Near-Optimal No-Regret Learning Dynamics for General Convex Games

Gabriele Farina*
Carnegie Mellon University
Pittsburgh, PA 15213
gfarina@cs.cmu.edu

Ioannis Anagnostides*
Carnegie Mellon University
Pittsburgh, PA 15213
ianagnos@cs.cmu.edu

Haipeng Luo
University of Southern California
Los Angeles, CA 90007
haipengl@usc.edu

Chung-Wei Lee
University of Southern California
Los Angeles, CA 90007
leechung@usc.edu

Christian Kroer
Columbia University
New York, NY 10027
christian.kroer@columbia.edu

Tuomas Sandholm
Carnegie Mellon University
Strategy Robot, Inc.
Optimized Markets, Inc.
Strategic Machine, Inc.
Pittsburgh, PA 15213
sandholm@cs.cmu.edu

Abstract

A recent line of work has established uncoupled learning dynamics such that, when employed by all players in a game, each player's *regret* after T repetitions grows polylogarithmically in T, an exponential improvement over the traditional guarantees within the no-regret framework. However, so far these results have only been limited to certain classes of games with structured strategy spaces—such as normal-form and extensive-form games. The question as to whether $O(\text{polylog}\, T)$ regret bounds can be obtained for general convex and compact strategy sets—which occur in many fundamental models in economics and multiagent systems—while retaining efficient strategy updates is an important question. In this paper, we answer this in the positive by establishing the first uncoupled learning algorithm with $O(\log T)$ per-player regret in general *convex games*, that is, games with concave utility functions supported on arbitrary convex and compact strategy sets. Our learning dynamics are based on an instantiation of optimistic follow-the-regularized-leader over an appropriately *lifted* space using a *self-concordant regularizer* that is peculiarly not a barrier for the feasible region. Our learning dynamics are efficiently implementable given access to a proximal oracle for the convex strategy set, leading to $O(\log \log T)$ per-iteration complexity; we also give extensions when access to only a *linear* optimization oracle is assumed. Finally, we adapt our dynamics to guarantee $O(\sqrt{T})$ regret in the adversarial regime. Even in those special cases where prior results apply, our algorithm improves over the state-of-the-art regret bounds either in terms of the dependence on the number of iterations or on the dimension of the strategy sets.

*Equal contribution.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

1 Introduction

Regret minimization is a celebrated framework that has been central in the development of online learning and the theory of multiagent systems. Indeed, fundamental connections have been forged between no-regret learning and game-theoretic solution concepts [Freund and Schapire, 1999, Hart and Mas-Colell, 2000, Foster and Vohra, 1997, Roughgarden, 2015]. More broadly, regret is an intrinsic measure of performance in online learning and games. Furthermore, regret minimization algorithms have enjoyed a remarkable practical success, being a primary component in recent landmark results in AI [Bowling et al., 2015, Moravčík et al., 2017, Brown and Sandholm, 2017, 2019]. These advances were guided by game-theoretic principles, made possible by training the AI agents using *self-play* under regret-minimizing algorithms, an approach that has proven to be more scalable compared to linear programming techniques. Nevertheless, the traditional no-regret framework is overly pessimistic, insisting on modeling the environment in a fully adversarial way. While this well-understood worst-case view might be justifiable for applications such a security games, it could be far from optimal in more benign and *predictable* environments, including the setting of training agents using self-play. This begs the question: *What are the optimal performance guarantees we can obtain when learning agents are competing against each other in general games?*

This fundamental question was first formulated and addressed by Daskalakis et al. [2011] within the context of *zero-sum games*. Since then, there has been a considerable interest in extending their guarantee to more general settings [Rakhlin and Sridharan, 2013, Syrgkanis et al., 2015, Foster et al., 2016, Chen and Peng, 2020, Daskalakis and Golowich, 2022, Piliouras et al., 2022]. In particular, Daskalakis et al. [2021] recently established that when all players in a general *normal-form game* employ an *optimistic* variant of *multiplicative weights update (MWU)*, the regret of each player grows *nearly-optimally* as $O(\log^4 T)$ after T repetitions of the game, leading to an *exponential improvement* over the guarantees obtained using traditional techniques within the no-regret framework. However, while normal-form games are a common way to represent strategic interactions in theory, most settings of practical significance inevitably involve more complex strategy spaces. For those settings, any faithful approximation of the game using the normal form is typically inefficient, requiring an action space that is exponential in the natural parameters of the problem, thereby limiting the practical implications of those prior results. This motivates our central question:

Can we establish near-optimal, efficiently implementable, and strongly uncoupled no-regret learning dynamics in general convex games? (♣)

Convex games are a rich class of games wherein the strategy space of each player is an arbitrary convex and compact set, while the utility of each player is an arbitrary concave function (see Section 2 for a formal description). As such, convex games encompass normal-form and *extensive-form games*, but go well-beyond to many other fundamental settings in economic theory including routing games, resource allocation problems, and competition between firms. Our primary contribution in this paper is to substantially extend prior results to all such games, addressing Question (♣).

1.1 Our Contributions

In this paper we introduce a novel no-regret learning algorithm, which we coin *lifted log-regularized optimistic follow the regularized leader (`LRL-OFTRL`)*. `LRL-OFTRL` settles Question (♣) in the positive, as summarized in the following theorem.[2]

Theorem 1 (Detailed version in Theorem 4)**.** *Consider any general convex game. When all players employ our strongly uncoupled learning dynamics (`LRL-OFTRL`), the regret of each player grows as $O(\log T)$. At the same time, if the player is facing adversarial utilities we guarantee $O(\sqrt{T})$ regret.*

Importantly, our learning dynamics are efficiently implementable given access to a *proximal oracle* for the set (Equation (7)), requiring only $O(\log \log T)$ operations per-iteration (Theorem 5); such an oracle is weaker than the—relatively standard in convex optimization—*quadratic optimization oracle*. We also point out extensions under a weaker *linear optimization oracle*, albeit with a worse per-iteration complexity (Theorem 6). Our no-regret learning dynamics imply the first efficiently implementable and near-optimal regret guarantees in general convex games, significantly extending

[2] For simplicity in the exposition we use the $O(\cdot)$ notation in our introduction to suppress time-independent parameters that depend (polynomially) on the game; precise statements are deferred to Section 3.

Method	Applies to	Regret bound	Cost per iteration
OFTRL / OMD [Syrgkanis et al., 2015]	General convex set	$O(\sqrt{n}\mathfrak{R}T^{1/4})$	Regularizer- & oracle- dependent
OMWU [Daskalakis et al., 2021]	Simplex Δ^d	$O(n \log d \log^4 T)$	$O(d)$
Clairvoyant MWU [Piliouras et al., 2022]	Simplex Δ^d	$O(n \log d)$ Subsequence only ‡	$O(d)$
Kernelized OMWU [Farina et al., 2022]	Polytope $\Omega = \text{co}\mathcal{V}$ with $\mathcal{V} \subseteq \{0,1\}^d$	$O(n \log \lvert\mathcal{V}\rvert \log^4 T)$	$d \times$ cost of kernel
`LRL-OFTRL` **[This paper]**	General convex set $\mathcal{X} \subseteq \mathbb{R}^d$	$O(nd\lVert\mathcal{X}\rVert_1^2 \log T)$	Oracle-dependent: • $O(\log\log T)$ proximal oracle calls • $O(\text{poly}\, T)$ linear opt. oracle calls

Table 1: Comparison of prior results on minimizing external regret in games. For simplicity, we have suppressed dependencies on the smoothness and the range of the utilities. We use n to denote the number of players; T to denote the number of repetitions; $\mathfrak{R}$ to indicate a parameter that depends on the regularizer; co$\mathcal{V}$ to denote the convex hull of $\mathcal{V}$; and $\|\mathcal{X}\|_1$ to denote a bound on the maximum ℓ_1 norm of any strategy. ‡ Unlike all other algorithms, the full sequence of iterates produced by Clairvoyant MWU (CMWU) is not known to achieve sublinear regret. Rather, after running CMWU for T iterations, only a smaller subsequence of length $\Theta(T/\log T)$ iterates is known to attain the regret stated in the table. So, we remark that in order to achieve a comparable approximation of a coarse correlated equilibrium, CMWU needs to be run for $\Theta(T \log T)$ iterations.

the scope of prior $O(\text{polylog}\, T)$-regret guarantees [Daskalakis et al., 2021, Farina et al., 2022]; a comparison with prior approaches is included in Table 1. We remark that Theorem 1 establishes near-optimal regret both under *self-play*, and in the adversarial regime—meaning that the other players act so as to minimize the player's utility; the latter feature of adversarial robustness has been a central desideratum in this line of work (*e.g.*, see the discussion in [Kangarshahi et al., 2018, Daskalakis et al., 2011]).

Our proposed learning dynamics lie within the general framework of *optimistic* no-regret learning, pioneered by Chiang et al. [2012] and Rakhlin and Sridharan [2013]. We leverage the `OFTRL` algorithm of Syrgkanis et al. [2015], but with some important twists. First, as detailed in Algorithm 1, the `OFTRL` optimization step is performed over a "lifted" space. While prior work in online learning has employed similar in spirit approaches [Lee et al., 2020, Luo et al., 2022], our lifting is quite different, ensuring that the regret incurred by `OFTRL` is *nonnegative* (Theorem 2). Further, we employ a *logarithmic self-concordant regularizer*; interestingly, and perhaps surprisingly, this is not a *barrier* for the underlying feasible set. This deviates substantially from the typical use of self-concordant regularization (especially within the bandit setting [Abernethy et al., 2008, Wei and Luo, 2018, Bubeck et al., 2019]). A pictorial overview of our construction is given in the caption of Algorithm 1.

The use of the logarithmic regularizer serves two main purposes. First, we show that it guarantees *multiplicative stability* of the strategies, a refined notion of stability that is also leveraged in the work of Daskalakis et al. [2021]. Nonetheless, we are the first to leverage such properties in general domains, going well beyond the guarantees of (Optimistic) MWU on the simplex [Daskalakis et al., 2021]. Further, the *local norm* induced by the logarithmic regularizer enables us to cast regret bounds from the lifted space to the original space, while preserving the *RVU property* [Syrgkanis et al., 2015, Definition 3]. In turn, this implies near-optimal regret by establishing that the *second-order path lengths* up to time T are bounded by $O(\log T)$ (Theorem 3), building on a recent technique of Anagnostides et al. [2022a] which crucially leverages the nonnegativity of swap regret.[3]

[3]To see why nonnegativity is crucial, note that the RVU bound implies optimal *sum* of players' regrets [Syrgkanis et al., 2015]. Thus, nonnegativity would imply the same bound for each player's regret.

1.2 Further Related Work

The rich line of work pursuing improved regret guarantees in games was pioneered by Daskalakis et al. [2011]. Specifically, they developed *strongly uncoupled* learning dynamics so that the players' regrets grow as $O(\log T)$, an exponential improvement over the guarantee one could hope for in adversarial environments [Shalev-Shwartz, 2012, Cesa-Bianchi and Lugosi, 2006]. Their result was significantly simplified by Rakhlin and Sridharan [2013]—again in zero-sum games—who introduced a simple variant of *mirror descent* with a *recency bias*—a.k.a. *optimistic* mirror descent (OMD). It is worth noting that, beyond the benefits of optimism from an optimization standpoint [Polyak, 1987], recency bias has been experimentally documented in natural learning environments in economics [Fudenberg and Peysakhovich, 2014].

Subsequently, Syrgkanis et al. [2015] crystallized the *RVU property*, an adversarial regret bound applicable for a broad class of optimistic no-regret learning algorithms. Using that property, they showed that the individual regret of each player grows as $O(T^{1/4})$ in general games, thereby converging to the set of *coarse correlated equilibria* with a rate of $O(T^{-3/4})$. A near-optimal bound of $O(\text{polylog}(T))$ in normal-form games was finally established by Daskalakis et al. [2021], while Farina et al. [2022] generalized that result in a class of polyhedral games that includes extensive-form games. Some extensions of the previous results have also been established for the stronger notion of *no-swap-regret* learning dynamics in normal-form games [Chen and Peng, 2020, Anagnostides et al., 2022b,a]. In particular, our work builds on a very recent technique of Anagnostides et al. [2022a], which established $O(\log T)$ swap regret in normal-form games using as a regularizer a self-concordant *barrier* function. On the other hand, establishing even sublinear $o(T)$ swap regret in extensive-form games is a notorious open question. Finally, an interesting new approach for obtaining near-optimal external regret in normal-form games was recently proposed in concurrent work by Piliouras et al. [2022].[4]

Games with continuous strategy spaces have received a lot of attention in the literature; *e.g.*, see [Roughgarden and Schoppmann, 2015, Even-Dar et al., 2009, Harks and Klimm, 2011, Hsieh et al., 2021, Mertikopoulos and Zhou, 2019, Stein et al., 2011, Stoltz and Lugosi, 2007], and references therein. Such games encompass a wide variety of applications in economics and multiagent systems; we give several examples in Section 2. Indeed, in many applications of interest a faithful approximation of the game requires an extremely large or even infinite action space; such settings could be abstracted as *Littlestone games* in the sense of the recent work of Daskalakis and Golowich [2022].

2 No-Regret Learning and Convex Games

In this section we review the general setting of *convex games*[5] which encompasses a number of important applications, as explained in Section 2.2. We then formally define the framework of uncoupled and online no-regret learning in games in Section 2.3.

Notation We let $\mathbb{N} = \{1, 2, \dots, \}$ be the set of natural numbers. For a vector $\boldsymbol{x} \in \mathbb{R}^d$ we denote by $\boldsymbol{x}[r]$ its r-th coordinate, for some index $r \in [\![d]\!] := \{1, 2, \dots, d\}$. We will typically represent the players using subscripts; superscripts are reserved for the time index, denoted by the variable t.

2.1 Convex Games

Let $[\![n]\!] := \{1, 2, \dots, n\}$ be a set of players, with $n \in \mathbb{N}$. In a *convex game*, every player $i \in [\![n]\!]$ has a nonempty convex and compact set of strategies $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$. For a *joint strategy profile* $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \in \bigtimes_{j=1}^n \mathcal{X}_j$, the reward of player i is given by a continuously differentiable utility function $u_i : \bigtimes_{j=1}^n \mathcal{X}_j \to \mathbb{R}$ subject to the following standard assumption.

[4] An earlier version of the paper [Piliouras et al., 2021] proposed a preliminary and *not uncoupled* version of the Clairvoyant MWU algorithm whose iterates were guaranteed to be no-regret and require $O(d \log T)$ per-iteration complexity. The 2022 revision of that paper provides an *uncoupled* version with time-independent $O(d)$ per-iteration complexity, albeit at the cost of losing the no-regret guarantee on the entire sequence of iterates. See also footnote ‡ in Table 1.

[5] Sometimes these are referred to as *concave games* [Rosen, 1965] or *continuous games* [Hsieh et al., 2021].

Assumption 1 (Convex games)**.** *The utility function $u_i(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ of any player $i \in [\![n]\!]$ satisfies the following properties:*

1. *(Concavity) $u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$ is* concave *in $\boldsymbol{x}_i$ for $\boldsymbol{x}_{-i} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \dots, \boldsymbol{x}_n) \in \bigtimes_{j \neq i} \mathcal{X}_j$;*
2. *(Bounded gradients) for any $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \in \bigtimes_{j=1}^n \mathcal{X}_j$, $\|\nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)\|_\infty \le B$, for some parameter $B > 0$; and*
3. *(L-smoothness) there exists $L > 0$ so that for any two joint strategy profiles $\boldsymbol{x}, \boldsymbol{x}' \in \bigtimes_{j=1}^n \mathcal{X}_j$,*

$$\|\nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}) - \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}')\|_\infty \le L \sum_{j \in [\![n]\!]} \|\boldsymbol{x}_j - \boldsymbol{x}'_j\|_1.$$

2.2 Applications and Examples of Convex Games

Here we discuss several different classes of games which can all be analyzed under the common framework of convex games. For simplicity, we describe Cournot competion in the one-dimensional setting, but it can be readily generalized in more general domains. For more examples, we refer to [Even-Dar et al., 2009, Hsieh et al., 2021], and references therein.

Normal-Form Games In *normal-form games (NFGs)* every player $i \in [\![n]\!]$ has a finite and nonempty set of strategies $\mathcal{A}_i$. Player i's strategy set contains all probability distributions supported on $\mathcal{A}_i$; that is, $\mathcal{X}_i = \Delta(\mathcal{A}_i)$. The utility of player i can be expressed as the *multilinear* function $u_i(\boldsymbol{x}) \coloneqq \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{x}}[\mathcal{U}_i(\boldsymbol{a})]$, for some arbitrary function $\mathcal{U}_i : \bigtimes_{j=1}^n \mathcal{A}_j \to \mathbb{R}$.

Extensive-Form Games *Extensive-form games (EFGs)* generalize NFGs by capturing both sequential and simultaneous moves, stochasticity from the environment, as well as *imperfect information*. EFGs are abstracted on a directed tree. Once the game reaches a terminal (or leaf) node $z \in \mathcal{Z}$, each player $i \in [\![n]\!]$ receives a utility $\mathcal{U}_i(z)$, for some $\mathcal{U}_i : \mathcal{Z} \to \mathbb{R}$. The strategy space of each player $i \in [\![n]\!]$ can be compactly represented using the *sequence-form polytope* $\mathcal{Q}_i$ [Romanovskii, 1962, Koller et al., 1996]. If $p_c(z)$ is the probability of reaching terminal node $z \in \mathcal{Z}$ over "chance moves", the utility of player i can be expressed as $u_i(\boldsymbol{q}) \coloneqq \sum_{z \in \mathcal{Z}} p_c(z) \mathcal{U}_i(z) \prod_{j \in [\![n]\!]} \boldsymbol{q}_j[\sigma_{j,z}]$, where $\boldsymbol{q} = (\boldsymbol{q}_1, \dots, \boldsymbol{q}_n) \in \bigtimes_{j=1}^n \mathcal{Q}_j$ is the joint strategy profile, and $\boldsymbol{q}_j[\sigma_{j,z}]$ is the probability mass assigned to the last *sequence* $\sigma_{j,z}$ encountered by player j before reaching z. The smoothness and the concavity of the utilities follow directly from multilinearity; for a more detailed account on EFGs we refer the interested reader to the excellent book of Shoham and Leyton-Brown [2008].

Splittable Routing Games In these games [Roughgarden and Schoppmann, 2015] every player has to route a flow f_i from a source to a destination in an undirected graph $G = (V, E)$. Every edge $e \in E$ is associated with a latency function $\ell_e(f_e)$ mapping the amount of flow passing through the edge to some latency. The set of strategies of player i corresponds to the possible ways of "splitting" the flow f_i into paths from the source to the destination. Under suitable restrictions on the latency functions, those games satisfy Assumption 1 (see [Syrgkanis et al., 2015]).

Cournot Competition This game is played among n firms (players). Every firm i decides the *quantity* $s_i \in \mathcal{S}_i \subseteq \mathbb{R}_{\ge 0}$ of a common good to produce, where $\mathcal{S}_i$ is an interval. Further, a cost function $c_i : \mathcal{S}_i \to \mathbb{R}$ assigns a *production cost* to a given quantity, while $p : \bigtimes \mathcal{S}_i \to \mathbb{R}_{\ge 0}$ is the price of the good determined by the the joint choice of quantity $\boldsymbol{s} = (s_1, \dots, s_n)$ across the firms. Then, the utility of firm i is defined as $u_i(\boldsymbol{s}) \coloneqq s_i p(\boldsymbol{s}) - c_i(s_i)$. In *linear Cournot competition*, $p(\boldsymbol{s}) \coloneqq a - b\left(\sum_{i=1}^n s_i\right)$, for some $a, b > 0$, while the cost functions c_i are assumed to be smooth and convex [Even-Dar et al., 2009].

2.3 Online Linear Optimization and No-Regret Learning

In the *online learning framework* a learning agent has to select a strategy $\boldsymbol{x}^{(t)} \in \mathcal{X} \subseteq \mathbb{R}^d$ at every time $t \in \mathbb{N}$. Then, in the *full information* model, the learner receives as feedback from the environment a *linear* utility function $\boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{u}^{(t)} \rangle$, for some vector $\boldsymbol{u}^{(t)} \in \mathbb{R}^d$. The canonical measure of performance is the notion of *regret*, defined for a time horizon $T \in \mathbb{N}$ as follows.

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

That is, the performance of the agent is compared to the optimal *fixed* strategy in hindsight. It is important to note that *regret can be negative*. In the context of convex games, it is assumed that every player $i \in [\![n]\!]$ receives at time t the "linearized" utility function $\boldsymbol{x}_i \mapsto \langle \boldsymbol{x}_i, \boldsymbol{u}_i^{(t)} \rangle$, where $\boldsymbol{u}_i^{(t)} := \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}^{(t)})$. By concavity (Assumption 1),

$$\max_{\boldsymbol{x}_i^* \in \mathcal{X}_i} \sum_{t=1}^{T} \left(u_i(\boldsymbol{x}_i^*, \boldsymbol{x}_{-i}^{(t)}) - u_i(\boldsymbol{x}^{(t)}) \right) \le \max_{\boldsymbol{x}_i^* \in \mathcal{X}_i} \sum_{t=1}^{T} \langle \boldsymbol{x}_i^* - \boldsymbol{x}_i^{(t)}, \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}^{(t)}) \rangle.$$

As a result, a regret bound on the linearized regret—in the sense of (1)—automatically translates to a regret bound in the convex game.

Strongly Uncoupled Learning Dynamics In this setting, all learning dynamics are *uncoupled* in the sense of Hart and Mas-Colell [2003]: every player is oblivious to the other players' utilities. In fact, players need not have any prior knowledge about the game, even about their own utilities; this captures the condition of *strong uncoupledness* of Daskalakis et al. [2011], along with a suitable bound on the memory of each player.

3 Near-Optimal No-Regret Learning in Convex Games

In this section we describe our algorithm, *Log-Regularized Lifted Optimistic FTRL* (henceforth `LRL-OFTRL`). The central result of this section, Theorem 4, asserts that when all players learn using `LRL-OFTRL`, their regret only grows logarithmically with respect to the number of repetitions of the game. Detailed proofs for this section are available in Appendix A.

3.1 Setup

In the sequel, we will define and analyze the regret cumulated by `LRL-OFTRL` from the perspective of a generic player, omitting player subscripts.

We denote the set of strategies of the player by $\mathcal{X} \subseteq \mathbb{R}^d$. Without loss of generality, we will assume that $\mathcal{X} \subseteq [0, +\infty)^d$; otherwise, it suffices to first shift the set. Furthermore, we assume without loss of generality that there is no index $r \in [\![d]\!]$ such that $\boldsymbol{x}[r] = 0$ for all $\boldsymbol{x} \in \mathcal{X}$—if not, dropping the identically-zero dimension would not alter regret. We define the *lifting of set* $\mathcal{X}$ as the following set:

$$\mathbb{R}^{d+1} \supseteq \tilde{\mathcal{X}} := \{(\lambda, \boldsymbol{y}) : \lambda \in [0,1], \boldsymbol{y} \in \lambda\mathcal{X}\}. \tag{2}$$

Further, we define the ℓ_1-norm $\|\mathcal{X}\|_1$ of $\mathcal{X}$ as the maximum ℓ_1-norm of any vector $\boldsymbol{x} \in \mathcal{X}$, that is, $\|\mathcal{X}\|_1 := \max_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{x}\|_1$; for example, $\|\Delta^d\|_1 = 1$.

The *logarithmic regularizer* for $\mathbb{R}^{d+1}$ is the function

$$\mathcal{R}(\lambda, \boldsymbol{y}) := -\log \lambda - \sum_{r=1}^{d} \log \boldsymbol{y}[r], \qquad \forall (\lambda, \boldsymbol{y}) \in \mathbb{R}_{>0}^{d+1}.$$

Given any vector $(\lambda, \boldsymbol{y}) \in \tilde{\mathcal{X}} \cap \mathbb{R}_{>0}^{d+1}$, we denote with $\|\cdot\|_{(\lambda,\boldsymbol{y})}$ and $\|\cdot\|_{*,(\lambda,\boldsymbol{y})}$ the *local norms centered at* $(\lambda, \boldsymbol{y})$ induced by $\mathcal{R}(\lambda, \boldsymbol{y})$, defined as

$$\left\| \begin{pmatrix} a \\ \boldsymbol{z} \end{pmatrix} \right\|_{(\lambda,\boldsymbol{y})} := \sqrt{\left(\frac{a}{\lambda}\right)^2 + \sum_{r=1}^{d} \left(\frac{\boldsymbol{z}[r]}{\boldsymbol{y}[r]}\right)^2}, \qquad \left\| \begin{pmatrix} a \\ \boldsymbol{z} \end{pmatrix} \right\|_{*,(\lambda,\boldsymbol{y})} := \sqrt{(a\lambda)^2 + \sum_{r=1}^{d} (\boldsymbol{z}[r]\boldsymbol{y}[r])^2}$$

for any $(a, \boldsymbol{z}) \in \mathbb{R}^{d+1}$. These are the norms induced by the Hessian matrix of $\mathcal{R}$ at $(\lambda, \boldsymbol{y})$ and its inverse. It is a well-known fact that $\|\cdot\|_{*,(\lambda,\boldsymbol{y})}$ is the dual norm of $\|\cdot\|_{(\lambda,\boldsymbol{y})}$, and *vice versa*.

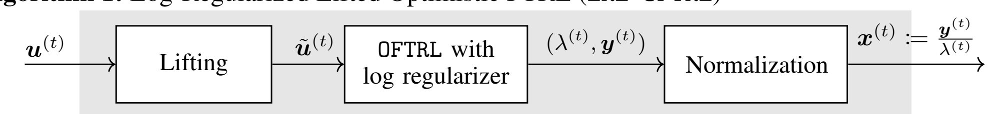
3.2 Overview of Our Algorithm

Our algorithm (Algorithm 1) leverages *optimistic follow the regularized leader* (`OFTRL`), a simple variant of `FTRL` introduced by Syrgkanis et al. [2015], but with some important twists. First, the optimization is performed over the lifting $\tilde{\mathcal{X}}$ of the set $\mathcal{X}$. More precisely, at every iteration the observed utility $\boldsymbol{u}^{(t)} \in \mathbb{R}^d$ will be transformed to $\tilde{\boldsymbol{u}}^{(t)} \in \mathbb{R}^{d+1}$ according to Line 6; this ensures that $\tilde{\boldsymbol{u}}^{(t)}$ is orthogonal to the vector $(1, \boldsymbol{x}^{(t)})$. Then, this utility vector $\tilde{\boldsymbol{u}}^{(t)}$ is given as input to a regret minimizer operating over $\tilde{\mathcal{X}}$, employing OFTRL under the logarithmic regularizer $\mathcal{R}(\lambda, \boldsymbol{y})$; this step is described in Line 3. We discuss how such an optimization problem can be solved efficiently in Section 3.5. Below we point out that Line 3 is indeed well-defined.

Proposition 1. *For any $\eta \geq 0$ and at all times $t \in \mathbb{N}$, the OFTRL optimization problem on Line 3 of Algorithm 1 admits a unique optimal solution $(\lambda^{(t)}, \boldsymbol{y}^{(t)}) \in \tilde{\mathcal{X}} \cap \mathbb{R}^{d+1}_{>0}$.*

Finally, given the iterate $(\lambda^{(t)}, \boldsymbol{y}^{(t)})$ output by the OFTRL step at time t, our regret minimizer over $\mathcal{X}$ selects the next strategy $\boldsymbol{x}^{(t)} := \boldsymbol{y}^{(t)}/\lambda^{(t)}$ (Line 4); this is indeed a valid strategy in $\mathcal{X}$ by definition of $\tilde{\mathcal{X}}$ in (2), as well as the fact that $\lambda^{(t)} > 0$ as asserted in Proposition 1.

Algorithm 1: Log-Regularized Lifted Optimistic FTRL (LRL-OFTRL)



Data: Learning rate η

1 Set $\tilde{\boldsymbol{U}}^{(1)}, \boldsymbol{u}^{(0)} \leftarrow \mathbf{0} \in \mathbb{R}^{d+1}$

2 **for** $t = 1, 2, \ldots, T$ **do**

3 Set $\begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \leftarrow \arg\max_{(\lambda, \boldsymbol{y}) \in \tilde{\mathcal{X}}} \left\{ \eta \left\langle \tilde{\boldsymbol{U}}^{(t)} + \tilde{\boldsymbol{u}}^{(t-1)}, \begin{pmatrix} \lambda \\ \boldsymbol{y} \end{pmatrix} \right\rangle + \log \lambda + \sum_{r=1}^{d} \log \boldsymbol{y}[r] \right\}$ [▷ OFTRL]

4 Play strategy $\boldsymbol{x}^{(t)} := \dfrac{\boldsymbol{y}^{(t)}}{\lambda^{(t)}} \in \mathcal{X}$ [▷ Normalization]

5 Observe $\boldsymbol{u}^{(t)} \in \mathbb{R}^d$

6 Set $\tilde{\boldsymbol{u}}^{(t)} \leftarrow \begin{pmatrix} -\langle \boldsymbol{u}^{(t)}, \boldsymbol{x}^{(t)} \rangle \\ \boldsymbol{u}^{(t)} \end{pmatrix}$ [▷ Lifting]

7 Set $\tilde{\boldsymbol{U}}^{(t+1)} \leftarrow \tilde{\boldsymbol{U}}^{(t)} + \tilde{\boldsymbol{u}}^{(t)}$

3.3 Regret Analysis

In this section, we study the regret of LRL-OFTRL under the idealized assumption that the optimization problem on Line 3 (OFTRL step) is solved exactly at each time t. In Section 3.5 we will relax that assumption, and study the regret of LRL-OFTRL when the solution to Line 3 is approximated using variants of Newton's method.

To study the regret Reg^T of LRL-OFTRL, defined in (1), it is useful to introduce the quantity $\tilde{\mathrm{Reg}}^T$, which measures the regret incurred *by the internal OFTRL algorithm* (Line 3) up to a time $T \in \mathbb{N}$ in the lifted space $\tilde{\mathcal{X}}$, *i.e.*,

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

As the following theorem clarifies, there is a strong connection between $\tilde{\mathrm{Reg}}^T$ and Reg^T.

Theorem 2. *For any time $T \in \mathbb{N}$ it holds that $\tilde{\mathrm{Reg}}^T = \max\{0, \mathrm{Reg}^T\}$. In particular, it follows that $\tilde{\mathrm{Reg}}^T \geq 0$ and $\mathrm{Reg}^T \leq \tilde{\mathrm{Reg}}^T$ for any $T \in \mathbb{N}$.*

The nonnegativity of $\tilde{\mathrm{Reg}}^T$ will be a crucial property in establishing Theorem 3. Further, Theorem 2 implies that a guarantee over the lifted space can be automatically translated to a regret bound over the original space $\mathcal{X}$. Now let

$$\| \cdot \|_t := \| \cdot \|_{(\lambda^{(t)}, \boldsymbol{y}^{(t)})} \qquad \text{and} \qquad \| \cdot \|_{*,t} := \| \cdot \|_{*,(\lambda^{(t)}, \boldsymbol{y}^{(t)})} \tag{3}$$

be the local norms centered at point $(\lambda^{(t)}, \boldsymbol{y}^{(t)})$ produced by `OFTRL` at time t (Line 3). In the next proposition we establish a refined RVU (Regret bounded by Variation in Utilities) bound in terms of this primal-dual norm pair.

Proposition 2 (RVU bound of `OFTRL` in local norms)**.** *Let $\tilde{\text{Reg}}^T$ be the regret cumulated up to time T by the internal `OFTRL` algorithm. If $\|\boldsymbol{u}^{(t)}\|_\infty \|\boldsymbol{x}\|_1 \leq 1$ at all times $t \in [\![T]\!]$, then for any time horizon $T \in \mathbb{N}$ and learning rate $\eta \leq \frac{1}{50}$,*

$$\tilde{\text{Reg}}^T \leq 4 + \frac{(d+1)\log T}{\eta} + 5\eta \sum_{t=1}^{T} \left\| \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\|_{*,t}^2 - \frac{1}{27\eta} \sum_{t=1}^{T-1} \left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_t^2 .$$

(We recall that $\tilde{\boldsymbol{u}}^{(0)} := \mathbf{0}$.) Proposition 2 differs from prior analogous results in that the regularizer is not a *barrier* over the feasible set. Next, we show that the iterates produced by `OFTRL` satisfy a refined notion of stability, which we refer to as *multiplicative stability*.

Proposition 3 (Multiplicative Stability)**.** *For any time $t \in \mathbb{N}$ and learning rate $\eta \leq \frac{1}{50}$, if $\|\boldsymbol{u}^{(t)}\|_\infty \|\boldsymbol{x}\|_1 \leq 1$,*

$$\left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_t \leq 22\eta.$$

Intuitively, this property ensures that coordinates of successive iterates will have a small multiplicative deviation. We leverage this refined notion of stability to establish the following crucial lemma.

Lemma 1. *For any time $t \in \mathbb{N}$ and learning rate $\eta \leq \frac{1}{50}$, if $\|\boldsymbol{u}^{(t)}\|_\infty \|\boldsymbol{x}\|_1 \leq 1$,*

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1 \leq 4\|\mathcal{X}\|_1 \left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_t .$$

Combining this lemma with Proposition 2 allows us to obtain an RVU bound for the original space $\mathcal{X}$, with no dependencies on local norms.

Corollary 1 (RVU bound in the original (unlifted) space)**.** *Fix any time $T \in \mathbb{N}$, and suppose that $\|\boldsymbol{u}^{(t)}\|_\infty \leq B$ for any $t \in [\![T]\!]$. If $\eta \leq \frac{1}{256B\|\mathcal{X}\|_1}$,*

$$\tilde{\text{Reg}}^T \leq 6B\|\mathcal{X}\|_1 + \frac{(d+1)\log T}{\eta} + 16\eta\|\mathcal{X}\|_1^2 \sum_{t=1}^{T-1} \|\boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}\|_\infty^2 - \frac{1}{512\eta\|\mathcal{X}\|_1^2} \sum_{t=1}^{T-1} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1^2 .$$

3.4 Main Result

So far, in Section 3.3, we have performed the analysis from the perspective of a single player, obtaining regret bounds that apply under an arbitrary sequence of utilities. Next, we assume that all players follow our dynamics such that the variation in one's utilities is now related to the variation in the joint strategies based on the smoothness condition of the utility function, connecting the last two terms of the RVU bound. Further leveraging the nonnegativity of the regrets in the lifted space, we establish that the second-order path lengths of the dynamics up to time T are bounded by $O(\log T)$:

Theorem 3. *Suppose that Assumption 1 holds for some parameters $B, L > 0$. If all players follow `LRL-OFTRL` with learning rate $\eta \leq \min\left\{ \frac{1}{256B\|\mathcal{X}\|_1}, \frac{1}{128nL\|\mathcal{X}\|_1^2} \right\}$, where $\|\mathcal{X}\|_1 := \max_{i \in [\![n]\!]} \|\mathcal{X}_i\|_1$, then*

$$\sum_{i=1}^{n} \sum_{t=1}^{T-1} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2 \leq 6144 n\eta B\|\mathcal{X}\|_1^3 + 1024 n(d+1)\|\mathcal{X}\|_1^2 \log T. \tag{4}$$

Here we made the mild assumption that each player knows the values of n, L, B and $\|\mathcal{X}\|_1$ in order to appropriately tune the learning rate; otherwise, similar guarantees are possible via a standard application of the doubling trick. It is interesting to point out that (4) holds even without the concavity condition (recall Assumption 1). We next leverage Theorem 3 to establish Theorem 1, the detailed version of which is given below.

Theorem 4 (Detailed Version of Theorem 1)**.** *Suppose that Assumption 1 holds for some parameters $B, L > 0$. If all players follow* `LRL-OFTRL` *with learning rate $\eta = \min\left\{\frac{1}{256B\|\mathcal{X}\|_1}, \frac{1}{128nL\|\mathcal{X}\|_1^2}\right\}$, then for any $T \in \mathbb{N}$ the regret Reg_i^T of each player $i \in [\![n]\!]$ can be bounded as*

$$\mathrm{Reg}_i^T \leq 12B\|\mathcal{X}\|_1 + 256(d+1)\max\left\{nL\|\mathcal{X}\|_1^2, 2B\|\mathcal{X}\|_1\right\}\log T. \tag{5}$$

Furthermore, the algorithm can be adaptive so that if player i is instead facing adversarial utilities, then $\mathrm{Reg}_i^T = O(\sqrt{T})$.

For clarity, below we cast (5) of Theorem 4 in normal-form games with utilities normalized in the range $[-1, 1]$, in which case we can take $B = 1$, $L = 1$ and $\|\mathcal{X}\|_1 = 1$.

Corollary 2 (Normal-form Games)**.** *Suppose that all players in a normal-form game with $n \geq 2$ follow* `LRL-OFTRL` *with learning rate $\eta = \frac{1}{128n}$. Then, for any $T \in \mathbb{N}$ and player $i \in [\![n]\!]$,*

$$\mathrm{Reg}_i^T \leq 12 + 256n(d+1)\log T.$$

3.5 Implementation and Iteration Complexity

In this section, we discuss the implementation and iteration complexity of `LRL-OFTRL`. The main difficulty in the implementation is the computation of the solution to the strictly concave *nonsmooth* constrained optimization problem in Line 3. We start by studying how the guarantees laid out in Theorem 4 are affected when the exact solution to the `OFTRL` problem (Line 4) in Algorithm 1 is replaced with an approximation. Specifically, suppose that at all times t the solution to the `OFTRL` step (Line 3) in Algorithm 1 is only *approximately* solved within tolerance $\epsilon^{(t)}$, in the sense that

$$\left\|\begin{pmatrix}\lambda^{(t)}\\ \boldsymbol{y}^{(t)}\end{pmatrix} - \begin{pmatrix}\lambda_\star^{(t)}\\ \boldsymbol{y}_\star^{(t)}\end{pmatrix}\right\|_{(\lambda_\star^{(t)}, \boldsymbol{y}_\star^{(t)})} \leq \epsilon^{(t)}, \tag{6}$$

where $(\lambda^{(t)}, \boldsymbol{y}^{(t)}) \in \mathbb{R}_{>0}^{d+1}$ and

$$\begin{pmatrix}\lambda_\star^{(t)}\\ \boldsymbol{y}_\star^{(t)}\end{pmatrix} := \operatorname*{arg\,max}_{(\lambda, \boldsymbol{y}) \in \tilde{\mathcal{X}}} \left\{\eta\left\langle \tilde{\boldsymbol{U}}^{(t)} + \tilde{\boldsymbol{u}}^{(t-1)}, \begin{pmatrix}\lambda\\ \boldsymbol{y}\end{pmatrix}\right\rangle + \log\lambda + \sum_{r=1}^{d}\log\boldsymbol{y}[r]\right\}.$$

Then, it can be proven directly from the definition of regret that the guarantees given in Corollary 1 deteriorate by an additive factor proportional to the sum of the tolerances $\sum_{t=1}^{T}\epsilon^{(t)}$. As an immediate corollary, when $\epsilon^{(t)} := \epsilon := 1/T$, the conclusion of Theorem 4 applies even when the solution to the optimization problem on Line 3 is only approximated up to ϵ tolerance. Therefore, to complete our construction, it suffices to show that it is indeed possible to efficiently compute approximate solutions to the `OFTRL` step (see Appendix A.5). In the remainder of this section, we show that this is indeed the case assuming access to two different type of oracles. It should be stressed that optimizing over a general convex set introduces several challenges not present under simplex domains, inevitably leading to an increased per-iteration complexity compared to algorithms designed specifically for normal-form games—such as OMWU.

Proximal Oracle First, we will assume access to a *proximal oracle* in local norm for the set $\tilde{\mathcal{X}}$, that is, access to a function that is able to compute the solution to the (positive-definite) quadratic optimization problem

$$\Pi_{\tilde{\boldsymbol{w}}}(\tilde{\boldsymbol{g}}) := \operatorname*{arg\,min}_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}}\left\{\tilde{\boldsymbol{g}}^\top\tilde{\boldsymbol{x}} + \frac{1}{2}\|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{w}}\|_{\tilde{\boldsymbol{w}}}^2\right\} = \operatorname*{arg\,min}_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}}\left\{\tilde{\boldsymbol{g}}^\top\tilde{\boldsymbol{x}} + \frac{1}{2}\sum_{r=1}^{d+1}\left(\frac{\tilde{\boldsymbol{x}}[r]}{\tilde{\boldsymbol{w}}[r]} - 1\right)^2\right\} \tag{7}$$

for arbitrary centers $\tilde{\boldsymbol{w}} \in \mathbb{R}_{>0}^{d+1}$ and gradients $\tilde{\boldsymbol{g}} \in \mathbb{R}^{d+1}$. For certain sets $\mathcal{X} \subseteq \mathbb{R}^d$, exact proximal oracles with polynomial complexity in the dimension d can be given. In particular, we show that this is the case when $\mathcal{X}$ is the strategy set of normal-form and extensive-form games by extending the approach of Gilpin [2009, pp. 128-133], as formalized below.

Proposition 4. *Let $\mathcal{X} \subseteq \mathbb{R}^d$ be the polytope of sequence-form strategies for a player in a perfect-recall extensive-form game. Then, the local proximal oracle $\Pi_{\tilde{\boldsymbol{w}}}(\tilde{\boldsymbol{g}})$ defined in (7) can be implemented exactly in time polynomial in the dimension d for any $\tilde{\boldsymbol{w}} \in \mathbb{R}_{>0}^{d+1}$ and $\tilde{\boldsymbol{g}} \in \mathbb{R}^{d+1}$.*

We provide the details and a more precise statement in Appendix B. In this context, the following guarantee employs the *proximal Newton algorithm* of Tran-Dinh et al. [2015]; see Algorithm 2.

Theorem 5 (Proximal Newton)**.** *Given any $\epsilon > 0$, it is possible to compute $(\lambda^{(t)}, \boldsymbol{y}^{(t)}) \in \tilde{\mathcal{X}} \cap \mathbb{R}^{d+1}_{>0}$ such that (6) holds for $\epsilon^{(t)} = \epsilon$ using $O(\log \log(1/\epsilon))$ operations and $O(\log \log(1/\epsilon))$ calls to the proximal oracle defined in Equation (7).*

Linear Maximization Oracle Moreover, we consider having access to a weaker *linear maximization oracle (LMO)* for the set $\mathcal{X}$:

$$\mathcal{L}_{\mathcal{X}}(\boldsymbol{u}) \coloneqq \arg\max_{\boldsymbol{x} \in \mathcal{X}} \langle \boldsymbol{x}, \boldsymbol{u} \rangle. \tag{8}$$

Such an oracle is more realistic in many settings [Jaggi, 2013], and it is particularly natural in the context of games, where it can be thought of as a *best response* oracle. We point out that an LMO for $\mathcal{X}$ automatically implies an LMO for $\tilde{\mathcal{X}}$. The following guarantee follows readily by applying the *Frank-Wolfe (projected) Newton method* [Liu et al., 2020, Algotihms 1 and 2].

Theorem 6 (Frank-Wolfe Newton)**.** *Given any $\epsilon > 0$, it is possible to compute $(\lambda^{(t)}, \boldsymbol{y}^{(t)}) \in \tilde{\mathcal{X}} \cap \mathbb{R}^{d+1}_{>0}$ such that (6) holds for $\epsilon^{(t)} = \epsilon$ using $O(\text{poly}(1/\epsilon))$ operations and $O(\text{poly}(1/\epsilon))$ calls to the LMO oracle defined in Equation (8).*

Experiments Finally, while our main contribution is of theoretical nature, we also support our theory by conducting experiments on some standard extensive-form games (Appendix C). The experiments verify that under `LRL-OFTRL` the regret of each player grows as $O(\log T)$.

4 Conclusions

In this paper we developed `LRL-OFTRL`, a novel no-regret learning algorithm. We showed that when all players in a general convex game employ `LRL-OFTRL`, the regret of each player grows only as $O(\log T)$, thereby significantly extending and strengthening the scope of all prior work. Further, our uncoupled no-regret learning dynamics can be efficiently implemented using, for example, a proximal oracle for the underlying feasible set.

One caveat of our framework applied to the special case of normal-form games is that the dependence on the dimension is linear (Corollary 2) as opposed to logarithmic [Daskalakis et al., 2021]. Whether the entropic regularizer—which induces OMWU—can be incorporated into our framework is an important open question. Another interesting avenue for future research would be to explore having access to different types of oracles. For example, is it possible to extend Theorem 4 using a *separation oracle* for the underlying set of strategies? If so, the ellipsoid algorithm [Bubeck, 2015] would be the obvious candidate en route to implementing `LRL-OFTRL`.

Acknowledgments

We are grateful to anonymous NeurIPS reviewers for many helpful comments. This material is based on work supported by the National Science Foundation under grants IIS-1718457, IIS-1901403, IIS-1943607, and CCF-1733556, and the ARO under award W911NF2010081. Christian Kroer is supported by the Office of Naval Research Young Investigator Program under grant N00014-22-1-2530.

References

Yoav Freund and Robert Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

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

Dean Foster and Rakesh Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.

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

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

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, 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*, May 2017.

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, page eaao1733, Dec. 2017.

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

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pages 3066–3074, 2013.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, pages 2989–2997, 2015.

Dylan J. Foster, Zhiyuan Li, Thodoris Lykouris, Karthik Sridharan, and Éva Tardos. Learning in games: Robustness of fast convergence. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pages 4727–4735, 2016.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

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

Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Beyond time-average convergence: Near-optimal uncoupled online learning via clairvoyant multiplicative weights update. *arXiv preprint*, 2022. URL `https://arxiv.org/abs/2111.14737v4`.

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

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 *International Conference on Machine Learning, ICML 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 6337–6357. PMLR, 2022.

Ehsan Asadi Kangarshahi, Ya-Ping Hsieh, Mehmet Fatih Sahin, and Volkan Cevher. Let's be honest: An optimal no-regret framework for zero-sum games. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 2493–2501. PMLR, 2018.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, pages 6–1, 2012.

Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, and Mengxiao Zhang. Bias no more: high-probability data-dependent regret bounds for adversarial bandits and mdps. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.

Haipeng Luo, Mengxiao Zhang, and Peng Zhao. Adaptive bandit convex optimization with heterogeneous curvature. In Po-Ling Loh and Maxim Raginsky, editors, *Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 1576–1612. PMLR, 2022.

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *In Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, 2008.

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

Sébastien Bubeck, Yuanzhi Li, Haipeng Luo, and Chen-Yu Wei. Improved path-length regret bounds for bandits. In *Conference on Learning Theory, COLT 2019, 25-28 June 2019*, volume 99 of *Proceedings of Machine Learning Research*, pages 508–528. PMLR, 2019.

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, 2022a.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2), 2012. ISSN 1935-8237.

Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

Boris T Polyak. Introduction to optimization. optimization software. *Inc., Publications Division, New York*, 1:32, 1987.

Drew Fudenberg and Alexander Peysakhovich. Recency, records and recaps: Learning and non-equilibrium behavior in a simple decision problem. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, page 971–986. Association for Computing Machinery, 2014.

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 *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, 2022*, pages 736–749. ACM, 2022b.

Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Optimal no-regret learning in general games: Bounded regret with unbounded step-sizes via clairvoyant MWU. *arXiv preprint*, 2021. URL `https://arxiv.org/abs/2111.14737v1`.

Tim Roughgarden and Florian Schoppmann. Local smoothness and the price of anarchy in splittable congestion games. *J. Econ. Theory*, 156:317–342, 2015.

Eyal Even-Dar, Yishay Mansour, and Uri Nadav. On the convergence of regret minimization dynamics in concave games. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009*, pages 523–532. ACM, 2009.

Tobias Harks and Max Klimm. Demand allocation games: Integrating discrete and continuous strategy spaces. In Ning Chen, Edith Elkind, and Elias Koutsoupias, editors, *Internet and Network Economics - 7th International Workshop, WINE 2011*, volume 7090 of *Lecture Notes in Computer Science*, pages 194–205. Springer, 2011.

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

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

Noah D. Stein, Pablo A. Parrilo, and Asuman E. Ozdaglar. Correlated equilibria in continuous games: Characterization and computation. *Games Econ. Behav.*, 71(2):436–455, 2011.

Gilles Stoltz and Gábor Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games Econ. Behav.*, 59(1):187–208, 2007.

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

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

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

Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93:1830–1836, 2003.

Andrew Gilpin. *Algorithms for abstracting and solving imperfect information games*. Doctoral dissertation, Computer Science Department, Carnegie Mellon University, Pittsburgh, 2009.

Quoc Tran-Dinh, Anastasios Kyrillidis, and Volkan Cevher. Composite self-concordant minimization. *J. Mach. Learn. Res.*, 16:371–416, 2015.

Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013*, volume 28 of *JMLR Workshop and Conference Proceedings*, pages 427–435. JMLR.org, 2013.

Deyi Liu, Volkan Cevher, and Quoc Tran-Dinh. A newton frank-wolfe method for constrained self-concordant minimization. *CoRR*, abs/2002.07003, 2020.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.*, 8(3-4):231–357, 2015.

Samid Hoda, Andrew Gilpin, Javier Peña, and Tuomas Sandholm. Smoothing techniques for computing Nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2), 2010.

H. W. Kuhn. Extensive games. *Proc. of the National Academy of Sciences*, 36:570–576, 1950.

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

Gabriele Farina, Chun Kai Ling, Fei Fang, and Tuomas Sandholm. Correlation in extensive-form games: Saddle-point formulation and benchmarks. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] see Sections 1 and 4
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 3

(b) Did you include complete proofs of all theoretical results? [Yes] See Appendices A and B

3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Appendix C
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Appendix C
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Appendix C
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Omitted Proofs

In this section we include all of the proofs omitted from the main body. For the convenience of the reader, we will restate each claim before proceeding with its proof.

A.1 Preliminary Proofs

We commence with the proof of Proposition 1.

Proposition 1. *For any $\eta \geq 0$ and at all times $t \in \mathbb{N}$, the* `OFTRL` *optimization problem on Line 3 of Algorithm 1 admits a unique optimal solution $(\lambda^{(t)}, \boldsymbol{y}^{(t)}) \in \tilde{\mathcal{X}} \cap \mathbb{R}^{d+1}_{>0}$.*

Proof. Uniqueness follow immediately from strict convexity. In the rest of the proof we focus on the existence part.

We start by showing that there exists a point $\tilde{x} \in \tilde{\mathcal{X}}$ whose coordinates are all strictly positive. By hypothesis (see Section 3.1), for every coordinate $r \in [\![d]\!]$, there exists a point $\boldsymbol{x}_r$ such that $\boldsymbol{x}_r[r] > 0$. Hence, by convexity of $\mathcal{X} \subseteq [0, +\infty)^d$ and by definition of $\tilde{\mathcal{X}}$, the point

$$(1, \boldsymbol{x}^\circ) := \left(1,\ \frac{1}{d}\sum_{r=1}^{d} \boldsymbol{x}_r\right).$$

is such that $(1, \boldsymbol{x}^\circ) \in \tilde{\mathcal{X}} \cap \mathbb{R}^d_{>0}$.

Let now M be the ℓ_∞ norm of the linear part in the `OFTRL` step (Line 3 of Algorithm 1). Then, a *lower bound* on the optimal value $v^\star$ of objective is obtained by plugging in the point $(1, \boldsymbol{x}^\circ)$ at least

$$v^\star \geq -M(1 + \|\mathcal{X}\|_1) + \sum_{r=1}^{d} \log \boldsymbol{x}^\circ[r]. \tag{9}$$

Let now

$$m := \exp\left\{-(2M + d)(1 + \|\mathcal{X}\|_1) + \sum_{r=1}^{d} \log \boldsymbol{x}^\circ[r]\right\} > 0. \tag{10}$$

We will show that any point $(\lambda, \boldsymbol{y}) \notin [m, +\infty) \cap \tilde{\mathcal{X}}$ cannot be optimal for the `OFTRL` objective. Indeed, take a point $(\lambda, \boldsymbol{y}) \notin [m, +\infty) \cap \tilde{\mathcal{X}}$. Then, at least one coordinate of $(\lambda, \boldsymbol{y})$ is strictly less than m. If $\lambda < m$, then the objective value at $(\lambda, \boldsymbol{y})$ is at most

$$\begin{aligned} M\lambda + M\|\mathcal{X}\|_1 + \log\lambda + \sum_{r=1}^{d} \log \boldsymbol{y}[r] &\leq M(1 + \|\mathcal{X}\|_1) + \log m + \sum_{r=1}^{d} \log \|\mathcal{X}\|_1 \\ &\leq M(1 + \|\mathcal{X}\|_1) + \log m + d(\|\mathcal{X}\|_1 - 1) \\ &< (M + d)(1 + \|\mathcal{X}\|_1) + \log m \\ &= -M(1 + \|\mathcal{X}\|_1) + \sum_{r=1}^{d} \log \boldsymbol{x}^\circ[r] && \text{(from (10))} \\ &\leq v^*, && \text{(from (9))} \end{aligned}$$

where the first inequality follows from upper bounding any coordinate of $\boldsymbol{y}$ with $\|\mathcal{X}\|_1$, and the second inequality follows from using the inequality $\log z \leq z - 1$, valid for all $z \in (0, +\infty)$. Similarly, if $\boldsymbol{y}[s] < m$ for some $s \in [\![d]\!]$, then we can upper bound the objective value at $(\lambda, \boldsymbol{y})$ as

$$\begin{aligned} M + M\|\mathcal{X}\|_1 + \log 1 + \sum_{r=1}^{d} \log \boldsymbol{y}[r] &\leq M(1 + \|\mathcal{X}\|_1) + \log m + \sum_{r=1}^{d} \log \|\mathcal{X}\|_1 \\ &\leq M(1 + \|\mathcal{X}\|_1) + (d - 1)(\|\mathcal{X}\|_1 - 1) + \log m \\ &< (M + d)(1 + \|\mathcal{X}\|_1) + \log m \leq v^*. \end{aligned}$$

So, in either case, we see that no optimal point can have any coordinate strictly less than m. Consequently, the maximizer of the OFTRL step lies in the set $\mathcal{S} \coloneqq [m, +\infty)^{d+1} \cap \tilde{\mathcal{X}}$. Since both $[m, +\infty)^{d+1}$ and $\tilde{\mathcal{X}}$ are closed, and since $\tilde{\mathcal{X}}$ is bounded by hypothesis, the set $\mathcal{S}$ is compact. Furthermore, note that $\mathcal{S}$ is nonempty, as $(1, \boldsymbol{x}^\circ) \in \mathcal{S}$, as for any $s \in [\![d]\!]$

$$\begin{aligned}
\log m &= -(2M + d)(1 + \|\mathcal{X}\|_1) + \sum_{r=1}^{d} \log \boldsymbol{x}^\circ[r] \\
&\leq -(2M + d)(1 + \|\mathcal{X}\|_1) + \log \boldsymbol{x}^\circ[s] + (d-1) \log \|\mathcal{X}\|_1 \\
&\leq -(2M + d)(1 + \|\mathcal{X}\|_1) + \log \boldsymbol{x}^\circ[s] + (d-1)(\|\mathcal{X}\|_1 - 1) \\
&\leq \log \boldsymbol{x}^\circ[s],
\end{aligned}$$

implying that $(1, \boldsymbol{x}^\circ) \in [m, +\infty)^{d+1}$. Since $\mathcal{S}$ is compact and nonempty and the objective function is continuous, the optimization problem attains an optimal solution on $\mathcal{S}$ by virtue of Weierstrass' theorem. □

Theorem 2. *For any time $T \in \mathbb{N}$ it holds that $\tilde{\mathrm{Reg}}^T = \max\{0, \mathrm{Reg}^T\}$. In particular, it follows that $\tilde{\mathrm{Reg}}^T \geq 0$ and $\mathrm{Reg}^T \leq \tilde{\mathrm{Reg}}^T$ for any $T \in \mathbb{N}$.*

Proof. First, by definition of $\tilde{\boldsymbol{u}}^{(t)}$ in Line 6, it follows that for any t,

$$\left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\rangle = \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} 1 \\ \boldsymbol{x}^{(t)} \end{pmatrix} \right\rangle = 0.$$

As a result, we have that $\max\{0, \mathrm{Reg}^T\}$ is equal to

$$\begin{aligned}
&\max\left\{0, \max_{\boldsymbol{x}^* \in \mathcal{X}} \sum_{t=1}^{T} \langle \boldsymbol{u}^{(t)}, \boldsymbol{x}^* - \boldsymbol{x}^{(t)} \rangle \right\} = \max\left\{0, \max_{\boldsymbol{x}^* \in \mathcal{X}} \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} 1 \\ \boldsymbol{x}^* \end{pmatrix} - \begin{pmatrix} 1 \\ \boldsymbol{x}^{(t)} \end{pmatrix} \right\rangle \right\} \\
&= \max\left\{0, \max_{\boldsymbol{x}^* \in \mathcal{X}} \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} 1 \\ \boldsymbol{x}^* \end{pmatrix} \right\rangle \right\} = \max_{(\lambda^*, \boldsymbol{y}^*) \in \tilde{\mathcal{X}}} \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} \lambda^* \\ \boldsymbol{y}^* \end{pmatrix} \right\rangle \\
&= \max_{(\lambda^*, \boldsymbol{y}^*) \in \tilde{\mathcal{X}}} \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} \lambda^* \\ \boldsymbol{y}^* \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\rangle = \tilde{\mathrm{Reg}}^T,
\end{aligned}$$

as we wanted to show. □

A.2 Analysis of OFTRL with Logarithmic Regularizer

For notational convenience, we define the log-regularizer $\psi : \tilde{\mathcal{X}} \to \mathbb{R}_{\geq 0}$ as

$$\psi(\tilde{\boldsymbol{x}}) \coloneqq -\frac{1}{\eta} \sum_{r=1}^{d+1} \log \tilde{\boldsymbol{x}}[r],$$

and its induced Bregman divergence

$$D_\psi(\tilde{\boldsymbol{x}} \,\|\, \tilde{\boldsymbol{z}}) \coloneqq \frac{1}{\eta} \sum_{r=1}^{d+1} h\left(\frac{\tilde{\boldsymbol{x}}[r]}{\tilde{\boldsymbol{z}}[r]}\right), \quad \text{where } h(a) = a - 1 - \ln(a).$$

Moreover, we define

$$\tilde{\boldsymbol{x}}^{(t)} = \arg\max_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}} -F_t(\tilde{\boldsymbol{x}}) = \arg\min_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}} F_t(\tilde{\boldsymbol{x}}), \quad \text{where } F_t(\tilde{\boldsymbol{x}}) = -\left\langle \tilde{\boldsymbol{U}}^{(t)} + \tilde{\boldsymbol{u}}^{(t-1)}, \tilde{\boldsymbol{x}} \right\rangle + \psi(\tilde{\boldsymbol{x}}). \tag{11}$$

We note that F_t is a convex function for each t and $\tilde{\boldsymbol{x}}^{(t)}$ is exactly equal to $\begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix}$ computed by Algorithm 1. Further, we define an auxiliary sequence $\{\tilde{\boldsymbol{z}}^{(t)}\}_{t=1,2,\ldots}$ defined as follows.

$$\tilde{\boldsymbol{z}}^{(t)} = \arg\max_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}} -G_t(\tilde{\boldsymbol{x}}) = \arg\min_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}} G_t(\tilde{\boldsymbol{x}}), \quad \text{where } G_t(\tilde{\boldsymbol{x}}) = -\left\langle \tilde{\boldsymbol{U}}^{(t)}, \tilde{\boldsymbol{x}} \right\rangle + \psi(\tilde{\boldsymbol{x}}). \tag{12}$$

Similarly, G_t is a convex function for each t. We also recall the primal and dual norm notation:

$$\|\tilde{\boldsymbol{z}}\|_t = \sum_{r=1}^{d+1} \left(\frac{\tilde{\boldsymbol{z}}[r]}{\tilde{\boldsymbol{x}}^{(t)}[r]} \right)^2, \quad \|\tilde{\boldsymbol{z}}\|_{*,t} = \sum_{r=1}^{d+1} \left(\tilde{\boldsymbol{x}}^{(t)}[r] \tilde{\boldsymbol{z}}[r] \right)^2.$$

Finally, for a $(d+1) \times (d+1)$ positive definite matrix $\mathbf{M}$, we use $\|\tilde{\boldsymbol{z}}\|_{\mathbf{M}}$ to denote the induced quadratic norm $\sqrt{\tilde{\boldsymbol{z}}^\top \mathbf{M} \tilde{\boldsymbol{z}}}$. We are now ready to establish Proposition 2.

Proposition 2 (RVU bound of `OFTRL` in local norms)**.** *Let $\tilde{\text{Reg}}^T$ be the regret cumulated up to time T by the internal `OFTRL` algorithm. If $\|\boldsymbol{u}^{(t)}\|_\infty \|\boldsymbol{x}\|_1 \leq 1$ at all times $t \in [\![T]\!]$, then for any time horizon $T \in \mathbb{N}$ and learning rate $\eta \leq \frac{1}{50}$,*

$$\tilde{\text{Reg}}^T \leq 4 + \frac{(d+1)\log T}{\eta} + 5\eta \sum_{t=1}^{T} \left\| \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\|_{*,t}^2 - \frac{1}{27\eta} \sum_{t=1}^{T-1} \left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_t^2.$$

Proof of Proposition 2. For any comparator $\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}$, define $\tilde{\boldsymbol{x}}' = \frac{T-1}{T} \cdot \tilde{\boldsymbol{x}} + \frac{1}{T} \cdot \tilde{\boldsymbol{x}}^{(1)} \in \tilde{\mathcal{X}}$, where we recall $\tilde{\boldsymbol{x}}^{(1)} = \arg\min_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}} F_1(\tilde{\boldsymbol{x}}) = \arg\min_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}} \psi(\tilde{\boldsymbol{x}})$. Then, we have

$$\begin{aligned}
\sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle &= \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}', \tilde{\boldsymbol{u}}^{(t)} \right\rangle + \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}}' - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle \\
&= \frac{1}{T} \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle + \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}}' - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle \\
&\leq 4 + \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}}' - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle,
\end{aligned}$$

where the last inequality follows from Cauchy-Schwarz together with the assumption that $\|\boldsymbol{u}^{(t)}\|_\infty \leq \frac{1}{\|\mathcal{X}\|_1}$.

Now, by standard Optimistic FTRL analysis (see Lemma 2), the last term $\sum_{t=1}^{T} \langle \tilde{\boldsymbol{x}}' - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \rangle$ (cumulative regret against $\tilde{\boldsymbol{x}}'$) is bounded by

$$\begin{aligned}
\sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}}' - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle \leq \psi(\tilde{\boldsymbol{x}}') - \psi(\tilde{\boldsymbol{x}}^{(1)}) &+ \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{z}}^{(t+1)} - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\rangle \\
&- \sum_{t=1}^{T} \left(D_\psi \left(\tilde{\boldsymbol{x}}^{(t)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t)} \right) + D_\psi \left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)} \right) \right).
\end{aligned}$$

For the term $\psi(\tilde{\boldsymbol{x}}') - \psi(\tilde{\boldsymbol{x}}^{(1)})$, a direct calculation using definitions shows

$$\psi(\tilde{\boldsymbol{x}}') - \psi(\tilde{\boldsymbol{x}}^{(1)}) = \frac{1}{\eta} \sum_{i=1}^{d+1} \log \frac{\tilde{\boldsymbol{x}}^{(1)}[i]}{\tilde{\boldsymbol{x}}'[i]} \leq \frac{d+1}{\eta} \log T.$$

For the other terms, we apply Lemma 3 and Lemma 5, which completes the proof. □

Lemma 2. *The update rule (11) ensures the following for any $\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}$:*

$$\begin{aligned}
\sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle \leq \psi(\tilde{\boldsymbol{x}}) - \psi(\tilde{\boldsymbol{x}}^{(1)}) &+ \sum_{t=1}^{T} \left\langle \tilde{\boldsymbol{z}}^{(t+1)} - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\rangle \\
&- \sum_{t=1}^{T} \left(D_\psi \left(\tilde{\boldsymbol{x}}^{(t)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t)} \right) + D_\psi \left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)} \right) \right).
\end{aligned}$$

Proof. First note that for any convex function $F : \tilde{\mathcal{X}} \to \mathbb{R}$ and a minimizer $\tilde{\boldsymbol{x}}^\star$, we have for any $\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}$:

$$F(\tilde{\boldsymbol{x}}^\star) = F(\tilde{\boldsymbol{x}}) - \langle \nabla F(\tilde{\boldsymbol{x}}^\star), \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^\star \rangle - D_F(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}^\star) \leq F(\tilde{\boldsymbol{x}}) - D_F(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}^\star),$$

where D_F is the Bregman Divergence induced by F and the inequality is by the first-order optimality. Using this fact and the optimality of $\tilde{\boldsymbol{z}}^{(t)}$, we have

$$G_t(\tilde{\boldsymbol{z}}^{(t)}) \le G_t(\tilde{\boldsymbol{x}}^{(t)}) - D_\psi\left(\tilde{\boldsymbol{x}}^{(t)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t)}\right)$$
$$= F_t(\tilde{\boldsymbol{x}}^{(t)}) + \left\langle \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t-1)} \right\rangle - D_\psi\left(\tilde{\boldsymbol{x}}^{(t)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t)}\right)$$

Similarly, using the optimality of $\tilde{\boldsymbol{x}}^{(t)}$, we have

$$F_t(\tilde{\boldsymbol{x}}^{(t)}) \le F_t(\tilde{\boldsymbol{z}}^{(t+1)}) - D_\psi\left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)}\right)$$
$$= G_{t+1}(\tilde{\boldsymbol{z}}^{(t+1)}) + \left\langle \tilde{\boldsymbol{z}}^{(t+1)}, \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\rangle - D_\psi\left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)}\right)$$

Combining the inequalities and summing over t, we have

$$G_1(\tilde{\boldsymbol{z}}^{(1)}) \le G_{T+1}(\tilde{\boldsymbol{z}}^{(T+1)}) + \sum_{t=1}^{T}\left(\left\langle \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} \right\rangle + \left\langle \tilde{\boldsymbol{z}}^{(t+1)} - \tilde{\boldsymbol{x}}^{(t)}, \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\rangle\right)$$
$$+ \sum_{t=1}^{T}\left(-D_\psi\left(\tilde{\boldsymbol{x}}^{(t)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t)}\right) - D_\psi\left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)}\right)\right).$$

Observe that $G_1(\tilde{\boldsymbol{z}}^{(1)}) = \psi(\tilde{\boldsymbol{x}}^{(1)})$ and $G_{T+1}(\tilde{\boldsymbol{z}}^{(T+1)}) \le -\left\langle \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{U}}^{(T+1)} \right\rangle + \psi(\tilde{\boldsymbol{x}})$. Rearranging then proves the lemma. □

Lemma 3. *If $\eta \le \frac{1}{50}$, then we have*

$$\left\|\tilde{\boldsymbol{z}}^{(t+1)} - \tilde{\boldsymbol{x}}^{(t)}\right\|_t \le 5\eta\left\|\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\right\|_{*,t} \le 10\sqrt{2}\eta \le 15\eta, \tag{13}$$
$$\left\|\tilde{\boldsymbol{x}}^{(t+1)} - \tilde{\boldsymbol{x}}^{(t)}\right\|_t \le 5\eta\left\|2\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\right\|_{*,t} \le 15\sqrt{2}\eta \le 22\eta. \tag{14}$$

Proof. The second part of both inequalities is clear by definitions:

$$\left\|\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\right\|_{*,t}^2 = \left(\lambda^{(t)}\left(\left\langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \right\rangle - \left\langle \boldsymbol{x}^{(t-1)}, \boldsymbol{u}^{(t-1)} \right\rangle\right)\right)^2 + \sum_{r=1}^{d}\left(\boldsymbol{y}^{(t)}[r]\left(\boldsymbol{u}^{(t)}[r] - \boldsymbol{u}^{(t-1)}[r]\right)\right)^2$$
$$\le 4(\lambda^{(t)})^2 + \frac{4}{\|\mathcal{X}\|_1^2}\sum_{r=1}^{d}\left(\boldsymbol{y}^{(t)}[r]\right)^2 \le 8,$$

where we use $\left\langle \boldsymbol{x}^{(\tau)}, \boldsymbol{u}^{(\tau)} \right\rangle \le \|\boldsymbol{x}^{(\tau)}\|_1\|\boldsymbol{u}^{(\tau)}\|_\infty \le 1$ and $|\boldsymbol{u}^{(\tau)}[r]| \le \frac{1}{\|\mathcal{X}\|_1}$ for any time τ and any coordinate r by the assumption, and similarly,

$$\left\|2\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\right\|_{*,t}^2 = \left(\lambda^{(t)}\left(2\left\langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \right\rangle - \left\langle \boldsymbol{x}^{(t-1)}, \boldsymbol{u}^{(t-1)} \right\rangle\right)\right)^2 + \sum_{r=1}^{d}\left(\boldsymbol{y}^{(t)}[r]\left(2\boldsymbol{u}^{(t)}[r] - \boldsymbol{u}^{(t-1)}[r]\right)\right)^2$$
$$\le 9(\lambda^{(t)})^2 + \frac{9}{\|\mathcal{X}\|_1^2}\sum_{r=1}^{d}\left(\boldsymbol{y}^{(t)}[r]\right)^2 \le 18$$

To prove the first inequality in Eq. (13), let $\mathcal{E}_t = \left\{\tilde{\boldsymbol{x}} : \left\|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)}\right\|_t \le 5\eta\left\|\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\right\|_{*,t}\right\}$. Noticing that $\tilde{\boldsymbol{z}}^{(t+1)}$ is the minimizer of the convex function G_{t+1}, to show $\tilde{\boldsymbol{z}}^{(t+1)} \in \mathcal{E}_t$, it suffices to show that for all $\tilde{\boldsymbol{x}}$ on the boundary of $\mathcal{E}_t$, we have $G_{t+1}(\tilde{\boldsymbol{x}}) \ge G_{t+1}(\tilde{\boldsymbol{x}}^{(t)})$. Indeed, using

Taylor's theorem, for any such $\tilde{\boldsymbol{x}}$, there is a point $\boldsymbol{\xi}$ on the line segment between $\tilde{\boldsymbol{x}}^{(t)}$ and $\tilde{\boldsymbol{x}}$ such that

$$\begin{aligned}
G_{t+1}(\tilde{\boldsymbol{x}}) &= G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}) + \left\langle \nabla G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}), \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\rangle + \frac{1}{2} \left\| \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\|^2_{\nabla^2 G_{t+1}(\boldsymbol{\xi})} \\
&= G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}) - \left\langle \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}, \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\rangle + \left\langle \nabla F_t(\tilde{\boldsymbol{x}}^{(t)}), \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\rangle + \frac{1}{2} \left\| \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\|^2_{\nabla^2 \psi(\boldsymbol{\xi})} \\
&\geq G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}) - \left\langle \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}, \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\rangle + \frac{1}{2} \left\| \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\|^2_{\nabla^2 \psi(\boldsymbol{\xi})} && \text{(by the optimality of } \tilde{\boldsymbol{x}}^{(t)}) \\
&\geq G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}) - \left\| \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\|_{*,t} \left\| \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\|_t + \frac{1}{2} \left\| \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\|^2_{\nabla^2 \psi(\boldsymbol{\xi})}. && \text{(by Hölder's inequality)} \\
&\geq G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}) - \left\| \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\|_{*,t} \left\| \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\|_t + \frac{2}{9\eta} \left\| \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)} \right\|^2_t && (\star) \\
&= G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}) + \frac{5}{9}\eta \left\| \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\|^2_{*,t} && (\|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)}\|_t = 5\eta \|\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\|_{*,t}) \\
&\geq G_{t+1}(\tilde{\boldsymbol{x}}^{(t)}).
\end{aligned}$$

Here, the inequality $(\star)$ holds because Lemma 4 (together with the condition $\eta \leq \frac{1}{50}$) shows $\frac{1}{2}\tilde{\boldsymbol{x}}^{(t)}[i] \leq \tilde{\boldsymbol{x}}[i] \leq \frac{3}{2}\tilde{\boldsymbol{x}}^{(t)}[i]$, which implies $\frac{1}{2}\tilde{\boldsymbol{x}}^{(t)}[i] \leq \boldsymbol{\xi}[i] \leq \frac{3}{2}\tilde{\boldsymbol{x}}^{(t)}[i]$ as well, and thus $\nabla^2\psi(\boldsymbol{\xi}) \succeq \frac{4}{9}\nabla^2\psi(\tilde{\boldsymbol{x}}^{(t)})$. This finishes the proof for Eq. (13). The first inequality of Eq. (14) can be proven in the same manner. ☐

Proposition 3 (Multiplicative Stability)**.** *For any time $t \in \mathbb{N}$ and learning rate $\eta \leq \frac{1}{50}$, if $\|\boldsymbol{u}^{(t)}\|_\infty \|\boldsymbol{x}\|_1 \leq 1$,*

$$\left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_t \leq 22\eta.$$

Proof. The statement is proved in Lemma 3. ☐

Lemma 4. *If $\tilde{\boldsymbol{x}}$ satisfies $\|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)}\|_t \leq \frac{1}{2}$, then $\frac{1}{2}\tilde{\boldsymbol{x}}^{(t)}[i] \leq \tilde{\boldsymbol{x}}[i] \leq \frac{3}{2}\tilde{\boldsymbol{x}}^{(t)}[i]$ for every coordinate i.*

Proof. By definition, $\|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^{(t)}\|_t \leq \frac{1}{2}$ implies for any i, $\frac{|\tilde{\boldsymbol{x}}[i] - \tilde{\boldsymbol{x}}^{(t)}[i]|}{\tilde{\boldsymbol{x}}^{(t)}[i]} \leq \frac{1}{2}$, and thus $\frac{1}{2}\tilde{\boldsymbol{x}}^{(t)}[i] \leq \tilde{\boldsymbol{x}}[i] \leq \frac{3}{2}\tilde{\boldsymbol{x}}^{(t)}[i]$. ☐

Lemma 5. *If $\eta \leq \frac{1}{50}$, then we have*

$$\sum_{t=1}^{T} \left(D_\psi \left(\tilde{\boldsymbol{x}}^{(t)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t)} \right) + D_\psi \left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)} \right) \right) \geq \frac{1}{27\eta} \sum_{t=1}^{T-1} \left\| \tilde{\boldsymbol{x}}^{(t+1)} - \tilde{\boldsymbol{x}}^{(t)} \right\|^2_t.$$

Proof. Recall $h(a) = a - 1 - \ln(a)$ and $D_\psi(\tilde{\boldsymbol{x}} \,\|\, \tilde{\boldsymbol{z}}) = \frac{1}{\eta}\sum_{i=1}^{d+1} h\left(\frac{\tilde{\boldsymbol{x}}[i]}{\tilde{\boldsymbol{z}}[i]}\right)$. We proceed as

$$\begin{aligned}
&\sum_{t=1}^{T}\left(D_\psi\left(\tilde{\boldsymbol{x}}^{(t)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t)}\right) + D_\psi\left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)}\right)\right)\\
&\qquad\geq \sum_{t=1}^{T-1}\left(D_\psi\left(\tilde{\boldsymbol{x}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{z}}^{(t+1)}\right) + D_\psi\left(\tilde{\boldsymbol{z}}^{(t+1)} \,\middle\|\, \tilde{\boldsymbol{x}}^{(t)}\right)\right)\\
&\qquad= \frac{1}{\eta}\sum_{t=1}^{T-1}\sum_{i=1}^{d+1}\left(h\left(\frac{\tilde{\boldsymbol{x}}^{(t+1)}[i]}{\tilde{\boldsymbol{z}}^{(t+1)}[i]}\right) + h\left(\frac{\tilde{\boldsymbol{z}}^{(t+1)}[i]}{\tilde{\boldsymbol{x}}^{(t)}[i]}\right)\right)\\
&\qquad\geq \frac{1}{6\eta}\sum_{t=1}^{T-1}\sum_{i=1}^{d+1}\left(\frac{(\tilde{\boldsymbol{x}}^{(t+1)}[i] - \tilde{\boldsymbol{z}}^{(t+1)}[i])^2}{\left(\tilde{\boldsymbol{z}}^{(t+1)}[i]\right)^2} + \frac{(\tilde{\boldsymbol{z}}^{(t+1)}[i] - \tilde{\boldsymbol{x}}^{(t)}[i])^2}{(\tilde{\boldsymbol{x}}^{(t)}[i])^2}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left(h(y) \geq \tfrac{(y-1)^2}{6} \text{ for } y \in [\tfrac{1}{3}, 3]\right)\\
&\qquad\geq \frac{2}{27\eta}\sum_{t=1}^{T-1}\sum_{i=1}^{d+1}\left(\frac{(\tilde{\boldsymbol{x}}^{(t+1)}[i] - \tilde{\boldsymbol{z}}^{(t+1)}[i])^2}{\left(\tilde{\boldsymbol{x}}^{(t)}[i]\right)^2} + \frac{(\tilde{\boldsymbol{z}}^{(t+1)}[i] - \tilde{\boldsymbol{x}}^{(t)}[i])^2}{(\tilde{\boldsymbol{x}}^{(t)}[i])^2}\right)\\
&\qquad\geq \frac{1}{27\eta}\sum_{t=1}^{T-1}\sum_{i=1}^{d+1}\left(\frac{(\tilde{\boldsymbol{x}}^{(t+1)}[i] - \tilde{\boldsymbol{x}}^{(t)}[i])^2}{\left(\tilde{\boldsymbol{x}}^{(t)}[i]\right)^2}\right) = \frac{1}{27\eta}\sum_{t=1}^{T-1}\left\|\tilde{\boldsymbol{x}}^{(t+1)} - \tilde{\boldsymbol{x}}^{(t)}\right\|_t^2.
\end{aligned}$$

Here, the second and the third inequality hold because by Lemma 3 and Lemma 4, we have $\frac{1}{2} \leq \frac{\tilde{\boldsymbol{z}}^{(t+1)}[i]}{\tilde{\boldsymbol{x}}^{(t)}[i]} \leq \frac{3}{2}$ and $\frac{1}{2} \leq \frac{\tilde{\boldsymbol{x}}^{(t+1)}[i]}{\tilde{\boldsymbol{x}}^{(t)}[i]} \leq \frac{3}{2}$, and thus $\frac{1}{3} \leq \frac{\tilde{\boldsymbol{x}}^{(t+1)}[i]}{\tilde{\boldsymbol{z}}^{(t+1)}[i]} \leq 3$. □

A.3 RVU Bound in the Original Space

Next, we establish an RVU bound in the original (unlifted) space, namely Corollary 1. To this end, we first proceed with the proof of Lemma 1, which boils down to the following simple claim.

Lemma 6. *Let $(\lambda, \boldsymbol{y}), (\lambda', \boldsymbol{y}') \in \tilde{\mathcal{X}} \cap \mathbb{R}_{>0}^{d+1}$ be arbitrary points such that*

$$\left\|\begin{pmatrix}\lambda'\\ \boldsymbol{y}'\end{pmatrix} - \begin{pmatrix}\lambda\\ \boldsymbol{y}\end{pmatrix}\right\|_{(\lambda,\boldsymbol{y})} \leq \frac{1}{2}.$$

Then,

$$\left\|\frac{\boldsymbol{y}}{\lambda} - \frac{\boldsymbol{y}'}{\lambda'}\right\|_1 \leq 4\|\mathcal{X}\|_1 \cdot \left\|\begin{pmatrix}\lambda'\\ \boldsymbol{y}'\end{pmatrix} - \begin{pmatrix}\lambda\\ \boldsymbol{y}\end{pmatrix}\right\|_{(\lambda,\boldsymbol{y})}.$$

Proof. Let μ be defined as

$$\mu := \max\left\{\left|\frac{\lambda'}{\lambda} - 1\right|, \max_{r\in[\![d]\!]}\left|\frac{\boldsymbol{y}'[r]}{\boldsymbol{y}[r]} - 1\right|\right\}. \tag{15}$$

By definition,

$$\left|\frac{\lambda'}{\lambda} - 1\right| \leq \mu,$$

which in turn implies that

$$(1-\mu)\lambda \leq \lambda' \leq (1+\mu)\lambda. \tag{16}$$

Similarly, for any $r \in [\![d]\!]$,

$$(1-\mu)\boldsymbol{y}[r] \leq \boldsymbol{y}'[r] \leq (1+\mu)\boldsymbol{y}[r]. \tag{17}$$

As a result, combining (16) and (17) we get that for any $r \in [\![d]\!]$,

$$\frac{\boldsymbol{y}'[r]}{\lambda'} - \frac{\boldsymbol{y}[r]}{\lambda} \leq \left(\frac{1+\mu}{1-\mu} - 1\right)\frac{\boldsymbol{y}[r]}{\lambda} \leq 4\mu\frac{\boldsymbol{y}[r]}{\lambda} = 4\mu\boldsymbol{x}[r],$$

since $\mu \leq \frac{1}{2}$. Similarly, by (16) and (17),

$$\frac{\boldsymbol{y}[r]}{\lambda} - \frac{\boldsymbol{y}'[r]}{\lambda'} \leq \left(1 - \frac{1-\mu}{1+\mu}\right)\frac{\boldsymbol{y}[r]}{\lambda} \leq 2\mu\frac{\boldsymbol{y}[r]}{\lambda} = 2\mu\boldsymbol{x}[r].$$

Thus, it follows that

$$\left|\frac{\boldsymbol{y}'[r]}{\lambda'} - \frac{\boldsymbol{y}[r]}{\lambda}\right| \leq 4\mu\boldsymbol{x}[r],$$

in turn implying that

$$\|\boldsymbol{x}' - \boldsymbol{x}\|_1 = \sum_{r=1}^{d}\left|\frac{\boldsymbol{y}'[r]}{\lambda'} - \frac{\boldsymbol{y}[r]}{\lambda}\right| \leq 4\mu\sum_{r=1}^{d}\boldsymbol{x}[r] \leq 4\mu\|\mathcal{X}\|_1. \tag{18}$$

Moreover, by definition of (15),

$$(\mu)^2 \leq \left\|\begin{pmatrix}\lambda' \\ \boldsymbol{y}'\end{pmatrix} - \begin{pmatrix}\lambda \\ \boldsymbol{y}\end{pmatrix}\right\|_t^2.$$

Finally, combining this bound with (18) concludes the proof. □

Lemma 1. *For any time $t \in \mathbb{N}$ and learning rate $\eta \leq \frac{1}{50}$, if $\|\boldsymbol{u}^{(t)}\|_\infty\|\boldsymbol{x}\|_1 \leq 1$,*

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1 \leq 4\|\mathcal{X}\|_1\left\|\begin{pmatrix}\lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)}\end{pmatrix} - \begin{pmatrix}\lambda^{(t)} \\ \boldsymbol{y}^{(t)}\end{pmatrix}\right\|_t.$$

Proof. Since $\eta \leq \frac{1}{50}$ by assumption, we have

$$\left\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\right\|_t \leq 22\eta < \frac{1}{2}.$$

Hence, we are in the domain of applicability of Lemma 6, which immediately yields the statement. □

Corollary 1 (RVU bound in the original (unlifted) space)**.** *Fix any time $T \in \mathbb{N}$, and suppose that $\|\boldsymbol{u}^{(t)}\|_\infty \leq B$ for any $t \in [\![T]\!]$. If $\eta \leq \frac{1}{256B\|\mathcal{X}\|_1}$,*

$$\tilde{\mathrm{Reg}}^T \leq 6B\|\mathcal{X}\|_1 + \frac{(d+1)\log T}{\eta} + 16\eta\|\mathcal{X}\|_1^2\sum_{t=1}^{T-1}\|\boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}\|_\infty^2 - \frac{1}{512\eta\|\mathcal{X}\|_1^2}\sum_{t=1}^{T-1}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1^2.$$

Proof. At first, assume that $\|\boldsymbol{u}^{(t)}\|_\infty \leq 1/\|\mathcal{X}\|_1$. By definition of the induced dual local norm in (3),

$$\begin{aligned}
\|\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\|_{*,t}^2 &\leq (\langle\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\rangle - \langle\boldsymbol{x}^{(t-1)}, \boldsymbol{u}^{(t-1)}\rangle)^2(\lambda^{(t)})^2 + \sum_{r=1}^{d}(\boldsymbol{y}[r])^2(\boldsymbol{u}^{(t)}[r] - \boldsymbol{u}^{(t-1)}[r])^2 \\
&\leq (\langle\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\rangle - \langle\boldsymbol{x}^{(t-1)}, \boldsymbol{u}^{(t-1)}\rangle)^2 + \sum_{r=1}^{d}(\boldsymbol{x}[r])^2(\boldsymbol{u}^{(t)}[r] - \boldsymbol{u}^{(t-1)}[r])^2 \\
&\leq \left(\langle\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\rangle - \langle\boldsymbol{x}^{(t-1)}, \boldsymbol{u}^{(t-1)}\rangle\right)^2 + \|\mathcal{X}\|_1^2\|\boldsymbol{u}^{(t)} - \boldsymbol{u}^{(t-1)}\|_\infty^2, \qquad (19)
\end{aligned}$$

for any $t \geq 2$. Further, by Young's inequality,

$$\begin{aligned}
\left(\langle\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\rangle - \langle\boldsymbol{x}^{(t-1)}, \boldsymbol{u}^{(t-1)}\rangle\right)^2 &\leq 2\left(\langle\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} - \boldsymbol{u}^{(t-1)}\rangle\right)^2 + 2\left(\langle\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}, \boldsymbol{u}^{(t-1)}\rangle\right)^2 \\
&\leq 2\|\mathcal{X}\|_1^2\|\boldsymbol{u}^{(t)} - \boldsymbol{u}^{(t-1)}\|_\infty^2 + \frac{2}{\|\mathcal{X}\|_1^2}\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_1^2.
\end{aligned}$$

Combining with (19),

$$\|\tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)}\|_{*,t}^2 \leq 3\|\mathcal{X}\|_1^2\|\boldsymbol{u}^{(t)} - \boldsymbol{u}^{(t-1)}\|_\infty^2 + \frac{2}{\|\mathcal{X}\|_1^2}\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_1^2,$$

for $t \geq 2$, since $\|\boldsymbol{u}\|_\infty \leq \frac{1}{\|\mathcal{X}\|_1}$ (by assumption). Further, $\|\tilde{\boldsymbol{u}}^{(1)} - \tilde{\boldsymbol{u}}^{(0)}\|_{*,t}^2 = \|\tilde{\boldsymbol{u}}^{(1)}\|_{*,t}^2 \leq 2$. Combining with Proposition 2 and Lemma 1, we get that $\tilde{\text{Reg}}^T$ is upper bounded by

$$6 + \frac{(d+1)\log T}{\eta} + 16\eta\|\mathcal{X}\|_1^2 \sum_{t=1}^{T-1} \|\boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}\|_\infty^2 + \frac{1}{\|\mathcal{X}\|_1^2}\left(10\eta - \frac{1}{432\eta}\right)\sum_{t=1}^{T-1}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1^2$$

$$\leq 6 + \frac{(d+1)\log T}{\eta} + 16\eta\|\mathcal{X}\|_1^2 \sum_{t=1}^{T-1} \|\boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}\|_\infty^2 - \frac{1}{512\eta\|\mathcal{X}\|_1^2}\sum_{t=1}^{T-1}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1^2.$$

Finally, we relax the assumption that $\|\boldsymbol{u}^{(t)}\|_\infty \leq 1/\|\mathcal{X}\|_1$. In that case, one can reduce to the above analysis by first rescaling all utilities by the factor $1/(B\|\mathcal{X}\|_1)$—which in turn is equivalent to rescaling the learning rate η by $1/(B\|\mathcal{X}\|_1)$. We then need to correct for the fact that the norm of the difference of utilities gets rescaled by a factor $1/(B\|\mathcal{X}\|_1)^2$, and that the regret $\tilde{\text{Reg}}^T$ with respect to the original utilities is a factor $B\|\mathcal{X}\|_1$ larger than the regret measured on the rescaled utilities. Taking these considerations into account leads to the statement. □

A.4 Main Result: Proof of Theorem 4

Finally, we are ready to establish Theorem 4. To this end, the main ingredient is the bound on the second-order path lengths predicted by Theorem 3, which is recalled below.

Theorem 3. *Suppose that Assumption 1 holds for some parameters $B, L > 0$. If all players follow* `LRL-OFTRL` *with learning rate $\eta \leq \min\left\{\frac{1}{256B\|\mathcal{X}\|_1}, \frac{1}{128nL\|\mathcal{X}\|_1^2}\right\}$, where $\|\mathcal{X}\|_1 := \max_{i\in[\![n]\!]}\|\mathcal{X}_i\|_1$, then*

$$\sum_{i=1}^{n}\sum_{t=1}^{T-1}\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2 \leq 6144n\eta B\|\mathcal{X}\|_1^3 + 1024n(d+1)\|\mathcal{X}\|_1^2\log T. \tag{4}$$

Proof. By Assumption 1, it follows that for any player $i \in [\![n]\!]$,

$$\left(\|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty\right)^2 \leq \left(L\sum_{j=1}^{n}\|\boldsymbol{x}_j^{(t+1)} - \boldsymbol{x}_j^{(t)}\|_1\right)^2 \leq L^2 n \sum_{j=1}^{n}\|\boldsymbol{x}_j^{(t+1)} - \boldsymbol{x}_j^{(t)}\|_1^2,$$

by Jensen's inequality. Hence, by Corollary 1 the regret Reg_i^T of each player $i \in [\![n]\!]$ can be upper bounded by

$$6B\|\mathcal{X}\|_1 + \frac{(d+1)\log T}{\eta} + 16\eta\|\mathcal{X}\|_1^2 L^2 n\sum_{j=1}^{n}\sum_{t=1}^{T-1}\|\boldsymbol{x}_j^{(t+1)} - \boldsymbol{x}_j^{(t)}\|_1^2 - \frac{1}{512\eta\|\mathcal{X}\|_1^2}\sum_{t=1}^{T-1}\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2,$$

Summing over all players $i \in [\![n]\!]$, we have that

$$\sum_{i=1}^{n}\tilde{\text{Reg}}_i^T \leq 6nB\|\mathcal{X}\|_1 + n\frac{(d+1)\log T}{\eta} + \sum_{i=1}^{n}\left(16\eta\|\mathcal{X}\|_1^2 L^2 n^2 - \frac{1}{512\eta\|\mathcal{X}\|_1^2}\right)\sum_{t=1}^{T-1}\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2$$

$$\leq 6nB\|\mathcal{X}\|_1 + n\frac{(d+1)\log T}{\eta} - \frac{1}{1024\eta\|\mathcal{X}\|_1^2}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2,$$

since $\eta \leq \frac{1}{256nL\|\mathcal{X}\|_1^2}$. Finally, the theorem follows since $\sum_{i=1}^{n}\tilde{\text{Reg}}_i^T \geq 0$, which in turn follows directly from Theorem 2. □

Theorem 4 (Detailed Version of Theorem 1)**.** *Suppose that Assumption 1 holds for some parameters $B, L > 0$. If all players follow* `LRL-OFTRL` *with learning rate $\eta = \min\left\{\frac{1}{256B\|\mathcal{X}\|_1}, \frac{1}{128nL\|\mathcal{X}\|_1^2}\right\}$, then for any $T \in \mathbb{N}$ the regret Reg_i^T of each player $i \in [\![n]\!]$ can be bounded as*

$$\text{Reg}_i^T \leq 12B\|\mathcal{X}\|_1 + 256(d+1)\max\left\{nL\|\mathcal{X}\|_1^2, 2B\|\mathcal{X}\|_1\right\}\log T. \tag{5}$$

Furthermore, the algorithm can be adaptive so that if player i is instead facing adversarial utilities, then $\text{Reg}_i^T = O(\sqrt{T})$.

Proof. First of all, by Assumption 1 we have that for any player $i \in [\![n]\!]$,

$$\|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 \le \left(L \sum_{j=1}^n \|\boldsymbol{x}_j^{(t+1)} - \boldsymbol{x}_j^{(t)}\|_1 \right)^2 \le L^2 n \sum_{j=1}^n \|\boldsymbol{x}_j^{(t+1)} - \boldsymbol{x}_j^{(t)}\|_1^2.$$

Hence, summing over all t,

$$\begin{aligned}\sum_{t=1}^{T-1} \|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 &\le L^2 n \sum_{t=1}^{T-1} \sum_{j=1}^n \|\boldsymbol{x}_j^{(t+1)} - \boldsymbol{x}_j^{(t)}\|_1^2 \\ &\le 6144 n^2 L^2 \eta B \|\mathcal{X}\|_1^3 + 1024 n^2 L^2 (d+1) \|\mathcal{X}\|_1^2 \log T,\end{aligned}$$

where the last bound uses Theorem 3. As a result, from Corollary 1, if $\eta = \frac{1}{128 n L \|\mathcal{X}\|_1^2}$,

$$\begin{aligned}\tilde{\text{Reg}}_i^T &\le 6B\|\mathcal{X}\|_1 + \frac{(d+1)\log T}{\eta} + 16\eta \|\mathcal{X}\|_1^2 \sum_{t=1}^{T-1} \|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 \\ &\le 12 B \|\mathcal{X}\|_1 + 256(d+1) n L \|\mathcal{X}\|_1^2 \log T.\end{aligned}$$

Thus, the bound on Reg_i^T follows directly since $\text{Reg}_i^T \le \tilde{\text{Reg}}_i^T$ by Theorem 2. The case where $\eta = \frac{1}{256 B \|\mathcal{X}\|_1}$ is analogous.

Next, let us focus on the adversarial bound. Each player can simply check whether there exists a time $t \in [\![T]\!]$ such that

$$\sum_{\tau=1}^{t-1} \|\boldsymbol{u}_i^{(\tau+1)} - \boldsymbol{u}_i^{(\tau)}\|_\infty^2 > 6144 n^2 L^2 \eta B \|\mathcal{X}\|_1^3 + 1024 n^2 L^2 (d+1) \|\mathcal{X}\|_1^2 \log t. \tag{20}$$

In particular, we know from Theorem 3 that when all players follow the prescribed protocol (20) will never by satisfied. On the other hand, if there exists time t so that (20) holds, then it suffices to switch to any no-regret learning algorithm tuned to face adversarial utilities. □

A.5 Extending the Analysis under Approximate Iterates

In this subsection we describe how to extend our analysis, and in particular Theorem 4, when the `OFTRL` step of Algorithm 1 at time t is only computed with tolerance $\epsilon^{(t)}$, in the sense of (6). We start by extending Proposition 2 below.

Proposition 5 (Extension of Proposition 2)**.** *Let $\tilde{\text{Reg}}^T$ be the regret cumulated up to time T by the internal `OFTRL` algorithm producing approximate iterates $(\lambda^{(t)}, \boldsymbol{y}^{(t)}) \in \tilde{\mathcal{X}}$, for any $t \in [\![T]\!]$. Then, for any $T \in \mathbb{N}$ and learning rate $\eta \le \frac{1}{50}$,*

$$\begin{aligned}\tilde{\text{Reg}}^T \le 4 + \frac{(d+1)\log T}{\eta} + 5\eta \sum_{t=1}^T \left\| \tilde{\boldsymbol{u}}^{(t)} - \tilde{\boldsymbol{u}}^{(t-1)} \right\|_{*,t}^2 - \frac{1}{27\eta} \sum_{t=1}^{T-1} \left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)}, \boldsymbol{y}_\star^{(t)})} \\ + 2 \sum_{t=1}^T \left\| \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)}, \boldsymbol{y}_\star^{(t)})},\end{aligned}$$

where

$$\begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} := \operatorname*{arg\,max}_{(\lambda, \boldsymbol{y}) \in \tilde{\mathcal{X}}} \left\{ \eta \left\langle \tilde{\boldsymbol{U}}^{(t)} + \tilde{\boldsymbol{u}}^{(t-1)}, \begin{pmatrix} \lambda \\ \boldsymbol{y} \end{pmatrix} \right\rangle + \log \lambda + \sum_{r=1}^d \log \boldsymbol{y}[r] \right\}.$$

Proof. Fix any $(\lambda^*, \boldsymbol{y}^*) \in \tilde{\mathcal{X}}$. Then,

$$\begin{aligned}\sum_{t=1}^T \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} \lambda^* \\ \boldsymbol{y}^* \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\rangle &= \sum_{t=1}^T \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} \lambda^* \\ \boldsymbol{y}^* \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\rangle + \sum_{t=1}^T \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\rangle \\ &\le \sum_{t=1}^T \left\langle \tilde{\boldsymbol{u}}^{(t)}, \begin{pmatrix} \lambda^* \\ \boldsymbol{y}^* \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\rangle + 2 \sum_{t=1}^T \left\| \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)}, \boldsymbol{y}_\star^{(t)})},\end{aligned}$$

where the last inequality uses Hölder's inequality along with the fact that $\|\tilde{\boldsymbol{u}}^{(t)}\|_{*,(\lambda^{(t)},\boldsymbol{y}^{(t)})} \leq 2$, which in turn follows since $\|\boldsymbol{u}^{(t)}\|_\infty \|\mathcal{X}\|_1 \leq 1$ (by assumption). Finally, the proof follows as an immediate consequence of Proposition 2. □

We next proceed with the extension of Lemma 1.

Lemma 7 (Extension of Lemma 1)**.** *Suppose that $\epsilon^{(t)} \leq \frac{1}{8}$, for any $t \in [\![T]\!]$. Then, for any time $t \in [\![T-1]\!]$ and learning rate $\eta \leq \frac{1}{256}$,*

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1 \leq 8\|\mathcal{X}\|_1 \left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} + 16\|\mathcal{X}\|_1 \epsilon^{(t+1)} + 8\|\mathcal{X}\|_1 \epsilon^{(t)},$$

where $\boldsymbol{x}^{(t)} := \boldsymbol{y}^{(t)}/\lambda^{(t)}$.

Proof. First, by the triangle inequality,

$$\left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} \geq \left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} - \left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} - \left\| \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})}.$$

Now given that $\eta \leq \frac{1}{50}$, it follows from Proposition 3 that

$$\left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} \leq \frac{1}{2},$$

which in turn—combined with Lemma 4—implies that

$$\left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} \leq 2 \left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} \right\|_{(\lambda_\star^{(t+1)},\boldsymbol{y}_\star^{(t+1)})}.$$

Similarly, since $\epsilon^{(t)} \leq \frac{1}{8}$, it follows that

$$\left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} \geq \frac{1}{2} \left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_{(\lambda^{(t)},\boldsymbol{y}^{(t)})}.$$

As a result,

$$\left\| \begin{pmatrix} \lambda_\star^{(t+1)} \\ \boldsymbol{y}_\star^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_\star^{(t)} \\ \boldsymbol{y}_\star^{(t)} \end{pmatrix} \right\|_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} \geq \frac{1}{2} \left\| \begin{pmatrix} \lambda^{(t+1)} \\ \boldsymbol{y}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda^{(t)} \\ \boldsymbol{y}^{(t)} \end{pmatrix} \right\|_{(\lambda^{(t)},\boldsymbol{y}^{(t)})} - 2\epsilon^{(t+1)} - \epsilon^{(t)}. \tag{21}$$

Next, we will prove that

$$\max \left\{ \left| \frac{\lambda^{(t+1)}}{\lambda^{(t)}} - 1 \right|, \max_{r \in [\![d]\!]} \left| \frac{\boldsymbol{y}^{(t+1)}[r]}{\boldsymbol{y}^{(t)}[r]} - 1 \right| \right\} \leq \frac{1}{2}. \tag{22}$$

Indeed, since $\epsilon^{(t)}, \epsilon^{(t+1)} \leq \frac{1}{8}$, it holds that

$$\left| 1 - \frac{\lambda^{(t)}}{\lambda_\star^{(t)}} \right| \leq \frac{1}{8} \implies \frac{7}{8}\lambda_\star^{(t)} \leq \lambda^{(t)} \leq \frac{9}{8}\lambda_\star^{(t)},$$

and

$$\left| 1 - \frac{\lambda^{(t+1)}}{\lambda_\star^{(t+1)}} \right| \leq \frac{1}{8} \implies \frac{7}{8}\lambda_\star^{(t+1)} \leq \lambda^{(t+1)} \leq \frac{9}{8}\lambda_\star^{(t+1)}.$$

Furthermore, for $\eta \leq \frac{1}{256}$,

$$\left| 1 - \frac{\lambda_\star^{(t+1)}}{\lambda_\star^{(t)}} \right| \leq \frac{1}{10} \implies \frac{9}{10}\lambda_\star^{(t)} \leq \lambda_\star^{(t+1)} \leq \frac{11}{10}\lambda_\star^{(t)},$$

by Proposition 3 and Lemma 4. Thus,

$$\frac{2}{3}\lambda^{(t+1)} \leq \frac{7}{8}\frac{10}{11}\frac{8}{9}\lambda^{(t+1)} \leq \lambda^{(t)} \leq \frac{9}{8}\frac{10}{9}\frac{8}{7}\lambda^{(t+1)} \leq \frac{3}{2}\lambda^{(t+1)},$$

in turn implying that

$$\left|1 - \frac{\lambda^{(t+1)}}{\lambda^{(t)}}\right| \leq \frac{1}{2}.$$

Similarly, we conclude that for any $r \in [\![d]\!]$,

$$\left|1 - \frac{\boldsymbol{y}^{(t+1)}[r]}{\boldsymbol{y}^{(t)}[r]}\right| \leq \frac{1}{2},$$

confirming (22). Hence, following the proof of Lemma 6, we derive that

$$\left\|\begin{pmatrix}\lambda^{(t+1)}\\ \boldsymbol{y}^{(t+1)}\end{pmatrix} - \begin{pmatrix}\lambda^{(t)}\\ \boldsymbol{y}^{(t)}\end{pmatrix}\right\|_{(\lambda^{(t)},\boldsymbol{y}^{(t)})} \geq \frac{1}{4\|\mathcal{X}\|_1}\left\|\frac{\boldsymbol{y}^{(t+1)}}{\lambda^{(t+1)}} - \frac{\boldsymbol{y}^{(t)}}{\lambda^{(t)}}\right\|_1 = \frac{1}{4\|\mathcal{X}\|_1}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1.$$

Combining this bound with (21) concludes the proof. □

We also state the following immediate implication of Lemma 7.

Corollary 3. *Suppose that $\epsilon^{(t)} \leq \frac{1}{8}$, for any $t \in [\![T]\!]$. Then, for any $t \in [\![T-1]\!]$ and learning rate $\eta \leq \frac{1}{256}$,*

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_1^2 \leq 192\|\mathcal{X}\|_1^2 \left\|\begin{pmatrix}\lambda_\star^{(t+1)}\\ \boldsymbol{y}_\star^{(t+1)}\end{pmatrix} - \begin{pmatrix}\lambda_\star^{(t)}\\ \boldsymbol{y}_\star^{(t)}\end{pmatrix}\right\|^2_{(\lambda_\star^{(t)},\boldsymbol{y}_\star^{(t)})} + 768\|\mathcal{X}\|_1^2(\epsilon^{(t+1)})^2 + 192\|\mathcal{X}\|_1^2(\epsilon^{(t)})^2,$$

where $\boldsymbol{x}^{(t)} := \boldsymbol{y}^{(t)}/\lambda^{(t)}$.

As a result, combining this bound with Proposition 5 extends Corollary 1 with an error term proportional to $\sum_{t=1}^{T}\epsilon^{(t)}$. Finally, the rest of the extension is identical to our proof of Theorem 4.

B Implementation via Proximal Oracles

In this section we provide the omitted proofs from Section 3.5 regarding the implementation of `LRL-OFTRL` using proximal oracles (recall Equation (7)).

B.1 The Proximal Newton Method

In this subsection we describe the proximal Newton algorithm of Tran-Dinh et al. [2015], leading to Theorem 5 we presented in Section 3.5. More precisely, Tran-Dinh et al. [2015] studied the following composite minimization problem:

$$\min_{\tilde{\boldsymbol{x}} \in \mathbb{R}^{d+1}} \left\{F(\tilde{\boldsymbol{x}}) := f(\tilde{\boldsymbol{x}}) + g(\tilde{\boldsymbol{x}})\right\}, \tag{23}$$

where f is a (standard) self-concordant and convex function, and $g : \mathbb{R}^{d+1} \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed and convex function. In our setting, we will let g be defined as

$$g(\tilde{\boldsymbol{x}}) := \begin{cases} 0 & \text{if } \tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}, \\ +\infty & \text{otherwise.} \end{cases}$$

Further, for a given time $t \in \mathbb{N}$, we let

$$f : \tilde{\boldsymbol{x}} \mapsto -\eta\left\langle \tilde{\boldsymbol{U}}^{(t)} + \tilde{\boldsymbol{u}}^{(t-1)}, \tilde{\boldsymbol{x}}\right\rangle - \sum_{r=1}^{d+1}\log\tilde{\boldsymbol{x}}[r].$$

Before we describe the proximal Newton method, let us define $\tilde{\boldsymbol{s}}_k$ as follows.

$$\tilde{\boldsymbol{s}}_k := \arg\min_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}} \left\{ f(\tilde{\boldsymbol{x}}_k) + (\nabla f(\tilde{\boldsymbol{x}}_k))^\top (\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_k) + \frac{1}{2} (\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_k)^\top \nabla^2 f(\tilde{\boldsymbol{x}}_k) (\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_k) \right\}, \tag{24}$$

for some $\tilde{\boldsymbol{x}}_k \in \mathbb{R}^{d+1}_{>0}$. We point out that the optimization problem (24) can be trivially solved when we have access to a (local) proximal oracle—given in Equation (7).

In this context, the proximal Newton method of Tran-Dinh et al. [2015] is given in Algorithm 2. Their algorithm proceeds in two phases. In the first phase we perform *damped steps* of proximal Newton until we reach the region of quadratic convergence. Afterwards, we perform *full steps* of proximal Newton until the desired precision $\epsilon > 0$ has been reached. Below we summarize the main guarantee regarding Algorithm 2, namely [Tran-Dinh et al., 2015, Theorem 9].

Theorem 7 ([Tran-Dinh et al., 2015]). *Algorithm 2 returns $\tilde{\boldsymbol{x}}_K \in \mathbb{R}^{d+1}_{>0}$ such that $\|\tilde{\boldsymbol{x}}_K - \tilde{\boldsymbol{x}}^*\|_{\tilde{\boldsymbol{x}}^*} \leq 2\epsilon$ after at most*

$$K = \left\lfloor \frac{f(\tilde{\boldsymbol{x}}_0) - f(\tilde{\boldsymbol{x}}^*)}{0.017} \right\rfloor + \left\lfloor 1.5 \ln \ln \left(\frac{0.28}{\epsilon} \right) \right\rfloor + 2$$

iterations, for any $\epsilon > 0$, where $\tilde{\boldsymbol{x}}^* = \arg\min_{\tilde{\boldsymbol{x}}} F(\tilde{\boldsymbol{x}})$, for the composite function F defined in (23).

To establish Theorem 5 from this guarantee, it suffices to initialize Algorithm 2 at every iteration $t \geq 2$ with $\tilde{\boldsymbol{x}}_0 := \tilde{\boldsymbol{x}}^{(t-1)} = (\lambda^{(t-1)}, \boldsymbol{y}^{(t-1)})$. Then, as long as $\epsilon^{(t-1)}$ is sufficiently small, the number of iterations predicted by Theorem 7 will be bounded by $O(\log \log(1/\epsilon))$, in turn establishing Theorem 5.

Algorithm 2: Proximal Newton [Tran-Dinh et al., 2015]

Data: Initial point $\tilde{\boldsymbol{x}}_0$
Precision $\epsilon > 0$
Constant $\sigma := 0.2$

```
for k = 1, ..., K do
    Obtain the proximal Newton direction d̃_k ← s̃_k − x̃_k, where s̃_k is defined in (24)
    Set λ_k ← ‖d̃_k‖_{x̃_k}
    if λ_k > 0.2 then
        x̃_{k+1} ← x̃_k + α_k d̃_k, where α_k := (1 + λ_k)^{-1}        [▷ Damped Step]
    else if λ_k > ε then
        x̃_{k+1} ← x̃_k + d̃_k                                        [▷ Full Step]
    else
        return x̃_k
```

B.2 Proximal Oracle for Normal-Form and Extensive-Form Games

In order to show that the proximal oracle of Section 3.5 can be implemented efficiently for probability simplexes (*i.e.*, the strategy sets of normal-form games) and sequence-form strategy spaces (*i.e.*, the strategy sets of extensive-form games), we will prove a slightly stronger result concerning *treeplex* sets, of which sequence-form strategy spaces are instances.

Definition 1. *A set $Q \subseteq [0, +\infty)^d$, $d \geq 1$, is* treeplex *if it is:*

1. *a simplex $Q = \Delta^d$;*
2. *a Cartesian product of treeplex sets $Q_1 \times \cdots \times Q_K$; or*
3. *(Branching operation) a set of the form*

$$\triangle(Q_1, \dots, Q_K) := \{(\boldsymbol{x}, \boldsymbol{x}[1]\boldsymbol{q}_1, \dots, \boldsymbol{x}[K]\boldsymbol{q}_K) : \boldsymbol{x} \in \Delta^K, \boldsymbol{q}_k \in Q_k \ \ \forall k \in [\![K]\!]\},$$

where $Q_1, \dots, Q_K$ are treeplex.

We will show that any treeplex Q is such that $[0,1]Q$ admits an efficient (positive-definite) quadratic optimization oracle. This is sufficient, since it is well-known that every sequence-form strategy space $\mathcal{X}$ is treeplex (e.g., Hoda et al. [2010]) and therefore, by definition, so is the set $\{(1, \boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$.

Introduce the *value function*

$$V_Q(t; \boldsymbol{g}, \boldsymbol{w}) \coloneqq \min_{\boldsymbol{x} \in tQ} \left\{ -\boldsymbol{g}^\top \boldsymbol{x} + \frac{1}{2} \sum_{r=1}^{d} \left(\frac{\boldsymbol{x}[r]}{\boldsymbol{w}[r]} \right)^2 \right\} \qquad (t \geq 0, \boldsymbol{w} > \boldsymbol{0}) \tag{25}$$

(note the rescaling by t in the domain of the minimization). We will be interested in the derivative of $V_Q(t; \boldsymbol{g}, \boldsymbol{w})$, which we will denote as[6]

$$\lambda_Q(t; \boldsymbol{g}, \boldsymbol{w}) \coloneqq \frac{d}{dt} V_Q(t; \boldsymbol{g}, \boldsymbol{w}).$$

Preliminaries on Strictly Monotonic Piecewise-Linear (SMPL) Functions

Definition 2 (SMPL function and standard representation)**.** *Given an interval $I \subseteq \mathbb{R}$ and a function $f : I \to \mathbb{R}$, we say that f is SMPL if it is strictly monotonically increasing and piecewise-linear on I.*

Definition 3 (Quasi-SMPL function)**.** *A quasi-SMPL function is a function $f : \mathbb{R} \to [0, +\infty)$ of the form $f(x) = [g(x)]^+$ where $g(x) : \mathbb{R} \to \mathbb{R}$ is SMPL and $[\,\cdot\,]^+ \coloneqq \max\{0, \cdot\}$.*

Definition 4. *Given a SMPL or quasi-SMPL function f, a* standard representation *for it is an expression of the form*

$$f(x) = \zeta + \alpha_0 x + \sum_{s=1}^{S} \alpha_s [x - \beta_s]^+,$$

valid for all x in the domain of f, where $S \in \mathbb{N} \cup \{0\}$ and $\beta_1 < \cdots < \beta_S$. The size of the standard representation is defined as the natural number S.

We now mention four basic results about SMPL and quasi-SMPL functions. The proofs are elementary and omitted.

Lemma 8. *Let $f : I \to \mathbb{R}$ be SMPL, and consider a standard representation of f of size S. Then, for any $\zeta \in \mathbb{R}$ and $\alpha \geq 0$, a standard representation for the SMPL function $I \ni x \mapsto \zeta + \alpha f(x)$ can be computed in $O(S + 1)$ time.*

Lemma 9. *The sum $f_1 + \cdots + f_n$ of n SMPL (resp., quasi-SMPL) functions $f_i : I \to \mathbb{R}$ is a SMPL (resp., quasi-SMPL) function $I \to \mathbb{R}$. Furthermore, if each f_i admits a standard representation of size S_i, then a standard representation of size at most $S_1 + \cdots + S_n$ for their sum can be computed in $O((S_1 + \cdots + S_n + 1) \log n)$ time.*

Lemma 10. *Let $f : \mathbb{R} \to \mathbb{R}$ be SMPL, and consider a standard representation of f of size S. Then, for any $\beta \in \mathbb{R}$, a standard representation of size at most S for the quasi-SMPL function $I \ni x \mapsto [f(x) - \beta]^+$ can be computed in $O(S + 1)$ time.*

Lemma 11. *The inverse $f^{-1} : \mathrm{range}(f) \to \mathbb{R}$ of a SMPL function $f : I \to \mathbb{R}$ is SMPL. Furthermore, if f admits a standard representation of size S, then a standard representation for f^{-1} of size at most S can be computed in $O(S + 1)$ time.*

Lemma 12. *Let $f : \mathbb{R} \to [0, +\infty)$ be quasi-SMPL. The restricted inverse $f^{-1} : (0, +\infty) \to \mathbb{R}$ of f is SMPL, where we restrict the domain to $(0, +\infty)$ because $f^{-1}(0)$ may be multivalued. Furthermore, if f admits a standard representation of size S, then a standard representation of size at most S for f^{-1} can be computed in $O(S + 1)$ time.*

Proof. We have $f(x) = [g(x)]^+$ where g is SMPL. It follows that the function $\bar{g} : I \to \mathbb{R}$ defined as $\bar{g}(x) = g(x)$ for the interval $I = \{x : g(x) > 0\}$ is SMPL as well. For any x such that $f(x) > 0$ we have $x \in I$, and thus $f^{-1} = g^{-1}$, and it follows from Lemma 11 that f^{-1} is SMPL. □

[6]For $t = 0$ we define $\lambda_Q(t; \boldsymbol{g}, \boldsymbol{w})$ in the usual way as

$$\lambda_Q(0; \boldsymbol{g}, \boldsymbol{w}) = \lim_{t \to 0^+} \frac{V_Q(t; \boldsymbol{g}, \boldsymbol{w}) - V_Q(0; \boldsymbol{g}, \boldsymbol{w})}{t} = \lim_{t \to 0^+} \frac{V_Q(t; \boldsymbol{g}, \boldsymbol{w})}{t}.$$

Lemma 13. *Let $f : [0, +\infty) \to \mathbb{R}$ be a SMPL function, and consider the function g that maps y to the unique solution to the equation $x = [y - f(x)]^+$. Then, g is quasi-SMPL and satisfies $g(y) = [(x+f)^{-1}(y)]^+$, where $(x+f)^{-1}$ denotes the inverse of the SMPL function $x \mapsto x + f([x]^+)$.*

Proof. For any $y \in \mathbb{R}$, the function $h_y : x \mapsto x - [y - f(x)]^+$ is clearly SMPL on $[0, +\infty)$. Furthermore, $h_y(0) \leq 0$ and $h_y(+\infty) = +\infty$, implying that $h_y(x) = 0$ has a unique solution. We now show that $g(y) = [(x + f)^{-1}(y)]^+$ is that solution, that is, it satisfies $g(y) = [y - f(g(y))]^+$ for all $y \in \mathbb{R}$. Fix any $y \in \mathbb{R}$ and let

$$\bar{g} := (x + f)^{-1}(y) \quad \iff \quad \bar{g} + f([\bar{g}]^+) = y \quad \iff \quad \bar{g} = y - f([\bar{g}]^+) \tag{26}$$

There are two cases:

- If $\bar{g} \geq 0$, then $g(y) = [\bar{g}]^+ = \bar{g}$, and so we have

  $$g(y) = [\bar{g}]^+ = [y - f([\bar{g}]^+)]^+ = [y - f(g(y))]^+,$$

  as we wanted to show.

- Otherwise, $\bar{g} < 0$ and $g(y) = 0$. From (26), the condition $\bar{g} < 0$ implies $y < f([\bar{g}]^+) = f(0)$. So, it is indeed the case that

  $$0 = g(y) = [y - f(0)]^+ = [y - f(g(0))]^+,$$

  as we wanted to show.

Finally, we note that the function $(x + f)^{-1} : \mathbb{R} \to \mathbb{R}$ is SMPL due to Lemma 11, implying that $g(y)$ is quasi-SMPL. ☐

Central result The following result is central in our analysis.

Lemma 14. *For any treeplex $Q \subseteq \mathbb{R}^d$, gradient $\boldsymbol{g} \in \mathbb{R}^d$, and center $\boldsymbol{w} \in \mathbb{R}^d_{>0}$, the function $t \mapsto \lambda_Q(t; \boldsymbol{g}, \boldsymbol{w})$ is SMPL, and a standard representation of it of size d can be computed in polynomial time in d.*

Proof. We will prove the result by structural induction on Q.

- First, we consider the case where Q is a Cartesian product,

  $$Q = Q_1 \times \cdots \times Q_K.$$

  In that case, the value function decomposes as follows

  $$V_Q(t; \boldsymbol{g}, \boldsymbol{w}) = \sum_{k=1}^{K} \min_{\boldsymbol{x}_k \in tQ_k} \left\{ -\boldsymbol{g}_k^\top \boldsymbol{x}_k + \frac{1}{2} \sum_{r=1}^{d_k} \left(\frac{\boldsymbol{x}_k[r]}{\boldsymbol{w}_k[r]} \right)^2 \right\} = \sum_{k=1}^{K} V_{Q_k}(t; \boldsymbol{g}_k, \boldsymbol{w}_k).$$

  By linearity of derivatives, we have

  $$\lambda_Q(t; \boldsymbol{g}, \boldsymbol{w}) = \sum_{k=1}^{K} \lambda_{Q_k}(t; \boldsymbol{g}_k, \boldsymbol{w}_k).$$

  From Lemma 9, we conclude that $\lambda_Q(t; \boldsymbol{g}, \boldsymbol{w})$ is a SMPL function with domain $[0, +\infty)$ which admits a standard representation of size at most $d = d_1 + \cdots + d_K$ computable in time $O(d \log K)$ starting from the standard representation of each of the $\lambda_{Q_k}(t; \boldsymbol{g}_k, \boldsymbol{w}_k)$.

- Second, consider the case where Q is a simplex or the result of a branching operation

  $$\triangle(Q_1, \ldots, Q_K) = \{(\boldsymbol{x}, \boldsymbol{x}[1]\boldsymbol{q}_1, \ldots, \boldsymbol{x}[K]\boldsymbol{q}_K) : \boldsymbol{x} \in \Delta^K, \boldsymbol{q}_k \in Q_k \ \ \forall k \in [\![K]\!]\},$$

  where $Q_k \in \mathbb{R}^{d_k}$. With a slighty abuse of notation, we will treat the two cases together, considering the K-simplex Δ^K as a branching operation over empty sets $Q_k = \emptyset$.

In this case, we can write

$$\boldsymbol{g} = (\boldsymbol{g}_\bullet[1], \dots, \boldsymbol{g}_\bullet[K], \boldsymbol{g}_1 \in \mathbb{R}^{d_1}, \cdots, \boldsymbol{g}_K \in \mathbb{R}^{d_K}), \text{ and}$$
$$\boldsymbol{w} = (\boldsymbol{w}_\bullet[1], \dots, \boldsymbol{w}_\bullet[K], \boldsymbol{w}_1 \in \mathbb{R}^{d_1}_{>0}, \cdots, \boldsymbol{w}_K \in \mathbb{R}^{d_K}_{>0}).$$

The value function then decomposes recursively as

$$\begin{aligned} V_Q(t; \boldsymbol{g}, \boldsymbol{w}) = \min_{\boldsymbol{x}_\bullet \in t\Delta^K} \Bigg\{ \Bigg(-\sum_{r=1}^{K} \boldsymbol{g}_\bullet[r]\boldsymbol{x}_\bullet[r] + \frac{1}{2}\sum_{r=1}^{K}\left(\frac{\boldsymbol{x}_\bullet[r]}{\boldsymbol{w}_\bullet[r]}\right)^2 \Bigg) \\ + \sum_{k=1}^{K} \min_{\boldsymbol{x}_k \in \boldsymbol{x}_\bullet[k] Q_k} \left\{ -\boldsymbol{g}_k^\top \boldsymbol{x}_k + \sum_{r=1}^{d_k}\left(\frac{\boldsymbol{x}_k[r]}{\boldsymbol{w}_k[r]}\right)^2 \right\} \Bigg\} \\ = \min_{\boldsymbol{x}_\bullet \in t\Delta^K} \Bigg\{ \Bigg(-\sum_{r=1}^{K} \boldsymbol{g}_\bullet[r]\boldsymbol{x}_\bullet[r] + \frac{1}{2}\sum_{r=1}^{K}\left(\frac{\boldsymbol{x}_\bullet[r]}{\boldsymbol{w}_\bullet[r]}\right)^2 \Bigg) + V_{Q_k}(\boldsymbol{x}_\bullet[k]; \boldsymbol{g}_k, \boldsymbol{w}_k) \Bigg\}. \end{aligned} \tag{27}$$

Suppose that for each $k \in [\![K]\!]$, $\lambda_{Q_k}(t; \boldsymbol{g}_k, \boldsymbol{w}_k)$ is piecewise linear and monotonically increasing in t. Now we consider the KKT conditions for $\boldsymbol{x}_\bullet$ in Equation (27):

$$\begin{aligned} -\boldsymbol{g}_\bullet[k] + \frac{\boldsymbol{x}_\bullet[k]}{\boldsymbol{w}_\bullet[k]^2} + \lambda_{Q_k}(\boldsymbol{x}_\bullet[k]; \boldsymbol{g}_k, \boldsymbol{w}_k) = \lambda_\bullet + \boldsymbol{\mu}[k] \qquad \forall k \in [\![K]\!] \quad &\text{(Stationarity)} \\ \boldsymbol{x}_\bullet \in t \cdot \Delta^K \quad &\text{(Primal feasibility)} \\ \lambda_\bullet \in \mathbb{R}, \boldsymbol{\mu} \in \mathbb{R}^d_{\geq 0} \quad &\text{(Dual feasibility)} \\ \boldsymbol{\mu}[k]\boldsymbol{x}_\bullet[k] = 0 \qquad \forall k \in [\![K]\!] \quad &\text{(Compl. slackness)} \end{aligned}$$

Solving for $\boldsymbol{x}_\bullet[k]$ in the stationarity condition, and using the conditions $\boldsymbol{x}_\bullet[k]\boldsymbol{\mu}[k] = 0$ and $\boldsymbol{\mu}[k] \geq 0$, it follows that for all $k \in [\![K]\!]$

$$\begin{aligned} \boldsymbol{x}_\bullet[k] &= \boldsymbol{w}_\bullet[k]^2 \Big(\lambda_\bullet + \boldsymbol{\mu}[k] + \boldsymbol{g}_\bullet[k] - \lambda_{Q_k}(\boldsymbol{x}_\bullet[k]; \boldsymbol{g}_k, \boldsymbol{w}_k) \Big) \\ &= \boldsymbol{w}_\bullet[k]^2 \Big[\lambda_\bullet + \boldsymbol{g}_\bullet[k] - \lambda_{Q_k}(\boldsymbol{x}_\bullet[k]; \boldsymbol{g}_k, \boldsymbol{w}_k) \Big]^+. \end{aligned} \tag{28}$$

Strict monotonicity and piecewise-linearity of $\boldsymbol{x}_\bullet[k]$ as a function of $\lambda_\bullet$. Given the preliminaries on SMPL functions, it is now immediate to see that $\boldsymbol{x}_\bullet[k]$ is unique as a function of $\lambda_\bullet$. Indeed, note that (28) can be rewritten as

$$\boldsymbol{x}_\bullet[k] = \Big[(\boldsymbol{w}_\bullet[k]^2)\lambda_\bullet - \boldsymbol{w}_\bullet[k]^2(-\boldsymbol{g}_\bullet[k] + \lambda_{Q_k}(\boldsymbol{x}_\bullet[k]; \boldsymbol{g}_k, \boldsymbol{w}_k)) \Big]^+,$$

which is a fixed-point problem of the form studied in Lemma 13 for $y = (\boldsymbol{w}_\bullet[k]^2)\lambda_\bullet$ and function f_k defined as

$$f_k(\boldsymbol{x}_\bullet[k]) = \boldsymbol{w}_\bullet[k]^2(-\boldsymbol{g}_\bullet[k] + \lambda_{Q_k}(\boldsymbol{x}_\bullet[k]; \boldsymbol{g}_k, \boldsymbol{w}_k)),$$

which is clearly SMPL by inductive hypothesis. Hence, the unique solution to the previous fixed-point equation is given by the quasi-SMPL function

$$g_k : \lambda_\bullet \mapsto \frac{1}{\boldsymbol{w}_\bullet[k]^2} \left[(\boldsymbol{x}_\bullet[k] + f_k)^{-1}(\lambda_\bullet) \right]^+,$$

a standard representation of which can be computed in time $O(d+1)$ by combining the results of Lemmas 8, 10 and 11 given that a standard representation of $\lambda_{Q_k}(t; \boldsymbol{g}_k, \boldsymbol{w}_k)$ of size d is available by inductive hypothesis.

Strict monotonicity and piecewise-linearity of $\lambda_\bullet$ as a function of t. At this stage, we know that given any value of the dual variable $\lambda_\bullet$, the unique value of the coordinate $\boldsymbol{x}_\bullet[k]$ that solves the KKT system can be computed using the quasi-SMPL function g_k. In turn, this means that we can remove the primal variables $\boldsymbol{x}_\bullet$ from the KKT system, leaving us a system in $\lambda_\bullet$ and t only. We now show that the solution $\lambda_\bullet^\star$ of that system is a SMPL function of $t \in [0, +\infty)$.

Indeed, the value of $\lambda_\bullet^\star$ that solves the KKT system has to satisfy the primal feasibility condition

$$t = \sum_{k=1}^{K} \boldsymbol{x}_\bullet[k] = \sum_{k=1}^{K} g_k(\lambda_\bullet).$$

Fix any $t > 0$. The right-hand side of the equation is a sum of quasi-SMPL functions. Hence, from Lemma 9, we have that the right-hand side has a standard representation of size at most $K + \sum_{k=1}^{K} d_k = d$ can be computed in time $O(d \log K)$. Furthermore, from Lemma 12, we have that the $\lambda_\bullet^\star$ that satisfies the equation is unique, and in fact that the mapping $(0, +\infty) \ni t \mapsto \lambda_\bullet^\star$ is SMPL with standard representation of size at most d.

Relating $\lambda_\bullet$ and $\lambda_Q(t; \boldsymbol{g}, \boldsymbol{w})$. Since $\lambda_\bullet^\star(t)$ is the coefficient on t in the Lagrangian relaxation of (27), it is a subgradient of $V_Q(t; \boldsymbol{g}, \boldsymbol{w})$, and since there is a unique solution, we get that it is the derivative, that is,

$$\lambda_\bullet^\star(t) = \lambda_Q(t; \boldsymbol{g}, \boldsymbol{w})$$

for all $t \in (0, +\infty)$. To conclude the proof by induction, we then need to analyze the case $t = 0$, which has so far been excluded. When $t = 0$, the feasible set tQ is a singleton, and $V_Q(0; \boldsymbol{g}, \boldsymbol{w}) = 0$. Since $V_Q(t; \boldsymbol{g}, \boldsymbol{w})$ is continuous on $[0, +\infty)$, and since $\lim_{t \to 0^+} \lambda_Q(t; \boldsymbol{g}, \boldsymbol{w}) = \lim_{t \to 0^+} \lambda_\bullet^\star(t)$ exists since $\lambda_\bullet^\star(t)$ is piecewise-linear, then by the mean value theorem,

$$\lambda_Q(0; \boldsymbol{g}, \boldsymbol{w}) = \lim_{t \to 0^+} \lambda_\bullet^\star(t),$$

that is, the continuous extension of $\lambda_\bullet^\star$ must be (right) derivative of $V_Q(t; \boldsymbol{g}, \boldsymbol{w})$ in 0. As extending continuously $\lambda_\bullet^\star(t)$ clearly does not alter its being SMPL nor its standard representation, we conclude the proof of the inductive case.

□

Lemma 14 also provides a constructive way of computing the argmin of (25) in polynomial time for any $t \in [0, +\infty)$. To conclude the construction of the proximal oracle, it is then enough to show how to pick the optimal value of $t \in [0, 1]$ that minimizes

$$\min_{\boldsymbol{x} \in [0,1]Q} \left\{ -\boldsymbol{g}^\top \boldsymbol{x} + \frac{1}{2} \sum_{r=1}^{d} \left(\frac{\boldsymbol{x}[r]}{\boldsymbol{w}[r]} \right)^2 \right\} = \min_{t \in [0,1]} V_Q(t; \boldsymbol{g}, \boldsymbol{w}).$$

That is easy starting from the derivative $\lambda_Q(t; \boldsymbol{g}, \boldsymbol{w})$, which is a SMPL function by Lemma 14. Indeed, if $\lambda_Q(0; \boldsymbol{g}, \boldsymbol{w}) \geq 0$, then by monotonicity of the derivative we know that the optimal value of t is $t = 0$. Else, if $\lambda_Q(1; \boldsymbol{g}, \boldsymbol{w}) \leq 0$, again by monotonicity we know that the optimal value of t is $t = 1$. Else, there exists a unique value of $t \in (0, 1)$ at which the derivative of the objective is 0, and such a value can be computed exactly using Lemma 11.

C Experimental Results

In this section we provide preliminary experimental results in order to verify our theoretical findings, and in particular the per-player regret bound established in Theorem 4. More specifically, we investigate the behavior of our learning dynamics (`LRL-OFTRL`) in four standard extensive-form games used in the literature: 2-player and 3-player *Kuhn poker* [Kuhn, 1950]; 2-player *Goofspiel* [Ross, 1971];[7] and the baseline version of (2-player) *Sheriff* [Farina et al., 2019]. From those games, only 2-player Kuhn poker is a zero-sum game. Our findings are summarized in Figure 1.

[7] We consider instances of Goofspiel with $r = 3$ cards and *limited information*—the actions of the other player are only observed at the end of the game. Also, we note that the tie-breaking mechanism makes the game general-sum.

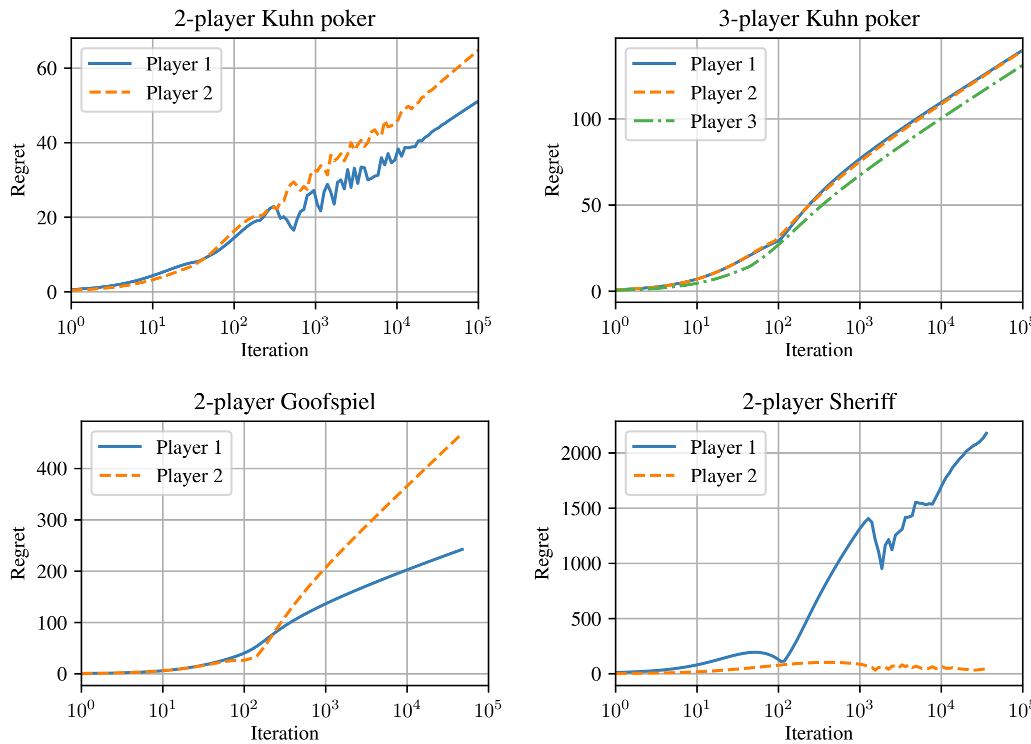


Figure 1: The regret of the players when they follow our learning dynamics, `LRL-OFTRL`; after a very mild tuning process, we selected the same learning rate $\eta := 0.5$ for all games. The x-axis indexes the iteration, while the y-axis the regret. The scale on the x-axis is *logarithmic*. We observe that the regret of each player grows as $O(\log T)$, verifying Theorem 4.