Journal of Machine Learning Research 27 (2026) 1-36 Submitted 9/25; Revised 7/26; Published 7/26

Alternating Regret for Online Convex Optimization

Soumita Hait* HAIT@USC.EDU
Department of Computer Science
University of Southern California
Los Angeles, CA 90089, USA

Ping Li* PINGLEE@STU.SUFE.EDU.CN
Institute for Theoretical Computer Science
Shanghai University of Finance and Economics
Shanghai 200433, China

Haipeng Luo* HAIPENGL@USC.EDU
Department of Computer Science
University of Southern California
Los Angeles, CA 90089, USA

Mengxiao Zhang* MENGXIAO-ZHANG@UIOWA.EDU
Department of Business Analytics
Tippie College of Business
University of Iowa
Iowa City, IA 52242, USA

Editor: Vianney Perchet

Abstract

Motivated by alternating learning dynamics in two-player games, a recent work by Cevher et al. (2023) shows that $o(\sqrt{T})$ alternating regret is possible for any T-round adversarial Online Linear Optimization (OLO) problem, and left as an open question whether the same is true for general Online Convex Optimization (OCO). We answer this question in the affirmative by showing that the continuous Hedge algorithm achieves $\widetilde{\mathcal{O}}(d^{\frac{2}{3}}T^{\frac{1}{3}})$ alternating regret for any adversarial d-dimensional OCO problems. We show that this implies an alternating learning dynamic that finds a Nash equilibrium for any convex-concave zero-sum games or a coarse correlated equilibrium for any convex two-player general-sum games at a rate of $\widetilde{\mathcal{O}}(d^{\frac{2}{3}}/T^{\frac{2}{3}})$. To further improve the time complexity and/or the dimension dependence, we propose another simple algorithm, Follow-the-Regularized-Leader with a regularizer whose convex conjugate is 3rd-order smooth, for OCO with smooth and self-concordant loss functions (such as linear or quadratic losses). We instantiate our algorithm with different regularizers and show that, for example, when the decision set is the ℓ_2 ball, our algorithm achieves $\widetilde{\mathcal{O}}(T^{\frac{2}{5}})$ alternating regret with no dimension dependence (and a better $\widetilde{\mathcal{O}}(T^{\frac{1}{3}})$ bound for quadratic losses). We complement our results by showing some algorithm-specific alternating regret lower bounds, including a somewhat surprising $\Omega(\sqrt{T})$ lower bound for a Regret Matching variant that is widely used in alternating learning dynamics.

Keywords: online learning, online convex optimization, learning in games, alternating dynamics, alternating regret

∗. Authors are listed in alphabetical order.

©2026 Soumita Hait, Ping Li, Haipeng Luo, and Mengxiao Zhang.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v27/25-2294.html.

1. Introduction

Online Convex Optimization (OCO) is a heavily-studied framework for online learning (Zinkevich, 2003). In this framework, a learner sequentially interacts with an adversary for T rounds. In each round t, the learner selects an action x_t from a fixed decision space $\mathcal{X}$, after which the adversary decides a convex loss function $f_t : \mathcal{X} \mapsto \mathbb{R}$. The learner then suffers loss $f_t(x_t)$ and observes f_t. The standard performance measure for the learner is the worst-case *regret*, defined as the difference between their total loss over T rounds and that of a best fixed action in hindsight.

One important application of OCO algorithms is for learning in large-scale games and finding approximate equilibria efficiently. The fundamental connection between online learning and game solving dates back to Foster and Vohra (1997); Freund and Schapire (1999); Hart and Mas-Colell (2000) and has been driving some of the recent AI breakthroughs such as superhuman-level poker agents (Bowling et al., 2015; Moravčík et al., 2017; Brown and Sandholm, 2018, 2019). The most basic result in this area is that in a two-player zero-sum convex-concave game, if both players repeatedly play the game using an OCO algorithm to make their decisions, then their time-averaged strategy is an approximate Nash Equilibrium (NE) with the duality gap being the average regret of the OCO algorithm. Since the worst-case regret for most problems is $\mathcal{O}(\sqrt{T})$, this implies an $\mathcal{O}(1/\sqrt{T})$ convergence rate to NE.

There has been a surge of studies on how to improve over this $\mathcal{O}(1/\sqrt{T})$ convergence rate in recent years. One of the simplest approaches is to apply *alternation*, which requires minimal change to the original approach: let the two players make their decisions in turn instead of simultaneously. Such a simple adjustment turns out to be highly efficient in speeding up the convergence and is a default method in solving large-scale extensive-form games (Tammelin, 2014).

However, why alternation works so well has been a mystery until some recent progress. Specifically, Wibisono et al. (2022) show that in normal-form games (where losses are linear), applying alternation with standard algorithms such as Hedge (Freund and Schapire, 1997) achieves a faster convergence rate of $\mathcal{O}(1/T^{\frac{2}{3}})$. To analyze alternation, they propose a new regret notion, later termed *alternating regret* by Cevher et al. (2023), and show that the convergence rate in an alternating learning dynamic is governed by the average alternating regret of the two players. In a word, alternating regret can be seen as the sum of standard regret plus *cheating regret* where the learner's loss at time t is evaluated at the next decision x_{t+1}, that is, $f_t(x_{t+1})$. Since x_{t+1} is computed with the knowledge of f_t, cheating regret can be easily made negative, making alternating regret potentially much smaller than standard regret and explaining the faster $\mathcal{O}(1/T^{\frac{2}{3}})$ convergence.

Motivated by this work, Cevher et al. (2023) further investigate alternating regret for the special case of Online Linear Optimization (OLO) where all loss functions are linear. They show that, even when facing a *completely adversarial* environment, alternating regret can be made as small as $\mathcal{O}(T^{\frac{1}{3}})$ when the decision set $\mathcal{X}$ is a simplex, or even $\mathcal{O}(\log T)$ when $\mathcal{X}$ is an ℓ_2 ball (while the standard regret for both cases is $\Omega(\sqrt{T})$). The latter result immediately implies that applying their algorithm to an alternating learning dynamic for a zero-sum game defined over an ℓ_2 ball leads to $\widetilde{\mathcal{O}}(1/T)$ convergence.

However, the results of Cevher et al. (2023) are restricted to linear losses, and they explicitly ask the question whether similar results can be derived for convex losses. Indeed, while for standard regret, a convex loss f_t can be reduced to a linear loss $x \mapsto \langle \nabla f_t(x_t), x \rangle$, this is no longer the case for alternating regret (see Section 2.2 for detailed explanation).

1.1 Contributions.

In this work, we answer this open question in the affirmative and present several results that significantly advance the understanding of alternating regret and alternating learning dynamics in two-player games, including both zero-sum and general-sum settings. A summary of our results is provided in Table 1.1. Our contributions are as follows: Our concrete contributions are:

- In Section 3, we start with an observation that the vanilla Hedge algorithm (Freund and Schapire, 1997) already achieves $\mathcal{O}(T^{\frac{1}{3}} \log^{\frac{2}{3}} d)$ alternating regret for an adversarial d-expert problem (that is, OLO over a simplex), a result that is implicitly shown in Wibisono et al. (2022). This observation not only already improves over the $\mathcal{O}(T^{\frac{1}{3}} \log^{\frac{4}{3}}(dT))$ bound of the arguably more complicated algorithm by Cevher et al. (2023), but more importantly also inspires us to consider the *continuous Hedge* algorithm for general OCO. Indeed, via a nontrivial generalization of the analysis, we manage to show that continuous Hedge achieves $\widetilde{\mathcal{O}}(d^{\frac{2}{3}} T^{\frac{1}{3}})$ alternating regret for any convex decision set $\mathcal{X}$ and any bounded convex loss functions.

- This general result implies that for any convex-concave zero-sum games, applying continuous Hedge in an alternating learning dynamic leads to $\widetilde{\mathcal{O}}(d^{\frac{2}{3}}/T^{\frac{2}{3}})$ convergence to NE. Moreover, we generalize this implication to two-player general-sum games with convex losses, showing the same convergence rate for coarse correlated equilibria. As far as we know, this is the first result for alternating learning dynamics in general-sum games.

- Although continuous Hedge can be implemented in polynomial time using existing log-concave samplers, the time complexity can be high. Moreover, its alternating regret bound depends on the dimension d even when $\mathcal{X}$ is an ℓ_2 ball. To address these issues, in Section 4, we propose another simple algorithm that applies Follow-the-Regularized-Leader (FTRL) to the original convex losses (instead of their linearized version) with a Legendre regularizer whose convex conjugate is third-order smooth. Based on the decision set, we instantiate the algorithm with different regularizers and analyze its alternating regret for OCO with smooth and self-concordant loss functions (which include linear and quadratic losses). For example, when $\mathcal{X}$ is an ℓ_2 ball, our algorithm achieves $\widetilde{\mathcal{O}}(T^{\frac{2}{5}})$ alternating regret without any dimension dependence, which can be further improved to $\widetilde{\mathcal{O}}(T^{\frac{1}{3}})$ for quadratic losses.

- Finally, in Section 5, we discuss some algorithm-specific alternating regret lower bounds, focusing on the special case with linear losses over a simplex. We start by showing that $\Omega(T^{\frac{1}{3}})$ alternating regret is unavoidable for Hedge, at least not with a fixed learning rate. We then turn our attention to a widely-used algorithm, called Predictive Regret Matching$^+$ (Farina et al., 2021), and show that despite its amazing empirical performance in alternating learning dynamics, its alternating regret is in fact $\Omega(\sqrt{T})$ when facing an

Table 1: This table provides a summary of all the alternating regret bounds in this paper. Here, SC stands for self-concordant losses and C is the self-concordance parameter ($C = 0$ for quadratic losses). All the results are in $\mathcal{O}(\cdot)$ notation hiding problem dependent constants, and the results for the smooth and self-concordant OCO setting also hides polylog-factors in T.

Setting	**Domain**	**Prior works**	**This paper**
OLO	Δ_d	$T^{1/3}\log^{4/3}(dT)$ Cevher et al. (2023)	$T^{1/3}\log^{2/3} d$ (Theorem 3)
OCO	$\mathcal{X} \subseteq \mathbb{R}^d$	N/A	$d^{2/3}T^{1/3}\log^{2/3} T$ (Theorem 4)
Smooth + SC OCO	$\mathcal{X} \subseteq \mathbb{R}^d$	N/A	$\max\{d^{3/5}(CT)^{2/5}, d^{2/3}T^{1/3}\}$ (Corollary 12)
	$\mathcal{B}_2^d(1)$	N/A	$\max\{(CT)^{2/5}, T^{1/3}\}$ (Corollary 14)
	Δ_d	N/A	$\max\{(CT)^{2/5}\log^{3/5} d, T^{1/3}\log^{2/3} d\}$ (Corollary 16)

adversarial environment. This implies that to demystify its excellent performance, one has to take the game structure into account. Moreover, using the same loss sequence, we show that another standard algorithm called Optimistic Online Gradient Descent (OOGD) (Rakhlin and Sridharan, 2013) also suffers $\Omega(\sqrt{T})$ alternating regret.

1.2 Related Work.

We refer the reader to Hazan et al. (2016); Orabona (2019) for a general introduction to OCO. The continuous Hedge algorithm dates back to Cover's universal portfolio algorithm (Cover, 1991), and is known to achieve $\widetilde{\mathcal{O}}(\sqrt{dT})$ (standard) regret for general convex losses (see e.g., Narayanan and Rakhlin (2010)) and $\mathcal{O}(d\log T)$ (standard) regret for exp-concave losses (Hazan et al., 2007). OCO with self-concordant loss functions have been studied in Zhang et al. (2017) under the context of dynamic regret.

As mentioned, there is a surge of studies on learning dynamics that converge faster than $1/\sqrt{T}$, such as Daskalakis et al. (2011); Rakhlin and Sridharan (2013); Syrgkanis et al. (2015); Chen and Peng (2020); Daskalakis et al. (2021); Farina et al. (2022b); Anagnostides et al. (2022a,b); Farina et al. (2022a). All these works achieve acceleration by improving the standard $\mathcal{O}(\sqrt{T})$ regret bound, which is only possible by exploiting the fact that the environment is not completely adversarial but controlled by other players who deploy a similar learning algorithm.

Alternation, on the other hand, is a much simpler approach to achieve acceleration, but the reason why it works so well is not well understood. Earlier works can only show that it works at least as well as the standard simultaneous dynamics (Tammelin, 2014; Burch et al., 2019) or it works strictly better but without a good characterization on how much

better (Grand-Clément and Kroer, 2024). Wibisono et al. (2022) are the first to show a substantial convergence improvement brought by alternation via the notion of alternating regret. Their $\mathcal{O}(1/T^{\frac{2}{3}})$ convergence rate was later improved to $\mathcal{O}(1/T^{\frac{4}{5}})$ by Katona et al. (2024). While this latter result heavily relies on the game structure, as we point out in Section 3, the alternating regret bound of Wibisono et al. (2022) in fact holds even in the adversarial setting. Focusing on OLO, Cevher et al. (2023) further provide more results on $o(\sqrt{T})$ alternating regret in an adversarial environment. These results make the analysis for alternating learning dynamics much simpler since one can simply ignore how the opponent behaves and how the players' decisions are entangled. Our work significantly advances our understanding on how far this approach can go by generalizing their results to convex losses as well as providing lower bounds for popular algorithms.

2. Preliminaries

2.1 General Notations.

For a convex function $f : \mathcal{X} \mapsto \mathbb{R}$ and $x, y \in \mathcal{X} \subseteq \mathbb{R}^d$, define the Bregman divergence between x and y with respect to f as $D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$. Define the convex conjugate of f, denoted by $f^* : \mathbb{R}^d \mapsto \mathbb{R}$, as $f^*(y) = \sup_{x \in \mathcal{X}} \{\langle x, y \rangle - f(x)\}$. For a vector $v \in \mathbb{R}^d$, denote v_i as the i-th entry of the vector. Define $\mathbb{R}^d_+$ to be the d-dimensional Euclidean space in the positive orthant and $\Delta_d = \{x \in \mathbb{R}^d_+ : \sum_{i=1}^d x_i = 1\}$ to be the $(d-1)$-dimensional simplex. We also denote $\Delta_{\mathcal{X}}$ to be the set of probability densities over a convex domain $\mathcal{X} \subset \mathbb{R}^d$, i.e., $\Delta_{\mathcal{X}} = \left\{p : \mathcal{X} \to \mathbb{R}_+ \mid \int_{x \in \mathcal{X}} p(x)\, dx = 1\right\}$. Let $\mathcal{B}_2^d(1) = \{x \in \mathbb{R}^d, \|x\|_2 \leq 1\}$ be the ℓ_2 unit ball, and e_i be the one-hot vector in an appropriate dimension with the i-th entry being 1 and remaining entries being 0. For a bounded convex domain $\mathcal{X}$, define its diameter as $\max_{x', x \in \mathcal{X}} \|x - x'\|_2$. Define $\mathrm{int}(\mathcal{X})$ as the interior of a convex domain $\mathcal{X}$ and $\partial(\mathcal{X})$ as the boundary of $\mathcal{X}$. The notation $\widetilde{\mathcal{O}}(\cdot)$ is used to hide all logarithmic terms.

2.2 Adversarial Online Convex Optimization

In adversarial Online Convex Optimization (OCO), at each round $t \in [T]$, the learner picks an action $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$ for some convex domain $\mathcal{X}$, and then the adversary picks a convex loss function $f_t : \mathcal{X} \mapsto \mathbb{R}$ and reveals it to the learner. The standard regret measures the difference between the total loss suffered by the learner and that of a best fixed decision in hindsight: $\mathrm{Reg} \triangleq \max_{u \in \mathcal{X}} \mathrm{Reg}(u)$ where $\mathrm{Reg}(u) \triangleq \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(u)$. On the other hand, the main focus of our work is *alternating regret* (Wibisono et al., 2022; Cevher et al., 2023), defined as follows: $\mathrm{Reg}_{\mathrm{Alt}} \triangleq \max_{u \in \mathcal{X}} \mathrm{Reg}_{\mathrm{Alt}}(u)$ where

$$\mathrm{Reg}_{\mathrm{Alt}}(u) \triangleq \sum_{t=1}^{T} \left(f_t(x_t) + f_t(x_{t+1})\right) - 2\sum_{t=1}^{T} f_t(u)$$

is the sum of standard regret $\mathrm{Reg}(u)$ and *cheating regret* $\mathrm{Reg}_{\mathrm{Cht}}(u) \triangleq \sum_{t=1}^T f_t(x_{t+1}) - \sum_{t=1}^T f_t(u)$.[1] The cheating regret evaluates the learner's loss at time t using the *next* decision

1. For notational convenience, we include x_{T+1}, the output of the algorithm at the beginning of round $T+1$, in the definition.

x_{t+1} that is computed with the knowledge of f_t (hence cheating), and thus intuitively can be easily made negative. The hope is that it is negative enough so that the alternating regret is order-different from the standard regret. Indeed, for the special case of Online Linear Optimization (OLO) where all f_t's are linear functions, Cevher et al. (2023) show a seperation between standard regret Reg and alternating regret $\mathrm{Reg}_{\mathrm{Alt}}$: while the minimax bound for the former is $\Theta(\sqrt{T})$, the latter can be as small as $\widetilde{\mathcal{O}}(T^{\frac{1}{3}})$ when $\mathcal{X}$ is a simplex and $\mathcal{O}(\log T)$ when $\mathcal{X}$ is an ℓ_2 ball.

However, as explicitly mentioned in Cevher et al. (2023), extension to general convex functions is highly unclear. To see this, first recall the standard OCO to OLO reduction for standard regret: $\mathrm{Reg}(u) \le \sum_{t=1}^{T} \langle \nabla f_t(x_t), x_t - u \rangle$, which is due to the convexity of f_t and reveals that it is sufficient to consider an OLO instance with linear loss $x \mapsto \langle \nabla f_t(x_t), x \rangle$ at time t. However, when applying the same trick to alternating regret: that is, bounding $\mathrm{Reg}_{\mathrm{Alt}}(u)$ by $\sum_{t=1}^{T} \langle \nabla f_t(x_t), x_t - u \rangle + \langle \nabla f_t(x_{t+1}), x_{t+1} - u \rangle$, one quickly realizes that this does not correspond to the alternating regret of any OLO instance. In particular, the alternating regret for an OLO instance with linear loss $x \mapsto \langle \nabla f_t(x_t), x \rangle$ at time t is instead: $\sum_{t=1}^{T} \langle \nabla f_t(x_t), x_t - u \rangle + \langle \nabla f_t(x_t), x_{t+1} - u \rangle$. One might hope to use the fact that x_t and x_{t+1} are close in standard OCO algorithms to resolve this discrepancy, but we point out that they are still not close enough to lead to $o(\sqrt{T})$ regret. Indeed, x_t and x_{t+1} are usually η-close for an algorithm with a learning rate η, making $f_t(x_t)$ and $f_t(x_{t+1})$ also only η-close for a smooth f_t. This contributes to an $\mathcal{O}(\eta T)$ term in the final alternating regret bound, which, together with the standard $\mathcal{O}(\frac{1}{\eta})$ term in regret analysis, leads to an $\Omega(\sqrt{T})$ bound.

2.3 Alternating Learning Dynamics in Games

The motivation of studying alternating regret stems from learning in games with *alternation*. Specifically, consider a two-player general-sum game where $u_1(x, y), u_2(x, y) : \mathcal{X} \times \mathcal{Y} \mapsto [-1, 1]$ are the loss functions for x-player and y-player, respectively. $u_1(x, y)$ is convex in x for any $y \in \mathcal{Y}$, and $u_2(x, y)$ is convex in y for any $x \in \mathcal{X}$. A pair of strategies $(\bar{x}, \bar{y})$ is an ε-NE (Nash Equilibrium) if no player has more than ε incentive to unilaterally deviate from it: $u_1(\bar{x}, \bar{y}) \le \min_{x \in \mathcal{X}} u_1(x, \bar{y}) + \varepsilon$ and $u_2(\bar{x}, \bar{y}) \le \min_{y \in \mathcal{Y}} u_2(\bar{x}, y) + \varepsilon$. A joint distribution $\mathcal{P}$ over $\mathcal{X} \times \mathcal{Y}$ is called an ε-CCE (Coarse Correlated Equilibrium) if in expectation no player has more than ε incentive to unilaterally deviate from a strategy drawn from $\mathcal{P}$ to a fixed strategy: $\mathbb{E}_{(x,y)\sim\mathcal{P}}[u_1(x, y)] \le \min_{x' \in \mathcal{X}} \mathbb{E}_{(x,y)\sim\mathcal{P}}[u_1(x', y)] + \varepsilon$ and $\mathbb{E}_{(x,y)\sim\mathcal{P}}[u_2(x, y)] \le \min_{y' \in \mathcal{Y}} \mathbb{E}_{(x,y)\sim\mathcal{P}}[u_2(x, y')] + \varepsilon$.

An efficient and popular way to find these equilibria is via repeated play using online learning algorithms. For example, in a standard (simultaneous) learning dynamic, both players use an OCO algorithm to make their decision $x_t \in \mathcal{X}$ and $y_t \in \mathcal{Y}$ for round t simultaneously and observe their loss function $u_1(x, y_t)$ and $u_2(x_t, y)$ respectively. After interacting for T rounds, it is well known that the uniform distribution over $(x_1, y_1), \ldots, (x_T, y_T)$ is an ε-CCE with ε being the average regret. Moreover, in the special case of zero-sum games ($u_2 = -u_1$), the average strategy $(\frac{1}{T}\sum_{t=1}^{T} x_t, \frac{1}{T}\sum_{t=1}^{T} y_t)$ is an ε-NE for the same ε (see e.g., Freund and Schapire, 1999). Since one can ensure $\mathcal{O}(\sqrt{T})$ standard regret, a convergence rate of $\mathcal{O}(1/\sqrt{T})$ to these equilibria is immediate.

There are many studies on how to accelerate such $\mathcal{O}(1/\sqrt{T})$ convergence, with alternation probably being the simplest and most practically popular one. Specifically, the

Algorithm 1: Alternating Learning Dynamic using OCO

Input: a two-player game with loss functions $u_1, u_2 : \mathcal{X} \times \mathcal{Y} \mapsto [-1, 1]$.
Input: OCO algorithms Alg_x (for $\mathcal{X}$) and Alg_y (for $\mathcal{Y}$).
for $t = 1$ ***to*** T **do**
  Alg_x decides $x_t \in \mathcal{X}$ using past loss functions $u_1(x, y_1), \ldots, u_1(x, y_{t-1})$
  Alg_y decides $y_t \in \mathcal{Y}$ using past loss functions $u_2(x_1, y), \ldots, u_2(x_{t-1}, y)$ and $u_2(x_t, y)$

difference in an alternating learning dynamic compared to a simultaneous one is that at each round t, the x-player first makes their decision x_t, and then the y-player, *seeing* x_t, makes their decision y_t; see Algorithm 1.

In the special case of zero-sum games ($u_2 = -u_1$), Wibisono et al. (2022); Cevher et al. (2023) show that the average strategy of such an alternating learning dynamic is an ε-NE with ε now being the average *alternating* regret. While they only prove this for linear losses, it is straightforward to generalize it to convex losses, which is included below for completeness.

Theorem 1 *In the alternating learning dynamic described by Algorithm 1 for a zero-sum game ($u_2 = -u_1$), suppose that Alg_x and Alg_y guarantee a worst-case alternating regret bound $\mathrm{Reg}^x_{\mathrm{Alt}}$ and $\mathrm{Reg}^y_{\mathrm{Alt}}$ respectively. Then, the average strategy $(\frac{1}{T}\sum_{t=1}^T x_t, \frac{1}{T}\sum_{t=1}^T y_t)$ is an ε-NE with $\varepsilon = \mathcal{O}\left(\frac{\mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}}}{T}\right)$.*

Proof Let y_0 be the output of Alg_y with an empty sequence of loss functions. According to Algorithm 1, the loss function for the x-player at round t is given by $u_1(\cdot, y_t)$, while the loss function for the y-player at the same round is $u_2(x_t, \cdot)$. By definition of alternating regret and using $u_2 = -u_1$, we know that for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$\begin{aligned}
&\sum_{t=1}^{T} \left(u_1(x_t, y_t) + u_1(x_{t+1}, y_t)\right) - 2\sum_{t=1}^{T} u_1(x, y_t) \le \mathrm{Reg}^x_{\mathrm{Alt}}, \\
&\sum_{t=0}^{T-1} \left(u_2(x_{t+1}, y_t) + u_2(x_{t+1}, y_{t+1})\right) - 2\sum_{t=0}^{T-1} u_2(x_{t+1}, y) \\
&= 2\sum_{t=0}^{T-1} u_1(x_{t+1}, y) - \sum_{t=0}^{T-1} \left(u_1(x_{t+1}, y_t) + u_1(x_{t+1}, y_{t+1})\right) \le \mathrm{Reg}^y_{\mathrm{Alt}}.
\end{aligned}$$

Taking a summation over both equations, we obtain that

$$\begin{aligned}
2\sum_{t=1}^{T} u_1(x_t, y) - 2\sum_{t=1}^{T} u_1(x, y_t) &\le \mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}} + u_1(x_1, y_0) - u_1(x_{T+1}, y_T) \\
&\le \mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}} + 2.
\end{aligned}$$

Since $u_1(x, y)$ is convex in x and concave in y, using Jensen's inequality, we know that for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$u_1\left(\overline{x}_T, y\right) - u_1(x, \overline{y}_T) \leq \frac{\mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}} + 2}{2T} = \mathcal{O}\left(\frac{\mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}}}{T}\right),$$

where $\overline{x}_T = \frac{1}{T}\sum_{t=1}^T x_t$ and $\overline{y}_T = \frac{1}{T}\sum_{t=1}^T y_t$. This further means

$$0 \leq \max_{y \in \mathcal{Y}} u_1(\overline{x}_T, y) - u_1(\overline{x}_T, \overline{y}_T) \leq \mathcal{O}\left(\frac{\mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}}}{T}\right),$$
$$0 \leq u_1(\overline{x}_T, \overline{y}_T) - \min_{x \in \mathcal{X}} u_1(x, \overline{y}_T) \leq \mathcal{O}\left(\frac{\mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}}}{T}\right),$$

which finishes the proof by the definition of an ε-NE with $\varepsilon = \mathcal{O}\left(\frac{\mathrm{Reg}^x_{\mathrm{Alt}} + \mathrm{Reg}^y_{\mathrm{Alt}}}{T}\right)$. ■

Therefore, if alternating regret can be made $o(\sqrt{T})$, a faster than $1/\sqrt{T}$ convergence rate is achieved. In fact, such an implication generalizes to CCE for two-player general-sum games as well, which is unknown before to the best of our knowledge.

Theorem 2 *In the alternating learning dynamic described by Algorithm 1 (for general-sum games), suppose that* Alg_x *and* Alg_y *guarantee a worst-case alternating regret bound* $\mathrm{Reg}^x_{\mathrm{Alt}}$ *and* $\mathrm{Reg}^y_{\mathrm{Alt}}$ *respectively. Then, the uniform distribution over* $\{(x_t, y_t), (x_{t+1}, y_t)\}_{t \in [T]}$ *is an* ε*-CCE with* $\varepsilon = \mathcal{O}\left(\frac{\max\{\mathrm{Reg}^x_{\mathrm{Alt}}, \mathrm{Reg}^y_{\mathrm{Alt}}\}}{T}\right)$.

Proof Similar to the proof of Theorem 1, let y_0 be the output of Alg_y with an empty sequence of loss functions. By definition of alternating regret, we know that for any $x' \in \mathcal{X}$ and $y' \in \mathcal{Y}$,

$$\sum_{t=1}^{T} \left(u_1(x_t, y_t) + u_1(x_{t+1}, y_t)\right) - 2\sum_{t=1}^{T} u_1(x', y_t) \leq \mathrm{Reg}^x_{\mathrm{Alt}}, \tag{1}$$
$$\sum_{t=0}^{T-1} \left(u_2(x_{t+1}, y_t) + u_2(x_{t+1}, y_{t+1})\right) - 2\sum_{t=0}^{T-1} u_2(x_{t+1}, y') \leq \mathrm{Reg}^y_{\mathrm{Alt}}. \tag{2}$$

Define $\mathcal{P}$ to be the uniform distribution over $\{(x_t, y_t), (x_{t+1}, y_t)\}_{t \in [T]}$. Adding $u_2(x_{T+1}, y_T) - u_2(x_1, y_0) + u_2(x_1, y') - u_2(x_{T+1}, y') \in [-4, 4]$ on both sides of Eq. (2) and dividing $2T$ on both sides of Eq. (1) and Eq. (2), we know that for any $x' \in \mathcal{X}$ and $y' \in \mathcal{Y}$,

$$\mathbb{E}_{(x,y)\sim\mathcal{P}}\left[u_1(x, y)\right] - \mathbb{E}_{(x,y)\sim\mathcal{P}}\left[u_1(x', y)\right] \leq \frac{\mathrm{Reg}^x_{\mathrm{Alt}}}{2T},$$
$$\mathbb{E}_{(x,y)\sim\mathcal{P}}\left[u_2(x, y)\right] - \mathbb{E}_{(x,y)\sim\mathcal{P}}\left[u_2(x, y')\right] \leq \frac{\mathrm{Reg}^y_{\mathrm{Alt}} + 4}{2T}.$$

This finishes the proof by the definition of an ε-CCE with $\varepsilon = \mathcal{O}\left(\frac{\max\{\mathrm{Reg}^x_{\mathrm{Alt}}, \mathrm{Reg}^y_{\mathrm{Alt}}\}}{T}\right)$. ■

We remark that it is unclear how to generalize Theorem 2 to a general-sum game with more than two players. We conjecture that such a generalization requires either a new alternation scheme or a new concept of regret.

Given these connections between alternating regret and the convergence of alternating learning dynamics, the rest of the paper focuses on understanding what alternating regret bounds are achievable in an adversarial environment.

3. $o(\sqrt{T})$ Alternating Regret for OCO

In this section, we propose an OCO algorithm with $\widetilde{\mathcal{O}}(T^{\frac{1}{3}})$ alternating regret. Our idea is based on the following observation for the special case of the expert problem where $\mathcal{X} = \Delta_d$ and $f_t(x) = \langle \ell_t, x \rangle$, $\|\ell_t\|_\infty \le 1$ for all $t \in [T]$. While Wibisono et al. (2022) show that the vanilla Hedge algorithm (Freund and Schapire, 1997) achieves $\widetilde{\mathcal{O}}(T^{\frac{1}{3}})$ alternating regret when played against itself in a zero-sum game, the same result in fact holds for any adversarial expert problem, which was not made explicit in their work. We summarize this observation in the following theorem.

Theorem 3 (Implicit in Wibisono et al., 2022) *For an adversarial OLO problem with $\mathcal{X} = \Delta_d$ and $f_t(x) = \langle \ell_t, x \rangle$, $\|\ell_t\|_\infty \le 1$ for all $t \in [T]$, the Hedge algorithm that plays $p_t \in \Delta_d$ in round t such that $p_{t,i} \propto \exp\left(-\eta \sum_{\tau=1}^{t-1} \ell_{\tau,i}\right)$ with $\eta = T^{-\frac{1}{3}} \log^{\frac{1}{3}} d$ ensures $\mathrm{Reg}_{\mathrm{Alt}} = \mathcal{O}(T^{\frac{1}{3}} \log^{\frac{2}{3}} d)$.*

Proof [Sketch] Via standard analysis of Hedge, its alternating regret against any u is equal to

$$\frac{2(\mathrm{KL}(u, p_1) - \mathrm{KL}(u, p_{T+1}))}{\eta} + \frac{1}{\eta} \sum_{t=1}^{T} \left(\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)\right). \tag{3}$$

Since p_1 is the uniform distribution over d experts by definition, the first term

$$\frac{2(\mathrm{KL}(u, p_1) - \mathrm{KL}(u, p_{T+1}))}{\eta}$$

can be bounded by $\frac{2 \log d}{\eta}$. It remains to control the second term, which we call a KL commutator following the term Bregman commutator from Wibisono et al. (2022). Direct calculation shows that

$$\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t) = D_{G^*}(-\eta L_t, -\eta L_{t-1}) - D_{G^*}(-\eta L_{t-1}, -\eta L_t),$$

where $G^*(w) = \log(\sum_{t=1}^{T} \exp(w_i))$ (the convex conjugate of negative entropy) and $L_t = \sum_{\tau \le t} \ell_\tau$. Then, applying the key Lemma A.2 of Wibisono et al. (2022) (included as Lemma 19 in Appendix A) which bounds the KL commutator of a 3rd-order smooth function (see Definition 7), we arrive at $D_{G^*}(-\eta L_t, -\eta L_{t-1}) - D_{G^*}(-\eta L_{t-1}, -\eta L_t) \le \frac{4}{3}\|\eta \ell_t\|_\infty^3 \le \frac{4}{3}\eta^3$, since $G^*(x)$ is 8-smooth of order 3 as proven in Wibisono et al. (2022, Example A.3). Plugging this bound to Eq. (3) and picking the optimal η finishes the proof. (The full proof is deferred to Appendix A.) ■

Algorithm 2: Continuous Hedge

Input: Parameter $\eta > 0$

for $t = 1$ ***to*** T **do**

The learner computes $p_t \in \Delta_{\mathcal{X}}$ such that $p_t(x) \propto \exp\left(-\eta \sum_{\tau=1}^{t-1} f_\tau(x)\right)$

The learner chooses $x_t = \int_{x \in \mathcal{X}} p_t(x) x dx$ and observes f_t.

This bound not only improves upon the $\mathcal{O}(T^{\frac{1}{3}} \log^{\frac{4}{3}} d)$ guarantee of a much more complicated algorithm by Cevher et al. (2023), but perhaps more importantly, also inspires us to consider whether a continuous version of Hedge, called continuous Hedge (Algorithm 2), can handle the general OCO setting well. Indeed, for standard regret, continuous Hedge achieves an $\widetilde{\mathcal{O}}(\sqrt{dT})$ bound for any OCO instance (Narayanan and Rakhlin, 2010), generalizing Hedge's $\mathcal{O}(\sqrt{T \log d})$ standard regret from OLO to OCO. Fortunately, in a similar way and via non-trivial analysis, we manage to show that the same generalization carries over for alternating regret, as shown below.

Theorem 4 *For any OCO instance with* $\max_{x \in \mathcal{X}} |f_t(x)| \leq 1$ *for all* $t \in [T]$*, Algorithm 2 with* $\eta = \min\{1, T^{-\frac{1}{3}} (d \log T)^{\frac{1}{3}}\}$ *achieves* $\mathrm{Reg}_{\mathrm{Alt}} = \mathcal{O}(d^{\frac{2}{3}} T^{\frac{1}{3}} \log^{\frac{2}{3}} T)$.

The general idea of the proof follows that of Theorem 3, which is to bound the alternating regret by the KL commutator $\sum_{t=1}^{T} (\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t))$. The main technical part of the proof is then to further bound the per-round KL-divergence commutator $\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)$ by $\mathcal{O}(\eta^3)$ for the continuous distribution p_t, p_{t+1} over the whole convex domain, which is done via a proof that deviates from that of Wibisono et al. (2022, Lemma A.2). See Appendix A for details.

We have thus shown that $\widetilde{\mathcal{O}}(d^{\frac{2}{3}} T^{\frac{1}{3}})$ alternating regret is achievable for general OCO, resolving the open question asked by Cevher et al. (2023) and bypassing the obstacle that OCO cannot be reduced to OLO in this case. Note that we do not require either Lipschitzness or smoothness of the loss functions, but only bounded function value. Moreover, combining Theorem 4 with Theorem 1 and Theorem 2, we immediately obtain the following convergence result on an alternating learning dynamic for general two-player games with convex loss functions.

Corollary 5 *Consider an alternating learning dynamic described by Algorithm 1 where the loss function* $u_1(\cdot, y), u_2(x, \cdot) \in [-1, 1]$ *are convex for all* $x \in \mathcal{X}$ *and* $y \in \mathcal{Y}$*, and* Alg_x *and* Alg_y *are Algorithm 2 with* $\eta = (d \log T)^{\frac{1}{3}} T^{-\frac{1}{3}}$ *(and* $f_t(\cdot)$ *being* $u_1(\cdot, y_t)$ *for* Alg_x *and* $u_2(x_{t+1}, \cdot)$ *for* Alg_y*). Then the time-averaged strategy* $(\frac{1}{T} \sum_{t=1}^{T} x_t, \frac{1}{T} \sum_{t=1}^{T} y_t)$ *is an* $\widetilde{\mathcal{O}}(d^{\frac{2}{3}} T^{-\frac{2}{3}})$*-NE when the game is zero-sum; otherwise, the uniform distribution over the strategies* $\{(x_t, y_t), (x_{t+1}, y_t)\}_{t \in [T]}$ *is an* $\widetilde{\mathcal{O}}(d^{\frac{2}{3}} T^{-\frac{2}{3}})$*-CCE.*

While there exist (simultaneous) learning dynamics that achieve faster convergence for games with convex losses (Syrgkanis et al., 2015; Farina et al., 2022a), our results are the first to show that better than $1/\sqrt{T}$ convergence is possible using simple alternating learning dynamics.

4. Another Algorithm for Smooth and Self-Concordant Losses

One issue with continuous Hedge is that its implementation requires a log concave sampler (e.g., Lovász and Vempala (2003, 2006b); Narayanan and Rakhlin (2010)), which can be expensive when the dimension d is high (even though the time complexity is polynomial in d).[2] Another issue is that its alternating regret bound has dimension dependence even when $\mathcal{X}$ is an ℓ_2 ball where it is well-known that d-independent standard regret bound is possible. To mitigate these issues, in this section, we propose yet another general algorithm and instantiate it for different decision spaces. Our result here, however, only applies to a restricted class of convex losses.

Specifically, we start by providing some basic definitions on functions (throughout this section, all norms considered are ℓ_2-norm).

Definition 6 *We say that a differentiable function $f : \mathcal{X} \mapsto \mathbb{R}$ is L-Lipschitz if for any $x \in \mathcal{X}$, $\|\nabla f(x)\|_2 \leq L$, β-smooth if for any $x, y \in \mathcal{X}$, $D_f(x, y) \leq \frac{\beta}{2}\|x - y\|_2^2$, σ-strongly convex if for any $x, y \in \mathcal{X}$, $D_f(x, y) \geq \frac{\sigma}{2}\|x - y\|_2^2$.*

Definition 7 (3rd-order smoothness) *A function $\psi : \mathcal{X} \mapsto \mathbb{R}$ is said to be M-*smooth of order 3 *if it is three times differentiable, and its third-order derivative at any $w \in \mathcal{X}$ satisfies $|\nabla^3\psi(w)[h, h, h]| \leq M,\ \forall h \in \mathbb{R}^d, \|h\|_2 = 1$.*

Definition 8 *A convex function $f : \mathcal{X} \mapsto \mathbb{R}$ for a convex body $\mathcal{X}$ is Legendre if it is differentiable and strictly convex in int($\mathcal{X}$) and also $\|\nabla f(x)\|_2 \to \infty$ as $x \to \partial\mathcal{X}$ (the boundary of $\mathcal{X}$).*

Definition 9 (Self-concordance) *(Nesterov, 1994) A function $f : \mathcal{X} \mapsto \mathbb{R}$ is said to be C-self-concordant for $C \geq 0$, if it is three times differentiable, and for all $x \in \mathcal{X}$ and all $h \in \mathbb{R}^d$, the following inequality holds: $|\nabla^3 f(x)[h, h, h]| \leq 2C\left(\nabla^2 f(x)[h, h]\right)^{3/2}$.*

Our algorithm is the classical and efficient Follow-the-Regularized-Leader algorithm that, with a learning rate $\eta > 0$ and a regularizer $\psi : \mathcal{X} \mapsto \mathbb{R}$, plays

$$x_t = \operatorname*{argmin}_{x \in \mathcal{X}} \sum_{\tau=1}^{t-1} f_\tau(x) + \frac{1}{\eta}\psi(x)$$

in round t; see Algorithm 3. However, there are two important elements: first, we apply FTRL directly to the original loss functions f_τ instead of its linearized version $x \mapsto \langle \nabla f_\tau(x_\tau), x\rangle$, in light of the issue discussed in Section 2.2; second, it is critical that our regularizer ψ satisfies the following conditions.

Assumption 1 *The regularizer ψ is Legendre and σ-strongly convex within domain $\mathcal{X}$. Moreover, its convex conjugate $\psi^*(w) = \sup_{x \in \mathcal{X}} \langle w, x\rangle - \psi(x)$ is M-smooth of order 3.*

We will analyze Algorithm 3 under the following assumption on the loss functions.

Assumption 2 *For all $t \in [T]$, the loss function f_t is L-Lipschitz , β-smooth and C-self-concordant.*

2. We provide a discussion on the runtime complexity of continuous Hedge in Appendix A.1.

Algorithm 3: FTRL for minimizing alternating regret

Input: learning rate $\eta > 0$.
Input: a regularizer ψ satisfying Assumption 1.
for $t = 1$ ***to*** T **do**
  Play $x_t = \operatorname{argmin}_{x \in \mathcal{X}} \sum_{\tau=1}^{t-1} f_\tau(x) + \frac{1}{\eta}\psi(x)$ and observe f_t.

To provide concrete examples, linear and convex quadratic loss functions satisfy Assumption 2 with $C = 0$. For quadratic functions in the form of $x^\top Ax + bx + c$, we note that applying Online Gradient Descent achieves $\mathcal{O}\left(\frac{d}{\alpha}\log T\right)$ standard regret (Hazan et al., 2007), where $\alpha \propto \frac{\lambda_{\min}(A+A^\top)}{\lambda_{\max}(A+A^\top)}$ can be small if the matrix $A + A^\top$ is ill-conditioned. In contrast, our alternating regret bound enjoys a better dependence on d and is independent of α. Furthermore, there are also cases where the quadratic loss is not even exp-concave (for example, when $A + A^\top$ has eigenvalue 0 with a corresponding eigenvector u such that $b^\top u \neq 0$). In such scenarios, Online Newton Step (Hazan et al., 2007) cannot apply, but our guarantees still hold. Beyond linear and quadratic cases, the logistic loss $\log(1 + \exp(-y\langle z, x\rangle))$ (with a bounded domain and bounded parameters y, z) is another example of a commonly used loss function with a constant self-concordance factor, as we show in Lemma 28. (We note that online learning with self-concordant losses has been previously considered by Zhang et al. (2017), albeit in the context of dynamic regret).

While our algorithm appears to be standard and has been studied in the literature for standard regret (see e.g., Orabona (2019)), analyzing its alternating regret requires new ideas based on the properties of the regularizer. Specifically, we prove the following general alternating regret bound.

Theorem 10 *Under Assumption 2, Algorithm 3 guarantees the following bound on* $\mathrm{Reg}_{\mathrm{Alt}}(u)$ *for any* $u \in \mathcal{X}$*,*

$$\mathcal{O}\left(\frac{\psi(u) - \psi(x_1)}{\eta} + \frac{CL^3}{\sigma^{3/2}}\eta^{3/2}T + \left(ML^3 + \frac{\beta L^2}{\sigma^2}\right)\eta^2 T\right). \tag{4}$$

Proof Define $F_t(x) \triangleq \sum_{\tau=1}^{t-1} f_\tau(x) + \frac{1}{\eta}\psi(x)$, so that $x_t = \operatorname{argmin}_{x\in\mathcal{X}} F_t(x)$. Note that

$$\begin{aligned}
-\sum_{t=1}^{T} f_t(u) &= \frac{1}{\eta}\psi(u) - \frac{1}{\eta}\psi(x_1) + F_{T+1}(x_{T+1}) - F_{T+1}(u) + \sum_{t=1}^{T}\left(F_t(x_t) - F_{t+1}(x_{t+1})\right) \\
&\leq \frac{\psi(u) - \psi(x_1)}{\eta} + \sum_{t=1}^{T}\left(F_t(x_t) - F_{t+1}(x_{t+1})\right),
\end{aligned}$$

where the last inequality comes from $F_{T+1}(x_{T+1}) \leq F_{T+1}(u)$ by the optimality of x_{T+1}. Substituting it in the alternating regret definition, we get

$$\mathrm{Reg}_{\mathrm{Alt}}(u) \leq \frac{2\left(\psi(u) - \psi(x_1)\right)}{\eta} + \sum_{t=1}^{T}\left(2F_t(x_t) - 2F_{t+1}(x_{t+1}) + f_t(x_t) + f_t(x_{t+1})\right). \tag{5}$$

Since ψ is Legendre, we must have $\nabla F_t(x_t) = \mathbf{0}$ by the optimality of x_t, where $\mathbf{0}$ is the zero vector. Further using the definition of Bregman divergence, we obtain

$$D_{F_{t+1}}(x_t, x_{t+1}) = F_{t+1}(x_t) - F_{t+1}(x_{t+1}) = F_t(x_t) + f_t(x_t) - F_{t+1}(x_{t+1})$$
$$\text{and } D_{F_t}(x_{t+1}, x_t) = F_t(x_{t+1}) - F_t(x_t) = F_{t+1}(x_{t+1}) - f_t(x_{t+1}) - F_t(x_t).$$

Therefore, the second term in the right-hand side of Eq. (5) can be written as

$$\sum_{t=1}^{T} (2F_t(x_t) - 2F_{t+1}(x_{t+1}) + f_t(x_t) + f_t(x_{t+1})) = \sum_{t=1}^{T} \left(D_{F_{t+1}}(x_t, x_{t+1}) - D_{F_t}(x_{t+1}, x_t)\right)$$
$$= \sum_{t=1}^{T} \left(D_{f_t}(x_t, x_{t+1}) + D_{F_t}(x_t, x_{t+1}) - D_{F_t}(x_{t+1}, x_t)\right).$$

Since f_t is β-smooth, the first term above, $D_{f_t}(x_t, x_{t+1})$, is at most $\frac{\beta}{2}\|x_t - x_{t+1}\|_2^2$, which can be further bounded by $\frac{\beta\eta^2 L^2}{2\sigma^2}$ due to a standard stability argument of FTRL (see Lemma 24). For the second and third terms, we proceed as

$$\begin{aligned}
&D_{F_t}(x_t, x_{t+1}) - D_{F_t}(x_{t+1}, x_t) \\
&= D_{F_t^*}(\nabla F_t(x_{t+1}), \nabla F_t(x_t)) - D_{F_t^*}(\nabla F_t(x_t), \nabla F_t(x_{t+1})) \\
&= D_{F_t^*}(\nabla F_t(x_{t+1}), \mathbf{0}) - D_{F_t^*}(\mathbf{0}, \nabla F_t(x_{t+1})) && \text{(since } \nabla F_t(x_t) = \mathbf{0}) \\
&\le \frac{1}{6}\left(\frac{2C\eta^{3/2}}{\sigma^{3/2}} + M\eta^2\right)\|\nabla F_t(x_{t+1})\|_2^3 \le \mathcal{O}\left(\left(\frac{C\eta^{3/2}}{\sigma^{3/2}} + M\eta^2\right)L^3\right).
\end{aligned}$$

Here, the first equality uses Lemma 23, a basic property of Bregman divergence of convex conjugate; the first inequality uses the fact that F_t^* is $\left(\frac{2C\eta^{3/2}}{\sigma^{3/2}} + M\eta^2\right)$-smooth of order 3 (see the technical Lemma 25), the place where we require the self-concordance of the loss functions and the 3rd-order smoothness of ψ^*, along with the aforementioned Lemma 19 taken from Wibisono et al. (2022); the last inequality comes from $\nabla F_t(x_{t+1}) = \nabla F_{t+1}(x_{t+1}) - \nabla f_t(x_{t+1}) = -\nabla f_t(x_{t+1})$ and the Lipschitzness of f_t. Plugging all bounds into Eq. (5) finishes the proof. ■

Comparing to the analysis for Theorem 3, the main difference in Theorem 10 is how we control the stability/commutator term. Specifically, instead of bounding this term using the property of KL-commutator in Theorem 3, in Theorem 10, we utilize the property of self-concordance and the 3rd-order smoothness of ψ^*. Then, by picking certain ψ according to different domain $\mathcal{X}$, we are able to control the penalty term $\psi(u) - \psi(x_1)$ with no polynomial-in-d dependency (as we will see in Section 4.2 and Section 4.3).

Focusing on the dependence on T and C, Theorem 10 yields an alternating regret bound of $\widetilde{\mathcal{O}}(\frac{1}{\eta} + (C\eta^{3/2} + \eta^2)T)$. With the optimal tuning of η, this simplifies to $\widetilde{\mathcal{O}}((CT)^{2/5} + T^{1/3})$. Specifically, when $C = 0$ (e.g., for linear and convex quadratic losses), we recover a bound of $\widetilde{\mathcal{O}}(T^{1/3})$; otherwise, the bound is of order $\widetilde{\mathcal{O}}(T^{2/5})$. We also remark that Algorithm 3 can be implemented efficiently in the standard interior-point oracle model. Consider the capped epigraph formulation $K_t \triangleq \{(x, s) : F_t(x) \le s \le U_t)\}$ where $U_t = F_t(x_0) + 1$ and $x_0 \in \mathcal{X}$,

making sure that K_t has a strictly feasible point in its interior. Assume that K_t admits a ν-self-concordant barrier and that we have access to its second-order oracle, an initial centered interior point, and the basic oracles needed to define the epigraph: a separation oracle for $\mathcal{X}$ and value/first-order oracles for F_t. Then, applying the path-following interior-point method of Nesterov (2004, Theorem 4.2.9) to $\min_{(x,s)\in K_t} s$ returns $(\widehat{x}_t, \widehat{s}_t)$ with $F_t(\widehat{x}_t) \le \widehat{s}_t$ such that $\widehat{s}_t - \min_{(x,s)\in K_t} s \le \varepsilon$ in $\mathcal{O}\left(\sqrt{\nu}\log\frac{1}{\varepsilon}\right)$ iterations. Since every convex body in $\mathbb{R}^{d+1}$ admits a self-concordant barrier with parameter $\nu = \mathcal{O}(d)$ (Bubeck and Eldan, 2015; Chewi, 2023), assuming access to the corresponding barrier oracle yields an ε-accurate update in $\mathcal{O}(\sqrt{d}\log(1/\varepsilon))$ iterations; in particular, $\varepsilon = 1/T$ requires $\mathcal{O}(\sqrt{d}\log T)$ iterations per update.

In the following, we provide concrete instantiations of ψ to fully characterize the dependence on the remaining parameters.

4.1 Entropic Barrier Regularizer for General $\mathcal{X}$

In this section, we consider a general compact and convex body $\mathcal{X}$ with a bounded diameter D. Our observation is that the *entropic barrier* of $\mathcal{X}$, proposed by Bubeck and Eldan (2015), is a valid regularizer for our algorithm as shown below.

Lemma 11 *The convex conjugate of the entropic barrier for $\mathcal{X}$,*

$$\psi^*(\theta) = \log\left(\int_{x\in\mathcal{X}} \exp\left(\langle\theta, x\rangle\right) dx\right),$$

satisfies Assumption 1 with parameters $\sigma = \frac{1}{D^2}$ and $M = D^3$.

Proof Given any $\theta \in \mathbb{R}^d$, define $p_\theta(x) = \exp\left(\langle\theta, x\rangle - \psi^*(x)\right)\mathbb{1}\left\{x \in \mathcal{X}\right\}$ to be the exponential distribution over $\mathcal{X}$ with parameter θ. Then, according to Lemma 1 of Bubeck and Eldan (2015), we know that

$$\begin{aligned}
\nabla^3\psi^*(\theta) &= \mathbb{E}_{X\sim p_\theta}\left[(X - x(\theta))\otimes(X - x(\theta))\otimes(X - x(\theta))\right], \\
\nabla^2\psi(x) &= \left(\mathbb{E}_{X\sim p_\theta}\left[(X - x)(X - x)^\top\right]\right)^{-1},
\end{aligned}$$

where $x(\theta) = \mathbb{E}_{X\sim p_\theta}[X]$. Therefore, for any $h \in \mathbb{R}^d$ such that $\|h\|_2 \le 1$, we know that

$$\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\frac{\partial^3\psi^*(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}h_ih_jh_k \le \max_{x,y\in\mathcal{X}} |\langle h, x - y\rangle|^3 \le D^3,$$

where D is the diameter of $\mathcal{X}$.
In addition, since the maximum eigenvalue of $\nabla^{-2}\psi(x) = \mathbb{E}_{X\sim p_\theta}\left[(X - x)(X - x)^\top\right]$ is upper bounded by D^2, we know that $\nabla^2\psi(x)$ has minimum eigenvalue lower bounded by $1/D^2$. Combining both arguments, we see that ψ satisfies Assumption 1 with $M = D^3$ and $\sigma = 1/D^2$. ■

Then, combining Theorem 10 and Lemma 11, we obtain the following corollary when $\mathcal{X}$ is a general compact convex body.

Corollary 12 *Let $\mathcal{X}$ be a compact convex body with diameter D. Under Assumption 2, Algorithm 3 with $\eta = \min\left\{(d\log T)^{\frac{2}{5}}(CL^3D^3)^{-\frac{2}{5}}T^{-\frac{2}{5}}, (d\log T)^{\frac{1}{3}}(L^3D^3+\beta L^2D^2)^{-\frac{1}{3}}T^{-\frac{1}{3}}\right\}$ and ψ being the entropic barrier for $\mathcal{X}$ guarantees*

$$\mathrm{Reg}_{\mathrm{Alt}} \le \mathcal{O}\left(\max\left\{(d\log T)^{\frac{3}{5}}(CL^3D^3)^{\frac{2}{5}}T^{\frac{2}{5}}, (d\log T)^{\frac{2}{3}}(L^3D^3+\beta L^2D^2)^{\frac{1}{3}}T^{\frac{1}{3}}\right\}\right).$$

Proof For any fixed $u \in \mathcal{X}$, we decompose $\mathrm{Reg}_{\mathrm{Alt}}(u)$ into

$$\mathrm{Reg}_{\mathrm{Alt}}(u') + \left[2\sum_{t=1}^{T}\left(f_t(u') - f_t(u)\right)\right], \tag{6}$$

where $u' = (1-\varepsilon)u + \varepsilon x_1 \in \mathcal{X}$ for $\varepsilon = \frac{1}{T}$ and $x_1 = \mathrm{argmin}_{x\in\mathcal{X}} \frac{\psi(x)}{\eta}$. Since f_t is convex, $f_t(u') - f_t(u) \le \frac{1}{T}\left(f_t(x_1) - f_t(u)\right) \le \frac{LD}{T}$ for all $t \in [T]$. Hence, the second term above can be bounded by $2LD$. For the first term $\mathrm{Reg}_{\mathrm{Alt}}(u')$, we apply Theorem 10 together with Lemma 11 to bound it as

$$\mathcal{O}\left(\frac{\psi(u') - \psi(x_1)}{\eta} + CL^3D^3\eta^{3/2}T + \left(L^3D^3 + \beta L^2D^2\right)\eta^2 T\right). \tag{7}$$

Finally, since ψ is an $\mathcal{O}(d)$ self-concordant barrier of $\mathcal{X}$ (Bubeck and Eldan, 2015; Chewi, 2023), we have $\psi(u') - \psi(x_1) = \mathcal{O}(d\log T)$ (see Lemma 22). Plugging in the value of η then finishes the proof. ■

Again, for loss functions with self-concordance parameter $C = 0$ (e.g., linear or convex quadratic losses), we achieve $\widetilde{\mathcal{O}}(d^{\frac{2}{3}}T^{\frac{1}{3}})$ alternating regret, matching the performance of continuous Hedge. In fact, for linear losses, the two algorithms are simply equivalent (Bubeck and Eldan, 2015), but this equivalence does not extend to general convex losses.

4.2 The ℓ_2 Ball Case

Next, we consider the case where the feasible domain $\mathcal{X} = \mathcal{B}_2^d(1)$ and pick $\psi(x) = -\log(1 - \|x\|_2^2)$. The following lemma shows that ψ satisfies Assumption 1.

Lemma 13 *The convex conjugate of ψ is defined as $\psi^*(w) = \sqrt{1+\|w\|_2^2} - \log(1 + \sqrt{1+\|w\|_2^2})$, and it satisfies Assumption 1 with parameters $\sigma = 2$ and $M = 4$.*

Proof It is direct to see that $\psi(x) = -\log\left(1 - \|x\|_2^2\right)$ is a Legendre function. Now, we show that it is 2-strongly convex with respect to the ℓ_2-norm. Direct calculation shows that

$$\nabla\psi(x) = \frac{2x}{1-\|x\|_2^2},$$
$$\nabla^2\psi(x) = \frac{2\mathbf{I}}{1-\|x\|_2^2} + \frac{4xx^\top}{\left(1-\|x\|_2^2\right)^2},$$

where $\mathbf{I}$ represents the d-dimensional identity matrix. Thus, for any $h \in \mathbb{R}^d$ and $\|h\|_2 = 1$, we have

$$\nabla^2\psi(x)[h,h] = \frac{2\|h\|_2^2}{1-\|x\|_2^2} + \frac{4\left(\langle x, h\rangle\right)^2}{\left(1-\|x\|_2^2\right)^2} \ge 2.$$

It remains to show that ψ^* is 4-smooth of order 3 with respect to the ℓ_2-norm. Below, we calculate the derivatives of ψ^*:

$$\nabla\psi^*(w) = \frac{w}{1+\sqrt{1+\|w\|_2^2}},$$
$$\nabla^2\psi^*(w) = \frac{I}{1+\sqrt{1+\|w\|_2^2}} - \frac{ww^\top}{\left(1+\sqrt{1+\|w\|_2^2}\right)^2\cdot\sqrt{1+\|w\|_2^2}}.$$

Denoting $g(w) = \sqrt{1+\|w\|_2^2}$, we have

$$\nabla^2\psi^*(w) = \frac{\mathbf{I}}{1+g(w)} - \frac{ww^\top}{(1+g(w))^2\cdot g(w)}.$$

Then, we have

$$\nabla^3\psi^*(w) = -\frac{\mathrm{Sym}(\mathbf{I}\otimes w)}{g(w)\,(1+g(w))^2} + \frac{(1+3g(w))\,w\otimes w\otimes w}{(1+g(w))^3\,g^3(w)}.$$

Here, $\mathrm{Sym}(\mathbf{I}\otimes w)$ denotes symmetrization over all permutations of the indices of the tensor product $I\otimes w$. Then, we check the condition in Definition 7 for any $h\in\mathbb{R}^d$ and $\|h\|_2 = 1$.

$$\begin{aligned}
&\sum_{i=1}^d\sum_{j=1}^d\sum_{k=1}^d \frac{\partial^3\psi(w)}{\partial w_i\partial w_j\partial w_k}h_ih_jh_k \\
&= \sum_{i=1}^d\sum_{j=1}^d\sum_{k=1}^d -\frac{(\delta_{ij}w_k+\delta_{ik}w_j+\delta_{jk}w_i)\,h_ih_jh_k}{g(w)\,(1+g(w))^2} + \frac{(1+3g(w))\,w_ih_iw_jh_jw_kh_k}{g^3(w)\,(1+g(w))^3} \\
&= -\frac{3\sum_{i=1}^d w_ih_i\cdot\sum_{k=1}^d h_k^2}{g(w)\,(1+g(w))^2} + \frac{(1+3g(w))\left(\sum_{i=1}^d w_ih_i\right)^3}{g^3(w)\,(1+g(w))^3}.
\end{aligned}$$

By Cauchy-Schwarz inequality, we have $|\langle w,h\rangle| \le \|w\|_2\cdot\|h\|_2$. Combining with the fact that $\|h\|_2 = 1$, we know that

$$\begin{aligned}
&\sum_{i=1}^d\sum_{j=1}^d\sum_{k=1}^d \frac{\partial^3\psi(w)}{\partial w_i\partial w_j\partial w_k}h_ih_jh_k \\
&\le \frac{3\|w\|_2}{g(w)\,(1+g(w))^2} + \frac{(1+3g(w))\,\|w\|_2^3}{g^3(w)\,(1+g(w))^3}.
\end{aligned}$$

Recall that $g(w) = \sqrt{1+\|w\|_2^2}$, it holds that $g(w)\ge\max(1,\|w\|_2)$. Therefore, we have

$$\sum_{i=1}^d\sum_{j=1}^d\sum_{k=1}^d \frac{\partial^3\psi(w)}{\partial w_i\partial w_j\partial w_k}h_ih_jh_k \le 4.$$

■

Thus, we get the following dimension independent regret bound from Theorem 10.

Corollary 14 *Let $\mathcal{X} = \mathcal{B}_2^d(1)$. Under Assumption 2, Algorithm 3 with regularizer $\psi(x) = -\log(1-\|x\|_2^2)$ and $\eta = \min\left\{(\log T)^{\frac{2}{5}}(CL^3)^{-\frac{2}{5}}T^{-\frac{2}{5}}, (\log T)^{\frac{1}{3}}(L^3+\beta L^2)^{-\frac{1}{3}}T^{-\frac{1}{3}}\right\}$ guarantees $\mathrm{Reg}_{\mathrm{Alt}} \le \mathcal{O}\left(\max\left\{(\log T)^{\frac{3}{5}}(CL^3)^{\frac{2}{5}}T^{\frac{2}{5}}, (\log T)^{\frac{2}{3}}(L^3+\beta L^2)^{\frac{1}{3}}T^{\frac{1}{3}}\right\}\right)$.*

Proof We decompose $\mathrm{Reg}_{\mathrm{Alt}}(u)$ for any $u \in \mathcal{X}$, as shown in Eq. (6), and choose u' similarly. The second term can be bounded by $4L$ and the first term by Eq. (7) again. Since ψ is a 1-self-concordant barrier (shown in section 6.2.1 of (Nesterov, 1994)), we have $\psi(u') - \psi(x_1) = \mathcal{O}(\log T)$ by Lemma 22. Finally, plugging in $\sigma = 2$ and $M = 4$ and using the value of η leads us to the final regret bound. ■

As a direct consequence of Corollary 14, our algorithm yields dimension-free alternating regret bounds for several standard loss functions. First, when the loss functions f_t are linear or quadratic (where $C = 0$), we obtain an $\widetilde{\mathcal{O}}(T^{\frac{1}{3}})$ bound. This completely removes the dimension dependence, strictly improving over the $\widetilde{\mathcal{O}}(d^{\frac{2}{3}}T^{\frac{1}{3}})$ bound achieved by Algorithm 2 as shown in Theorem 4. Furthermore, consider the case where the loss functions are the logistic loss, defined as $f_t(x) = \log\big(1+\exp(-y_t\langle z_t, x\rangle)\big)$ with bounded domains $x \in \mathcal{X} \triangleq \{x \in \mathbb{R}^d : \|x\|_2 \le 1\}$, $y_t \in \{-1, +1\}$, and $\|z_t\|_2 \le 1$. Under these conditions, we show that f_t is $\sinh(1/2)$-self-concordant on $\mathcal{X}$ (see Appendix C for the proof). This implies that we achieve an $\widetilde{\mathcal{O}}(T^{\frac{2}{5}})$ alternating regret guarantee for the logistic loss, which, importantly, remains independent of the dimension d.

4.3 The Simplex Case

Finally, we consider the case where the feasible domain is the $(d-1)$-dimensional simplex Δ_d. Specifically, to represent Δ_d as a convex body in $(d-1)$-dimensional space, we set $\mathcal{X} = \{x \in \mathbb{R}_+^{d-1}, \sum_{i=1}^{d-1} x_i \le 1\}$. We instantiate Algorithm 3 with the entropy regularizer $\psi(x) = \sum_{i=1}^{d-1} x_i \log x_i + (1 - \sum_{i=1}^{d-1} x_i)\log(1 - \sum_{i=1}^{d-1} x_i)$. The following lemma shows that ψ satisfies Assumption 1 with $\sigma = 1$ and $M = 8$.

Lemma 15 *$\psi(x)$ is Legendre and 1-strongly convex with respect to ℓ_2-norm. The convex conjugate of $\psi(x)$ is $\psi^*(w) = \log\left(1 + \sum_{i=1}^{d-1} e^{w_i}\right)$, which is 8-smooth of order 3 with respect to ℓ_2-norm.*

Proof Direct calculation shows that the gradient and Hessian of $\psi(x)$ are as follows:

$$\frac{\partial \psi}{\partial x_j} = \log x_j + 1 - \log(1 - \sum_{i=1}^{d-1} x_i) - 1 = \log x_j - \log(1 - \sum_{i=1}^{d-1} x_i),$$

$$\nabla^2\psi(x) = \begin{bmatrix} \frac{1}{x_1} & & & \\ & \frac{1}{x_2} & & \\ & & \cdots & \\ & & & \frac{1}{x_{d-1}} \end{bmatrix} + \frac{1}{1 - \sum_{i=1}^{d-1} x_i} \cdot \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \succeq \begin{bmatrix} \frac{1}{x_1} & & & \\ & \frac{1}{x_2} & & \\ & & \cdots & \\ & & & \frac{1}{x_{d-1}} \end{bmatrix}.$$

Since $x_i \le 1$ for all $i \in [d-1]$, we have $\nabla^2\psi(x) \succeq I$, meaning that $\psi(x)$ is 1-strongly convex with respect to $\|\cdot\|_2$.

Direct calculation also shows that $\psi^*(w) = \log\left(1 + \sum_{i=1}^{d-1}\exp(w_i)\right)$ is the convex conjugate of $\psi(x)$. Define a distribution p_w over $[d]$ where $p_w(i) = \exp(w_i - \psi^*(w))$ for $i \in [d-1]$ and $p_w(d) = \exp(-\psi^*(w))$. For any $v \in \mathbb{R}^{d-1}$, $\nabla^3\psi^*(w)[v,v,v]$ can be calculated as follows:

$$\nabla^3\psi^*(w)[v,v,v] = \mathbb{E}_{I\sim p_w}[(v_I - \bar{v})^3],$$

where $v_d = 0$ and $\bar{v} = \sum_{i=1}^{d-1} v_i \cdot p_w(i)$. Therefore, when $\|v\|_2 = 1$, we have $|v_I - \bar{v}| \le 2$ for all $I \in [d]$ and

$$|\nabla^3\psi^*(w)[v,v,v]| \le \mathbb{E}_{I\sim p_w}[|v_I - \bar{v}|^3] \le 2^3 = 8.$$

This implies that ψ^* is 8-smooth of order 3 with respect to ℓ_2-norm. ■

Combining Theorem 10 and Lemma 15 leads to the following guarantee for the simplex case, which has no polynomial dependence on the dimension.

Corollary 16 *Let $\mathcal{X} = \{x \in \mathbb{R}_+^{d-1}, \sum_{i=1}^{d-1} x_i \le 1\}$. Under Assumption 2, Algorithm 3 with $\eta = \min\left\{(\log d)^{\frac{2}{5}}(CL^3)^{-\frac{2}{5}}T^{-\frac{2}{5}}, (\log d)^{\frac{1}{3}}(L^3+\beta L^2)^{-\frac{1}{3}}T^{-\frac{1}{3}}\right\}$ and regularizer $\psi(x) = \sum_{i=1}^{d-1} x_i \log x_i + (1 - \sum_{i=1}^{d-1} x_i)\log(1 - \sum_{i=1}^{d-1} x_i)$ guarantees*

$$\mathrm{Reg}_{\mathrm{Alt}} \le \mathcal{O}\left(\max\left\{(\log d)^{\frac{3}{5}}(CL^3)^{\frac{2}{5}}T^{\frac{2}{5}}, (\log d)^{\frac{2}{3}}(L^3+\beta L^2)^{\frac{1}{3}}T^{\frac{1}{3}}\right\}\right).$$

Proof Using Theorem 10, the alternating regret of Hedge can be bounded by:

$$\mathrm{Reg}_{\mathrm{Alt}} \le \mathcal{O}\left(\frac{B_\psi}{\eta} + CL^3\eta^{3/2}T + \left(L^3 + \beta L^2\right)\eta^2 T\right),$$

where $B_\psi = \max_{u_1,u_2\in\mathcal{X}}(\psi(u_1) - \psi(u_2))$. The maximum value of the negative entropy regularizer is 0, attained at a pure strategy point, and the minimum value is $-\log d$, attained by the uniform distribution over the actions. Thus, $B_\psi = \log d$. Plugging in the value of η finishes the proof. ■

Once again, for losses with $C = 0$, our regret bound is of order $\widetilde{\mathcal{O}}(T^{\frac{1}{3}})$ (with only logarithmic dependence on d hidden in $\widetilde{\mathcal{O}}(\cdot)$), improving over the $\widetilde{\mathcal{O}}(d^{\frac{2}{3}}T^{\frac{1}{3}})$ bound of continuous Hedge.

5. Alternating Regret Lower Bounds

While our work significantly advances our understanding on alternating regret upper bounds, there are no existing alternating regret lower bounds at all. Towards closing this gap, we make an initial attempt by considering the special case of the expert problem (OLO over simplex Δ_d) and providing two algorithm-specific lower bounds. In the first result, echoing Theorem 3, we show that the worst-case alternating regret of Hedge is $\Omega(T^{\frac{1}{3}})$, thereby giving a tight characterization of Hedge's alternating regret (at least for the case with a fixed learning rate).

Theorem 17 *For $d \geq 3$ and any $\eta > 0$, there exists a loss sequence $\{\ell_t\}_{t=1}^T$ where $\ell_t \in [0,1]^d$ for all $t \in [T]$, such that Hedge suffers $\mathrm{Reg}_{\mathrm{Alt}} = \Omega(T^{\frac{1}{3}})$.*

The full proof is deferred to Appendix D. To provide a sketch, we consider two environments for $d = 3$. In the first environment, the loss vectors are cycling among the three one-hot loss vectors, i.e. $\ell_{3k-2} = (1,0,0), \ell_{3k-1} = (0,1,0), \ell_{3k} = (0,0,1)$ for $k = 1, 2, \ldots, \frac{T}{3}$. We then calculate the closed form of p_t, which also turns out to cycle among three vectors. Direct calculation then shows that Hedge achieves an $\Omega(\eta^2 T)$ alternating regret. In the second environment, the loss vectors are all $(1,0,0)$. Via a direct calculation, we show that the alternating regret is $\Omega(\frac{1}{\eta})$. Combining both cases proves that the worst case alternating regret is $\Omega(\max\{\eta^2 T, \frac{1}{\eta}\}) = \Omega(T^{\frac{1}{3}})$.

We then move on to another widely-used algorithm called Predictive Regret Matching$^+$ (PRM$^+$) (Farina et al., 2021). Cai et al. (2025) show that, unlike its other variants, PRM$^+$ coupled with alternation has empirically shown fast last-iterate convergence in two-player zero-sum games. However, the reason behind this is not well known. One may wonder whether the fast convergence rate of PRM$^+$ is due to small alternating regret in the adversarial setting. Unfortunately, we show that this is not the case: in the adversarial setting, PRM$^+$ actually achieves $\Omega(\sqrt{T})$ alternating regret in the worst case, which is even worse than Hedge. In fact, the same is true for another popular algorithm called Optimistic Online Gradient Descent (OOGD) (Rakhlin and Sridharan, 2013), even though it is well known that in a game setting with simultaneous play, OOGD achieves $o(\sqrt{T})$ regret. We summarize these lower bounds in the following theorem.

Theorem 18 *There exists a loss sequence $\{\ell_t\}_{t=1}^T$ where $\ell_t \in [-4,4]^d$ for all $t \in [T]$, such that $\mathrm{Reg}_{\mathrm{Alt}} = \Omega(\sqrt{T})$ for OOGD and PRM$^+$.*

The proof is deferred to Appendix D. Specifically, the loss sequence we create for both algorithms is alternating between $(4,0)$ and $(-2,0)$. Then, we show that for OOGD with learning rate $\eta > 0$, when $t \leq \Theta(\frac{1}{\eta})$, the alternating regret is $\Theta(\frac{1}{\eta})$; when $t \geq \Theta(\frac{1}{\eta})$, p_t starts to alternate between $[2\eta, 1-2\eta]$ and $[0,1]$, leading to $\Theta(\eta T)$ alternating regret. Combining both cases, the overall alternating regret is $\Omega(\frac{1}{\eta} + \eta T) = \Omega(\sqrt{T})$. For PRM$^+$, we can also show that under this loss sequence, the prediction of PRM$^+$ converges to the optimal decision $(0,1)$ at a rate of $1/\sqrt{T}$ approximately, leading to an $\Omega(\sqrt{T})$ alternating regret. The implication of this result is that, to demystify the practical success of PRM$^+$, one must take the game structure into account.

6. Conclusion and Open Problems

In this work, following up on Cevher et al. (2023) and addressing their main open question, we show that $\mathcal{O}(T^{1/3})$ alternating regret is possible for general Online Convex Optimization via simple algorithms. The main open problem remains to be: what the minimax optimal alternating regret is for OLO/OCO? This appears to be highly nontrivial even for very special cases — for example, even for 1-dimensional OLO over $[-1,1]$ where $\mathcal{O}(\log T)$ alternating regret is achieved by Cevher et al. (2023), figuring out whether the optimal bound is $\Theta(\log T)$ or $\Theta(1)$ appears to require new techniques.

Acknowledgments

HL thanks Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, and Weiqiang Zheng for discussions related to regret matching with alternation. He is supported by NSF awards IIS-1943607 and CCF-2551666.

Appendix A. Omitted Details in Section 3

In this section, we show the omitted proofs in Section 3. For completeness, we first include Lemma A.2 in Wibisono et al. (2022), which provides an upper bound for the Bregman commutator for functions with 3-rd order smoothness.

Lemma 19 (Lemma A.2 of Wibisono et al. (2022)) *Assuming ψ^* is M-smooth of order 3 with respect to norm $\|\cdot\|_2$, we have for all $w, w' \in \mathbb{R}^d$,*

$$|D_{\psi^*}(w, w') - D_{\psi^*}(w', w)| \le \frac{M}{6}\|w - w'\|_2^3.$$

In the following, we provide the proof of Theorem 3.

Proof [Proof of Theorem 3] We aim to show that

$$\langle p_t - u, \ell_t \rangle = \frac{1}{\eta}\left(\mathrm{KL}(u, p_t) - \mathrm{KL}(u, p_{t+1}) + \mathrm{KL}(p_t, p_{t+1})\right), \tag{8}$$

$$\langle p_{t+1} - u, \ell_t \rangle = \frac{1}{\eta}\left(\mathrm{KL}(u, p_t) - \mathrm{KL}(u, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)\right). \tag{9}$$

Note that the dynamic of Hedge implies that $p_{t+1,i} = \frac{p_{t,i}\exp(-\eta\ell_{t,i})}{\sum_{j=1}^d p_{t,j}\exp(-\eta\ell_{t,j})}$. Direct calculation shows that

$$\begin{aligned}
&\frac{1}{\eta}\left(\mathrm{KL}(u, p_t) - \mathrm{KL}(u, p_{t+1}) + \mathrm{KL}(p_t, p_{t+1})\right) \\
&= \frac{1}{\eta}\sum_{i=1}^d (p_{t,i} - u_i)\cdot \log\frac{p_{t,i}}{p_{t+1,i}} \\
&= \sum_{i=1}^d (p_{t,i} - u_i)\left(\ell_{t,i} + \frac{1}{\eta}\log\left(\sum_{j=1}^d p_{t,j}\exp(-\eta\ell_{t,j})\right)\right) \\
&= \langle p_t - u, \ell_t\rangle, \\
&\frac{1}{\eta}\left(\mathrm{KL}(u, p_t) - \mathrm{KL}(u, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)\right) \\
&= \frac{1}{\eta}\sum_{i=1}^d (p_{t+1,i} - u_i)\cdot \log\frac{p_{t,i}}{p_{t+1,i}} \\
&= \sum_{i=1}^d (p_{t+1,i} - u_i)\left(\ell_{t,i} + \frac{1}{\eta}\log\left(\sum_{j=1}^d p_{t,j}\exp(-\eta\ell_{t,j})\right)\right) \\
&= \langle p_{t+1} - u, \ell_t\rangle.
\end{aligned}$$

Combining Eq. (8) and Eq. (9) for all $t \in [T]$, we obtain that

$$\mathrm{Reg}_{\mathrm{Alt}}(u) = \frac{2(\mathrm{KL}(u, p_1) - \mathrm{KL}(u, p_{T+1}))}{\eta} + \frac{1}{\eta}\sum_{t=1}^{T}\left(\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)\right). \tag{10}$$

Since p_1 is the uniform distribution over all d experts, we can bound $\mathrm{KL}(u, p_1) - \mathrm{KL}(u, p_{T+1}) \le \log d$. To control the term $\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)$, define $G(x) = \log(\sum_{i=1}^{d}\exp(x_i))$ for $x \in \mathbb{R}^d$ and $L_t = \sum_{\tau \le t}\ell_\tau$. Direct calculation shows that

$$\begin{aligned}
&D_G(-\eta L_t, -\eta L_{t-1}) - D_G(-\eta L_{t-1}, -\eta L_t)\\
&= 2\log\left(\sum_{i=1}^{d}\exp(-\eta L_{t,i})\right) - 2\log\left(\sum_{i=1}^{d}\exp(-\eta L_{t-1,i})\right)\\
&\quad + \sum_{i=1}^{d} p_{t+1,i}\cdot(\eta\ell_{t,i}) + \sum_{i=1}^{d} p_{t,i}\cdot(\eta\ell_{t,i})\\
&= \mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t).
\end{aligned}$$

Applying Lemma A.2 of Wibisono et al. (2022), we know that $\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t) \le \frac{4}{3}\|\eta\ell_t\|_\infty^3 \le \frac{4}{3}\eta^3$, where the first inequality is because $G(x)$ is 8-smooth of order 3 as proven in Example A.3 Wibisono et al. (2022). Plugging the above to Eq. (10) and picking the optimal η finishes the proof. ■

In the following, we show the proof for Theorem 4.

Proof [Proof of Theorem 4] By definition, we know that $p_t(x) = \frac{1}{Z_t}\exp(-\eta F_{t-1}(x))$ and $p_{t+1}(x) = \frac{1}{Z_{t+1}}\exp(-\eta F_t(x))$, where $Z_t = \int_{x\in\mathcal{X}}\exp(-\eta F_{t-1}(x))dx$ and $F_t(x) = \sum_{\tau\le t} f_\tau(x)$. Then, direct calculation shows that

$$\begin{aligned}
\frac{1}{\eta}\mathrm{KL}(p_t, p_{t+1}) &= \frac{1}{\eta}\int_{x\in\mathcal{X}} p_t(x)\log\frac{p_t(x)}{p_{t+1}(x)}dx\\
&= \int_{x\in\mathcal{X}} p_t(x)\left(f_t(x) + \frac{1}{\eta}\log\frac{Z_{t+1}}{Z_t}\right)dx\\
&= \int_{x\in\mathcal{X}} p_t(x) f_t(x)dx + \frac{1}{\eta}\left(\log Z_{t+1} - \log Z_t\right)\\
\frac{1}{\eta}\mathrm{KL}(p_{t+1}, p_t) &= \frac{1}{\eta}\int_{x\in\mathcal{X}} p_{t+1}(x)\log\frac{p_{t+1}(x)}{p_t(x)}dx\\
&= \int_{x\in\mathcal{X}} p_{t+1}(x)\left(-f_t(x) + \frac{1}{\eta}\log\frac{Z_t}{Z_{t+1}}\right)dx\\
&= -\int_{x\in\mathcal{X}} p_{t+1}(x) f_t(x)dx + \frac{1}{\eta}(\log Z_t - \log Z_{t+1})
\end{aligned}$$

Combining the above two inequalities, we have

$$\int_{x\in\mathcal{X}}(p_t(x) + p_{t+1}(x))f_t(x)dx + \frac{2}{\eta}(\log Z_{T+1} - \log Z_1) = \frac{1}{\eta}\left(\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)\right). \tag{11}$$

According to the proof of Theorem 3.1 in Bubeck (2011), we know that

$$\log Z_{T+1} - \log Z_1 \geq -d \log \frac{1}{\gamma} - \eta \sum_{t=1}^{T} f_t(u) - \eta T \gamma,$$

for any $u \in \mathcal{X}$ and $\gamma \in [0,1]$. Picking $\gamma = \frac{1}{T}$ and using the fact that $\int_{x\in\mathcal{X}} p_t(x) f_t(x) dx \geq f_t(x_t)$ and $\int_{x\in\mathcal{X}} p_{t+1}(x) f_t(x) dx \geq f_t(x_{t+1})$ due to convexity, we know that

$$\mathrm{Reg}_{\mathrm{Alt}} \leq \frac{d \log T}{\eta} + \frac{1}{\eta} \sum_{t=1}^{T} \left(\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t) \right). \tag{12}$$

Next, we bound $\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t)$. Define $H(F) = \log \int_{x\in\mathcal{X}} \exp(-\eta F(x)) dx$. By definition of $\mathrm{KL}(p,q)$, we know that

$$\mathrm{KL}(p_t, p_{t+1}) = \int_{x\in\mathcal{X}} p_t(x) \log \frac{p_t(x)}{p_{t+1}(x)} dx = H(F_{t+1}) - H(F_t) + \eta \cdot \mathbb{E}_{x\sim p_t} \left[f_t(x) \right]. \tag{13}$$

Define $p_{t,s}$ to be the distribution where $p_{t,s}(x) \propto \exp(-\eta((1-s)F_{t-1}(x) + sF_t(x)))$ and $Z_{t,s} = \int_{x\in\mathcal{X}} \exp(-\eta((1-s)F_{t-1}(x) + sF_t(x))) dx$. By definition, we have $p_{t,0} = p_t$, $p_{t,1} = p_{t+1}$, $Z_{t,0} = Z_t$ and $Z_{t,1} = Z_{t+1}$. Direct calculation shows that

$$\begin{aligned}
\nabla_s \log Z_{t,s} &= \frac{\nabla_s Z_{t,s}}{Z_{t,s}} = \frac{1}{Z_{t,s}} \int_{x\in\mathcal{X}} -\eta f_t(x) \exp(-\eta((1-s)F_{t-1}(x) + sF_t(x))) dx \\
&= -\mathbb{E}_{x\sim p_{t,s}}[\eta f_t(x)], \qquad (14) \\
\nabla_s^2 \log Z_{t,s} &= \frac{\int_{x\in\mathcal{X}} \eta^2 f_t(x)^2 \exp(-\eta((1-s)F_{t-1}(x) + sF_t(x))) dx}{Z_{t,s}} \\
&\quad - \left(\frac{\int_{x\in\mathcal{X}} \eta f_t(x) \exp(-\eta((1-s)F_{t-1}(x) + sF_t(x))) dx}{Z_{t,s}} \right)^2 \\
&= \mathrm{Var}_{x\sim p_{t,s}}[\eta f_t(x)]. \qquad (15)
\end{aligned}$$

According to Fundamental theorem of calculus, we know that

$$\begin{aligned}
\log Z_{t,1} &= \log Z_{t,0} + \int_0^1 \nabla_s \log Z_{t,s} ds \\
&= \log Z_{t,0} + [\nabla_s \log Z_{t,s}]_{s=0} + \int_0^1 \int_0^\alpha \nabla_s^2 \log Z_{t,s} ds d\alpha \\
&= \log Z_{t,0} + [\nabla_s \log Z_{t,s}]_{s=0} + \int_0^1 \left(\int_s^1 d\alpha \right) \nabla_s^2 \log Z_{t,s} ds \\
&= \log Z_{t,0} + [\nabla_s \log Z_{t,s}]_{s=0} + \int_0^1 (1-s) \nabla_s^2 \log Z_{t,s} ds.
\end{aligned}$$

Then, based on Eq. (13), we can obtain that

$$\mathrm{KL}(p_t, p_{t+1}) = \log Z_{t,1} - \log Z_{t,0} - [\nabla_s \log Z_{t,s}]_{s=0}$$

$$= \int_0^1 (1-s)[\nabla_s^2 \log Z_{t,s}]ds$$
$$= \int_0^1 (1-s)\mathrm{Var}_{x\sim p_{t,s}}[\eta f_t(x)]ds,$$

where the third equality is due to Eq. (15). Similarly, we can obtain that

$$\begin{aligned}\mathrm{KL}(p_{t+1}, p_t) &= \log Z_{t,0} - \log Z_{t,1} - [\nabla_s \log Z_{t,s}]_{s=1} \\ &= \int_0^1 (1-s)\mathrm{Var}_{x\sim p_{t,1-s}}[-\eta f_t(x)]ds \\ &= \int_0^1 s'\mathrm{Var}_{x\sim p_{t,s'}}[\eta f_t(x)]ds',\end{aligned}$$

where the last equality is by a variable replacement $s' = 1 - s$ and $\mathrm{Var}[x] = \mathrm{Var}[-x]$.

Combining the above two equalities, we know that

$$\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t) = \int_0^1 (1-2s)\mathrm{Var}_{x\sim p_{t,s}}[\eta f_t(x)]ds. \tag{16}$$

To further analyze $\int_0^1 (1-2s)\mathrm{Var}_{x\sim p_{t,s}}[\eta f_t(x)]ds$, let $v = s - \frac{1}{2}$ and $p_{t,v} \propto \exp(-(\frac{1}{2} - v)\eta F_t(x) - (\frac{1}{2} + v)\eta F_{t+1}(x))$ with an abuse of notation. Then, we have

$$\begin{aligned}&\frac{1}{2}\int_0^1 (1-2s)\mathrm{Var}_{x\sim p_{t,s}}[\eta f_t(x)]ds \\ &= -\int_{-\frac{1}{2}}^{\frac{1}{2}} v \cdot \mathrm{Var}_{x\sim p_{t,v}}[\eta f_t(x)]dv \\ &= -\int_0^{\frac{1}{2}} v \cdot \mathrm{Var}_{x\sim p_{t,v}}[\eta f_t(x)]dv - \int_{-\frac{1}{2}}^0 v \cdot \mathrm{Var}_{x\sim p_{t,v}}[\eta f_t(x)]dv \\ &= \int_0^{\frac{1}{2}} v\left(\mathrm{Var}_{x\sim p_{t,-v}}[\eta f_t(x)] - \mathrm{Var}_{x\sim p_{t,v}}[\eta f_t(x)]\right)dv.\end{aligned} \tag{17}$$

We then show how to bound $\mathrm{Var}_{x\sim p_{t,-v}}[\eta f_t(x)] - \mathrm{Var}_{x\sim p_{t,v}}[\eta f_t(x)]$. For notational convenience, let $p_{t,-v} = q_1$ and $p_{t,v} = q_2$.

Direct calculation shows that

$$\begin{aligned}&\mathrm{Var}_{x\sim q_1}[\eta f_t(x)] - \mathrm{Var}_{x\sim q_2}[\eta f_t(x)] \\ &= \eta^2\left(\mathbb{E}_{q_1}[f_t(x)^2] - \mathbb{E}_{q_2}[f_t(x)^2]\right) - \eta^2(\mathbb{E}_{q_1}[f_t(x)] + \mathbb{E}_{q_2}[f_t(x)])(\mathbb{E}_{q_1}[f_t(x)] - \mathbb{E}_{q_2}[f_t(x)]) \\ &\le \eta^2 \max_{x\in\mathcal{X}} f_t^2(x)\|q_1 - q_2\|_{TV} + 2\eta^2 \max_{x\in\mathcal{X}} f_t^2(x)\|q_1 - q_2\|_{TV} \\ &\le 3\eta^2\|q_1 - q_2\|_{TV},\end{aligned} \tag{18}$$

where we use $f_t(x) \in [-1, 1]$. To further bound $\|q_1 - q_2\|_{TV}$, we define $G(x) = -\frac{1}{2}(F_t(x) + F_{t+1}(x))$ for notational convenience. By definition of q_1 and q_2, we know that for any $x \in \mathcal{X}$

$$\frac{q_1(x)}{q_2(x)} = \frac{\int_{x'} \exp(G(x') + \eta v f_t(x'))dx'}{\int_{x'} \exp(G(x') - \eta v f_t(x'))dx'} \cdot \frac{\exp(G(x) - \eta v f_t(x))}{\exp(G(x) + \eta v f_t(x))}$$

$$\leq \frac{\int_{x'} \exp(G(x') - \eta v f_t(x'))dx'}{\int_{x'} \exp(G(x') - \eta v f_t(x'))dx'} \cdot \frac{\exp(G(x) + v f_t(x))}{\exp(G(x) + v f_t(x))} \exp(\eta)$$
$$= \exp(\eta),$$

where the inequality is because $v \leq \frac{1}{2}$ and $f_t(x) \in [-1, 1]$. Similarly, we can also show that

$$\frac{q_1(x)}{q_2(x)} = \frac{\int_{x'} \exp(G(x') + \eta v f_t(x'))dx'}{\int_{x'} \exp(G(x') - \eta v f_t(x'))dx'} \cdot \frac{\exp(G(x) - \eta v f_t(x))}{\exp(G(x) + \eta v f_t(x))}$$
$$\geq \frac{\int_{x'} \exp(G(x') - \eta v f_t(x'))dx'}{\int_{x'} \exp(G(x') - \eta v f_t(x'))dx'} \cdot \frac{\exp(G(x) + v f_t(x))}{\exp(G(x) + v f_t(x))} \exp(-\eta)$$
$$= \exp(-\eta).$$

Therefore, we know that $\|q_1 - q_2\|_{TV} \leq \exp(\eta) - 1 \leq 2\eta$ since $\eta \leq 1$. Combining Eq. (16), Eq. (17), and Eq. (18), we conclude that

$$\mathrm{KL}(p_t, p_{t+1}) - \mathrm{KL}(p_{t+1}, p_t) = \mathcal{O}(\eta^3). \tag{19}$$

Combining Eq. (12) and Eq. (19) and picking η optimally finishes the proof.

■

A.1 Discussion on the runtime complexity of continuous Hedge

In this subsection, we analyze the runtime complexity of Algorithm 2. Implementing continuous Hedge requires access to a log-concave density in each round, since the loss functions are convex. Given a convex body $\mathcal{X}$ and a target total variation (TV) error $\varepsilon_{\mathrm{TV}}$ per round, in order to sample from an $\varepsilon_{\mathrm{TV}}$-TV approximation of the log-concave target distribution, one may employ the Hit-and-Run walk in $\mathbb{R}^d$ (Lovász and Vempala, 2006a). Under the standard oracle model for log-concave sampling, including membership access to $\mathcal{X}$ and value access to the unnormalized log-density, Theorem 1.1 of Lovász and Vempala (2006a) bounds the TV mixing time by a polynomial in the dimension d and the geometric conditioning R/r, where $\mathcal{X}$ contains a ball of radius r and is contained in a ball of radius R, with only a polylogarithmic dependence on $1/\varepsilon_{\mathrm{TV}}$. Specifically, after $m = \widetilde{\mathcal{O}}(d^2(R/r)^2 \operatorname{polylog}(1/\varepsilon_{\mathrm{TV}}))$ steps, the chain is within $\mathcal{O}(\varepsilon_{\mathrm{TV}})$ TV distance from the target distribution. Under standard rounding or well-roundedness assumptions where $R/r = \mathcal{O}(\sqrt{d})$, this process yields $\mathcal{O}(d^3 \operatorname{polylog}(1/\varepsilon_{\mathrm{TV}}))$ steps per sample. This per-sample cost is sufficient for a randomized implementation with an expected regret guarantee. Taking $\varepsilon_{\mathrm{TV}} = 1/T$ makes the accumulated TV error negligible, and the per-round sampler-oracle complexity is $\mathcal{O}(d^3 \operatorname{polylog} T)$.

This per-sample cost, however, does not by itself compute the action x_t used by Algorithm 2, which is the mean of p_t. To implement the algorithm as stated, one needs to approximate $x_t = \mathbb{E}_{X \sim p_t}[X]$. Standard Hoeffding's inequality shows that an estimate of this mean with error at most $\varepsilon_{\mathrm{mean}}$ generally requires $\widetilde{\mathcal{O}}(1/\varepsilon_{\mathrm{mean}}^2)$ approximately independent samples. Under an additional Lipschitzness assumption on the losses, playing an $\varepsilon_{\mathrm{mean}}$-accurate approximation of x_t incurs $O(T\varepsilon_{\mathrm{mean}})$ extra loss over T rounds. Taking $\varepsilon_{\mathrm{mean}} = T^{-2/3}$ makes this approximation error $\mathcal{O}(T^{1/3})$, which is absorbed by the $T^{1/3}$-scale alternating regret bound. This requires $\widetilde{\mathcal{O}}(T^{4/3})$ samples per round, which leads to a per-round runtime complexity of $\widetilde{\mathcal{O}}(T^{4/3}d^3)$.

Appendix B. Preliminary for Self-Concordant Barrier

In this section, we provide a preliminary for self-concordant barrier functions that is useful in our analysis.

Definition 20 *Let $\psi : \mathrm{int}(\Omega) \to \mathbb{R}$ be a C^3-smooth convex function. ψ is called a self-concordant barrier on Ω if it satisfies:*

- *$\psi(x_i) \to \infty$ as $i \to \infty$ for any sequence $x_1, x_2, \cdots \in \mathrm{int}(\Omega) \subset \mathbb{R}^d$ converging to the boundary of Ω;*
- *for all $w \in \mathrm{int}(\Omega)$ and $h \in \mathbb{R}^d$, the following inequality always holds:*

$$\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} \frac{\partial^3 \psi(w)}{\partial w_i \partial w_j \partial w_k} h_i h_j h_k \le 2\|h\|^3_{\nabla^2 \psi(w)}.$$

We further call ψ is a ν-self-concordant barrier if it satisfies the conditions above and also

$$\langle \nabla \psi(w), h \rangle \le \sqrt{\nu}\|h\|_{\nabla^2 \psi(w)}$$

for all $w \in \mathrm{int}(\Omega)$ and $h \in \mathbb{R}^d$.

To deal with the infinite range of the barrier ψ on Ω, we define a shrunk version of the decision domain Ω using Minkowsky functions.

Definition 21 *Define the Minkowsky function $\pi_w : \Omega \mapsto \mathbb{R}$ associated with a point $w \in \mathrm{int}(\Omega)$ and Ω as:*

$$\pi_w(u) = \inf\left\{ s > 0 \;\middle|\; w + \frac{u - w}{s} \in \Omega \right\}.$$

Lemma 22 (Proposition 2.3.2 in (Nesterov, 1994)) *Let ψ be a ν-self-concordant barrier on $\Omega \in \mathbb{R}^d$. Then, for any $u, w \in \mathrm{int}(\Omega)$, we have*

$$\psi(u) - \psi(w) \le \nu \log\left(\frac{1}{1 - \pi_w(u)}\right).$$

This means that for all $u \in \Omega' \triangleq \{(1-\varepsilon)x + \varepsilon w : x \in \Omega\}$ where $\varepsilon > 0$, we have

$$\psi(u) - \psi(w) \le \nu \log\left(1/\varepsilon\right),$$

since $\pi_w(u) \le 1 - \varepsilon$.

Appendix C. Omitted Details in Section 4

In this section, we provide omitted details in Section 4. We first recall the standard conjugate-duality identity for Bregman divergences. This identity allows us to move between the primal and dual spaces when analyzing the FTRL iterates.

Lemma 23 (Conjugate Duality) *Suppose that $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ is differentiable and strictly convex. Let $\psi^*(w) = \sup_{x\in\mathbb{R}^d} (\langle x, w\rangle - \psi(x))$. Then, for any $x_1, x_2 \in \mathbb{R}^d$, we have $D_\psi(x_1, x_2) = D_{\psi^*}(\nabla\psi(x_2), \nabla\psi(x_1))$.*

Proof Direct calculation shows that

$$\begin{aligned} D_\psi(x_1, x_2) &= \psi(x_1) - \psi(x_2) - \langle \nabla\psi(x_2), x_1 - x_2\rangle \\ &= \psi(x_1) - \psi(x_2) - \langle \nabla\psi(x_2), x_1 - x_2\rangle + \langle x_1, \nabla\psi(x_1)\rangle - \langle x_1, \nabla\psi(x_1)\rangle \\ &= (\langle \nabla\psi(x_2), x_2\rangle - \psi(x_2)) - (\langle \nabla\psi(x_1), x_1\rangle - \psi(x_1)) - \langle x_1, \nabla\psi(x_2) - \nabla\psi(x_1)\rangle \\ &= D_{\psi^*}(\nabla\psi(x_2), \nabla\psi(x_1)), \end{aligned}$$

where the last equality is because $\nabla\psi^*(\nabla\psi(x)) = x$ and $\psi(x) + \psi^*(\nabla\psi(x)) = \langle x, \nabla\psi(x)\rangle$. ■

We next control the movement of consecutive FTRL iterates. The following stability bound relates this movement directly to the gradient magnitude of the newly observed loss.

Lemma 24 (Stability of FTRL) *Algorithm 3 ensures $\|x_t - x_{t+1}\|_2 \le \frac{\eta}{\sigma}\|\nabla f_t(x_t)\|_2$, for all t.*

Proof Let $F_{t+1}(x) \triangleq \sum_{\tau=1}^t f_\tau(x) + \frac{1}{\eta}\psi(x)$. Since x_{t+1} is the minimizer of F_{t+1} in $\mathcal{X}$, using first-order optimality, we have $F_{t+1}(x_{t+1}) \le F_{t+1}(x) - D_{F_{t+1}}(x, x_{t+1})$ for all $x \in \mathcal{X}$. Picking $x = x_t$, we get

$$F_{t+1}(x_{t+1}) - F_t(x_t) \le f_t(x_t) - D_{F_{t+1}}(x_t, x_{t+1}). \tag{20}$$

Similarly, for F_t, we have

$$F_t(x_t) - F_{t+1}(x_{t+1}) \le -f_t(x_{t+1}) - D_{F_t}(x_{t+1}, x_t). \tag{21}$$

Summing up Eq. (20) and Eq. (21), we can obtain that

$$f_t(x_t) - f_t(x_{t+1}) \ge D_{F_{t+1}}(x_t, x_{t+1}) + D_{F_t}(x_{t+1}, x_t) \ge \frac{\sigma}{\eta}\|x_t - x_{t+1}\|_2^2,$$

where the last inequality is because ψ is σ-strongly convex with respect to $\|\cdot\|_2$ within domain $\mathcal{X}$, implying that both F_t and F_{t+1} are $\frac{\sigma}{\eta}$-strongly convex. Using the convexity of f_t, we further have $f_t(x_t) - f_t(x_{t+1}) \le \langle \nabla f_t(x_t), x_t - x_{t+1}\rangle \le \|\nabla f_t(x_t)\|_2 \cdot \|x_t - x_{t+1}\|_2$. Combining these two inequalities, we get $\|x_t - x_{t+1}\|_2 \le \frac{\eta}{\sigma}\|\nabla f_t(x_t)\|_2$. ■

The next lemma is the main smoothness ingredient in the analysis. It shows that self-concordance of the losses, together with third-order smoothness of the conjugate regularizer, yields a third-order smoothness bound for the conjugate of the cumulative FTRL objective.

Lemma 25 *Let $F_t(x) \triangleq \sum_{\tau=1}^t f_\tau(x) + \frac{1}{\eta}\psi(x)$. If ψ is Legendre, σ-strongly convex and its convex conjugate ψ^* is M-smooth of order 3 with respect to $\|\cdot\|_2$, and f_τ is C-self-concordant for all $\tau \in [t]$, then F_t^* is $\left(\frac{2C\eta^{3/2}}{\sigma^{3/2}} + M\eta^2\right)$-smooth of order 3 with respect to $\|\cdot\|_2$.*

Proof By definition of convex conjugate, we know that

$$\nabla F_t^*(y) = (\nabla F_t)^{-1}(y),$$
$$\nabla^2 F_t^*(y) = \left(\nabla^2 F_t(x)\right)^{-1},$$

where $x = (\nabla F_t)^{-1}(y)$, the dual variable of y. To obtain the third derivative, we differentiate both sides of the equation above and obtain that:

$$\nabla^3 F_t^*(y)[h,h,h] = -\nabla^3 F_t(x)\left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right], \forall\ h \in \mathbb{R}^d.$$

We want to bound $|\nabla^3 F_t^*(y)[h,h,h]|$ for all h. Taking $A = \frac{1}{\eta}\nabla^2\psi(x)$ and $B = \nabla^2\left(\sum_{\tau=1}^t f_\tau(x)\right)$, we have

$$\begin{aligned}
&- \nabla^3 \frac{\psi(x)}{\eta}\left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right] \\
&= -\nabla^3 \frac{\psi(x)}{\eta}\left[A^{-1}A(A+B)^{-1}h, A^{-1}A(A+B)^{-1}h, A^{-1}A(A+B)^{-1}h\right] \\
&= \nabla^3\left(\frac{\psi}{\eta}\right)^*(y)[A(A+B)^{-1}h, A(A+B)^{-1}h, A(A+B)^{-1}h] \\
&\le \eta^2 M\|A(A+B)^{-1}h\|_2^3 && \text{(by Lemma 26)} \\
&\le \eta^2 M\|h\|_2^3. && \text{(by Lemma 27)}
\end{aligned}$$

Since for all $\tau \in [t]$, f_τ is C-self-concordant, we have

$$\begin{aligned}
&- \sum_{\tau=1}^t \nabla^3 f_\tau(x)\left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right] \\
&\le \sum_{\tau=1}^t 2C\left(\nabla^2 f_\tau(x)\left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right]\right)^{3/2} && \text{(by Definition 9)} \\
&\le 2C\left(\sum_{\tau=1}^t \nabla^2 f_\tau(x)\left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right]\right)^{3/2} && \text{(since } \nabla^2 f_\tau(x) \text{ is PSD)} \\
&= 2C\left(B\left[(A+B)^{-1}h, (A+B)^{-1}h\right]\right)^{3/2} \\
&\le 2C\left((A+B)\left[(A+B)^{-1}h, (A+B)^{-1}h\right]\right)^{3/2} && \text{(since } A \text{ is PSD)} \\
&\le 2C\left(\frac{\eta}{\sigma}\|h\|_2^2\right)^{3/2} && \text{(since } A \succeq \tfrac{\sigma}{\eta} I) \\
&\le \frac{2C\eta^{3/2}}{\sigma^{3/2}}\|h\|_2^3.
\end{aligned}$$

Now we are ready to estimate the third-order smoothness of F_t^*. Expanding the third derivative of F_t in terms of the derivatives of f_τ and $\frac{1}{\eta}\psi$, and bounding each of them, gives us the following:

$$\nabla^3 F_t^*(y)[h,h,h] = -\nabla^3 F_t(x)\left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right]$$

$$
\begin{aligned}
&= -\sum_{\tau=1}^{t} \nabla^3 f_\tau(x) \left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right] \\
&- \nabla^3 \frac{1}{\eta}\psi(x) \left[(\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h, (\nabla^2 F_t(x))^{-1}h\right] \\
&\leq \left(\frac{2C\eta^{3/2}}{\sigma^{3/2}} + M\eta^2\right)\|h\|_2^3.
\end{aligned}
$$

Therefore, F_t^* is $\left(\frac{2C\eta^{3/2}}{\sigma^{3/2}} + M\eta^2\right)$-smooth of order 3 with respect to $\|\cdot\|_2$. ■

To complete the preceding argument, we record two elementary auxiliary facts. The first describes how third-order smoothness of a conjugate scales when the primal regularizer is rescaled by the learning rate.

Lemma 26 *If ψ^* is M-smooth of order 3 with respect to $\|\cdot\|_2$, then the convex conjugate of $\frac{1}{\eta}\psi$ is $\eta^2 M$-smooth of order 3.*

Proof By the definition of convex conjugate, we know that

$$\left(\frac{1}{\eta}\psi\right)^*(y) = \sup_x \left(\langle y, x\rangle - \frac{1}{\eta}\psi(x)\right).$$

Rearranging the terms, we can obtain that

$$\eta \cdot \left(\frac{1}{\eta}\psi\right)^*(y) = \sup_x \left(\langle \eta y, x\rangle - \psi(x)\right) = \psi^*(\eta y).$$

Taking the third derivative, we have

$$\nabla^3 \left(\frac{1}{\eta}\psi\right)^*(y) = \eta^2 \nabla^3 \psi(\eta y),$$

which concludes the result in the lemma. ■

The second auxiliary fact controls the linear transformation that appears when comparing the Hessian of the regularizer with the Hessian of the full objective.

Lemma 27 *Assume that A is a positive definite matrix, while B is a positive semi-definite matrix. Then, it holds that the maximum singular value of $A(A+B)^{-1}$ is bounded by 1.*

Proof Let $C = A + B$. Then, we know that C is a positive definite matrix and

$$A(A+B)^{-1} = (C-B)C^{-1} = I - BC^{-1}.$$

Since C^{-1} is symmetric, there exists orthogonal matrix M such that $M^{-1}C^{-1}M = D$, where D is a diagonal matrix with positive diagonal entries. The singular value of $A(A+B)^{-1}$ is same as

$$M^{-1}A(A+B)^{-1}M = I - B'D$$

where $B' = M^{-1}BM$ is also symmetric. Since similarity transformation does not change the singular value, we only need to calculate the singular value of the following matrix:

$$D^{\frac{1}{2}}(I - B'D)D^{-\frac{1}{2}} = I - D^{\frac{1}{2}}B'D^{\frac{1}{2}},$$

where $D^{\frac{1}{2}}B'D^{\frac{1}{2}}$ is symmetric. Therefore, the singular value of $I - D^{\frac{1}{2}}B'D^{\frac{1}{2}}$ is bounded by 1, which implies that the maximum singular value of $A(A+B)^{-1}$ is bounded by 1. ■

Finally, we verify that the logistic loss used in our application satisfies the required self-concordance condition on the unit Euclidean ball. This allows the preceding third-order smoothness result to be applied directly to logistic losses.

Lemma 28 *Suppose that $\|z_t\|_2 \le 1$ and $\|x\|_2 \le 1$. For the logistic loss $f_t(x) = \log(1 + \exp(-y_t\langle z_t, x\rangle))$, where $y_t \in \{-1, 1\}$, we have*

$$\left|D^3 f_t(x)[h,h,h]\right| \le 2\sinh\left(\frac{1}{2}\right)\left(D^2 f_t(x)[h,h]\right)^{3/2}$$

for every direction h.

Proof Direct differentiation gives

$$D^2 f_t(x)[h,h] = \frac{e^{\langle y_t z_t, x\rangle}}{\left(1 + e^{\langle y_t z_t, x\rangle}\right)^2}\langle y_t z_t, h\rangle^2$$

and

$$D^3 f_t(x)[h,h,h] = \frac{e^{\langle y_t z_t, x\rangle}\left(1 - e^{\langle y_t z_t, x\rangle}\right)}{\left(1 + e^{\langle y_t z_t, x\rangle}\right)^3}\langle y_t z_t, h\rangle^3.$$

If $\langle y_t z_t, h\rangle = 0$, the desired inequality is immediate. Otherwise,

$$\frac{\left|D^3 f_t(x)[h,h,h]\right|}{\left(D^2 f_t(x)[h,h]\right)^{3/2}} = e^{-\langle y_t z_t, x\rangle/2}\left|1 - e^{\langle y_t z_t, x\rangle}\right| = 2\sinh\left(\frac{|\langle y_t z_t, x\rangle|}{2}\right) \le 2\sinh\left(\frac{1}{2}\right),$$

where the last inequality follows from $|\langle y_t z_t, x\rangle| \le \|z_t\|_2\|x\|_2 \le 1$. This proves the claim. ■

Appendix D. Omitted Details in Section 5

In this section, we show omitted proofs in Section 5. We first prove Theorem 17, which shows that Hedge suffers $\Omega(T^{\frac{1}{3}})$ alternating regret in the expert problem.

Proof [Proof of Theorem 17] We prove the lower bound by constructing two environments and showing that if the Hedge algorithm achieves $\mathcal{O}(T^{1/3})$ alternating regret for one environment, then it will suffer $\Omega(T^{1/3})$ alternating regret for the other one.

Environment 1: We consider the time horizon to be $3T$ and 3 actions with the loss vector cycling between the three basis vectors in $\mathbb{R}^3 : (1,0,0), (0,1,0), (0,0,1)$ and we have $\min_{i\in[3]}\sum_{t=1}^{3T}\ell_{t,i} = T$. Direct calculation shows that Hedge with learning rate $\eta > 0$

predict the p_t sequence as follows: $p_{3t-2} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), p_{3t-1} = \left(\frac{e^{-\eta}}{2+e^{-\eta}}, \frac{1}{2+e^{-\eta}}, \frac{1}{2+e^{-\eta}}\right), p_{3t} = \left(\frac{e^{-\eta}}{1+2e^{-\eta}}, \frac{e^{-\eta}}{1+2e^{-\eta}}, \frac{1}{1+2e^{-\eta}}\right)$, $t \in [T]$. Thus, we can bound the alternating regret as follows:

$$\mathrm{Reg}_{\mathrm{Alt}} = T\left(\frac{2}{3} + \frac{1+e^{-\eta}}{2+e^{-\eta}} + \frac{1+e^{-\eta}}{1+2e^{-\eta}}\right) - 2T = \frac{T(1-e^{-\eta})^2}{3(2+e^{-\eta})(1+2e^{-\eta})} \geq \frac{T(1-e^{-\eta})^2}{27}.$$

To proceed, note that when $\eta \geq 1$, the above inequality already means that $\mathrm{Reg}_{\mathrm{Alt}} = \Omega(T)$. Therefore, we only consider the case when $\eta \leq 1$. Using the fact that $e^{-\eta} \leq 1 - \eta + \frac{\eta^2}{2}$ for $\eta \geq 0$, we can further lower bound $\frac{T(1-e^{-\eta})^2}{27}$ by $\mathrm{Reg}_{\mathrm{Alt}} \geq \frac{T(\eta - \frac{\eta^2}{2})^2}{27} \geq \frac{\eta^2 T}{108}$, where the last inequality is due to $\eta \leq 1$. Therefore, we know that $\mathrm{Reg}_{\mathrm{Alt}} = \Omega(\eta^2 T)$.

Environment 2: We consider the time horizon to be T and 3 actions with the loss vector ℓ_t being $(1, 0, 0)$ for all rounds. Here, the benchmark $\min_{i\in[3]} \sum_{t=1}^{T} \ell_{t,i}$ is 0, and $p_{t,1} = \frac{e^{-\eta T}}{2+e^{-\eta T}}$ for all $t \in [T]$. In this case, if $\eta \leq \frac{2}{T}$, we know that $p_{t,1} \geq \frac{e^{-2}}{2+e^{-2}}$ for all $t \in [T]$ and the algorithm will be suffering $\Omega(T)$ regret. When $1 \geq \eta \geq \frac{2}{T}$, we have:

$$\begin{aligned}
\mathrm{Reg}_{\mathrm{Alt}} &= \sum_{t=0}^{T} \frac{2e^{-\eta t}}{2+e^{-\eta t}} - \frac{1}{3} - \frac{e^{-\eta T}}{2+e^{-\eta T}} \\
&\geq \sum_{t=1}^{T-1} \frac{2}{1+2e^{\eta t}} \geq \int_{1}^{T} \frac{2}{3e^{\eta t}} dt \\
&= \frac{2}{3\eta}\left[-e^{-\eta t}\right]\Big|_{1}^{T} \geq \frac{e^{-1}}{3\eta},
\end{aligned}$$

where the second inequality uses $e^{\eta t} \geq 1$ and the last inequality uses $\eta \leq 1$. Therefore, we have $\mathrm{Reg}_{\mathrm{Alt}} = \Omega\left(\frac{1}{\eta}\right)$. Combining both environments, we know that Hedge with learning rate η suffers a $\Omega(\max\{\frac{1}{\eta}, \eta^2 T\})$, leadning to a $\Omega(T^{\frac{1}{3}})$ lower bound. ■

Next, we show that OOGD and PRM$^+$ suffer $\Omega(\sqrt{T})$ alternating regret in the adversarial environment.

Proof [Proof of Theorem 18] To prove the lower bound for both algorithms, we use the following loss sequence to show the lower bound: for $k \geq 0$, consider

$$\ell_{2k} = \begin{bmatrix} 4 \\ \\ 0 \end{bmatrix}, \quad \ell_{2k+1} = \begin{bmatrix} -2 \\ \\ 0 \end{bmatrix}.$$

We start with the proof for OOGD. The process of OOGD with learning rate $\eta > 0$ is as follows. Let $\widehat{p} = \frac{1}{2} \cdot \mathbf{1}$ and $m_1 = \mathbf{0}$ where $\mathbf{0}$ and $\mathbf{1}$ are vectors with all components equal to 0 and 1. At each round t, OOGD selects $p_t = \mathrm{argmin}_{p\in\Delta_2}\{\langle p, m_t\rangle + \frac{1}{2\eta}\|p - \widehat{p}_t\|_2^2\}$. Then, after receiving ℓ_t, OOGD updates $\widehat{p}_{t+1}$ to be $\mathrm{argmin}_{p\in\Delta_2}\{\langle p, \ell_t\rangle + \frac{1}{2\eta}\|p - \widehat{p}_t\|_2^2\}$ and sets $m_{t+1} = \ell_t$.

Without loss of generality, we assume that $\frac{1}{\eta}$ and T are even. To show OOGD suffers $\Omega(\sqrt{T})$ alternating regret, direct calculation shows that when $t \le \frac{1}{\eta} - 3$, we have

$$p_t = \begin{cases} [\frac{1}{2}, \frac{1}{2}] & \text{when } t = 1, \\ [\frac{1}{2} - k\eta + 3\eta, \frac{1}{2} + k\eta - 3\eta] & \text{when } t = 2k, \\ [\frac{1}{2} - k\eta - 2\eta, \frac{1}{2} + k\eta + 2\eta] & \text{when } t = 2k+1; \end{cases} \tag{22}$$

when $t \ge \frac{1}{\eta} + 6$, we have

$$p_t = \begin{cases} [2\eta, 1 - 2\eta] & \text{when } t = 2k, \\ [0, 1] & \text{when } t = 2k+1. \end{cases} \tag{23}$$

Based on the analytic form of p_t as shown in Eq. (22) and Eq. (23), when $t \le \frac{1}{\eta} - 3$, the alternating regret can be calculated as

$$\sum_{k=1}^{\frac{1}{2\eta}-2} \langle p_{2k} - e_2, \ell_{2k-1} + \ell_{2k} \rangle + \langle p_{2k+1} - e_2, \ell_{2k} + \ell_{2k+1} \rangle = \sum_{k=1}^{\frac{1}{2\eta}-2} (2 - 4k\eta + 2\eta) = \frac{1}{2\eta} - 8\eta.$$

When $t \ge \frac{1}{\eta} + 6$, the alternating regret is

$$\sum_{k=\frac{1}{2\eta}+3}^{T/2-1} \langle p_{2k} - e_2, \ell_{2k-1} + \ell_{2k} \rangle + \langle p_{2k+1} - e_2, \ell_{2k} + \ell_{2k+1} \rangle = 2\eta T - 2 - 12\eta.$$

Combining the above two cases, we know that

$$\mathrm{Reg}_{\mathrm{Alt}} = C + \frac{1}{2\eta} + 2\eta T - 2 - 20\eta,$$

where C is the regret for the remaining rounds (which is only a constant). This finishes our proof of the $\Omega(\sqrt{T})$ alternating regret lower bound for OOGD.

Next, we prove our results for PRM$^+$ under the same loss sequence. The procedure of PRM$^+$ is as follows: Let $\widehat{R}_1 = R_1 = r_0 = \mathbf{0}$, and for $t \ge 1$, PRM$^+$ selects p_t to be $\widehat{R}_t / \|\widehat{R}_t\|_1$ where $\widehat{R}_t = [R_t + r_{t-1}]^+$ and update R_{t+1} to be $[R_t + r_t]^+$, where $r_t = \langle p_t, \ell_t \rangle \mathbf{1}_d - \ell_t$.

To simplify notation, denote $4\alpha_k = R_{2k+1,2}$ for $k \in [\frac{T}{2} - 1]$. Lemma 29 shows that for $k \ge 5$, α_k follows the following recurrence relation: $\alpha_{k+1} = \alpha_k + \frac{1}{1+\alpha_k}$. We use this recurrence to compute the values of p_t for all rounds $t > 10$.

Thus, using Lemma 29, the alternating loss (standard loss + cheating loss) for the rounds $t = 2k+1$ and $t = 2k+2$ can be calculated as

$$\langle p_{2k+1}, \ell_{2k} + \ell_{2k+1} \rangle + \langle p_{2k+2}, \ell_{2k+1} + \ell_{2k+2} \rangle = \frac{2}{(1 + \alpha_k)}.$$

Since the action 2 always has a loss of 0, the benchmark here is 0. Therefore, the alternating regret is:

$$\mathrm{Reg}_{\mathrm{Alt}} = C' + \sum_{k=5}^{\frac{T}{2}} \frac{2}{1 + \alpha_k}, \tag{24}$$

where C' is a constant bounding the regret for the first 10 rounds. To estimate the quantity above, we prove $\alpha_k \leq 2\sqrt{k} - 1$ by induction. The base case $k = 5$ can be verified by direct calculation. Now, let us assume that the claim holds for k. Then, for $k + 1$,

$$\alpha_{k+1} = \alpha_k + \frac{1}{1+\alpha_k} \leq 2\sqrt{k} - 1 + \frac{1}{2\sqrt{k}} \leq \frac{4k+1}{2\sqrt{k}} - 1 \leq 2\sqrt{k+1} - 1.$$

The first inequality comes from the fact that the function $f(x) = x + \frac{1}{1+x}$ is monotonically increasing for $x \geq 0$. Substituting it in Eq. (24), we get

$$\mathrm{Reg}_{\mathrm{Alt}} \geq C' + \sum_{k=5}^{\frac{T}{2}} \frac{1}{\sqrt{k}} = \Theta(\sqrt{T}).$$

Therefore, $\mathrm{Reg}_{\mathrm{Alt}} = \Omega(\sqrt{T})$ for the loss sequence proposed. ■

Lemma 29 *Suppose that the loss vector sequence satisfies that* $\ell_{2k} = \left[\begin{smallmatrix} 4 \\ 0 \end{smallmatrix}\right], \ell_{2k+1} = \left[\begin{smallmatrix} -2 \\ 0 \end{smallmatrix}\right]$ *for* $k \geq 0$. *Then, PRM*$^+$ *guarantees that for* $k \geq 5$

$$\begin{aligned}
p_{2k+1} &= \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], p_{2k+2} = \begin{bmatrix} \frac{1}{1+\alpha_k} \\ \frac{\alpha_k}{1+\alpha_k} \end{bmatrix}, \\
R_{2k+2} &= \left[\begin{smallmatrix} 2 \\ 4\alpha_k \end{smallmatrix}\right], R_{2k+3} = \begin{bmatrix} 0 \\ 4\alpha_k + \frac{4}{1+\alpha_k} \end{bmatrix}, \\
\widehat{R}_{2k+2} &= \left[\begin{smallmatrix} 4 \\ 4\alpha_k \end{smallmatrix}\right], \widehat{R}_{2k+3} = \begin{bmatrix} 0 \\ 4\alpha_k + \frac{8}{1+\alpha_k} \end{bmatrix},
\end{aligned}$$

where $\alpha_5 > 2$ *is certain constant and* $\alpha_{k+1} = \alpha_k + \frac{1}{1+\alpha_k}$ *for* $k \geq 5$.

Proof It can be verified that Lemma 29 holds true when $k = 5$. For $k > 5$, we prove by induction. Suppose Lemma 29 holds for k. Then, for $k + 1$, we have,

$$\begin{aligned}
p_{2k+3} &= \frac{\widehat{R}_{2k+3}}{\|\widehat{R}_{2k+3}\|_1} = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \\
R_{2k+4} &= [R_{2k+3} + r_{2k+3}]^+ = \begin{bmatrix} 2 \\ 4\alpha_k + \frac{4}{1+\alpha_k} \end{bmatrix} = \left[\begin{smallmatrix} 2 \\ 4\alpha_{k+1} \end{smallmatrix}\right], \\
r_{2k+3} &= \langle p_{2k+3}, \ell_{2k+3} \rangle \mathbf{1}_d - \ell_{2k+3} = \left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right], \\
\widehat{R}_{2k+4} &= [R_{2k+4} + r_{2k+3}]^+ = \left[\begin{smallmatrix} 4 \\ 4\alpha_{k+1} \end{smallmatrix}\right].
\end{aligned}$$

Since $\alpha_k \geq 2$, we have $\alpha_{k+1} \geq 2$ as well. Using this, we can see that

$$\begin{aligned}
p_{2k+4} &= \frac{\widehat{R}_{2k+4}}{\|\widehat{R}_{2k+4}\|_1} = \begin{bmatrix} \frac{1}{1+\alpha_{k+1}} \\ \frac{\alpha_{k+1}}{1+\alpha_{k+1}} \end{bmatrix}, \\
r_{2k+4} &= \langle p_{2k+4}, \ell_{2k+4} \rangle \mathbf{1}_d - \ell_{2k+4} = \begin{bmatrix} -\frac{4\alpha_{k+1}}{1+\alpha_{k+1}} \\ \frac{4}{1+\alpha_{k+1}} \end{bmatrix},
\end{aligned}$$

$$R_{2k+5} = [R_{2k+4} + r_{2k+4}]^+ = \begin{bmatrix} 0 \\ 4\alpha_{k+1} + \frac{4}{1+\alpha_{k+1}} \end{bmatrix},$$
$$\widehat{R}_{2k+5} = [R_{2k+5} + r_{2k+4}]^+ = \begin{bmatrix} 0 \\ 4\alpha_{k+1} + \frac{8}{1+\alpha_{k+1}} \end{bmatrix},$$

where the third equality uses the fact that $\frac{\alpha_{k+1}}{1+\alpha_{k+1}} \geq \frac{2}{3} > \frac{1}{2}$. Thus, the claim is true for $k+1$, and hence it holds for all $k \geq 5$. ■

References

Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2022a.

Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $o(\log t)$ swap regret in multiplayer games. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022b.

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

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

Noam Brown and Tuomas Sandholm. Superhuman ai for multiplayer poker. *Science*, 365 (6456):885–890, 2019.

Sébastien Bubeck. Introduction to online optimization. *Lecture notes*, 2:1–86, 2011.

Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Proceedings of the 28th Conference on Learning Theory (COLT)*, 2015.

Neil Burch, Matej Moravcik, and Martin Schmid. Revisiting cfr+ and alternating updates. *Journal of Artificial Intelligence Research*, 64:429–443, 2019.

Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Weiqiang Zheng. Last-iterate convergence properties of regret-matching algorithms in games. In *Proceedings of the 13th International Conference on Learning Representations (ICLR)*, 2025.

Volkan Cevher, Ashok Cutkosky, Ali Kavis, Georgios Piliouras, Stratis Skoulakis, and Luca Viano. Alternation makes the adversary weaker in two-player games. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:18263–18290, 2023.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:18990–18999, 2020.

Sinho Chewi. The entropic barrier is n-self-concordant. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2020–2022*, pages 209–222. Springer, 2023.

Thomas M Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 235–254. SIAM, 2011.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pages 5363–5371, 2021.

Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning dynamics for general convex games. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:39076–39089, 2022a.

Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 6337–6357. PMLR, 2022b.

Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40–55, 1997.

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

Julien Grand-Clément and Christian Kroer. Solving optimization problems with blackwell approachability. *Mathematics of Operations Research*, 49(2):697–728, 2024.

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

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

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

Jonas Katona, Xiuyuan Wang, and Andre Wibisono. A symplectic analysis of alternating mirror descent. *arXiv preprint arXiv:2405.03472*, 2024.

László Lovász and Santosh Vempala. The geometry of logconcave functions and an o*(n3) sampling algorithm (technical report msr-tr-2003-04). *Microsoft Research*, 2003.

László Lovász and Santosh Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 57–68. IEEE, 2006a.

László Lovász and Santosh Vempala. Simulated annealing in convex bodies and an o*(n4) volume algorithm. *Journal of Computer and System Sciences*, 72(2):392–417, 2006b.

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

Hariharan Narayanan and Alexander Rakhlin. Random walk approach to regret minimization. *Advances in Neural Information Processing Systems (NeurIPS)*, 23, 2010.

Y Nesterov. Interior point polynomial algorithms in convex programming. *SIAM Studies in Applied Mathematics*, 1994.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 2004.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

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

Andre Wibisono, Molei Tao, and Georgios Piliouras. Alternating mirror descent for constrained min-max games. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:35201–35212, 2022.

Lijun Zhang, Tianbao Yang, Jinfeng Yi, Rong Jin, and Zhi-Hua Zhou. Improved dynamic regret for non-degenerate functions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pages 928–936, 2003.