Faster Game Solving via Predictive Blackwell Approachability: Connecting Regret Matching and Mirror Descent

Gabriele Farina,[1] **Christian Kroer,**[2] **Tuomas Sandholm**[1,3,4,5]
[1]Computer Science Department, Carnegie Mellon University, [2]IEOR Department, Columbia University
[3]Strategic Machine, Inc., [4]Strategy Robot, Inc., [5]Optimized Markets, Inc.
gfarina@cs.cmu.edu, christian.kroer@columbia.edu, sandholm@cs.cmu.edu

Abstract

Blackwell approachability is a framework for reasoning about repeated games with vector-valued payoffs. We introduce *predictive* Blackwell approachability, where an estimate of the next payoff vector is given, and the decision maker tries to achieve better performance based on the accuracy of that estimator. In order to derive algorithms that achieve predictive Blackwell approachability, we start by showing a powerful connection between four well-known algorithms. *Follow-the-regularized-leader (FTRL)* and *online mirror descent (OMD)* are the most prevalent regret minimizers in online convex optimization. In spite of this prevalence, the *regret matching (RM)* and *regret matching$^+$ (RM$^+$)* algorithms have been preferred in the practice of solving large-scale games (as the local regret minimizers within the counterfactual regret minimization framework). We show that RM and RM$^+$ are the algorithms that result from running FTRL and OMD, respectively, to select the halfspace to force at all times in the underlying Blackwell approachability game. By applying the predictive variants of FTRL or OMD to this connection, we obtain predictive Blackwell approachability algorithms, as well as predictive variants of RM and RM$^+$. In experiments across 18 common zero-sum extensive-form benchmark games, we show that predictive RM$^+$ coupled with counterfactual regret minimization converges vastly faster than the fastest prior algorithms (CFR$^+$, DCFR, LCFR) across all games but two of the poker games, sometimes by two or more orders of magnitude.

1 Introduction

Extensive-form games (EFGs) are the standard class of games that can be used to model sequential interaction, outcome uncertainty, and imperfect information. Operationalizing these models requires algorithms for computing game-theoretic equilibria. A recent success of EFGs is the use of Nash equilibrium for several recent poker AI milestones, such as essentially solving the game of limit Texas hold'em (Bowling et al. 2015), and beating top human poker pros in no-limit Texas hold'em with the *Libratus* AI (Brown and Sandholm 2017). A central component of all recent poker AIs has been a fast iterative method for computing approximate Nash equilibrium at scale. The leading approach is the *counterfactual regret minimization (CFR)* framework, where the problem of minimizing regret over a player's strategy space of an EFG is decomposed into a set of regret-minimization problems over probability simplexes (Zinkevich et al. 2007; Farina, Kroer, and Sandholm 2019c). Each simplex represents the probability over actions at a given decision point. The CFR setup can be combined with any regret minimizer for the simplexes. If both players in a zero-sum EFG repeatedly play each other using a CFR algorithm, the average strategies converge to a Nash equilibrium. Initially *regret matching* (RM) was the prevalent simplex regret minimizer used in CFR. Later, it was found that by alternating strategy updates between the players, taking linear averages of strategy iterates over time, and using a variation of RM called *regret-matching$^+$ (RM$^+$)* (Tammelin 2014) leads to significantly faster convergence in practice. This variation is called CFR$^+$. Both CFR and CFR$^+$ guarantee convergence to Nash equilibrium at a rate of $T^{-1/2}$. CFR$^+$ has been used in every milestone in developing poker AIs in the last decade (Bowling et al. 2015; Moravčík et al. 2017; Brown and Sandholm 2017, 2019b). This is in spite of the fact that its theoretical rate of convergence is the same as that of CFR with RM (Tammelin 2014; Farina, Kroer, and Sandholm 2019a; Burch, Moravcik, and Schmid 2019), and there exist algorithms which converge at a faster rate of T^{-1} (Hoda et al. 2010; Kroer et al. 2020; Farina, Kroer, and Sandholm 2019b). In spite of this theoretically-inferior convergence rate, CFR$^+$ has repeatedly performed favorably against T^{-1} methods for EFGs (Kroer, Farina, and Sandholm 2018b; Kroer et al. 2020; Farina, Kroer, and Sandholm 2019b; Gao, Kroer, and Goldfarb 2021). Similarly, the *follow-the-regularized-leader (FTRL)* and *online mirror descent (OMD)* regret minimizers, the two most prominent algorithms in online convex optimization, can be instantiated to have a better dependence on dimensionality than RM$^+$ and RM, yet RM$^+$ has been found to be superior (Brown, Kroer, and Sandholm 2017).

There has been some interest in connecting RM to the more prevalent (and more general) online convex optimization algorithms such as OMD and FTRL, as well as classical first-order methods. Waugh and Bagnell (2015) showed that RM is equivalent to Nesterov's dual averaging algorithm (which is an offline version of FTRL), though this equivalence requires specialized step sizes that are proven correct by invoking the correctness of RM itself. Burch (2018) stud-

ies RM and RM$^+$, and contrasts them with mirror descent and other prox-based methods.

We show a strong connection between RM, RM$^+$, and FTRL, OMD. This connection arises via *Blackwell approachability*, a framework for playing games with vector-valued payoffs, where the goal is to get the average payoff to approach some convex target set. Blackwell originally showed that this can be achieved by repeatedly *forcing* the payoffs to lie in a sequence of halfspaces containing the target set (Blackwell 1956). Our results are based on extending an equivalence between approachability and regret minimization (Abernethy, Bartlett, and Hazan 2011). We show that RM and RM$^+$ are the algorithms that result from running FTRL and OMD, respectively, to select the halfspace to force at all times in the underlying Blackwell approachability game. The equivalence holds for any constant step size. Thus, RM and RM$^+$, the two premier regret minimizers in EFG solving, turn out to follow exactly from the two most prevalent regret minimizers from online optimization theory. This is surprising for several reasons:

- RM$^+$ was originally discovered as a heuristic modification of RM in order to avoid accumulating large negative regrets. In contrast, OMD and FTRL were developed separately from each other.
- When applying FTRL and OMD directly to the strategy space of each player, Farina, Kroer, and Sandholm (2019b, 2020) found that FTRL seems to perform better than OMD, even when using stochastic gradients. This relationship is reversed here, as RM$^+$ is *vastly* faster numerically than RM.
- The dual averaging algorithm (whose simplest variant is an offline version of FTRL), was originally developed in order to have increasing weight put on more recent gradients, as opposed to OMD which has constant or decreasing weight (Nesterov 2009). Here this relationship is reversed: OMD (which we show has a close link to RM$^+$) thresholds away old negative regrets, whereas FTRL keeps them around. Thus OMD ends up being *more* reactive to recent gradients in our setting.
- FTRL and OMD both have a step-size parameter that needs to be set according to the magnitude of gradients, while RM and RM$^+$ are parameter free (which is a desirable feature from a practical perspective). To reconcile this seeming contradiction, we show that the step-size parameter does not affect which halfspaces are forced, so any choice of step size leads to RM and RM$^+$.

Leveraging our connection, we study the algorithms that result from applying predictive variants of FTRL and OMD to choosing which halfspace to force. By applying predictive OMD we get the first predictive variant of RM$^+$, that is, one that has regret that depends on how good the sequence of predicted regret vectors is (as a side note of their paper, Brown and Sandholm (2019a) also tried a heuristic for optimism/predictiveness by counting the last regret vector twice in RM$^+$, but this does not yield a predictive algorithm). We call our regret minimizer *predictive regret matching*$^+$ (PRM$^+$). We go on to instantiate CFR with PRM$^+$ using the two standard techniques—alternation and quadratic averaging—-and find that it often converges much faster than CFR$^+$ and every other prior CFR variant, sometimes by several orders of magnitude. We show this on a large suite of common benchmark EFGs. However, we find that on poker games (except shallow ones), *discounted CFR (DCFR)* (Brown and Sandholm 2019a) is the fastest. We conclude that our algorithm based on PRM$^+$ yields the new state-of-the-art convergence rate for the remaining games. Our results also highlight the need to test on EFGs other than poker, as our non-poker results invert the superiority of prior algorithms as compared to recent results on poker.

2 Online Linear Optimization, Regret Minimizers, and Predictions

At each time t, an oracle for the *online linear optimization (OLO)* problem supports the following two operations, in order: NEXTSTRATEGY returns a point $\boldsymbol{x}^t \in \mathcal{D} \subseteq \mathbb{R}^n$, and OBSERVELOSS receives a *loss vector* $\boldsymbol{\ell}^t$ that is meant to evaluate the strategy $\boldsymbol{x}^t$ that was last output. Specifically, the oracle incurs a loss equal to $\langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle$. The loss vector $\boldsymbol{\ell}^t$ can depend on all past strategies that were output by the oracle. The oracle operates *online* in the sense that each strategy $\boldsymbol{x}^t$ can depend only on the decision $\boldsymbol{x}^1, \dots, \boldsymbol{x}^{t-1}$ output in the past, as well as the loss vectors $\boldsymbol{\ell}^1, \dots, \boldsymbol{\ell}^{t-1}$ that were observed in the past. No information about the future losses $\boldsymbol{\ell}^t, \boldsymbol{\ell}^{t+1}, \dots$ is available to the oracle at time t. The objective of the oracle is to make sure the *regret*

$$R^T(\hat{\boldsymbol{x}}) \coloneqq \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle - \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \hat{\boldsymbol{x}} \rangle = \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t - \hat{\boldsymbol{x}} \rangle,$$

which measures the difference between the total loss incurred up to time T compared to always using the *fixed* strategy $\hat{\boldsymbol{x}}$, does not grow too fast as a function of time T. Oracles that guarantee that $R^T(\hat{\boldsymbol{x}})$ grow sublinearly in T in the worst case for all $\hat{\boldsymbol{x}} \in \mathcal{D}$ (no matter the sequence of losses $\boldsymbol{\ell}^1, \dots, \boldsymbol{\ell}^T$ observed) are called *regret minimizers*. While most theory about regret minimizers is developed under the assumption that the domain $\mathcal{D}$ is *convex* and *compact*, in this paper we will need to consider sets $\mathcal{D}$ that are convex and closed, but unbounded (hence, not compact).

Incorporating Predictions

A recent trend in online learning has been concerned with constructing oracles that can incorporate *predictions* of the next loss vector $\boldsymbol{\ell}^t$ in the decision making (Chiang et al. 2012; Rakhlin and Sridharan 2013a,b). Specifically, a *predictive* oracle differs from a regular (that is, non-predictive) oracle for OLO in that the NEXTSTRATEGY function receives a *prediction* $\boldsymbol{m}^t \in \mathbb{R}^n$ of the next loss $\boldsymbol{\ell}^t$ at all times t. Conceptually, a "good" predictive regret minimizer should guarantee a superior regret bound than a non-predictive regret minimizer if $\boldsymbol{m}^t \approx \boldsymbol{\ell}^t$ at all times t. Algorithms exist that can guarantee this. For instance, it is always possible to construct an oracle that guarantees that $R^T = O(1+\sum_{t=1}^{T} \|\boldsymbol{\ell}^t - \boldsymbol{m}^t\|^2)$, which implies that the regret stays constant when $\boldsymbol{m}^t$ is clairvoyant. In fact, even stronger regret bounds can be attained: for example, Syrgkanis et al.

Algorithm 1: (Predictive) FTRL

$\boldsymbol{L}^0 \leftarrow \mathbf{0} \in \mathbb{R}^n$

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

▷ Set $\boldsymbol{m}^t = \mathbf{0}$ for non-predictive version

return $\arg\min_{\hat{\boldsymbol{x}} \in \mathcal{D}} \left\{ \langle \boldsymbol{L}^{t-1} + \boldsymbol{m}^t, \hat{\boldsymbol{x}} \rangle + \frac{1}{\eta} \varphi(\hat{\boldsymbol{x}}) \right\}$

function OBSERVELOSS($\boldsymbol{\ell}^t$)

$\boldsymbol{L}^t \leftarrow \boldsymbol{L}^{t-1} + \boldsymbol{\ell}^t$

Algorithm 2: (Predictive) OMD

$\boldsymbol{z}^0 \in \mathcal{D}$ such that $\nabla \varphi(\boldsymbol{z}^0) = \mathbf{0}$

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

▷ Set $\boldsymbol{m}^t = \mathbf{0}$ for non-predictive version

return $\arg\min_{\hat{\boldsymbol{x}} \in \mathcal{D}} \left\{ \langle \boldsymbol{m}^t, \hat{\boldsymbol{x}} \rangle + \frac{1}{\eta} D_\varphi(\hat{\boldsymbol{x}} \,\|\, \boldsymbol{z}^{t-1}) \right\}$

function OBSERVELOSS($\boldsymbol{\ell}^t$)

$\boldsymbol{z}^t \leftarrow \arg\min_{\hat{\boldsymbol{z}} \in \mathcal{D}} \left\{ \langle \boldsymbol{\ell}^t, \hat{\boldsymbol{z}} \rangle + \frac{1}{\eta} D_\varphi(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{t-1}) \right\}$

(2015) show that the sharper *Regret bounded by Variation in Utilities (RVU)* condition can be attained, while Farina et al. (2019a) focus on *stable-predictivity*.

FTRL, OMD, and their Predictive Variants

Follow-the-regularized-leader (FTRL) (Shalev-Shwartz and Singer 2007) and *online mirror descent (OMD)* are the two best known oracles for the online linear optimization problem. Their *predictive* variants are relatively new and can be traced back to the works by Rakhlin and Sridharan (2013a) and Syrgkanis et al. (2015). Since the original FTRL and OMD algorithms correspond to predictive FTRL and predictive OMD when the prediction $\boldsymbol{m}^t$ is set to the $\mathbf{0}$ vector at all t, the implementation of FTRL in Algorithm 1 and OMD in Algorithm 2 captures both algorithms. In both algorithm, $\eta > 0$ is an arbitrary step size parameter, $\mathcal{D} \subseteq \mathbb{R}^n$ is a convex and closed set, and $\varphi : \mathcal{D} \to \mathbb{R}_{\geq 0}$ is a 1-strongly convex differentiable regularizer (with respect to some norm $\|\cdot\|$). The symbol $D_\varphi(\,\|\,)$ used in OMD denotes the *Bregman divergence* associated with φ, defined as $D_\varphi(\boldsymbol{x} \,\|\, \boldsymbol{c}) \coloneqq \varphi(\boldsymbol{x}) - \varphi(\boldsymbol{c}) - \langle \nabla \varphi(\boldsymbol{c}), \boldsymbol{x} - \boldsymbol{c} \rangle$ for all $\boldsymbol{x}, \boldsymbol{c} \in \mathcal{D}$.

We state regret guarantees for (predictive) FTRL and (predictive) OMD in Proposition 1. Our statements are slightly more general than those by Syrgkanis et al. (2015), in that we (i) do not assume that the domain is a simplex, and (ii) do not use quantities that might be unbounded in non-compact domains $\mathcal{D}$. A proof of the regret bounds is in Appendix A of the full version of the paper[1] for FTRL and Appendix B for OMD.

Proposition 1. *At all times T, the regret cumulated by (predictive) FTRL (Algorithm 1) and (predictive) OMD (Algorithm 2) compared to* any *strategy $\hat{\boldsymbol{x}} \in \mathcal{D}$ is bounded as*

$$R^T(\hat{\boldsymbol{x}}) \leq \frac{\varphi(\hat{\boldsymbol{x}})}{\eta} + \eta \sum_{t=1}^{T} \|\boldsymbol{\ell}^t - \boldsymbol{m}^t\|_*^2 - \frac{1}{c\eta} \sum_{t=1}^{T-1} \|\boldsymbol{x}^{t+1} - \boldsymbol{x}^t\|^2,$$

where $c = 4$ for FTRL and $c = 8$ for OMD, and where $\|\cdot\|_*$ denotes the dual of the norm $\|\cdot\|$ with respect to which φ is 1*-strongly convex.*

Proposition 1 implies that, by appropriately setting the step size parameter (for example, $\eta = T^{-1/2}$), (predictive) FTRL and (predictive) OMD guarantee $R^T(\hat{\boldsymbol{x}}) = O(T^{1/2})$ for all $\hat{\boldsymbol{x}}$. Hence, (predictive) FTRL and (predictive) OMD are regret minimizers.

[1]The full version of this paper is at arxiv.org/abs/2007.14358.

3 Blackwell Approachability

Blackwell approachability (Blackwell 1956) generalizes the problem of playing a repeated two-player game to games whose utilites are vectors instead of scalars. In a Blackwell approachability game, at all times t, two players interact in this order: first, Player 1 selects an action $\boldsymbol{x}^t \in \mathcal{X}$; then, Player 2 selects an action $\boldsymbol{y}^t \in \mathcal{Y}$; finally, Player 1 incurs the vector-valued payoff $\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \in \mathbb{R}^d$, where $\boldsymbol{u}$ is a biaffine function. The sets $\mathcal{X}, \mathcal{Y}$ of player actions are assumed to be compact convex sets. Player 1's objective is to guarantee that the average payoff converges to some desired closed convex *target set* $S \subseteq \mathbb{R}^d$. Formally, given target set $S \subseteq \mathbb{R}^d$, Player 1's goal is to pick actions $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots \in \mathcal{X}$ such that no matter the actions $\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots \in \mathcal{Y}$ played by Player 2,

$$\min_{\hat{\boldsymbol{s}} \in S} \left\| \hat{\boldsymbol{s}} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \right\|_2 \to 0 \quad \text{as} \quad T \to \infty. \tag{1}$$

A central concept in the theory of Blackwell approachability is the following.

Definition 1 (Approachable halfspace, forcing function)**.** *Let $(\mathcal{X}, \mathcal{Y}, \boldsymbol{u}(\cdot,\cdot), S)$ be a Blackwell approachability game as described above and let $H \subseteq \mathbb{R}^d$ be a halfspace, that is, a set of the form $H = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{a}^\top \boldsymbol{x} \leq b\}$ for some $\boldsymbol{a} \in \mathbb{R}^d, b \in \mathbb{R}$. The halfspace H is said to be* forceable *if there exists a strategy of Player 1 that guarantees that the payoff is in H no matter the actions played by Player 2. In symbols, H is forceable if there exists $\boldsymbol{x}^* \in \mathcal{X}$ such that for all $\boldsymbol{y} \in \mathcal{Y}$, $\boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{y}) \in H$. When this is the case, we call action $\boldsymbol{x}^*$ a* forcing action *for H.*

Blackwell's *approachability theorem* (Blackwell 1956) states that goal (1) can be attained if and only if all halfspaces $H \supseteq S$ are forceable. Blackwell approachability has a number of applications and connections to other problems in the online learning and game theory literature (e.g., (Blackwell 1954; Foster 1999; Hart and Mas-Colell 2000)).

In this paper we leverage the Blackwell approachability formalism to draw new connections between FTRL and OMD with RM and RM$^+$, respectively. We also introduce predictive Blackwell approachability, and show that it can be used to develop new state-of-the-art algorithms for simplex domains and imperfect-information extensive-form zero-sum games.

Algorithm 3: From OLO to (predictive) approachability

Data: $\mathcal{D} \subseteq \mathbb{R}^n$ convex and closed, s.t. $\mathcal{K} := C^\circ \cap \mathbb{B}_2^n \subseteq \mathcal{D} \subseteq C^\circ$
$\mathcal{L}$ online linear optimization algorithm for domain $\mathcal{D}$

function NEXTSTRATEGY($\boldsymbol{v}^t$)
  ▷ Set $\boldsymbol{v}^t = \mathbf{0}$ for non-predictive version
  $\boldsymbol{\theta}^t \leftarrow \mathcal{L}$.NEXTSTRATEGY($-\boldsymbol{v}^t$)
  return $\boldsymbol{x}^t$ *forcing action for* $H^t := \{\boldsymbol{x} : \langle \boldsymbol{\theta}^t, \boldsymbol{x} \rangle \leq 0\}$

function RECEIVEPAYOFF($\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)$)
  $\mathcal{L}$.OBSERVELOSS($-\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)$)

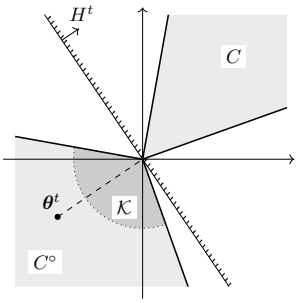


Figure 1: Pictorial depiction of Algorithm 3's inner working: at all times t, the algorithm plays a forcing action for the halfspace H^t induced by the last decision output by $\mathcal{L}$.

4 From Online Linear Optimization to Blackwell Approachability

Abernethy, Bartlett, and Hazan (2011) showed that it is always possible to convert a regret minimizer into an algorithm for a Blackwell approachability game (that is, an algorithm that chooses actions $\boldsymbol{x}^t$ at all times t in such a way that goal (1) holds no matter the actions $\boldsymbol{y}^1, \boldsymbol{y}^2, \dots$ played by the opponent).[2]

In this section, we slightly extend their constructive proof by allowing more flexibility in the choice of the domain of the regret minimizer. This extra flexibility will be needed to show that RM and RM$^+$ can be obtained directly from FTRL and OMD, respectively.

We start from the case where the target set in the Blackwell approachability game is a closed convex cone $C \subseteq \mathbb{R}^n$. As Proposition 2 shows, Algorithm 3 provides a way of playing the Blackwell approachability game that guarantees that (1) is satisfied (the proof is in Appendix C in the full version of the paper). In broad strokes, Algorithm 3 works as follows (see also Figure 1): the regret minimizer has as its decision space the polar cone to C (or a subset thereof), and its decision is used as the normal vector in choosing a halfspace to force. At time t, the algorithm plays a forcing action $\boldsymbol{x}^t$ for the halfspace H_t induced by the last decision $\boldsymbol{\theta}^t$ output by the OLO oracle $\mathcal{L}$. Then, $\mathcal{L}$ incurs the loss $-\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)$, where $\boldsymbol{u}$ is the payoff function of the Blackwell approachability game.

Proposition 2. *Let $(\mathcal{X}, \mathcal{Y}, \boldsymbol{u}(\cdot,\cdot), C)$ be an approachability game, where $C \subseteq \mathbb{R}^n$ is a closed convex cone, such that each halfspace $H \supseteq C$ is approachable (Definition 1). Let $\mathcal{K} := C^\circ \cap \mathbb{B}_2^n$, where $C^\circ = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 0 \; \forall \boldsymbol{y} \in C\}$ denotes the* polar cone *to C and $\mathbb{B}_2^n := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2 \leq 1\}$ is the unit ball. Finally, let $\mathcal{L}$ be an oracle for the OLO problem (for example, the FTRL or OMD algorithm) whose domain of decisions is any closed convex set $\mathcal{D}$, such that $\mathcal{K} \subseteq \mathcal{D} \subseteq C^\circ$. Then, at all times T, the distance between the average payoff cumulated by Algorithm 3 and the target cone C is upper bounded as*

$$\min_{\hat{\boldsymbol{s}} \in C} \left\| \hat{\boldsymbol{s}} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \right\|_2 \leq \frac{1}{T} \max_{\hat{\boldsymbol{x}} \in \mathcal{K}} R_{\mathcal{L}}^T(\hat{\boldsymbol{x}}),$$

where $R_{\mathcal{L}}^T(\hat{\boldsymbol{x}})$ is the regret cumulated by $\mathcal{L}$ up to time T compared to always playing $\hat{\boldsymbol{x}} \in \mathcal{K}$.

As $\mathcal{K}$ is compact, by virtue of $\mathcal{L}$ being a regret minimizer, $1/T \cdot \max_{\hat{\boldsymbol{x}} \in \mathcal{K}} R^T(\hat{\boldsymbol{x}}) \to 0$ as $T \to \infty$, Algorithm 3 satisfies the Blackwell approachability goal (1). The fact that Proposition 2 applies only to conic target sets does not limit its applicability. Indeed, Abernethy, Bartlett, and Hazan (2011) showed that any Blackwell approachability game with a non-conic target set can be efficiently transformed to another one with a conic target set. In this paper, we only need to focus on conic target sets.

The construction by Abernethy, Bartlett, and Hazan (2011) coincides with Proposition 2 in the special case where the domain $\mathcal{D}$ is set to $\mathcal{D} = \mathcal{K}$. In the next section, we will need our added flexibility in the choice of $\mathcal{D}$: in order to establish the connection between RM$^+$ and OMD, it is necessary to set $\mathcal{D} = C^\circ \neq \mathcal{K}$.

5 Connecting FTRL, OMD with RM, RM$^+$

Constructing a regret minimizer for a simplex domain $\Delta^n := \{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} : \|\boldsymbol{x}\|_1 = 1\}$ can be reduced to constructing an algorithm for a particular Blackwell approachability game $\Gamma := (\Delta^n, \mathbb{R}^n, \boldsymbol{u}(\cdot,\cdot), \mathbb{R}^n_{\leq 0})$ that we now describe (Hart and Mas-Colell 2000). For all $i \in \{1, \dots, n\}$, the i-th component of the vector-valued payoff function $\boldsymbol{u}$ measures the change in regret incurred at time t, compared to always playing the i-th vertex $\boldsymbol{e}_i$ of the simplex. Formally, $\boldsymbol{u} : \Delta^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) = \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1} - \boldsymbol{\ell}^t, \tag{2}$$

where $\mathbf{1}$ is the n-dimensional vector whose components are all 1. It is known that Γ is such that the halfspace $H_{\boldsymbol{a}} := \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{a} \rangle \leq 0\} \supseteq \mathbb{R}^n_{\leq 0}$ is forceable (Definition 1) for all $\boldsymbol{a} \in \mathbb{R}^n_{\geq 0}$. A forcing action for $H_{\boldsymbol{a}}$ is given by $\boldsymbol{g}(\boldsymbol{a}) := \boldsymbol{a}/\|\boldsymbol{a}\|_1 \in \Delta^n$ when $\boldsymbol{a} \neq \mathbf{0}$; when $\boldsymbol{a} = \mathbf{0}$, any $\boldsymbol{x} \in \Delta^n$ is a forcing action. The following is known.

Lemma 1. *The regret $R^T(\hat{\boldsymbol{x}}) = \frac{1}{T} \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t - \hat{\boldsymbol{x}} \rangle$ cumulated up to any time T by the decisions $\boldsymbol{x}^1, \dots, \boldsymbol{x}^T \in \Delta^n$ compared to any $\hat{\boldsymbol{x}} \in \Delta^n$ is related to the distance of the average Blackwell payoff from the target cone $\mathbb{R}^n_{\leq 0}$ as*

$$\frac{1}{T} R^T(\hat{\boldsymbol{x}}) \leq \min_{\hat{\boldsymbol{s}} \in \mathbb{R}^n_{\leq 0}} \left\| \hat{\boldsymbol{s}} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) \right\|_2. \tag{3}$$

[2] Gordon's Lagrangian Hedging (Gordon 2005, 2007) partially overlaps with the construction by Abernethy, Bartlett, and Hazan (2011). We did not investigate to what extent the *predictive* point of view we adopted in the paper could apply to Gordon's result.

Algorithm 4: (Predictive) regret matching

1 $\boldsymbol{r}^0 \leftarrow \mathbf{0} \in \mathbb{R}^n,\ \ \boldsymbol{x}^0 \leftarrow \mathbf{1}/n \in \Delta^n$

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

▷ Set $\boldsymbol{m}^t = \mathbf{0}$ for non-predictive version

3 $\quad \boldsymbol{\theta}^t \leftarrow [\boldsymbol{r}^{t-1} + \langle \boldsymbol{m}^t, \boldsymbol{x}^{t-1} \rangle \mathbf{1} - \boldsymbol{m}^t]^+$

4 $\quad$ **if** $\boldsymbol{\theta}^t \neq \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \,/\, \|\boldsymbol{\theta}^t\|_1$

5 $\quad$ **else** $\quad$ **return** $\boldsymbol{x}^t \leftarrow$ arbitrary point in Δ^n

6 **function** OBSERVELOSS($\boldsymbol{\ell}^t$)

7 $\quad \boldsymbol{r}^t \leftarrow \boldsymbol{r}^{t-1} + \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1} - \boldsymbol{\ell}^t$

So, a strategy for the Blackwell approachability game Γ is a regret-minimizing strategy for the simplex domain Δ^n.

When the approachability game Γ is solved by means of the constructive proof of Blackwell's approachability theorem (Blackwell 1956), one recovers a particular regret minimizer for the domain Δ^n known as the *regret matching (RM)* algorithm (Hart and Mas-Colell 2000). The same cannot be said for the closely related RM$^+$ algorithm (Tammelin 2014), which converges significantly faster in practice than RM, as has been reported many times.

We now uncover deep and surprising connections between RM, RM$^+$ and the OLO algorithms FTRL, OMD by solving Γ using Algorithm 3. Let $\mathcal{L}_\eta^{\text{ftrl}}$ be the FTRL algorithm instantiated over the conic domain $\mathcal{D} = \mathbb{R}^n_{\geq 0}$ with the 1-strongly convex regularizer $\varphi(\boldsymbol{x}) = 1/2 \, \|\boldsymbol{x}\|_2^2$ and an arbitrary step size parameter η. Similarly, let $\mathcal{L}_\eta^{\text{omd}}$ be the OMD algorithm instantiated over the same domain $\mathcal{D} = \mathbb{R}^n_{\geq 0}$ with the same convex regularizer $\varphi(\boldsymbol{x}) = 1/2 \, \|\boldsymbol{x}\|_2^2$. Since $\mathbb{R}^n_{\geq 0} = (\mathbb{R}^n_{\leq 0})^\circ$, $\mathcal{D}$ satisfies the requirements of Proposition 2. So, $\mathcal{L}_\eta^{\text{ftrl}}$ and $\mathcal{L}_\eta^{\text{omd}}$ can be plugged into Algorithm 3 to compute a strategy for the Blackwell approachability game Γ. When that is done, the following can be shown (all proofs for this section are in Appendix D in the full version of the paper).

Theorem 1 (FTRL reduces to RM)**.** *For all $\eta > 0$, when Algorithm 3 is set up with $\mathcal{D} = \mathbb{R}^n_{\geq 0}$ and regret minimizer $\mathcal{L}_\eta^{\text{ftrl}}$ to play Γ, it produces the same iterates as the RM algorithm.*

Theorem 2 (OMD reduces to RM$^+$)**.** *For all $\eta > 0$, when Algorithm 3 is set up with $\mathcal{D} = \mathbb{R}^n_{\geq 0}$ and regret minimizer $\mathcal{L}_\eta^{\text{omd}}$ to play Γ, it produces the same iterates as the RM$^+$ algorithm.*

Pseudocode for RM and RM$^+$ is given in Algorithms 4 and 5 (when $\boldsymbol{m}^t = \mathbf{0}$). In hindsight, the equivalence between RM and RM$^+$ with FTRL and OMD is clear. The computation of $\boldsymbol{\theta}^t$ on Line 3 in both PRM and PRM$^+$ corresponds to the closed-form solution for the minimization problems of Line 4 in FTRL and Line 3 in OMD, respectively, in accordance with Line 2 of Algorithm 3. Next, Lines 4 and 5 in both PRM and PRM$^+$ compute the forcing action required in Line 3 of Algorithm 3 using the function $\boldsymbol{g}$ defined above. Finally, in accordance with Line 6 of Algorithm 3, Line 7 of PRM corresponds to Line 6 of FTRL, and Line 7 of PRM$^+$ to Line 5 of OMD.

Algorithm 5: (Predictive) regret matching$^+$

1 $\boldsymbol{z}^0 \leftarrow \mathbf{0} \in \mathbb{R}^n,\ \ \boldsymbol{x}^0 \leftarrow \mathbf{1}/n \in \Delta^n$

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

▷ Set $\boldsymbol{m}^t = \mathbf{0}$ for non-predictive version

3 $\quad \boldsymbol{\theta}^t \leftarrow [\boldsymbol{z}^{t-1} + \langle \boldsymbol{m}^t, \boldsymbol{x}^{t-1} \rangle \mathbf{1} - \boldsymbol{m}^t]^+$

4 $\quad$ **if** $\boldsymbol{\theta}^t \neq \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \,/\, \|\boldsymbol{\theta}^t\|_1$

5 $\quad$ **else** $\quad$ **return** $\boldsymbol{x}^t \leftarrow$ arbitrary point in Δ^n

6 **function** OBSERVELOSS($\boldsymbol{\ell}^t$)

7 $\quad \boldsymbol{z}^t \leftarrow [\boldsymbol{z}^{t-1} + \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1} - \boldsymbol{\ell}^t]^+$

6 Predictive Blackwell Approachability, and Predictive RM and RM$^+$

It is natural to wonder whether it is possible to devise an algorithm for Blackwell approachability games that is able to guarantee faster convergence to the target set when good predictions of the next vector payoff are available. We call this setup *predictive Blackwell approachability*. We answer the question in the positive by leveraging Proposition 2. Since the loss incurred by the regret minimizer is $\boldsymbol{\ell}^t := -\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)$ (Line 5 in Algorithm 3), any prediction $\boldsymbol{v}^t$ of the payoff $\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)$ is naturally a prediction about the next loss incurred by the underlying regret minimizer $\mathcal{L}$ used in Algorithm 3. Hence, as long as the prediction is propagated as in Line 2 in Algorithm 3, Proposition 2 holds verbatim. In particular, we prove the following. All proofs for this section are in Appendix E in the full version of the paper.

Proposition 3. *Let $(\mathcal{X}, \mathcal{Y}, \boldsymbol{u}(\cdot,\cdot), S)$ be a Blackwell approachability game, where every halfspace $H \supseteq S$ is approachable (Definition 1). For all T, given predictions $\boldsymbol{v}^t$ of the payoff vectors, there exist algorithms for playing the game (that is, pick $\boldsymbol{x}^t \in \mathcal{X}$ at all t) that guarantee*

$$\min_{\hat{\boldsymbol{s}} \in S} \left\| \hat{\boldsymbol{s}} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \right\|_2 \leq \frac{1}{\sqrt{T}} \left(1 + \frac{2}{T} \sum_{t=1}^{T} \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) - \boldsymbol{v}^t\|_2^2 \right).$$

We now focus on how predictive Blackwell approachability ties into our discussion of RM and RM$^+$. In Section 5 we showed that when Algorithm 3 is used in conjunction with FTRL and OMD on the Blackwell approachability game Γ of Section 5, the iterates coincide with those of RM and RM$^+$, respectively. In the rest of this section we investigate the use of *predictive* FTRL and *predictive* OMD in that framework. Specifically, we use predictive FTRL and predititive OMD as the regret minimizers to solve the Blackwell approachability game introduced in Section 5, and coin the resulting predictive regret minimization algorithms for simplex domains *predictive regret matching (PRM)* and *predictive regret matching$^+$ (PRM$^+$)*, respectively. Ideally, starting from the prediction $\boldsymbol{m}^t$ of the next loss, we would want the prediction $\boldsymbol{v}^t$ of the next utility in the equivalent Blackwell game Γ (Section 5) to be $\boldsymbol{v}^t = \langle \boldsymbol{m}^t, \boldsymbol{x}^t \rangle \mathbf{1} - \boldsymbol{m}^t$ to maintain symmetry with (2). However, $\boldsymbol{v}^t$ is computed before $\boldsymbol{x}^t$ is computed, and $\boldsymbol{x}^t$ depends on $\boldsymbol{v}^t$, so the previous expression requires the computation of a fixed point. To sidestep this issue, we let

$$\boldsymbol{v}^t := \langle \boldsymbol{m}^t, \boldsymbol{x}^{t-1} \rangle \mathbf{1} - \boldsymbol{m}^t$$

instead. We give pseudocode for PRM and PRM$^+$ as Algorithms 4 and 5. In the rest of this section, we discuss formal guarantees for PRM and PRM$^+$.

Theorem 3 (Correctness of PRM, PRM$^+$)**.** *Let $\mathcal{L}_\eta^{\text{ftrl}^*}$ and $\mathcal{L}_\eta^{\text{omd}^*}$ denote the predictive FTRL and predictive OMD algorithms instantiated with the same choice of regularizer and domain as in Section 5, and predictions $\boldsymbol{v}^t$ as defined above for the Blackwell approachability game Γ. For all $\eta > 0$, when Algorithm 3 is set up with $\mathcal{D} = \mathbb{R}^n_{\geq 0}$, the regret minimizer $\mathcal{L}_\eta^{\text{ftrl}^*}$ (resp., $\mathcal{L}_\eta^{\text{omd}^*}$) to play Γ, it produces the same iterates as the PRM (resp., PRM$^+$) algorithm. Furthermore, PRM and PRM$^+$ are regret minimizer for the domain Δ^n, and at all times T satisfy the regret bound*

$$R^T(\hat{\boldsymbol{x}}) \leq \sqrt{2}\left(\sum_{t=1}^{T} \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) - \boldsymbol{v}^t\|_2^2\right)^{1/2}.$$

At a high level, the main insight behind the regret bound of Theorem 3 is to combine Proposition 2 with the guarantees of predictive FTRL and predictive OMD (Proposition 1). In particular, combining (3) with Proposition 2, we find that the regret R^T cumulated by the strategies $\boldsymbol{x}^1, \dots, \boldsymbol{x}^T$ produced up to time T by PRM and PRM$^+$ satisfies

$$\frac{1}{T}\max_{\hat{\boldsymbol{x}} \in \Delta^n} R^T(\hat{\boldsymbol{x}}) \leq \frac{1}{T}\max_{\hat{\boldsymbol{x}} \in \mathbb{R}^n_{\geq 0} \cap \mathbb{B}_2^n} R^T_{\mathcal{L}}(\hat{\boldsymbol{x}}), \tag{4}$$

where $\mathcal{L} = \mathcal{L}_\eta^{\text{ftrl}^*}$ for PRM and $\mathcal{L} = \mathcal{L}_\eta^{\text{omd}^*}$ for PRM$^+$. Since the domain of the maximization on the right hand side is a subset of the domain $\mathcal{D} = \mathbb{R}^n_{\geq 0}$ of $\mathcal{L}$, the bound in Proposition 1 holds, and in particular

$$\begin{aligned}\max_{\hat{\boldsymbol{x}} \in \Delta^n} R^T(\hat{\boldsymbol{x}}) &\leq \max_{\hat{\boldsymbol{x}} \in \mathbb{R}^n_{\geq 0} \cap \mathbb{B}_2^n} \left\{\frac{\|\hat{\boldsymbol{x}}\|_2^2}{2\eta} + \eta \sum_{t=1}^{T} \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) - \boldsymbol{v}^t\|_2^2\right\} \\ &\leq \left(\frac{1}{2\eta} + \eta \sum_{t=1}^{T} \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) - \boldsymbol{v}^t\|_2^2\right), \end{aligned} \tag{5}$$

where in the first inequality we used the fact that $\varphi(\hat{\boldsymbol{x}}) = \|\hat{\boldsymbol{x}}\|_2^2/2$ by construction and in the second inequality we used the definition of unit ball $\mathbb{B}_2^n$. Finally, using the fact that the iterates produced by PRM and PRM$^+$ do not depend on the chosen step size $\eta > 0$ (first part of Theorem 3), we conclude that (5) must hold true for any $\eta > 0$, and so in particular also the $\eta > 0$ that minimizes the right hand side:

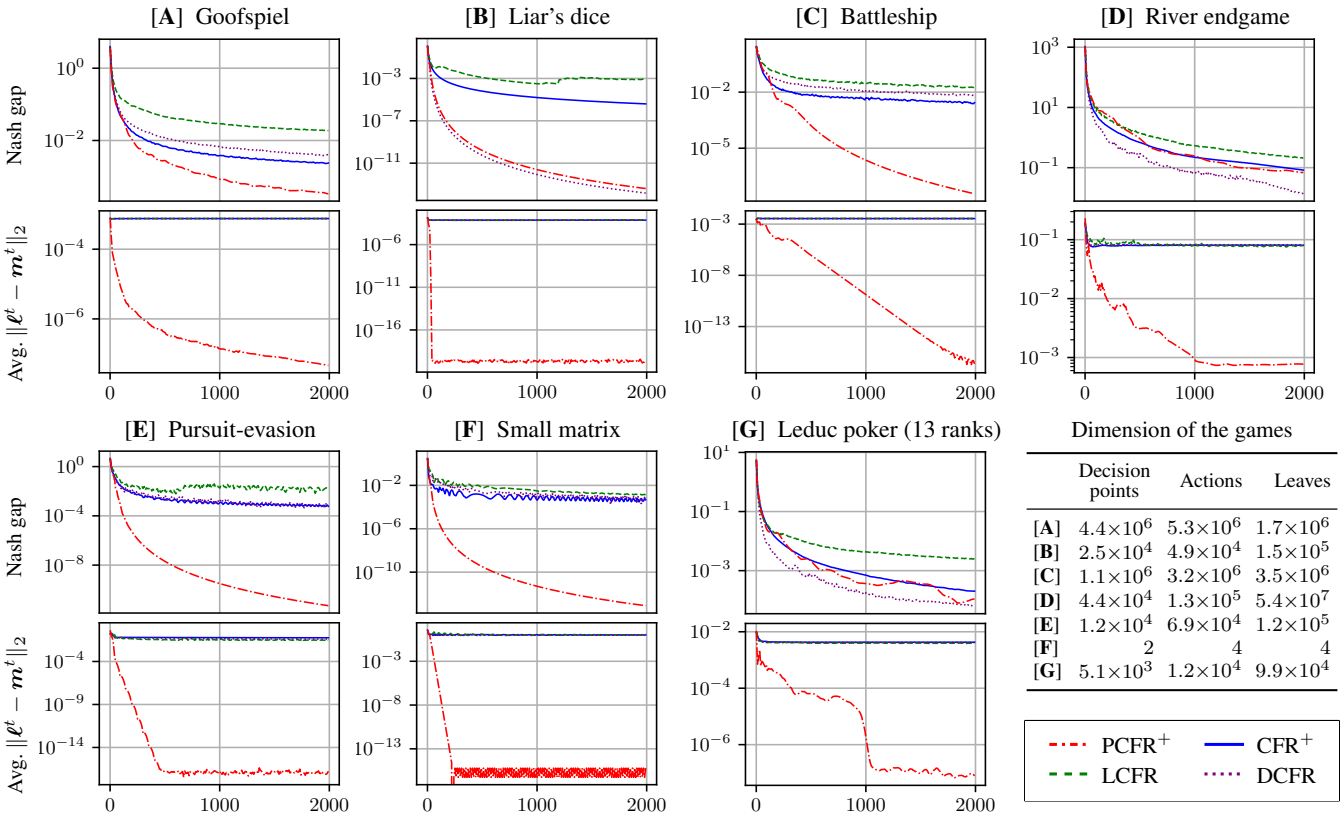
$$\begin{aligned}\max_{\hat{\boldsymbol{x}} \in \Delta^n} R^T(\hat{\boldsymbol{x}}) &\leq \inf_{\eta > 0}\left\{\frac{1}{2\eta} + \eta \sum_{t=1}^{T} \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) - \boldsymbol{v}^t\|_2^2\right\} \\ &= \sqrt{2}\left(\sum_{t=1}^{T} \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) - \boldsymbol{v}^2\|_2^2\right)^{1/2}.\end{aligned}$$

7 Experiments

We conduct experiments on solving two-player zero-sum games. As mentioned previously, for EFGs the CFR framework is used for decomposing regrets into local regret minimization problems at each simplex corresponding to a decision point in the game (Zinkevich et al. 2007; Farina, Kroer, and Sandholm 2019a), and we do the same. However, as the regret minimizer for each local decision point, we use PRM$^+$ instead of RM. In addition, we apply two heuristics that usually lead to better practical performance: we use quadratic averaging of the strategy iterates, that is, we average the sequence-form strategies $\boldsymbol{x}^1, \dots, \boldsymbol{x}^T$ using the formula $\frac{6}{T(T+1)(2T+1)}\sum_{t=1}^{T} t^2 \boldsymbol{x}^t$, and we use the *alternating updates* scheme. We call this algorithm PCFR$^+$. We compare PCFR$^+$ to the prior state-of-the-art CFR variants: CFR$^+$ (Tammelin 2014), *Discounted CFR (DCFR)* with its recommended parameters (Brown and Sandholm 2019a), and *Linear CFR (LCFR)* (Brown and Sandholm 2019a).

We conduct the experiments on common benchmark games. We show results on seven games in the main body of the paper. An additional 11 games are shown in the appendix of the full version of the paper. The experiments shown in the main body are representative of those in the appendix. A description of all the games is in Appendix G in the full version of the paper, and the results are shown in Figure 2. The x-axis shows the number of iterations of each algorithm. Every algorithm pays almost exactly the same cost per iteration, since the predictions require only one additional thresholding step in PCFR$^+$. For each game, the top plot shows on the y-axis the Nash gap, while the bottom plot shows the accuracy in our predictions of the regret vector, measured as the average ℓ_2 norm of the difference between the actual loss $\boldsymbol{\ell}^t$ received and its prediction $\boldsymbol{m}^t$ across all regret minimizers at all decision points in the game. For all non-predictive algorithms (CFR$^+$, LCFR, and DCFR), we let $\boldsymbol{m}^t = \boldsymbol{0}$. For our predictive algorithm, we set $\boldsymbol{m}^t = \boldsymbol{\ell}^{t-1}$ at all times $t \geq 2$ and $\boldsymbol{m}^1 = \boldsymbol{0}$. Both y-axes are in log scale. On Battleship and Pursuit-evasion, PCFR$^+$ is faster than the other algorithms by 3-6 orders of magnitude already after 500 iterations, and around 10 orders of magnitude after 2000 iterations. On Goofspiel, PCFR$^+$ is also significantly faster than the other algorithms, by 0.5-1 order of magnitude. Finally, in the River endgame, our only poker experiment here, PCFR$^+$ is slightly faster than CFR$^+$, but slower than DCFR. Finally, PRM$^+$ converges very rapidly on the *smallmatrix* game, a 2-by-2 matrix game where CFR$^+$ and other RM-based methods converge at a rate slower than T^{-1} (Farina, Kroer, and Sandholm 2019b). Across all non-poker games in the appendix, we also find that PCFR$^+$ beats the other algorithms, often by several orders of magnitude. We conclude that PCFR$^+$ seems to be the fastest method for solving non-poker EFGs. The only exception to the non-poker-game empirical rule is Liar's Dice (game [**B**]), where our predictive method performs comparably to DCFR. In the appendix, we also test CFR$^+$ with quadratic averaging (as opposed to the linear averaging that CFR$^+$ normally uses). This does not change any of our conclusions, except that for Liar's Dice, CFR$^+$ performs comparably to DCFR and PCFR$^+$ when using quadratic averaging (in fact, quadratic averaging hurts CFR$^+$ in every game except poker and Liar's Dice).

We tested on three poker games, the River endgame shown here (which is a real endgame encountered by the *Libratus* AI (Brown and Sandholm 2017) in the man-machine "Brains vs. Artificial Intelligence: Upping the Ante" com-



Dimension of the games

	Decision points	Actions	Leaves
[A]	4.4×10^6	5.3×10^6	1.7×10^6
[B]	2.5×10^4	4.9×10^4	1.5×10^5
[C]	1.1×10^6	3.2×10^6	3.5×10^6
[D]	4.4×10^4	1.3×10^5	5.4×10^7
[E]	1.2×10^4	6.9×10^4	1.2×10^5
[F]	2	4	4
[G]	5.1×10^3	1.2×10^4	9.9×10^4

Figure 2: Performance of PCFR$^+$, CFR$^+$, DCFR, and LCFR on five EFGs. In all plots, the x axis is the number of iterations of each algorithm. For each game, the top plot shows that the Nash gap on the y axis (on a log scale), the bottom plot shows and the average prediction error (on a log scale).

petition), as well as Kuhn and Leduc poker in the appendix. On Kuhn poker, PCFR$^+$ is extremely fast and the fastest of the algorithms. That game is known to be significantly easier than deeper EFGs for predictive algorithms (Farina, Kroer, and Sandholm 2019b). On Leduc poker as well as the River endgame, the predictions in PCFR$^+$ do not seem to help as much as in other games. On the River endgame, the performance is essentially the same as that of CFR$^+$. On Leduc poker, it leads to a small speedup over CFR$^+$. On both of those games, DCFR is fastest. In contrast, DCFR actually performs worse than CFR$^+$ in our non-poker experiments, though it is sometimes on par with CFR$^+$. In the appendix, where we try quadratic averaging in CFR$^+$, we find that for poker games this does speed up CFR$^+$, and allows it to be slightly faster than PCFR$^+$ on the River endgame and Leduc poker. We conclude that PCFR$^+$ is much faster than CFR$^+$ and DCFR on non-poker games, whereas on poker games DCFR is the fastest.

The convergence rate of PCFR$^+$ is closely related to how good the predictions $\boldsymbol{m}^t$ of $\boldsymbol{\ell}^t$ are. On Battleship and Pursuit-evasion, the predictions become extremely accurate very rapidly, and PCFR$^+$ converges at an extremely fast rate. On Goofspiel, the predictions are fairly accurate (the error is of the order 10^{-5}) and PCFR$^+$ is still significantly faster than the other algorithms. On the River endgame, the average prediction error is of the order 10^{-3}, and PCFR$^+$ performs on par with CFR$^+$, and slower than DCFR. Similar trends prevail in the experiments in the appendix. Additional experimental insights are described in the appendix.

8 Conclusions and Future Research

We extended Abernethy, Bartlett, and Hazan (2011)'s reduction of Blackwell approachability to regret minimization beyond the compact setting. This extended reduction allowed us to show that FTRL applied to the decision of which halfspace to force in Blackwell approachability is equivalent to the regret matching algorithm. OMD applied to the same problem turned out to be equivalent to RM$^+$. Then, we showed that the predictive variants of FTRL and OMD yield predictive algorithms for Blackwell approachability, as well as predictive variants of RM and RM$^+$. Combining PRM$^+$ with CFR, we introduced the PCFR$^+$ algorithm for solving EFGs. Experiments across many common benchmark games showed that PCFR$^+$ outperforms the prior state-of-the-art algorithms on non-poker games by orders of magnitude.

This work also opens future directions. Can PRM$^+$ guarantee T^{-1} convergence on matrix games like optimistic FTRL and OMD, or do the less stable updates prevent that? Can one develop a predictive variant of DCFR, which is faster on poker domains? Can one combine DCFR and PCFR$^+$, so DCFR would be faster initially but PCFR$^+$ would overtake? If the cross-over point could be approximated, this might yield a best-of-both-worlds algorithm.

Acknowledgments

This material is based on work supported by the National Science Foundation under grants IIS-1718457, IIS-1901403, and CCF-1733556, and the ARO under award W911NF2010081. Gabriele Farina is supported by a Facebook fellowship.

References

Abernethy, J.; Bartlett, P. L.; and Hazan, E. 2011. Blackwell Approachability and No-Regret Learning are Equivalent. In *COLT*, 27–46.

Blackwell, D. 1954. Controlled random walks. In *Proceedings of the international congress of mathematicians*, volume 3, 336–338.

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

Bošanskỳ, B.; and Čermák, J. 2015. Sequence-form algorithm for computing Stackelberg equilibria in extensive-form games. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Bošanskỳ, B.; Kiekintveld, C.; Lisý, V.; and Pěchouček, M. 2014. An Exact Double-Oracle Algorithm for Zero-Sum Extensive-Form Games with Imperfect Information. *Journal of Artificial Intelligence Research* 829–866.

Bowling, M.; Burch, N.; Johanson, M.; and Tammelin, O. 2015. Heads-up Limit Hold'em Poker is Solved. *Science* 347(6218).

Brown, N.; Kroer, C.; and Sandholm, T. 2017. Dynamic Thresholding and Pruning for Regret Minimization. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Brown, N.; and Sandholm, T. 2017. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science* eaao1733.

Brown, N.; and Sandholm, T. 2019a. Solving imperfect-information games via discounted regret minimization. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Brown, N.; and Sandholm, T. 2019b. Superhuman AI for multiplayer poker. *Science* 365(6456): 885–890.

Burch, N. 2018. *Time and space: Why imperfect information games are hard*. Ph.D. thesis, University of Alberta.

Burch, N.; Moravcik, M.; and Schmid, M. 2019. Revisiting CFR+ and alternating updates. *Journal of Artificial Intelligence Research* 64: 429–443.

Chiang, C.-K.; Yang, T.; Lee, C.-J.; Mahdavi, M.; Lu, C.-J.; Jin, R.; and Zhu, S. 2012. Online optimization with gradual variations. In *Conference on Learning Theory*, 6–1.

Farina, G.; Kroer, C.; Brown, N.; and Sandholm, T. 2019a. Stable-Predictive Optimistic Counterfactual Regret Minimization. In *International Conference on Machine Learning (ICML)*.

Farina, G.; Kroer, C.; and Sandholm, T. 2019a. Online Convex Optimization for Sequential Decision Processes and Extensive-Form Games. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Farina, G.; Kroer, C.; and Sandholm, T. 2019b. Optimistic Regret Minimization for Extensive-Form Games via Dilated Distance-Generating Functions. In *Advances in Neural Information Processing Systems*, 5222–5232.

Farina, G.; Kroer, C.; and Sandholm, T. 2019c. Regret Circuits: Composability of Regret Minimizers. In *International Conference on Machine Learning*, 1863–1872.

Farina, G.; Kroer, C.; and Sandholm, T. 2020. Stochastic regret minimization in extensive-form games. In *International Conference on Machine Learning (ICML)*.

Farina, G.; Ling, C. K.; Fang, F.; and Sandholm, T. 2019b. Correlation in Extensive-Form Games: Saddle-Point Formulation and Benchmarks. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Foster, D. P. 1999. A proof of calibration via Blackwell's approachability theorem. *Games and Economic Behavior* 29(1-2): 73–78.

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

Gordon, G. J. 2005. No-regret algorithms for structured prediction problems. Technical report, Carnegie-Mellon University, Computer Science Department, Pittsburgh PA USA.

Gordon, G. J. 2007. No-regret algorithms for online convex programs. In *Advances in Neural Information Processing Systems*, 489–496.

Hart, S.; and Mas-Colell, A. 2000. A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica* 68: 1127–1150.

Hoda, S.; Gilpin, A.; Peña, J.; and Sandholm, T. 2010. Smoothing Techniques for Computing Nash Equilibria of Sequential Games. *Mathematics of Operations Research* 35(2).

Kroer, C.; Farina, G.; and Sandholm, T. 2018a. Robust Stackelberg Equilibria in Extensive-Form Games and Extension to Limited Lookahead. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Kroer, C.; Farina, G.; and Sandholm, T. 2018b. Solving Large Sequential Games with the Excessive Gap Technique. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*.

Kroer, C.; Waugh, K.; Kılınç-Karzan, F.; and Sandholm, T. 2020. Faster algorithms for extensive-form game solving via improved smoothing functions. *Mathematical Programming* 179(1): 385–417.

Kuhn, H. W. 1950. A Simplified Two-Person Poker. In Kuhn, H. W.; and Tucker, A. W., eds., *Contributions to the Theory of Games*, volume 1 of *Annals of Mathematics Studies, 24*, 97–103. Princeton, New Jersey: Princeton University Press.

Lanctot, M.; Waugh, K.; Zinkevich, M.; and Bowling, M. 2009. Monte Carlo Sampling for Regret Minimization in Extensive Games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*.

Lisỳ, V.; Lanctot, M.; and Bowling, M. 2015. Online Monte Carlo counterfactual regret minimization for search in imperfect information games. In *Proceedings of the 2015 international conference on autonomous agents and multiagent systems*, 27–36.

Moravčík, M.; Schmid, M.; Burch, N.; Lisý, V.; Morrill, D.; Bard, N.; Davis, T.; Waugh, K.; Johanson, M.; and Bowling, M. 2017. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* 356(6337): 508–513.

Nesterov, Y. 2009. Primal-dual subgradient methods for convex problems. *Mathematical programming* 120(1): 221–259.

Rakhlin, A.; and Sridharan, K. 2013a. Online Learning with Predictable Sequences. In *Conference on Learning Theory*, 993–1019.

Rakhlin, S.; and Sridharan, K. 2013b. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, 3066–3074.

Ross, S. M. 1971. Goofspiel—the game of pure strategy. *Journal of Applied Probability* 8(3): 621–625.

Shalev-Shwartz, S.; and Singer, Y. 2007. A primal-dual perspective of online learning algorithms. *Machine Learning* 69(2-3): 115–142.

Southey, F.; Bowling, M.; Larson, B.; Piccione, C.; Burch, N.; Billings, D.; and Rayner, C. 2005. Bayes' Bluff: Opponent Modelling in Poker. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*.

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

Tammelin, O. 2014. Solving large imperfect information games using CFR+. *arXiv preprint arXiv:1407.5042* .

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

Waugh, K.; and Bagnell, D. 2015. A Unified View of Large-scale Zero-sum Equilibrium Computation. In *Computer Poker and Imperfect Information Workshop at the AAAI Conference on Artificial Intelligence (AAAI)*.

Zinkevich, M.; Bowling, M.; Johanson, M.; and Piccione, C. 2007. Regret Minimization in Games with Incomplete Information. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*.