META-LEARNING IN GAMES

Keegan Harris[*]
Carnegie Mellon University
keeganh@cs.cmu.edu

Ioannis Anagnostides[*]
Carnegie Mellon University
ianagnos@cs.cmu.edu

Gabriele Farina
FAIR, Meta AI
gfarina@meta.com

Mikhail Khodak
Carnegie Mellon University
mkhodak@cs.cmu.edu

Zhiwei Steven Wu
Carnegie Mellon University
zhiweiw@cs.cmu.edu

Tuomas Sandholm
Carnegie Mellon University
Strategy Robot, Inc.
Optimized Markets, Inc.
Strategic Machine, Inc.
sandholm@cs.cmu.edu

ABSTRACT

In the literature on game-theoretic equilibrium finding, focus has mainly been on solving a single game in isolation. In practice, however, strategic interactions—ranging from routing problems to online advertising auctions—evolve dynamically, thereby leading to many similar games to be solved. To address this gap, we introduce *meta-learning* for equilibrium finding and learning to play games. We establish the first meta-learning guarantees for a variety of fundamental and well-studied classes of games, including two-player zero-sum games, general-sum games, and Stackelberg games. In particular, we obtain rates of convergence to different game-theoretic equilibria that depend on natural notions of similarity between the sequence of games encountered, while at the same time recovering the known single-game guarantees when the sequence of games is arbitrary. Along the way, we prove a number of new results in the single-game regime through a simple and unified framework, which may be of independent interest. Finally, we evaluate our meta-learning algorithms on endgames faced by the poker agent *Libratus* against top human professionals. The experiments show that games with varying stack sizes can be solved significantly faster using our meta-learning techniques than by solving them separately, often by an order of magnitude.

1 INTRODUCTION

Research on game-theoretic equilibrium computation has primarily focused on solving a single game in isolation. In practice, however, there are often many similar games which need to be solved. One use-case is the setting where one wants to find an equilibrium for each of multiple game variations—for example poker games where the players have various sizes of chip stacks. Another use-case is strategic interactions that evolve dynamically: in online advertising auctions, the advertiser's value for different keywords adapts based on current marketing trends (Nekipelov et al., 2015); routing games—be it Internet routing or physical transportation—reshape depending on the topology and the cost functions of the underlying network (Hoefer et al., 2011); and resource allocation problems (Johari and Tsitsiklis, 2004) vary based on the values of the goods/services. Successful agents in such complex decentralized environments must effectively learn how to incorporate past experience from previous strategic interactions in order to adapt their behavior to the current and future tasks.

Meta-learning, or *learning-to-learn* (Thrun and Pratt, 1998), is a common formalization for machine learning in dynamic single-agent environments. In the meta-learning framework, a learning agent faces a sequence of tasks, and the goal is to use knowledge gained from previous tasks in order to improve performance on the current task at hand. Despite rapid progress in this line of work, prior results have not been tailored to tackle multiagent settings. This begs the question: *Can players obtain provable performance improvements when meta-learning across a sequence of games?* We answer this

[*] Equal contribution.

question in the affirmative by introducing meta-learning for equilibrium finding and learning to play games, and providing the first performance guarantees in a number of fundamental multiagent settings.

1.1 Overview of Our Results

Our main contribution is to develop a general framework for establishing the first provable guarantees for meta-learning in games, leading to a comprehensive set of results in a variety of well-studied multiagent settings. In particular, our results encompass environments ranging from two-player zero-sum games with general constraint sets (and multiple extensions thereof), to general-sum games and Stackelberg games. See Table 1 for a summary of our results. Our refined guarantees are parameterized based on natural similarity metrics between the sequence of games. For example, in zero-sum games we obtain last-iterate rates that depend on the variance of the Nash equilibria (Theorem 3.2); in potential games based on the deviation of the potential functions (Theorem 3.4); and in Stackelberg games our regret bounds depend on the similarity of the leader's optimal commitment in hindsight (Theorem 3.8). All of these measures are algorithm-independent, and tie naturally to the underlying game-theoretic solution concepts.

Importantly, our algorithms are agnostic to how similar the games are, but are nonetheless specifically designed to adapt to the similarity. Our guarantees apply under a broad class of no-regret learning algorithms, such as *optimistic mirror descent (`OMD`)* (Chiang et al., 2012; Rakhlin and Sridharan, 2013b), with the important twist that each player employs an additional regret minimizer for meta-learning the parameterization of the base-learner; the latter component builds on the meta-learning framework of Khodak et al. (2019). For example, in zero-sum games we leverage an initialization-dependent *RVU bound* (Syrgkanis et al., 2015) in order to meta-learn the initialization of `OMD` across the sequences of games, leading to per-game convergence rates to Nash equilibria that closely match our refined lower bound (Theorem 3.3). More broadly, in the worst-case—*i.e.*, when the sequence of games is arbitrary—we recover the near-optimal guarantees known for static games, but as the similarity metrics become more favorable we establish significant gains in terms of convergence to different notions of equilibria.

Along the way, we also obtain new insights and results even from a single-game perspective, including convergence rates of `OMD` and the *extra-gradient method* in Hölder continuous variational inequalities (Rakhlin and Sridharan, 2013a), and certain nonconvex-nonconcave problems such as those considered by (Diakonikolas et al., 2021) and stochastic games. Further, our analysis is considerably simpler than prior techniques and unifies several prior results. Finally, in Section 4 we evaluate our techniques on a series of poker endgames faced by the poker agent *Libratus* (Brown and Sandholm, 2018) against top human professionals. The experiments show that our meta-learning algorithms offer significant gains compared to solving each game in isolation, often by an order of magnitude.

Table 1: A summary of our key theoretical results on meta-learning in games.

Class of games	**Specific problem**	**Key results**	**Location**
Zero-sum games	Bilinear saddle-point problems	Theorems 3.1 and 3.2	Section 3.1
	Hölder continuous VIs	Theorems C.17 and C.34	Appendices C.2 and C.6
	Lower bound	Theorem 3.3	Section 3.1
General-sum games	Potential games	Theorem 3.4	Section 3.2
	(Coarse) Correlated equilibria	Theorems D.7 and D.10	Appendices D.2 and D.3
	Approximately optimal welfare	Theorem 3.6	Section 3.2
Stackelberg games	Security games	Theorem 3.8	Section 3.3

1.2 Related Work

While most prior work on learning in games posits that the underlying game remains invariant, there is ample motivation for studying games that gradually change over time, such as online advertising (Nekipelov et al., 2015; Lykouris et al., 2016; Nisan and Noti, 2017) or congestion games (Hoefer et al., 2011; Bertrand et al., 2020; Meigs et al., 2017). Indeed, a number of prior works study the performance of learning algorithms in time-varying zero-sum games (Zhang et al., 2022b; Fiez et al., 2021b; Duvocelle et al., 2022; Cardoso et al., 2019); there, it is natural to espouse dynamic notions of regret (Yang et al., 2016; Zhao et al., 2020). A work closely related to ours is the recent paper by Zhang et al. (2022b), which provides regret bounds in time-varying bilinear saddle-point problems parameterized by the similarity of the payoff matrices and the equilibria of those games. In contrast to our meta-learning setup, they study a more general setting in which the game can change arbitrarily from round-to-round. While our problem can be viewed a special type of a time-varying game in which the boundaries between different games are fixed and known, algorithms designed for generic time-varying games will not perform as well in our setting, as they do not utilize this extra information. As a result, we view these results as complementary to ours. For a more detailed discussion, see Appendix A.

An emerging paradigm for modeling such considerations is meta-learning, which has gained increasing popularity in the machine learning community in recent years; for a highly incomplete set of pointers, we refer to (Balcan et al., 2015b; Al-Shedivat et al., 2018; Finn et al., 2017; 2019; Balcan et al., 2019; Li et al., 2017; Chen et al., 2022), and references therein. Our work constitutes the natural coalescence of meta-learning with the line of work on (decentralized) online learning in games. Although, as we pointed out earlier, learning in dynamic games has already received considerable attention, we are the first (to our knowledge) to formulate and address such questions within the meta-learning framework; *c.f.*, see (Kayaalp et al., 2020; 2021; Li et al., 2022). Finally, our methods may be viewed within the *algorithms with predictions* paradigm (Mitzenmacher and Vassilvitskii, 2020): we speed up equilibrium computation by learning to predict equilibria across multiple games, with the task-similarity being the measure of prediction quality. For further related work, see Appendix A.

2 Our Setup: Meta-Learning in Games

Notation We use boldface symbols to represent vectors and matrices. Subscripts are typically reserved to indicate the player, while superscripts usually correspond to the iteration or the index of the task. We let $\mathbb{N} := \{1, 2, \dots, \}$ be the set of natural numbers. For $T \in \mathbb{N}$, we use the shorthand notation $[\![T]\!] := \{1, 2, \dots, T\}$. For a nonempty convex and compact set $\mathcal{X}$, we denote by $\Omega_{\mathcal{X}}$ its ℓ_2-diameter: $\Omega_{\mathcal{X}} := \max_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{x} - \boldsymbol{x}'\|_2$. Finally, to lighten the exposition we use the $O(\cdot)$ notation to suppress factors that depend polynomially on the natural parameters of the problem.

The general setup We consider a setting wherein players interact in a sequence of T repeated games (or *tasks*), for some $\mathbb{N} \ni T \gg 1$. Each task itself consists of $m \in \mathbb{N}$ iterations. Any fixed task t corresponds to a multiplayer game $\mathcal{G}^{(t)}$ between a set $[\![n]\!]$ of players, with $n \geq 2$; it is assumed for simplicity in the exposition that n remains invariant across the games, but some of our results apply more broadly. Each player $k \in [\![n]\!]$ selects a strategy $\boldsymbol{x}_k$ from a convex and compact set of strategies $\mathcal{X}_k \subseteq \mathbb{R}^{d_k}$ with nonempty relative interior. For a given joint strategy profile $\boldsymbol{x} := (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \in \bigtimes_{k=1}^{n} \mathcal{X}_k$, there is a multilinear utility function $u_k : \boldsymbol{x} \mapsto \langle \boldsymbol{x}_k, \boldsymbol{u}_k(\boldsymbol{x}_{-k}) \rangle$ for each player k, where $\boldsymbol{x}_{-k} := (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}, \boldsymbol{x}_{k+1}, \dots, \boldsymbol{x}_n)$. We will also let $L > 0$ be a Lipschitz parameter of each game, in the sense that for any player $k \in [\![n]\!]$ and any two strategy profiles $\boldsymbol{x}_{-k}, \boldsymbol{x}'_{-k} \in \bigtimes_{k' \neq k} \mathcal{X}_{k'}$,

$$\|\boldsymbol{u}_k(\boldsymbol{x}_{-k}) - \boldsymbol{u}_k(\boldsymbol{x}'_{-k})\|_2 \leq L \|\boldsymbol{x}_{-k} - \boldsymbol{x}'_{-k}\|_2. \tag{1}$$

Here, we use the ℓ_2-norm for convenience in the analysis; (1) can be translated to any equivalent norm. Finally, for a joint strategy profile $\boldsymbol{x} \in \bigtimes_{k=1}^{n} \mathcal{X}_k$, the *social welfare* is defined as $\mathrm{SW}(\boldsymbol{x}) := \sum_{k=1}^{n} u_k(\boldsymbol{x})$, so that $\mathrm{OPT} := \max_{\boldsymbol{x} \in \bigtimes_{k=1}^{n} \mathcal{X}_k} \mathrm{SW}(\boldsymbol{x})$ denotes the optimal social welfare.

A concrete example encompassed by our setup is that of *extensive-form games*. More broadly, it captures general games with concave utilities (Rosen, 1965; Hsieh et al., 2021).

Online learning in games Learning proceeds in an online fashion as follows. At every iteration $i \in [\![m]\!]$ of some underlying game t, each player $k \in [\![n]\!]$ has to select a strategy $\boldsymbol{x}_k^{(t,i)} \in \mathcal{X}_k$. Then, in the full information setting, the player observes as feedback the utility corresponding to the other players' strategies at iteration i; namely, $\boldsymbol{u}_k^{(t,i)} := \boldsymbol{u}_k(\boldsymbol{x}_{-k}^{(t,i)}) \in \mathbb{R}^{d_k}$. For convenience, we will assume that $\|\boldsymbol{u}_k(\boldsymbol{x}_{-k}^{(t,i)})\|_\infty \leq 1$. The canonical measure of performance in online learning is that of *external regret*, comparing the performance of the learner with that of the optimal fixed strategy in hindsight:

Definition 2.1 (Regret). *Fix a player $k \in [\![n]\!]$ and some game $t \in [\![T]\!]$. The (external) regret of player k is defined as*

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

We will let $\mathring{\boldsymbol{x}}_k^{(t)}$ be an optimum-in-hindsight strategy for player k in game t; ties are broken arbitrarily, but according to a fixed rule (*e.g.*, lexicographically). In the meta-learning setting, our goal will be to optimize the average performance—typically measured in terms of convergence to different game-theoretic equilibria—across the sequence of games.

Optimistic mirror descent Suppose that $\mathcal{R}_k : \mathcal{X}_k \to \mathbb{R}$ is a 1-strongly convex regularizer with respect to a norm $\| \cdot \|$. We let $\mathcal{B}_{\mathcal{R}_k}(\boldsymbol{x}_k \parallel \boldsymbol{x}_k') := \mathcal{R}_k(\boldsymbol{x}_k) - \mathcal{R}_k(\boldsymbol{x}_k') - \langle \nabla \mathcal{R}_k(\boldsymbol{x}_k'), \boldsymbol{x}_k - \boldsymbol{x}_k' \rangle$ denote the *Bregman divergence* induced by $\mathcal{R}_k$, where $\boldsymbol{x}_k'$ is in the relative interior of $\mathcal{X}_k$. *Optimistic mirror descent (*`OMD`*)* (Chiang et al., 2012; Rakhlin and Sridharan, 2013b) is parameterized by a prediction $\boldsymbol{m}_k^{(t,i)} \in \mathbb{R}^{d_k}$ and a learning rate $\eta > 0$, and is defined at every iteration $i \in \mathbb{N}$ as follows.

$$\boldsymbol{x}_k^{(t,i)} := \arg\max_{\boldsymbol{x}_k \in \mathcal{X}_k} \left\{ \langle \boldsymbol{x}_k, \boldsymbol{m}_k^{(t,i)} \rangle - \frac{1}{\eta} \mathcal{B}_{\mathcal{R}_k}(\boldsymbol{x}_k \parallel \hat{\boldsymbol{x}}_k^{(t,i-1)}) \right\},$$
$$\hat{\boldsymbol{x}}_k^{(t,i)} := \arg\max_{\hat{\boldsymbol{x}}_k \in \mathcal{X}_k} \left\{ \langle \hat{\boldsymbol{x}}_k, \boldsymbol{u}_k^{(t,i)} \rangle - \frac{1}{\eta} \mathcal{B}_{\mathcal{R}_k}(\hat{\boldsymbol{x}}_k \parallel \hat{\boldsymbol{x}}_k^{(t,i-1)}) \rangle \right\}.$$

Further, $\hat{\boldsymbol{x}}_k^{(1,0)} := \arg\min_{\hat{\boldsymbol{x}}_k \in \mathcal{X}_k} \mathcal{R}_k(\hat{\boldsymbol{x}}_k) =: \boldsymbol{x}_k^{(1,0)}$, and $\boldsymbol{m}_k^{(t,1)} := \boldsymbol{u}_k(\boldsymbol{x}_{-k}^{(t,0)})$. Under Euclidean regularization, $\mathcal{R}_k(\boldsymbol{x}_k) := \frac{1}{2}\|\boldsymbol{x}_k\|_2^2$, we will refer to `OMD` as *optimistic gradient descent (*`OGD`*)*.

3 Meta-Learning How to Play Games

In this section, we present our main theoretical results: provable guarantees for online and decentralized meta-learning in games. We commence with zero-sum games in Section 3.1, and we then transition to general-sum games (Section 3.2) and Stackelberg (security) games (Section 3.3).

3.1 Zero-Sum Games

We first highlight our results for bilinear saddle-point problems (BSPPs), which take the form $\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} \boldsymbol{x}^\top \mathbf{A} \boldsymbol{y}$, where $\mathbf{A} \in \mathbb{R}^{d_x \times d_y}$ is the payoff matrix of the game. A canonical application for this setting is on the solution of zero-sum imperfect-information extensive-form games (Romanovskii, 1962; Koller and Megiddo, 1992), as we explore in our experiments (Section 4). Next we describe a number of extensions to gradually more general settings, and we conclude with our lower bound (Theorem 3.3). The proofs from this subsection are included in Appendix C.

We first derive a refined meta-learning convergence guarantee for the average of the players' strategies. Below, we denote by $V_x^2 := \frac{1}{T} \min_{\boldsymbol{x} \in \mathcal{X}} \sum_{t=1}^T \|\mathring{\boldsymbol{x}}^{(t)} - \boldsymbol{x}\|_2^2$ the task similarity metric for player x, written in terms of the optimum-in-hindsight strategies; analogous notation is used for player y.

Theorem 3.1 (Informal; Detailed Version in Corollary C.2)**.** *Suppose that both players employ* `OGD` *with a suitable (fixed) learning rate and follow the leader over previous optimum-in-hindsight strategies for the initialization. Then, the game-average duality gap of the players' average strategies is bounded by*

$$\frac{1}{T} \sum_{t=1}^T \frac{1}{m} \left(\mathrm{Reg}_x^{(t,m)} + \mathrm{Reg}_y^{(t,m)} \right) \le \frac{2L}{m} \left(V_x^2 + V_y^2 \right) + \frac{8L(1 + \log T)}{mT} \left(\Omega_{\mathcal{X}}^2 + \Omega_{\mathcal{Y}}^2 \right). \quad (2)$$

Here, the second term in the right-hand side of (2) becomes negligible for a large number of games T, while the first term depends on the task similarity measures. For any sequence of games, Theorem 3.1 nearly matches the lower bound in the single-task setting (Daskalakis et al., 2015), but our guarantee can be significantly better when $V_x^2, V_y^2 \ll 1$. To achieve this, the basic idea is to use—on top of `OGD`—a "meta" regret minimization algorithm that, for each player, learns a sequence of initializations by taking the average of the past optima-in-hindsight, which is equivalent to *follow the leader (*`FTL`*)* over the regret upper-bounds of the within-task algorithm; see Algorithm 1 (in Appendix B) for pseudocode of the meta-version of `OGD` we consider. Similar results can be obtained more broadly for `OMD` (*c.f.*, see Appendices D.2 and D.3). We also obtain analogous refined bounds for the *individual* regret of each player (Corollary C.4).

One caveat of Theorem 3.1 is that the underlying task similarity measure could be algorithm-dependent, as the optimum-in-hindsight for each player could depend on the other player's behavior. To address this, we show that if the meta-learner can initialize using *Nash equilibria (NE)* (recall Definition C.5) from previously seen games, the game-average last-iterate rates gracefully decrease with the similarity of the Nash equilibria of those games. More precisely, if $\boldsymbol{z} := (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y} =: \mathcal{Z}$, we let $V_{\text{NE}}^2 := \frac{1}{T} \max_{\boldsymbol{z}^{(1,\star)},\dots,\boldsymbol{z}^{(T,\star)}} \min_{\boldsymbol{z} \in \mathcal{Z}} \sum_{t=1}^{T} \|\boldsymbol{z}^{(t,\star)} - \boldsymbol{z}\|_2^2$, where $\boldsymbol{z}^{(t,\star)}$ is any Nash equilibrium of the t-th game. As we point out in the sequel, we also obtain results under a more favorable notion of task similarity that does not depend on the worst sequence of NE.

Theorem 3.2 (Informal; Detailed Version in Theorem C.8)**.** *When both players employ* `OGD` *with a suitable (fixed) learning rate and* `FTL` *over previous NE strategies for the initialization, then*

$$\bar{m} \leq \frac{2V_{NE}^2}{\epsilon^2} + \frac{8(1 + \log T)}{T\epsilon^2} \left(\Omega_{\mathcal{X}}^2 + \Omega_{\mathcal{Y}}^2\right)$$

iterations suffice to reach an $O(\epsilon)$-approximate Nash equilibrium on average across the T games.

Theorem 3.2 recovers the optimal $m^{-1/2}$ rates for `OGD` (Golowich et al., 2020a;b) under an arbitrary sequence of games, but offers substantial gains in terms of the average iteration complexity when the Nash equilibria of the games are close. For example, when they lie within a ball of ℓ_2-diameter $\sqrt{\delta(\Omega_{\mathcal{X}}^2 + \Omega_{\mathcal{Y}}^2)}$, for some $\delta \in (0, 1]$, Theorem 3.2 improves upon the rate of `OGD` by at least a multiplicative factor of $1/\delta$ as $T \to \infty$. While *generic*—roughly speaking, randomly perturbed—zero-sum (normal-form) games have a unique Nash equilibrium (van Damme, 1987), the worst-case NE similarity metric used in Theorem 3.2 can be loose under multiplicity of equilibria. For that reason, in Appendix C.1.2 we further refine Theorem 3.2 using the most favorable sequence of Nash equilibria; this requires that players know each game after its termination, which is arguably a well-motivated assumption in some applications. We further remark that Theorem 3.2 can be cast in terms of the similarity $V_x^2 + V_y^2$, instead of V_{NE}^2, using the parameterization of Theorem 3.1. Finally, since the base-learner can be viewed as an algorithm with predictions—the number of iterations to compute an approximate NE is smaller if the Euclidean error of a prediction of it (the initialization) is small—Theorem 3.2 can also be viewed as *learning* these predictions (Khodak et al., 2022) by targeting that error measure.

Extensions Moving beyond bilinear saddle-point problems, we extend our results to gradually broader settings. First, in Appendix C.2 we apply our techniques to general variational inequality problems under a Lipschitz continuous operator for which the so-called *MVI property* (Mertikopoulos et al., 2019) holds. Thus, Theorems 3.1 and 3.2 are extended to settings such as smooth convex-concave games and zero-sum polymatrix (multiplayer) games (Cai et al., 2016). Interestingly, extensions are possible even under the *weak MVI property* (Diakonikolas et al., 2021), which captures certain "structured" nonconvex-nonconcave games. In a similar vein, we also study the challenging setting of Shapley's stochastic games (Shapley, 1953) (Appendix C.5). There, we show that there exists a time-varying—instead of constant—but non-vanishing learning rate schedule for which `OGD` reaches minimax equilibria, thereby leading to similar extensions in the meta-learning setting. Next, we relax the underlying Lipschitz continuity assumption underpinning the previous results by instead imposing only α-Hölder continuity (recall Definition C.32). We show that in such settings `OGD` enjoys a rate of $m^{-\alpha/2}$ (Theorem C.34), which is to our knowledge a new result; in the special case where $\alpha = 1$, we recover the recently established $m^{-1/2}$ rates. Finally, while we have focused on the `OGD` algorithm, our techniques apply to other learning dynamics as well. For example, in Appendix C.7 we show that the extensively studied extra-gradient (`EG`) algorithm (Korpelevich, 1976) can be analyzed in a unifying way with `OGD`, thereby inheriting all of the aforementioned results under `OGD`; this significantly broadens the implications of (Mokhtari et al., 2020), which only applied in certain unconstrained problems. Perhaps surprisingly, although `EG` is *not* a no-regret algorithm, our analysis employs a regret-based framework using a suitable proxy for the regret (see Theorem C.35).

Lower bound We conclude this subsection with a lower bound, showing that our guarantee in Theorem 3.1 is essentially sharp under a broad range of our similarity measures. Our result significantly refines the single-game lower bound of Daskalakis et al. (2015) by constructing an appropriate distribution over sequences of zero-sum games.

Theorem 3.3 (Informal; Precise Version in Theorem C.39)**.** *For any $\epsilon > 0$, there exists a distribution over sequences of T zero-sum games, with a sufficiently large $T = T(\epsilon)$, such that*

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\text{Reg}_x^{(t,m)} + \text{Reg}_y^{(t,m)}] \geq \frac{1}{2} \left(V_x^2 + V_y^2\right) - \epsilon = \frac{1}{2} V_{NE}^2 - \epsilon.$$

3.2 General-Sum Games

In this subsection, we switch our attention to general-sum games. Here, unlike zero-sum games, no-regret learning algorithms are instead known to generally converge—in a time-average sense—to *correlated equilibrium* concepts, which are more permissive than the Nash equilibrium. Nevertheless, there are structured classes of general-sum games for which suitable dynamics do reach Nash equilibria; perhaps the most notable example being that of *potential games*. In this context, we first obtain meta-learning guarantees for potential games, parameterized by the similarity of the potential functions. Then, we derive meta-learning algorithms with improved guarantees for convergence to correlated and *coarse* correlated equilibria. Finally, we conclude this subsection with improved guarantees of convergence to near-optimal—in terms of social welfare—equilibria. Proofs from this subsection are included in Appendices B and D.

Potential games A potential game is endowed with the additional property of admitting a potential: a player-independent function that captures the player's benefit from unilaterally deviating from any given strategy profile (Definition D.2). In our meta-learning setting, we posit a sequence of potential games $(\Phi^{(t)})_{1 \leq t \leq t}$, each described by its potential function. Unlike our approach in Section 3.1, a central challenge here is that the potential function is in general nonconcave/nonconvex, precluding standard regret minimization approaches. Instead, we find that by initializing at the previous last-iterate the dynamics still manage to adapt based on the similarity $V_\Delta := \frac{1}{T}\sum_{t=1}^{T-1} \Delta(\Phi^{(t)}, \Phi^{(t+1)})$, where $\Delta(\Phi, \Phi') := \max_{\boldsymbol{x}}(\Phi(\boldsymbol{x}) - \Phi'(\boldsymbol{x}))$, which captures the deviation of the potential functions. This initialization has the additional benefit of being agnostic to the boundaries of different tasks. Unlike our results in Section 3.1, the following guarantee applies even for vanilla (*i.e.*, non-optimistic) projected gradient descent (`GD`).

Theorem 3.4 (Informal; Detailed Version in Corollary D.5). *`GD` with suitable parameterization requires $O\left(\frac{V_\Delta}{\epsilon^2} + \frac{\Phi_{max}}{\epsilon^2 T}\right)$ iterations to reach an ϵ-approximate Nash equilibrium on average across the T potential games, where $\max_{\boldsymbol{x},t} |\Phi^{(t)}(\boldsymbol{x})| \leq \Phi_{max}$.*

Theorem 3.4 matches the known rate of `GD` for potential games in the worst case, but offers substantial gains in terms of the average iteration complexity when the games are similar. For example, if $|\Phi^{(t)}(\boldsymbol{x}) - \Phi^{(t-1)}(\boldsymbol{x})| \leq \alpha$, for all $\boldsymbol{x} \in \bigtimes_{k=1}^{n} \mathcal{X}_k$ and $t \geq 2$, then $O(\alpha/\epsilon^2)$ iterations suffice to reach an ϵ-approximate NE on an average game, as $T \to +\infty$. Such a scenario may arise in, *e.g.*, a sequence of routing games if the cost functions for each edge change only slightly between games.

Convergence to correlated equilibria In contrast, for general games the best one can hope for is to obtain improved rates for convergence to correlated or coarse correlated equilibria (Hart and Mas-Colell, 2000; Blum and Mansour, 2007). It is important to stress that learning correlated equilibria is fundamentally different than learning Nash equilibria—which are product distributions. For example, for the former any initialization—which is inevitably a product distribution in the case of uncoupled dynamics—could fail to exploit the learning in the previous task (Proposition D.1): unlike Nash equilibria, correlated equilibria (in general) cannot be decomposed for each player, thereby making uncoupled methods unlikely to adapt to the similarity of CE. Instead, our task similarity metrics depend on the optima-in-hindsight for each player. Under this notion of task similarity, we obtain task-average guarantees for CCE by meta-learning the initialization (by running `FTL`) and the learning rate (by running the `EWOO` method of Hazan et al. (2007) over a sequence of regret upper bounds) of *optimistic hedge* (Daskalakis et al., 2021) (Theorem D.7)—`OMD` with entropic regularization. Similarly, to obtain guarantees for CE, we use the *no-swap-regret* construction of Blum and Mansour (2007) in conjuction with the logarithmic barrier (Anagnostides et al., 2022a) (Theorem D.10).

3.2.1 Social Welfare Guarantees

We conclude this subsection with meta-learning guarantees for converging to near-optimal equilibria (Theorem 3.6). Let us first recall the following central definition.

Definition 3.5 (Smooth games (Roughgarden, 2015)). *A game $\mathcal{G}$ is (λ, μ)-smooth, with $\lambda, \mu > 0$, if there exists a strategy profile $\boldsymbol{x}^\star \in \bigtimes_{k=1}^{n} \mathcal{X}_k$ such that for any $\boldsymbol{x} \in \bigtimes_{k=1}^{n} \mathcal{X}_k$,*

$$\sum_{k=1}^{n} u_k(\boldsymbol{x}_k^\star, \boldsymbol{x}_{-k}) \geq \lambda \mathrm{OPT} - \mu \mathrm{SW}(\boldsymbol{x}), \tag{3}$$

where OPT *is the optimal social welfare and* $\mathrm{SW}(\boldsymbol{x})$ *is the social welfare of joint strategy profile* $\boldsymbol{x}$*.*

Smooth games capture a number of important applications, including network congestion games (Awerbuch et al., 2013; Christodoulou and Koutsoupias, 2005) and simultaneous auctions (Christodoulou et al., 2016; Roughgarden et al., 2017) (see Appendix B for additional examples); both of those settings are oftentimes non-static in real-world applications, thereby motivating our meta-learning considerations. In this context, we assume that there is a sequence of smooth games $(\mathcal{G}^{(t)})_{1\leq t\leq T}$, each of which is $(\lambda^{(t)}, \mu^{(t)})$-smooth (Definition 3.5).

Theorem 3.6 (Informal; Detailed Version in Theorem B.11)**.** *If all players use* `OGD` *with suitable parameterization in a sequence of T games $(\mathcal{G}^{(t)})_{1\leq t\leq T}$, each of which is $(\lambda^{(t)}, \mu^{(t)})$-smooth, then*

$$\frac{1}{mT}\sum_{t=1}^{T}\sum_{i=1}^{m}\mathrm{SW}(\boldsymbol{x}^{(t,i)}) \geq \frac{1}{T}\sum_{t=1}^{T}\frac{\lambda^{(t)}}{1+\mu^{(t)}}\mathrm{OPT}^{(t)} - \frac{2L\sqrt{n-1}}{m}\sum_{k=1}^{n}V_k^2 - \widetilde{O}\left(\frac{1}{mT}\right), \tag{4}$$

where $\mathrm{OPT}^{(t)}$ *is the optimal social welfare attainable at game* $\mathcal{G}^{(t)}$ *and* $\widetilde{O}(\cdot)$ *hides logarithmic terms.*

The first term in the right-hand side of (4) is the average robust PoA in the sequence of games, while the third term vanishes as $T \to \infty$. The orchestrated learning dynamics reach approximately optimal equilibria much faster when the underlying task similarity is small; without meta-learning one would instead obtain the m^{-1} rate known from the work of Syrgkanis et al. (2015). Theorem 3.6 is established by first providing a refined guarantee for the *sum of the players regrets* (Theorem B.3), and then translating that guarantee in terms of the social welfare using the smoothness condition for each game (Proposition B.10). Our guarantees are in fact more general, and apply for any suitable linear combination of players' utilities (see Corollary B.12).

3.3 Stackelberg (Security) Games

To conclude our theoretical results, we study meta-learning in repeated Stackelberg games. Following the convention of Balcan et al. (2015a), we present our results in terms of Stackelberg security games, although our results apply to general Stackelberg games as well (see (Balcan et al., 2015a, Section 8) for details on how such results extend).

Stackelberg security games A repeated Stackelberg security game is a sequential interaction between a defender and m attackers. In each round, the defender commits to a mixed strategy over d targets to protect, which induces a *coverage probability vector* $\boldsymbol{x} \in \Delta^d$ over targets. After having observed coverage probability vector, the attacker *best responds* by attacking some target $b(\boldsymbol{x}) \in [\![d]\!]$ in order to maximize their utility in expectation. Finally, the defender's utility is some function of their coverage probability vector $\boldsymbol{x}$ and the target attacked $b(\boldsymbol{x})$.

It is a well-known fact that no-regret learning in repeated Stackelberg games is not possible without any prior knowledge about the sequence of followers (Balcan et al., 2015a, Section 7), so we study the setting in which each attacker belongs to one of k possible *attacker types*. We allow sequence of attackers to be adversarially chosen from the k types, and assume the attacker's type is revealed to the leader after each round. We adapt the methodology of Balcan et al. (2015a) to our setting by meta-learning the initialization and learning rate of the multiplicative weights update (henceforth `MWU`) run over a finite (but exponentially-large) set of *extreme points* $\mathcal{E} \subset \Delta^d$.[1] Each point $\boldsymbol{x} \in \mathcal{E}$ corresponds to a leader mixed strategy, and $\mathcal{E}$ can be constructed in such a way that it will always contain a mixed strategy which is arbitrarily close to the optima-in-hindsight for each task.[2]

Our results are given in terms of guarantees on the task-average *Stackelberg regret*, which measures the difference in utility between the defender's deployed sequence of mixed strategies and the optima-in-hindsight, given that the attacker best responds.

Definition 3.7 (Stackelberg Regret)**.** *Denote attacker $f^{(t,i)}$'s best response to mixed strategy $\boldsymbol{x}$ as $b_{f^{(t,i)}}(\boldsymbol{x})$. The Stackelberg regret of the attacker in a repeated Stackelberg security game t is*

$$\mathrm{StackReg}^{(t,m)}(\mathring{\boldsymbol{x}}^{(t)}) = \sum_{i=1}^{m}\langle \mathring{\boldsymbol{x}}^{(t)}, \boldsymbol{u}^{(t)}(b_{f^{(t,i)}}(\mathring{\boldsymbol{x}}^{(t)}))\rangle - \langle \boldsymbol{x}^{(t,i)}, \boldsymbol{u}^{(t)}(b_{f^{(t,i)}}(\boldsymbol{x}^{(t,i)}))\rangle.$$

[1] This is likely unavoidable, as Li et al. (2016) show computing a Stackelberg strategy is strongly NP-Hard.

[2] For a precise definition of how to construct $\mathcal{E}$, we point the reader to (Balcan et al., 2015a, Section 4).

In contrast to the standard notion of regret (Definition 2.1), Stackelberg regret takes into account the extra structure in the defender's utility in Stackelberg games; namely that it is a function of the defender's current mixed strategy (through the attacker's best response).

Theorem 3.8 (Informal; Detailed Version in Theorem E.1)**.** *Given a sequence of T repeated Stackelberg security games with d targets, k attacker types, and within-game time-horizon m, running* `MWU` *over the set of extreme points $\mathcal{E}$ as defined in Balcan et al. (2015a) with suitable initialization and sequence of learning rates achieves task-averaged expected Stackelberg regret*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\text{StackReg}^{(t,m)}] = O(\sqrt{H(\bar{\boldsymbol{y}})m}) + o_T(\mathsf{poly}(m, |\mathcal{E}|)),$$

where the sequence of attackers in each task can be adversarially chosen, the expectation is with respect to the randomness of `MWU`*, $\bar{\boldsymbol{y}} := \frac{1}{T}\sum_{t=1}^{T}\mathring{\boldsymbol{y}}^{(t)}$, where $\mathring{\boldsymbol{y}}^{(t)}$ is the optimum-in-hindsight distribution over mixed strategies in $\mathcal{E}$ for game t, $H(\bar{\boldsymbol{y}})$ is the Shannon entropy of $\bar{\boldsymbol{y}}$, and $o_T(1)$ suppresses terms which decay with T.*

$H(\bar{\boldsymbol{y}}) \leq \log|\mathcal{E}|$, so in the worst-case our algorithm asymptotically matches the $O(\sqrt{m\log|\mathcal{E}|})$ performance of the algorithm of Balcan et al. (2015a). Entropy $H(\bar{\boldsymbol{y}})$ is small whenever the same small set of mixed strategies are optimal for the sequence of T Stackelberg games. For example, if in each task the adversary chooses from $s \ll k$ attacker types who are only interested in attacking $u \ll d$ targets (unbeknownst to the meta-learner), $H(\bar{\boldsymbol{y}}) = O(s^2 u \log(su))$. In Stackelberg security games $|\mathcal{E}| = O((2^d + kd^2)^d d^k)$, so $\log|\mathcal{E}| = O(d^2 k \log(dk))$. Finally, the distance between the set of optimal strategies does not matter, as $\bar{\boldsymbol{y}}$ is a categorical distribution over a discrete set of mixed strategies.

4 Experiments

In this section, we evaluate our meta-learning techniques in two River endgames that occurred in the *Brains vs AI* competition (Brown and Sandholm, 2018). We use the two public endgames that were released by the authors,[3] denoted 'Endgame A' and 'Endgame B,' each corresponding to a zero-sum extensive-form game. For each of these endgames, we produced $T := 200$ individual tasks by varying the size of the stacks of each player according to three different *task sequencing setups*:[4]

1. (*random* stacks) In each task we select stack sizes for the players by sampling uniformly at random a multiple of 100 in the range $[1000, 20000]$.
2. (*sorted* stacks) Task $t \in \{1, \dots, 200\}$ corresponds to solving the endgame where the stack sizes are set to the amount $t \times 100$ for each player.
3. (*alternating* stacks) We sequence the stack amounts of the players as follows: in task 1, the stacks are set to 100; in task 2 to $200,000$; in task 3 to 200; in task 4 to $199,900$; and so on.

For each endgame, we tested the performance when both players (1) employ `OGD` while meta-learning the initialization (Theorem 3.1) with $\boldsymbol{m}_x^{(t,1)} = \boldsymbol{0}_{d_x}$ and $\boldsymbol{m}_y^{(t,1)} = \boldsymbol{0}_{d_y}$, (2) employ `OGD` while setting the initialization equal to the last iterate of the previous task (see Remark B.8), and (3) use the vanilla initialization of `OGD`—*i.e.*, the players treat each game separately. For each game, players run $m := 1000$ iterations. The ℓ_2 projection to the *sequence-form polytope* (Romanovskii, 1962; Koller and Megiddo, 1992)—the strategy set of each player in extensive-form games—required for the steps of `OGD` is implemented via an algorithm originally described by Gilpin et al. (2012), and further clarified in (Farina et al., 2022, Appendix B). We tried different learning rates for the players selected from the set $\{0.1, 0.01, 0.001\}$. Figure 1 illustrates our results for $\eta := 0.01$, while the others are deferred to Appendix F. In the table at the top of Figure 1 we highlight several parameters of the endgames including the board configuration, the dimensions of the players' strategy sets—*i.e.*, the sequences—and the number of nonzero elements in each payoff matrix. Because of the scale of the games, we used the *Kronecker sparsification* algorithm of Farina and Sandholm (2022, Technique A) in order to accelerate the training.

[3] Obtained from `https://github.com/Sandholm-Lab/LibratusEndgames`.

[4] While in the general meta-learning setup it is assumed that the number of tasks is large but per-task data is limited (*i.e.*, $T \gg m$), we found that setting $T := 200$ was already sufficient to see substantial benefits.

			Sequences		**Decision Points**		**Payoff Matrix**
Game	**Board**	**Pot**	Pl. 1	Pl. 2	Pl. 1	Pl. 2	num. nonzeros
Endgame A	J♠ K♠ 5♣ Q♠ 7♦	3,700	18,789	19,237	6,710	6,870	14,718,298
Endgame B	4♠ 8♥ 10♣ 9♥ 2♠	500	46,875	47,381	16,304	16,480	62,748,525

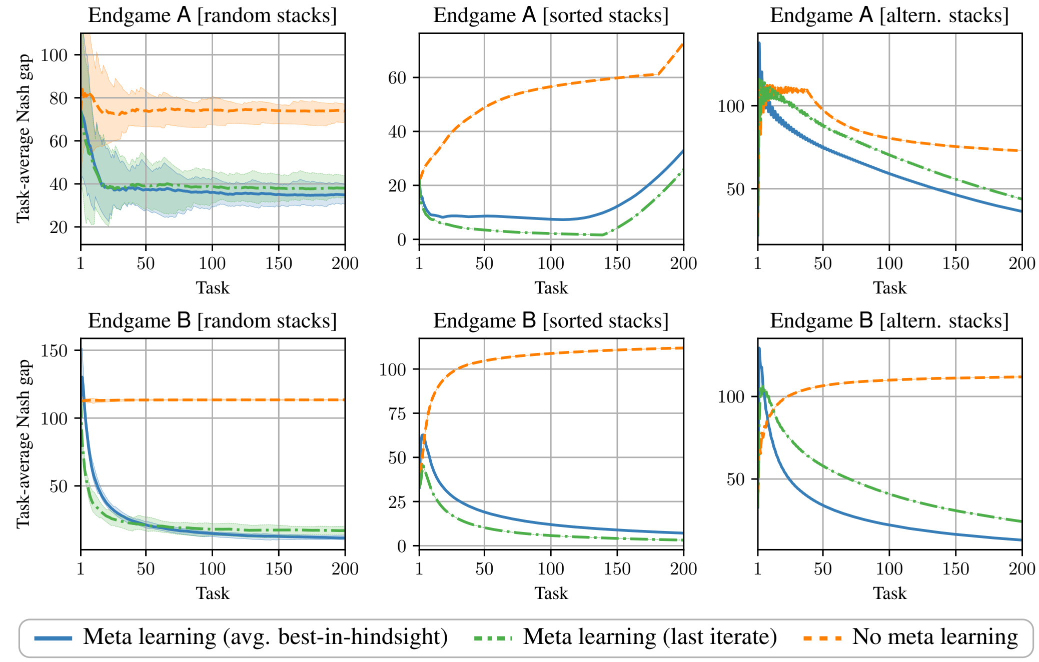

Figure 1: (Top) Parameters of each endgame. (Bottom) The task-averaged NE gap of the players' average strategies across 200 tasks, 2 endgames, and 3 different stack orderings. Both players use OGD with $\eta := 0.01$. For the random stacks, we repeated each experiment 10 times with different random seeds. The plots show the mean (thick line) as well as the minimum and maximum values. We see that across all task sequencing setups, meta-learning the initialization (using either technique) leads to up to an order of magnitude better performance compared to vanilla OGD. When stacks are sorted, initializing to the last iterate of the previous game obtains the best performance, whereas when stacks are alternated or random, initializing according to Theorem 3.1 performs best.

5 Conclusions and Future Research

In this paper, we introduced the study of meta-learning in games. In particular, we considered many of the most central game classes—including zero-sum games, potential games, general-sum multi-player games, and Stackelberg security games—and obtained provable performance guarantees expressed in terms of natural measures of similarity between the games. Experiments on several sequences of poker endgames that were actually played in the *Brains vs AI* competition (Brown and Sandholm, 2018) show that meta-learning the initialization improves performance even by an order of magnitude.

Our results open the door to several exciting directions for future research, including meta-learning in other settings for which single-game results are known, such as general nonconvex-nonconcave min-max problems (Suggala and Netrapalli, 2020), the nonparametric regime (Daskalakis and Golowich, 2022), and partial feedback (such as bandit) models (Wei and Luo, 2018; Hsieh et al., 2022; Balcan et al., 2022; Osadchiy et al., 2022). Another interesting, yet challenging, avenue for future research would be to consider strategy sets that can vary across tasks.

ACKNOWLEDGEMENTS

We are grateful to the anonymous ICLR reviewers for valuable feedback. KH is supported by a NDSEG Fellowship. IA and TS are supported by NSF grants IIS-1901403 and CCF-1733556, and the ARO under award W911NF2010081. MK is supported by a Meta Research PhD Fellowship. ZSW is supported in part by the NSF grant FAI-1939606, a Google Faculty Research Award, a J.P. Morgan Faculty Award, a Meta Research Award, and a Mozilla Research Grant. The authors would like to thank Nina Balcan for helpful discussions throughout the course of the project. IA is grateful to Ioannis Panageas for insightful discussions regarding Appendix C.5.

REFERENCES

Jacob D. Abernethy, Kevin A. Lai, Kfir Y. Levy, and Jun-Kun Wang. Faster rates for convex-concave games. In *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 1595–1625. PMLR, 2018.

Alekh Agarwal, Sham M. Kakade, Jason D. Lee, and Gaurav Mahajan. On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *J. Mach. Learn. Res.*, 22: 98:1–98:76, 2021.

Maruan Al-Shedivat, Trapit Bansal, Yura Burda, Ilya Sutskever, Igor Mordatch, and Pieter Abbeel. Continuous adaptation via meta-learning in nonstationary and competitive environments. In *6th International Conference on Learning Representations, ICLR 2018*. OpenReview.net, 2018.

Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $O(\log T)$ swap regret in multiplayer games. *arXiv preprint arXiv:2204.11417*, 2022a.

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On last-iterate convergence beyond zero-sum games. In *International Conference on Machine Learning, ICML 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 536–581. PMLR, 2022b.

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

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

Waïss Azizian, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. The last-iterate convergence rate of optimistic mirror descent in stochastic variational inequalities. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, pages 326–358. PMLR, 2021.

Maria-Florina Balcan, Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Commitment without regrets: Online learning in stackelberg security games. In *Proceedings of the sixteenth ACM conference on economics and computation*, pages 61–78, 2015a.

Maria-Florina Balcan, Avrim Blum, and Santosh S. Vempala. Efficient representations for lifelong learning and autoencoding. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015*, volume 40 of *JMLR Workshop and Conference Proceedings*, pages 191–210. JMLR.org, 2015b.

Maria-Florina Balcan, Mikhail Khodak, and Ameet Talwalkar. Provable guarantees for gradient-based meta-learning. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, volume 97 of *Proceedings of Machine Learning Research*, pages 424–433. PMLR, 2019.

Maria-Florina Balcan, Keegan Harris, Mikhail Khodak, and Zhiwei Steven Wu. Meta-learning adversarial bandits. *arXiv preprint arXiv:2205.14128*, 2022.

Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. *J. Mach. Learn. Res.*, 6:1705–1749, 2005.

Heinz H. Bauschke, Walaa M. Moursi, and Xianfu Wang. Generalized monotone operators and their averaged resolvents. *Math. Program.*, 189:55–74, 2021.

Nathalie Bertrand, Nicolas Markey, Suman Sadhukhan, and Ocan Sankur. Dynamic network congestion games. In *40th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2020*, volume 182 of *LIPIcs*, pages 40:1–40:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

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

Adam Block, Yuval Dagan, Noah Golowich, and Alexander Rakhlin. Smoothed online learning is as easy as statistical learning. In *Conference on Learning Theory, 2-5 July 2022*, volume 178 of *Proceedings of Machine Learning Research*, pages 1716–1786. PMLR, 2022.

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(6), 2007.

Noam Brown and Tuomas Sandholm. Regret transfer and parameter optimization. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014*, pages 594–601. AAAI Press, 2014.

Noam Brown and Tuomas Sandholm. Regret-based pruning in extensive-form games. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, pages 1972–1980, 2015a.

Noam Brown and Tuomas Sandholm. Simultaneous abstraction and equilibrium finding in games. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, pages 489–496. AAAI Press, 2015b.

Noam Brown and Tuomas Sandholm. Strategy-based warm starting for regret minimization in games. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 432–438. AAAI Press, 2016.

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

Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, 2019*, pages 1829–1836. AAAI Press, 2019.

Yang Cai and Weiqiang Zheng. Accelerated single-call methods for constrained min-max optimization. *CoRR*, abs/2210.03096, 2022.

Yang Cai, Ozan Candogan, Constantinos Daskalakis, and Christos H. Papadimitriou. Zero-sum polymatrix games: A generalization of minmax. *Math. Oper. Res.*, 41(2):648–655, 2016.

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

Adrian Rivera Cardoso, Jacob D. Abernethy, He Wang, and Huan Xu. Competing against nash equilibria in adversarially changing zero-sum games. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, volume 97 of *Proceedings of Machine Learning Research*, pages 921–930. PMLR, 2019.

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

Nicolò Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order bounds for prediction with expert advice. *Mach. Learn.*, 66(2-3):321–352, 2007.

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

Xi Chen, Christos H. Papadimitriou, and Binghui Peng. Memory bounds for continual learning. *CoRR*, abs/2204.10830, 2022.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *COLT 2012 - The 25th Annual Conference on Learning Theory, 2012*, volume 23 of *JMLR Proceedings*, pages 6.1–6.20. JMLR.org, 2012.

George Christodoulou and Elias Koutsoupias. The price of anarchy of finite congestion games. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 67–73. ACM, 2005.

George Christodoulou, Annamária Kovács, and Michael Schapira. Bayesian combinatorial auctions. *J. ACM*, 63(2):11:1–11:19, 2016.

Patrick L. Combettes and Teemu Pennanen. Proximal methods for cohypomonotone operators. *SIAM J. Control. Optim.*, 43(2):731–742, 2004.

Constantinos Daskalakis and Noah Golowich. Fast rates for nonparametric online learning: from realizability to learning in games. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, 2022*, pages 846–859. ACM, 2022.

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

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. *Games Econ. Behav.*, 92:327–348, 2015.

Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. In *6th International Conference on Learning Representations, ICLR 2018*. OpenReview.net, 2018.

Constantinos Daskalakis, Dylan J. Foster, and Noah Golowich. Independent policy gradient methods for competitive reinforcement learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021*, pages 27604–27616, 2021.

Jelena Diakonikolas, Constantinos Daskalakis, and Michael I. Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021*, volume 130 of *Proceedings of Machine Learning Research*, pages 2746–2754. PMLR, 2021.

John C. Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, 2011.

Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. Multiagent online learning in time-varying games. *Mathematics of Operations Research*, 2022.

Gabriele Farina and Tuomas Sandholm. Fast payoff matrix sparsification techniques for structured extensive-form games. In *AAAI Conference on Artificial Intelligence*, 2022.

Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning for general convex games. *CoRR*, abs/2206.08742, 2022.

Tanner Fiez, Lillian J. Ratliff, Eric Mazumdar, Evan Faulkner, and Adhyyan Narang. Global convergence to local minmax equilibrium in classes of nonconvex zero-sum games. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021*, pages 29049–29063, 2021a.

Tanner Fiez, Ryann Sim, Stratis Skoulakis, Georgios Piliouras, and Lillian J. Ratliff. Online learning in periodic zero-sum games. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021*, pages 10313–10325, 2021b.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1126–1135. PMLR, 2017.

Chelsea Finn, Aravind Rajeswaran, Sham M. Kakade, and Sergey Levine. Online meta-learning. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, volume 97 of *Proceedings of Machine Learning Research*, pages 1920–1930. PMLR, 2019.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.

Yuan Gao, Christian Kroer, and Donald Goldfarb. Increasing iterate averaging for solving saddle-point problems. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 7537–7544. AAAI Press, 2021.

Angeliki Giannou, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Panayotis Mertikopoulos. 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*, pages 2147–2148. PMLR, 2021.

Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. In *7th International Conference on Learning Representations, ICLR 2019*. OpenReview.net, 2019.

Andrew Gilpin, Javier Peña, and Tuomas Sandholm. First-order algorithm with $\mathcal{O}(\ln(1/\epsilon))$ convergence for ϵ-equilibrium in two-person zero-sum games. *Math. Program.*, 133(1-2):279–298, 2012.

Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020a.

Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman E. Ozdaglar. 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*, pages 1758–1784. PMLR, 2020b.

Nika Haghtalab, Yanjun Han, Abhishek Shetty, and Kunhe Yang. Oracle-efficient online learning for beyond worst-case adversaries. *CoRR*, abs/2202.08549, 2022.

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

Sergiu Hart and David Schmeidler. Existence of correlated equilibria. *Mathematics of Operations Research*, 14(1):18–25, 1989.

Jason D. Hartline, Vasilis Syrgkanis, and Éva Tardos. No-regret learning in bayesian games. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, pages 3061–3069, 2015.

Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: regret bounded by variation in costs. *Mach. Learn.*, 80(2-3):165–188, 2010.

Elad Hazan and Satyen Kale. Better algorithms for benign bandits. *J. Mach. Learn. Res.*, 12: 1287–1311, 2011.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192, 2007.

Martin Hoefer, Vahab S. Mirrokni, Heiko Röglin, and Shang-Hua Teng. Competitive routing over time. *Theor. Comput. Sci.*, 412(39):5420–5432, 2011.

Josef Hofbauer and William H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.

Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019*, pages 6936–6946, 2019.

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

Yu-Guan Hsieh, Kimon Antonakopoulos, Volkan Cevher, and Panayotis Mertikopoulos. No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation. *CoRR*, abs/2206.06015, 2022.

Ramesh Johari and John N. Tsitsiklis. Efficiency loss in a network resource allocation game. *Mathematics of Operations Research*, 29(3):407–435, 2004.

Mert Kayaalp, Stefan Vlaski, and Ali H. Sayed. Dif-maml: Decentralized multi-agent meta-learning. *CoRR*, abs/2010.02870, 2020.

Mert Kayaalp, Stefan Vlaski, and Ali H Sayed. Distributed meta-learning with networked agents. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pages 1361–1365. IEEE, 2021.

Mikhail Khodak, Maria-Florina F Balcan, and Ameet S Talwalkar. Adaptive gradient-based meta-learning methods. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Mikhail Khodak, Maria-Florina Balcan, Ameet Talwalkar, and Sergei Vassilvitskii. Learning predictions for algorithms with predictions. In *Advances in Neural Information Processing Systems*, 2022. To appear.

Robert Kleinberg, Georgios Piliouras, and Éva Tardos. Multiplicative updates outperform generic no-regret learning in congestion games: extended abstract. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009*, pages 533–542. ACM, 2009.

Daphne Koller and Nimrod Megiddo. The complexity of two-person zero-sum games in extensive form. *Games and Economic Behavior*, 4(4):528–552, 1992.

Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

Christian Kroer and Tuomas Sandholm. A unified framework for extensive-form game abstraction with bounds. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018*, pages 613–624, 2018.

Stefanos Leonardos, Will Overman, Ioannis Panageas, and Georgios Piliouras. Global convergence of multi-agent policy gradient in markov potential games. In *The Tenth International Conference on Learning Representations, ICLR 2022*. OpenReview.net, 2022.

Shuangtong Li, Tianyi Zhou, Xinmei Tian, and Dacheng Tao. Learning to collaborate in decentralized learning of personalized models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9766–9775, 2022.

Yuqian Li, Vincent Conitzer, and Dmytro Korzhyk. Catcher-evader games. *arXiv preprint arXiv:1602.01896*, 2016.

Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few shot learning. *CoRR*, abs/1707.09835, 2017.

Brendan Lucier and Allan Borodin. Price of anarchy for greedy auctions. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010*, pages 537–553. SIAM, 2010.

Haipeng Luo and Robert E. Schapire. Achieving all with no parameters: Adanormalhedge. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015*, volume 40 of *JMLR Workshop and Conference Proceedings*, pages 1286–1304. JMLR.org, 2015.

Thodoris Lykouris, Vasilis Syrgkanis, and Éva Tardos. Learning and efficiency in games with dynamic population. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016*, pages 120–129. SIAM, 2016.

H. Brendan McMahan. 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*, pages 525–533. JMLR.org, 2011.

Emily Meigs, Francesca Parise, and Asuman E. Ozdaglar. Learning dynamics in stochastic routing games. In *55th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2017*, pages 259–266. IEEE, 2017.

Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *7th International Conference on Learning Representations, ICLR 2019*. OpenReview.net, 2019.

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

Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. In Tim Roughgarden, editor, *Beyond the Worst-Case Analysis of Algorithms*, pages 646–662. Cambridge University Press, 2020.

Aryan Mokhtari, Asuman E. Ozdaglar, and Sarath Pattathil. 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*, pages 1497–1507. PMLR, 2020.

Denis Nekipelov, Vasilis Syrgkanis, and Éva Tardos. Econometrics for learning agents. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC '15*, pages 1–18. ACM, 2015.

J. V. Neumann. A model of general economic equilibrium. *The Review of Economic Studies*, 13(1): 1–9, 1945.

Noam Nisan and Gali Noti. An experimental evaluation of regret-based econometrics. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017*, pages 73–81. ACM, 2017.

Ilya Osadchiy, Kfir Y Levy, and Ron Meir. Online meta-learning in adversarial multi-armed bandits. *arXiv preprint arXiv:2205.15921*, 2022.

Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Optimal no-regret learning in general games: Bounded regret with unbounded step-sizes via clairvoyant MWU. *CoRR*, abs/2111.14737, 2021.

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

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013*, pages 3066–3074, 2013b.

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

J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, 1965.

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

Tim Roughgarden, Vasilis Syrgkanis, and Éva Tardos. The price of anarchy in auctions. *J. Artif. Intell. Res.*, 59:59–101, 2017.

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

Arun Sai Suggala and Praneeth Netrapalli. Follow the perturbed leader: Optimism and fast parallel algorithms for smooth minimax games. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020.

Vasilis Syrgkanis and Éva Tardos. Composable and efficient mechanisms. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, 2013*, pages 211–220. ACM, 2013.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, pages 2989–2997, 2015.

Oskari Tammelin. Solving large imperfect information games using CFR+. *CoRR*, abs/1407.5042, 2014.

Sebastian Thrun and Lorien Pratt. *Learning to learn*. Springer Science & Business Media, 1998.

Eric van Damme. *Stability and Perfection of Nash Equilibria*. Springer-Verlag, Berlin, Heidelberg, 1987.

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

Jun-Kun Wang, Jacob D. Abernethy, and Kfir Y. Levy. No-regret dynamics in the fenchel game: A unified framework for algorithmic convex optimization. *CoRR*, abs/2111.11309, 2021.

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

Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *9th International Conference on Learning Representations, ICLR 2021*. OpenReview.net, 2021.

Andre Wibisono, Molei Tao, and Georgios Piliouras. Alternating mirror descent for constrained min-max games. *CoRR*, abs/2206.04160, 2022.

Tianbao Yang, Lijun Zhang, Rong Jin, and Jinfeng Yi. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 449–457. JMLR.org, 2016.

Hugh Zhang, Adam Lerer, and Noam Brown. Equilibrium finding in normal-form games via greedy regret minimization. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, 2022*, pages 9484–9492. AAAI Press, 2022a.

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

Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Dynamic regret of convex and smooth functions. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020.