On Last-Iterate Convergence Beyond Zero-Sum Games

Ioannis Anagnostides [1]   **Ioannis Panageas** [2]   **Gabriele Farina** [1]   **Tuomas Sandholm** [1 3 4 5]

Abstract

Most existing results about *last-iterate convergence* of learning dynamics are limited to two-player zero-sum games, and only apply under rigid assumptions about what dynamics the players follow. In this paper we provide new results and techniques that apply to broader families of games and learning dynamics. First, we show that in a class of games that includes constant-sum polymatrix and strategically zero-sum games, the trajectories of dynamics such as *optimistic mirror descent (OMD)* exhibit a *boundedness* property, which holds even when players employ different algorithms and prediction mechanisms. This property enables us to obtain $O(1/\sqrt{T})$ rates and optimal $O(1)$ regret bounds. Our analysis also reveals a surprising property: OMD either reaches arbitrarily close to a Nash equilibrium or it outperforms the *robust price of anarchy* in efficiency. Moreover, for potential games we establish convergence to an ϵ-equilibrium after $O(1/\epsilon^2)$ iterations for mirror descent under a broad class of regularizers, as well as optimal $O(1)$ regret bounds for OMD variants. Our framework also extends to near-potential games, and unifies known analyses for distributed learning in Fisher's market model. Finally, we analyze the convergence, efficiency, and robustness of *optimistic gradient descent (OGD)* in general-sum continuous games.

1. Introduction

No-regret learning and game theory share an intricately connected history tracing back to Blackwell's seminal *approachability theorem* (Blackwell, 1956; Abernethy et al., 2011), leading to fundamental connections between no-regret learning and game-theoretic equilibrium concepts. For example,

[1]Carnegie Mellon University [2]University of California Irvine [3]Strategy Robot, Inc. [4]Optimized Markets, Inc. [5]Strategic Machine, Inc. Correspondence to: Ioannis Anagnostides <ianagnos@cs.cmu.edu>.

it is folklore that if both players in a zero-sum game employ a no-regret algorithm, the *time average* of their strategies will eventually converge to a *Nash equilibrium (NE)*. However, typical guarantees within the no-regret framework provide no insights about the *final* state of the system. This begs the question: *Will the agents eventually play according to an equilibrium strategy?* In general, the answer to this question is *no*: Broad families of regret minimization algorithms such as *mirror descent (MD)* are known to exhibit recurrent or even chaotic behavior (Sato et al., 2002; Sandholm, 2010; Mertikopoulos et al., 2018).

In an attempt to stabilize the chaotic behavior of traditional no-regret learning algorithms and ameliorate the notoriously tedious training process of *generative adversarial networks (GANs)* (Goodfellow et al., 2014), Daskalakis et al. (2018) discovered that *optimistic gradient descent (OGD)* (Popov, 1980) guarantees *last-iterate* convergence in unconstrained bilinear "games".[1] Thereafter, their result has been extended along several lines (*e.g.*, (Daskalakis & Panageas, 2019; Mertikopoulos et al., 2019; Wei et al., 2021b; Golowich et al., 2020a; Azizian et al., 2021)). Last-iterate convergence is also central in economics (Milgrom & Roberts, 1990), and goes back to the fundamental question of what it really means to "learn in a game". Indeed, it is unclear how a time average guarantee is meaningful from an economic standpoint. Nevertheless, the known guarantees only apply for restricted classes of games such as two-player zero-sum games. As a result, it is natural to ask whether last-iterate convergence could be a *universal* phenomenon in games.

Unfortunately, this cannot be the case: if every player employs a no-regret algorithm and the individual strategies converge, this would necessarily mean that the limit point is a Nash equilibrium.[2] But this is precluded by fundamental impossibility results (Hart & Mas-Colell, 2003; Rubinstein, 2016; Chen et al., 2009; Babichenko, 2014; Daskalakis et al., 2009), at least in a polynomial number of iterations. This observation offers an additional crucial motivation for convergence since Nash equilibria can be exponentially more efficient compared to *correlated equilibria* (Blum et al., 2008).

---

[1]We call them "games" with an abuse of terminology. They are not games in the game-theoretic sense.

[2]To see this, observe that the average product distribution of play—which is a *coarse correlated equilibrium*—will converge to a product distribution, and hence, a Nash equilibrium.

In light of this, we ask the following question: *For which classes of games beyond two-player zero-sum games can we guarantee last-iterate convergence?* Characterizing dynamics beyond zero-sum games is a recognized important open question in the literature (Candogan et al., 2013). In fact, even in GANs many practitioners employ a *general-sum* objective due to its practical superiority. However, many existing techniques are tailored to the min-max structure of the problem.

Another drawback of existing last-iterate guarantees is that they make restrictive assumptions about the dynamics the players follow; namely, they employ exactly the *same learning algorithm*. However, we argue that in many settings such a premise is unrealistic under independent and decentralized players. Indeed, to quote from the work of Candogan et al. (2013): *"The limit behavior of dynamic processes where players adhere to different update rules is [...] an open question, even for potential games"*. In this paper, we make progress towards addressing these fundamental questions.

1.1 Overview of Our Contributions   The existing techniques employed to show last-iterate convergence are inherently different from the ones used to analyze regret. Indeed, it is tacitly accepted that these two require a different treatment; this viewpoint is reflected by Mertikopoulos et al. (2018): *"a regret-based analysis cannot distinguish between a self-stabilizing system, and one with recurrent cycles"*.

Our first contribution is to challenge this conventional narrative. We show that for a broad class of games the *regret bounded by variation in utilities* (RVU) property established by Syrgkanis et al. (2015) implies that the second-order path lengths are bounded (Theorem 3.1) for a broad family of dynamics including *optimistic mirror descent (OMD)* (Rakhlin & Sridharan, 2013). We use this bound to obtain optimal $O(1)$ individual regret bounds (Corollary 3.3), and, more importantly, to show that $O(1/\epsilon^2)$ iterations suffice to reach an ϵ-approximate Nash equilibrium (Theorem 3.4). This characterization holds for a class of games intricately connected to the *minimax theorem*, and includes *constant-sum polymatrix* (Kearns et al., 2001) and *strategically zero-sum* games (Moulin & Vial, 1978).

Furthermore, our results hold even if players employ *different OMD variants* (with smooth regularizers) and even with *advanced predictions*, as long as an appropriate RVU bound holds. Additionally, our techniques apply under arbitrary (convex and compact) constrained sets, thereby allowing for a direct extension, for example, to *extensive-form games*. We also illustrate how our framework can be applied to smooth min-max optimization. Such results appear hard to obtain with prior techniques. Also, in cases where prior techniques applied, our proofs are considerably simpler. Overall, our framework inherits all of the robustness and the simplicity deriving from a regret-based analysis.

Our approach also reveals an intriguing additional result: OMD variants either converge arbitrarily close to a Nash equilibrium, or they outperform the *robust price of anarchy* in terms of efficiency (Roughgarden, 2015) (see Theorem 3.8 for a formal statement). This is based on the observation that lack of last-iterate convergence can be leveraged to show that the sum of the players' regrets at a sufficiently large time T will be at most $-CT$, for some parameter $C > 0$. This substantially refines the result of Syrgkanis et al. (2015) using the rather underappreciated fact that *regret can be negative*. In fact, the further the dynamics are from a Nash equilibrium, the larger is the improvement compared to the robust price of anarchy.

Next, we study the convergence of *mirror descent (MD)* learning dynamics in (weighted) potential games. We give a new potential argument applicable even if players employ different regularizers (Theorem 4.3), thereby addressing an open question of Candogan et al. (2013) regarding heterogeneous dynamics in potential games. Such results for no-regret algorithms were known when all players employed variants of *multiplicative weights update* (Palaiopanos et al., 2017; Héliou et al., 2017), though in (Héliou et al., 2017) the analysis requires vanishing learning rates.

Our potential argument implies a boundedness property for the trajectories, (similarly to Theorem 3.1 we previously discussed); we use this property to show that $O(1/\epsilon^2)$ iterations suffice to reach an ϵ-Nash equilibrium. We also show a similar boundedness property for *optimistic* variants, which is then used—along with the RVU property—to show optimal $O(1)$ regret for each individual player (Theorem 4.6). To our knowledge, this is the first result showing optimal $O(1)$ individual regret in potential games, and it is based on the observation that last-iterate convergence can be leveraged to show improved regret bounds. In a sense, this is the "converse" of the technique employed in Section 3. Our potential argument also extends to *near-potential games* (Candogan et al., 2013), implying convergence to approximate Nash equilibria (Theorem 4.10). Importantly, our framework is general enough to unify results from distributed learning in Fisher's market model as well (Birnbaum et al., 2011).

Finally, motivated by applications such as GANs, we study the convergence of *optimistic gradient descent (OGD)* in *continuous games*. We characterize the class of two-player games for which OGD converges, extending the known prior results (Theorem 5.1). Also, we show that OGD can be arbitrarily inefficient beyond zero-sum games (Proposition 5.2). More precisely, under OGD players can fail to coordinate even if the objective presents a "clear" coordination aspect. Finally, in Theorem 5.4 we use our techniques to characterize convergence in multiplayer settings as well.

For the convenience of the reader we have summarized our key results in Table 1.

Setting	Results
Games with nonnegative regrets (Sections 3.1 and 3.2)	• Bounded second-order path lengths (Theorem 3.1) • Optimal regret (Corollary 3.3) • $O(1/\sqrt{T})$ rates to Nash equilibria (NE) (Theorem 3.4)
Smooth games (Section 3.3)	• Outperforming the (robust) price of anarchy (Theorem 3.8)
Potential and near-potential games (Section 4)	• Optimal regret (Theorem 4.6) • $O(1/\sqrt{T})$ rates to NE (Theorems 4.7 and 4.10) • Convergence in Fisher markets (Appendix B.2)
Unconstrained general-sum games (Section 5)	• Convergence of OGD (Theorems 5.1 and 5.4) • Inefficiency of OGD (Proposition 5.2) • Instability of first-order methods (Theorem C.11)

Table 1. Overview of our main results.

1.2 Further Related Work   The related literature on the subject is too vast to cover in its entirety; below we emphasize certain key contributions.

For *constrained* zero-sum games Daskalakis & Panageas (2019) established asymptotic last-iterate convergence for *optimistic multiplicative weights update*, although they had to assume the existence of a *unique* Nash equilibrium. Our techniques do not require a uniqueness assumption, which is crucial for *equilibrium refinements* (Damme, 1987). Wei et al. (2021b) improved several aspects of the result of Daskalakis & Panageas (2019), showing a surprising linear rate of convergence, albeit with a dependence on condition number-like quantities which were left unspecified. Some of the latter results were extended for the substantially more complex class of extensive-form games by Lee et al. (2021). While the aforementioned results apply under a *time-invariant* learning rate, there has also been a substantial amount of work considering a *vanishing* scheduling (*e.g.*, (Mertikopoulos et al., 2019; Mertikopoulos & Zhou, 2019; Zhou et al., 2017; 2018; Hsieh et al., 2021)). Our results apply under a constant learning rate, which has been extensively argued to be desirable in the literature (Bailey & Piliouras, 2019; Golowich et al., 2020a). Last-iterate convergence has also been recently documented in certain settings from competitive *reinforcement learning* (Daskalakis et al., 2020; Wei et al., 2021a; Leonardos et al., 2022).

Beyond zero-sum games, our knowledge remains much more limited. Cheung & Piliouras (2020); Cheung & Tao (2021) made progress by establishing chaotic behavior for instances of OMD in *bimatrix games*. Last-iterate convergence under *follow the regularizer leader (FTRL)* is known to be rather elusive (Vlatakis-Gkaragkounis et al., 2020), occurring only under *strict Nash equilibria* (Giannou et al., 2021). On the positive side, in (Hsieh et al., 2021), last-iterate convergence for a variant of OMD with an adaptive learning rate was established for *variationally stable games*,

a class of games that includes zero-sum convex-concave. The convergence of learning dynamics has also recently received attention in several auction settings; we refer to the work of Feng et al. (2021) and Deng et al. (2022), as well as references therein.

In the unconstrained setting, the behavior of OGD is by now well-understood in bilinear zero-sum games (Liang & Stokes, 2019; Mokhtari et al., 2020b; Zhang & Yu, 2020). Various results have also been shown for *convex-concave* landscapes (*e.g.*, (Mertikopoulos et al., 2019)), and *cocoercive games* (*e.g.*, (Lin et al., 2020)), but covering this literature goes beyond our scope. For *monotone variational inequalities (VIs)* problems (Harker & Pang, 1990), Golowich et al. (2020a) showed last-iterate rates of $O(1/\sqrt{T})$, which is also tight for their considered setting. Note that this rate is slower than that of the time average for the *extra-gradient (EG)* method (Golowich et al., 2020b). Lastly, last-iterates rates for variants of OMD have been obtained by Azizian et al. (2021) in a certain stochastic VI setup.

2. Preliminaries

Conventions   In the main body we allow the $O(\cdot)$ notation to suppress game-dependent parameters *polynomial* in the game; precise statements are given in the Appendix. We use subscripts to indicate players, and superscripts for time indices. The j-th coordinate of $\boldsymbol{x} \in \mathbb{R}^d$ is denoted by $\boldsymbol{x}(j)$.

Normal-Form Games   In this paper we primarily focus on *normal-form games (NFGs)*. We let $[n] \coloneqq \{1, \dots, n\}$ be the set of players. Each player $i \in [n]$ has a set of available *actions* $\mathcal{A}_i$. The *joint action profile* is denoted by $\boldsymbol{a} \coloneqq (a_1, \dots, a_n) \in \prod_{i \in [n]} \mathcal{A}_i$. For a given $\boldsymbol{a}$ each player i receives some (normalized) *utility* $u_i(\boldsymbol{a})$, where $u_i : \prod_{i \in [n]} \mathcal{A}_i \to [-1, 1]$. Players are allowed to randomize, and we denote by $\boldsymbol{x}_i(a_i)$ the probability that player i will chose action $a_i \in \mathcal{A}_i$, so that $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i) \coloneqq$

$\left\{\boldsymbol{x} \in \mathbb{R}_{\geq 0}^{|\mathcal{A}_i|} : \sum_{a_i \in \mathcal{A}_i} \boldsymbol{x}(a_i) = 1\right\}$. The *expected utility* of player i can be expressed as $u_i(\boldsymbol{x}) := \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{x}}[u_i(\boldsymbol{a})]$, where $\boldsymbol{x} := (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is the *joint (mixed) strategy profile*. The *social welfare* for a given $\boldsymbol{a}$ is defined as the sum of the players' utilities: $\mathrm{SW}(\boldsymbol{a}) := \sum_{i=1}^n u_i(\boldsymbol{a})$. As usual, we overload notation to denote $\mathrm{SW}(\boldsymbol{x}) := \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{x}} \mathrm{SW}(\boldsymbol{a})$. We also let $\mathrm{OPT} := \max_{\boldsymbol{a} \in \mathcal{A}} \mathrm{SW}(\boldsymbol{a})$ be the maximum social welfare.

Definition 2.1 (Approximate Nash equilibrium). A joint strategy profile $(\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*) \in \prod_{i \in [n]} \Delta(\mathcal{A}_i)$ is an ϵ-*approximate (mixed) Nash equilibrium* if for any player $i \in [n]$ and any unilateral deviation $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)$,

$$u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}^*) \leq u_i(\boldsymbol{x}^*) + \epsilon.$$

Regret Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex and compact set. A *regret minimization algorithm* produces at every time $t \in \mathbb{N}$ a strategy $\boldsymbol{x}^{(t)} \in \mathcal{X}$, and receives as feedback from the environment a (linear) utility function $u^{(t)}(\boldsymbol{x}) : \boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{u}^{(t)} \rangle$. The *cumulative regret* (or simply regret) Reg^T up to time T measures the performance of the regret minimization algorithm compared to the optimal *fixed strategy in hindsight*:

$$\mathrm{Reg}^T := \max_{\boldsymbol{x}^* \in \mathcal{X}} \left\{ \sum_{t=1}^T \langle \boldsymbol{x}^*, \boldsymbol{u}^{(t)} \rangle \right\} - \sum_{t=1}^T \langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle.$$

Optimistic Dynamics Following a recent trend in online learning (Rakhlin & Sridharan, 2013), we study *optimistic* (or *predictive*) algorithms. Let $\mathcal{R}$ be a 1-strongly convex *regularizer*—or *distance generating function (DGF)*—with respect to some norm $\|\cdot\|$, continuously differentiable on the relative interior of the feasible set (Rockafellar, 1970). We denote by $D_{\mathcal{R}}$ the *Bregman divergence* generated by $\mathcal{R}$; that is, $D_{\mathcal{R}}(\boldsymbol{x}, \boldsymbol{y}) := \mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y}) - \langle \nabla \mathcal{R}(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle$. Canonical DGFs include the *squared Euclidean* $\mathcal{R}(\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{x}\|_2^2$ which generates the *squared Euclidean distance*, as well as the *negative entropy* $\mathcal{R}(\boldsymbol{x}) = \sum_{r=1}^d \boldsymbol{x}(r)(\log \boldsymbol{x}(r) - 1)$ which generates the *Kulllback-Leibler divergence*. *Optimistic mirror descent (OMD)* has an internal prediction $\boldsymbol{m}^{(t)}$ at every time $t \in \mathbb{N}$, so that the update rule takes the following form for $t \geq 0$:

$$\boldsymbol{x}^{(t+1)} := \arg\max_{\boldsymbol{x} \in \mathcal{X}} \eta \langle \boldsymbol{x}, \boldsymbol{m}^{(t+1)} \rangle - D_{\mathcal{R}}(\boldsymbol{x}, \widehat{\boldsymbol{x}}^{(t)});$$
$$\widehat{\boldsymbol{x}}^{(t+1)} := \arg\max_{\widehat{\boldsymbol{x}} \in \mathcal{X}} \eta \langle \widehat{\boldsymbol{x}}, \boldsymbol{u}^{(t+1)} \rangle - D_{\mathcal{R}}(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{x}}^{(t)}), \quad \text{(OMD)}$$

where $\widehat{\boldsymbol{x}}^{(0)} := \arg\min_{\widehat{\boldsymbol{x}} \in \mathcal{X}} \mathcal{R}(\widehat{\boldsymbol{x}})$. We also consider the *optimistic follow the regularizer leader (OFTRL)* algorithm of Syrgkanis et al. (2015), defined as follows.

$$\boldsymbol{x}^{(t+1)} := \arg\max_{\boldsymbol{x} \in \mathcal{X}} \eta \left\langle \boldsymbol{x}, \boldsymbol{m}^{(t+1)} + \sum_{\tau=1}^t \boldsymbol{u}_i^{(\tau)} \right\rangle - \mathcal{R}(\boldsymbol{x}).$$

Unless specified otherwise, it will be assumed that $\boldsymbol{m}^{(t+1)} := \boldsymbol{u}^{(t)}$ for $t \geq 1$, and $\boldsymbol{m}^{(1)} := \boldsymbol{0}$. Syrgkanis et al. (2015) established that OMD and OFTRL have the so-called RVU property, which we recall below.

Definition 2.2 (RVU Property). A regret minimizing algorithm satisfies the (α, β, γ)-RVU property under a pair of dual norms[3] $(\|\cdot\|, \|\cdot\|_*)$ if for any sequence of utility vectors $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(T)}$ its regret Reg^T can be bounded as

$$\mathrm{Reg}^T \leq \alpha + \beta \sum_{t=1}^T \|\boldsymbol{u}^{(t)} - \boldsymbol{u}^{(t-1)}\|_*^2 - \gamma \sum_{t=1}^T \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|^2,$$

where $\alpha, \beta, \gamma > 0$ are time-independent parameters, and $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(T)}$ is the sequence of produced iterates.

Proposition 2.3 (Syrgkanis et al., 2015). *OFTRL with η and $\boldsymbol{m}^{(t)} = \boldsymbol{u}^{(t-1)}$ satisfies the RVU property with $(\alpha, \beta, \gamma) = (\frac{\Omega}{\eta}, \eta, \frac{1}{4\eta})$, where $\Omega := \sup_{\boldsymbol{x}} \mathcal{R}(\boldsymbol{x}) - \inf_{\boldsymbol{x}} \mathcal{R}(\boldsymbol{x})$. OMD with η and $\boldsymbol{m}^{(t)} = \boldsymbol{u}^{(t-1)}$ satisfies the RVU property with $(\alpha, \beta, \gamma) = (\frac{\Omega}{\eta}, \eta, \frac{1}{8\eta})$, where $\Omega := \sup_{\boldsymbol{x}} D_{\mathcal{R}}(\boldsymbol{x}, \widehat{\boldsymbol{x}}^{(0)})$.*

3. Optimistic Learning in Games

In this section we study optimistic learning dynamics. We first show that the RVU bound implies a strong boundedness property for the trajectories under a broad class of games.

Theorem 3.1. *Suppose that every player $i \in [n]$ employs a regret minimizing algorithm satisfying the $(\alpha_i, \beta_i, \gamma_i)$-RVU property with $\gamma_i \geq 2(n-1) \sum_{j \neq i} \beta_j$, for all $i \in [n]$, under the pair of dual norms $(\|\cdot\|_1, \|\cdot\|_\infty)$. If the game is such that $\sum_{i=1}^n \mathrm{Reg}_i^T \geq 0$, for any $T \in \mathbb{N}$, it holds that*

$$\sum_{i=1}^n \sum_{t=1}^T \gamma_i \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq 2 \sum_{i=1}^n \alpha_i. \quad (1)$$

The ℓ_1-norm in Theorem 3.1 is only used for convenience; (1) immediately extends under any equivalent norm. Now the requirement of Theorem 3.1 related to the RVU property can be satisfied by OMD and OFTRL, as implied by Proposition 2.3. Concretely, if all players employ (OMD) with the same $\eta > 0$, it would suffice to take $\eta \leq \frac{1}{4(n-1)}$. In light of this, applying Theorem 3.1 only requires that the game is such that $\sum_{i=1}^n \mathrm{Reg}_i^T \geq 0$. Next, we show that this is indeed the case for certain important classes of games.

3.1 Games with Nonnegative Sum of Regrets In *polymatrix games* there is an underlying graph so that every node is associated with a given player and every edge corresponds to a two-person game. The utility of a player is the sum of the utilities obtained from each individual game played with its neighbors on the graph (Kearns et al., 2001). In a *polymatrix zero-sum game*, every two-person game is

---

[3]The dual of norm $\|\cdot\|$ is defined as $\|\boldsymbol{v}\|_* := \sup_{\|\boldsymbol{u}\| \leq 1} \langle \boldsymbol{u}, \boldsymbol{v} \rangle$.

zero-sum. More generally, in a *zero-sum polymatrix game* the sum of *all* players' utilities has to be 0.

Strategically zero-sum games, introduced by Moulin & Vial (1978), is the subclass of bimatrix games which are *strategically equivalent* to a zero-sum game (see Definition A.5 for a formal statement). For this class of games, Moulin & Vial (1978) showed that *fictitious play* converges to a Nash equilibrium. We will extend their result under a broad class of no-regret learning dynamics. An important feature of strategically zero-sum games is that it constitutes the *exact class of games* whose fully-mixed Nash equilibria cannot be improved by, *e.g.*, a *correlation scheme* (Aumann, 1974).

Proposition 3.2 (Abridged; Full Version in Proposition A.10). *For the following classes of games it holds that $\sum_{i=1}^{n} \mathrm{Reg}_i^T \geq 0$:*

1. *Two-player zero-sum games;*
2. *Polymatrix zero-sum games;*
3. *Constant-sum Polymatrix games;*
4. *Strategically zero-sum games;*[4]
5. *Polymatrix strategically zero-sum games;*[4]

As suggested in the full version (Proposition A.10), this class of games appears to be intricately connected with the admission of a *minimax theorem*. As such, it also captures certain nonconvex-nonconcave landscapes such as *quasiconvex-quasiconcave games* (Sion, 1958) and *stochastic games* (Shapley, 1953), but establishing last-iterate convergence for those settings goes beyond the scope of this paper. Note that Von Neumann's minimax theorem for two-player zero-sum games can be generalized to the polymatrix games we are considering (Daskalakis & Papadimitriou, 2009; Cai & Daskalakis, 2011; Cai et al., 2016). Interestingly, our framework also has implications when the *duality gap* is small; see Remark A.11.

3.2 Implications   Having established the richness of the class of games we are capturing, we present implications deriving from Theorem 3.1, and extensions thereof. First, we show that it implies optimal $O(1)$ *individual regret*.

Corollary 3.3 (Optimal Regret Bound). *In the setting of Theorem 3.1 with $\alpha_i = \alpha$, $\beta_i = \beta$, and $\gamma_i = \gamma$, the individual regret of each player $i \in [n]$ can be bounded as*

$$\mathrm{Reg}_i^T \leq \alpha + \frac{2n(n-1)\alpha\beta}{\gamma}.$$

For example, under (OMD) with $\eta = \frac{1}{4(n-1)}$ this corollary implies that $\mathrm{Reg}_i^T \leq 8n\Omega_i \leq 8n\log|\mathcal{A}_i|$, for any $i \in [n]$. Moreover, we also obtain an $O(1/\epsilon^2)$ bound on the number of iterations so that the last iterate is an ϵ-Nash equilibrium.

Theorem 3.4 (Abridged; Full Version in Theorem A.12). *Suppose that each player employs (OMD) with a suitable*

---

[4]See Remark A.7.

regularizer. If $\sum_{i=1}^{n} \mathrm{Reg}_i^T \geq 0$, then for any $\epsilon > 0$ and after $T = O(1/\epsilon^2)$ iterations there is joint strategy $\boldsymbol{x}^{(t)}$, with $t \in [T]$, which is an ϵ-approximate Nash equilibrium.

This theorem requires smoothness of the regularizer used by each player (which could be different), satisfied by a broad family that includes the Euclidean DGF. Interestingly, the argument does not appear to hold under non-smooth regularizers such as the negative entropy DGF, even if one works with common local norms. We also remark that while the bound in Theorem 3.4 applies for the best iterate, it is always possible to terminate once a desired precision has been reached. Asymptotic last-iterate convergence follows immediately from our techniques; see Remark A.15.

Advanced Predictions   Our last-iterate guarantees are the first to apply even if players employ more advanced prediction mechanisms. Indeed, Syrgkanis et al. (2015) showed that a qualitatively similar RVU bound can be derived under

1. *H-step recency bias*: Given a window of size H, we define $\boldsymbol{m}^{(t)} := \sum_{\tau=t-H}^{t-1} \boldsymbol{u}^{(\tau)}/H$;
2. *Geometrically discounted recency bias*: For $\delta \in (0,1)$, we define $\boldsymbol{m}^{(t)} := \frac{1}{\sum_{\tau=0}^{t-1} \delta^{-\tau}} \sum_{\tau=0}^{t-1} \delta^{-\tau} \boldsymbol{u}^{(\tau)}$.

In Proposition A.2 we show the boundedness property for the trajectories under such prediction mechanisms, and we experimentally evaluate their performance in Appendix D.

Bilinear Saddle-Point Problems   Our framework also has direct implications for *extensive-form games (EFGs)* where the strategy space of each player is no longer a probability simplex. In particular, computing a Nash equilibrium in zero-sum EFGs can be formulated as a *bilinear saddle-point problem (BSPP)*

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} \boldsymbol{x}^{\top} \mathbf{A} \boldsymbol{y}, \tag{2}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are convex and compact sets associated with the sequential decision process faced by each player (Romanovskii, 1962; von Stengel, 1996; Koller et al., 1996).

Proposition 3.5 (Abridged; Full Version in Proposition A.22). *Suppose that both players in a BSPP employ (OMD) with $\eta \leq \frac{1}{4\|\mathbf{A}\|_2}$, where $\|\cdot\|_2$ denotes the spectral norm. Then, after $T = O(1/\epsilon^2)$ iterations there is a pair $(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)})$ with $t \in [T]$ which is an $O(\epsilon)$-approximate Nash equilibrium.*

We are not aware of any prior *polynomial-time* guarantees for the last iterate of no-regret learning dynamics in EFGs. Asymptotic last-iterate convergence also follows immediately. Proposition 3.5 is verified through experiments on benchmark EFGs in Section 6 (Figure 1).

Remark 3.6 (Convex-Concave Games). Our framework also applies to smooth min-max optimization via a well-known linearization trick; see Appendix A.1. Implications for unconstrained setups are also immediate; see Remark A.19.

3.3 Convergence of the Social Welfare

We conclude this section with a surprising new result presented in Theorem 3.8. Let us first recall the concept of a *smooth game*.

Definition 3.7 (Smooth Game; Roughgarden, 2015)**.** A utility-maximizing game is (λ, μ)-*smooth* if for any two action profiles $\boldsymbol{a}, \boldsymbol{a}^* \in \mathcal{A}$ it holds that

$$\sum_{i=1}^{n} u_i(a_i^*, \boldsymbol{a}_{-i}) \geq \lambda \operatorname{SW}(\boldsymbol{a}^*) - \mu \operatorname{SW}(\boldsymbol{a}).$$

The smoothness condition imposes a bound on the "externality" of any unilateral deviation. The *robust price of anarchy (PoA)* of a game Γ is defined as $\rho := \sup_{\lambda, \mu} \lambda / (1 + \mu)$, where the supremum is taken over all possible λ, μ for which Γ is (λ, μ)-smooth. Smooth games with favorable smoothness parameters include simultaneous second-price auctions (Christodoulou et al., 2016), valid utility games (Vetta, 2002), and congestion games with affine costs (Awerbuch et al., 2013; Christodoulou & Koutsoupias, 2005). We refer to (Roughgarden, 2015) for an additional discussion.

The importance of Roughgarden's smoothness framework is that it guarantees convergence (in a time average sense) of no-regret learning to outcomes with *approximately optimal welfare*, where the approximation depends on the robust PoA. In particular, the convergence is controlled by the sum of the players' regrets (Syrgkanis et al., 2015, Proposition 2). We use this property to obtain the following result.

Theorem 3.8 (Abridged; Full Version in Theorem A.17)**.** *Suppose that each player in a (λ, μ)-smooth game Γ employs* (OMD) *with a suitable regularizer and learning rate $\eta > 0$. For any $\gamma > 0$ and a sufficiently large number of iterations $T = \Omega(1/\gamma^2)$, either of the following occurs:*

1. *There is an iterate $\boldsymbol{x}^{(t)}$ with $t \in [T]$ which is an $O(\gamma)$-approximate Nash equilibrium;*
2. *Otherwise, it holds that*

$$\frac{1}{T} \sum_{t=1}^{T} \operatorname{SW}(\boldsymbol{x}^{(t)}) \geq \frac{\lambda}{1 + \mu} \operatorname{OPT} + \frac{1}{1 + \mu} \frac{\gamma^2}{16\eta},$$

where (λ, μ) are the smoothness parameters of Γ.

In words, the dynamics either approach arbitrarily close to a Nash equilibrium, or the time average of the social welfare outperforms the robust price of anarchy. Either of these implications is remarkable.

4. Convergence with the Potential Method

In this section we show optimal regret bounds (Theorem 4.6) and last-iterate rates (Theorem 4.7) for the fundamental class of *potential games* (Monderer & Shapley, 1996; Rosenthal,

1973) under mirror descent (MD) with suitable regularizers. In fact, our approach is general enough to capture under a unifying framework distributed learning in Fisher's market model (Birnbaum et al., 2011), while we expect that further applications will be identified in the future. Finally, we show that our approach can also be applied in *near-potential* games, in the precise sense of (Candogan et al., 2013), showing convergence to approximate Nash equilibria.

We commence by formally describing the class of games we are considering in this section; in Proposition 4.2 we show that it incorporates typical potential games.

Definition 4.1. Let $\Phi : \prod_{i=1}^{n} \Delta(\mathcal{A}_i) \ni (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \mapsto \mathbb{R}$ be a bounded function, with $\max_{\boldsymbol{x}} |\Phi(\boldsymbol{x})| \leq \Phi_{\max}$, for which there exists $L > 0$ such that for any $\boldsymbol{x}, \widetilde{\boldsymbol{x}}$ it holds that

$$\Phi(\boldsymbol{x}) \leq \Phi(\widetilde{\boldsymbol{x}}) - \langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \widetilde{\boldsymbol{x}} - \boldsymbol{x} \rangle + L \|\widetilde{\boldsymbol{x}} - \boldsymbol{x}\|_2^2. \quad (3)$$

Moreover, let g_i be a strictly increasing function for each $i \in [n]$. A game Γ is $(g_1, \ldots, g_n)$-*potential* if for all $i \in [n]$, $a_i \in \mathcal{A}_i$, and $\boldsymbol{x}_{-i} \in \prod_{j \neq i} \Delta(\mathcal{A}_j)$ we have that

$$\frac{\partial \Phi(\boldsymbol{x})}{\partial \boldsymbol{x}_i(a_i)} = g_i(u_i(a_i, \boldsymbol{x}_{-i})). \quad (4)$$

A few remarks are in order. First, (3) imposes a "one-sided" smoothness condition. This relaxation turns out to be crucial to encompass settings such as linear Fisher markets (Birnbaum et al., 2011). Moreover, (4) prescribes applying a monotone transformation to the utility. While the identity mapping $g_i : x \mapsto x$ suffices to cover typical potential games, a logarithmic transformation is required to capture the celebrated *proportional response* dynamics (Birnbaum et al., 2011; Wu & Zhang, 2007) in markets.

Proposition 4.2. *Let Γ be a game for which there exists a function $\Phi : \prod_{i=1}^{n} \mathcal{A}_i \to \mathbb{R}$ with*

$$\Phi(\boldsymbol{a}) - \Phi(a_i', \boldsymbol{a}_{-i}) = \boldsymbol{w}_i(u_i(\boldsymbol{a}) - u_i(a_i', \boldsymbol{a}_{-i})),$$

where $\boldsymbol{w} \in \mathbb{R}_{>0}^n$. Then, Γ is a potential game in the sense of Definition 4.1 with $L = \frac{1}{2} \Phi_{\max} \sum_{i=1}^{n} |\mathcal{A}_i|$.

In this context, we will assume that all players update their strategies using mirror descent (MD) for $t \geq 0$:

$$\boldsymbol{x}_i^{(t+1)} := \underset{\boldsymbol{x} \in \Delta(\mathcal{A}_i)}{\arg\max} \, \eta \langle g_i(\boldsymbol{u}_i^{(t)}), \boldsymbol{x} \rangle - D_{\mathcal{R}_i}(\boldsymbol{x}, \boldsymbol{x}_i^{(t)}). \quad \text{(MD)}$$

In accordance to Definition 4.1, players apply (coordinate-wise) the transformation g_i to the observed utility. Importantly, we will allow players to employ different regularizers, as long as $\mathcal{R}_i$ is 1-strongly convex with respect to $\|\cdot\|_2$. This trivially holds under the Euclidean DGF, while it also holds under negative entropy. We are ready to establish that the potential function increases along non-stationary orbits:

Theorem 4.3. *Suppose that each player employs* (MD) *with a 1-strongly convex regularizer with respect to* $\|\cdot\|_2$, *and learning rate* $\eta = \frac{1}{2L}$, *where* L *is defined as in Proposition 4.2. Then, for any* $t \geq 1$ *it holds that*

$$\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \geq \frac{1}{2\eta} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 \geq 0. \quad (5)$$

For large values of learning rate η variants of (MD) are known to exhibit chaotic behavior in potential games (Bielawski et al., 2021; Palaiopanos et al., 2017). Theorem 4.3 also implies the following boundedness property for the trajectories from a telescopic summation of (5):

Corollary 4.4. *In the setting of Theorem 4.3 it holds that*

$$\frac{1}{2\eta} \sum_{t=1}^{T-1} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \leq \Phi(\boldsymbol{x}^{(T)}) - \Phi(\boldsymbol{x}^{(1)}) \leq 2\,\Phi_{\max}.$$

In the sequel, to argue about the regret incurred by each player and the convergence to Nash equilibria, we tacitly assume that g_i is the identity map. The first implication of Corollary 4.4 is an $O(\sqrt{T})$ bound on the individual regrets.

Corollary 4.5. *In the setting of Theorem 4.3, with* $\mathcal{R}_i \coloneqq \frac{1}{2}\|\boldsymbol{x}\|_2^2$, *it holds that the regret of each player* $i \in [n]$ *is such that* $\mathrm{Reg}_i^T = O(\sqrt{T})$.

Note that we establish vanishing $O(1/\sqrt{T})$ average regret under *constant* learning rate, thereby deviating from the traditional regret analysis of (MD). More importantly, with a more involved argument we show optimal individual regret under optimistic multiplcative weights update (OMWU):

Theorem 4.6 (Optimal Regret for Potential Games)**.** *Suppose that each player employs OMWU with a sufficiently small learning rate* $\eta > 0$. *Then, the regret of each player* $i \in [n]$ *is such that* $\mathrm{Reg}_i^T = O(1)$.

This theorem is based on showing a suitable boundedness property (Theorem B.5), which is then appropriately combined with Proposition 2.3. While Theorem 4.6 is stated for OMWU, the proof readily extends well-beyond. In this way, when the underlying game is potential, we substantially strengthen and simplify the result of Daskalakis et al. (2021). Moreover, we also obtain a bound on the number of iterations required to reach an approximate Nash equilibrium.

Theorem 4.7 (Abridged; Full Version in Theorem B.6)**.** *Suppose that each player employs* (MD) *with a suitable regularizer. Then, after* $O(1/\epsilon^2)$ *iterations there is a strategy* $\boldsymbol{x}^{(t)}$ *which is an* ϵ-*approximate Nash equilibrium.*

This theorem has a similar flavour to Theorem 3.4, and it is the first of its kind for the class of potential games. Finally, in settings where the potential function is concave we show a rate of convergence of $O(1/T)$.

Proposition 4.8. *In the setting of Theorem 4.3, if the potential function* Φ *is also concave it holds that*

$$\Phi(\boldsymbol{x}^*) - \Phi(\boldsymbol{x}^{(T+1)}) \leq \frac{2L}{T} \sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(1)}).$$

This result is stronger than the standard convergence guarantee in smooth convex optimization since Definition 4.1 only imposes a one-sided condition. While concavity does not hold in typical potential games, it applies to games such as Fisher's market model; see our remark below.

Remark 4.9. In Appendix B.2 we explain how our framework naturally captures the analysis of Birnbaum et al. (2011) for distributed dynamics in Fisher's market model.

4.1 Near-Potential Games Finally, we illustrate the robustness of our framework by extending our results to *near-potential* games (Candogan et al., 2013). Roughly speaking, a game is near-potential if it is close—in terms of the maximum possible utility improvement through unilateral deviations (MPD)—to a potential game; we defer the precise definition to Appendix B.1. It is also worth noting that our result immediately extends under different distance measures. In this context, we prove the following theorem.

Theorem 4.10 (Abridged; Full Version in Theorem B.12)**.** *Consider a* δ-*near-potential game where each player employs* (MD) *with suitable regularizer. Then, there exists a potential function* Φ *which increases as long as* $\boldsymbol{x}^{(t)}$ *is not an* $O(\sqrt{\delta})$-*Nash equilibrium.*

5. Continuous Games

Sections 3 and 4 primarily focused on classes of games stemming from applications in game theory. In this section we shift our attention to *continuous "games"*, strategic interactions motivated by applications such as GANs. Specifically, we study the convergence properties of *optimistic gradient descent (OGD)* beyond two-player zero-sum games. Recall that OGD in unconstrained settings can be expressed using the following update rule for $t \geq 1$:

$$\boldsymbol{x}_i^{(t+1)} = \boldsymbol{x}_i^{(t)} + 2\eta \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}^{(t)}) - \eta \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}^{(t-1)}), \quad \text{(OGD)}$$

for any player $i \in [n]$, where $u_i : \prod_{i \in [n]} \mathcal{X}_i \to \mathbb{R}$ is assumed to be continuously differentiable.

5.1 Two-Player Games We first study two-player games. Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d \times d}$ be matrices so that under strategies $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$ the utilities of players $\mathcal{X}$ and $\mathcal{Y}$ are given by the bilinear form $\boldsymbol{x}^\top \mathbf{A} \boldsymbol{y}$ and $\boldsymbol{x}^\top \mathbf{B} \boldsymbol{y}$ respectively. As

is common in this line of work, the matrices are assumed to be square and non-singular. In line with the application of interest, we mostly focus on the *unconstrained setting*, where $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{X} = \mathbb{R}^d$; some of our results are also applicable when $\mathcal{X}$ and $\mathcal{Y}$ are balls in $\mathbb{R}^d$, as we make clear in the sequel. A point $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an equilibrium if

$$
\begin{aligned}
\boldsymbol{x}^\top \mathbf{A} \boldsymbol{y}^* &\leq (\boldsymbol{x}^*)^\top \mathbf{A} \boldsymbol{y}^*, \quad \forall \boldsymbol{x} \in \mathcal{X}; \\
(\boldsymbol{x}^*)^\top \mathbf{B} \boldsymbol{y} &\leq (\boldsymbol{x}^*)^\top \mathbf{B} \boldsymbol{y}^*, \quad \forall \boldsymbol{y} \in \mathcal{Y}.
\end{aligned}
\tag{6}
$$

Even when $\mathbf{B} = -\mathbf{A}$, this seemingly simple saddle-point problem is—with the addition of appropriate regularization—powerful enough to capture problems such as linear regression, empirical risk minimization, and robust optimization (Du & Hu, 2019). Further, studying such games relates to the *local convergence* of complex games encountered in practical applications (Liang & Stokes, 2019). We also point out that our techniques can address the addition of quadratic regularization. In this regime, our first contribution is to extend the known regime for which OGD retains stability:

Theorem 5.1 (Abridged; Full Version in Theorem C.3). *Suppose that the matrix $\mathbf{A}^\top \mathbf{B}$ has strictly negative (real) eigenvalues. Then, for a sufficiently small learning rate $\eta > 0$, (OGD) converges linearly to an equilibrium.*

The proof of this theorem is based on transforming the dynamics of (OGD) to the *frequency domain* via the Z-transform, in order to derive the characteristic equation of the dynamical system. When the game is zero-sum, the condition of the theorem holds since the matrix $-\mathbf{A}^\top \mathbf{A}$ is symmetric and negative definite. As such, Theorem 5.1 extends the known results in the literature. The technique we employ can also reveal the rate of convergence in terms of the eigenvalues of $\mathbf{A}^\top \mathbf{B}$. The first question stemming from Theorem 5.1 is whether the condition of stability captures games which are, in some sense, fully competitive. Our next result answers this question in the negative.

Indeed, it turns out that the condition of stability in Theorem 5.1 also includes games with a coordination aspect, but under (OGD) the players will fail to coordinate. In turn, we show that this implies that (OGD) can be *arbitrarily inefficient*. More precisely, for strategies $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$, we define the social welfare as $\mathrm{SW}(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \boldsymbol{x}^\top \mathbf{A} \boldsymbol{y} + \boldsymbol{x}^\top \mathbf{B} \boldsymbol{y}$. We show the following result (see Figure 8 for an illustration).

Proposition 5.2. *For any sufficiently large $R > 0$, there exist games such that (OGD) converges under any initialization to an equilibrium $(\boldsymbol{x}^{(\infty)}, \boldsymbol{y}^{(\infty)})$ such that $\mathrm{SW}(\boldsymbol{x}^{(\infty)}, \boldsymbol{y}^{(\infty)}) = 0$, while there exist an equilibrium $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ with $\mathrm{SW}(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq 2R^2$.*

This holds when $\mathcal{X}$ and $\mathcal{Y}$ are compact balls on $\mathbb{R}^d$ and parameter $R > 0$ controls their radius. Proposition 5.2 seems

to suggest that the stabilizing properties of (OGD) come at a dramatic loss of efficiency beyond zero-sum games.

Another interesting implication is that an arbitrarily small perturbation from a zero-sum game can destabilize (OGD):

Proposition 5.3. *For any $\epsilon > 0$ there exists a game $(\mathbf{A}, \mathbf{B})$ with $\|\mathbf{A} + \mathbf{B}\|_F \leq \epsilon$ for which (OGD) diverges, while the dynamics converge for the game $(\mathbf{A}, -\mathbf{A})$.*

Thus, even if a game $(\mathbf{A}, \mathbf{B})$ is arbitrarily close to a zero-sum in the Frobenius norm, (OGD) may still diverge. To put it differently, a small noise in one of the payoff matrices can dramatically alter the behavior of the system. This phenomenon is illustrated and further discussed in Figure 9.

5.2 Multiplayer Games   Moreover, we also characterize the (OGD) dynamics in polymatrix games. Such multiplayer interactions have already received considerable attention in the literature on GANs (*e.g.*, see (?), and references therein), but the behavior of the dynamics is poorly understood even in structured games (Kalogiannis et al., 2021). Our next theorem characterizes the important case where a single player 1 plays against $n - 1$ different players, numbered from 2 to n. In particular, the utility of player 1 under strategies $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is given by $\sum_{j \neq i} \boldsymbol{x}_i^\top \mathbf{A}_{1,j} \boldsymbol{x}_j$, while the utility of player j is given by $\boldsymbol{x}_j^\top \mathbf{A}_{j,1} \boldsymbol{x}_1$. Our next result extends Theorem 5.1.

Theorem 5.4 (Abridged; Full Version in Theorem C.7). *If $\mathbf{M} \coloneqq \sum_{j \neq 1} \mathbf{A}_{1,j} \mathbf{A}_{j,1}$ has strictly negative (real) eigenvalues, there exists a sufficiently small learning rate $\eta > 0$ depending only on the spectrum of the matrix $\mathbf{M}$ such that (OGD) converges with linear rate to an equilibrium.*

This condition is again trivially satisfied in polymatrix zero-sum games. In fact, Theorem 5.4 is an instance of a more general characterization for polymatrix unconstrained games, as we further explain in Remark C.9.

Remark 5.5 (Beyond OGD: Stability of First-Order Methods). In light of the result established in Theorem 5.1, it is natural to ask whether its condition is necessary. In Theorem C.11 we show that the presence of positive eigenvalues in the spectrum of matrix $\mathbf{A}^\top \mathbf{B}$ necessarily implies instability even under a *generic class* of first-order methods.

6. Experiments

In this section we numerically investigated the last-iterate convergence of (OMD) in two zero-sum extensive-form games (EFGs). Recall that computing a Nash equilibrium in this setting can be expressed as the bilinear saddle-point problem of (2). When both players employ (OMD) with Euclidean regularization and $\eta \leq \frac{1}{4\|\mathbf{A}\|_2}$, where $\|\mathbf{A}\|_2$ is the spectral norm of $\mathbf{A}$, our results guarantee last-iterate convergence in terms of the saddle-point gap.

We experimented on two standard poker benchmarks known as *Kuhn poker* (Kuhn, 1950) and *Leduc poker* (Southey et al., 2005). For each benchmark game, we ran (OMD) with Euclidean regularization and three different values of η for 10000 iterations. After each iteration t, we computed the saddle point gap corresponding to the iterates at time t, as well as to the average iterates up to time t. The results are shown in Figure 1. As expected, we observe that both the average and the last iterate converge in terms of the saddle-point gap. Moreover, we see that larger values of learning rate lead to substantially faster convergence, illustrating the importance of obtaining sharp theoretical bounds.

Additional experiments related to our theoretical results have been included in Appendix D.
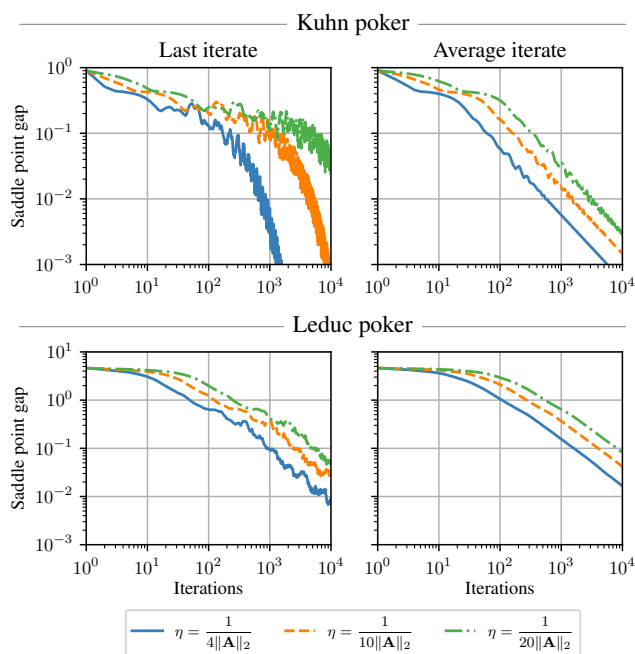

Figure 1. The saddle-point gap of the last iterate and the average iterate of (OMD) dynamics in Kuhn and Leduc poker.

Acknowledgements

References

Abernethy, J. D., Bartlett, P. L., and Hazan, E. Blackwell approachability and no-regret learning are equivalent. In *COLT 2011 - The 24th Annual Conference on Learning Theory*, volume 19 of *JMLR Proceedings*, pp. 27–46. JMLR.org, 2011.

Adler, I., Daskalakis, C., and Papadimitriou, C. H. A note on strictly competitive games. In *Internet and Network Economics, 5th International Workshop, WINE 2009*, volume 5929 of *Lecture Notes in Computer Science*, pp. 471–474. Springer, 2009.

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

Awerbuch, B., Azar, Y., and Epstein, A. The price of routing unsplittable flow. *SIAM J. Comput.*, 42(1):160–177, 2013.

Azizian, W., Iutzeler, F., Malick, J., and Mertikopoulos, P. The last-iterate convergence rate of optimistic mirror descent in stochastic variational inequalities. In *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, pp. 326–358. PMLR, 2021.

Babichenko, Y. Query complexity of approximate nash equilibria. In *Symposium on Theory of Computing, STOC 2014*, pp. 535–544. ACM, 2014.

Bailey, J. P. and Piliouras, G. Fast and furious learning in zero-sum games: Vanishing regret with non-vanishing step sizes. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 12977–12987, 2019.

Bielawski, J., Chotibut, T., Falniowski, F., Kosiorowski, G., Misiurewicz, M., and Piliouras, G. Follow-the-regularized-leader routes to chaos in routing games. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*, volume 139 of *Proceedings of Machine Learning Research*, pp. 925–935. PMLR, 2021.

Birnbaum, B. E., Devanur, N. R., and Xiao, L. Distributed algorithms via gradient descent for fisher markets. In *Proceedings 12th ACM Conference on Electronic Commerce (EC-2011)*, pp. 127–136. ACM, 2011.

Blackwell, D. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1 – 8, 1956.

Blum, A., Hajiaghayi, M., Ligett, K., and Roth, A. Regret minimization and the price of total anarchy. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, 2008*, pp. 373–382. ACM, 2008.

Cai, Y. and Daskalakis, C. On minmax theorems for multiplayer games. In *Proceedings of the Twenty-Second*

Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, pp. 217–234. SIAM, 2011.

Cai, Y., Candogan, O., Daskalakis, C., and Papadimitriou, C. Zero-sum polymatrix games: A generalization of minmax. *Mathematics of Operations Research*, 41(2): 648–655, 2016.

Candogan, O., Ozdaglar, A. E., and Parrilo, P. A. A projection framework for near-potential games. In *Proceedings of the 49th IEEE Conference on Decision and Control, CDC 2010*, pp. 244–249. IEEE, 2010.

Candogan, O., Menache, I., Ozdaglar, A. E., and Parrilo, P. A. Flows and decompositions of games: Harmonic and potential games. *Math. Oper. Res.*, 36(3):474–503, 2011.

Candogan, O., Ozdaglar, A. E., and Parrilo, P. A. Dynamics in near-potential games. *Games Econ. Behav.*, 82:66–90, 2013.

Chen, C., Surana, A., Bloch, A. M., and Rajapakse, I. Multilinear control systems theory. *SIAM Journal on Control and Optimization*, 59(1):749–776, 2021.

Chen, G. and Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

Chen, X., Deng, X., and Teng, S. Settling the complexity of computing two-player nash equilibria. *J. ACM*, 56(3): 14:1–14:57, 2009.

Cheung, Y. K. and Piliouras, G. Chaos, extremism and optimism: Volume analysis of learning in games. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020.

Cheung, Y. K. and Tao, Y. Chaos of learning beyond zero-sum and coordination via game decompositions. In *9th International Conference on Learning Representations, ICLR 2021*. OpenReview.net, 2021.

Christodoulou, G. and Koutsoupias, E. The price of anarchy of finite congestion games. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, 2005*, pp. 67–73. ACM, 2005.

Christodoulou, G., Kovács, A., and Schapira, M. Bayesian combinatorial auctions. *J. ACM*, 63(2):11:1–11:19, 2016.

Damme, E. V. *Stability and Perfection of Nash Equilibria*. Berlin: Springer-Verlag, 1987.

Daskalakis, C. and Panageas, I. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In Blum, A. (ed.), *10th Innovations in Theoretical Computer Science Conference, ITCS 2019*, volume 124 of *LIPIcs*, pp. 27:1–27:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

Daskalakis, C. and Papadimitriou, C. H. On a network generalization of the minmax theorem. In *Automata, Languages and Programming, 36th Internatilonal Colloquium, ICALP 2009*, volume 5556 of *Lecture Notes in Computer Science*, pp. 423–434. Springer, 2009.

Daskalakis, C., Fabrikant, A., and Papadimitriou, C. H. The game world is flat: The complexity of nash equilibria in succinct games. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006*, volume 4051 of *Lecture Notes in Computer Science*, pp. 513–524. Springer, 2006.

Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. The complexity of computing a nash equilibrium. *SIAM J. Comput.*, 39(1):195–259, 2009.

Daskalakis, C., Ilyas, A., Syrgkanis, V., and Zeng, H. Training gans with optimism. In *6th International Conference on Learning Representations, ICLR 2018*. OpenReview.net, 2018.

Daskalakis, C., Foster, D. J., and Golowich, N. Independent policy gradient methods for competitive reinforcement learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020l*, 2020.

Daskalakis, C., Fishelson, M., and Golowich, N. Near-optimal no-regret learning in general games. *CoRR*, abs/2108.06924, 2021.

Deng, X., Hu, X., Lin, T., and Zheng, W. Nash convergence of mean-based learning algorithms in first price auctions. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pp. 141–150. ACM, 2022.

Diakonikolas, J. and Wang, P. Potential function-based framework for making the gradients small in convex and min-max optimization. *CoRR*, abs/2101.12101, 2021.

Du, S. S. and Hu, W. Linear convergence of the primal-dual gradient method for convex-concave saddle point problems without strong convexity. In *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019*, volume 89 of *Proceedings of Machine Learning Research*, pp. 196–205. PMLR, 2019.

Eisenberg, E. and Gale, D. Consensus of Subjective Probabilities: The Pari-Mutuel Method. *The Annals of Mathematical Statistics*, 30(1):165 – 168, 1959.

Feng, Z., Guruganesh, G., Liaw, C., Mehta, A., and Sethi, A. Convergence analysis of no-regret bidding algorithms in repeated auctions. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pp. 5399–5406. AAAI Press, 2021.

Giannou, A., Vlatakis-Gkaragkounis, E., and Mertikopoulos, P. Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. In *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, pp. 2147–2148. PMLR, 2021.

Golowich, N., Pattathil, S., and Daskalakis, C. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *Advances in Neural Information Processing Systems 2020*, 2020a.

Golowich, N., Pattathil, S., Daskalakis, C., and Ozdaglar, A. E. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In *Conference on Learning Theory, COLT 2020*, volume 125 of *Proceedings of Machine Learning Research*, pp. 1758–1784. PMLR, 2020b.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., and Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 2014*, pp. 2672–2680, 2014.

Harker, P. T. and Pang, J.-S. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Math. Program.*, 48(1–3):161–220, 1990.

Hart, S. and Mas-Colell, A. Uncoupled dynamics do not lead to nash equilibrium. *The American Economic Review*, 93(5):1830–1836, 2003.

Héliou, A., Cohen, J., and Mertikopoulos, P. Learning with bandit feedback in potential games. In *Advances in Neural Information Processing Systems 30, 2017*, pp. 6369–6378, 2017.

Hsieh, Y., Antonakopoulos, K., and Mertikopoulos, P. Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. In *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, volume 134 of *Proceedings of Machine Learning Research*, pp. 2388–2422. PMLR, 2021.

Kalogiannis, F., Panageas, I., and Vlatakis-Gkaragkounis, E. Teamwork makes von neumann work: Min-max optimization in two-team zero-sum games. *CoRR*, abs/2111.04178, 2021.

Kearns, M. J., Littman, M. L., and Singh, S. P. Graphical models for game theory. In *UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, 2001*, pp. 253–260. Morgan Kaufmann, 2001.

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

Kuhn, H. W. A simplified two-person poker. In Kuhn, H. W. and Tucker, A. W. (eds.), *Contributions to the Theory of Games*, volume 1 of *Annals of Mathematics Studies, 24*, pp. 97–103. Princeton University Press, Princeton, New Jersey, 1950.

Lee, C.-W., Kroer, C., and Luo, H. Last-iterate convergence in extensive-form games. In *NeurIPS*, 2021.

Leonardos, S., Overman, W., Panageas, I., and Piliouras, G. Global convergence of multi-agent policy gradient in markov potential games. *ICLR*, 2022.

Liang, T. and Stokes, J. Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks. In *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019*, volume 89 of *Proceedings of Machine Learning Research*, pp. 907–915. PMLR, 2019.

Lin, T., Zhou, Z., Mertikopoulos, P., and Jordan, M. I. Finite-time last-iterate convergence for multi-agent learning in games. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6161–6171. PMLR, 2020.

McMahan, H. B. Follow-the-regularized-leader and mirror descent: Equivalence theorems and L1 regularization. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011*, volume 15 of *JMLR Proceedings*, pp. 525–533. JMLR.org, 2011.

Mertikopoulos, P. and Zhou, Z. Learning in games with continuous action sets and unknown payoff functions. *Math. Program.*, 173(1-2):465–507, 2019.

Mertikopoulos, P., Papadimitriou, C. H., and Piliouras, G. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018*, pp. 2703–2717. SIAM, 2018.

Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C., Chandrasekhar, V., and Piliouras, G. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *7th International Conference on Learning Representations, ICLR 2019*. OpenReview.net, 2019.

Milgrom, P. and Roberts, J. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58(6):1255–1277, 1990.

Mokhtari, A., Ozdaglar, A., and Pattathil, S. Convergence rate of $\mathcal{O}(1/k)$ for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 30:3230–3251, 2020a.

Mokhtari, A., Ozdaglar, A. E., and Pattathil, S. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1497–1507. PMLR, 2020b.

Monderer, D. and Shapley, L. S. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.

Moulin, H. and Vial, J. Strategically zero-sum games: The class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory*, 1978.

Palaiopanos, G., Panageas, I., and Piliouras, G. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pp. 5872–5882, 2017.

Panageas, I., Tröbst, T., and Vazirani, V. V. Combinatorial algorithms for matching markets via nash bargaining: One-sided, two-sided and non-bipartite. *CoRR*, abs/2106.02024, 2021.

Popov, L. D. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical notes of the Academy of Sciences of the USSR*, 28:845–848, 1980.

Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In *COLT 2013 - The 26th Annual Conference on Learning Theory*, volume 30 of *JMLR Workshop and Conference Proceedings*, pp. 993–1019. JMLR.org, 2013.

Rockafellar, R. T. *Convex analysis*. Princeton University Press, Princeton, N. J., 1970.

Romanovskii, I. Reduction of a game with complete memory to a matrix game. *Soviet Mathematics*, 3, 1962.

Rosenthal, R. W. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

Roughgarden, T. Intrinsic robustness of the price of anarchy. *J. ACM*, 62(5):32:1–32:42, 2015.

Rubinstein, A. Settling the complexity of computing approximate two-player nash equilibria. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016*, pp. 258–265. IEEE Computer Society, 2016.

Sandholm, W. H. *Population Games and Evolutionary Dynamics*. MIT Press, 2010.

Sato, Y., Akiyama, E., and Farmer, J. D. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7):4748–4751, 2002.

Shalev-Shwartz, S. Online learning and online convex optimization. *Found. Trends Mach. Learn.*, 4(2):107–194, 2012.

Shapley, L. S. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.

Shmyrev, V. An algorithm for finding equilibrium in the linear exchange model with fixed budgets. *Journal of Applied and Industrial Mathematics*, 3:505–518, 10 2009. doi: 10.1134/S1990478909040097.

Sion, M. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171 – 176, 1958.

Southey, F., Bowling, M. H., Larson, B., Piccione, C., Burch, N., Billings, D., and Rayner, D. C. Bayes? bluff: Opponent modelling in poker. In *UAI '05, Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence*, pp. 550–558. AUAI Press, 2005.

Syrgkanis, V., Agarwal, A., Luo, H., and Schapire, R. E. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, pp. 2989–2997, 2015.

Vetta, A. Nash equilibria in competitive societies, with applications to facility location, traffic routing and auctions. In *43rd Symposium on Foundations of Computer Science (FOCS 2002)*, pp. 416. IEEE Computer Society, 2002.

Vlatakis-Gkaragkounis, E., Flokas, L., Lianeas, T., Mertikopoulos, P., and Piliouras, G. No-regret learning and mixed nash equilibria: They do not mix. In *Advances in Neural Information Processing Systems 33 2020l*, 2020.

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

Wei, C., Lee, C., Zhang, M., and Luo, H. Last-iterate convergence of decentralized optimistic gradient descent/ascent in infinite-horizon competitive markov games. In *Conference on Learning Theory, COLT 2021*, volume 134 of

Proceedings of Machine Learning Research, pp. 4259–4299. PMLR, 2021a.

Wei, C., Lee, C., Zhang, M., and Luo, H. Linear last-iterate convergence in constrained saddle-point optimization. In *9th International Conference on Learning Representations, ICLR 2021*. OpenReview.net, 2021b.

Wu, F. and Zhang, L. Proportional response dynamics leads to market equilibrium. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 2007*, pp. 354–363. ACM, 2007.

Zhang, G. and Yu, Y. Convergence of gradient methods on bilinear zero-sum games. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

Zhang, L. Proportional response dynamics in the fisher market. *Theor. Comput. Sci.*, 412(24):2691–2698, 2011.

Zhou, Z., Mertikopoulos, P., Moustakas, A. L., Bambos, N., and Glynn, P. W. Mirror descent learning in continuous games. In *56th IEEE Annual Conference on Decision and Control, CDC 2017, Melbourne, Australia, December 12-15, 2017*, pp. 5776–5783. IEEE, 2017.

Zhou, Z., Mertikopoulos, P., Athey, S., Bambos, N., Glynn, P. W., and Ye, Y. Learning in games with lossy feedback. In *Advances in Neural Information Processing Systems 2018,*, pp. 5140–5150, 2018.

A. Proofs from Section 3

In this section we give complete proofs for our results in Section 3. For the convenience of the reader we will restate the claims before proceeding with their proofs. We begin with Theorem 3.1.

Theorem 3.1. *Suppose that every player $i \in [n]$ employs a regret minimizing algorithm satisfying the $(\alpha_i, \beta_i, \gamma_i)$-RVU property with $\gamma_i \geq 2(n-1) \sum_{j \neq i} \beta_j$, for all $i \in [n]$, under the pair of dual norms $(\|\cdot\|_1, \|\cdot\|_\infty)$. If the game is such that $\sum_{i=1}^n \text{Reg}_i^T \geq 0$, for any $T \in \mathbb{N}$, it holds that*

$$\sum_{i=1}^n \sum_{t=1}^T \gamma_i \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq 2 \sum_{i=1}^n \alpha_i. \tag{1}$$

Proof. The proof commences similarly to (Syrgkanis et al., 2015, Theorem 4). By assumption, we know that for any player $i \in [n]$ it holds that

$$\text{Reg}_i^T \leq \alpha_i + \beta_i \sum_{t=1}^T \|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty^2 - \gamma_i \sum_{t=1}^T \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2. \tag{7}$$

Furthermore, the middle term in the RVU property can be bounded using the following simple claim.

Claim A.1. *For any player $i \in [n]$ it holds that*

$$\|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty \leq \sum_{j \neq i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_1.$$

Proof. By the normalization assumption on the utilities we know that $|\boldsymbol{u}_i(\cdot)| \leq 1$. Thus, from the triangle inequality it follows that

$$\|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty \leq \sum_{\boldsymbol{a}_{-i} \in \mathcal{A}_{-i}} \left| \prod_{j \neq i} \boldsymbol{x}_j^{(t)}(a_j) - \prod_{j \neq i} \boldsymbol{x}_j^{(t-1)}(a_j) \right|.$$

The induced term can be recognized as (two times) the total variation distance of two product distributions. Hence, a standard application of the sum-product inequality implies that

$$\sum_{\boldsymbol{a}_{-i} \in \mathcal{A}_{-i}} \left| \prod_{j \neq i} \boldsymbol{x}_j^{(t)}(a_j) - \prod_{j \neq i} \boldsymbol{x}_j^{(t-1)}(a_j) \right| \leq \sum_{j \neq i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_1.$$

$\square$

Moreover, a direct application of Young's inequality to the bound of the previous claim gives us that

$$\|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty^2 \leq (n-1) \sum_{j \neq i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_1^2. \tag{8}$$

As a result, if we plug-in this bound to (7) we may conclude that for all players $i \in [n]$ it holds that

$$\text{Reg}_i^T \leq \alpha_i + (n-1)\beta_i \sum_{t=1}^T \sum_{j \neq i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_1^2 - \gamma_i \sum_{t=1}^T \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2.$$

Summing these inequalities for all players $i \in [n]$ yields that

$$\sum_{i=1}^n \text{Reg}_i^T \leq \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \left((n-1) \sum_{j \neq i} \beta_j - \gamma_i \right) \sum_{t=1}^T \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \gamma_i \sum_{t=1}^T \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2,$$

where we used the assumption that $\gamma_i \geq 2(n-1) \sum_{j \neq i} \beta_j$ for any $i \in [n]$. Finally, the theorem follows from a rearrangement of the final bound, using the assumption that $\sum_{i=1}^n \text{Reg}_i^T \geq 0$. $\square$

Proposition A.2. *Suppose that each player employs OFTRL with the same learning rate $\eta > 0$. Moreover, suppose that the game is such that $\sum_{i=1}^{n} \text{Reg}_i^T \geq 0$. Then,*

1. *If all players use H-step recency bias (Item 1) and $\eta \leq \frac{1}{4(n-1)H}$ it holds that*

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq 8 \sum_{i=1}^{n} \Omega_i.$$

2. *If all players use geometrically δ-discounted recency bias (Item 2) and $\eta \leq \frac{(1-\delta)^{3/2}}{4(n-1)}$ it holds that*

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq 16 \sum_{i=1}^{n} \Omega_i.$$

Proof. Let us first establish Item 1. We will use the following RVU bound shown by Syrgkanis et al. (2015, Proposition 9).

Proposition A.3 (Syrgkanis et al., 2015). *OFTRL with learning rate $\eta > 0$ and H-step recency bias (Item 1) satisfies the RVU property with $(\alpha, \beta, \gamma) = (\frac{\Omega}{\eta}, \eta H^2, \frac{1}{4\eta})$, where $\Omega := \sup_{\boldsymbol{x} \in \mathcal{X}} \mathcal{R}(\boldsymbol{x}) - \inf_{\boldsymbol{x} \in \mathcal{X}} \mathcal{R}(\boldsymbol{x})$.*

As a result, we may conclude from this bound that the regret of each player $i \in [n]$ can be bounded as

$$\text{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \eta H^2 \sum_{t=1}^{T} \|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty^2 - \frac{1}{4\eta} \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2,$$

where $\Omega_i := \sup_{\boldsymbol{x} \in \mathcal{X}_i} \mathcal{R}(\boldsymbol{x}) - \inf_{\boldsymbol{x} \in \mathcal{X}} \mathcal{R}(\boldsymbol{x})$. Next, using Claim A.1 the previous bound implies that

$$\text{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \eta(n-1) H^2 \sum_{t=1}^{T} \sum_{j \neq i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_1^2 - \frac{1}{4\eta} \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2.$$

Summing these bounds for all players $i \in [n]$ gives us that

$$\sum_{i=1}^{n} \text{Reg}_i^T \leq \frac{1}{\eta} \sum_{i=1}^{n} \Omega_i + \sum_{i=1}^{n} \left(\eta(n-1)^2 H^2 - \frac{1}{4\eta} \right) \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq \frac{1}{\eta} \sum_{i=1}^{n} \Omega_i - \frac{1}{8\eta} \sum_{t=1}^{T} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2,$$

for $\eta \leq \frac{1}{4(n-1)H}$. Thus, Item 1 follows immediately under the assumption that $\sum_{i=1}^{n} \text{Reg}_i^T \geq 0$. Next, we proceed with the proof of Item 2. We use the following RVU bound shown by Syrgkanis et al. (2015, Proposition 10).

Proposition A.4 (Syrgkanis et al., 2015). *OFTRL with learning rate $\eta > 0$ and geometrically δ-discounted recency bias (Item 2) satisfies the RVU property with $(\alpha, \beta, \gamma) = (\frac{\Omega}{\eta}, \frac{\eta}{(1-\delta)^3}, \frac{1}{8\eta})$, where $\Omega := \sup_{\boldsymbol{x} \in \mathcal{X}} \mathcal{R}(\boldsymbol{x}) - \inf_{\boldsymbol{x} \in \mathcal{X}} \mathcal{R}(\boldsymbol{x})$.*

Thus, if player $i \in [n]$ uses geometrically δ-discounted recency bias it follows that its regret can be bounded as

$$\text{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \frac{\eta}{(1-\delta)^3} \sum_{t=1}^{T} \|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2.$$

Combining this bound with Claim A.1 yields that

$$\text{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \frac{\eta(n-1)}{(1-\delta)^3} \sum_{j \neq i} \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2.$$

Summing these bounds for all players $i \in [n]$ gives us that

$$\sum_{i=1}^{n} \text{Reg}_i^T \leq \frac{1}{\eta} \sum_{i=1}^{n} \Omega_i + \sum_{i=1}^{n} \left(\frac{\eta(n-1)^2}{(1-\delta)^3} - \frac{1}{8\eta} \right) \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq \frac{1}{\eta} \sum_{i=1}^{n} \Omega_i - \frac{1}{16\eta} \sum_{t=1}^{T} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2,$$

for $\eta \leq \frac{(1-\delta)^{3/2}}{4(n-1)}$. Finally, the fact that $\sum_{i=1}^{n} \text{Reg}_i^T \geq 0$ along with a rearrangement of the previous bound concludes the proof. $\square$

Next, we proceed with the proof of Proposition 3.2. First, we state some preliminary facts about strategically zero-sum games, a subclass of bimatrix games. It should be stressed that bimatrix games is already a very general class of games. For example, Daskalakis et al. (2006) have shown that for *succinctly representable* multiplayer games, the problem of computing Nash equilibria can be reduced to the two-player case. Below we give the formal definition of a strategically zero-sum game.

Definition A.5 (Strategically Zero-Sum Games (Moulin & Vial, 1978))**.** The bimatrix games $(\mathbf{A}, \mathbf{B})$ and $(\mathbf{C}, \mathbf{D})$, defined on the same action space, are *strategically equivalent* if for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ and $\boldsymbol{y}, \boldsymbol{y}' \in \mathcal{Y}$,

$$\boldsymbol{x}^\top \mathbf{A} \boldsymbol{y} \geq (\boldsymbol{x}')^\top \mathbf{A} \boldsymbol{y} \iff \boldsymbol{x}^\top \mathbf{C} \boldsymbol{y} \geq (\boldsymbol{x}')^\top \mathbf{C} \boldsymbol{y};$$
$$\boldsymbol{x}^\top \mathbf{B} \boldsymbol{y} \geq \boldsymbol{x}^\top \mathbf{B} \boldsymbol{y}' \iff \boldsymbol{x}^\top \mathbf{D} \boldsymbol{y} \geq \boldsymbol{x}^\top \mathbf{D} \boldsymbol{y}'.$$

A game is *strategically zero-sum* if it is strategically equivalent to some zero-sum game.

In words, consider a pair of strategies $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$. Player $\mathcal{X}$ can order her strategy space based on the (expected) payoff if player $\mathcal{Y}$ was to play $\boldsymbol{y}$, and similarly, player $\mathcal{Y}$ can order her strategy space under the assumption that player $\mathcal{X}$ will play $\boldsymbol{x}$. In strategically equivalent games this ordering is preserved.

Without any loss, we will consider *non-trivial games* in the sense of (Moulin & Vial, 1978, Definition 2). The following characterization (Moulin & Vial, 1978, Theorem 2) will be crucial for the proof of Proposition 3.2.

Theorem A.6 (Moulin & Vial, 1978)**.** *Let $(\mathbf{A}, \mathbf{B})$ be a non-trivial $n \times m$ strategically zero-sum game. Then, there exist $\lambda > 0, \mu > 0$ and a compatible matrix $\mathbf{C}$ such that*

- $\mathbf{A} = \lambda \mathbf{C} + \boldsymbol{v}_a \mathbf{1}_m^\top$, *for some $\boldsymbol{v}_a \in \mathbb{R}^n$;*

- $\mathbf{B} = -\mu \mathbf{C} + \mathbf{1}_n \boldsymbol{v}_b^\top$, *for some $\boldsymbol{v}_b \in \mathbb{R}^m$.*

Here we used the notation $\mathbf{1}_n$ to denote the all-ones vector in $\mathbb{R}^n$. The converse of Theorem A.6 also holds. It is also worth pointing out strategically zero-sum games can be "embedded" in a zero-sum game with imperfect information (Moulin & Vial, 1978).

Remark A.7. For the purpose of Proposition 3.2 (and Proposition A.10) it will be assumed that $\lambda = \mu$, neutralizing any "scale imbalances" in the payoff matrices of the two players. We remark that if this is not the case, our result for strategically zero-sum games (Item 4) still holds by considering an appropriately *weighted* sum of regrets. In this case one should select different learning rates for the two players in order to cancel out the underlying scale imbalance. Nonetheless, such an extension does not appear to work for polymatrix strategically zero-sum games (Item 5). The latter fact could be partly justified by the surprising result that even polymatrix *strictly competitive games* are PPAD-hard (Cai & Daskalakis, 2011, Theorem 1.2).

Remark A.8 (Strictly Competitive Games). Strictly competitive games form a subclass of strategically zero-sum games. More precisely, a two-person game is *strictly competitive* if it has the following property: if both players change their mixed strategies, then either there is no change in the expected payoffs, or one of the two expected payoffs increases and the other decreases. It was formally shown by Adler et al. (2009) that a game $(\mathbf{A}, \mathbf{B})$ is strictly competitive if and only if $\mathbf{B}$ is an *affine variant* of matrix $\mathbf{A}$; that is, $\mathbf{B} = \lambda \mathbf{A} + \mu \mathbf{1}$, where $\lambda > 0$, μ is unrestricted, and $\mathbf{1}$ denotes the all-ones matrix.

Before we proceed with the proof of Proposition 3.2, let us first make the following simple observation.

Observation A.9. *FTRL and MD, and optimistic variants thereof, employed on a strategically zero-sum game with suitable learning rates are equivalent to the dynamics in a "hidden" zero-sum game.*

Thus, this observation reassures us that OFTRL and OMD dynamics in strategically zero-sum games inherit all of the favorable last-iterate convergence guarantees known for zero-sum games. We are now ready to prove Proposition 3.2, the extended version of which is included below.

Proposition A.10. *Full Version of Proposition 3.2 For the following classes of games it holds that $\sum_{i=1}^n \mathrm{Reg}_i^T \geq 0$:*

1. *Two-player zero-sum games;*
2. *Polymatrix zero-sum games;*
3. *Constant-sum Polymatrix games;*
4. *Strategically zero-sum games;*[5]

---

[5]See Remark A.7.

5. *Polymatrix strategically zero-sum games;*[5]
6. *Convex-concave (zero-sum) games;*
7. *Zero-sum games with objective $f(\boldsymbol{x}, \boldsymbol{y})$ such that*

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} f(\boldsymbol{x}, \boldsymbol{y}) = \max_{\boldsymbol{y} \in \mathcal{Y}} \min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}, \boldsymbol{y});$$

8. *Quasiconvex-quasiconcave (zero-sum) games;*
9. *Zero-sum stochastic games.*

Proof. We proceed separately for each class of games.

- For Item 1 let us denoted by $\mathbf{A}$ the matrix associated with the zero-sum game, so that player $\mathcal{X}$ is the "minimizer" and player $\mathcal{Y}$ is the "maximizer". Further, let us denote by $\bar{\boldsymbol{x}} := (\boldsymbol{x}^{(1)} + \dots \boldsymbol{x}^{(T)})/T$ and $\bar{\boldsymbol{y}} := (\boldsymbol{y}^{(1)} + \dots \boldsymbol{y}^{(T)})/T$ the time average of the strategies employed by the two players respectively up to time $T \in \mathbb{N}$. Then, we have that

$$\operatorname{Reg}_{\mathcal{X}}^T + \operatorname{Reg}_{\mathcal{Y}}^T = \max_{\boldsymbol{x}^* \in \mathcal{X}} \left\langle \boldsymbol{x}^*, -\mathbf{A} \sum_{t=1}^T \boldsymbol{y}^{(t)} \right\rangle + \sum_{t=1}^T \langle \boldsymbol{x}^{(t)}, \mathbf{A} \boldsymbol{y}^{(t)} \rangle + \max_{\boldsymbol{y}^* \in \mathcal{Y}} \left\langle \boldsymbol{y}^*, \mathbf{A}^\top \sum_{t=1}^T \boldsymbol{x}^{(t)} \right\rangle - \sum_{t=1}^T \langle \boldsymbol{y}^{(t)}, \mathbf{A}^\top \boldsymbol{x}^{(t)} \rangle$$

$$= T \left(\max_{\boldsymbol{y}^* \in \mathcal{Y}} \langle \boldsymbol{y}^*, \mathbf{A}^\top \bar{\boldsymbol{x}} \rangle - \min_{\boldsymbol{x}^* \in \mathcal{X}} \langle \boldsymbol{x}^*, \mathbf{A} \bar{\boldsymbol{y}} \rangle \right) \geq T \left(\langle \bar{\boldsymbol{y}}, \mathbf{A}^\top \bar{\boldsymbol{x}} \rangle - \langle \bar{\boldsymbol{x}}, \mathbf{A} \bar{\boldsymbol{y}} \rangle \right) = 0.$$

- For Item 2 let us denoted by $\mathcal{N}_i \subseteq [n]$ the neighborhood of the node associated with each player $i \in [n]$ so that the (expected) utility of player $i \in [n]$ under a joint (mixed) strategy vector $\boldsymbol{x} := (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ can be expressed as $\sum_{j \in \mathcal{N}_i} \boldsymbol{x}_i^\top \mathbf{A}_{i,j} \boldsymbol{x}_j$, where $\mathbf{A}_{i,j}$ is the payoff matrix associated with the edge $i \to j$. Observe that since every edge corresponds to a zero-sum game, it holds that $\mathbf{A}_{i,j} = -\mathbf{A}_{j,i}^\top$. Moreover, we let $\bar{\boldsymbol{x}}_i := (\boldsymbol{x}_i^{(1)} + \dots + \boldsymbol{x}_i^{(T)})/T$ be the time average of player i's strategies up to time T. We have that

$$\frac{1}{T} \sum_{i=1}^n \operatorname{Reg}_i^T = \frac{1}{T} \sum_{i=1}^n \left(\max_{\boldsymbol{x}_i^* \in \mathcal{X}_i} \left\{ \left\langle \boldsymbol{x}_i^*, \sum_{t=1}^T \sum_{j \in \mathcal{N}_i} \mathbf{A}_{i,j} \boldsymbol{x}_j^{(t)} \right\rangle \right\} - \sum_{t=1}^T \left\langle \boldsymbol{x}_i^{(t)}, \sum_{j \in \mathcal{N}_i} \mathbf{A}_{i,j} \boldsymbol{x}_j^{(t)} \right\rangle \right)$$

$$= \sum_{i=1}^n \max_{\boldsymbol{x}_i^* \in \mathcal{X}_i} \left\{ \left\langle \boldsymbol{x}_i^*, \sum_{j \in \mathcal{N}_i} \mathbf{A}_{i,j} \bar{\boldsymbol{x}}_j \right\rangle \right\} - \frac{1}{T} \sum_{i=1}^n \sum_{t=1}^T \left\langle \boldsymbol{x}_i^{(t)}, \sum_{j \in \mathcal{N}_i} \mathbf{A}_{i,j} \boldsymbol{x}_j^{(t)} \right\rangle$$

$$= \sum_{i=1}^n \max_{\boldsymbol{x}_i^* \in \mathcal{X}_i} \left\{ \left\langle \boldsymbol{x}_i^*, \sum_{j \in \mathcal{N}_i} \mathbf{A}_{i,j} \bar{\boldsymbol{x}}_j \right\rangle \right\} \tag{9}$$

$$\geq \sum_{i=1}^n \left\langle \bar{\boldsymbol{x}}_i, \sum_{j \in \mathcal{N}_i} \mathbf{A}_{i,j} \bar{\boldsymbol{x}}_j \right\rangle = 0, \tag{10}$$

where (9) and (10) follow from the fact that $\mathbf{A}_{i,j} = -\mathbf{A}_{j,i}^\top$ for any $i \neq j$.

- For Item 3 the proof is analogous to Item 2 using the assumption that

$$\sum_{i=1}^n u_i(\boldsymbol{x}) = \sum_{i=1}^n \left\langle \boldsymbol{x}_i, \sum_{j \in \mathcal{N}_i} \mathbf{A}_{i,j} \boldsymbol{x}_j \right\rangle = C,$$

for any $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \in \prod_{i \in [n]} \Delta(\mathcal{A}_i)$, where C is some arbitrary constant.

- For Item 4 we let $(\mathbf{A}, \mathbf{B})$ be the underlying (bimatrix) strategically zero-sum game, with $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times m}$. We will use Theorem A.6 under the assumption that $\lambda = \mu$; see Remark A.7. That is, we know that there exists a (compatible) matrix $\mathbf{C}$ and $\lambda > 0$ such that $\mathbf{A} = \lambda \mathbf{C} + \boldsymbol{v}_a \mathbf{1}_m^\top$ and $\mathbf{B} = -\lambda \mathbf{C} + \mathbf{1}_n \boldsymbol{v}_b^\top$, where $\boldsymbol{v}_a \in \mathbb{R}^n$ and $\boldsymbol{v}_b \in \mathbb{R}^m$. Similarly to

the proof for Item 1, we have that

$$
\begin{aligned}
\mathrm{Reg}_{\mathcal{X}}^{T} + \mathrm{Reg}_{\mathcal{Y}}^{T} &= \max_{\boldsymbol{x}^{*}\in\Delta^{n}} \left\langle \boldsymbol{x}^{*}, \mathbf{A}\sum_{t=1}^{T}\boldsymbol{y}^{(t)} \right\rangle - \sum_{t=1}^{T}\langle \boldsymbol{x}^{(t)}, \mathbf{A}\boldsymbol{y}^{(t)}\rangle + \max_{\boldsymbol{y}^{*}\in\Delta^{m}} \left\langle \boldsymbol{y}^{*}, \mathbf{B}^{\top}\sum_{t=1}^{T}\boldsymbol{x}^{(t)} \right\rangle - \sum_{t=1}^{T}\langle \boldsymbol{y}^{(t)}, \mathbf{B}^{\top}\boldsymbol{x}^{(t)}\rangle
\end{aligned}
$$

$$
= \max_{\boldsymbol{x}^{*}\in\Delta^{n}} \left\{ \lambda \left\langle \boldsymbol{x}^{*}, \mathbf{C}\sum_{t=1}^{T}\boldsymbol{y}^{(t)} \right\rangle + T\langle \boldsymbol{x}^{*}, \boldsymbol{v}_{a}\rangle \right\} - \lambda\sum_{t=1}^{T}\left\langle \boldsymbol{x}^{(t)}, \mathbf{C}\boldsymbol{y}^{(t)} \right\rangle - \sum_{t=1}^{T}\left\langle \boldsymbol{x}^{(t)}, \boldsymbol{v}_{a} \right\rangle \quad (11)
$$

$$
+ \max_{\boldsymbol{y}^{*}\in\Delta^{m}} \left\{ -\lambda \left\langle \boldsymbol{y}^{*}, \mathbf{C}^{\top}\sum_{t=1}^{T}\boldsymbol{x}^{(t)} \right\rangle + T\langle \boldsymbol{y}^{*}, \boldsymbol{v}_{b}\rangle \right\} + \lambda\sum_{t=1}^{T}\left\langle \boldsymbol{y}^{(t)}, \mathbf{C}^{\top}\boldsymbol{x}^{(t)} \right\rangle - \sum_{t=1}^{T}\left\langle \boldsymbol{y}^{(t)}, \boldsymbol{v}_{b} \right\rangle \quad (12)
$$

$$
= T\left(\max_{\boldsymbol{x}^{*}\in\Delta^{n}} \{\lambda\langle \boldsymbol{x}^{*}, \mathbf{C}\bar{\boldsymbol{y}}\rangle + \langle \boldsymbol{x}^{*}, \boldsymbol{v}_{a}\rangle\} - \langle \bar{\boldsymbol{x}}, \boldsymbol{v}_{a}\rangle + \max_{\boldsymbol{y}^{*}\in\Delta^{m}} \{-\lambda\langle \boldsymbol{y}^{*}, \mathbf{C}^{\top}\bar{\boldsymbol{x}}\rangle + \langle \boldsymbol{y}^{*}, \boldsymbol{v}_{b}\rangle\} - \langle \bar{\boldsymbol{y}}, \boldsymbol{v}_{b}\rangle \right)
$$

$$
\geq T(\lambda\langle \bar{\boldsymbol{x}}, \mathbf{C}\bar{\boldsymbol{y}}\rangle - \lambda\langle \bar{\boldsymbol{y}}, \mathbf{C}^{\top}\bar{\boldsymbol{x}}\rangle) = 0.
$$

where in (11) we used the fact that $\mathbf{A} = \lambda\mathbf{C} + \boldsymbol{v}_{a}\mathbf{1}_{m}^{\top}$ and that $\mathbf{1}_{m}^{\top}\boldsymbol{y}^{(t)} = 1$ since $\boldsymbol{y}^{(t)} \in \Delta^{m}$, and (12) follows given that $\mathbf{B} = -\lambda\mathbf{C} + \boldsymbol{v}_{b}\mathbf{1}_{m}^{\top}$ and that $\mathbf{1}_{n}^{\top}\boldsymbol{x}^{(t)} = 1$ since $\boldsymbol{x}^{(t)} \in \Delta^{n}$.

- For Item 5 the proof follows analogously to Item 2 and Item 4; see Remark A.7.

- For Item 6 let $f(\boldsymbol{x}, \boldsymbol{y})$ be any *convex-concave function*; that is, $f(\boldsymbol{x}, \boldsymbol{y})$ is convex with respect to $\boldsymbol{x} \in \mathcal{X}$ for any fixed $\boldsymbol{y} \in \mathcal{Y}$ and concave with respect to $\boldsymbol{y} \in \mathcal{Y}$ for any fixed $\boldsymbol{x} \in \mathcal{X}$. Moreover, let us assume that player $\mathcal{X}$ is the "minimizer" and player $\mathcal{X}$ the "maximizer". We have that

$$
\begin{aligned}
\mathrm{Reg}_{\mathcal{X}}^{T} + \mathrm{Reg}_{\mathcal{Y}}^{T} &= -\min_{\boldsymbol{x}^{*}\in\mathcal{X}} \left\{ \sum_{t=1}^{T} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{(t)}) \right\} + \sum_{t=1}^{T} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) + \max_{\boldsymbol{y}^{*}\in\mathcal{Y}} \left\{ \sum_{t=1}^{T} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{*}) \right\} - \sum_{t=1}^{T} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) \\
&\geq -\sum_{t=1}^{T} f(\bar{\boldsymbol{x}}, \boldsymbol{y}^{(t)}) + \sum_{t=1}^{T} f(\boldsymbol{x}^{(t)}, \bar{\boldsymbol{y}}) \geq -Tf(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) + Tf(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) = 0,
\end{aligned}
$$

where the last inequality follows from Jensen's inequality, applicable since $f(\boldsymbol{x}, \boldsymbol{y})$ is convex-concave.

- For Item 7 we assume that $\mathcal{X}$ and $\mathcal{Y}$ are convex and compact sets. Then, we have that we have that

$$
\begin{aligned}
\frac{\mathrm{Reg}_{\mathcal{X}}^{T} + \mathrm{Reg}_{\mathcal{Y}}^{T}}{T} &= \max_{\boldsymbol{y}^{*}\in\mathcal{Y}} \left\{ \frac{1}{T}\sum_{t=1}^{T} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{*}) \right\} - \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \frac{1}{T}\sum_{t=1}^{T} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{(t)}) \right\} \\
&\geq \max_{\boldsymbol{y}^{*}\in\mathcal{Y}} \min_{\boldsymbol{x}^{*}\in\mathcal{X}} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{*}) - \min_{\boldsymbol{x}^{*}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{*}) = 0, \quad (13)
\end{aligned}
$$

where (13) uses the following inequalities:

$$
\frac{f(\boldsymbol{x}^{(1)}, \boldsymbol{y}^{*}) + \cdots + f(\boldsymbol{x}^{(T)}, \boldsymbol{y}^{*})}{T} \geq \min_{\boldsymbol{x}^{*}\in\mathcal{X}} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{*});
$$

$$
\frac{f(\boldsymbol{x}^{*}, \boldsymbol{y}^{(1)}) + \cdots + f(\boldsymbol{x}^{*}, \boldsymbol{y}^{(T)})}{T} \leq \max_{\boldsymbol{y}^{*}\in\mathcal{Y}} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{*}).
$$

- For Item 8 it is assumed that $f(\boldsymbol{x}, \boldsymbol{y})$ is lower semicontinuous and quasiconvex with respect to $\boldsymbol{x} \in \mathcal{X}$ for any fixed $\boldsymbol{y} \in \mathcal{Y}$, and upper semicontinuous and quasiconcave with respect to $\boldsymbol{y} \in \mathcal{Y}$ for any fixed $\boldsymbol{x} \in \mathcal{X}$, where $\mathcal{X}$ and $\mathcal{Y}$ are convex and compact sets. By Sion's minimax theorem (Sion, 1958) we know that

$$
\max_{\boldsymbol{y}^{*}\in\mathcal{Y}} \min_{\boldsymbol{x}^{*}\in\mathcal{X}} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{*}) = \min_{\boldsymbol{x}^{*}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x}^{*}, \boldsymbol{y}^{*}),
$$

and the conclusion follows from Item 7.

- For Item 9 the claim follows directly by Item 7 by virtue of Shapley's theorem (Shapley, 1953).

□

Remark A.11 (Duality Gap). Consider a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Weak duality implies that

$$\min_{\boldsymbol{x}^* \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} f(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq \max_{\boldsymbol{y}^* \in \mathcal{Y}} \min_{\boldsymbol{x}^* \in \mathcal{X}} f(\boldsymbol{x}^*, \boldsymbol{y}^*).$$

The value $\text{gap} := \min_{\boldsymbol{x}^* \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} f(\boldsymbol{x}^*, \boldsymbol{y}^*) - \max_{\boldsymbol{y}^* \in \mathcal{Y}} \min_{\boldsymbol{x}^* \in \mathcal{X}} f(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq 0$ is called the *duality gap* of f. Analogously to the proof of Proposition 3.2 (Item 7) we may conclude that

$$\text{Reg}_{\mathcal{X}}^T + \text{Reg}_{\mathcal{Y}}^T \geq -\text{gap } T.$$

Now observe that if we relax the condition of Theorem 3.1 so that $\sum_{i=1}^n \text{Reg}_i^T \geq -\text{gap } T$, then for $\alpha_i = \alpha$, $\beta_i = \beta$, and $\gamma_i = \gamma$ it follows that

$$\sum_{t=1}^T \sum_{i=1}^n \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq \frac{2n\alpha}{\gamma} + \frac{2\,\text{gap}}{\gamma}T.$$

This observation, along with the argument of Theorem 3.4, imply that for suitable regularizers and for a sufficiently large T there will exist a joint strategy vector $\boldsymbol{x}^{(t)}$ with $t \in [T]$ which is an $O(\sqrt{\text{gap}})$-Nash equilibrium, as long as $\sum_{i=1}^n \text{Reg}_i^T \geq -\text{gap } T$, where $\text{gap} > 0$. In fact, this argument is analogous to that we provide for near-potential games in Theorem 4.10.

Corollary 3.3 (Optimal Regret Bound). *In the setting of Theorem 3.1 with $\alpha_i = \alpha$, $\beta_i = \beta$, and $\gamma_i = \gamma$, the individual regret of each player $i \in [n]$ can be bounded as*

$$\text{Reg}_i^T \leq \alpha + \frac{2n(n-1)\alpha\beta}{\gamma}.$$

Proof. First of all, when $\gamma_i = \gamma$ and $\alpha_i = \alpha$ for all $i \in [n]$, the bound we obtained in Theorem 3.1 can be simplified as

$$\sum_{i=1}^n \sum_{t=1}^T \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq \frac{2n\alpha}{\gamma}. \tag{14}$$

Moreover, the RVU property implies that the regret of each individual player $i \in [n]$ can be bounded as

$$\text{Reg}_i^T \leq \alpha + \beta \sum_{t=1}^T \|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty^2 - \gamma \sum_{t=1}^T \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2 \leq \alpha + (n-1)\beta \sum_{j \neq i} \sum_{t=1}^T \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_1^2, \tag{15}$$

where we used Claim A.1 and the fact that $\gamma > 0$. As a result, plugging-in bound (14) (which follows from Theorem 3.1) to (15) concludes the proof. □

Next, we proceed with the proof of Theorem 3.4, the detailed version of which is given below.

Theorem A.12 (Full Version of Theorem 3.4). *Suppose that each player employs (OMD) with (i) pair of norms $(\|\cdot\|, \|\cdot\|_*)$ such that $\|\boldsymbol{x}\| \geq C\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{x}\|_* \leq C_*\|\boldsymbol{x}\|_\infty$ for any $\boldsymbol{x} \in \mathcal{X}_i$; (ii) G_i-smooth regularizer $\mathcal{R}_i$; and (iii) learning rate $\eta \leq \frac{C}{4C_*(n-1)}$. Moreover, suppose that the game is such that $\sum_{i=1}^n \text{Reg}_i^T \geq 0$ for any $T \in \mathbb{N}$. Then, for any $\epsilon > 0$, after $T > \lceil \frac{8}{\epsilon^2} \sum_{i=1}^n \Omega_i \rceil$ iterations there exists an iterate $\boldsymbol{x}^{(t)}$ with $t \in [T]$ which is an*

$$\epsilon \left(C_* + 2 \frac{\max_{i \in [n]} \{G_i \Omega_i'\}}{\eta} \right)$$

approximate Nash equilibrium, where $\Omega_i := \sup_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_i} D_{\mathcal{R}_i}(\boldsymbol{x}, \boldsymbol{y})$ and $\Omega_i' := \sup_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_i} \|\boldsymbol{x} - \boldsymbol{y}\|$.

Proof. We will use the following refined RVU bound, extracted from (Syrgkanis et al., 2015, Proof of Theorem 18):

Proposition A.13 (Syrgkanis et al., 2015). *If a player i employs (OMD), it holds that*

$$\text{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \eta \sum_{t=1}^T \|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_*^2 - \frac{1}{4\eta} \sum_{t=1}^T \left(\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\|^2 + \|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\|^2 \right).$$

Using the norm-equivalence bounds $\|x\| \geq C\|x\|_1$ and $\|x\|_* \leq C_*\|x\|_\infty$ to the bound of Proposition A.13 yields that

$$\text{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \eta C_*^2 \sum_{t=1}^T \|u_i^{(t)} - u_i^{(t-1)}\|_\infty^2 - \frac{1}{8\eta}C^2 \sum_{t=1}^T \|x_i^{(t)} - \widehat{x}_i^{(t)}\|_1^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|_1^2$$

$$- \frac{1}{8\eta} \sum_{t=1}^T \|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2$$

Moreover, combining this bound with Claim A.1 implies that

$$\text{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \eta C_*^2(n-1) \sum_{t=1}^T \sum_{j \neq i} \|x_j^{(t)} - x_j^{(t-1)}\|_1^2 - \frac{1}{8\eta}C^2 \sum_{t=1}^T \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|_1^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|_1^2 \right)$$

$$- \frac{1}{8\eta} \sum_{i=1}^T \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2 \right)$$

$$\leq \frac{\Omega_i}{\eta} + \eta C_*^2(n-1) \sum_{t=1}^T \sum_{j \neq i} \|x_j^{(t)} - x_j^{(t-1)}\|_1^2 - \frac{C^2}{16\eta} \sum_{t=1}^T \|x_i^{(t)} - x_i^{(t-1)}\|_1^2$$

$$- \frac{1}{8\eta} \sum_{i=1}^T \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2 \right),$$

where we used the fact that

$$\sum_{t=1}^T \|x_i^{(t)} - x_i^{(t-1)}\|^2 \leq 2 \sum_{t=1}^T \|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + 2 \sum_{t=1}^T \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|_1^2,$$

which in turn follows from Young's inequality. As a result, we may conclude that

$$\sum_{i=1}^n \text{Reg}_i^T \leq \sum_{i=1}^n \frac{\Omega_i}{\eta} + \left(\eta C_*^2(n-1)^2 - \frac{C^2}{16\eta} \right) \sum_{i=1}^n \sum_{t=1}^T \|x_i^{(t)} - x_i^{(t-1)}\|_1^2$$

$$- \frac{1}{8\eta} \sum_{i=1}^n \sum_{i=1}^T \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2 \right).$$

Thus, for learning rate $\eta \leq \frac{C}{4C_*(n-1)}$ it follows that

$$0 \leq \sum_{i=1}^n \text{Reg}_i^T \leq \frac{1}{\eta} \sum_{i=1}^n \Omega_i - \frac{1}{8\eta} \sum_{i=1}^n \sum_{i=1}^T \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2 \right),$$

implying that

$$\sum_{t=1}^T \sum_{i=1}^n \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2 \right) \leq 8 \sum_{i=1}^n \Omega_i. \tag{16}$$

Now assume that for all $t \in [T]$ it holds that $\sum_{i=1}^n \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2 \right) > \epsilon^2$. In this case, it follows from (16) that

$$\epsilon^2 T \leq 8 \sum_{i=1}^n \Omega_i \implies T \leq \frac{8}{\epsilon^2} \sum_{i=1}^n \Omega_i.$$

Thus, for $T > \left\lceil \frac{8}{\epsilon^2} \sum_{i=1}^n \Omega_i \right\rceil$ it must be the case that there exists $t \in [T]$ such that

$$\sum_{i=1}^n \left(\|x_i^{(t)} - \widehat{x}_i^{(t)}\|^2 + \|x_i^{(t)} - \widehat{x}_i^{(t-1)}\|^2 \right) \leq \epsilon^2.$$

In turn, this implies that for any $i \in [n]$,

(i) $\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\| \leq \epsilon$;

(ii) $\|\widehat{\boldsymbol{x}}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\|^2 \leq 2\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\|^2 + 2\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\|^2 \leq 2\epsilon^2 \implies \|\widehat{\boldsymbol{x}}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\| \leq 2\epsilon.$

Finally, we show the following claim which will conclude the proof.

Claim A.14. *In the setting of Theorem A.12, suppose that*

$$\sqrt{\sum_{i=1}^{n} \left(\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\|^2 + \|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\|^2 \right)} \leq \epsilon.$$

Then, it follows that $\boldsymbol{x}^{(t)}$ is an

$$\epsilon \left(C_* + 2\frac{\max_{i \in [n]} G_i \Omega_i'}{\eta} \right)$$

approximate Nash equilibrium.

Proof. Observe that the maximization problem associated with (OMD) can be expressed in the following variational inequality form:

$$\left\langle \boldsymbol{u}_i^{(t)} - \frac{1}{\eta} \left(\nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t)}) - \nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t-1)}) \right), \widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_i^{(t)} \right\rangle \leq 0, \quad \forall \widehat{\boldsymbol{x}}_i \in \mathcal{X}_i,$$

for any $i \in [n]$. Thus, it follows that

$$\langle \boldsymbol{u}_i^{(t)}, \widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_i^{(t)} \rangle \leq \frac{1}{\eta} \langle \nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t)}) - \nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t-1)}), \widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_i^{(t)} \rangle$$

$$\leq \frac{1}{\eta} \|\nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t)}) - \nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t-1)})\|_* \|\widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_i^{(t)}\| \tag{17}$$

$$\leq 2\epsilon \frac{G_i \Omega_i'}{\eta}, \tag{18}$$

where (17) follows from the Cauchy-Schwarz inequality, and (18) uses the fact that $\|\nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t)}) - \nabla \mathcal{R}_i(\widehat{\boldsymbol{x}}_i^{(t-1)})\|_* \leq G_i \|\widehat{\boldsymbol{x}}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\| \leq 2\epsilon G_i$, which follows from the smoothness assumption on the regularizer $\mathcal{R}_i$. (18) also uses the notation $\Omega_i' := \sup_{\boldsymbol{x},\boldsymbol{y} \in \mathcal{X}_i} \|\boldsymbol{x} - \boldsymbol{y}\|$. As a result, we have established that for any player $i \in [n]$ it holds that for any $\widehat{\boldsymbol{x}}_i \in \mathcal{X}_i$,

$$\langle \boldsymbol{u}_i^{(t)}, \widehat{\boldsymbol{x}}_i^{(t)} \rangle \geq \langle \boldsymbol{u}_i^{(t)}, \widehat{\boldsymbol{x}}_i \rangle - 2\epsilon \frac{G_i \Omega_i'}{\eta}. \tag{19}$$

Moreover, we also have that

$$\left| \langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)} \rangle \right| \leq \|\boldsymbol{u}_i^{(t)}\|_* \|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\| \leq \epsilon C_*,$$

where we used the fact that $\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\| \leq \epsilon$, and that $\|\boldsymbol{u}_i^{(t)}\|_\infty \leq 1$ (by the normalization hypothesis). Plugging-in the last bound to (19) gives us that

$$\langle \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \rangle \geq \langle \widehat{\boldsymbol{x}}_i^{(t)}, \boldsymbol{u}_i^{(t)} \rangle - \epsilon C_* \geq \langle \widehat{\boldsymbol{x}}_i, \boldsymbol{u}_i^{(t)} \rangle - \epsilon C_* - 2\epsilon \frac{G_i \Omega_i'}{\eta},$$

for any $\widehat{\boldsymbol{x}}_i \in \mathcal{X}_i$ and player $i \in [n]$. As a result, the proof follows by definition of approximate Nash equilibria. $\square$

$\square$

Remark A.15 (Last-Iterate Convergence). Leveraging our argument in the proof of Theorem A.12 it follows that for any $\epsilon > 0$ there exists a sufficiently large $T = T(\epsilon)$ so that $\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\|_1 \leq \epsilon$ and $\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\| \leq \epsilon$, for any $i \in [n]$ and $t \geq T$. In turn, under smooth regularizers this implies that any $\boldsymbol{x}^{(t)}$ with $t \geq T = T(\epsilon)$ will be an $O(\epsilon)$-approximate Nash equilibrium by virtue of Claim A.14, establishing last-iterate convergence. On the other hand, our techniques do not seem to imply *pointwise convergence*.

Next, we give the proof of Theorem 3.8, the detailed version of which is given in Theorem A.17. To this end, we will require the following proposition shown by Syrgkanis et al. (2015), which is a slight refinement of a result due to Roughgarden (2015).

Proposition A.16 (Syrgkanis et al., 2015). *Consider a (λ, μ)-smooth game. If each player $i \in [n]$ incurs regret at most Reg_i^T, then*

$$\frac{1}{T} \sum_{t=1}^{T} \text{SW}(\boldsymbol{x}^{(t)}) \geq \frac{\lambda}{1+\mu} \text{OPT} - \frac{1}{1+\mu} \frac{1}{T} \sum_{i=1}^{n} \text{Reg}_i^T.$$

Theorem A.17 (Full Version of Theorem 3.8). *Suppose that each player employs (OMD) with (i) pair of norms $(\|\cdot\|, \|\cdot\|_*)$ such that $\|\boldsymbol{x}\| \geq C\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{x}\|_* \leq C_*\|\boldsymbol{x}\|_\infty$ for any $\boldsymbol{x} \in \mathcal{X}_i$; (ii) G_i-smooth regularizer $\mathcal{R}_i$; and (iii) learning rate $\eta \leq \frac{C}{4C_*(n-1)}$. Moreover, fix any $\gamma > 0$ and consider T iterations of the dynamics with $T \geq \frac{16 \sum_{i=1}^{n} \Omega_i}{\gamma^2}$, where $\Omega_i := \sup_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_i} D_{\mathcal{R}_i}(\boldsymbol{x}, \boldsymbol{y})$. Then, either of the following occurs:*

1. *There exists an iterate $\boldsymbol{x}^{(t)}$ with $t \in [T]$ which is*

$$\gamma \left(C_* + 2 \frac{\max_{i \in [n]} G_i \Omega_i'}{\eta} \right)$$

   approximate Nash equilibrium;

2. *Otherwise, it holds that*

$$\frac{1}{T} \sum_{t=1}^{T} \text{SW}(\boldsymbol{x}^{(t)}) \geq \frac{\lambda}{1+\mu} \text{OPT} + \frac{\gamma^2}{16\eta}.$$

Proof. Similarly to the proof of Theorem A.12, we may conclude that for $\eta \leq \frac{C}{4C_*(n-1)}$ it holds that

$$\sum_{i=1}^{n} \text{Reg}_i^T \leq \frac{1}{\eta} \sum_{i=1}^{n} \Omega_i - \frac{1}{8\eta} \sum_{t=1}^{T} \sum_{i=1}^{n} \left(\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\|^2 + \|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\|^2 \right). \tag{20}$$

Now if there exists $t \in [T]$ such that

$$\sum_{i=1}^{n} \left(\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\|^2 + \|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t-1)}\|^2 \right) \leq \gamma^2,$$

it follows from Claim A.14 that $\boldsymbol{x}^{(t)}$ is a

$$\gamma \left(C_* + 2 \frac{\max_{i \in [n]} G_i \Omega_i'}{\eta} \right)$$

approximate Nash equilibrium. In the contrary case, we may conclude from (20) that

$$\sum_{i=1}^{n} \text{Reg}_i^T \leq \frac{1}{\eta} \sum_{i=1}^{n} \Omega_i - \frac{1}{8\eta} \sum_{t=1}^{T} \gamma^2 \leq -\frac{1}{16\eta} \gamma^2 T,$$

as long as $T \geq \frac{16 \sum_{i=1}^{n} \Omega_i}{\gamma^2}$. Thus, we may conclude from Proposition A.16 that

$$\frac{1}{T} \sum_{t=1}^{T} \text{SW}(\boldsymbol{x}^{(t)}) \geq \frac{\lambda}{1+\mu} \text{OPT} + \frac{1}{1+\mu} \frac{\gamma^2}{16\eta}.$$

$\square$

A.1 Smooth Convex-Concave Games

In this subsection we explain how our framework can also be applied in the context of *smooth min-max optimization*. To be precise, let $f(\boldsymbol{x}, \boldsymbol{y}) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a continuously differentiable *convex-concave* function; that is, $f(\boldsymbol{x}, \boldsymbol{y})$ is convex with respect to $\boldsymbol{x} \in \mathcal{X}$ for any fixed $\boldsymbol{y} \in \mathcal{Y}$ and concave with respect to $\boldsymbol{y} \in \mathcal{Y}$ for any fixed $\boldsymbol{x} \in \mathcal{X}$. We will also make the following standard ℓ_2-smoothness assumptions:

$$
\begin{aligned}
\|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y}) - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y}')\|_2 &\le L\|\boldsymbol{y} - \boldsymbol{y}'\|_2, \quad \forall \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}, \boldsymbol{y}' \in \mathcal{Y}; \\
\|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y}) - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}', \boldsymbol{y})\|_2 &\le L\|\boldsymbol{x} - \boldsymbol{x}'\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y};
\end{aligned} \tag{21}
$$

and

$$
\begin{aligned}
\|\nabla_{\boldsymbol{y}} f(\boldsymbol{x}, \boldsymbol{y}) - \nabla_{\boldsymbol{y}} f(\boldsymbol{x}, \boldsymbol{y}')\|_2 &\le L\|\boldsymbol{y} - \boldsymbol{y}'\|_2, \quad \forall \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}, \boldsymbol{y}' \in \mathcal{Y}; \\
\|\nabla_{\boldsymbol{y}} f(\boldsymbol{x}, \boldsymbol{y}) - \nabla_{\boldsymbol{y}} f(\boldsymbol{x}', \boldsymbol{y})\|_2 &\le L\|\boldsymbol{x} - \boldsymbol{x}'\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}.
\end{aligned} \tag{22}
$$

For notational simplicity we are using a common smoothness parameter L in (21) and (22), but a more refined analysis follows directly from our techniques.

The key idea is the well-known fact that one can construct a regret minimizer for convex utility functions via a regret minimizer for linear utilities (*e.g.*, see (McMahan, 2011)). Indeed, we claim that the incurred regret Reg^T under convex utility functions can be bounded by the regret $\mathrm{Reg}_\mathcal{L}$ of an algorithm observing the tangent plane of the utility function at every decision point. To see this, let $\nabla_{\boldsymbol{x}} u^{(t)}(\boldsymbol{x}^{(t)})$ be the gradient of a convex and continuously differentiable utility function $u^{(t)}$ on some convex domain $\mathcal{X}$.[6] Then, by convexity, we have that

$$
u^{(t)}(\boldsymbol{x}) \ge \boldsymbol{u}^{(t)}(\boldsymbol{x}^{(t)}) + \langle \nabla_{\boldsymbol{x}} u^{(t)}(\boldsymbol{x}^{(t)}), \boldsymbol{x} - \boldsymbol{x}^{(t)} \rangle, \quad \forall \boldsymbol{x} \in \mathcal{X}.
$$

From this inequality it is easy to conclude that

$$
\mathrm{Reg}^T \le \mathrm{Reg}_\mathcal{L}^T, \tag{23}
$$

as we claimed.

Now let us assume, for concreteness, that each player employs (OMD) with Euclidean regularization, while observing the *linearized utility functions* based on the aforementioned scheme. Then, Proposition 2.3 implies that the individual regret of each player can be bounded as

$$
\begin{aligned}
\mathrm{Reg}_{\mathcal{X},\mathcal{L}} &\le \frac{\Omega_\mathcal{X}}{\eta} + \eta \sum_{t=1}^{T} \|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t-1)}, \boldsymbol{y}^{(t-1)})\|_2^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2; \\
\mathrm{Reg}_{\mathcal{Y},\mathcal{L}} &\le \frac{\Omega_\mathcal{Y}}{\eta} + \eta \sum_{t=1}^{T} \|\nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) - \nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t-1)}, \boldsymbol{y}^{(t-1)})\|_2^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2,
\end{aligned} \tag{24}
$$

where $\Omega_\mathcal{X}$ and $\Omega_\mathcal{Y}$ are defined as in Proposition 2.3. Now it follows from (21) and Young's inequality that

$$
\begin{aligned}
\|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t-1)}, \boldsymbol{y}^{(t-1)})\|_2^2 &\le 2\|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t-1)})\|_2^2 + \\
&\quad 2\|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t-1)}) - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(t-1)}, \boldsymbol{y}^{(t-1)})\|_2^2 \\
&\le 2L^2 \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 + 2L^2 \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2.
\end{aligned} \tag{25}
$$

Similarly, it follows from (22) and Young's inequality that

$$
\begin{aligned}
\|\nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) - \nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t-1)}, \boldsymbol{y}^{(t-1)})\|_2^2 &\le 2\|\nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) - \nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t-1)})\|_2^2 + \\
&\quad 2\|\nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t-1)}) - \nabla_{\boldsymbol{y}} f(\boldsymbol{x}^{(t-1)}, \boldsymbol{y}^{(t-1)})\|_2^2 \\
&\le 2L^2 \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 + 2L^2 \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2.
\end{aligned} \tag{26}
$$

---

[6]The same technique applies more generally using any *subgradient* at the decision point.

As a result, plugging (25) and (26) to the RVU bound of (24) gives us that

$$\text{Reg}_{\mathcal{X},\mathcal{L}} + \text{Reg}_{\mathcal{Y},\mathcal{L}} \leq \frac{\Omega_{\mathcal{X}} + \Omega_{\mathcal{Y}}}{\eta} + \left(4\eta L^2 - \frac{1}{8\eta}\right)\sum_{t=1}^{T}\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 + \left(4\eta L^2 - \frac{1}{8\eta}\right)\sum_{t=1}^{T}\|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2. \quad (27)$$

Letting $\eta = \frac{1}{8L}$ in (27) implies that

$$\text{Reg}_{\mathcal{X},\mathcal{L}} + \text{Reg}_{\mathcal{Y},\mathcal{L}} \leq 8L(\Omega_{\mathcal{X}} + \Omega_{\mathcal{Y}}) - \frac{L}{2}\sum_{t=1}^{T}\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 - \frac{L}{2}\sum_{t=1}^{T}\|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2. \quad (28)$$

Moreover, we know from Proposition 3.2 that $\text{Reg}_{\mathcal{X}}^T + \text{Reg}_{\mathcal{Y}}^T \geq 0$, in turn implying that $\text{Reg}_{\mathcal{X},\mathcal{L}}^T + \text{Reg}_{\mathcal{Y},\mathcal{L}}^T \geq 0$ by virtue of (23). As a result, we may conclude from (28) the following theorem.

Theorem A.18. *Let $f(\boldsymbol{x}, \boldsymbol{y})$ be a continuously differentiable convex-concave function satisfying the L-smoothness condition of (21) and (22). If both players employ (OMD) with Euclidean regularization and $\eta = \frac{1}{8L}$ it holds that*

$$\sum_{t=1}^{T}\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 + \sum_{t=1}^{T}\|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2 \leq 16(\Omega_{\mathcal{X}} + \Omega_{\mathcal{Y}}).$$

This theorem combined with (24) directly gives us that every player incurs $O(1)$ individual regret. Further, a bound on the number of iterations required to reach an approximate equilibrium can be derived similarly to Theorem 3.4.

Remark A.19 (Unconstrained Setting). While our framework based on regret minimization—and in particular Theorem A.18—requires $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ to be bounded, extensions are possible to the unconstrained setup as well. Indeed, as was pointed out by Golowich et al. (2020a, Remark 4), it suffices to use the fact that the iterates of (unconstrained) optimistic gradient descent remain within a bounded ball that depends only the initialization; the later property was shown in (Mokhtari et al., 2020a, Lemma 4). This can be combined with Theorem A.18 to bound the norm of the gradients $\Lambda^{(t)} := \|\nabla_{\boldsymbol{x}}f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)})\| + \|\nabla_{\boldsymbol{y}}f(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)})\|$ at time t. More precisely, Theorem A.18 implies that after T iterations there exists time $t \in [T]$ such that $\Lambda^{(t)} = O(1/\sqrt{T})$; cf., see Golowich et al. (2020a). We point out that $\Lambda^{(t)}$ is, perhaps, the most common measure for last-iterate convergence in min-max optimization (Diakonikolas & Wang, 2021).

Remark A.20 (Curvature Exploitation). The linearization trick we employed can be suboptimal as the regret minimization algorithm could fail to exploit the *curvature* of the utility functions; *e.g.*, this would be the case if the objective function f is strongly-convex-strongly-concave. Extending our framework to address this issue is an interesting direction for the future.

A.2 Bilinear Saddle-Point Problems

In this subsection we show that our framework has direct implications for extensive-form games (EFGs). A comprehensive overview of extensive-form games would lead us beyond the scope of this paper. Instead, we will focus on two-player zero-sum games wherein the computation of a Nash equilibrium can be formulated as a *bilinear saddle-point problem* (BSPP). A BSPP can be expressed as

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

where $\mathcal{X}$ and $\mathcal{Y}$ are convex and compact sets, and $\mathbf{A} \in \mathbb{R}^{n \times m}$. In the case of EFGs, $\mathcal{X}$ and $\mathcal{Y}$ are the *sequence-form strategy polytopes* of the sequential decision process faced by the two players, and $\mathbf{A}$ is a matrix with the leaf payoffs of the game. We first show the following.

Proposition A.21. *Suppose that both players in a BSPP employ (OMD) with Euclidean regularizer and $\eta \leq \frac{1}{4\|\mathbf{A}\|_2}$, where $\|\mathbf{A}\|_2$ is the spectral norm of $\mathbf{A}$. Then, it holds that*

$$\sum_{t=1}^{T}\left(\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 + \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2\right) \leq 16(\Omega_{\mathcal{X}} + \Omega_{\mathcal{Y}}).$$

Proof. Given that player $\mathcal{X}$ employs (OMD) with Euclidean regularization, Proposition 2.3 implies that[7]

$$\text{Reg}_{\mathcal{X}}^T \le \frac{\Omega_{\mathcal{X}}}{\eta} + \eta \sum_{t=1}^{T} \|\mathbf{A}\boldsymbol{y}^{(t)} - \mathbf{A}\boldsymbol{y}^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2$$

$$\le \frac{\Omega_{\mathcal{X}}}{\eta} + \eta \|\mathbf{A}\|_2^2 \sum_{t=1}^{T} \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2. \tag{29}$$

Similarly, we have that

$$\text{Reg}_{\mathcal{Y}}^T \le \frac{\Omega_{\mathcal{Y}}}{\eta} + \eta \sum_{t=1}^{T} \|\mathbf{A}^\top \boldsymbol{x}^{(t)} - \mathbf{A}^\top \boldsymbol{x}^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2$$

$$\le \frac{\Omega_{\mathcal{Y}}}{\eta} + \eta \|\mathbf{A}\|_2^2 \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=1}^{T} \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2, \tag{30}$$

where we used the fact that $\|\mathbf{A}^\top\|_2 = \|\mathbf{A}\|_2$. Thus, summing (29) and (30) yields that

$$0 \le \text{Reg}_{\mathcal{X}}^T + \text{Reg}_{\mathcal{Y}}^T \le \frac{\Omega_{\mathcal{X}} + \Omega_{\mathcal{Y}}}{\eta} + \left(\eta \|\mathbf{A}\|_2^2 - \frac{1}{8\eta}\right) \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 + \left(\eta \|\mathbf{A}\|_2^2 - \frac{1}{8\eta}\right) \sum_{t=1}^{T} \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2.$$

$$\le \frac{\Omega_{\mathcal{X}} + \Omega_{\mathcal{Y}}}{\eta} - \frac{1}{16\eta} \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 - \frac{1}{16\eta} \sum_{t=1}^{T} \|\boldsymbol{y}^{(t)} - \boldsymbol{y}^{(t-1)}\|_2^2,$$

where we used the fact that $\eta \le \frac{1}{4\|\mathbf{A}\|_2}$. A rearrangement of the final bound completes the proof. □

Moreover, we can employ the argument of Theorem 3.4 to also bound the number of iterations required to reach an ϵ-Nash equilibrium of the BSPP. In the following claim we use the $O(\cdot)$ notation to suppress (universal) constants.

Proposition A.22 (Full Version of Proposition 3.5). *Suppose that both players in a BSPP employ* (OMD) *with Euclidean regularization and* $\eta \le \frac{1}{4\|\mathbf{A}\|_2}$. *Then, after* $T = O\left(\frac{\Omega_{\mathcal{X}} + \Omega_{\mathcal{Y}}}{\epsilon^2}\right)$ *iterations there is a joint strategy* $(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)})$ *with* $t \in [T]$ *which is an* $O(\epsilon \max\{\Omega_{\mathcal{X}}, \Omega_{\mathcal{Y}}\} \|\mathbf{A}\|_2)$-*approximate Nash equilibrium of the BSPP.*

B. Proofs from Section 4

In this section we give complete proofs for our results in Section 4. We commence with the simple proof of Proposition 4.2, which is recalled below.

Proposition 4.2. *Let* Γ *be a game for which there exists a function* $\Phi : \prod_{i=1}^{n} \mathcal{A}_i \to \mathbb{R}$ *with*

$$\Phi(\boldsymbol{a}) - \Phi(a_i', \boldsymbol{a}_{-i}) = \boldsymbol{w}_i(u_i(\boldsymbol{a}) - u_i(a_i', \boldsymbol{a}_{-i})),$$

where $\boldsymbol{w} \in \mathbb{R}_{>0}^n$. *Then,* Γ *is a potential game in the sense of Definition 4.1 with* $L = \frac{1}{2} \Phi_{\max} \sum_{i=1}^{n} |\mathcal{A}_i|$.

Proof. By assumption, we know that there exists a (bounded) function Φ such that

$$\Phi(\boldsymbol{a}) - \Phi(a_i', \boldsymbol{a}_{-i}) = \boldsymbol{w}_i(u_i(\boldsymbol{a}) - u_i(a_i', \boldsymbol{a}_{-i})). \tag{31}$$

That is, Γ is a *weighted potential game*: the difference in utility deriving from a deviation of a player $i \in [n]$ is translated to the exact same deviation in the potential function, modulo the scaling factor $\boldsymbol{w}_i$. We will first show that the condition of (31) can be translated for the mixed extensions as well. As is common, in the sequel we slightly abuse notation by using the same symbols for the mixed extensions of the potential function and the utility function of each player.

---

[7]Although Proposition 2.3 was only stated for simplexes by Syrgkanis et al. (2015), the proof readily extends for arbitrary convex and compact sets.

Claim B.1. *For any* $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \prod_{i \in [n]} \Delta(\mathcal{A}_i)$ *it holds that*

$$\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) = \boldsymbol{w}_i(u_i(\boldsymbol{x}) - u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i})), \tag{32}$$

where $\Phi(\boldsymbol{x}) := \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{x}}[\Phi(\boldsymbol{a})] = \sum_{(a_1, \ldots, a_n) \in \mathcal{A}} \Phi(a_1, \ldots, a_n) \prod_{i=1}^{n} \boldsymbol{x}_i(a_i)$.

Proof. We have that

$$
\begin{aligned}
\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{x}_i', \boldsymbol{x}_{-i}) &= \sum_{\boldsymbol{a} \in \mathcal{A}} \Phi(\boldsymbol{a})(\boldsymbol{x}_i(a_i) - \boldsymbol{x}_i'(a_i)) \prod_{j \neq i} \boldsymbol{x}_j(a_j) \\
&= \sum_{a_i \in \mathcal{A}_i} (\boldsymbol{x}_i(a_i) - \boldsymbol{x}_i'(a_i)) \sum_{\boldsymbol{a}_{-i} \in \mathcal{A}_{-i}} \Phi(a_i, \boldsymbol{a}_{-i}) \prod_{j \neq i} \boldsymbol{x}_j(a_j) \\
&= \boldsymbol{w}_i \sum_{a_i \in \mathcal{A}_i} (\boldsymbol{x}_i(a_i) - \boldsymbol{x}_i'(a_i)) \sum_{\boldsymbol{a}_{-i} \in \mathcal{A}_{-i}} u_i(a_i, \boldsymbol{a}_{-i}) \prod_{j \neq i} \boldsymbol{x}_j(a_j) \\
&= \boldsymbol{w}_i \left(\sum_{\boldsymbol{a} \in \mathcal{A}} u_i(\boldsymbol{a}) \prod_{j=1}^{n} \boldsymbol{x}_j(a_j) - \sum_{\boldsymbol{a} \in \mathcal{A}} u_i(\boldsymbol{a}) \boldsymbol{x}_i'(a_i) \prod_{k \neq i}^{n} \boldsymbol{x}_k(a_k) \right) \\
&= \boldsymbol{w}_i(u_i(\boldsymbol{x}) - u_i(\boldsymbol{x}_i', \boldsymbol{x}_{-i})),
\end{aligned}
\tag{33}
$$

where (33) follows from a rearrangement of the terms and the assumption of (31). $\qquad \square$

Now observe that from Claim B.1 we may conclude that

$$
\begin{aligned}
\frac{\partial \Phi(\boldsymbol{x})}{\partial \boldsymbol{x}_i(a_i)} &= \boldsymbol{w}_i \frac{\partial u_i(\boldsymbol{x})}{\partial \boldsymbol{x}_i(a_i)} = \boldsymbol{w}_i \frac{\partial}{\partial \boldsymbol{x}_i(a_i)} \left(\sum_{\boldsymbol{a} \in \mathcal{A}} u_i(\boldsymbol{a}) \prod_{j=1}^{n} \boldsymbol{x}_j(a_j) \right) \\
&= \boldsymbol{w}_i \sum_{\boldsymbol{a}_{-i} \in \mathcal{A}_{-i}} u_i(a_i, \boldsymbol{a}_{-i}) \prod_{j \neq i} \boldsymbol{x}_j(a_j) = \boldsymbol{w}_i u_i(a_i, \boldsymbol{x}_{-i}).
\end{aligned}
$$

This verifies condition (4) from Definition 4.1. Thus, it remains to establish the smoothness of the potential, in the sense of (3). To this end, we first recall the following simple fact.

Fact B.2. *Let* $g : \mathbb{R}^d \to \mathbb{R}$ *be a twice continuously differentiable function such that the spectral norm of the Hessian* $\nabla^2 g$ *is upper bounded by some* $M > 0$. *Then, for any* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ *it holds that*

$$\|\nabla g(\boldsymbol{x}) - \nabla g(\boldsymbol{y})\|_2 \leq M \|\boldsymbol{x} - \boldsymbol{y}\|_2.$$

Hence, it suffices to bound the operator norm of the Hessian of Φ. This is shown in the following lemma, where we crucially leverage the multilinearity of the (mixed) potential function.

Lemma B.3. *It holds that* $\|\nabla^2 \Phi\|_2 \leq \Phi_{\max} \sum_{i=1}^{n} |\mathcal{A}_i|$.

Proof. First of all, it is easy to see that for all $i \in [n]$ and $a_i, a_i' \in \mathcal{A}_i$, it holds that

$$\frac{\partial^2 \Phi}{\partial \boldsymbol{x}_i(a_i) \partial \boldsymbol{x}_i(a_i')} = 0.$$

On the other hand, for $i \neq j \in [n]$ and $a_i \in \mathcal{A}_i$ and $a_j \in \mathcal{A}_j$ it holds that

$$\frac{\partial^2 \Phi}{\partial \boldsymbol{x}_i(a_i) \partial \boldsymbol{x}_j(a_j)} = \sum_{\boldsymbol{a}_{-i,-j} \in \mathcal{A}_{-i,-j}} \Phi(a_i, a_j, \boldsymbol{a}_{-i,-j}) \prod_{k \neq i,j} \boldsymbol{x}_k(a_k);$$

here we used the notation $\boldsymbol{a}_{-i,-j} := \prod_{k \neq i,j} a_k$ and $\mathcal{A}_{-i,-j} := \prod_{k \neq i,j} \mathcal{A}_k$. Thus, applying the triangle inequality yields that

$$\left| \frac{\partial^2 \Phi}{\partial \boldsymbol{x}_i(a_i) \partial \boldsymbol{x}_j(a_j)} \right| = \left| \sum_{\boldsymbol{a}_{-i,-j}} \Phi(a_i, a_j, \boldsymbol{a}_{-i,-j}) \prod_{k \neq i,j} \boldsymbol{x}_k(a_k) \right| \leq \Phi_{\max} \sum_{\boldsymbol{a}_{-i,-j}} \prod_{k \neq i,j} \boldsymbol{x}_k(a_k) = \Phi_{\max}, \tag{34}$$

where the final equality holds by the normalization constraint of the induced product distribution. Thus, for any $\boldsymbol{z} \in \mathbb{R}^d$, where $d = \sum_{i=1}^{n} |\mathcal{A}_i|$, we have that

$$\|\nabla^2 \Phi \boldsymbol{z}\|_2 \leq \Phi_{\max} \sqrt{d \left(\sum_{r=1}^{d} \boldsymbol{z}(r) \right)^2} \leq \Phi_{\max} \sqrt{d^2 \sum_{r=1}^{d} \boldsymbol{z}^2(r)} = \Phi_{\max} d \|\boldsymbol{z}\|_2,$$

where we used (34) in the first bound, and Jensen's inequality in the second one. As a result, we have shown that $\|\nabla^2 \Phi\|_2 \leq \Phi_{\max} \sum_{i=1}^{n} |\mathcal{A}_i|$, concluding the proof. $\square$

Finally, combining this lemma with Fact B.2 we arrive at the following corollary, which concludes the proof of Proposition 4.2.

Corollary B.4. *Let* $L = \frac{1}{2} \Phi_{\max} \sum_{i=1}^{n} |\mathcal{A}_i|$. *Then, for any* $\boldsymbol{x}, \widetilde{\boldsymbol{x}} \in \prod_{i \in [n]} \Delta(\mathcal{A}_i)$ *it holds that*

$$\Phi(\widetilde{\boldsymbol{x}}) \leq \Phi(\boldsymbol{x}) + \langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \widetilde{\boldsymbol{x}} - \boldsymbol{x} \rangle + L \|\widetilde{\boldsymbol{x}} - \boldsymbol{x}\|_2^2;$$
$$-\Phi(\widetilde{\boldsymbol{x}}) \leq -\Phi(\boldsymbol{x}) - \langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \widetilde{\boldsymbol{x}} - \boldsymbol{x} \rangle + L \|\widetilde{\boldsymbol{x}} - \boldsymbol{x}\|_2^2.$$

$\square$

Theorem 4.3. *Suppose that each player employs* (MD) *with a 1-strongly convex regularizer with respect to* $\|\cdot\|_2$, *and learning rate* $\eta = \frac{1}{2L}$, *where L is defined as in Proposition 4.2. Then, for any $t \geq 1$ it holds that*

$$\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \geq \frac{1}{2\eta} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 \geq 0. \tag{5}$$

Proof. First of all, since every regularizer $\mathcal{R}_i$ is 1-strongly convex with respect to $\|\cdot\|_2$, we obtain from Definition 4.1 that

$$\Phi(\boldsymbol{x}) \leq \Phi(\widetilde{\boldsymbol{x}}) - \langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \widetilde{\boldsymbol{x}} - \boldsymbol{x} \rangle + L \sum_{i=1}^{n} \|\widetilde{\boldsymbol{x}}_i - \boldsymbol{x}_i\|_2^2$$
$$\leq \Phi(\widetilde{\boldsymbol{x}}) - \langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \widetilde{\boldsymbol{x}} - \boldsymbol{x} \rangle + 2L \sum_{i=1}^{n} D_{\mathcal{R}_i}(\widetilde{\boldsymbol{x}}_i, \boldsymbol{x}_i). \tag{35}$$

Moreover, we know from the update rule of (MD) that for any player $i \in [n]$ it holds that

$$\boldsymbol{x}_i^{(t+1)} = \arg\max_{\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)} \left\{ \langle \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}_i - \boldsymbol{x}_i^{(t)} \rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i, \boldsymbol{x}_i^{(t)}) \right\}, \tag{36}$$

where we used the fact that $g(\boldsymbol{u}_i) = \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x})$ (Definition 4.1). Also, from (35) we obtain that

$$\Phi(\boldsymbol{x}^{(t)}) \leq \Phi(\boldsymbol{x}^{(t+1)}) - \langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \rangle + 2L \sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}). \tag{37}$$

Now the update rule of (36) implies that

$$\langle \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) \geq \frac{1}{2\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2, \tag{38}$$

for all $i \in [n]$, where we used the 1-strong convexity of $\mathcal{R}_i$ with respect to $\|\cdot\|_2$ (quadratic growth). As a result, summing (38) for all $i \in [n]$ yields that

$$\left\langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \right\rangle - \frac{1}{\eta} \sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) \geq \frac{1}{2\eta} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2.$$

Finally, plugging this bound with $\eta = \frac{1}{2L}$ to (37) and rearranging the terms concludes the proof. □

Corollary 4.5. *In the setting of Theorem 4.3, with $\mathcal{R}_i := \frac{1}{2}\|\boldsymbol{x}\|_2^2$, it holds that the regret of each player $i \in [n]$ is such that* $\mathrm{Reg}_i^T = O(\sqrt{T})$.

Proof. It is well-known (*e.g.*, see (Shalev-Shwartz, 2012)) that the cumulative regret Reg^T of (MD) can be bounded as

$$\mathrm{Reg}^T \leq \frac{\Omega}{\eta} + \sum_{t=1}^{T} \|\boldsymbol{u}^{(t)}\|_* \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|, \tag{39}$$

where $\Omega := \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{X}} D_{\mathcal{R}}(\boldsymbol{x}, \boldsymbol{y})$, and $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(t)}$ represents the sequence of utility vectors observed by the regret minimizer. Note that with a conventional abuse of notation, it is assumed that $\boldsymbol{x}^{(0)} := \boldsymbol{0}$ in (39). Now we know from Corollary 4.4 that for $\eta = \frac{1}{2L}$ it holds that

$$\sum_{t=2}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 \leq \sum_{i=1}^{n} \sum_{t=2}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2 \leq \frac{2\,\Phi_{\max}}{L}.$$

Thus, an application of the Cauchy-Schwarz inequality implies that

$$\sum_{t=2}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2 \leq \sqrt{(T-1)\sum_{t=2}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2^2} \leq \sqrt{T}\sqrt{\frac{2\,\Phi_{\max}}{L}}. \tag{40}$$

Finally, from (39) it follows that for any player $i \in [n]$ it holds that

$$\mathrm{Reg}_i^T \leq \frac{\Omega_i}{\eta} + \|\boldsymbol{u}_i\|_2 \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_2 \leq \frac{\Omega_i}{\eta} + \|\boldsymbol{u}_i\|_2 \left(\|\boldsymbol{x}^{(1)}\|_2 + \sqrt{T}\sqrt{\frac{2\,\Phi_{\max}}{L}}\right) = O(\sqrt{T}),$$

where we used the notation $\|\boldsymbol{u}_i\|_2 := \max_{t\in[T]} \|\boldsymbol{u}_i^{(t)}\|_2$. This concludes the proof. □

Theorem 4.6 (Optimal Regret for Potential Games)**.** *Suppose that each player employs OMWU with a sufficiently small learning rate $\eta > 0$. Then, the regret of each player $i \in [n]$ is such that $\mathrm{Reg}_i^T = O(1)$.*

Proof. First of all, it is well-known that OMWU on the simplex can be expressed with the following update rule for $t \geq 1$:

$$\boldsymbol{x}_i^{(t+1)}(a_i) = \frac{\exp\left(2\eta \boldsymbol{u}_i^{(t)}(a_i) - \eta \boldsymbol{u}_i^{(t-1)}(a_i)\right)}{\sum_{a_i'\in\mathcal{A}_i} \exp\left(2\eta \boldsymbol{u}_i^{(t)}(a_i') - \eta \boldsymbol{u}_i^{(t-1)}(a_i')\right) \boldsymbol{x}_i^{(t)}(a_i')} \boldsymbol{x}_i^{(t)}(a_i), \tag{OMWU}$$

for any action $a_i \in \mathcal{A}_i$ and player $i \in [n]$. By convention, $\boldsymbol{x}_i^{(0)}$ and $\boldsymbol{x}_i^{(1)}$ are initialized from the uniform distribution over $\mathcal{A}_i$, for all $i \in [n]$. Now we claim that the update rule of (OMWU) is equivalent to

$$\boldsymbol{x}_i^{(t+1)} = \arg\max_{\boldsymbol{x}_i\in\Delta(\mathcal{A}_i)} \left\{ \left\langle 2\nabla_{\boldsymbol{x}_i}\Phi(\boldsymbol{x}^{(t)}) - \nabla_{\boldsymbol{x}_i}\Phi(\boldsymbol{x}^{(t-1)}), \boldsymbol{x}_i - \boldsymbol{x}_i^{(t)} \right\rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i, \boldsymbol{x}_i^{(t)}) \right\},$$

assuming that $\mathcal{R}_i$ is the negative entropy DGF for all $i \in [n]$. By 1-strong convexity of $\mathcal{R}_i$, this implies that

$$\left\langle 2\nabla_{\boldsymbol{x}_i}\Phi(\boldsymbol{x}^{(t)}) - \nabla_{\boldsymbol{x}_i}\Phi(\boldsymbol{x}^{(t-1)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) \geq \frac{1}{2\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2. \tag{41}$$

Moreover, we have that

$$\left| \left\langle \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t)}) - \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t-1)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle \right| \le \|\nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t)}) - \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t-1)})\|_2 \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \tag{42}$$

$$= \|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_2 \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \tag{43}$$

$$\le \sqrt{|\mathcal{A}_i|} \|\boldsymbol{u}_i^{(t)} - \boldsymbol{u}_i^{(t-1)}\|_\infty \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \tag{44}$$

$$\le \sqrt{|\mathcal{A}_i|} \sum_{j \ne i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_2 \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \tag{45}$$

$$\le \frac{1}{2} \sqrt{|\mathcal{A}_i|} \sum_{j \ne i} \left(\|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_2^2 + \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 \right) \tag{46}$$

$$= \frac{1}{2} \sqrt{|\mathcal{A}_i|} \left((n-1)\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 + \sum_{j \ne i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_2^2 \right),$$

where (42) follows from Cauchy-Schwarz; (43) is a consequence of Definition 4.1; (44) is derived from Hölder's inequality (equivalence of norms); (45) uses Claim A.1; and (46) is an application of Young's inequality. Thus, combining this bound to (41) yields that

$$\left\langle \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) \ge \left(\frac{1}{2\eta} - \frac{1}{2}\sqrt{|\mathcal{A}_i|}(n-1) \right) \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2$$

$$- \frac{1}{2}\sqrt{|\mathcal{A}_i|} \sum_{j \ne i} \|\boldsymbol{x}_j^{(t)} - \boldsymbol{x}_j^{(t-1)}\|_2^2.$$

Summing these inequalities for all $i \in [n]$ gives us that

$$\left\langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \right\rangle - \frac{1}{\eta} \sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) \ge$$

$$\sum_{i=1}^{n} \left(\frac{1}{2\eta} - \frac{1}{2}\sqrt{|\mathcal{A}_i|}(n-1) \right) \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - \frac{1}{2} \left(\sum_{j \ne i} \sqrt{|\mathcal{A}_j|} \right) \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_2^2.$$

Thus, from (35) for $\eta \le \frac{1}{2L}$ we conclude that

$$\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \ge \sum_{i=1}^{n} \left(\left(\frac{1}{2\eta} - \frac{1}{2}\sqrt{|\mathcal{A}_i|}(n-1) \right) \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - \frac{1}{2} \left(\sum_{j \ne i} \sqrt{|\mathcal{A}_j|} \right) \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_2^2 \right).$$

As a result, as long as

$$\eta \le \min \left\{ \frac{1}{2L}, \frac{1}{2\sqrt{|\mathcal{A}_i|}(n-1)} \right\},$$

for all $i \in [n]$, the previous bound implies that

$$\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \ge \sum_{i=1}^{n} \left(\frac{1}{4\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - \frac{1}{2} \left(\sum_{j \ne i} \sqrt{|\mathcal{A}_j|} \right) \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|_2^2 \right).$$

Thus, a telescopic summation leads to the following conclusion:

Theorem B.5. *Suppose that each player $i \in [n]$ employs (OMWU) with learning rate $\eta > 0$ such that*

$$\eta \le \min \left\{ \frac{1}{2L}, \frac{1}{2\sqrt{|\mathcal{A}_i|}(n-1)}, \frac{1}{4 \sum_{j \ne i} \sqrt{|\mathcal{A}_j|}} \right\},$$

for all $i \in [n]$, where L is defined as in Proposition 4.2. Then, it holds that

$$\frac{1}{8\eta} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 \le \Phi(\boldsymbol{x}^{(T)}) - \Phi(\boldsymbol{x}^{(1)}) \le 2\,\Phi_{\max}\,.$$

Finally, Theorem B.5 along with the RVU bound (which holds for (OMWU) as specified in Proposition 2.3) and Claim A.1 conclude the proof of the theorem. $\square$

Next, we proceed with the proof of Theorem 4.7, the detailed version of which is given below.

Theorem B.6 (Full Version of Theorem 4.7). *Suppose that each player i employs (MD) with $\mathcal{R}_i(\boldsymbol{x})$ such that $\nabla \mathcal{R}_i$ is G-Lipschitz continuous with respect to the ℓ_2-norm, and $\eta = \frac{1}{2L}$, where L is defined as in Proposition 4.2. Then, after $O(1/\epsilon^2)$ iterations there exists a joint strategy $\boldsymbol{x}^{(t)}$ which is an ϵ-approximate Nash equilibrium.*

For the proof of this theorem we will use the following simple claim.

Claim B.7. *Suppose that each player employs (MD) with learning rate $\eta > 0$ and regularizer $\mathcal{R}_i$ such that $\nabla \mathcal{R}_i$ is G-Lipschitz continuous with respect to the ℓ_2-norm. If $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2 \le \epsilon$, it holds that $\boldsymbol{x}^{(t)}$ is an*

$$\epsilon \left(\frac{G\Omega}{\eta} + \sqrt{|\mathcal{A}|} \right)$$

approximate Nash equilibrium, where $\Omega := \max_{i \in [n]} \sup_{\boldsymbol{x}_i, \boldsymbol{x}_i' \in \Delta(\mathcal{A}_i)} \|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2$, and $|\mathcal{A}| := \max_{i \in [n]} |\mathcal{A}_i|$.

Proof. The argument proceeds similarly to the proof of Theorem A.12. First, observe that for each $i \in [n]$ it holds that $\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \le \epsilon$. By definition of (MD) we have that

$$\boldsymbol{x}_i^{(t+1)} = \arg\max_{\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)} \left\{ \langle \boldsymbol{x}_i, \boldsymbol{u}_i^{(t)} \rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i, \boldsymbol{x}_i^{(t)}) \right\}.$$

This maximization problem can be equivalently expressed in the following variational inequality form:

$$\left\langle \boldsymbol{u}_i^{(t)} - \frac{1}{\eta} \left(\nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t+1)}) - \nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t)}) \right), \widehat{\boldsymbol{x}}_i - \boldsymbol{x}_i^{(t+1)} \right\rangle \le 0, \quad \forall \widehat{\boldsymbol{x}}_i \in \Delta(\mathcal{A}_i),$$

for any $i \in [n]$. Thus, it follows that

$$\langle \boldsymbol{u}_i^{(t)}, \widehat{\boldsymbol{x}}_i - \boldsymbol{x}_i^{(t+1)} \rangle \le \frac{1}{\eta} \langle \nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t+1)}) - \nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t)}), \widehat{\boldsymbol{x}}_i - \boldsymbol{x}_i^{(t+1)} \rangle$$

$$\le \frac{1}{\eta} \|\nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t+1)}) - \nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t)})\|_2 \|\widehat{\boldsymbol{x}}_i - \boldsymbol{x}_i^{(t+1)}\|_2 \tag{47}$$

$$\le \epsilon \frac{G\Omega_i}{\eta}. \tag{48}$$

where (47) follows from the Cauchy-Schwarz inequality, and (48) uses the fact that $\|\nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t+1)}) - \nabla \mathcal{R}_i(\boldsymbol{x}_i^{(t)})\|_2 \le G\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \le \epsilon G$, which follows from the assumption that $\nabla \mathcal{R}_i$ is G-Lipschitz continuous. Also note that (48) uses the notation $\Omega_i := \sup_{\boldsymbol{x}_i, \boldsymbol{x}_i' \in \Delta(\mathcal{A}_i)} \|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2$. As a result, we have established that for any player $i \in [n]$ it holds that for any $\widehat{\boldsymbol{x}}_i \in \Delta(\mathcal{A}_i)$,

$$\langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t+1)} \rangle \ge \langle \boldsymbol{u}_i^{(t)}, \widehat{\boldsymbol{x}}_i \rangle - \epsilon \frac{G\Omega_i}{\eta}. \tag{49}$$

Moreover, it also follows that

$$\left| \langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \rangle \right| \le \|\boldsymbol{u}_i^{(t)}\|_2 \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \le \sqrt{|\mathcal{A}_i|}\epsilon,$$

where we used the fact that $\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2 \leq \epsilon$, and that $\|\boldsymbol{u}_i^{(t)}\|_\infty \leq 1$ (by the normalization hypothesis). Plugging the last bound to (49) gives us that

$$\langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t)} \rangle \geq \langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t+1)} \rangle - \epsilon\sqrt{|\mathcal{A}_i|} \geq \langle \boldsymbol{u}_i^{(t)}, \widehat{\boldsymbol{x}}_i \rangle - \epsilon\frac{G\Omega_i}{\eta} - \epsilon\sqrt{|\mathcal{A}_i|}, \tag{50}$$

for any $\widehat{\boldsymbol{x}}_i \in \Delta(\mathcal{A}_i)$ and player $i \in [n]$. Finally, the proof follows by definition of approximate Nash equilibrium. □

Proof of Theorem B.6. Suppose that $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2 > \epsilon$ for all $t \in [T]$. Corollary 4.4 implies that

$$4\eta\,\Phi_{\max} \geq \sum_{t=1}^{T-1} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \geq (T-1)\epsilon^2 \implies T \leq \frac{4\eta\,\Phi_{\max}}{\epsilon^2} + 1. \tag{51}$$

Hence, for $T > \left\lceil \frac{4\eta\,\Phi_{\max}}{\epsilon^2} \right\rceil + 1$ it must be the case that there exists $t \in [T]$ such that $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2 \leq \epsilon$. Thus, the theorem follows directly from Claim B.7. □

Proposition 4.8. *In the setting of Theorem 4.3, if the potential function Φ is also concave it holds that*

$$\Phi(\boldsymbol{x}^*) - \Phi(\boldsymbol{x}^{(T+1)}) \leq \frac{2L}{T} \sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(1)}).$$

Proof. The proof proceeds similarly to that in (Birnbaum et al., 2011, Lemma 4). We will first require an auxiliary lemma regarding the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & g(\boldsymbol{x}) - D_{\mathcal{R}}(\boldsymbol{x}, \boldsymbol{y}); \\ \text{subject to} \quad & \boldsymbol{x} \in C, \end{aligned} \tag{52}$$

where g is a concave function on a convex and compact domain C.

Lemma B.8 (*e.g.*, see (Chen & Teboulle, 1993)). *Let $\widehat{\boldsymbol{x}}$ be the optimal solution to the optimization problem (52). Then, it holds that*

$$g(\boldsymbol{x}) - D_{\mathcal{R}}(\boldsymbol{x}, \boldsymbol{y}) \leq g(\widehat{\boldsymbol{x}}) - D_{\mathcal{R}}(\widehat{\boldsymbol{x}}, \boldsymbol{y}) - D_{\mathcal{R}}(\boldsymbol{x}, \widehat{\boldsymbol{x}}).$$

Next, we will apply this lemma to conclude that for all $i \in [n]$ it holds that

$$\eta\langle \nabla_{\boldsymbol{x}_i}\Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}_i^* - \boldsymbol{x}_i^{(t)} \rangle - D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(t)}) \leq \eta\langle \nabla_{\boldsymbol{x}_i}\Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \rangle - D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) - D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(t+1)}). \tag{53}$$

Moreover, from (35) it follows that

$$-\Phi(\boldsymbol{x}^{(t+1)}) \leq -\Phi(\boldsymbol{x}^{(t)}) - \langle \nabla_{\boldsymbol{x}}\Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \rangle + 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)})$$

$$\leq -\Phi(\boldsymbol{x}^{(t)}) - \langle \nabla_{\boldsymbol{x}}\Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}^* - \boldsymbol{x}^{(t)} \rangle + 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(t)}) - 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(t+1)}) \tag{54}$$

$$\leq -\Phi(\boldsymbol{x}^*) + 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}^*, \boldsymbol{x}_i^{(t)}) - 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(t+1)}), \tag{55}$$

where (54) follows from (53) for $\eta = \frac{1}{2L}$, and (55) follows by concavity of Φ. As a result, summing (55) for all $t \in [T]$ and removing the telescopic terms yields that

$$\sum_{t=1}^{T} \left(\Phi(\boldsymbol{x}^*) - \Phi(\boldsymbol{x}^{(t+1)}) \right) \leq 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(1)}) - 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(T+1)}) \leq 2L\sum_{i=1}^{n} D_{\mathcal{R}_i}(\boldsymbol{x}_i^*, \boldsymbol{x}_i^{(1)}), \tag{56}$$

where in the last inequality we used the well-known fact that $D_{\mathcal{R}_i}(\cdot, \cdot) \geq 0$. Finally, Theorem 4.3 implies that

$$\sum_{t=1}^{T} \Phi(\boldsymbol{x}^*) - \Phi(\boldsymbol{x}^{(t+1)}) \geq T\Phi(\boldsymbol{x}^*) - T\Phi(\boldsymbol{x}^{(T+1)}),$$

and plugging this bound to (56) concludes the proof after a rearrangement. □

B.1 Near-Potential Games

In this section we extend some of the results established for (weighted) potential games to *near-potential* games—in the sense of Candogan et al. (2013)—by proving Theorem 4.10. Intuitively, a game Γ is said to be near-potential if it is *close* to some potential game. While there are many natural ways to measure the distance between two games, here we follow the one suggested by Candogan et al. (2013); namely, the *maximum pairwise difference*:

Definition B.9 (Maximum Pairwise Difference (MPD); (Candogan et al., 2013))**.** Let Γ and $\widehat{\Gamma}$ be two (normal-form) games. The *maximum pairwise difference* between Γ and $\widehat{\Gamma}$ is defined as

$$d(\Gamma, \widehat{\Gamma}) := \sup_{i \in [n], \boldsymbol{a} \in \mathcal{A}} |(u_i(a_i, \boldsymbol{a}_{-i}) - u_i(a_i', \boldsymbol{a}_{-i})) - (\widehat{u}_i(a_i, \boldsymbol{a}_{-i}) - \widehat{u}_i(a_i', \boldsymbol{a}_{-i}))|,$$

where $u_i : \prod_{i \in [n]} \mathcal{A}_i \to [-1, 1]$ and $\widehat{u}_i : \prod_{i \in [n]} \mathcal{A}_i \to [-1, 1]$ are the utility functions associated with player i for Γ and $\widehat{\Gamma}$.

This definition tacitly posits that the games are compatible in the sense that the set of actions available to each player coincide. MPD captures the difference between two games in terms of the maximum possible utility improvement through unilateral deviations. Different distance measures can be considered without qualitatively altering the rest of the analysis. Armed with Definition B.9, we are ready to introduce the concept of a near-potential game.

Definition B.10 (Near-Potential Game; (Candogan et al., 2013))**.** A game Γ is δ-*near-potential* if there exists a (compatible) potential game $\widehat{\Gamma}$ such that $d(\Gamma, \widehat{\Gamma}) \leq \delta$.

We remark that (Candogan et al., 2010; 2011) have devised a framework for efficiently finding the nearest potential game to a given game when the distance is measured in terms of the MPD. Specifically, they show that this can be formulated as a convex optimization problem (Candogan et al., 2010; 2011). Nevertheless, it is clear that following a given update rule in a game Γ does not require any sort of knowledge regarding the closest potential game $\widehat{\Gamma}$; the potential function of $\widehat{\Gamma}$ will only be used for the purpose of the analysis. We begin by pointing out the following simple claim.

Claim B.11. *Let Γ be any δ-near-potential game with utilities $u_i : \prod_{i \in [n]} \mathcal{A}_i \to [-1, 1]$, for all $i \in [n]$. Moreover, let Φ be the potential function of a game $\widehat{\Gamma}$ such that $d(\Gamma, \widehat{\Gamma}) = \delta$. Then,*

$$\frac{\partial \Phi(\boldsymbol{x})}{\partial \boldsymbol{x}_i(a_i)} = u_i(a_i, \boldsymbol{x}_{-i}) + e_i(a_i; \boldsymbol{x}_{-i}), \tag{57}$$

where $|e_i(a_i, \boldsymbol{x}_{-i})| = O(\delta)$, for any $i \in [n]$, $a_i \in \mathcal{A}_i$, and $\boldsymbol{x}_{-i} \in \prod_{j \neq i} \Delta(\mathcal{A}_j)$.

Now we are ready to prove Theorem 4.10, the full version of which is given below.

Theorem B.12 (Full Version of Theorem 4.10)**.** *Consider a δ-nearly-potential game wherein every player employs (MD) with learning rate $\eta = \frac{1}{2L}$, where L is a parameter associated with the closest potential game, and regularizer $\mathcal{R}_i$ which is 1-strongly convex and G-smooth with respect to $\|\cdot\|_2$. Then, there exists a bounded potential function Φ which increases as long as $\boldsymbol{x}^{(t)}$ is not an $O(\sqrt{\delta})$-approximate Nash equilibrium.*

Proof. Since players employ (MD), we know that the update rule of each player $i \in [n]$ takes the form

$$\boldsymbol{x}_i^{(t+1)} = \arg\max_{\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)} \left\{ \left\langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i - \boldsymbol{x}_i^{(t)} \right\rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i, \boldsymbol{x}_i^{(t)}) \right\}.$$

By the 1-strong convexity of $\mathcal{R}_i$ with respect to $\|\cdot\|_2$, this implies that

$$\left\langle \boldsymbol{u}_i^{(t)}, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) \geq \frac{1}{2\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2. \tag{58}$$

However, by Claim B.11 we also know that

$$\boldsymbol{u}_i^{(t)} = \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t)}) + \boldsymbol{e}_i^{(t)},$$

where Φ is the potential function of the potential game within δ distance from the original game, and $\boldsymbol{e}_i^{(t)}$ is a vector such that $\|\boldsymbol{e}_i^{(t)}\|_\infty = C\delta$, for some parameter $C > 0$. Thus, combing this fact with (58) yields that

$$
\begin{aligned}
\left\langle \nabla_{\boldsymbol{x}_i} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle - \frac{1}{\eta} D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) &\geq \frac{1}{2\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - \left\langle \boldsymbol{e}_i^{(t)}, \boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)} \right\rangle \\
&\geq \frac{1}{2\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - \|\boldsymbol{e}_i^{(t)}\|_\infty \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1 \\
&\geq \frac{1}{2\eta} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - 2C\delta, \qquad (59)
\end{aligned}
$$

where we used the fact that the ℓ_1 diameter of $\Delta(\mathcal{A}_i)$ is 2 in the last derivation. Summing (59) for all $i \in [n]$ gives us that

$$
\left\langle \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}^{(t)}), \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \right\rangle - \frac{1}{\eta} \sum_{i=1}^n D_{\mathcal{R}_i}(\boldsymbol{x}_i^{(t+1)}, \boldsymbol{x}_i^{(t)}) \geq \frac{1}{2\eta} \sum_{i=1}^n \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - 2Cn\delta.
$$

Thus, using the smoothness condition of (35) with $\eta = \frac{1}{2L}$ implies that

$$
\Phi(\boldsymbol{x}^{(t+1)}) - \Phi(\boldsymbol{x}^{(t)}) \geq \frac{1}{2\eta} \sum_{i=1}^n \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_2^2 - 2Cn\delta.
$$

As a result, we conclude that Φ increases as long as

$$
\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2 \geq 2\sqrt{\eta Cn\delta},
$$

and the theorem follows directly from Claim B.7. $\qquad \square$

B.2 Fisher Markets

In this subsection we illustrate how the framework we developed in Section 4 unifies the work of Birnbaum et al. (2011) related to distributed dynamics in Fisher's classical market model. While the following exposition focuses on the linear counterpart of Fisher's model, we refer the interested reader to the work of Birnbaum et al. (2011) for an elegant extension to the *spending constraints model*. Regarding the motivation for studying distributed dynamics in markets, let us quote from the work of Birnbaum et al. (2011):

> *"Algorithmic results in a centralized model of computation do not directly address the question of market dynamics: how might agents interacting in a market arrive at an equilibrium? Here, the quest is for simple and distributed algorithms that are guaranteed to converge fast. Such distributed algorithms are especially applicable when the agents involved are automated, and one has to prescribe a particular protocol for them to follow"*

In this context, we commence by recalling the underlying model. The exposition in the sequel follows that in (Birnbaum et al., 2011). In the linear Fisher's market model there are n agents (bidders) and m (infinitely) *divisible* goods. It is assumed—without any loss of generality—that there is a unit supply from each good. Every agent has an overall *budget* B_i, circumscribing its buying power. The goal of each agent $i \in [n]$ is to maximize its own utility, which, for a given *allocation vector* $\boldsymbol{x} \in \mathbb{R}_{\geq 0}^{n \times m}$, is defined as $\sum_j \boldsymbol{x}_i(j) \boldsymbol{u}_i(j)$, where $\boldsymbol{u}_i(j)$ represents the *utility* of that agent for a unit of good j.

Equilibrium Conditions Consider an allocation vector $\boldsymbol{x} \in \mathbb{R}_{\geq 0}^{n \times m}$ and a *price vector* $\boldsymbol{p} \in \mathbb{R}_{\geq 0}^m$. The pair $(\boldsymbol{x}, \boldsymbol{p})$ is said to be an *equilibrium* if the following conditions are met.

1. *Buyer optimality*: Each agent $i \in [n]$ maximizes its own utility subject to the budget constraints. Formally, for the given price vector $\boldsymbol{p}$, it has to be that the allocation vector $\boldsymbol{x}$ maximizes the following (linear) program:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{j \in [m]} \boldsymbol{u}_i(j) \boldsymbol{x}_i(j), \\
\text{subject to} \quad & \sum_{j \in [m]} \boldsymbol{p}(j) \boldsymbol{x}_i(j) \leq B_i; \\
& \boldsymbol{x}_i(j) \geq 0, \quad \forall j \in [m].
\end{aligned}
$$

2. *Market clearance*: It must be the case that $\sum_{i\in[n]} \boldsymbol{x}_i(j) = 1$, for all $j \in [m]$.

Convex Formulation of Equilibria Equilibria in linear Fisher markets can be determined via the seminal *Eisenberg-Gale convex program* (Eisenberg & Gale, 1959). As in (Birnbaum et al., 2011), here we focus on an alternative formulation due to Shmyrev (2009), which is given below.

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i,j} \boldsymbol{b}_i(j) \log \boldsymbol{u}_i(j) - \sum_{j\in[m]} \boldsymbol{p}(j) \log \boldsymbol{p}(j), \\
\text{subject to} \quad & \sum_{i\in[n]} \boldsymbol{b}_i(j) = \boldsymbol{p}(j), \quad \forall j \in [m]; \\
& \sum_{j\in[m]} \boldsymbol{b}_i(j) = B_i, \quad \forall \in [n]; \\
& \boldsymbol{b}_i(j) \geq 0, \quad \forall i \in [n], j \in [m].
\end{aligned}
\tag{60}
$$

In this convex program each variable $\boldsymbol{b}_i(j)$ represents the amount spent by agent i on good j, so that for a given solution to (60), the induced allocation $\boldsymbol{x}_i(j)$ is given by $\boldsymbol{b}_i(j)/\boldsymbol{p}(j)$. Observe that for any feasible solution to (60) the markets always clears in the sense of Item 2. Moreover, Shmyrev (2009) showed that the optimal solution of (60) is such that each buyer is allocated an optimal bundle of goods, guaranteeing both equilibrium conditions for the associated Fisher market. In the sequel, and for the sake of simplicity, we will let $B_i = 1$.

In this context, we will relate linear Fisher markets with the setting we presented in Section 4. To this end, let $\Phi(\boldsymbol{b})$ be the objective function of (60):

$$
\Phi(\boldsymbol{b}) = \sum_{i,j} \boldsymbol{b}_i(j) \log \boldsymbol{u}_i(j) - \sum_{j\in[m]} \boldsymbol{p}(j) \log \boldsymbol{p}(j) = \sum_{i,j} \boldsymbol{b}_i(j) \log \left(\frac{\boldsymbol{u}_i(j)}{\boldsymbol{p}(j)} \right),
$$

where we used the fact that $\boldsymbol{p}(j) = \sum_{i\in[n]} \boldsymbol{b}_i(j)$, which in turn follows by the feasibility constraints of (60). Now a direct calculation shows that for any $i \in [n]$, each component $j \in [m]$ on the corresponding gradient is such that

$$
(\nabla\Phi(\boldsymbol{b}))_{i,j} = \log \left(\frac{\boldsymbol{u}_i(j)}{\boldsymbol{p}(j)} \right) - 1.
$$

Thus, observe that Φ is not always smooth on the feasible region

$$
S := \left\{ \boldsymbol{b} \in \mathbb{R}^{n\times m} : \sum_{j\in[m]} \boldsymbol{b}_i(j) = B_i, \forall i \in [n], \boldsymbol{b}_i(j) \geq 0, \forall(i,j) \in [n] \times [m] \right\}.
$$

Nevertheless, (Birnbaum et al., 2011, Lemma 7) establishes that Φ satisfies the weaker "one-sided" smoothness condition of Definition 4.1 with respect to the Kullback-Leibler divergence. Moreover, let $g : x \mapsto \log x - 1$ be a monotone transformation, so that

$$
g^{-1} \left((\nabla\Phi(\boldsymbol{b}))_{i,j} \right) = \frac{\boldsymbol{u}_i(j)}{\boldsymbol{p}(j)} = \frac{\boldsymbol{u}_i(j)\boldsymbol{x}_i(j)}{\boldsymbol{b}_i(j)}.
\tag{61}
$$

But the latter expression can be thought of as the utility that agent i received from item j per fraction of budget invested to j. As a result, if every agent employs (MD) with negative entropy DGF, monotone transformation $g : x \mapsto \log x - 1$, and utility vector the feedback suggested by Equation (61), it can be shown that bids are updated using the following update rule:

$$
\boldsymbol{b}_i(j) = \frac{\boldsymbol{u}_i(j)\boldsymbol{x}_i(j)}{\sum_{k\in[m]} \boldsymbol{u}_i(k)\boldsymbol{x}_i(k)} B_i.
\tag{PR}
$$

These dynamics are known us *Proportional Response*, and they were introduced by Wu & Zhang (2007) (see also (Zhang, 2011)). (PR) can be seen as typical multiplicative weights update after applying the monotone transformation g. In this way, these distributed dynamics are tantamount to optimizing Smyrev's convex program (60), which converges to an equilibrium with $O(1/T)$ rate by concavity of Φ; see Proposition 4.8.

An interesting direction would be to incorporate into this framework further and more general market models such as the *Arrow-Debreu (exchange)* version; *e.g.*, see (Panageas et al., 2021) for some recent developments related in spirit to the approach we presented in this section.

C. Proofs from Section 5

In this section we provide the proofs omitted from Section 5. From a technical standpoint we will rely on a fundamental tool from signal processing and control theory; namely, the *Z-transform*. Recall that the (*bilateral*) *Z*-transform of a sequence $(\boldsymbol{x}^{(t)})$ in $\mathbb{R}^d$ is defined as

$$\boldsymbol{X}(z) := \mathcal{Z}\left\{\boldsymbol{x}^{(t)}\right\} = \sum_{t=-\infty}^{\infty} \boldsymbol{x}^{(t)} z^{-t}, \tag{62}$$

where $z \in \mathbb{C}^*$ is assumed to be in the *region of convergence*:

$$\mathrm{ROC} := \left\{ z \in \mathbb{C}^* : \left| \sum_{t=-\infty}^{\infty} \boldsymbol{x}^{(t)} z^{-t} \right| < \infty \right\}.$$

Observe that we define the Z-transform coordinate-wise. Our analysis will leverage the following well-known properties which follow directly from the definition of (62).

Property C.1 (Linearity). *Let $(\boldsymbol{x}^{(t)})_{t=-\infty}^{\infty}$ and $(\boldsymbol{y}^{(t)})_{t=-\infty}^{\infty}$ be sequences in $\mathbb{R}^d$. Then,*

$$\mathcal{Z}\left\{\boldsymbol{x}^{(t)} + \boldsymbol{y}^{(t)}\right\} = \mathcal{Z}\left\{\boldsymbol{x}^{(t)}\right\} + \mathcal{Z}\left\{\boldsymbol{y}^{(t)}\right\}.$$

Property C.2 (Time-Delay Property). *Let $(\boldsymbol{x}^{(t)})_{t=-\infty}^{\infty}$ be a sequence in $\mathbb{R}^d$, and some $t_0 \in \mathbb{R}$. Then, it holds that*

$$\mathcal{Z}\left\{\boldsymbol{x}^{(t-t_0)}\right\} = z^{-t_0} \mathcal{Z}\left\{\boldsymbol{x}^{(t)}\right\}.$$

We are now ready to prove Theorem 5.1, the detailed version of which is given below.

Theorem C.3 (Detailed Version of Theorem 5.1). *Let $\mathbf{A}$ and $\mathbf{B}$ be square and full-rank $d \times d$ matrices, and $\gamma := \rho(\mathbf{A}^\top \mathbf{B})$, where $\rho(\cdot)$ denotes the spectral radius. If the matrix $\mathbf{A}^\top \mathbf{B}$ has strictly negative (real) eigenvalues, it holds that for any learning rate $\eta \leq \frac{1}{2\sqrt{\gamma}}$ (OGD) converges with linear rate to an equilibrium.*

Proof. First, observe that $\nabla_{\boldsymbol{x}}(\boldsymbol{x}^\top \mathbf{A} \boldsymbol{y}) = \mathbf{A}\boldsymbol{y}$ and $\nabla_{\boldsymbol{y}}(\boldsymbol{x}^\top \mathbf{B} \boldsymbol{y}) = \mathbf{B}^\top \boldsymbol{x}$. Thus, (OGD) can be expressed as the following linear dynamical system:

$$\begin{aligned}
\boldsymbol{x}^{(t+1)} &= \boldsymbol{x}^{(t)} + 2\eta \mathbf{A} \boldsymbol{y}^{(t)} - \eta \mathbf{A} \boldsymbol{y}^{(t-1)}; \\
\boldsymbol{y}^{(t+1)} &= \boldsymbol{y}^{(t)} + 2\eta \mathbf{B}^\top \boldsymbol{x}^{(t)} - \eta \mathbf{B}^\top \boldsymbol{x}^{(t-1)}.
\end{aligned} \tag{63}$$

To analyze its convergence properties, we will transfer (63) to the z-space. To this end, let $\boldsymbol{X}(z)$ and $\boldsymbol{Y}(z)$ be the Z-transform of the sequence $(\boldsymbol{x}^{(t)})$ and $(\boldsymbol{y}^{(t)})$ respectively; here it is tacitly assumed that z belongs to the region of convergence. Thus, using linearity (Property C.1) and the time-delay property (Property C.2), it follows from (63) that

$$\begin{aligned}
z\boldsymbol{X}(z) &= \boldsymbol{X}(z) + 2\eta \mathbf{A} \boldsymbol{Y}(z) - \eta z^{-1} \mathbf{A} \boldsymbol{Y}(z); \\
z\boldsymbol{Y} &= \boldsymbol{Y}(z) + 2\eta \mathbf{B}^\top \boldsymbol{X}(z) - \eta z^{-1} \mathbf{B}^\top \boldsymbol{X}(z).
\end{aligned} \tag{64}$$

Note that we also used the fact that $\mathcal{Z}\left\{\mathbf{A}\boldsymbol{y}^{(t)}\right\} = \mathbf{A}\boldsymbol{Y}(z)$ and $\mathcal{Z}\left\{\mathbf{B}^\top \boldsymbol{x}^{(t)}\right\} = \mathbf{B}^\top \boldsymbol{X}(z)$, which follow from linearity of the Z-transform. In this way, we have transferred the *difference equation* (63) to the *algebraic equation* (64). From the latter algebraic system, we may uncouple these equations to conclude that

$$\begin{aligned}
z^2(z-1)^2 \boldsymbol{X}(z) &= (2\eta z - \eta)\mathbf{A}z(z-1)\boldsymbol{Y}(z) = (2\eta z - \eta)\mathbf{A}(2\eta z - \eta)\mathbf{B}^\top \boldsymbol{X}(z) = \eta^2(2z-1)^2 \mathbf{A}\mathbf{B}^\top \boldsymbol{X}(z); \\
z^2(z-1)^2 \boldsymbol{Y}(z) &= (2\eta z - \eta)\mathbf{B}^\top z(z-1)\boldsymbol{X}(z) = (2\eta z - \eta)\mathbf{B}^\top(2\eta z - \eta)\mathbf{A}\boldsymbol{Y}(z) = \eta^2(2z-1)^2 \mathbf{B}^\top \mathbf{A}\boldsymbol{Y}(z).
\end{aligned} \tag{65}$$

As a result, we have derived the *characteristic equation* of the dynamical system (63):

Proposition C.4. *The characteristic equation of* (63) *can be expressed as*

$$\det(\mathbf{I}_d(z(z-1))^2 - \eta^2(2z-1)^2\mathbf{A}^\top\mathbf{B}) = 0. \tag{66}$$

In particular, if $\chi(z)$ is the characteristic polynomial of matrix $\mathbf{A}^\top\mathbf{B}$, it follows that $z \in \mathbb{C}^*$ satisfies (66) *if and only if*

$$\chi\left(\left(\frac{1}{\eta}\frac{z(z-1)}{2z-1}\right)^2\right) = 0. \tag{67}$$

Note that we used the fact that the matrices $\mathbf{A}\mathbf{B}^\top$ and $\mathbf{B}^\top\mathbf{A}$ have identical spectrum, which in turn follows since $\mathbf{A}$ and $\mathbf{B}$ are assumed to be non-singular, as well as the property $\det(\mathbf{M}) = \det(\mathbf{M}^\top)$ in order to derive (66) from (65). To see the second part of the claim in Proposition C.4, first observe that $z = \frac{1}{2}$ is never a solution to (66). Thus, from the multilinearity of the determinant, it holds that

$$\det(\mathbf{I}_d(z(z-1))^2 - \eta^2(2z-1)^2\mathbf{A}^\top\mathbf{B}) = 0 \iff \det\left(\left(\frac{1}{\eta}\frac{z(z-1)}{2z-1}\right)^2\mathbf{I}_d - \mathbf{A}^\top\mathbf{B}\right) = 0.$$

As a result, the equivalence asserted in (67) follows by recalling that $\chi(\lambda) = \det(\lambda\mathbf{I}_d - \mathbf{A}^\top\mathbf{B})$ since $\chi(\lambda)$ is the characteristic polynomial of matrix $\mathbf{A}^\top\mathbf{B}$.

Having derived the characteristic equation of the dynamical system (Proposition C.4), it remains to derive its solutions. To do this, let $-\lambda$ be such that $\chi(-\lambda) = 0$; by assumption, we know that $\lambda \in \mathbb{R}_{>0}$. Now this eigenvalue induces solutions of the following form:

$$\left(\frac{1}{\eta}\frac{z(z-1)}{2z-1}\right)^2 = -\lambda \iff \frac{z(z-1)}{2z-1} = \pm\eta\sqrt{\lambda}j \iff \begin{cases} z^2 + z(-1 - 2\eta\sqrt{\lambda}j) + \eta\sqrt{\lambda}j = 0; \\ z^2 + z(-1 + 2\eta\sqrt{\lambda}j) - \eta\sqrt{\lambda}j = 0. \end{cases} \tag{68}$$

where $j \in \mathbb{C}$ denotes the imaginary unit. Now observe that $z^2 + z(-1 - 2\eta\sqrt{\lambda}j) + \eta\sqrt{\lambda}j = 0 \iff \bar{z}^2 + \bar{z}(-1 + 2\eta\sqrt{\lambda}j) - \eta\sqrt{\lambda}j = 0$, where $\bar{z}$ denotes the complex conjugate of z. Hence, it suffices to solve only the first equation since their solutions are equivalent in terms of magnitude—in particular, they are complex conjugates. Thus, we obtain the following solutions:

$$z_\pm = \frac{1 + 2\eta\sqrt{\lambda}j \pm \sqrt{1 - 4\eta^2\lambda}}{2}. \tag{69}$$

Moreover, by assumption we know that $\eta \leq \frac{1}{2\sqrt{\gamma}}$, which in turn implies that $\eta \leq \frac{1}{2\sqrt{\lambda}} \iff 1 - 4\eta^2\lambda \geq 0$; this holds by definition of $\gamma := \rho(\mathbf{A}^\top\mathbf{B})$. Thus, it follows that

$$|z_+|^2 = \frac{1}{4}\left((1 + \sqrt{1 - 4\eta^2\lambda})^2 + (2\eta\sqrt{\lambda})^2\right) = \frac{1}{2}\left(1 + \sqrt{1 - 4\eta^2\lambda}\right) < 1;$$

$$|z_-|^2 = \frac{1}{4}\left((1 - \sqrt{1 - 4\eta^2\lambda})^2 + (2\eta\sqrt{\lambda})^2\right) = \frac{1}{2}\left(1 - \sqrt{1 - 4\eta^2\lambda}\right) < 1.$$

This implies that all of the solutions to the characteristic equation (67) lie within the unit circle on the complex plane. Thus, the theorem follows from the fundamental theorem of linear dynamical systems. $\qquad\square$

Proposition 5.2. *For any sufficiently large $R > 0$, there exist games such that* (OGD) *converges under any initialization to an equilibrium $(\boldsymbol{x}^{(\infty)}, \boldsymbol{y}^{(\infty)})$ such that* $\mathrm{SW}(\boldsymbol{x}^{(\infty)}, \boldsymbol{y}^{(\infty)}) = 0$, *while there exist an equilibrium $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ with $\mathrm{SW}(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq 2R^2$.*

Proof. We consider the game described with the following matrices:

$$\mathbf{A} := \begin{bmatrix} 1 & -2 \\ -1 & 1 \end{bmatrix}; \mathbf{B} := \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{70}$$

We also assume that the constraints sets are such that $\mathcal{X} = \mathcal{B}_1(\mathbf{0}, R)$ and $\mathcal{Y} = \mathcal{B}_1(\mathbf{0}, R)$, for a sufficiently large radius $R > 0$. Here $\mathcal{B}_1$ denotes the closed ℓ_1-ball; we only use the ℓ_1-norm for mathematical convenience. We will first consider

the *unconstrained dynamics*. In particular, we will show that the condition of Theorem 5.1 is met. Indeed, a direct calculation reveals that

$$\mathbf{A}^\top \mathbf{B} = \begin{bmatrix} 0 & 2 \\ -1 & -3 \end{bmatrix}.$$

Now the characteristic equation of $\mathbf{A}^\top \mathbf{B}$ reads $\lambda(\lambda + 3) + 2 = 0 \iff \lambda_1 = -2, \lambda_2 = -1$. Thus, $\mathbf{A}^\top \mathbf{B}$ has negative (real) eigenvalues. As a result, Theorem 5.1 implies that for a sufficiently small learning rate (OGD) converges under any initialization—assuming unconstrained domains (see Figure 8). Moreover, the following claim characterizing its limit points is immediate.

Claim C.5. *Suppose that* (63) *converges to a point* $(\boldsymbol{x}^{(\infty)}, \boldsymbol{y}^{(\infty)})$. *Then, it holds that* $\mathbf{A}\boldsymbol{y}^{(\infty)} = \mathbf{0}$ *and* $\mathbf{B}^\top \boldsymbol{x}^{(\infty)} = \mathbf{0}$.

In particular, given that the matrices $\mathbf{A}$ and $\mathbf{B}$ are full-rank, this claim implies that $\boldsymbol{x}^{(\infty)} = \mathbf{0}$ and $\boldsymbol{y}^{(\infty)} = \mathbf{0}$. Moreover, for a sufficiently large $R > 0$ we know that the projected dynamics on $\mathcal{X}$ and $\mathcal{Y}$ (respectively) will coincide with the unconstrained dynamics. As a result, we have shown that (OGD) constrained on $\mathcal{X}$ and $\mathcal{Y}$ will converge to $(\mathbf{0}, \mathbf{0})$, which is clearly an equilibrium point. However, there exists a much more efficient equilibrium:

Claim C.6. *Let* $\boldsymbol{x}^* = (R, 0) \in \mathcal{B}_1(\mathbf{0}, R)$ *and* $\boldsymbol{y}^* = (R, 0) \in \mathcal{B}_1(\mathbf{0}, R)$. *Then,* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an equilibrium of the game* (70).

Proof. When player $\mathcal{Y}$ plays $\boldsymbol{y}^*$ it follows that $\mathbf{A}\boldsymbol{y}^* = (R, -R)$. Thus, it follows that $\boldsymbol{x}^\top \mathbf{A}\boldsymbol{y}^* \leq \|\boldsymbol{x}\|_1 \|\mathbf{A}\boldsymbol{y}^*\|_\infty \leq R^2 = (\boldsymbol{x}^*)^\top \mathbf{A}\boldsymbol{y}^*$, for any $\boldsymbol{x} \in \mathcal{B}_1(\mathbf{0}, R)$. Thus, $\boldsymbol{x}^*$ is indeed a best response to $\boldsymbol{y}^*$. Similarly, when player $\mathcal{X}$ plays $\boldsymbol{x}^*$ it follows that $\mathbf{B}^\top \boldsymbol{x}^* = (R, R)$. Thus, we conclude that $(\boldsymbol{x}^*)^\top \mathbf{B}\boldsymbol{y} \leq \|\boldsymbol{y}\|_1 \|\mathbf{B}^\top \boldsymbol{x}^*\|_\infty \leq R^2 = (\boldsymbol{x}^*)^\top \mathbf{B}\boldsymbol{y}^*$, for any $\boldsymbol{y} \in \mathcal{B}_1(\mathbf{0}, R)$. This means that $\boldsymbol{y}^*$ is indeed a best response to $\boldsymbol{x}^*$, verifying our claim. $\square$

Finally, we have that $\mathrm{SW}(\boldsymbol{x}^*, \boldsymbol{y}^*) = (\boldsymbol{x}^*)^\top \mathbf{A}\boldsymbol{y}^* + (\boldsymbol{x}^*)^\top \mathbf{B}\boldsymbol{y}^* = 2R^2$. This concludes the proof. $\square$

Proposition 5.3. *For any* $\epsilon > 0$ *there exists a game* $(\mathbf{A}, \mathbf{B})$ *with* $\|\mathbf{A} + \mathbf{B}\|_F \leq \epsilon$ *for which* (OGD) *diverges, while the dynamics converge for the game* $(\mathbf{A}, -\mathbf{A})$.

Proof. Fix any $\epsilon > 0$. We consider the game $(\mathbf{A}, \mathbf{B})$ described with the following payoff matrices:

$$\mathbf{A} := \begin{bmatrix} 1 & 0 \\ 0 & \epsilon/2 \end{bmatrix}; \mathbf{B} := \begin{bmatrix} -1 & 0 \\ 0 & \epsilon/2 \end{bmatrix}. \tag{71}$$

Observe that $\|\mathbf{A} + \mathbf{B}\|_F = \epsilon$.[8] Let us first argue about convergence of (OGD) in the game $(\mathbf{A}, -\mathbf{A})$. To this end, observe that the matrix $\mathbf{A}^\top \mathbf{A}$ has only positive eigenvalues. Thus, we know from Theorem 5.1 that (OGD) converges for $\eta \leq \frac{1}{2}$ since the spectral radius of $\mathbf{A}^\top \mathbf{A}$ is 1. On the other hand, for the game $(\mathbf{A}, \mathbf{B})$ it holds that the matrix $\mathbf{A}^\top \mathbf{B}$ has a positive eigenvalue; namely, $\lambda = \epsilon^2/4$. But from Proposition C.4 it can be shown that this implies that the characteristic equation of the associated dynamical system has a solution $z \in \mathbb{C}^*$ with $|z| > 1$. In turn, this implies that (OGD) will diverge under any non-trivial initialization; see Figure 9 for a graphical illustration and a discussion about this phenomenon. $\square$

Next, we proceed with the proof of Theorem 5.4, the detailed version of which is recalled below.

Theorem C.7 (Detailed Version of Theorem 5.4). *Let* $\{\mathbf{A}_{1,j}\}_{j=2}^n$ *and* $\{\mathbf{A}_{j,1}\}_{j=2}^n$ *be square matrices such that* $\det(\mathbf{M}) \neq 0$, *where* $\mathbf{M} := \sum_{j \neq 1} \mathbf{A}_{1,j} \mathbf{A}_{j,1}$. *Moreover, let* $\gamma := \rho(\mathbf{A})$, *where* $\rho(\cdot)$ *denotes the spectral radius. If* $\mathbf{M}$ *has strictly negative (real) eigenvalues, it holds that for any learning rate* $\eta \leq \frac{1}{2\sqrt{\gamma}}$ (OGD) *converges with linear rate to an equilibrium.*

Proof. First of all, recall that we have that $u_1(\boldsymbol{x}) = \sum_{j \neq 1} \boldsymbol{x}_1^\top \mathbf{A}_{1,j} \boldsymbol{x}_j$, while $u_j(\boldsymbol{x}) = \boldsymbol{x}_j^\top \mathbf{A}_{j,1} \boldsymbol{x}_1$ for $j \neq 1$. Thus, it follows that $\nabla_{\boldsymbol{x}_1} u_1(\boldsymbol{x}) = \sum_{j \neq 1} \mathbf{A}_{1,j} \boldsymbol{x}_j$, and $\nabla_{\boldsymbol{x}_j} u_j(\boldsymbol{x}) = \mathbf{A}_{j,1} \boldsymbol{x}_1$ for $j \neq 1$. As a result, (OGD) can be cast as

$$\begin{aligned} \boldsymbol{x}_1^{(t+1)} &= \boldsymbol{x}_1^{(t)} + 2\eta \sum_{j \neq 1} \mathbf{A}_{1,j} \boldsymbol{x}_j^{(t)} - \eta \sum_{j \neq 1} \mathbf{A}_{1,j} \boldsymbol{x}_j^{(t-1)}; \\ \boldsymbol{x}_j^{(t+1)} &= \boldsymbol{x}_j^{(t)} + 2\eta \mathbf{A}_{j,1} \boldsymbol{x}_1^{(t)} - \eta \mathbf{A}_{j,1} \boldsymbol{x}_1^{(t-1)}, \forall j \neq 1. \end{aligned} \tag{72}$$

---

[8]Recall that the Frobenius norm of a matrix $\mathbf{M}$ is defined as $\|\mathbf{M}\|_F := \|(\mathbf{M})^\flat\|_2$, where $(\mathbf{M})^\flat$ is the standard vectorization of $\mathbf{M}$.

Transferring these dynamics to the z-space yields that

$$z\boldsymbol{X}_1(z) = \boldsymbol{X}_1(z) + 2\eta \sum_{j \neq 1} \mathbf{A}_{1,j} \boldsymbol{X}_j(z) - \eta z^{-1} \sum_{j \neq 1} \mathbf{A}_{1,j} \boldsymbol{X}_j(z);$$
$$z\boldsymbol{X}_j(z) = \boldsymbol{X}_j(z) + 2\eta \mathbf{A}_{j,1} \boldsymbol{X}_1(z) - \eta z^{-1} \mathbf{A}_{j,1} \boldsymbol{X}_1(z), \forall j \neq 1.$$

(73)

From the second equation it follows that $(z^2 - z)\boldsymbol{X}_j(z) = \eta(2z - 1)\mathbf{A}_{j,1}\boldsymbol{X}_1(z)$, for all $j \neq 1$. Thus, plugging this to the first equation of (73) yields that

$$z^2(z-1)^2 \boldsymbol{X}_1(z) = \eta^2(2z-1)^2 \sum_{j \neq 1} \mathbf{A}_{1,j}\mathbf{A}_{j,1}\boldsymbol{X}_1(z) \iff \left((z(z-1))^2 \mathbf{I}_d - \eta^2(2z-1)^2 \sum_{j \neq 1} \mathbf{A}_{1,j}\mathbf{A}_{j,1} \right) \boldsymbol{X}_1(z) = 0.$$

From this, it is possible to conclude the characteristic equation of the associated dynamical system:

Proposition C.8. *The characteristic equation of the dynamical system* (72) *can be expressed as*

$$\det \left((z(z-1))^2 \mathbf{I}_d - \eta^2(2z-1)^2 \sum_{j \neq 1} \mathbf{A}_{1,j}\mathbf{A}_{j,1} \right) = 0.$$

Observe that if all the poles of $\boldsymbol{X}_1(z)$ lie within the unit circle of the complex plane, the same holds for each $\boldsymbol{X}_j(z)$, for all $j \neq 1$, modulo at most a single pole at $z = 1$. This follows given that $(z^2 - z)\boldsymbol{X}_j(z) = \eta(2z - 1)\mathbf{A}_{j,1}\boldsymbol{X}_1(z)$, for all $j \neq 1$. Finally, the rest of the theorem follows analogously to Theorem C.3, while it is direct to verify that, assuming convergence, the limit points satisfy the equilibrium conditions. $\square$

Remark C.9. More broadly, the convergence of (OGD) can be determined in terms of the spectrum of the matrix

$$\mathbf{M} := \begin{pmatrix} \mathbf{0}_d & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \dots & \mathbf{A}_{1,n} \\ \mathbf{A}_{2,1} & \mathbf{0}_d & \mathbf{A}_{2,3} & \dots & \mathbf{A}_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{n,1} & \mathbf{A}_{n,2} & \mathbf{A}_{n,3} & \dots & \mathbf{0}_d \end{pmatrix}.$$

Indeed, Theorem 5.4 can be seen as an instance of such a broader characterization for *separable* continuous games. As we point out in Remark C.10, obtaining such results in multilinear settings would require different techniques.

Remark C.10 (Beyond Polymatrix Games). We believe that results in unconstrained multilinear n-player games (akin to standard NFGs, but without constraints) can be obtained using recent advances from *multilinear control theory* (Chen et al., 2021). In particular, stability could be determined in terms of the tensor of the underlying game.

Our final result for Section 5 shows that a generic class of first-order fails to guarantee stability in all general-sum games, even with two players. More precisely, we consider the following method.

$$\boldsymbol{x}_i^{(t+1)} = \sum_{\tau=0}^{T} \alpha^{(\tau)} \boldsymbol{x}_i^{(t-\tau)} + \sum_{\tau=0}^{T} \beta^{(\tau)} \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}^{(t-\tau)}).$$

(HGD)

This method—which will be referred to as *historical gradient descent (HGD)*—can be described with the ordered set of coefficients $A := (\alpha^{(0)}, \dots, \alpha^{(T)})$ and $B := (\beta^{(0)}, \dots, \beta^{(T)})$. In particular, (HGD) contains (OGD) when $A = (1)$ and $B = (2\eta, -\eta)$, but it goes well-beyond this method. For the purpose of our analysis we are going to represent an (HGD) algorithm using the following two polynomials.

$$S(z) := \alpha^{(0)} + \alpha^{(1)} z^{-1} + \dots + \alpha^{(T)} z^{-T};$$
$$G(z) := \beta^{(0)} + \beta^{(1)} z^{-1} + \dots + \beta^{(T)} z^{-T}.$$

For example, (OGD) is associated with polynomials $S(z) = 1$ and $G(z) = 2\eta - \eta z^{-1}$. However, the considered class of dynamics contains certain trivial algorithms. For example, the update rule of an (HGD) algorithm for which $G(z) \equiv 0$ would not depend on the observed utility at any iteration. To address such trivialities, we impose a condition which ensures

that an (HGD) algorithm is interesting from a game-theoretic standpoint. In particular, we say that an (HGD) algorithm is *regular* if $S(1) = 1$ and $G(1) \neq 0$. Then, it is easy to see that, if both players employ a regular (HGD) algorithm and the dynamics converge, the limit points will be equilibria in the sense of (6). We will also assume without any loss that the polynomials $G(z)$ and $S(z) - z$ have no common roots. We are now ready to state and prove our main theorem on (HGD) algorithms.

Theorem C.11. *For any regular algorithm in* (HGD) *there exists a game for which the method will diverge under any non-trivial initialization.*

Proof. Consider a game $(\mathbf{A}, \mathbf{B})$. The dynamics of the underlying dynamical system in the z-space can be expressed as follows.

$$zX(z) = S(z)X(z) + G(z)\mathbf{A}Y(z); \qquad (z - S(z))X(z) = G(z)\mathbf{A}Y(z);$$
$$zY(z) = S(z)Y(z) + G(z)\mathbf{B}^\top X(z). \qquad \Longleftrightarrow \qquad (z - S(z))Y(z) = G(z)\mathbf{B}^\top X(z).$$

From this representation, we may conclude that the characteristic equation of the dynamical system can be expressed as

$$\det\left((z - S(z))^2 \mathbf{I}_d - (G(z))^2 \mathbf{A}^\top \mathbf{B}\right) = 0 \iff \chi\left(\left(\frac{z - S(z)}{G(z)}\right)^2\right) = 0,$$

where $\chi(z)$ represents the characteristic polynomial of matrix $\mathbf{A}^\top \mathbf{B}$. Note that our previous derivation uses the assumption that the polynomials $S(z) - z$ and $G(z)$ do not share any common roots, in turn implying that no root of $G(z)$ can satisfy the equation $\det\left((z - S(z))^2 \mathbf{I}_d - (G(z))^2 \mathbf{A}^\top \mathbf{B}\right) = 0$. Now consider any $\epsilon > 0$ so that $G(1 + \epsilon) \neq 0$ and $S(1 + \epsilon) \neq 1 + \epsilon$. If the game $(\mathbf{A}, \mathbf{B})$ is such that the matrix $\mathbf{A}^\top \mathbf{B}$ has the positive eigenvalue

$$\lambda = \left(\frac{1 + \epsilon - S(1 + \epsilon)}{G(1 + \epsilon)}\right)^2,$$

then it follows that $z = 1 + \epsilon$ is a solution to the characteristic equation of the system. But given that $|z| > 1$, this necessarily implies that the dynamics will diverge by the fundamental theorem of linear dynamical systems. $\square$

While there is a plethora of control-theoretic techniques that could stabilize the dynamics, we have argued (Proposition 5.2) that stability might be at odds with efficiency—welfare maximization. Understanding the fundamental tension between different solution concepts is an important question for the future.

D. Experiments

In this section we corroborate some our theoretical results through experiments on several games.

D.1 Normal-Form Games   We start by illustrating the last-iterate convergence when players employ different and potentially advanced prediction mechanisms in normal-form games. We are first going to assume that both players employ (OMD) with prediction $m^{(t)}$ set to either H-step recency bias (Item 1), or geometrically δ-discounted recency bias (Item 2). We remark that while Syrgkanis et al. (2015) established RVU bounds only for OFTRL under such predictions, it is immediate to extend these bounds with qualitatively similar results for OMD. That is, the bounds on the learning rate obtained in Proposition A.2 for OFTRL are analogous to the corresponding ones for OMD. Inspired by the work of Daskalakis et al. (2021), we will also experiment with predictions inducing H-order differences in the RVU bound:

(i) 1-order: $m^{(t)} := u^{(t-1)}$;

(ii) 2-order: $m^{(t)} := 2u^{(t-1)} - u^{(t-2)}$;

(iii) 3-order: $m^{(t)} := 3u^{(t-1)} - 3u^{(t-2)} + u^{(t-3)}$;

(iv) 4-order: $m^{(t)} := 4u^{(t-1)} - 6u^{(t-2)} + 4u^{(t-3)} - u^{(t-4)}$.

We point out that our techniques immediate imply last-iterate convergence under H-order predictions, for a sufficiently small learning rate $\eta = \eta(H)$.

Zero-Sum Games We first illustrate the behavior of the dynamics for two-player zero-sum games. We let $\mathbf{A}_j$ represent the cost matrix for player $\mathcal{X}$—and subsequently the payoff matrix for player $\mathcal{Y}$—for $j \in \{1, 2, 3\}$, defined as follows.

$$\mathbf{A}_1 := \begin{bmatrix} 1 & -1 & -1 \\ -1 & -1 & 0 \\ -0.5 & 0 & -1 \end{bmatrix}; \mathbf{A}_2 := \begin{bmatrix} 1 & -2 & -1 \\ -1 & 1 & 0 \\ -0.5 & 1 & -1 \end{bmatrix}; \mathbf{A}_3 := \begin{bmatrix} -1 & 1 & -1 \\ 0 & 0.5 & -1 \\ 0.3 & -0.5 & -0.5 \end{bmatrix}. \tag{74}$$

The (OMD) dynamics when *both* players employ H-step recency bias are illustrated and discussed in Figure 2. Figure 3 illustrates the behavior of the dynamics when only *one* of the two players employs H-step recency bias. The geometrically δ-discounting prediction mechanism is depicted in Figure 4, while Figure 5 shows the dynamics under H-order predictions.
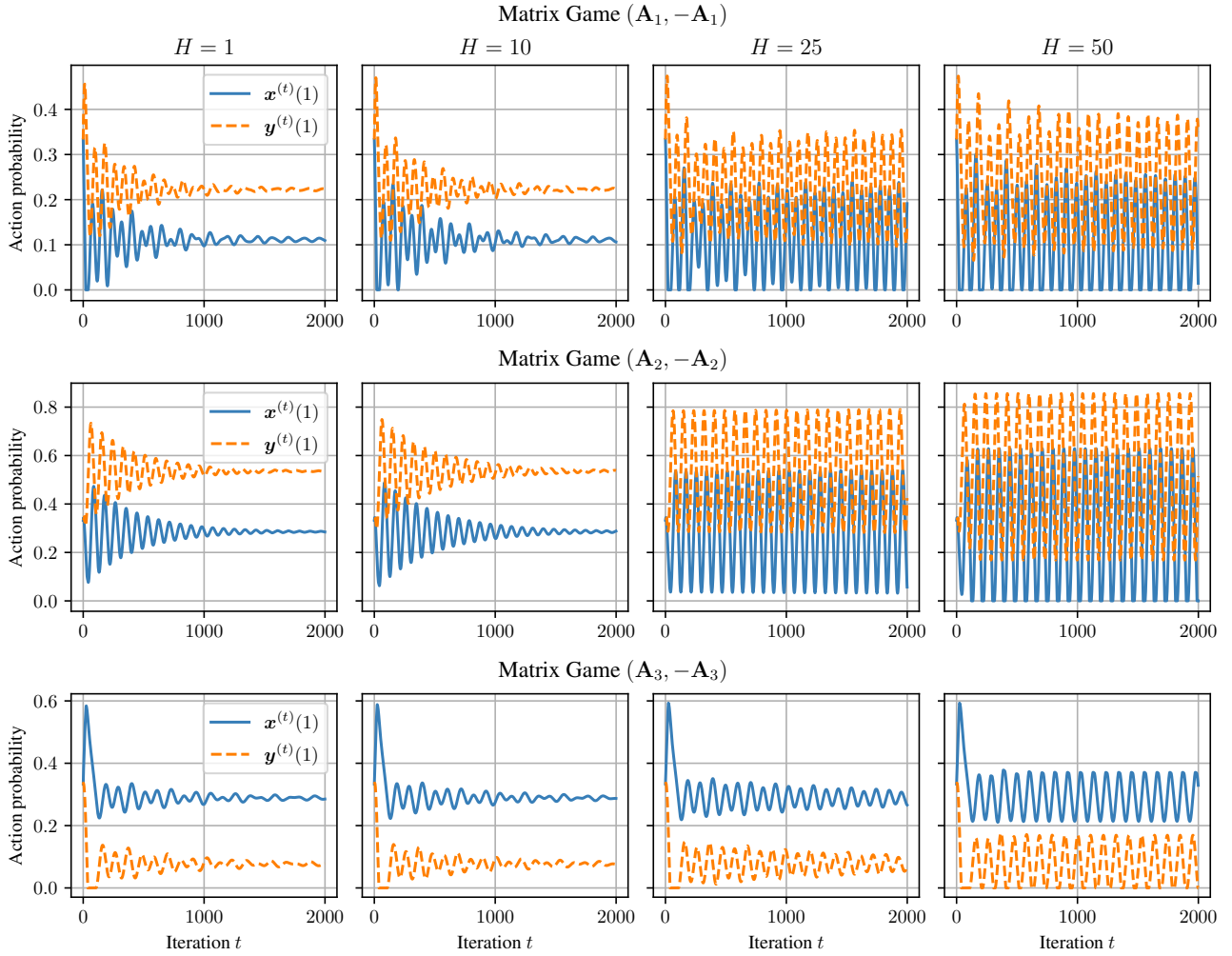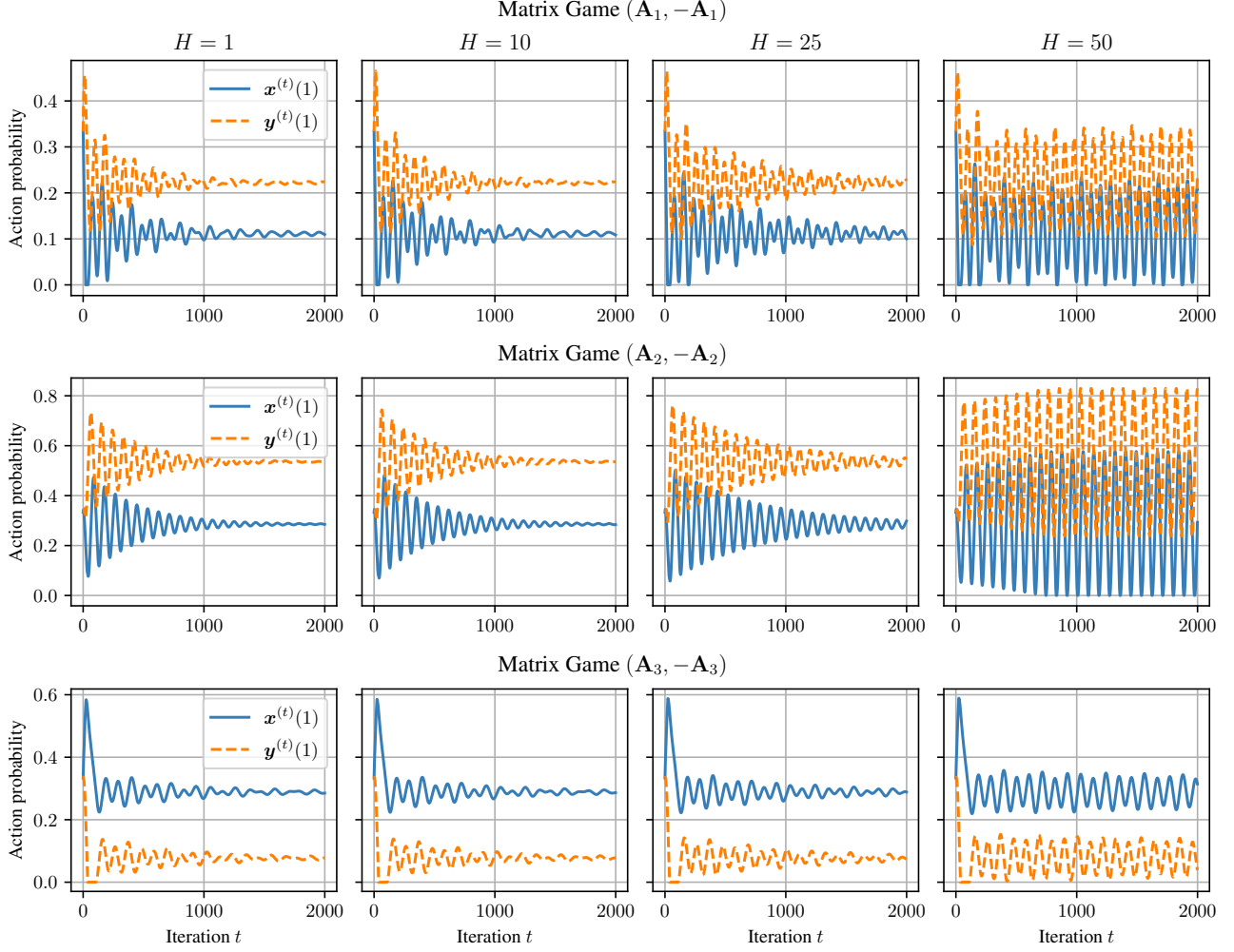

Figure 2. The (OMD) dynamics in the zero-sum games described in (74) for 2000 iterations. The y-axis corresponds to the probability with which each player selects the first action 1. Both players use Euclidean regularization with $\eta = 0.05$ and H-step recency bias (Item 1) for different values of $H \in \{1, 10, 25, 50\}$. The dynamics under small values of the prediction window H are qualitatively similar. On the other hand, as suggested by Proposition A.2, larger values of H can lead to instability, confirming that the learning rate has to be adapted to H.

Figure 3. The (OMD) dynamics in the zero-sum games described in (74) for 2000 iterations. The y-axis corresponds to the probability with which each player selects the first action 1. Both players use Euclidean regularization with $\eta = 0.05$. Unlike Figure 2, player $\mathcal{X}$ uses one-step recency bias, while player $\mathcal{Y}$ continues to be using H-step recency bias for differnt values of $H \in \{1, 10, 25, 50\}$. We observe that larger values of H could lead to unstable behavior if η is not selected sufficiently small.

Strategically Zero-Sum Games Next, we experiment with strategically zero-sum games (Definition A.5). The cost matrix for player $\mathcal{X}$ in each game is still given by (74), but now we assume that the cost matrix of player $\mathcal{Y}$ are given as follows.

$$\mathbf{B}_1 \coloneqq \begin{bmatrix} -1 & 0.5 & 1 \\ 0 & .5 & .5 \\ -0.25 & 0 & 1 \end{bmatrix}; \mathbf{B}_2 \coloneqq \begin{bmatrix} 0.3 & 0 & 0.3 \\ -0.2 & .25 & 0.3 \\ -0.35 & 0.75 & 0.05 \end{bmatrix}; \mathbf{B}_3 \coloneqq \begin{bmatrix} 0.7 & 0.5 & 0.56 \\ 0.4 & 0.4 & 0.7 \\ 0.5 & 0.6 & 0.5 \end{bmatrix}. \tag{75}$$

To verify that the games $(\mathbf{A}_1, \mathbf{B}_1)$, $(\mathbf{A}_2, \mathbf{B}_2)$, and $(\mathbf{A}_3, \mathbf{B}_3)$ are strategically zero-sum, observe that

$$\mathbf{B}_1 = -0.5\mathbf{A}_1 + \mathbf{1}_3 \begin{bmatrix} -0.5 & 0 & 0.5 \end{bmatrix};$$
$$\mathbf{B}_2 = -0.5\mathbf{A}_2 + \mathbf{1}_3 \begin{bmatrix} -0.2 & 0.5 & -0.2 \end{bmatrix};$$
$$\mathbf{B}_3 = -0.2\mathbf{A}_3 + \mathbf{1}_3 \begin{bmatrix} 0.5 & 0.5 & 0.5 \end{bmatrix}.$$

where recall that $\mathbf{1}_3$ is the all-ones vector in $\mathbb{R}^3$. Thus, the fact that these games are strategically zero-sum follows from the
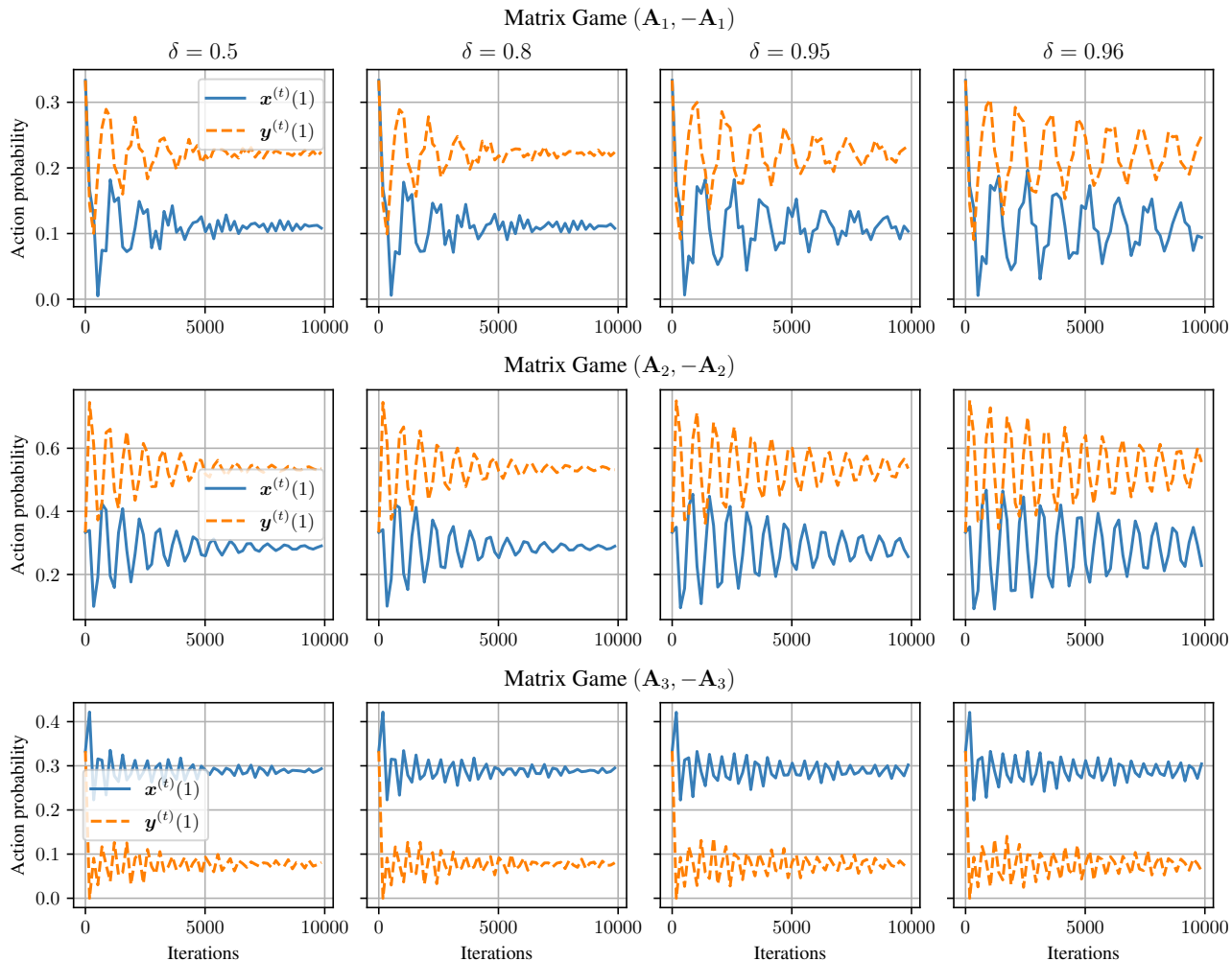
Figure 4. The (OMD) dynamics in the zero-sum games described in (74) for 10000 iterations. The y-axis corresponds to the probability with which each player selects the first action 1. Both players use Euclidean regularization with $\eta = 0.02$ and geometrically δ-discounted predictions (Item 2) for different values of $\delta \in \{0.5, 0.8, 0.95, 0.96\}$. We observe that while δ approaches to 1 the dynamics become more oscillatory.

converse of Theorem A.6. In particular, the game $(\mathbf{A}_3, \mathbf{B}_3)$ is strictly competitive (see Remark A.8). The (OMD) dynamics in these games are illustrated in Figure 6.

We also experiment with different players using different regularization. In particular, we assume that player $\mathcal{X}$ uses the Euclidean DGF, while player $\mathcal{Y}$ uses the negative entropy DGF. While our last-iterate guarantees do not appear to apply for this case due to the lack of smoothness of the entropic regularizer (close to the boundary), Figure 7 illustrates that the dynamics still converge.

D.2 Continuous Games In this subsection with provide experiments on continuous unconstrained games. In particular, we illustrate the behavior of the dynamics for the games constructed for Proposition 5.2 and Proposition 5.3 in Figure 8 and Figure 9 respectively.
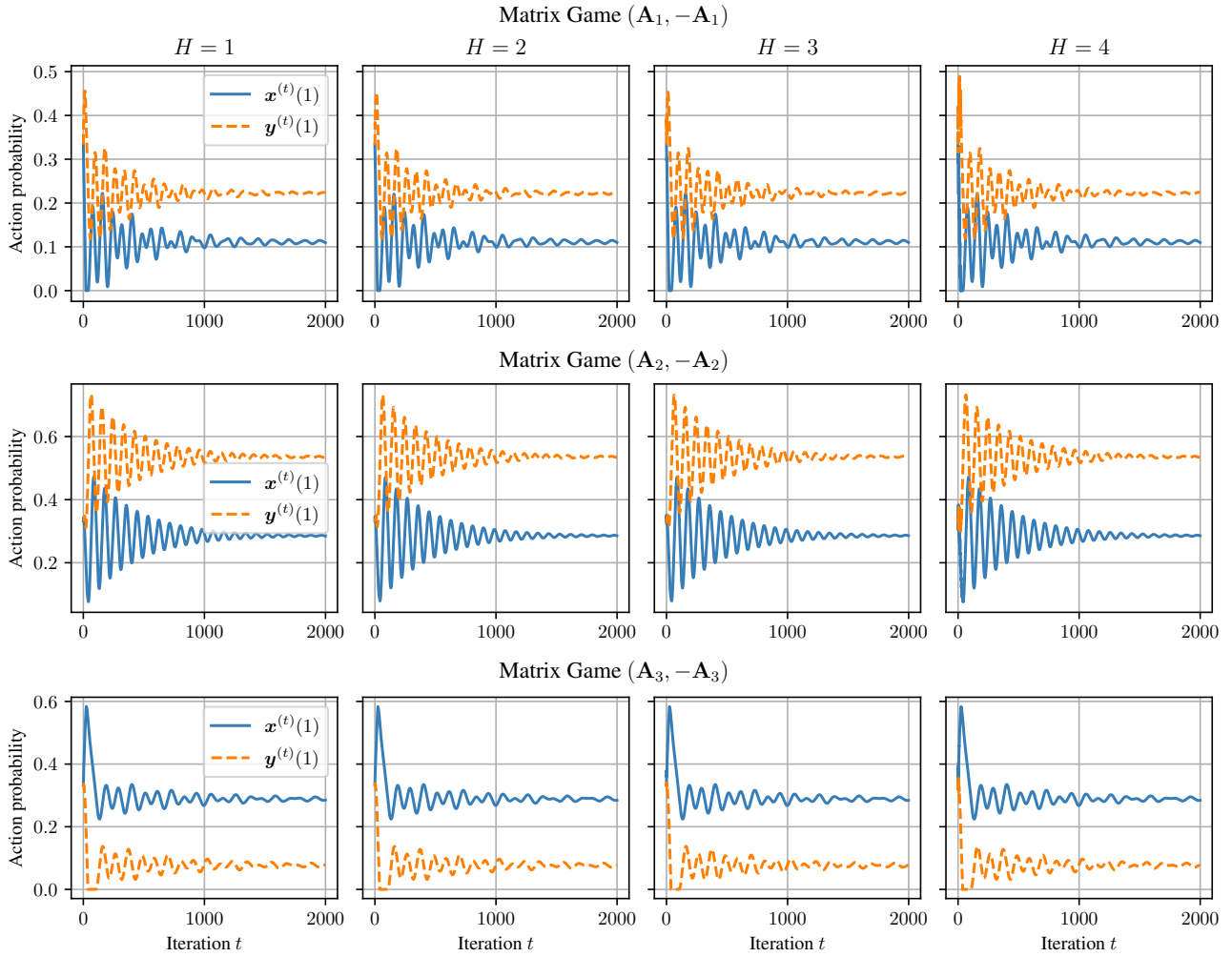
Figure 5. The (OMD) dynamics in the zero-sum games described in (74) for 2000 iterations. The y-axis corresponds to the probability with which each player selects the first action 1. Both players use Euclidean regularization with $\eta = 0.05$ and H-order predictions for different values of $H \in \{1, 2, 3, 4\}$. Interestingly, the dynamics exhibit almost identical behavior to that with one recency bias—equivalently, 1-order predictions.
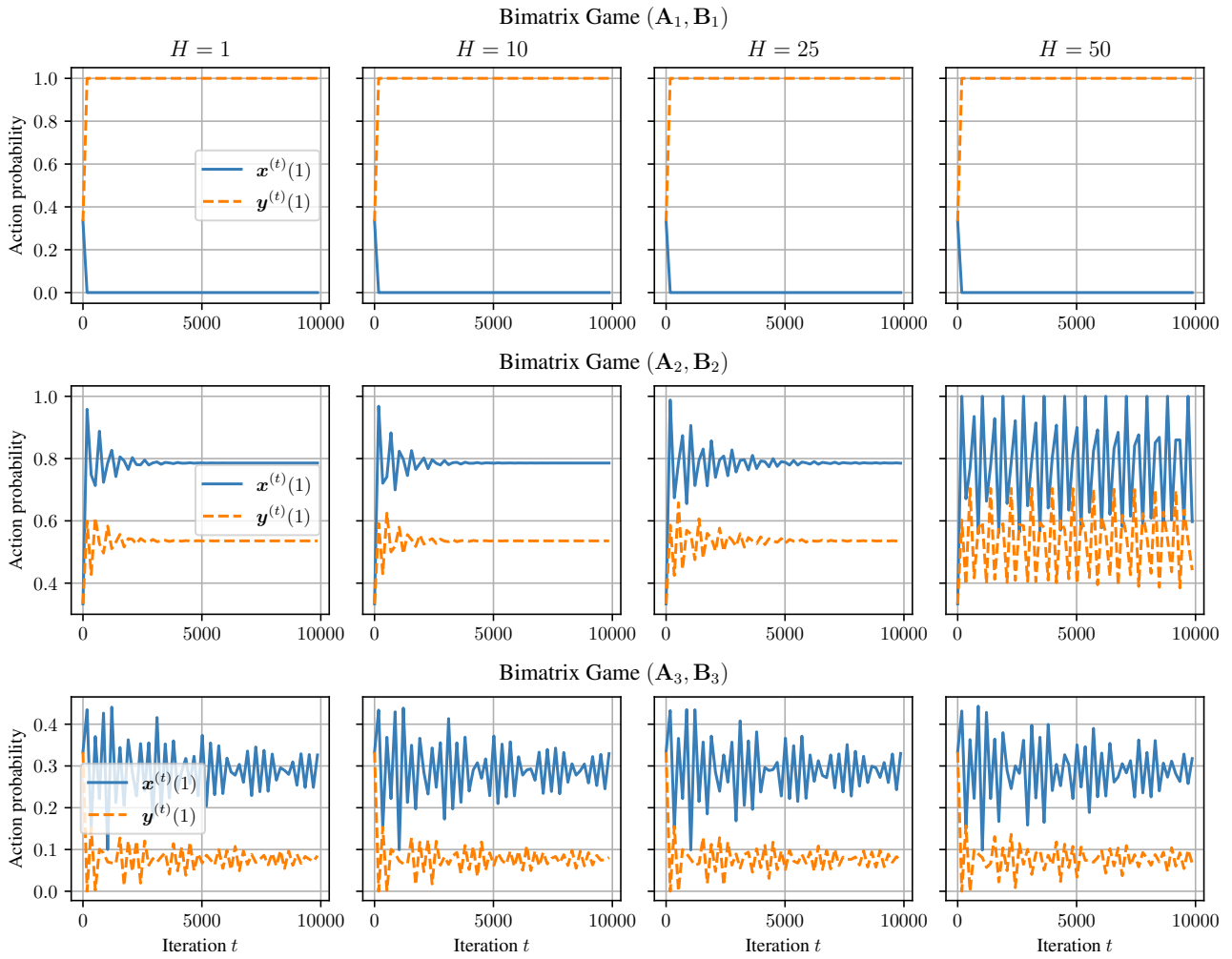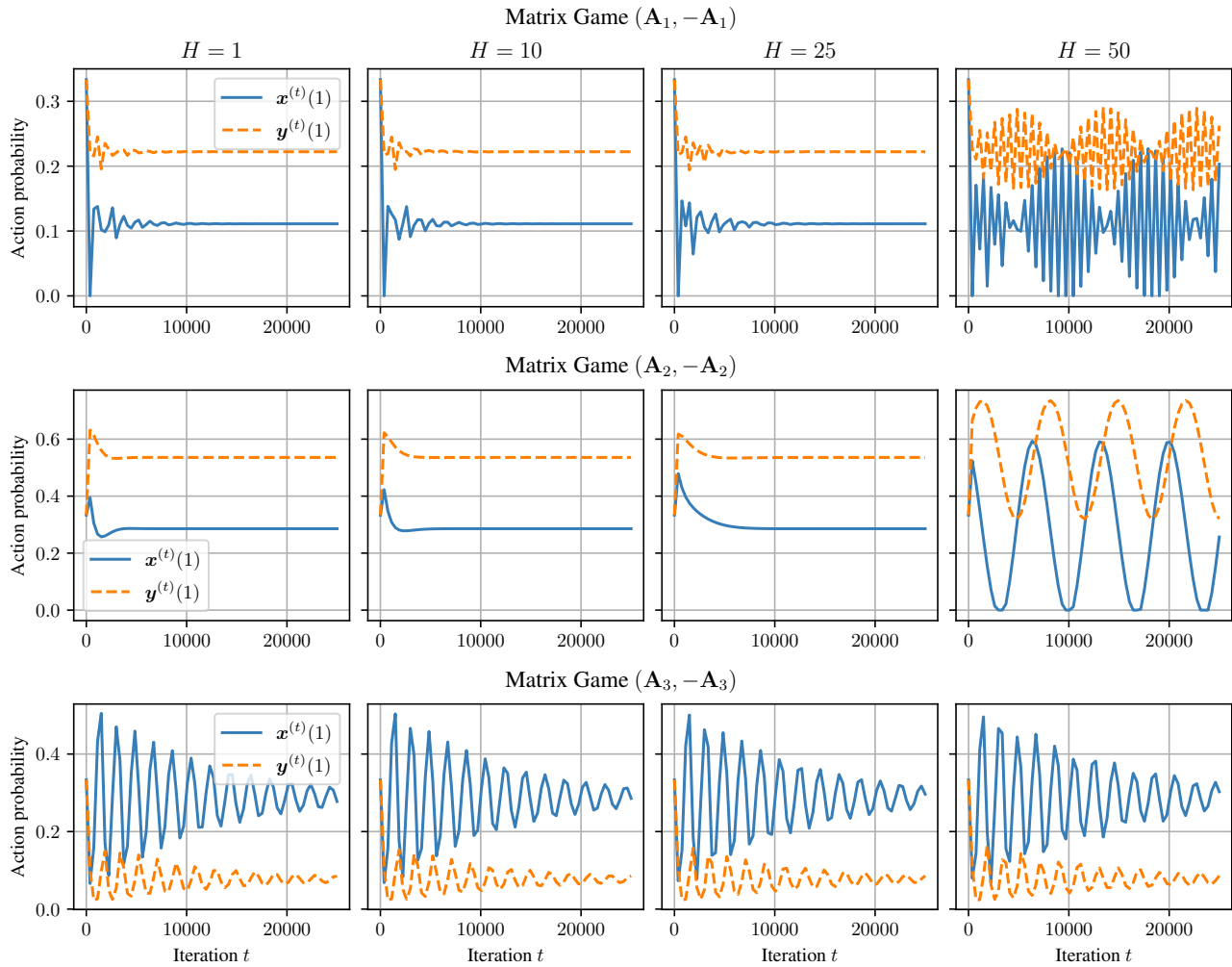
Figure 6. The (OMD) dynamics in the strategically zero-sum cost-minimization games described in (74) and (75) for 10000 iterations. The y-axis corresponds to the probability with which each player selects the first action 1. Both players use Euclidean regularization with $\eta = 0.05$ and H-step recency bias for different values of $H \in \{1, 10, 25, 50\}$. The dynamics under small values of the prediction window H are qualitatively similar. In contrast, observe that larger values of H can lead to instability.

Figure 7. The (OMD) dynamics in the zero-sum games described in (74). The y-axis corresponds to the probability with which each player selects the first action 1. Player $\mathcal{X}$ uses the Euclidean DGF, while player $\mathcal{Y}$ uses the negative entropy DGF, both with $\eta = 0.05$ and H-order predictions. We observe that the dynamics exhibit convergent behavior.
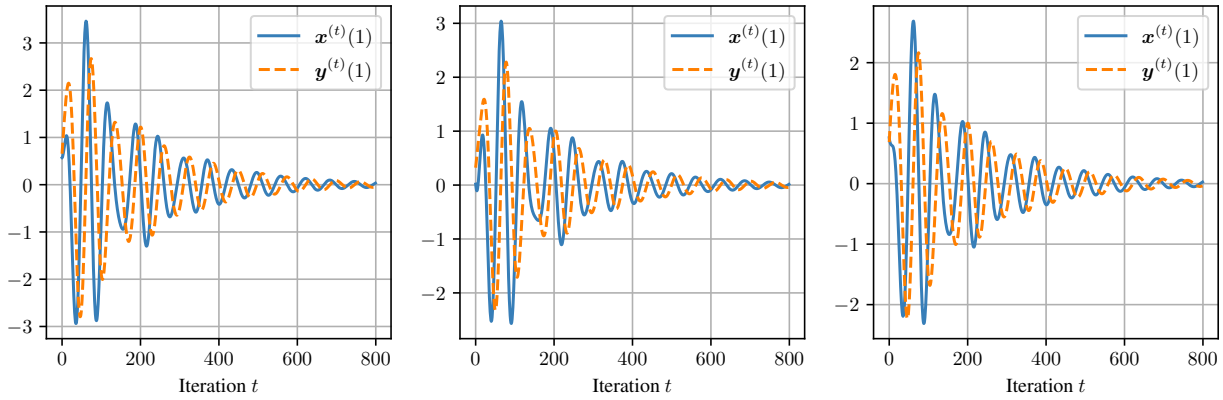
Figure 8. The (OGD) dynamics for the continuous game described in (70) for 800 iterations. The y-axis illustrates the first coordinate of $\boldsymbol{x} \in \mathbb{R}^2$ and $\boldsymbol{y} \in \mathbb{R}^2$ respectively. Different columns correspond to different random initializations. As predicted by Theorem 5.1, and subsequently Proposition 5.2, the dynamics converge to the point $(\boldsymbol{0}, \boldsymbol{0})$, which leads to each player receiving 0 utility.
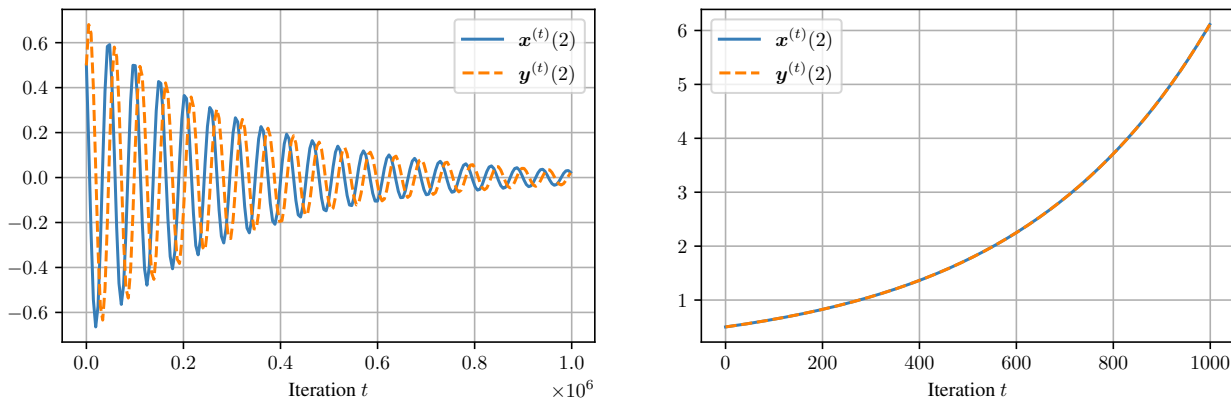

Figure 9. The (OGD) dynamics for the continuous game described in (71) for $\epsilon = 0.05$. In particular, the left image corresponds to the game $(\mathbf{A}, -\mathbf{A})$ for 10^6 iterations, while the right one to the game $(\mathbf{A}, \mathbf{B})$ for 1000 iterations. Although $\|\mathbf{A} + \mathbf{B}\|_F = \epsilon$, the two systems exhibit completely different behavior. This appears to be related to the fact that although the dynamics in the game $(\mathbf{A}, -\mathbf{A})$ converge linearly, the rate of convergence is very close to 1. The same phenomenon occurs for any $\epsilon > 0$, as predicted by Proposition 5.3.