Efficient Learning in Polyhedral Games via Best-Response Oracles*

Darshan Chakrabarti[1], **Gabriele Farina**[2], **Christian Kroer**[1]

[1]Columbia University
[2]MIT
dc3595@columbia.edu, gfarina@mit.edu, ck2945@columbia.edu

Abstract

We study online learning and equilibrium computation in games with polyhedral decision sets, a property shared by normal-form games (NFGs) and extensive-form games (EFGs), when the learning agent is restricted to utilizing a best-response oracle. We show how to achieve constant regret in zero-sum games and $O(T^{1/4})$ regret in general-sum games while using only $O(\log t)$ best-response queries at a given iteration t, thus improving over the best prior result, which required $O(T)$ queries per iteration. Moreover, our framework yields the first last-iterate convergence guarantees for self-play with best-response oracles in zero-sum games. This convergence occurs at a linear rate, though with a condition-number dependence. We go on to show a $O(1/\sqrt{T})$ best-iterate convergence rate without such a dependence. Our results build on linear-rate convergence results for variants of the Frank-Wolfe (FW) algorithm for strongly convex and smooth minimization problems over polyhedral domains. These FW results depend on a condition number of the polytope, known as facial distance. In order to enable application to settings such as EFGs, we show two broad new results: 1) the facial distance for polytopes of the form $\{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \mid \mathbf{A}\boldsymbol{x} = \boldsymbol{b}\}$ is at least $\gamma/\sqrt{k}$ where γ is the minimum value of a nonzero coordinate of a vertex in the polytope and $k \leq n$ is the number of tight inequality constraints in the optimal face, and 2) the facial distance for polytopes of the form $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \mathbf{C}\boldsymbol{x} \leq \boldsymbol{d}, \boldsymbol{x} \geq \mathbf{0}$ where $\boldsymbol{x} \in \mathbb{R}^n$, $\mathbf{C} \geq \mathbf{0}$ is a nonzero integral matrix, and $\boldsymbol{d} \geq \mathbf{0}$, is at least $1/(\|\mathbf{C}\|_\infty \sqrt{n})$. This yields the first such results for several problems, such as sequence-form polytopes, flow polytopes, and matching polytopes.

1 Introduction

Learning in games is a well-studied framework in which agents iteratively refine their strategies through repeated interactions with their environment. One natural way for agents to iteratively refine their strategies is by best-responding. This idea can be applied in many forms, the simplest and earliest instance of which was fictitious play (FP) (Brown 1951). This algorithm involves the agent observing the strategies played by the opponent and then playing a strategy that corresponds to the best response to the average of the observed strategies. FP was shown to converge (Robinson 1951), but its convergence rate can, in the worst case, scale quite poorly with the number of actions available to each player (Daskalakis and Pan 2014). It is then natural to ask what are the best convergence guarantees that can be obtained for the computation of Nash equilibria in two-player zero-sum games or coarse correlated equilibria in multiplayer games when agents are learning through a best-response oracle.

In the online learning community, methods based only on best-response oracles are special cases of methods based on a *linear minimization oracle* (LMO), which can be queried for points that minimize a linear objective over the feasible set. Such methods are known as *projection-free* methods because they avoid potentially expensive projections onto the feasible set. Projection-free online learning algorithms might perform multiple LMO calls per iteration, so our paper and related literature are concerned not only with the number of iterations T of online learning but also the total number of LMO calls, which we will denote by N. Because LMOs for polyhedral decision sets essentially correspond to best-response oracles (BROs), we will use these two terms interchangeably.

Follow the Perturbed Leader (FTPL) (Kalai and Vempala 2005) was the first such algorithm. When used by both players in two-player zero-sum games, it yields a $O(1/\sqrt{T})$ convergence rate to Nash equilibrium, with a single LMO call in each iteration. More recently, Suggala and Netrapalli (2020) proposed an optimistic variant of FTPL (OFTPL); OFTPL achieves a $O(1/T)$ rate of convergence to Nash equilibrium but requires $O(T)$ LMO calls per iteration, thus corresponding to a $O(1/\sqrt{N})$ rate as a function of the total number of LMO calls. Online Frank-Wolfe (OFW) (Hazan and Kale 2012) is another projection-free algorithm, based on the well-studied Frank-Wolfe (FW) algorithm (Frank and Wolfe 1956). While it does not require multiple calls to an LMO, it can only achieve a $O(T^{-1/3})$ rate for two-player zero-sum games. We aim to break the $O(1/\sqrt{N})$ barrier in terms of convergence towards two-player zero-sum equilibria and beyond. We focus on the setting of polyhedral games, a class containing both normal-form and extensive-form games. Our primary contribution is an online learning method which enjoys $O(1/T)$ average individual regret in zero-sum games and $O(1/T^{3/4})$ average individual regret in general-sum games while only requiring $O(\log t)$ LMO

*The full version of this paper (including appendices) can be found at https://arxiv.org/abs/2312.03696.

calls in iteration t. Table 1 compares our algorithm with other algorithms with the best-known guarantees for the setting we consider. In the table, we also include a non-optimistic version of our algorithm; despite having worse theoretical guarantees than existing projection-free algorithms, it outperforms them in our numerical experiments.

Independent of work in projection-free online learning, there has also been substantial work developing BRO-based algorithms in the game-solving community. Most prominent is the Double Oracle (DO) algorithm (McMahan, Gordon, and Blum 2003) for computing Nash equilibria in two-player zero-sum games, which uses BROs and a meta-solver for computing Nash equilibria in a restricted game formed by the returned iterates. More recently, the Policy Space Response Oracle framework (Lanctot et al. 2017), a generalization of Double Oracle, has laid the foundation for the design of DO-variants. These algorithms only have theoretical guarantees for computing Nash equilibria in two-player zero-sum games, require a meta-solver for solving the restricted game composed of the strategies chosen by the players at each iteration, and thus are centralized, and do not have convergence rate guarantees for convergence to equilibria.

The optimization community has also done substantial work on developing projection-free methods, spurred by the work of Frank and Wolfe (1956). Guarantees for FW typically assume the function f being optimized is smooth (has Lipschitz gradient) and convex, the domain $\mathcal{X}$ being optimized over is convex and compact, and that the algorithm has access to a first-order oracle for the function which returns gradients $\nabla f(\boldsymbol{x})$ at a queried point $\boldsymbol{x}$ and an LMO which returns solutions to minimization problems of the form $\operatorname{argmin}_{\boldsymbol{x}\in\mathcal{X}}\langle \boldsymbol{c}, \boldsymbol{x}\rangle$ for any choice of $\boldsymbol{c} \in \mathbb{R}^n$. Given an initial iterate $\boldsymbol{x}^{(0)}$, it produces new iterates given by the following update rule:

$$\boldsymbol{x}^{(t+1)} = \frac{t}{t+2}\boldsymbol{x}^{(t)} + \frac{2}{t+2}\operatorname*{argmin}_{\boldsymbol{x}'\in\mathcal{X}}\langle \nabla f(\boldsymbol{x}^{(t)}), \boldsymbol{x}'\rangle.$$

In recent years, there has been work on developing FW-based approaches to saddle-point computation (e.g., Gidel, Jebara, and Lacoste-Julien (2017); Lan and Zhou (2016)). However, Gidel, Jebara, and Lacoste-Julien (2017) only have fast convergence guarantees for strongly convex-concave objectives and Lan and Zhou (2016) are only able to provide $O(1/\sqrt{N})$ convergence to saddle-points. On the other hand, our method is able to leverage a FW variant, away-step Frank-Wolfe (AFW), to achieve faster convergence rates.

An extended discussion of related work is given in Section 1.2 of the full version of the paper.

1.1 Contributions

We present a projection-free online learning method, Approximate Reflected Online Mirror Descent using away-step Frank-Wolfe (AFW-ROMD), for learning over compact and convex polyhedral decision sets. AFW-ROMD uses reflected online mirror descent (ROMD), an optimistic online learning algorithm which requires a prediction of the next loss, instantiated with the Euclidean regularizer. The proximal problem for this ROMD setup is thus strongly convex and smooth. Using the linear convergence of AFW for polyhedral domains, we implement approximate steps of ROMD using only a logarithmic number of AFW iterations. We show that even with approximate steps, ROMD still yields an approximate *regret bounded by variation in utilities* (RVU) bound (Syrgkanis et al. 2015), a form of regret guarantee that depends on how much the observed utilities vary, and which has enabled proving constant regret of *optimistic* learning dynamics in two-player zero-sum games and $O(T^{1/4})$ regret in general games (Syrgkanis et al. 2015). While using FW-based methods to approximate proximal steps has been previously studied, pioneered by work of Lan and Zhou (2016), it is a surprising blind spot in the literature that the connection to regret guarantees for games has not previously been made.

We then apply our AFW-ROMD algorithm to learning in games using only BROs (as opposed to typical learning algorithms, which require access to Euclidean projection or other proximal oracles). We show that when every player employs AFW-ROMD, it is possible to converge to a Nash equilibrium in a two-player zero-sum game at a rate of $O(\log N/N)$, where N is the number of BRO calls. In contrast, the best prior results converged at a rate of $O(1/\sqrt{N})$ (Suggala and Netrapalli 2020). More generally, we show that AFW-ROMD requires only $O(\log t)$ best-response queries at each self-play iteration t while guaranteeing constant social regret, as well as $O(T^{1/4})$ regret for each player after T total iterations of self-play. We go on to study the *last-iterate convergence* properties of AFW-ROMD in self-play in zero-sum settings. We show that, indeed, it is possible to retain last-iterate convergence under AFW-ROMD, and in fact, it converges at a linear rate (up to error induced by the approximate proximal computation). To the best of our knowledge, these are both the first last-iterate convergence and the first linear-rate convergence results for self-play dynamics that purely rely on best-response oracles. As with existing linear-rate results in the literature (Tseng 1995; Gilpin, Peña, and Sandholm 2012; Wei et al. 2021), we have a dependence on a condition number inherent to the game, which is generally hard to evaluate. We show that if one wishes to avoid this condition-number dependence, then self-play with AFW-ROMD still achieves a $O(1/\sqrt{T})$ *best-iterate* convergence. Our convergence results to Nash equilibria and coarse correlated equilibria are summarized below using the following two informal theorems:

Theorem 1 (Informal; Full version in Theorems 6, 8 and 9)**.** *Using* AFW-ROMD *an ϵ'-Nash equilibrium in a two-player zero-sum polyhedral game can be computed in $O(1/\epsilon')$ iterations and $O(\frac{1}{\epsilon'}\log\frac{1}{\epsilon'})$ LMO calls. Furthermore, in two-player zero-sum games,* AFW-ROMD *will produce an iterate which is an ϵ'-Nash equilibrium in $O(\frac{1}{\epsilon'^2}\log\frac{1}{\epsilon'})$ LMO calls without a dependence on a problem-dependent constant.* AFW-ROMD *will produce an iterate which is an ϵ'-Nash equilibrium in $O(\log\frac{1}{\epsilon'})$ iterations, with this convergence rate having a dependence on a game-dependent constant. Exact dependence on problem parameters can be found in the formal theorems.*

Theorem 2 (Informal; Full version in Theorem 7)**.** *In multiplayer polyhedral games,* AFW-ROMD *can be used to compute an ϵ'-coarse correlated equilibrium using $O(\frac{1}{\epsilon'^{\frac{4}{3}}}\log\frac{1}{\epsilon'})$ LMO calls. Exact dependence on problem parameters can be*

found in the formal theorem.

The linear convergence of AFW on polyhedral domains is crucial to our result, but the particular rate depends on the facial distance constant of the polytope in question (in addition to the strong convexity and smoothness constants). To that end, we show two novel lower bounds on the facial distance of a polytope. Our first result concerns polytopes that can be described in the form $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \mathbf{0}$ where $\boldsymbol{x} \in \mathbb{R}^n$. Let γ be the minimum value of a nonzero coordinate of a vertex in the polytope. Then, we show that the facial distance is at least $\gamma/\sqrt{n}$.

Moreover, if the optimal solution lies in a face such that $k \leq n$ inequality constraints $x_i \geq 0$ are tight, then the rate dependence in AFW can be tightened to $\frac{\gamma}{\sqrt{k}}$. This theorem immediately implies several useful results, including a bound $1/\sqrt{n}$ on the facial distance of the *sequence-form polytope*, which is the polytope describing the set of feasible strategies of a player in an EFG, where n is the number of sequences for the player. It also implies similar results for flow polytopes and matching polytopes. The fact that the facial distance is only square-root power small in the dimension of the problem ensures that the convergence rate of linearly convergent FW variants over these polytopes does not scale poorly as the ambient dimension of the problem increases. Our second result concerns an *integral* polytope $\mathcal{X}$ given by $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \mathbf{C}\boldsymbol{x} \leq \boldsymbol{d}, \boldsymbol{x} \geq \mathbf{0}$ where $\boldsymbol{x} \in \mathbb{R}^n$, where $\mathbf{C} \geq \mathbf{0}$ is a nonzero integral matrix, and $\boldsymbol{d} \geq \mathbf{0}$. In particular, for integral polytopes, we are able to handle inequality constraints. In that case, we show that the facial distance is at least $1/(\|\mathbf{C}\|_\infty \sqrt{n})$.

Finally, we conduct experiments to demonstrate that our algorithm performs well in practice relative to other projection-free algorithms when computing Nash and coarse correlated equilibria in polyhedral games.

2 Notation and Preliminaries

We will use $\|\cdot\|_p$ to denote the $\boldsymbol{\ell}_p$-norm and $\|\cdot\|$ without subscript to denote $\|\cdot\|_2$. Any norm-dependent quantity (e.g., diameter, facial distance, strong convexity, and smoothness) will be with respect to the Euclidean norm (which is self-dual) unless otherwise noted. Because we are principally concerned with using these algorithms for equilibrium computation in games, we will use subscripts to indicate a set or constant corresponding to a particular agent. We will use $[n]$ to denote the set $\{1, \ldots, n\}$, and L-smoothness refers to Lipschitz continuity of the gradient, with modulus L.

2.1 Online Linear Optimization

In online learning, an agent i repeatedly interacts with an environment, aiming to minimize its *regret*. At each time t, the agent chooses a strategy $\boldsymbol{x}_i^{(t)}$ from a given feasible set $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ and then receives a loss vector $\boldsymbol{\ell}_i^{(t)} \in \mathcal{X}_i \to \mathbb{R}$. The loss is allowed to depend adversarially on $\boldsymbol{x}_i^{(t)}$. The agent then pays a cost of $\langle \boldsymbol{\ell}_i^{(t)}, \boldsymbol{x}_i^{(t)} \rangle$. The (cumulative) regret $\mathrm{Reg}_i^{(T)}$ after T iterations is defined as $\max_{\boldsymbol{x}' \in \mathcal{X}_i} \sum_{t=1}^{T} \langle \boldsymbol{\ell}_i^{(t)}, \boldsymbol{x}_i^{(t)} \rangle - \langle \boldsymbol{\ell}_i^{(t)}, \boldsymbol{x}' \rangle$, and average regret is defined as regret divided by the number of iterations. We will assume that losses are bounded and normalized: $\|\boldsymbol{\ell}_i^{(t)}\| \leq 1$ for all $t \in [T]$.

In order to achieve desired regret guarantees, online learning algorithms typically require some form of regularization. While FTPL achieves this regularization through randomization, the framework of algorithms utilizing approximate prox calls that we present will require access to a regularizer $\varphi_i : \mathcal{X}_i \to \mathbb{R}$, which is 1-strongly convex and L_i smooth on $\mathcal{X}_i$. The Bregman divergence between $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_i$ is denoted by $\mathcal{D}_{\varphi_i}(\boldsymbol{x} \,\|\, \boldsymbol{y})$. Furthermore, we define $\Omega_i := \sup_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_i} \mathcal{D}_{\varphi_i}(\boldsymbol{x} \,\|\, \boldsymbol{y})$ and $D_i := \sup_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_i} \|\boldsymbol{x} - \boldsymbol{y}\|$. $\delta(\mathcal{X}_i)$ will be used for the facial distance of $\mathcal{X}_i$, defined in Section 2.4. For a given set $\mathcal{X}$, and a point $\boldsymbol{x} \in \mathcal{X}$, we denote $\mathrm{dist}(\boldsymbol{x}, \mathcal{X}) := \inf_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{x} - \boldsymbol{x}'\|$ and in the case that $\mathcal{X}$ is compact, define $\Pi_{\mathcal{X}}(\boldsymbol{x}) = \mathrm{argmin}_{\boldsymbol{x}' \in \mathcal{X}} \|\boldsymbol{x} - \boldsymbol{x}'\|$.

Online Mirror Descent (OMD) is an algorithm which performs a single proximal computation at every iteration of the algorithm, generating iterates as follows:

$$\boldsymbol{x}_i^{(t+1)} = \operatorname*{argmin}_{\boldsymbol{x}_i \in \mathcal{X}_i} \left\{ \langle \boldsymbol{\ell}_i^{(t)}, \boldsymbol{x}_i \rangle + \frac{\mathcal{D}_{\varphi_i}(\boldsymbol{x}_i \| \boldsymbol{x}_i^{(t)})}{\eta} \right\}.$$

It enjoys $O(1/\sqrt{T})$ average regret (e.g., Hazan et al. (2016); Orabona (2019)).

Reflected Online Mirror Descent (ROMD) is an optimistic version of OMD which utilizes a prediction $\boldsymbol{m}_i^{(t+1)}$ of the next loss $\boldsymbol{\ell}_i^{(t+1)}$ to generate the iterate at time $t+1$:

$$\boldsymbol{x}_i^{(t+1)} = \operatorname*{argmin}_{\boldsymbol{x}_i \in \mathcal{X}_i} \left\{ \langle \boldsymbol{\ell}_i^{(t)} + \boldsymbol{m}_i^{(t+1)} - \boldsymbol{m}_i^{(t)}, \boldsymbol{x}_i \rangle + \frac{\mathcal{D}_{\varphi_i}(\boldsymbol{x}_i \| \boldsymbol{x}_i^{(t)})}{\eta} \right\}.$$

It is common to use the last observed loss as the prediction for the next loss: set $\boldsymbol{m}_i^{(t+1)}$ equal to $\boldsymbol{\ell}_i^{(t)}$. In this case, ROMD achieves $O(1/T)$ average regret (Malitsky 2015; Joulani, György, and Szepesvári 2017) in self-play. Since $\boldsymbol{m}_i^{(t+1)}$ is the prediction of a loss $\boldsymbol{\ell}_i^{(t+1)}$ which is assumed to have norm bounded by 1, we will assume that $\|\boldsymbol{m}_i^{(t)}\| \leq 1$ for all $t \in [T]$.

Syrgkanis et al. (2015) introduce the notion of Regret bounded by Variation in Utilities (RVU), recalled next, and demonstrate that algorithms with this property exhibit faster convergence to equilibria in games.

Definition 1 (RVU (Syrgkanis et al. 2015))**.** *A learning algorithm for Player i is said to satisfy the RVU property if for some $\alpha, \beta, \gamma > 0$, and all possible $\boldsymbol{\ell}_i^{(1)}, \ldots, \boldsymbol{\ell}_i^{(T)}$,*

$$\mathrm{Reg}_i^{(T)} \leq \alpha + \beta \sum_{t=1}^{T} \|\boldsymbol{\ell}_i^{(t)} - \boldsymbol{\ell}_i^{(t-1)}\|^2 - \gamma \sum_{t=1}^{T} \|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}_i^{(t-1)}\|^2.$$

ROMD satisfies this inequality with $\alpha = \Omega/\eta$, $\beta = \eta$, $\gamma = 1/4\eta$; we are not aware of a reference for this, but it can be shown very similarly to known results for optimistic OMD. Later we will show in Lemma 2 that our approximate ROMD in Algorithm 1 still satisfies the RVU property.

2.2 Game-Theoretic Notions

Normal-form games (NFGs) model single-shot simultaneous interactions among a set of agents denoted by $[N]$. The agents

Algorithm	∇ computations at iteration t	LMO calls at iteration t	Social regret $\sum_i \mathrm{Reg}_i^{(T)}$	Avg. social regret $\sum_i \frac{1}{T}\mathrm{Reg}_i^{(T)}$
FTPL (Kalai and Vempala 2005)	$O(1)$	$O(1)$	$O(\sqrt{T})$	$O(1/\sqrt{N})$
OFTPL (Suggala and Netrapalli 2020)	$O(1)$	$O(T)$	$O(1)$	$O(1/\sqrt{N})$
AFW-OMD [this paper]	$O(1)$	$O(\log t)$	$O(\sqrt{T})$	$O(\sqrt{\log N/N})$
AFW-ROMD [this paper]	$O(1)$	$O(\log t)$	$O(1)$	$O(\log N/N)$

Table 1: A comparison of the number of gradient (∇) computations, number of LMO calls, cumulative regret (as a function of the total number of iterations T), and average regret (as a function of total LMO calls N) of various projection-free algorithms. In two-player zero-sum games, average social regret upper bounds the duality gap to Nash equilibrium for the averaged iterates.

each have a set of possible actions $\mathcal{A}_i$ and a normalized utility function $u_i : \prod_{i\in[N]} \mathcal{A}_i \to [-1, 1]$, the latter specifying their payoff for a given choice of actions by each of the agents. The game is said to be *zero-sum* if $\sum_{i\in[n]} u_i(\boldsymbol{a}) = 0$ for all $\boldsymbol{a} \in \prod_{i\in[N]} \mathcal{A}_i$. A mixed strategy $\boldsymbol{x}_i$ for Player i, is a probability distribution over $\mathcal{A}_i$; $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)$. We can extend the domain of u_i to be over $\Delta(\mathcal{A}_i)$ by taking the expectation of the utility function over the distribution over $\mathcal{A}_i$ induced by $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)$.

Nash equilibrium is the de facto notion of equilibrium in NFGs, and the problem of computing a Nash equilibrium (NE) in two-player zero-sum games can be formulated as a bilinear saddle-point problem (BSPP):

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} \langle \mathbf{A}\boldsymbol{x}, \boldsymbol{y}\rangle. \tag{BSPP}$$

In this case, $\mathcal{X}$ and $\mathcal{Y}$ are the space of mixed strategies for Player 1 and Player 2, respectively, and $\mathbf{A}$ encodes the utility of Player 2 for a given choice of strategies for both players. The duality gap ξ of $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ for (BSPP) can be defined as $\max_{\boldsymbol{y}\in\mathcal{Y}} \langle \mathbf{A}\bar{\boldsymbol{x}}, \boldsymbol{y}\rangle - \min_{\boldsymbol{x}\in\mathcal{X}} \langle \mathbf{A}\boldsymbol{x}, \bar{\boldsymbol{y}}\rangle$. This quantity is typically used to measure the quality of a solution; in the case of two-player zero-sum games, a duality gap of ϵ' corresponds to an ϵ'-NE (and thus is also known as Nash gap).

Definition 2 (ϵ'-coarse correlated equilibrium)**.** *An ϵ'-coarse correlated equilibrium (ϵ'-CCE) is defined as $\boldsymbol{x} \in \Delta(\prod_{i\in[N]} \mathcal{A}_i)$ such that*

$$\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{x}}[u_i(\boldsymbol{a})] \geq \mathbb{E}_{\boldsymbol{a}_{-i}\sim\boldsymbol{x}}[u_i(\boldsymbol{a}'_i, \boldsymbol{a}_{-i})] - \epsilon'$$

for all players $i \in [N]$, for all $\boldsymbol{a}'_i \in \mathcal{A}_i$, for $\epsilon' \geq 0$; $\epsilon' = 0$ corresponds to an exact CCE.

Extensive-form games (EFGs) are a generalization of normal-form games, which also allow for modeling of sequential moves (and also private and/or imperfect information and stochasticity). Nash equilibrium computation for two-player zero-sum EFGs can also be formulated as (BSPP), by letting $\mathcal{X}$ and $\mathcal{Y}$ be convex polytopes known as *sequence-form polytopes* (Romanovskii 1962; Koller, Megiddo, and von Stengel 1996; von Stengel 1996), and letting $\mathbf{A}$ be a matrix representing the utility of Player 2 for a given choice of strategies for both players. An important property that sequence-form polytopes have is they can be expressed in the form $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0$; this will allow us to characterize the facial distance of these polytopes. Additional background on EFGs can be found in Appendix B.

When we are analyzing general multiplayer polyhedral games involving N agents, we will let $\mathcal{X}_i \subset \mathbb{R}^{n_i}$ denote the convex and compact polyhedral set of strategies for the i^{th} player, where $i \in [N]$, and let $\boldsymbol{x}_i \in \mathcal{X}_i$ represent their strategy. For NFGs, $\mathcal{X}_i$ is $\Delta(\mathcal{A}_i)$, the set of mixed strategies for i, while for EFGs, $\mathcal{X}_i$ is the sequence-form polytope for Player i. In the case of two-player zero-sum NFGs or EFGs, in which case Nash equilibrium computation corresponds to (BSPP), we will let $\mathcal{X} = \mathcal{X}_1$, $\mathcal{Y} = \mathcal{X}_2$, and $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. In this case, we will use $\mathcal{Z}^*$ to denote the set of solutions to the BSPP. We define the vector field $\boldsymbol{F}(\boldsymbol{z}) = (\mathbf{A}^T\boldsymbol{y}, -\mathbf{A}\boldsymbol{x})$ for $\boldsymbol{z} \in \mathcal{Z}$. Without loss of generality, we assume that $\boldsymbol{F}$ is smooth with constant 1 (the payoff matrix $\mathbf{A}$ can be scaled to ensure this is the case).

2.3 Saddle-Point Metric-Subregularity

Problems of the form (BSPP) satisfy a condition known as *Saddle-Point Metric Subregularity* (Wei et al. 2021) as long as $\mathcal{X}$ and $\mathcal{Y}$ are convex polytopes (as is the case for NFGs and EFGs).

Definition 3 (Saddle-Point Metric Subregularity)**.** *The SP-MS condition is satisfied if for any $\boldsymbol{z} \in \mathcal{Z} \setminus \mathcal{Z}^*$ with $\boldsymbol{z}^* = \Pi_{\mathcal{Z}}(\boldsymbol{z})$ for some $\beta \geq 0$ and $\nu > 0$,*

$$\sup_{\boldsymbol{z}'\in\mathcal{Z}} \frac{\langle \boldsymbol{F}(\boldsymbol{z}), \boldsymbol{z} - \boldsymbol{z}'\rangle}{\|\boldsymbol{z} - \boldsymbol{z}'\|} \geq \nu \|\boldsymbol{z} - \boldsymbol{z}^*\|^{\beta+1}. \tag{SP-MS}$$

For two-player zero-sum polyhedral games, there exists $\nu \geq 0$ so that Inequality (SP-MS) holds with $\beta = 0$; given a choice of a game, we will use ν to refer to this problem-dependent constant. Wei et al. (2021) use this condition to demonstrate linear last-iterate convergence of certain online learning algorithms. Earlier works (Tseng 1995; Gilpin, Peña, and Sandholm 2012) showed linear last-iterate convergence using error bounds, and Wei et al. (2021) note that there is a close correspondence between the SP-MS condition and error bound techniques for bilinear polyhedral settings.

2.4 Frank-Wolfe and Facial Distance

Frank-Wolfe is a projection-free algorithm that converges with rate $O(1/T)$ for smooth convex functions over convex compact sets. In certain situations, faster convergence rates can be obtained; for example, when the function is strongly convex and the optimal solution lies in the relative interior of the feasible set, the original FW algorithm is linearly convergent (Guélat and Marcotte 1986). In the case

where the optimal solution is not in the interior, away-step Frank Wolfe (AFW) is a variant of Frank-Wolfe which was shown to achieve linear convergence for strongly convex objectives over polyhedral sets (Wolfe 1970; Guélat and Marcotte 1986; Lacoste-Julien and Jaggi 2015). Pseudocode for AFW is provided in Appendix A. The linear rate of AFW and several other linearly convergent variants of FW depends on a condition number of the polytope known as the facial distance. It can be defined concisely using a theorem from (Pena and Rodriguez 2019):

Definition 4 (Facial distance, Pena and Rodriguez (2019))**.**

$$\delta(\mathcal{P}) = \min_{\substack{\mathcal{F} \in \text{faces}(\mathcal{P}) \\ \emptyset \subseteq \mathcal{F} \subseteq \mathcal{P}}} \text{dist}(\mathcal{F}, \text{Conv}(\text{Vert}(\mathcal{P}) \setminus \mathcal{F})).$$

For linearly convergent methods for strongly convex functions, the ratio between the strong convexity modulus and smoothness constant appears as a term in the linear rate. For linearly convergent FW variants, both the former ratio and the ratio between the facial distance and diameter over the polyhedral domain appear as terms in the linear rate.

Theorem 3 (Convergence rate of AFW for strongly convex functions over polyhedral sets (Lacoste-Julien and Jaggi 2015)[1])**.** *In order to compute an ϵ-optimal solution to a 1-strongly convex, L-smooth function over a convex polytope that has diameter D and facial distance δ,* AFW *requires $O(\frac{LD^2}{\delta^2} \log \frac{LD}{\epsilon})$ LMO calls.*

This dependence on the facial distance to diameter ratio makes it desirable to have a lower bound on the facial distance, to ensure that the linear rate scales well with the size of the problem.

3 New Results on Polyhedral Facial Distance

In this section, we present two theorems which characterize lower bounds on the facial distance in special cases where the constraints of the polytope can be written in a certain form. The proofs are deferred to Appendix C.

Theorem 4. *Let $\mathcal{P}$ be a polytope given by $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \mathbf{0}$ where $\boldsymbol{x} \in \mathbb{R}^n$. Let γ be the minimum value of a nonzero coordinate of a vertex. Then $\delta(\mathcal{P}) \geq \frac{\gamma}{\sqrt{n}}$. Moreover, if the optimal solution lies in a face $\mathcal{F}$ such that k coordinates are zero (i.e., $|\{i : x_i = 0\}| = k$), then $\delta(\mathcal{P}) \geq \frac{\gamma}{\sqrt{k}}$.*

Corollary 1. *For any $0/1$-polytope $\mathcal{P}$, $\delta(\mathcal{P}) \geq \frac{1}{\sqrt{n}}$.*

The corollary follows by noting that Theorem 4 holds with $\gamma = 1$ for $0/1$ polytopes. Note that simplices (the strategy spaces of NFGs), sequence-form polytopes (the strategy spaces of EFGs), flow polytopes, and matching polytopes are all $0/1$ polytopes. The fact that the facial distance decreases only at a square-root rate in the dimension of the problem ensures reasonable scaling of AFW as the ambient dimension of the problem increases.

In the case we are dealing with integral polytopes (polytopes with integral vertices), we can also handle inequality constraints beyond the non-negativity constraints.

[1] *Pyramidal width* was used in the original linear-rate result (Lacoste-Julien and Jaggi 2015). It was shown to be equivalent to facial distance by Pena and Rodriguez (2019).

Theorem 5. *Let $\mathcal{P}$ be an integral polytope given by $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \mathbf{C}\boldsymbol{x} \leq \boldsymbol{d}, \boldsymbol{x} \geq \mathbf{0}$ where $\boldsymbol{x} \in \mathbb{R}^n$, with $\mathbf{C} \geq \mathbf{0}$ a nonzero integral matrix, and $\boldsymbol{d} \geq \mathbf{0}$. Then $\delta(\mathcal{P}) \geq \frac{1}{\|\mathbf{C}\|_\infty \sqrt{n}}$.*

We note that facial distance and essentially equivalent notions, have been considered non-trivial to evaluate (Garber and Meshi 2016; Bashiri and Zhang 2017; Braun et al. 2022) for most polytopes besides hypercubes, unit ℓ_1 balls and simplices. Thus, our lower bounds contribute to a more complete characterization of convergence rates of linearly convergent FW variants over a fairly broad class of polytopes.

4 Approximate Reflected Online Mirror Descent Using Away-Step Frank-Wolfe

In this section, we propose an algorithmic framework that uses approximate proximal updates instead of exact proximal updates. First, we will show that such an approximate variant of ROMD still retains many of the nice properties of ROMD, up to the error in the approximation oracle. Then, we propose the use of linearly convergent variants of FW for implementing the approximate proximal step, specifically when the regularizer is smooth, and strongly convex (as is the case with the Euclidean regularizer) and the decision set is a convex polytope, which is the case in NFGs and EFGs. We abstract away the concept of computing an approximate proximal update using what we call an approximate proximal oracle (APO).

Definition 5 (Approximate proximal oracle)**.** *An $\text{APO}_{\mathcal{X}}$, given a choice of convex and compact set $\mathcal{X}$, takes as input a function $f : \mathcal{X} \to \mathbb{R}$, a L-smooth and 1-strongly convex regularizer φ, a prox center $\boldsymbol{x}_c$, and a desired accuracy $\epsilon \geq 0$, and returns $\boldsymbol{x}' \in \mathcal{X}$ such that*

$$f(\boldsymbol{x}') + \mathcal{D}_\varphi(\boldsymbol{x}' \,\|\, \boldsymbol{x}_c) \leq \min_{\boldsymbol{x} \in \mathcal{X}} \Big\{ f(\boldsymbol{x}) + \mathcal{D}_\varphi(\boldsymbol{x} \,\|\, \boldsymbol{x}_c) \Big\} + \epsilon. \tag{APO}$$

While our framework can be adapted to a variety of online learning algorithms, we illustrate the framework using ROMD in Algorithm 1. $\boldsymbol{x}_i^{(t)}$ is the iterate returned by the algorithm at iteration t, $\boldsymbol{\ell}_i^{(t)}$ is the loss received at iteration t, and $\boldsymbol{m}_i^{(t)}$ is the prediction of the loss to be used at iteration t. Proofs for results in this section are deferred to Appendices D and E. The subscript i is dropped in statements about a single regret-minimizing agent applying the algorithm; the only exceptions are this paragraph and the pseudocode in Algorithm 1.

We present the following lemma, which characterizes the cumulative regret of Algorithm 1 when using an APO.

Lemma 1. *Let $\boldsymbol{x}^* = \text{argmax}_{\boldsymbol{x}' \in \mathcal{X}} \sum_{t=1}^T \langle \boldsymbol{\ell}^{(t)}, \boldsymbol{x}^{(t)} \rangle - \langle \boldsymbol{\ell}^{(t)}, \boldsymbol{x}' \rangle$. Algorithm 1 yields*

$$\begin{aligned} \text{Reg}^{(T)} \leq & \sum_{t=1}^T \|\boldsymbol{\ell}^{(t)} - \boldsymbol{m}^{(t)}\| \cdot \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\| \\ & - \frac{1}{2\eta} \sum_{t=1}^T \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|^2 + \frac{1}{\eta} \mathcal{D}_\varphi\Big(\boldsymbol{x}^* \,\|\, \boldsymbol{x}^{(0)}\Big) \\ & + \big\langle \boldsymbol{m}^{(1)}, \boldsymbol{x}^{(2)} - \boldsymbol{x}^* \big\rangle + \sum_{t=1}^T \frac{\epsilon^{(t)}}{\eta}. \end{aligned}$$

Algorithm 1: Reflected Gradient OMD with Approximate Proximal Computation (for a generic Player i)

Data: $\mathcal{X}_i \subseteq \mathbb{R}^n$: convex and compact set, $\varphi_i : \mathcal{X} \to \mathbb{R}_{\geq 0}$: L_i-smooth, 1-strongly convex, $\eta_i > 0$: step-size parameter, $\epsilon_i^{(t)}$: desired accuracy of prox call at each t, $\text{APO}_{\mathcal{X}_i}$: an APO for $\mathcal{X}_i$, $\boldsymbol{x}_i^{(0)} \in \mathcal{X}_i$

$\boldsymbol{\ell}_i^{(0)} = \boldsymbol{m}_i^{(0)} = \mathbf{0}$;

function NEXTSTRATEGY($\boldsymbol{m}_i^{(t)}$)

$\quad \text{APO}_{\mathcal{X}_i}\Big(-\eta_i\Big\langle \boldsymbol{\ell}_i^{(t-1)} + \boldsymbol{m}_i^{(t)} - \boldsymbol{m}_i^{(t-1)}, \cdot \Big\rangle, \varphi_i, \boldsymbol{x}_i^{(t-1)}, \epsilon_i^{(t)}\Big)$

In the rest of this section, for the sake of brevity, we state our results using $\epsilon^{(t)} = \frac{1}{t^2}$, and the last observed loss as the prediction $\boldsymbol{m}^{(t)} = \boldsymbol{\ell}^{(t-1)}$. In this case, we obtain the following refined result:

Lemma 2. *Algorithm 1 with $\epsilon^{(t)} = \frac{1}{t^2}$-optimal prox computations at each time step and using $\boldsymbol{m}^{(t)} = \boldsymbol{\ell}^{(t-1)}$ yields*

$$\text{Reg}^{(T)} \leq \frac{\Omega+2}{\eta} + \eta \sum_{t=1}^{T} \|\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}\|^2 - \frac{1}{4\eta} \sum_{t=1}^{T} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|^2.$$

In particular, this satisfies the RVU property with parameters $\alpha = \frac{\Omega+2}{\eta}, \beta = \eta, \gamma = \frac{1}{4\eta}$.

We refer to Algorithm 1 instantiated with AFW (Algorithm 2 in the full version of the paper) as AFW-ROMD. Lemma 2 immediately enables us to state several results on Algorithm 1's ergodic convergence to equilibrium and characterize average regret in terms of LMO calls for AFW-ROMD.

Theorem 6. *An ϵ'-Nash equilibrium in any two-player zero-sum polyhedral game can be computed in $O(1/\epsilon')$ iterations of Algorithm 1. This corresponds to $O(\max_{i\in\{1,2\}} \frac{1}{\epsilon'} \frac{L_i D_i^2}{\delta_i^2} \log\left[\frac{L_i D_i}{\epsilon'}\right])$ LMO calls when using* AFW-ROMD.

Theorem 7. *An ϵ'-CCE in any N-player general-sum polyhedral game can be computed in $O(1/{\epsilon'}^{\frac{4}{3}})$ iterations of Algorithm 1. This corresponds to $O(\max_{i\in[N]} \frac{1}{{\epsilon'}^{\frac{4}{3}}} \frac{L_i D_i^2}{\delta_i^2} \log\left[\frac{L_i D_i}{\epsilon'}\right])$ LMO calls when using* AFW-ROMD.

4.1 Last-Iterate Convergence

Next, we obtain asymptotic last-iterate convergence to an approximate Nash equilibrium, when $\sum_{i=1}^{N} \text{Reg}_i^{(t)} \geq 0$ for any $t \in \mathbb{N}$, adapting a result from Anagnostides et al. (2022). A wide class of games, including two-player NFGs and EFGs, polymatrix zero-sum games, constant-sum polymatrix games, strategically zero-sum games, and polymatrix strategically zero-sum games satisfy this condition on social regret (Anagnostides et al. 2022); thus, our result holds for this class of games as well.

Theorem 8. *For any N-player general-sum polyhedral game, given $\epsilon \in (0, 1)$, let Player i employ the above framework with $\epsilon_i^{(t)} = \epsilon^2$ and $\boldsymbol{m}_i^{(t)} = \boldsymbol{\ell}_i^{(t-1)}$. Let $\eta_{\max} \leq \frac{1}{2\sqrt{2}(N-1)}$ where $\eta_{\max} = \max_{i\in[N]} \eta_i$ and suppose $\sum_{i=1}^{N} \text{Reg}_i^{(t)} \geq 0$ for any $t \in \mathbb{N}$. Define $\alpha_i = \left(\frac{1}{\eta_i} + \frac{2\Omega_i}{\eta_i}(L_i + N - 1) + 1\right)$. Then, after $T > \left\lceil \frac{8\eta_{\max}}{\epsilon^2} \sum_{i=1}^{N} \frac{(\Omega_i+2)}{\eta_i} \right\rceil$ iterations, there exists $\boldsymbol{x}^{(t)}$ with $t \in [T]$ which is an $\epsilon\Big(\max_{i\in[N]} \sqrt{2\eta_i}(\frac{2L_i D_i}{\eta_i} + 3) + \alpha_i\Big)$-Nash equilibrium.* AFW-ROMD *will yield an iterate that is an ϵ'-Nash equilibrium in $O\Big(\max_{j\in[N]} \Big\{ \frac{\eta_{\max}\alpha_j^2}{\epsilon'^2} \sum_{i=1}^{N} (\frac{\Omega_i+2}{\eta_i}) \frac{L_i D_i^2}{\delta_i^2} \log\left[\frac{L_i D_i \alpha_j}{\epsilon'}\right] \Big\}\Big)$ LMO calls when $\epsilon \leq \min_{i\in[N]} \frac{\epsilon'}{\alpha_i}$.*

In the two-player zero-sum case, we also obtain last-iterate linear-rate convergence to ϵ'-Nash equilibria when instantiated with the Euclidean regularizer, $\varphi_i(\boldsymbol{x}_i) = \frac{1}{2}\|\boldsymbol{x}_i\|_2^2$ for $i \in \{1, 2\}$, for any choice of ϵ'.

Theorem 9. *In any two-player zero-sum polyhedral game, both players employing Algorithm 1 with $\boldsymbol{m}_i^{(t)} = \boldsymbol{\ell}_i^{(t-1)}$, $\epsilon_i^{(t)} = \epsilon$, $\varphi_i(\boldsymbol{x}_i) = \frac{1}{2}\|\boldsymbol{x}_i\|_2^2$, and $\eta_i = \eta \leq \frac{1}{4}$ yields linear last-iterate convergence to a $\frac{(16+C_1)\epsilon + 32\max_{i\in\{1,2\}} \sqrt{2\eta\epsilon}(2L_i D_i + 3\eta)}{C_2}$-Nash equilibrium, where ν is a game-dependent constant associated with the SP-MS condition, $C_1 = 2(1 + \frac{4\eta^2\nu^2}{25})$, and $C_2 = \min(\frac{1}{2}, \frac{\eta^2\nu^2}{25})$:*

$$\text{dist}(\boldsymbol{z}^{(t)}, \mathcal{Z}^*)^2 \leq 2\left(1 + \frac{C_2}{4}\right)^{-t} \text{dist}(\boldsymbol{z}^{(1)}, \mathcal{Z}^*)^2 + \frac{(16+C_1)\epsilon + 32\max_{i\in\{1,2\}} \sqrt{2\eta\epsilon}(2L_i D_i + 3\eta)}{C_2}.$$

In the same setting ($\boldsymbol{m}_i^{(t)} = \boldsymbol{\ell}_i^{(t-1)}$, $\epsilon_i^{(t)} = \epsilon$, and $\eta_i = \eta \leq \frac{1}{4}$), if it is assumed that both players are applying AFW-ROMD*, then they can achieve linear last-iterate convergence to a $\frac{48+C_1}{C_2}\epsilon$-Nash equilibrium, with the same definitions for ν, C_1, C_2:*

$$\text{dist}(\boldsymbol{z}^{(t)}, \mathcal{Z}^*)^2 \leq 2\left(1 + \frac{C_2}{4}\right)^{-t} \text{dist}(\boldsymbol{z}^{(1)}, \mathcal{Z}^*)^2 + \frac{48 + C_1}{C_2}\epsilon.$$

AFW-ROMD *requires*

$$O\left(\max_{i\in\{1,2\}} \frac{\log \frac{2C_2+48+C_1}{C_2\epsilon'}}{\log \frac{4+C_2}{4}} \frac{L_i D_i^2}{\delta_i^2} \log\left[\frac{(2C_2+48+C_1) L_i D_i}{C_2\epsilon'}\right]\right)$$

LMO calls to compute an ϵ'-NE. Furthermore, the approximate solution it returns will have support of size $O\Big(\max_{i\in\{1,2\}} \frac{L_i D_i^2}{\delta_i^2} \log\left[\frac{L_i D_i (2C_2+48+C_1)}{C_2\epsilon'}\right]\Big)$.

5 Experimental Results and Discussion

We conduct experiments on standard EFG benchmarks to demonstrate the numerical performance of our algorithm relative to known algorithms from the literature. Details of games are provided in Appendix F. In addition to evaluating AFW-ROMD, we also consider its non-optimistic variant (taken by letting $\boldsymbol{m}^{(t)} = \mathbf{0}$ for all $t \in [T]$). We call this algorithm AFW-OMD since it corresponds to using AFW as an APO in

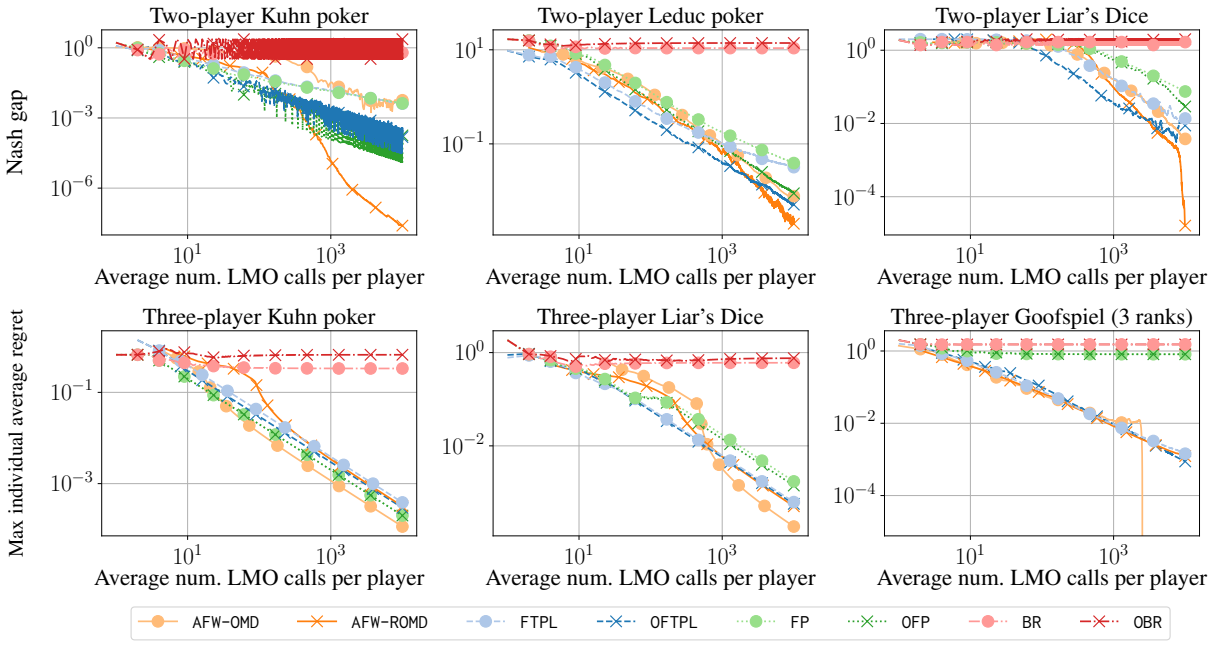

Figure 1: Convergence to equilibrium (Nash eq. and CCE) as a function of average LMO calls per player for AFW-OMD, AFW-ROMD, FTPL, OFTPL, FP, OFP, BR, and OBR.

vanilla OMD. We use the Euclidean regularizer, $\varphi_i(\boldsymbol{x}_i) = \frac{1}{2}\|\boldsymbol{x}_i\|_2^2$ for $i \in [N]$.

We compare against FTPL and OFTPL, fictitious play (FP) (Brown 1951) and best-response dynamics (BR), the latter two being unregularized variants of FTRL/FTPL and OMD, respectively. Finally, we also compare to *optimistic* versions of fictitious play (OFP) and best-response dynamics (OBR). We provide pseudocode for all of these algorithms in Appendix G; the pseudocode for these algorithms explicitly demonstrates that only one LMO call is required per iteration of these algorithms. Because FP and BR represent unregularized variants of FTRL/FTPL and OMD, they can be thought of as letting the stepsize be arbitrarily large for FTRL and OMD respectively (or letting the noise be zero for FTPL in the case of the former); we choose to depict their performance to evaluate the limiting behavior of FTRL/FTPL and OMD. We note that, unlike FTPL, OFTPL, and our algorithms, FP may converge at a *very* slow rate even in two-player zero-sum games (Daskalakis and Pan 2014), and BR may not converge at all.

For AFW-OMD, AFW-ROMD, FTPL, and OFTPL, we try $\eta \in 0.01 \cdot 2^{[14]}$, where η is the stepsize for our algorithms, while η is the noise used for FTPL and OFTPL. Additionally, we try uniform, linear, and quadratic iterate averaging for all algorithms, as well as last-iterate. Non-uniform averaging schemes are known to often outperform uniform averaging when solving BSPPs (Tammelin et al. 2015; Gao, Kroer, and Goldfarb 2021). Note that we demonstrate theoretical guarantees for last-iterate convergence of AFW-ROMD, whereas the other algorithms are not known to have such guarantees. Moreover, in the case of averaging, we examine the effects of applying *adaptive restarting* in Appendix H. Adaptive restarts are known to lead to linear convergence for polyhedral BSPPs for some algorithms, as they satisfy a *sharpness* property (Applegate et al. 2023; Fercoq 2022; Tseng 1995; Gilpin, Peña, and Sandholm 2012).

For our algorithms and (O)FTPL, we restrict the number of LMO calls per iteration to be in $\{1, 2, 3, 4, 5, 10, 20, 100, 200\}$. Note that the original presentations of OFTPL and FTPL set the termination criterion for an iteration of the algorithm based on the number of LMO calls ($4T$ and 1, respectively, for the convergence guarantees provided for each algorithm). On the other hand, the more natural termination criterion for our algorithm is the accuracy to which the approximate proximal call is to be computed. Nevertheless, we find that using the number of LMO calls as a termination criterion generally works best for our algorithms as well. Furthermore, we use warmstarting for our algorithm, which involves initializing the active set of AFW in the current iteration of our algorithms with the active set of AFW in the previous iteration. We provide complete pseudocode for adaptive restarting and various iterate averaging schemes in Appendix G. In the *Warmstart and Termination Criteria Ablation* subsection, we demonstrate that restricting the number of LMO calls per outer iteration and warmstarting generally leads to better performance for our algorithms. We conduct further ablation on the averaging scheme in Appendix H.

For each of the six algorithms, we use the choice of stepsize, number of LMO calls, and averaging, which generally leads to the best performance for each game. We provide additional graphs in Appendix H demonstrating that the performance of our algorithms relative to the others generally holds irrespective of the choice of the averaging. All of our experiments are run until the average number of LMO calls

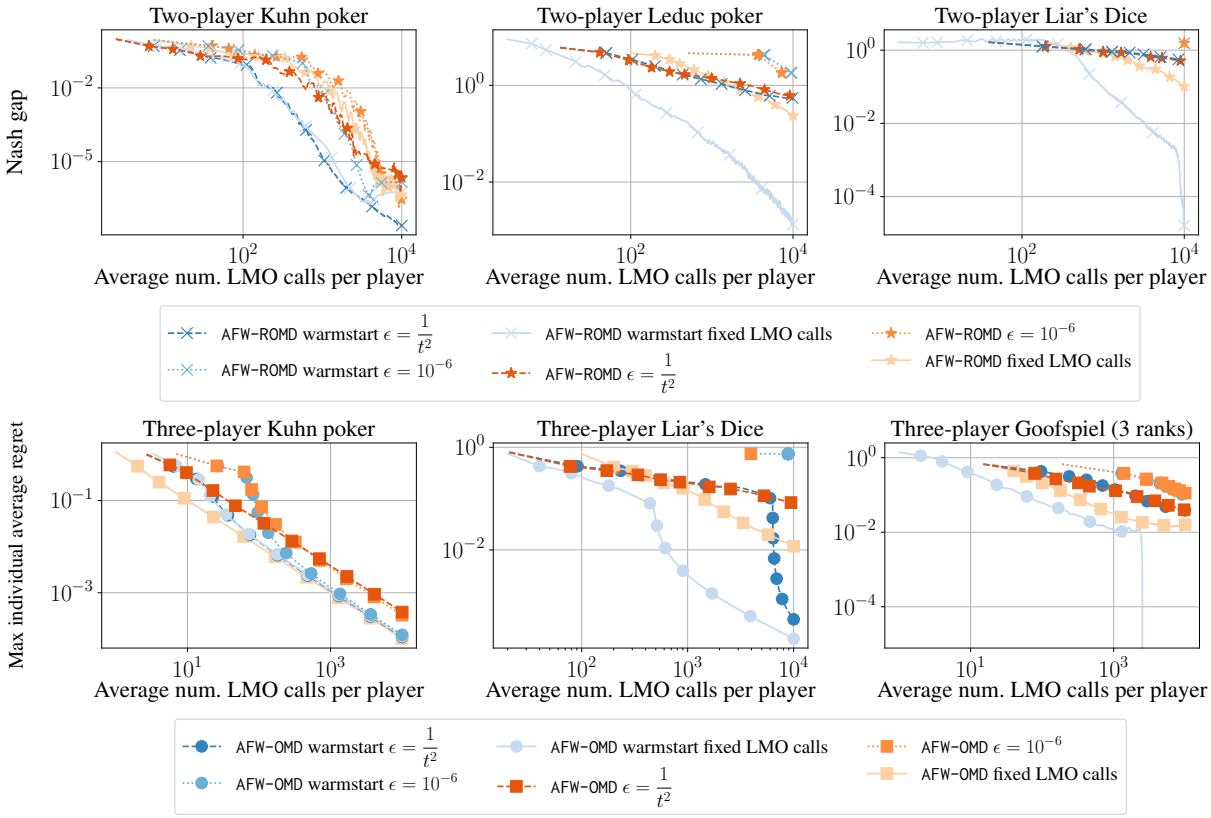

Figure 2: Convergence to equilibrium (Nash and CCE) as a function of average per-player LMO calls for AFW-ROMD (top row, Nash) and AFW-OMD (bottom row, CCE) for different APO termination criteria, and with and without warmstarting. The best averaging scheme is chosen for each warmstart and termination criterion setting.

for each player is 10^4.

We show the results of running our algorithms on two-player Kuhn poker, two-player Leduc poker, two-player Liar's Dice, three-player Kuhn poker, three-player Liar's Dice, and three-player Goofspiel (3 ranks) in Figure 1, seeking to compute Nash equilibria in the former three games and CCE in the latter three games. In the case of NE computation, AFW-ROMD outperforms existing algorithms in all three games. In Kuhn and Liar's Dice, we observe that (O)FTPL erratically reach small Nash gaps before returning to an iterate which has a high duality gap. This erratic behavior is because the last-iterate averaging is used for FTPL in Kuhn and for both OFTPL and FTPL in Liar's Dice. Despite (O)FTPL often achieving low duality gap in those two games, AFW-ROMD performs well relative to them (and the other algorithm). As noted above, we evaluate all of these algorithms for fixed choices of averaging schemes in Appendix H. Optimism is clearly helpful for all of the algorithms besides BR. In the case of CCE computation, we measure the maximum individual player's average regret since a bound of ϵ' on all players' average regrets corresponds to an ϵ'-CCE. Again, in all of the games, our algorithms are competitive with existing algorithms. However, this time it is our non-optimistic algorithm, AFW-OMD, that performs best. Interestingly, for the other sets of algorithms, the effect of optimism is minimal but does not hurt, whereas it hurts for our algorithm.

5.1 Warmstart and Termination Criteria Ablation

We also test the performance of our algorithm when using different choices of termination criteria for the approximate prox call and the choice of whether to warmstart. We only show the ablation for AFW-ROMD for NE computation and for AFW-OMD for CCE computation since these were our respective best-performing algorithms for each of these sets of experiments. The ablation of AFW-ROMD for CCE computation and AFW-OMD for NE computation is deferred to Appendix H. It can be seen in Figure 2, that in two-player Leduc poker and two-player Liar's Dice, using warmstarting and a fixed number of LMO calls per iteration leads to the best performance. In the multiplayer setting, it can be observed again that using warmstarting and a fixed number of LMO calls leads to best performance of our algorithm in three-player Liar's Dice and three-player Goofspiel (3 ranks). For both two-player and three-player Kuhn, the choice of using warmstarting and a fixed number of LMO calls is competitive.

Acknowledgements

Darshan Chakrabarti was supported by the National Science Foundation Graduate Research Fellowship Program under

award number DGE-2036197. Christian Kroer was supported by the Office of Naval Research awards N00014-22-1-2530 and N00014-23-1-2374, and the National Science Foundation awards IIS-2147361 and IIS-2238960.

References

Anagnostides, I.; Panageas, I.; Farina, G.; and Sandholm, T. 2022. On last-iterate convergence beyond zero-sum games. In *International Conference on Machine Learning (ICML)*.

Applegate, D.; Hinder, O.; Lu, H.; and Lubin, M. 2023. Faster first-order primal-dual methods for linear programming using restarts and sharpness. *Mathematical Programming*, 201(1-2): 133–184.

Bashiri, M. A.; and Zhang, X. 2017. Decomposition-invariant conditional gradient for general polytopes with line search. In *Neural Information Processing Systems (NIPS)*.

Braun, G.; Carderera, A.; Combettes, C. W.; Hassani, H.; Karbasi, A.; Mokhtari, A.; and Pokutta, S. 2022. Conditional gradient methods. *arXiv preprint arXiv:2211.14103*.

Brown, G. W. 1951. Iterative solution of games by fictitious play. *Activity Analysis of Production and Allocation*, 13(1): 374.

Daskalakis, C.; and Pan, Q. 2014. A counter-example to Karlin's strong conjecture for fictitious play. In *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*.

Fercoq, O. 2022. Quadratic error bound of the smoothed gap and the restarted averaged primal-dual hybrid gradient. *arXiv preprint arXiv:2206.03041*.

Frank, M.; and Wolfe, P. 1956. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2): 95–110.

Gao, Y.; Kroer, C.; and Goldfarb, D. 2021. Increasing Iterate Averaging for Solving Saddle-Point Problems. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Garber, D.; and Meshi, O. 2016. Linear-memory and decomposition-invariant linearly convergent conditional gradient algorithm for structured polytopes. In *Neural Information Processing Systems (NIPS)*.

Gidel, G.; Jebara, T.; and Lacoste-Julien, S. 2017. Frank-Wolfe algorithms for saddle point problems. In *Artificial Intelligence and Statistics (AISTATS)*.

Gilpin, A.; Peña, J.; and Sandholm, T. 2012. First-order algorithm with $\mathcal{O}(\ln(1/\epsilon))$ convergence for ϵ-equilibrium in two-person zero-sum games. *Mathematical Programming*, 133(1): 279–298.

Guélat, J.; and Marcotte, P. 1986. Some comments on Wolfe's 'away step'. *Mathematical Programming*, 35(1): 110–119.

Hazan, E.; and Kale, S. 2012. Projection-free online learning. In *International Coference on Machine Learning (ICML)*.

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

Joulani, P.; György, A.; and Szepesvári, C. 2017. A Modular Analysis of Adaptive (Non-)Convex Optimization: Optimism, Composite Objectives, and Variational Bounds. In *International Conference on Algorithmic Learning Theory (ALT)*.

Kalai, A.; and Vempala, S. 2005. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3): 291–307.

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

Lacoste-Julien, S.; and Jaggi, M. 2015. On the global linear convergence of Frank-Wolfe optimization variants. In *Neural Information Processing Systems (NIPS)*.

Lan, G.; and Zhou, Y. 2016. Conditional gradient sliding for convex optimization. *SIAM Journal on Optimization*, 26(2): 1379–1409.

Lanctot, M.; Zambaldi, V.; Gruslys, A.; Lazaridou, A.; Tuyls, K.; Pérolat, J.; Silver, D.; and Graepel, T. 2017. A unified game-theoretic approach to multiagent reinforcement learning. In *Neural Information Processing Systems (NIPS)*.

Malitsky, Y. 2015. Projected reflected gradient methods for monotone variational inequalities. *SIAM Journal on Optimization*, 25(1): 502–520.

McMahan, H. B.; Gordon, G. J.; and Blum, A. 2003. Planning in the presence of cost functions controlled by an adversary. In *International Conference on Machine Learning (ICML)*.

Orabona, F. 2019. A Modern Introduction to Online Learning. *arXiv preprint arXiv:1912.13213*.

Pena, J.; and Rodriguez, D. 2019. Polytope conditioning and linear convergence of the Frank–Wolfe algorithm. *Mathematics of Operations Research*, 44(1): 1–18.

Robinson, J. 1951. An Iterative Method of Solving a Game. *Annals of Mathematics*, 54(2): 296–301.

Romanovskii, I. 1962. Reduction of a game with complete memory to a matrix game. *Soviet Mathematics*, 3: 678–681.

Suggala, A.; and Netrapalli, P. 2020. Follow the perturbed leader: Optimism and fast parallel algorithms for smooth minimax games. In *Neural Information Processing Systems (NeurIPS)*.

Syrgkanis, V.; Agarwal, A.; Luo, H.; and Schapire, R. E. 2015. Fast convergence of regularized learning in games. In *Neural Information Processing Systems (NIPS)*.

Tammelin, O.; Burch, N.; Johanson, M.; and Bowling, M. 2015. Solving heads-up limit Texas Hold'em. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

Tseng, P. 1995. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2): 237–252.

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

Wei, C.-Y.; Lee, C.-W.; Zhang, M.; and Luo, H. 2021. Linear Last-Iterate Convergence in Constrained Saddle-Point Optimization. In *International Conference on Learning Representations (ICLR)*.

Wolfe, P. 1970. Convergence theory in nonlinear programming. *Integer and Nonlinear Programming*, 1–36.