \frac{B_\alpha(m)}{(1-\alpha)\eta T}\right\}.$$

For the negative α-Tsallis entropy, we have

$$\phi_\alpha(u) = -\frac{u^\alpha - u}{\alpha(1-\alpha)}, \qquad \phi'_\alpha(u) = \frac{1}{\alpha(1-\alpha)} - \frac{1}{1-\alpha}u^{\alpha-1}.$$

Hence, Lemma 18 for any $t \geq 2$,

$$\left(\frac{1-r_t}{m-1}\right)^{\alpha-1} - r_t^{\alpha-1} = (1-\alpha)\eta t \Delta_x \,.$$

For $r \in (0,1)$, define

$$F_\alpha(r) \coloneqq \left(\frac{1-r}{m-1}\right)^{\alpha-1} - r^{\alpha-1} \,.$$

Since we have

$$F'_\alpha(r) = (1-\alpha)\left(\frac{1}{m-1}\left(\frac{1-r}{m-1}\right)^{\alpha-2} + r^{\alpha-2}\right) > 0 \,,$$

the function F_α is strictly increasing on $(0,1)$. We also have $F_\alpha(1/2) = B_\alpha(m)$, and

$$F_\alpha(r_t) = (1-\alpha)\eta t \Delta_x \leq (1-\alpha)\eta T \Delta_x \leq B_\alpha(m) = F_\alpha(1/2)$$

for all $t \in [T]$. Hence, using this equality with the monotonicity of F_α and $r_1 = 1/m \leq 1/2$, we have $r_t \leq 1/2$ for $t \in [T]$.

Therefore, the regret of the x-player is lower bounded as

$$\mathsf{Reg}_T^x = \sum_{t=1}^T \Delta_x(1-r_t) \geq \sum_{t=1}^T \frac{\Delta_x}{2} = \frac{T\Delta_x}{2} \geq \frac{1}{2}\min\left\{T, \frac{B_\alpha(m)}{(1-\alpha)\eta}\right\} = \frac{1}{2}\min\left\{T, \frac{(2(m-1))^{1-\alpha} - 2^{1-\alpha}}{(1-\alpha)\eta}\right\},$$

where we used the choice of Δ_x. For the y-player, we define Δ_y analogously with m, η replaced by n, η', and applying the same calculation on the same payoff matrix, we can obtain the desired regret lower bound for the y-player. This completes the proof. □

F EXTENSION OF UPPER BOUNDS TO CONVEX–CONCAVE GAMES

In this section, we show that the argument used to derive the improved regret upper bounds in Section 3 can be extended to general Hölder-smooth convex–concave games. We consider a static convex–concave saddle-point problem of the form

$$\min_{x \in X} \max_{y \in Y} f(x,y) \,, \tag{36}$$

where $X \subset \mathbb{R}^{d_x}$ and $Y \subset \mathbb{R}^{d_y}$ are nonempty, closed, bounded convex sets, and $f\colon X \times Y \to \mathbb{R}$ is convex in x for each fixed y and concave in y for each fixed x. We equip X and Y with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, and denote by $\|\cdot\|_{X,*}$ and $\|\cdot\|_{Y,*}$ their dual norms. We assume that f is block-wise Hölder smooth with respect to these norms, that is, there exist nonnegative constants $\nu_{xx}, \nu_{xy}, \nu_{yx}, \nu_{yy} \in (0,1]$ and $L_{xx}, L_{xy}, L_{yx}, L_{yy} \geq 0$ such that for any $x, x' \in X$ and any $y, y' \in Y$,

$$\begin{aligned} \|\nabla_x f(x,y) - \nabla_x f(x',y)\|_{X,*} &\leq L_{xx}\|x-x'\|_X^{\nu_x^x} \,, \quad \|\nabla_x f(x,y) - \nabla_x f(x,y')\|_{X,*} \leq L_{xy}\|y-y'\|_Y^{\nu_x^y} \,, \\ \|\nabla_y f(x,y) - \nabla_y f(x',y)\|_{Y,*} &\leq L_{yx}\|x-x'\|_X^{\nu_y^x} \,, \quad \|\nabla_y f(x,y) - \nabla_y f(x,y')\|_{Y,*} \leq L_{yy}\|y-y'\|_Y^{\nu_y^y} \,. \end{aligned} \tag{37}$$

In particular, L_{xx} and L_{yy} quantify the *self-smoothness* of f in x and y, respectively, while L_{xy} and L_{yx} quantify the *cross-smoothness*.

We study (36) through a repeated-game interpretation. At each round $t = 1, 2, \ldots, T$: the x-player chooses $x_t \in X$ and y-player chooses $y_t \in Y$ simultaneously, and then the x-player observes gradient $g_t \coloneqq \nabla_x f(x_t, y_t)$, while the y-player observes $\ell_t \coloneqq \nabla_y f(x_t, y_t)$. We define the regrets of the two players as in the case of learning in two-player zero-sum games:

$$\mathsf{Reg}_T^x = \sum_{t=1}^T f(x_t, y_t) - \min_{x \in X} \sum_{t=1}^T f(x, y_t) \,, \quad \mathsf{Reg}_T^y = \max_{y \in Y} \sum_{t=1}^T f(x_t, y) - \sum_{t=1}^T f(x_t, y_t) \,.$$

Given any $(x, y) \in X \times Y$, the (primal–dual) gap is defined as

$$\mathsf{Gap}(x, y) = \sup_{y' \in Y} f(x, y') - \inf_{x' \in X} f(x', y)\,.$$

If (x^*, y^*) is a saddle point, then $\mathsf{Gap}(x^*, y^*) = 0$ and $f(x^*, y) \le f(x^*, y^*) \le f(x, y^*)$ for all $(x, y) \in X \times Y$.

As in the case of the two-player zero-sum games, it is known that the averaged iterates of online learning algorithms $\bar{x}_T = \frac{1}{T}\sum_{t=1}^T x_t$ and $\bar{y}_T = \frac{1}{T}\sum_{t=1}^T y_t$, satisfy

$$\mathsf{Gap}(\bar{x}_T, \bar{y}_T) \le \frac{\mathsf{Reg}_T^x + \mathsf{Reg}_T^y}{T}\,.$$

Algorithms The x-player runs OFTRL on X with a regularizer that is β_x-strongly convex with respect to $\|\cdot\|_X$ and learning rate $\eta > 0$, and the y-player runs OFTRL on Y with a regularizer that is β_y-strongly convex with respect to $\|\cdot\|_Y$ and learning rate $\eta' > 0$ (see Lemma 23 for the general regret bound):

$$x_t \in \arg\min_{x \in X}\left\{\left\langle x, \nabla_x f(x_{t-1}, y_{t-1}) + \sum_{s=1}^{t-1} \nabla_x f(x_s, y_s)\right\rangle + \frac{1}{\eta}\psi_X(x)\right\},$$
$$y_t \in \arg\max_{y \in Y}\left\{\left\langle y, \nabla_y f(x_{t-1}, y_{t-1}) + \sum_{s=1}^{t-1} \nabla_y f(x_s, y_s)\right\rangle - \frac{1}{\eta'}\psi_Y(y)\right\}.$$

We define B_x and B_y by $B_x = \sup_{x,x' \in X}(\psi_X(x) - \psi_X(x'))$ and $B_y = \sup_{y,y' \in Y}(\psi_Y(y) - \psi_Y(y'))$, and assume that $B_x, B_y < \infty$.

Theorem 21. *Consider convex–concave smooth games with Hölder smoothness in (37). Let $\widetilde{B}_x = B_x/\beta_x$, $\widetilde{B}_y = B_y/\beta_y$. Let the averaged iterates be $\bar{x}_T = \frac{1}{T}\sum_{t=1}^T x_t$, and $\bar{y}_T = \frac{1}{T}\sum_{t=1}^T y_t$. Then, the above algorithm with an appropriate choice of $\eta, \eta' > 0$ guarantees that the social regret is upper bounded as*

$$\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \le C_{xx}\, T^{\frac{1-\nu_{xx}}{2}} + C_{yy}\, T^{\frac{1-\nu_{yy}}{2}} + C_{xy}\, T^{\frac{1-\nu_{xy}}{2}} + C_{yx}\, T^{\frac{1-\nu_{yx}}{2}}\,,$$

where

$$C_{xx} = c_{\mathrm{reg}}(\nu_{xx})\, 4^{\frac{1+\nu_{xx}}{2}} L_{xx} \widetilde{B}_x^{\frac{1+\nu_{xx}}{2}}\,, \quad C_{yy} = c_{\mathrm{reg}}(\nu_{yy})\, 4^{\frac{1+\nu_{yy}}{2}} L_{yy} \widetilde{B}_y^{\frac{1+\nu_{yy}}{2}}\,,$$
$$C_{xy} = c_{\mathrm{reg}}(\nu_{xy})\, 8^{\frac{1+\nu_{xy}}{2}} L_{xy} \widetilde{B}_x^{\frac{1}{2}} \widetilde{B}_y^{\frac{\nu_{xy}}{2}}\,, \quad C_{yx} = c_{\mathrm{reg}}(\nu_{yx})\, 8^{\frac{1+\nu_{yx}}{2}} L_{yx} \widetilde{B}_y^{\frac{1}{2}} \widetilde{B}_x^{\frac{\nu_{yx}}{2}}\,,$$

where $c_{\mathrm{reg}}(\nu) = \frac{2}{1+\nu}((1-\nu)/(1+\nu))^{\frac{\nu-1}{2}}$ for $\nu \in (0,1)$ and $c_{\mathrm{reg}}(1) = 1$. Consequently, the duality gap at $(\bar{x}_T, \bar{y}_T)$ is upper bounded as

$$\mathsf{Gap}(\bar{x}_T, \bar{y}_T) = \sup_{y \in Y} f(\bar{x}_T, y) - \inf_{x \in X} f(x, \bar{y}_T)$$
$$\lesssim L_{xx} \widetilde{B}_x^{\frac{1+\nu_{xx}}{2}} T^{-\frac{1+\nu_{xx}}{2}} + L_{yy} \widetilde{B}_y^{\frac{1+\nu_{yy}}{2}} T^{-\frac{1+\nu_{yy}}{2}} + L_{xy} \widetilde{B}_x^{\frac{1}{2}} \widetilde{B}_y^{\frac{\nu_{xy}}{2}} T^{-\frac{1+\nu_{xy}}{2}} + L_{yx} \widetilde{B}_x^{\frac{\nu_{yx}}{2}} \widetilde{B}_y^{\frac{1}{2}} T^{-\frac{1+\nu_{yx}}{2}}\,.$$

Note that β_x and β_y denote the strong convexity parameters of the regularizers used in OFTRL, not the strong convexity parameters of the convex–concave game.

Corollary 22. *Consider smooth convex–concave games. Let the averaged iterates be $\bar{x}_T = \frac{1}{T}\sum_{t=1}^T x_t$, and $\bar{y}_T = \frac{1}{T}\sum_{t=1}^T y_t$. Then, the duality gap at $(\bar{x}_T, \bar{y}_T)$ is upper bounded as*

$$\mathsf{Gap}(\bar{x}_T, \bar{y}_T) = \sup_{y \in Y} f(\bar{x}_T, y) - \inf_{x \in X} f(x, \bar{y}_T)$$
$$\le \frac{4L_{xx}\, B_x/\beta_x + 4L_{yy}\, B_y/\beta_y}{T} + \frac{8\max\{L_{xy}, L_{yx}\}\sqrt{B_x B_y/(\beta_x\beta_y)}}{T}\,.$$

Before proving the theorem, we prepare several lemmas.

Lemma 23. *Let $W \subseteq \mathbb{R}^d$ be a convex body and $\phi\colon \mathbb{R} \to \mathbb{R}$. Let $\psi_t(w) = \frac{1}{\eta_t}\sum_{i=1}^d \phi(w(i))$ be a regularizer with nonincreasing learning rate η_t and suppose that the function $\sum_{i=1}^d \phi(\cdot)$ is β-strongly-convex with respect to a norm $\|\cdot\|$. Let $w_t \in \arg\min_{w\in W}\{\langle w, m_t + \sum_{s=1}^{t-1} z_s\rangle + \psi_t(w)\}$. Then, for any $w^* \in W$ and $c_1, \dots, c_T > 0$, it holds that*

$$\sum_{t=1}^T \langle w_t - w^*, z_t\rangle \le \frac{B_\phi}{\eta_{T+1}} + \sum_{t=1}^T \frac{\eta_t}{2c_t\beta}\|z_t - m_t\|_*^2 - \sum_{t=1}^T \frac{(1-c_t)\beta}{2\eta_t}\|w_t - w_{t+1}\|^2\,,$$

where $B_\phi = \sup_{w\in W}\sum_{i=1}^d \phi(w)$.

This lemma generalizes Tsuchiya et al. (2025, Lemma 17). Choosing $\eta_t = \eta$, $c_t = c$ for all $t \in [T]$ and $m_{T+1} = \mathbf{0}$ yields Lemma 2.

Proof. We will upper bound the RHS of (17) in Lemma 16. From the definition of B_ϕ and the assumption that $(\eta_t)_t$ is nonincreasing, we have

$$\psi_{T+1}(w^*) - \psi_1(w_1) + \sum_{t=1}^T(\psi_t(w_{t+1}) - \psi_{t+1}(w_{t+1})) \le \frac{B_\phi}{\eta_1} + \sum_{t=1}^T\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)B_\phi = \frac{B_\phi}{\eta_{T+1}}\,.$$

Fix an arbitrary $c_t > 0$. Then, we also have

$$\begin{aligned}
&\langle w_t - w_{t+1}, z_t - m_t\rangle - D_{\psi_t}(w_{t+1}, w_t)\\
&= \langle w_t - w_{t+1}, z_t - m_t\rangle - \frac{1}{\eta_t}D_{\sum_{i=1}^d\phi(\cdot)}(w_{t+1}, w_t)\\
&\le \|w_t - w_{t+1}\|\|z_t - m_t\|_* - \frac{\beta}{2\eta_t}\|w_t - w_{t+1}\|^2\\
&= \|w_t - w_{t+1}\|\|z_t - m_t\|_* - \frac{c_t\beta}{2\eta_t}\|w_t - w_{t+1}\|^2 - \frac{(1-c_t)\beta}{2\eta_t}\|w_t - w_{t+1}\|^2\\
&\le \frac{\eta_t}{2c_t\beta}\|z_t - m_t\|_*^2 - \frac{(1-c_t)\beta}{2\eta_t}\|w_t - w_{t+1}\|^2\,,
\end{aligned}$$

where the first inequality follows from Hölder's inequality and the fact that the function $\sum_{i=1}^d \phi(\cdot)$ is β-strongly convex with respect to $\|\cdot\|$, and the last inequality follows by considering the worst-case with respect to $\|w_t - w_{t+1}\|$ in the first two terms. Combining Lemma 16 with the above two inequalities completes the proof. □

The following lemma is a direct consequence of Zhao et al. (2025, Lemma 1) for Hölder smooth functions.

Lemma 24. *Suppose that (37) holds. Fix parameters $\delta_{xx}, \delta_{xy}, \delta_{yx}, \delta_{yy} > 0$ and define*

$$\bar{L}_{xx} \coloneqq \delta_{xx}^{\frac{\nu_x^x - 1}{1+\nu_x^x}}(L_{xx})^{\frac{2}{1+\nu_x^x}}\,,\quad \bar{L}_{xy} \coloneqq \delta_{xy}^{\frac{\nu_x^y - 1}{1+\nu_x^y}}(L_{xy})^{\frac{2}{1+\nu_x^y}}\,,\quad \bar{L}_{yx} \coloneqq \delta_{yx}^{\frac{\nu_y^x - 1}{1+\nu_y^x}}(L_{yx})^{\frac{2}{1+\nu_y^x}}\,,\quad \bar{L}_{yy} \coloneqq \delta_{yy}^{\frac{\nu_y^y - 1}{1+\nu_y^y}}(L_{yy})^{\frac{2}{1+\nu_y^y}}\,.$$

Then, for all $x, x' \in X$ and $y, y' \in Y$,

$$\begin{aligned}
\|\nabla_x f(x,y) - \nabla_x f(x',y)\|_{X,*}^2 &\le (\bar{L}_{xx})^2\|x - x'\|_X^2 + 4\bar{L}_{xx}\delta_{xx}\,,\\
\|\nabla_x f(x,y) - \nabla_x f(x,y')\|_{X,*}^2 &\le (\bar{L}_{xy})^2\|y - y'\|_Y^2 + 4\bar{L}_{xy}\delta_{xy}\,,\\
\|\nabla_y f(x,y) - \nabla_y f(x',y)\|_{Y,*}^2 &\le (\bar{L}_{yx})^2\|x - x'\|_X^2 + 4\bar{L}_{yx}\delta_{yx}\,,\\
\|\nabla_y f(x,y) - \nabla_y f(x,y')\|_{Y,*}^2 &\le (\bar{L}_{yy})^2\|y - y'\|_Y^2 + 4\bar{L}_{yy}\delta_{yy}\,.
\end{aligned}$$

To optimize $\delta_{xx}, \delta_{xy}, \delta_{yx}, \delta_{yy}$, we will use the following lemma.

Lemma 25. *Let $\nu \in (0,1]$ and define $\kappa(\nu) \coloneqq \frac{\nu-1}{1+\nu} \in [-1, 0]$. For any $A, B > 0$ and any $T \ge 1$, consider $F(\delta) = A\delta^{\kappa(\nu)} + BT\delta$ for $\delta \ge 0$. Then there exists $\delta^* \ge 0$ such that*

$$F(\delta^*) \le c_{\mathrm{reg}}(\nu)\, A^{\frac{1+\nu}{2}} B^{\frac{1-\nu}{2}} T^{\frac{1-\nu}{2}}\,,\quad \textit{where}\quad c_{\mathrm{reg}}(\nu) = \begin{cases} \frac{2}{1+\nu}\left(\frac{1-\nu}{1+\nu}\right)^{\frac{\nu-1}{2}} & \textit{when } \nu \in (0,1)\,,\\ 1 & \textit{when } \nu = 1\,. \end{cases}$$

In particular, when $\nu \in (0,1)$, we have $\delta^* > 0$.

Proof. We have $F'(\delta) = A\kappa\,\delta^{\kappa-1} + BT$. When $\nu \in (0,1)$, we have $\kappa < 0$ and thus the unique critical point is $\delta^* = \left(\frac{A(1-\nu)}{B(1+\nu)}\right)^{\frac{1+\nu}{2}} T^{-\frac{1+\nu}{2}}$, which is positive. Combining this with the definition of F, we have $F(\delta^*) = c_{\mathrm{reg}}(\nu) A^{\frac{1+\nu}{2}} B^{\frac{1-\nu}{2}} T^{\frac{1-\nu}{2}}$. When $\nu = 1$, the inequality trivially holds. □

Now we are ready to prove Theorem 21.

Proof of Theorem 21. Using Lemma 23 and the analysis similar to Section 3, we can show that

$$\mathsf{Reg}_T^x \le \frac{B_x}{\eta} + \frac{\eta}{2c\beta_x}\sum_{t=1}^T \|g_t - g_{t-1}\|_{X,*}^2 - \frac{(1-c)\beta_x}{2\eta}\sum_{t=2}^T \|x_t - x_{t-1}\|_X^2\,,$$
$$\mathsf{Reg}_T^y \le \frac{B_y}{\eta'} + \frac{\eta'}{2c'\beta_y}\sum_{t=1}^T \|\ell_t - \ell_{t-1}\|_{Y,*}^2 - \frac{(1-c')\beta_y}{2\eta'}\sum_{t=2}^T \|y_t - y_{t-1}\|_Y^2\,.$$

The second term is upper bounded as

$$\begin{aligned}\|g_t - g_{t-1}\|_{X,*}^2 &= \|\nabla_x f(x_t,y_t) - \nabla_x f(x_{t-1},y_{t-1})\|_{X,*}^2 \\ &\le 2\|\nabla_x f(x_t,y_t) - \nabla_x f(x_{t-1},y_t)\|_{X,*}^2 + 2\|\nabla_x f(x_{t-1},y_t) - \nabla_x f(x_{t-1},y_{t-1})\|_{X,*}^2 \\ &\le 2(\bar{L}_{xx})^2\|x_t - x_{t-1}\|_X^2 + 2(\bar{L}_{xy})^2\|y_t - y_{t-1}\|_Y^2 + 4(\bar{L}_{xx}\delta_{xx} + \bar{L}_{xy}\delta_{xy})\,.\end{aligned}$$

Similarly,

$$\|\ell_t - \ell_{t-1}\|_{Y,*}^2 \le 2(\bar{L}_{yy})^2\|y_t - y_{t-1}\|_Y^2 + 2(\bar{L}_{yx})^2\|x_t - x_{t-1}\|_X^2 + 4(\bar{L}_{yy}\delta_{yy} + \bar{L}_{yx}\delta_{yx})\,.$$

Therefore,

$$\begin{aligned}\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \le \frac{B_x}{\eta} + \frac{B_y}{\eta'} &+ \left(\frac{\eta(\bar{L}_{xx})^2}{c\beta_x} + \frac{\eta'(\bar{L}_{yx})^2}{c'\beta_y} - \frac{(1-c)\beta_x}{2\eta}\right)\sum_{t=1}^T \|x_t - x_{t-1}\|_X^2 \\ &+ \left(\frac{\eta'(\bar{L}_{yy})^2}{c'\beta_y} + \frac{\eta(\bar{L}_{xy})^2}{c\beta_x} - \frac{(1-c')\beta_y}{2\eta'}\right)\sum_{t=1}^T \|y_t - y_{t-1}\|_Y^2 \\ &+ 2T\left(\frac{\eta}{c\beta_x}(\bar{L}_{xx}\delta_{xx} + \bar{L}_{xy}\delta_{xy}) + \frac{\eta'}{c'\beta_y}(\bar{L}_{yy}\delta_{yy} + \bar{L}_{yx}\delta_{yx})\right).\end{aligned}$$

To simplify, define $\widetilde{B}_x = B_x/\beta_x$, $\widetilde{B}_y = B_y/\beta_y$, $\widetilde{\eta} = \eta/\beta_x$ and $\widetilde{\eta}' = \eta'/\beta_y$. Then, we have

$$\begin{aligned}\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \le \frac{\widetilde{B}_x}{\widetilde{\eta}} + \frac{\widetilde{B}_y}{\widetilde{\eta}'} &+ A_x\sum_{t=1}^T \|x_t - x_{t-1}\|_X^2 + A_y\sum_{t=1}^T \|y_t - y_{t-1}\|_Y^2 \\ &+ 2T\left(\frac{\eta}{c\beta_x}(\bar{L}_{xx}\delta_{xx} + \bar{L}_{xy}\delta_{xy}) + \frac{\eta'}{c'\beta_y}(\bar{L}_{yy}\delta_{yy} + \bar{L}_{yx}\delta_{yx})\right), \\ \text{where}\quad A_x = \frac{\widetilde{\eta}(\bar{L}_{xx})^2}{c} + \frac{\widetilde{\eta}'(\bar{L}_{yx})^2}{c'} - \frac{1-c}{2\widetilde{\eta}}\,,&\quad A_y = \frac{\widetilde{\eta}'(\bar{L}_{yy})^2}{c'} + \frac{\widetilde{\eta}(\bar{L}_{xy})^2}{c} - \frac{1-c'}{2\widetilde{\eta}'}\,.\end{aligned} \tag{38}$$

Now, a sufficient condition for $A_x \le 0$ and $A_y \le 0$ is

$$\frac{\widetilde{\eta}(\bar{L}_{xx})^2}{c} - \frac{1-c}{4\widetilde{\eta}} \le 0\,,\quad \frac{\widetilde{\eta}'(\bar{L}_{yx})^2}{c'} - \frac{1-c}{4\widetilde{\eta}} \le 0\,,\quad \frac{\widetilde{\eta}'(\bar{L}_{yy})^2}{c'} - \frac{1-c'}{4\widetilde{\eta}'} \le 0\,,\quad \frac{\widetilde{\eta}(\bar{L}_{xy})^2}{c} - \frac{1-c'}{4\widetilde{\eta}'} \le 0\,,$$

which is equivalent to

$$\widetilde{\eta}^2 \le \frac{c(1-c)}{4(\bar{L}_{xx})^2}\,,\quad \widetilde{\eta}'^2 \le \frac{c'(1-c')}{4(\bar{L}_{yy})^2}\,,\quad \widetilde{\eta}\widetilde{\eta}' \le \min\left\{\frac{(1-c)c'}{4(\bar{L}_{yx})^2}\,,\ \frac{(1-c')c}{4(\bar{L}_{xy})^2}\right\}. \tag{39}$$

Choose optimal learning rates and coefficients of negative terms In what follows, we will choose $\widetilde{\eta}$, $\widetilde{\eta}'$, c, c' satisfying (39). For simplicity, we let $c = c' = 1/2$. Then, these constraints can be written as

$$\widetilde{\eta}^2 \leq \frac{c(1-c)}{4(\bar{L}_{xx})^2} = \frac{1}{16(\bar{L}_{xx})^2}\,, \quad \widetilde{\eta}'^2 \leq \frac{c'(1-c')}{4(\bar{L}_{yy})^2} = \frac{1}{16(\bar{L}_{yy})^2}\,,$$
$$\widetilde{\eta}\widetilde{\eta}' \leq \min\left\{\frac{(1-c)c'}{4(\bar{L}_{yx})^2}\,, \frac{(1-c')c}{4(\bar{L}_{xy})^2}\right\} = \frac{1}{16}\min\left\{\frac{1}{(\bar{L}_{yx})^2}\,, \frac{1}{(\bar{L}_{xy})^2}\right\}.$$

Let $\bar{L}_{\mathrm{cross}} = \max\{\bar{L}_{xy}, \bar{L}_{yx}\}$. Then, choosing

$$\widetilde{\eta} = \min\left\{\frac{1}{4\bar{L}_{xx}}\,, \frac{1}{4\bar{L}_{\mathrm{cross}}}\sqrt{\frac{\widetilde{B}_x}{\widetilde{B}_y}}\right\}, \quad \widetilde{\eta}' = \min\left\{\frac{1}{4\bar{L}_{yy}}\,, \frac{1}{4\bar{L}_{\mathrm{cross}}}\sqrt{\frac{\widetilde{B}_y}{\widetilde{B}_x}}\right\} \tag{40}$$

satisfies all inequalities in (39), where we note that the third condition is satisfied since

$$\widetilde{\eta}\widetilde{\eta}' \leq \frac{1}{4\bar{L}_{\mathrm{cross}}}\sqrt{\frac{\widetilde{B}_x}{\widetilde{B}_y}} \cdot \frac{1}{4\bar{L}_{\mathrm{cross}}}\sqrt{\frac{\widetilde{B}_y}{\widetilde{B}_x}} = \frac{1}{16\bar{L}_{\mathrm{cross}}^2} = \frac{1}{16}\min\left\{\frac{1}{(\bar{L}_{yx})^2}\,, \frac{1}{(\bar{L}_{xy})^2}\right\}$$

Hence for these choices of (η, η', c, c'), we have $A_x \leq 0$ and $A_y \leq 0$, and thus continuing from (38) we have

$$\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \leq \frac{\widetilde{B}_x}{\widetilde{\eta}} + \frac{\widetilde{B}_y}{\widetilde{\eta}'} + 2T\left(\frac{\eta}{c\beta_x}(\bar{L}_{xx}\delta_{xx} + \bar{L}_{xy}\delta_{xy}) + \frac{\eta'}{c'\beta_y}(\bar{L}_{yy}\delta_{yy} + \bar{L}_{yx}\delta_{yx})\right).$$

Since

$$\frac{\widetilde{B}_x}{\widetilde{\eta}} \leq 4\bar{L}_{xx}\widetilde{B}_x + 4\bar{L}_{\mathrm{cross}}\sqrt{\widetilde{B}_x\widetilde{B}_y}\,, \quad \frac{\widetilde{B}_y}{\widetilde{\eta}'} \leq 4\bar{L}_{yy}\widetilde{B}_y + 4\bar{L}_{\mathrm{cross}}\sqrt{\widetilde{B}_x\widetilde{B}_y}\,,$$

we have

$$\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \leq 4\bar{L}_{xx}\,\widetilde{B}_x + 4\bar{L}_{yy}\,\widetilde{B}_y + 8\bar{L}_{\mathrm{cross}}\sqrt{\widetilde{B}_x\widetilde{B}_y} + 2T\left(\frac{\eta}{c\beta_x}(\bar{L}_{xx}\delta_{xx} + \bar{L}_{xy}\delta_{xy}) + \frac{\eta'}{c'\beta_y}(\bar{L}_{yy}\delta_{yy} + \bar{L}_{yx}\delta_{yx})\right), \tag{41}$$

where we recall $\widetilde{B}_x = \frac{B_x}{\beta_x}$, $\widetilde{B}_y = \frac{B_y}{\beta_y}$, and $\bar{L}_{\mathrm{cross}} = \max\{\bar{L}_{xy}, \bar{L}_{yx}\}$.

Recall that $c = c' = 1/2$ and $\widetilde{\eta} = \eta/\beta_x$, $\widetilde{\eta}' = \eta'/\beta_y$. Then

$$\frac{\eta}{c\beta_x} = \frac{2\eta}{\beta_x} = 2\widetilde{\eta}\,, \qquad \frac{\eta'}{c'\beta_y} = \frac{2\eta'}{\beta_y} = 2\widetilde{\eta}'\,, \tag{42}$$

and thus the last term in (41) becomes

$$2T\left(\frac{\eta}{c\beta_x}(\bar{L}_{xx}\delta_{xx} + \bar{L}_{xy}\delta_{xy}) + \frac{\eta'}{c'\beta_y}(\bar{L}_{yy}\delta_{yy} + \bar{L}_{yx}\delta_{yx})\right) = 4T\left(\widetilde{\eta}\bar{L}_{xx}\delta_{xx} + \widetilde{\eta}\bar{L}_{xy}\delta_{xy} + \widetilde{\eta}'\bar{L}_{yy}\delta_{yy} + \widetilde{\eta}'\bar{L}_{yx}\delta_{yx}\right).$$

By (40) and the definition of $\bar{L}_{\mathrm{cross}}$,

$$4\widetilde{\eta}\bar{L}_{xx} \leq 1\,, \quad 4\widetilde{\eta}'\bar{L}_{yy} \leq 1\,, \quad 4\widetilde{\eta}\bar{L}_{xy} \leq 4\widetilde{\eta}\bar{L}_{\mathrm{cross}} \leq \sqrt{\frac{\widetilde{B}_x}{\widetilde{B}_y}}\,, \quad 4\widetilde{\eta}'\bar{L}_{yx} \leq 4\widetilde{\eta}'\bar{L}_{\mathrm{cross}} \leq \sqrt{\frac{\widetilde{B}_y}{\widetilde{B}_x}}\,.$$

Hence,

$$4T\left(\widetilde{\eta}\bar{L}_{xx}\delta_{xx} + \widetilde{\eta}\bar{L}_{xy}\delta_{xy} + \widetilde{\eta}'\bar{L}_{yy}\delta_{yy} + \widetilde{\eta}'\bar{L}_{yx}\delta_{yx}\right) \leq T\delta_{xx} + T\delta_{yy} + T\sqrt{\frac{\widetilde{B}_x}{\widetilde{B}_y}}\,\delta_{xy} + T\sqrt{\frac{\widetilde{B}_y}{\widetilde{B}_x}}\,\delta_{yx}. \tag{43}$$

Combining (41) with (43) and using $\bar{L}_{\mathrm{cross}} = \max\{\bar{L}_{xy}, \bar{L}_{yx}\}$ yields

$$\begin{aligned}\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \leq\ & 4\bar{L}_{xx}\,\widetilde{B}_x + 4\bar{L}_{yy}\,\widetilde{B}_y + 8\bar{L}_{xy}\sqrt{\widetilde{B}_x\widetilde{B}_y} + 8\bar{L}_{yx}\sqrt{\widetilde{B}_x\widetilde{B}_y} \\ & + T\delta_{xx} + T\delta_{yy} + T\sqrt{\frac{\widetilde{B}_x}{\widetilde{B}_y}}\,\delta_{xy} + T\sqrt{\frac{\widetilde{B}_y}{\widetilde{B}_x}}\,\delta_{yx}\,.\end{aligned} \tag{44}$$

Optimize δ_{xx}, δ_{xy}, δ_{yx}, δ_{yy} Now we recall the definitions in Lemma 24:

$$\begin{aligned}
\bar{L}_{xx} &= L_{xx}^{\frac{2}{1+\nu_{xx}}}\,\delta_{xx}^{\kappa_{xx}}\,, & \kappa_{xx} &\coloneqq \frac{\nu_{xx}-1}{1+\nu_{xx}}\,,\\
\bar{L}_{yy} &= L_{yy}^{\frac{2}{1+\nu_{yy}}}\,\delta_{yy}^{\kappa_{yy}}\,, & \kappa_{yy} &\coloneqq \frac{\nu_{yy}-1}{1+\nu_{yy}}\,,\\
\bar{L}_{xy} &= L_{xy}^{\frac{2}{1+\nu_{xy}}}\,\delta_{xy}^{\kappa_{xy}}\,, & \kappa_{xy} &\coloneqq \frac{\nu_{xy}-1}{1+\nu_{xy}}\,,\\
\bar{L}_{yx} &= L_{yx}^{\frac{2}{1+\nu_{yx}}}\,\delta_{yx}^{\kappa_{yx}}\,, & \kappa_{yx} &\coloneqq \frac{\nu_{yx}-1}{1+\nu_{yx}}\,.
\end{aligned}$$

with $\nu_{xx},\nu_{yy},\nu_{xy},\nu_{yx}\in(0,1]$, so $\kappa_{xx},\kappa_{yy},\kappa_{xy},\kappa_{yx}\in[-1,0)$. Substituting these expressions into (44), we obtain

$$\mathsf{Reg}_T^x+\mathsf{Reg}_T^y\le F_{xx}(\delta_{xx})+F_{yy}(\delta_{yy})+F_{xy}(\delta_{xy})+F_{yx}(\delta_{yx})\,, \tag{45}$$

where

$$\begin{aligned}
F_{xx}(\delta_{xx}) &\coloneqq 4\widetilde{B}_x L_{xx}^{\frac{2}{1+\nu_{xx}}}\,\delta_{xx}^{\kappa_{xx}}+T\delta_{xx}\,,\\
F_{yy}(\delta_{yy}) &\coloneqq 4\widetilde{B}_y L_{yy}^{\frac{2}{1+\nu_{yy}}}\,\delta_{yy}^{\kappa_{yy}}+T\delta_{yy}\,,\\
F_{xy}(\delta_{xy}) &\coloneqq 8\sqrt{\widetilde{B}_x\widetilde{B}_y}L_{xy}^{\frac{2}{1+\nu_{xy}}}\,\delta_{xy}^{\kappa_{xy}}+T\sqrt{\frac{\widetilde{B}_x}{\widetilde{B}_y}}\,\delta_{xy}\,,\\
F_{yx}(\delta_{yx}) &\coloneqq 8\sqrt{\widetilde{B}_x\widetilde{B}_y}L_{yx}^{\frac{2}{1+\nu_{yx}}}\,\delta_{yx}^{\kappa_{yx}}+T\sqrt{\frac{\widetilde{B}_y}{\widetilde{B}_x}}\,\delta_{yx}\,.
\end{aligned}$$

The four quantities $\delta_{xx},\delta_{yy},\delta_{xy},\delta_{yx}$ are analysis parameters and can be optimized independently for each block.

We apply Lemma 25 to each $F_{\bullet\bullet}$ in (45). Define

$$\begin{aligned}
A_{xx} &\coloneqq 4\widetilde{B}_x L_{xx}^{\frac{2}{1+\nu_{xx}}}\,, & B_{xx} &\coloneqq 1\,,\\
A_{yy} &\coloneqq 4\widetilde{B}_y L_{yy}^{\frac{2}{1+\nu_{yy}}}\,, & B_{yy} &\coloneqq 1\,,\\
A_{xy} &\coloneqq 8\sqrt{\widetilde{B}_x\widetilde{B}_y}L_{xy}^{\frac{2}{1+\nu_{xy}}}\,, & B_{xy} &\coloneqq \sqrt{\frac{\widetilde{B}_x}{\widetilde{B}_y}}\,,\\
A_{yx} &\coloneqq 8\sqrt{\widetilde{B}_x\widetilde{B}_y}L_{yx}^{\frac{2}{1+\nu_{yx}}}\,, & B_{yx} &\coloneqq \sqrt{\frac{\widetilde{B}_y}{\widetilde{B}_x}}\,.
\end{aligned}$$

Then Lemma 25 implies that there exist choices $\delta_{xx}^*(T),\delta_{yy}^*(T),\delta_{xy}^*(T),\delta_{yx}^*(T)$ such that

$$\begin{aligned}
F_{xx}(\delta_{xx}^*(T)) &\le C_{xx}\,T^{\frac{1-\nu_{xx}}{2}}\,, & C_{xx} &\coloneqq c_{\mathrm{reg}}(\nu_{xx})\,A_{xx}^{\frac{1+\nu_{xx}}{2}}\,B_{xx}^{\frac{1-\nu_{xx}}{2}}\,,\\
F_{yy}(\delta_{yy}^*(T)) &\le C_{yy}\,T^{\frac{1-\nu_{yy}}{2}}\,, & C_{yy} &\coloneqq c_{\mathrm{reg}}(\nu_{yy})\,A_{yy}^{\frac{1+\nu_{yy}}{2}}\,B_{yy}^{\frac{1-\nu_{yy}}{2}}\,,\\
F_{xy}(\delta_{xy}^*(T)) &\le C_{xy}\,T^{\frac{1-\nu_{xy}}{2}}\,, & C_{xy} &\coloneqq c_{\mathrm{reg}}(\nu_{xy})\,A_{xy}^{\frac{1+\nu_{xy}}{2}}\,B_{xy}^{\frac{1-\nu_{xy}}{2}}\,,\\
F_{yx}(\delta_{yx}^*(T)) &\le C_{yx}\,T^{\frac{1-\nu_{yx}}{2}}\,, & C_{yx} &\coloneqq c_{\mathrm{reg}}(\nu_{yx})\,A_{yx}^{\frac{1+\nu_{yx}}{2}}\,B_{yx}^{\frac{1-\nu_{yx}}{2}}\,.
\end{aligned}$$

Expanding the expressions for $A_{\bullet\bullet}, B_{\bullet\bullet}$, we obtain

$$\begin{aligned}
C_{xx} &= c_{\mathrm{reg}}(\nu_{xx})4^{\frac{1+\nu_{xx}}{2}}\,L_{xx}\,\widetilde{B}_x^{\frac{1+\nu_{xx}}{2}}\,,\\
C_{yy} &= c_{\mathrm{reg}}(\nu_{yy})4^{\frac{1+\nu_{yy}}{2}}\,L_{yy}\,\widetilde{B}_y^{\frac{1+\nu_{yy}}{2}}\,,\\
C_{xy} &= c_{\mathrm{reg}}(\nu_{xy})8^{\frac{1+\nu_{xy}}{2}}\,L_{xy}\widetilde{B}_x^{\frac{1}{2}}\widetilde{B}_y^{\frac{\nu_{xy}}{2}}\,,\\
C_{yx} &= c_{\mathrm{reg}}(\nu_{yx})8^{\frac{1+\nu_{yx}}{2}}\,L_{yx}\widetilde{B}_y^{\frac{1}{2}}\widetilde{B}_x^{\frac{\nu_{yx}}{2}}\,.
\end{aligned}$$

Table 2: Eight learning dynamics compared in the numerical experiments and their learning rates. Here, $M = \log m$, $N = \log n$, $M' = M + 1/2$, $N' = N + 1/2$, and $D = \sqrt{M'N'} + \sqrt{MN}$. "indiv." is an abbreviation for "individual".

Name	Target	Upper bound	Proposition	Learning rate
▷ Strongly-uncoupled learning dynamics (cardinality-unaware)				
`U-Social`	Social regret	$2(M+N)+1$	Theorem 6	$\eta = 1/2, \eta' = 1/2$
`U-X-only`	x-player's regret	$M+N+1/2$	Eq. (7)	$\eta = 1, \eta' = 0$
`U-MaxInd-Cl`	Max of indiv. regrets	$3\sqrt{3}(M+N)+1/\sqrt{3}$	Theorem 10	$\eta = 1/(2\sqrt{3}), \eta' = 1/(2\sqrt{3})$
`U-MaxInd-Num`	Max of indiv. regrets	same as above	–	Numerically computed (Appendix B.4.4)
▷ Cardinality-aware strongly-uncoupled learning dynamics				
`A-Social`	Social regret	$2\sqrt{MN'} + 2\sqrt{M'N}$	Theorem 5	$\eta = \frac{\sqrt{MM'}}{D}, \eta' = \frac{\sqrt{NN'}}{D}$
`A-X-only`	x-player's regret	$2\sqrt{MN'}$	Eq. (6)	$\eta = \sqrt{M/N'}, \eta' = 0$
`A-MaxInd-Cl`	Max of indiv. regrets	$(20/3)(\sqrt{MN'} + \sqrt{M'N})$	Theorem 9	$\eta = \frac{\sqrt{MM'}}{2D}, \eta' = \frac{\sqrt{NN'}}{2D}$
`A-MaxInd-Num`	Max of indiv. regrets	$C(\sqrt{MN'} + \sqrt{M'N})$ $(C < \frac{20}{3})$	–	Numerically computed (see Appendix B.4.4)

Substituting these bounds into (45) and using the corresponding optimal choices of $\delta_{xx}, \delta_{yy}, \delta_{xy}, \delta_{yx}$, we obtain

$$\mathsf{Reg}_T^x + \mathsf{Reg}_T^y \leq C_{xx}\, T^{\frac{1-\nu_{xx}}{2}} + C_{yy}\, T^{\frac{1-\nu_{yy}}{2}} + C_{xy}\, T^{\frac{1-\nu_{xy}}{2}} + C_{yx}\, T^{\frac{1-\nu_{yx}}{2}},$$

which is the desired bound. □

G NUMERICAL EXPERIMENTS

In this section, we present numerical experiments comparing the performance of the eight learning dynamics investigated in this paper. We then demonstrate that, in settings where there is a gap between the sizes of m and n, as discussed in Section 1, being cardinality-aware indeed improves performance over the corresponding cardinality-unaware learning dynamics.

Setup The eight learning dynamics used for comparison in the experiments are summarized in Table 2. This table is obtained by modifying Table 1 to remove the lower-bound entries and to summarize the optimal learning rates η, η' for each target regret (*i.e.,* social, individual, or the maximum of the individual regrets). For the performance comparison, we used the payoff matrix defined in (9) with $\Delta_x = \Delta_y = 1$. We set the numbers of actions to $(m, n) = (2, 10^4)$ so that their difference is large, and the number of rounds to $T = 2000$. We evaluated four metrics: the social regret $\mathsf{Reg}_T^x + \mathsf{Reg}_T^y$, the maximum of the individual regrets $\max\{\mathsf{Reg}_T^x, \mathsf{Reg}_T^y\}$, the x-player's regret Reg_T^x, and the y-player's regret Reg_T^y.

Results The results are provided in Figure 3. For the social regret, the learning dynamic that minimizes the social regret under the cardinality-aware setting (`A-Social`) achieves the best performance, significantly improving upon the best cardinality-unaware algorithm (`U-Social`). For the maximum of the individual regrets, `A-Social` again achieves the best performance, followed by the approaches that directly minimize the maximum of the individual regrets under the cardinality-aware setting (`A-MaxInd-Cl` and `A-MaxInd-Num`). These results demonstrate that incorporating cardinality-awareness leads to clear performance improvements over the cardinality-unaware case.

It should be noted that `A-Social`, which achieved the best performance in terms of both social regret and the maximum of the individual regrets, is not theoretically guaranteed to upper bound the individual regret, as discussed in the main text. This suggests that the algorithm may have empirically achieved low individual regret under this particular instance, or that we can upper bound the individual regrets under this choice of learning rates. As discussed in Section 6, a more detailed investigation of this remains an important direction for future work.

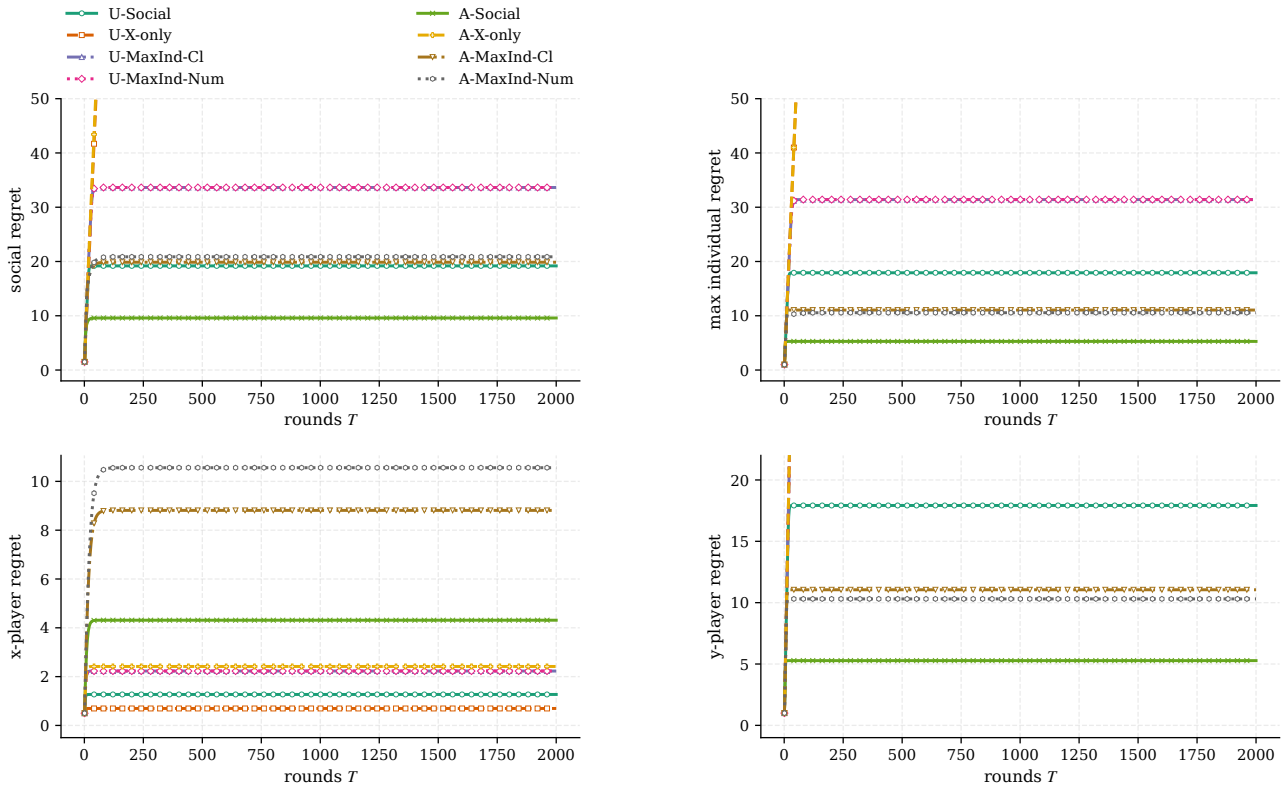

Figure 3: Regret versus the number of rounds for the learning dynamics based on the optimistic Hedge algorithm in the setting $m = 2$ and $n = 10^4$. From top-left to bottom-right: social regret, the maximum of the individual regrets, the regret of the x-player, and the regret of the y-player.