Near-Optimal No-Regret Learning for Correlated Equilibria in Multi-player General-Sum Games

Ioannis Anagnostides
ianagnos@cs.cmu.edu
Carnegie Mellon University
USA, Pittsburgh

Constantinos Daskalakis
costis@csail.mit.edu
Massachusetts Institute of Technology
USA, Boston

Gabriele Farina
gfarina@cs.cmu.edu
Carnegie Mellon University
USA, Pittsburgh

Maxwell Fishelson
maxfish@csail.mit.edu
Massachusetts Institute of Technology
USA, Boston

Noah Golowich
nzg@csail.mit.edu
Massachusetts Institute of Technology
USA, Boston

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

ABSTRACT

Recently, Daskalakis, Fishelson, and Golowich (DFG) (NeurIPS '21) showed that if all agents in a multi-player general-sum normal-form game employ Optimistic Multiplicative Weights Update (OMWU), the external regret of every player is $O(\text{polylog}(T))$ after T repetitions of the game. In this paper we extend their result from external regret to internal and swap regret, thereby establishing uncoupled learning dynamics that converge to an approximate correlated equilibrium at the rate of $\widetilde{O}\left(T^{-1}\right)$. This substantially improves over the prior best rate of convergence of $O(T^{-3/4})$ due to Chen and Peng (NeurIPS '20), and it is optimal up to polylogarithmic factors.

To obtain these results, we develop new techniques for establishing higher-order smoothness for learning dynamics involving fixed point operations. Specifically, we first establish that the no-internal-regret learning dynamics of Stoltz and Lugosi (Mach Learn '05) are equivalently simulated by no-external-regret dynamics on a combinatorial space. This allows us to trade the computation of the stationary distribution on a polynomial-sized Markov chain for a (much more well-behaved) linear transformation on an exponential-sized set, enabling us to leverage similar techniques as DGF to near-optimally bound the internal regret.

Moreover, we establish an $O(\text{polylog}(T))$ no-swap-regret bound for the classic algorithm of Blum and Mansour (BM) (JMLR '07). We do so by introducing a technique based on the Cauchy Integral Formula that circumvents the more limited combinatorial arguments of DFG. In addition to shedding clarity on the near-optimal regret guarantees of BM, our arguments provide insights into the various ways in which the techniques by DFG can be extended and leveraged in the analysis of more involved learning algorithms.

STOC '22, June 20–24, 2022, Rome, Italy

ACM ISBN 978-1-4503-9264-8/22/06…$15.00
https://doi.org/10.1145/3519935.3520031

CCS CONCEPTS

• **Theory of computation → Convergence and learning in games**.

KEYWORDS

Convergence of learning dynamics, internal regret, swap regret, correlated equilibria

ACM Reference Format:
Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. 2022. Near-Optimal No-Regret Learning for Correlated Equilibria in Multi-player General-Sum Games. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC '22), June 20–24, 2022, Rome, Italy.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3519935.3520031

1 INTRODUCTION

Online learning and game theory share an intricately connected history tracing back to Robinson's analysis of *fictitious play* [34], as well as Blackwell's seminal *approachability theorem* [5], which served as the advent of the modern *no-regret* framework [1, 23]. These connections have since led to the discovery of broad learning paradigms such as Online Mirror Descent, encompassing algorithms such as the celebrated Multiplicative Weights Update (MWU) [27]. Importantly, *uncoupled* learning dynamics overcome the often unreasonable assumption that players have perfect knowledge of the game, while they have also emerged as a central component in several recent landmark results in computational game solving [7, 28]. Moreover, another compelling feature of the no-regret framework is that it guarantees robustness even against *adversarial* opponents. Indeed, there are broad families of learning paradigms [37] that accumulate $O(\sqrt{T})$ regret after T iterations, a barrier which is known to be insuperable in fully adversarial environments [9]. However, this begs the question: *What if players do not face adversarial losses, but instead face* predictable *losses?*

This question was first addressed by Daskalakis, Deckelbaum and Kim [12]. They devised a decentralized variant of Nesterov's *excessive gap technique* [30], enjoying a near-optimal rate of convergence of $O(\log T/T)$ to Nash equilibrium when employed by

both players in a two-player zero-sum normal-form game. (For brevity we will henceforth omit the specification "normal-form" when referring to games.) At the same time, their algorithm also guarantees optimal (external) regret under worst-case losses. Subsequently, Rakhlin and Sridharan [32, 33] introduced an *optimistic* variant of Online Mirror Descent—considerably simpler than the algorithm proposed in [12]—achieving optimal convergence rate to Nash equilibrium, again in zero-sum games. Then Syrgkanis, Agarwal, Luo and Schapire [40] identified a broad class of *predictive* learning algorithms that induce no-regret learning dynamics in multi-player general-sum games that guarantee $O(T^{1/4})$ regret if followed by each player. This line of work culminated in a recent advancement by Daskalakis, Fishelson and Golowich [13], where it was shown that, when all players employ an optimistic variant of MWU, each player incurs only $O(\text{polylog}(T))$ regret. In turn, this implies that the average product distribution of play induced by optimistic MWU is an $\widetilde{O}\left(T^{-1}\right)$-approximate[1] *coarse correlated equilibrium* (*CCE*) after T repetitions of the game.

Yet, it is well-understood that a CCE prescribes a rather weak notion of equilibrium [21]. An arguably more compelling solution concept[2] in multi-player general-sum games is that of *correlated equilibria* (*CE*) [3]. Like CCE, it is known that CE can be computed through uncoupled learning dynamics. Thus, our paper is concerned with the following central question:

> *Are there learning dynamics that, if followed by all players in a multi-player general game, guarantee convergence with rate $\widetilde{O}\left(T^{-1}\right)$ to a correlated equilibrium?*

The main contribution of our paper is to answer this question in the affirmative. Unlike in the case of CCE, typical no-external-regret dynamics such as MWU are known not to guarantee convergence to CE. Instead, specialized *no-internal-regret* or *no-swap-regret* algorithms have to be employed to converge to CE [9]. Compared to no-external-regret dynamics, these learning dynamics are considerably more complex in that all known algorithms require the computation of the stationary distribution of a certain Markov chain at every iteration. Our main primary technical contribution is to develop techniques to overcome these additional challenges.

1.1 Contributions

Our work presents a refined analysis of the no-internal-regret algorithm of Stoltz and Lugosi [38], as well as the no-swap-regret algorithm of Blum and Mansour [6], both instantiated with Optimistic Multiplicative Weights Update (OMWU). Going forward, we will refer to those learning dynamics as SL-OMWU and BM-OMWU, respectively. Our primary contribution is to show that both of these algorithms exhibit a near-optimal convergence rate of $\widetilde{O}(T^{-1})$, settling our main question in the affirmative. More precisely, for SL-OMWU our main theorem is summarized as follows.

Theorem 1.1. *Consider a general-sum multi-player game with m players, with each player $i \in [[m]]$ having n_i actions. There exists a universal constant $C > 0$ such that, when all players select strategies according to algorithm SL-OMWU with step size $\eta = 1/(C \cdot m \log^4 T)$, the internal regret of every player $i \in [[m]]$ is bounded by $O\left(m \log n_i \log^4 T\right)$. As a result, the average product distribution of play is an $O\left((m \log n \log^4 T)/T\right)$-approximate correlated equilibrium.*

This matches, up to constant factors, the rate of convergence for coarse correlated equilibria as follows by the result in [13], and it is optimal, within the no-regret framework,[3] up to polylogarithmic factors [12]. This also substantially improves upon the $O(T^{-3/4})$ rate of convergence for correlated equilibria recently shown by Chen and Peng [11], both in terms of the dependence on n_i and T. Moreover, since swap regret on an n-simplex is trivially at most n times larger than internal regret (*e.g.*, see Blum and Mansour [6, pp. 1311]), Theorem 1.1 directly gives a bound in terms of swap regret as well, stated as follows.

Corollary 1.2. *If all players select strategies according to algorithm SL-OMWU, the swap regret of every player $i \in [[m]]$ is bounded by $O\left(m\, n_i \log n_i \log^4 T\right)$.*

For the popular and more involved algorithm BM-OMWU, our main theorem is summarized as follows.

Theorem 1.3. *Consider a general-sum multi-player game with m players, with each player $i \in [[m]]$ having n_i actions. There exists a universal constant $C > 0$ such that, when all players select strategies according to algorithm BM-OMWU with step size $\eta = 1/(C \cdot m\, n_i^3 \log^4(T))$, the swap regret of every player $i \in [[m]]$ is bounded by $O(m\, n_i^4 \log n_i \log^4(T))$. As a result, the average product distribution of play is an $O\left((m\, n^4 \log n \log^4 T)/T\right)$-approximate correlated equilibrium.*

Finally, we remark that SL-OMWU and BM-OMWU instantiated with the learning rates described in Theorems 1.1 and 1.3 guarantee near-optimal swap regret (in T) when all players use the same dynamics, but might not against general, adversarial losses. To guarantee near-optimal swap regret in both the adversarial and the non-adversarial regime, an adaptive choice of learning rate similar to that in [13] can be used (see Corollary 3.5).

1.2 Overview of Techniques

The recent work of Daskalakis, Fishelson and Golowich [13] identified *higher-order smoothness* of no-external-regret learning dynamics as a key property for obtaining near-optimal external regret bounds. In particular, they showed that for the no-external-regret dynamics OMWU, the higher-order differences (Definition 3.3) of the sequence of loss vectors decay exponentially at orders up to roughly $\log T$. However, establishing such higher-order smoothness for no-internal- and no-swap-regret learning dynamics is a considerable challenge since the known algorithms involve computing the stationary distribution of a certain Markov chain at every iteration.

[1] As usual, we use the notation $\widetilde{O}(\cdot)$ to suppress polylogarithmic factors of T. Also note that for simplicity, and with a slight abuse of notation, in our introductory section we sometimes use the big-O notation to hide game-specific parameters.

[2] In general-sum multi-player games it is typical to search for solution concepts more permissive than *Nash equilibria* [29] as the latter is known to be computationally intractable under reasonable assumptions [4, 10, 14, 18, 25, 36].

[3] Finding a correlated equilibrium can be phrased as a linear programming problem, and thus ϵ-approximate correlated equilibria can be found in time $\text{poly}(m, n, \log(1/\epsilon))$, where $n = \max_{i \in [[m]]} \{n_i\}$, for succinct multi-player games [31]. However, the procedure for doing so, *ellipsoid against hope*, cannot be phrased as uncoupled dynamics and is unlikely to be run by players competing in a repeated game.

Our main technical contribution is to develop new techniques to effectively address this challenge.

Proof of Theorem 1.1: Analyzing SL-OMWU. First, we show that internal regret minimization on an n-simplex can be simulated by no-external-regret dynamics on the combinatorial space of all n-node *directed trees* (Theorem 3.1). Our equivalence result enables us to trade the computation of a stationary distribution of a polynomial-sized Markov chain for a (much more well-behaved) linear transformation on an exponential-sized set. To our knowledge, this is the first no-internal-to-no-external-regret reduction that sidesteps the computation of stationary distributions of Markov chains, and might have applications beyond the characterization of higher-order smoothness of the dynamics. Based on our equivalence result, we then adapt and leverage the known higher-order smoothness techniques for no-external-regret dynamics [13]. We stress that our analysis is eventually brought back to a "low-dimensional" regret minimizer, instead of solely operating over the space of directed trees; this step is crucial for obtaining the logarithmic dependence on the number of actions of each player for no-internal-regret dynamics (Theorem 1.1).

The equivalence result mentioned in the previous paragraph arises as a consequence of the classic Markov chain tree theorem, which provides a closed-form combinatorial formula for the stationary distribution of an ergodic Markov chain, and crucially relies on the multiplicative structure of the update rule of (O)MWU. Specifically, we prove that the stationary distributions of certain Markov chains whose transition probabilities are updated through (O)MWU are themselves linear transformations of iterates produced by (O)MWU. Our equivalence gives a direct way to argue about the higher-order smoothness of stationary distributions of Markov chains, substantially extending the first-order smoothness observation of Chen and Peng [11]. Furthermore, we expect the equivalence to continue to hold beyond stationary distributions of Markov chains, to the more general problem of computing *fixed points* of linear transformations required in the framework of *Phi-regret* [19, 22, 39].

Proof of Theorem 1.3: Analyzing BM-OMWU. The techniques we described so far enable us to establish the near-optimal internal and swap regret bounds for SL-OMWU (Theorem 1.1 and Corollary 1.2), as well as the corresponding convergence to correlated equilibrium. However, different techniques are necessary to establish the regret bound of BM-OMWU (Theorem 1.3). At a high level, BM-OMWU runs n_i independent external regret minimizers for each player i, aggregates the outputs into a transition matrix of a Markov chain, and then computes its stationary distribution. Because of the independence between the n_i regret minimizers, it is unclear if an analogue of the simulation result (Theorem 3.1) holds.

Thus, rather than arguing indirectly in terms of a supplementary external-regret minimizer, we *directly* analyze the higher-order smoothness of a sequence of stationary distributions of Markov chains, at the cost of ultimately obtaining a worse dependence on the number of actions n_i in our swap regret bounds. Using the machinery of [13] (in particular, the boundedness chain rule of Lemma A.1), doing so boils down to obtaining a bound on the Taylor series coefficients of the function that maps the entries of an ergodic matrix $\mathbf{Q}$ to $\mathbf{Q}$'s stationary distribution. Taken literally, such a bound is not quite possible, since the stationary distribution may have singularities around the non-ergodic matrices. However, we show that by using the Markov chain tree theorem together with the multi-dimensional version of Cauchy's integral formula, it is possible to bound the Taylor series of the function mapping the *logarithms* of the entries of $\mathbf{Q}$ to its stationary distribution (see Lemma C.3). Leveraging the exponential-type structure of the OMWU updates, we then use this bound to obtain the desired guarantee on the higher-order differences of the stationary distributions (Lemma 4.1).

1.3 Further Related Work

No-internal-regret algorithms that require black-box access to a *single* no-external-regret minimizer are known in the literature [9, 38]. This is in contrast with the construction of Blum and Mansour [6] for the stronger notion of swap regret, which requires n independent no-external-regret minimizers—one per each action of the player. Nevertheless, both classes of algorithms involve computing the stationary distribution of a certain Markov chain at every iteration. The intrinsic complexity associated with the computation of a stationary distribution was arguably the main factor limiting our ability to give accelerated convergence guarantees for either class of algorithms. Indeed, while learning dynamics guaranteeing external regret bounded by $O(T^{1/4})$ have been known for several years [40], a matching bound for swap regret was only recently shown by Chen and Peng [11].

The setting studied in our paper (learning dynamics for correlated equilibrium) is substantially more challenging than the problem of giving accelerated learning dynamics for Nash equilibria in two-player zero-sum games, as well learning dynamics for *smooth games* [35]. Indeed, while in our setting the convergence to the equilibrium is driven by the *maximum* internal (or swap) regret cumulated by the players, in the latter two settings the quality metric is driven by the *sum* of the external regrets. As shown by Syrgkanis, Agarwal, Luo and Schapire [40], it is possible to guarantee a *constant sum* of external regrets under a *broad* class of *predictive* no-regret algorithms which includes optimistic OMD and optimistic FTRL under very general distance-generating functions. This is in contrast with the case of no-regret dynamics for CCE and CE, where it remains an open question to give broad classes of algorithms that can achieve near-optimal convergence.

Finally, we point out that optimistic variants of FTRL such as OMWU have been shown to also converge in the *last-iterate sense*, and the convergence is known to be linear[4] [15–17, 41], but this holds only for restricted classes of games, such as two-player zero-sum games.

2 PRELIMINARIES

Consider a finite *normal-form* game Γ consisting of a set of m players $[\![m]\!] \coloneqq \{1, 2, \ldots, m\}$ such that every player i has an *action space* $\mathcal{A}_i \coloneqq [\![n_i]\!]$, for some $n_i \in \mathbb{N}$. The *joint* action space will be represented with $\mathcal{A} \coloneqq \mathcal{A}_1 \times \cdots \times \mathcal{A}_m$. For a given action profile $\boldsymbol{a} = (a_1, \ldots, a_m) \in \mathcal{A}$, the *loss function* $\Lambda_i : \mathcal{A} \to [0, 1]$ specifies the loss of player i under the action profile $\boldsymbol{a}$; note that the

[4] However, the convergence rate of the last-iterate may *not* depend polynomially in the size of the game [41].

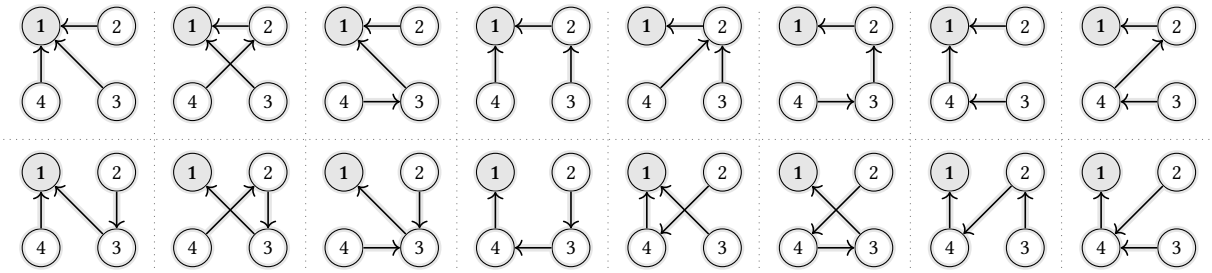

Figure 1: Set $\mathbb{T}_1^4$ of all directed trees rooted at node 1 in a graph with 4 nodes.

normalization of the losses comes without any loss of generality. A *mixed strategy* $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i)$ for a player $i \in [\![m]\!]$ is a probability distribution over i's action space $\mathcal{A}_i$, so that the coordinate $\boldsymbol{x}_i[j]$ indicates the probability that player i will select action $j \in [\![n_i]\!]$. A *deterministic strategy* refers to a mixed strategy supported on a single coordinate. Given a *joint* vector of mixed strategies $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_m)$, the *expected* loss $\boldsymbol{\ell}_i$ of player i is such that $\boldsymbol{\ell}_i[j] := \mathbb{E}_{\boldsymbol{a}_{-i} \sim \boldsymbol{x}_{-i}}[\Lambda_i(j, \boldsymbol{a}_{-i})]$, for all $j \in [\![n_i]\!]$, where we used the notation $\boldsymbol{a}_{-i}$ to denote the vector $\boldsymbol{a}$ excluding the coordinate corresponding to i's action; i.e. $\boldsymbol{a}_{-i} := (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_m)$.

Hindsight Rationality and Regret Minimization. A standard quality metric in the theory of learning in games is *hindsight rationality*. Hindsight rationality encodes the idea that a player has "learnt" to play the game when, looking back at the history of play, there is no transformation of their strategies that—applied to the whole history of play—would have led to strictly better utility for that player. This notion is operationalized through *Phi-regret*. Formally, the Φ_i-regret incurred by the sequence of strategies $\boldsymbol{x}_i^{(1)}, \dots, \boldsymbol{x}_i^{(T)} \in \Delta^{n_i}$ selected by player $i \in [\![m]\!]$ is defined as

$$\mathrm{Reg}_{\Phi_i}^T := \sum_{t=1}^{T} \langle \boldsymbol{x}_i^{(t)}, \boldsymbol{\ell}_i^{(t)} \rangle - \min_{\phi^* \in \Phi_i} \sum_{t=1}^{T} \langle \phi^*(\boldsymbol{x}_i^{(t)}), \boldsymbol{\ell}_i^{(t)} \rangle. \tag{1}$$

Notable special cases are identified based on the particular choice of transformations Φ_i, as follows.

(i) *External* regret (or simply *regret*) corresponds to the case where Φ_i is the set of all *constant* functions $\Phi_i^{\mathrm{const}} := \{\phi : \Delta^{n_i} \to \Delta^{n_i}, \phi(\boldsymbol{x}) = \phi(\boldsymbol{x}') \ \forall \boldsymbol{x}, \boldsymbol{x}' \in \Delta^{n_i}\}$.

(ii) *Internal* regret corresponds to the set of linear transformation $\Phi_i = \Phi_i^{\mathrm{int}}$ that transport probability mass from an action j to some other action k. Formally, Φ_i^{int} is the convex hull $\Phi_i^{\mathrm{int}} := \mathrm{co}\{\phi_{j\to k}\}_{j,k \in \mathcal{A}_i, j \neq k}$ of the functions $\phi_{j\to k}$ defined as

$$\phi_{j\to k} : \boldsymbol{x} \mapsto \boldsymbol{x} + (\boldsymbol{e}_k - \boldsymbol{e}_j)\, \boldsymbol{x}[j] = \left(\mathbf{I} + (\boldsymbol{e}_k - \boldsymbol{e}_j)\boldsymbol{e}_j^\top\right) \boldsymbol{x} =: \mathbf{E}_{j\to k}\, \boldsymbol{x}. \tag{2}$$

(iii) *Swap* regret corresponds to the set Φ_i^{swap} of all linear transformations $\Delta^{n_i} \ni \boldsymbol{x} \mapsto \mathbf{Q}^\top \boldsymbol{x}$ where $\mathbf{Q}$ is a *row-stochastic matrix*, that is, a non-negative matrix whose rows all sum to 1.

Connections with Solution Concepts. There exist connections between the different notions of hindsight rationality described above and game-theoretic solution concepts, including correlated equilibria (the focus of this paper), whose definition is recalled next [20, 23].

Definition 2.1 (Correlated Equilibrium). A probability distribution $\boldsymbol{\mu}$ over $\Pi_1 \times \cdots \times \Pi_m$ is said to be *an ϵ-correlated equilibrium*, where $\epsilon > 0$, if for every player $i \in [\![m]\!]$ and every $\phi \in \Phi_i^{\mathrm{int}}$,

$$\mathbb{E}_{(\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_m) \sim \boldsymbol{\mu}} \left[\langle \boldsymbol{\ell}_i, \boldsymbol{\pi}_i \rangle - \langle \boldsymbol{\ell}_i, \phi(\boldsymbol{\pi}_i) \rangle \right] \leq \epsilon. \tag{3}$$

A folklore result states that the average product distribution of play of no-internal-regret players converges to an approximate *correlated equilibrium* (*CE*). More precisely, the following holds.

Theorem 2.2. *Consider T repetitions of play in a game where every player $i \in [\![m]\!]$ employs an algorithm that produces strategies $\boldsymbol{x}_i^{(1)}, \dots, \boldsymbol{x}_i^{(T)} \in \Delta^{n_i}$ with internal regret IntReg_i^T. Then, the average product distribution of play $\bar{\boldsymbol{\mu}} := \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_1^{(t)} \otimes \cdots \otimes \boldsymbol{x}_m^{(t)}$ is an $O(\max_i \mathrm{IntReg}_i^T / T)$-CE.*

Optimistic Multiplicative Weights Update. The *optimistic* variant of MWU, called *Optimistic Multiplicative Weights Update (OMWU)*[5] is a particular instantiation of the more general *optimistic FTRL* algorithm. At time $t = 1$, OMWU outputs the uniform distribution $\boldsymbol{x}^{(1)} = \frac{1}{n}\mathbf{1} \in \Delta^n$. Then, at each time $t \geq 1$, it computes the next iterate $\boldsymbol{x}^{(t+1)}$ according to the update rule

$$\boldsymbol{x}^{(t+1)}[j] \propto \exp\{-\eta(2\boldsymbol{\ell}^{(t)}[j] - \boldsymbol{\ell}^{(t-1)}[j])\}\, \boldsymbol{x}^{(t)}[j], \tag{OMWU}$$

where $\eta > 0$ is a *learning rate* parameter and $\boldsymbol{\ell}^{(t)}$ is the feedback loss vector observed at time t (we conventionally let $\boldsymbol{\ell}^{(0)} = \mathbf{0}$).

The Markov Chain Tree Theorem. Given an n-state ergodic (*i.e.*, aperiodic and irreducible) Markov chain with (row-stochastic) transition matrix $\mathbf{Q}$, the classic Markov chain tree theorem provides a closed-form solution for its *stationary distribution*, that is, the unique distribution $\boldsymbol{\pi} \in \Delta^n$ such that $\boldsymbol{\pi}^\top \mathbf{Q} = \boldsymbol{\pi}^\top$. The result requires the notion of a *directed rooted tree* (a.k.a. *arborescence*), recalled next.

Definition 2.3. Let $\mathcal{T} = (V, E)$, with $V = [\![n]\!]$, be an n-node directed graph. $\mathcal{T}$ is a *directed tree rooted at $j \in V$* if (i) it contains no cycle (including self-loops); (ii) every node $V \setminus \{j\}$ has *exactly one* outgoing edge; and (iii) node j has *no* outgoing edges. We denote the set of all n-node directed trees rooted at $j \in [\![n]\!]$ with the symbol $\mathbb{T}_j^n$. Furthermore, we let $\mathbb{T}^n := \mathbb{T}_1^n \cup \cdots \cup \mathbb{T}_n^n$.

It is well-known that $|\mathbb{T}_j^n| = n^{n-2}$ for all $j \in [\![n]\!]$, and consequently $|\mathbb{T}^n| = n^{n-1}$ [8]. An example for $n = 4$ is given in Figure 1. In

[5] The OMWU algorithm is sometimes referred to as *Optimistic Hedge* in the literature.

this context, the Markov chain tree theorem asserts that the vector $(\Sigma_1, \ldots, \Sigma_n)$ of quantities

$$\Sigma_j \coloneqq \sum_{\mathcal{T} \in \mathbb{T}_j} \prod_{(a,b) \in E(\mathcal{T})} Q[a,b] \qquad (j \in [[n]]) \quad (4)$$

is proportional to the stationary distribution. Specifically, we have the following.

Theorem 2.4 (Markov chain tree theorem). *The stationary distribution $\boldsymbol{\pi}$ of an n-state ergodic Markov chain satisfies $\boldsymbol{\pi}[j] = \Sigma_j / \Sigma$ for all $j \in [[n]]$, where $\Sigma \coloneqq \Sigma_1 + \cdots + \Sigma_n$.*

For a proof of the Markov chain tree theorem, we refer the interested reader to the works of Anantharam and Tsoucas [2], Kruckman, Greenwald and Wicks [26].

3 NEAR-OPTIMAL NO-INTERNAL-REGRET DYNAMICS FOR CORRELATED EQUILIBRIA

In this section we establish our main result regarding algorithm SL-OMWU, namely Theorem 1.1. In particular, we start in Section 3.1 by formalizing one of our key insights: SL-OMWU can be equivalently thought of as a certain linear transformation of the output of an external regret minimizer operating over the combinatorial space of directed trees; this equivalence is formalized in Theorem 3.1. Next, in Section 3.2 we leverage this connection to bound the internal regret of SL-OMWU using an extension of the techniques developed in [13].

3.1 Equivalence Result

Before we proceed with the statement and proof of Theorem 3.1, we first summarize SL-OMWU in Algorithm 1. In addition, in Algorithm 2 we present an external regret minimizer over the space of all directed trees (a.k.a. arborescences). Both Algorithms 1 and 2 use the symbol $\boldsymbol{x}^{(t)}$ to denote the output strategies; this choice is justified by the following theorem (see also Figure 2).

Theorem 3.1. *For any learning rate $\eta > 0$, Algorithms 1 and 2 produce the same strategies $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(T)} \in \Delta^n$, assuming that they observe the same sequence of losses $\boldsymbol{\ell}^{(1)}, \ldots, \boldsymbol{\ell}^{(T)} \in \mathbb{R}^n$.*

Proof. We will inductively show that the following property holds.

Lemma 3.2. *At all times $t \geq 1$, the following conditions hold:*

(1) there exists $N^{(t)} \in \mathbb{R}$ such that

$$\prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t)}[a \to b] = N^{(t)} \boldsymbol{X}^{(t)}[\mathcal{T}] \qquad \forall \mathcal{T} \in \mathbb{T}^n, \quad (5)$$

where $\boldsymbol{p}^{(t)}$ and $\boldsymbol{X}^{(t)}$ are as defined in Algorithm 1 and Algorithm 2, respectively; and

(2) the strategies produced by Algorithms 1 and 2 are the same.

Condition 2 of Lemma 3.2 immediately implies the statement. First, we establish the base case $t = 1$. The first iterate of OMWU is always the uniform distribution; so, $\boldsymbol{p}^{(1)} = 1/n(n-1)\,\mathbf{1} \in \Delta^{n(n-1)}$. Hence, the matrix $\mathbf{M}^{(1)}$ (Line 3 of Algorithm 1) has entries $\mathbf{M}[j,k] = 1/n(n-1)$ for all $j \neq k \in [[n]]$. Correspondingly, the (unique) stationary distribution of $\mathbf{M}^{(1)}$ is the uniform distribution $\boldsymbol{x}^{(1)} = \frac{1}{n}\mathbf{1} \in \Delta^n$. We now verify that the same iterate is produced by Algorithm 2. Since $\boldsymbol{X}^{(1)}$ is computed using OMWU, $\boldsymbol{X}^{(1)} \in \Delta^{n^{n-1}}$ is the uniform distribution $\boldsymbol{X}^{(1)}[\mathcal{T}] = 1/n^{n-1}$. So, using the fact that $|\mathbb{T}_j^n| = n^{n-2}$ for all $j \in [[n]]$, each coordinate of the output strategy of Algorithm 2 is equal to $n^{n-2}/n^{n-1} = 1/n$, establishing the base case for Condition 2 of Lemma 3.2. Furthermore, since $\boldsymbol{p}^{(1)}$ and $\boldsymbol{X}^{(1)}$ are the uniform distributions over $\Delta^{n(n-1)}$ and $\Delta^{(n^{n-1})}$ respectively, Condition 1 follows directly from the fact that any directed tree $\mathcal{T} \in \mathbb{T}^n$ has exactly $n-1$ edges.

Next, we prove the inductive step. Assume that Lemma 3.2 holds at times $\tau = 1, \ldots, t$, for some $t \geq 1$. We will prove that it will hold at time $t+1$ as well. Since $\boldsymbol{p}^{(t+1)} \in \Delta^{n(n-1)}$ in Algorithm 1 is updated using (OMWU), for all $j \neq k \in [[n]]$

$$\frac{\boldsymbol{p}^{(t+1)}[j \to k]}{\boldsymbol{p}^{(t)}[j \to k]} \propto \exp\{-\eta(2\boldsymbol{L}^{(t)}[j \to k] - \boldsymbol{L}^{(t-1)}[j \to k])\}, \quad (6)$$

where the loss vector

$$\boldsymbol{L}^{(t)}[j \to k] \coloneqq \boldsymbol{x}^{(t)}[j](\boldsymbol{\ell}^{(t)}[k] - \boldsymbol{\ell}^{(t)}[j]) \quad (7)$$

is as defined on Line 8 of Algorithm 1. For convenience, we will denote the normalization parameter in Equation (6) as $S^{(t+1)}$. Similarly, since the strategies $\boldsymbol{X}^{(t)} \in \Delta^{(n^{n-1})}$ of Algorithm 2 are updated using (OMWU) with the *same* learning rate $\eta > 0$, we have that

$$\boldsymbol{X}^{(t+1)}[\mathcal{T}] \propto \exp\{-\eta(2\boldsymbol{\mathcal{L}}^{(t)}[\mathcal{T}] - \boldsymbol{\mathcal{L}}^{(t-1)}[\mathcal{T}])\}\boldsymbol{X}^{(t)}[\mathcal{T}], \quad (8)$$

for all $\mathcal{T} \in \mathbb{T}^n$, where $\boldsymbol{\mathcal{L}}^{(t)}$ is defined on Line 7 of Algorithm 2 as

$$\sum_{(a,b) \in E(\mathcal{T})} \boldsymbol{x}^{(t)}[a](\boldsymbol{\ell}^{(t)}[b] - \boldsymbol{\ell}^{(t)}[a]) = \sum_{(a,b) \in E(\mathcal{T})} \boldsymbol{L}^{(t)}[a \to b], \quad (9)$$

for all $\mathcal{T} \in \mathbb{T}^n$. (Note that in (7) and (9) we implicitly used the inductive hypothesis that the strategies of the two algorithms coincide at iterate t, as well as the assumption that the sequence of losses $\boldsymbol{\ell}^{(t)}$ observed by Algorithms 1 and 2 is the same.) To simplify the notation, we will use $\mathcal{S}^{(t+1)}$ to represent the denominator in (8). Next, we observe that for any $\mathcal{T} \in \mathbb{T}^n$ the term $\prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t+1)}[a \to b]$ is equal to

$$\left(\frac{1}{S^{(t+1)}}\right)^{n-1} \exp\left\{-\eta\left(2\boldsymbol{\mathcal{L}}^{(t)}[\mathcal{T}] - \boldsymbol{\mathcal{L}}^{(t-1)}[\mathcal{T}]\right)\right\} \times \prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t)}[a \to b], \quad (10)$$

where we used the fact that every directed tree $\mathcal{T}$ has exactly $n-1$ edges; the multiplicative properties of the exponential; and the definition of $\boldsymbol{\mathcal{L}}^{(t)}$ given in (9). From the inductive hypothesis, Condition 1 of Lemma 3.2 holds at time t; so, continuing (10), $\prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t+1)}[a \to b]$ is equal to

$$\left(\frac{1}{S^{(t+1)}}\right)^{n-1} \exp\left\{-\eta\left(2\boldsymbol{\mathcal{L}}^{(t)}[\mathcal{T}] - \boldsymbol{\mathcal{L}}^{(t-1)}[\mathcal{T}]\right)\right\} N^{(t)} \boldsymbol{X}^{(t)}[\mathcal{T}]$$

$$= \left(\frac{1}{S^{(t+1)}}\right)^{n-1} \mathcal{S}^{(t+1)} N^{(t)} \boldsymbol{X}^{(t+1)}[\mathcal{T}], \quad (11)$$

where (11) follows from (8) and the definition of $\mathcal{S}^{(t+1)}$. As a result, we have shown that Condition 1 of Lemma 3.2 holds at time $t+1$ for the parameter

$$N^{(t+1)} \coloneqq \left(\frac{1}{S^{(t+1)}}\right)^{n-1} \mathcal{S}^{(t+1)} N^{(t)}. \quad (12)$$

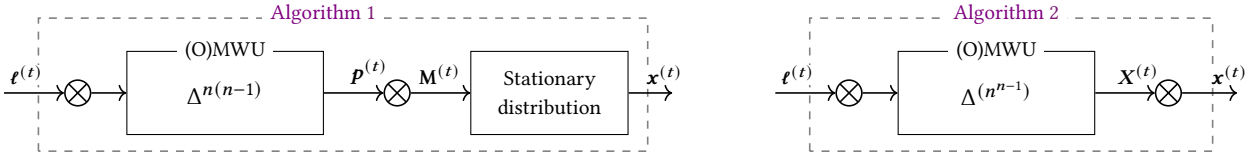

Figure 2: A schematic illustration of the equivalence result of Theorem 3.1; ⊗ in the figure represents a linear transformation.

Algorithm 1: Stoltz and Lugosi [38]

Data: $\mathcal{R}_\Delta$: OMWU algorithm for $\Delta^{n(n-1)}$ with $\eta > 0$

1 **function** *NextStrategy*()
2 $\quad \boldsymbol{p}^{(t)} \leftarrow \mathcal{R}_\Delta.\text{NextStrategy}()$
3 $\quad \mathbf{M}^{(t)} \leftarrow \sum_{j \neq k \in [[n]]} \boldsymbol{p}^{(t)}[j \to k]\, \mathbf{E}_{j \to k}^\top$
4 $\quad$ **return** $\boldsymbol{x}^{(t)} \leftarrow \text{StationDistrib}(\mathbf{M}^{(t)})$ [6]
5 **function** *ObserveUtility*($\boldsymbol{\ell}^{(t)}$)
6 $\quad \boldsymbol{L}^{(t)} \leftarrow \mathbf{0} \in \mathbb{R}^{n(n-1)}$
7 $\quad$ **for** $j \neq k \in [[n]]$ **do**
8 $\quad\quad \boldsymbol{L}^{(t)}[j \to k] \leftarrow \boldsymbol{x}^{(t)}[j](\boldsymbol{\ell}^{(t)}[k] - \boldsymbol{\ell}^{(t)}[j])$
9 $\quad \mathcal{R}_\Delta.\text{ObserveUtility}(\boldsymbol{L}^{(t)})$

Algorithm 2: Arborescence-based dynamics

Data: $\mathcal{R}_\Delta$: OMWU algorithm for $\Delta^{(n^{n-1})}$ with $\eta > 0$

1 **function** *NextStrategy*()
2 $\quad \boldsymbol{X}^{(t)} \leftarrow \mathcal{R}_\Delta.\text{NextStrategy}()$
3 $\quad$ **return** $\boldsymbol{x}^{(t)} \leftarrow \left(\sum_{\mathcal{T} \in \mathbb{T}_j^n} \boldsymbol{X}^{(t)}[\mathcal{T}]\right)_{j=1}^n$
4 **function** *ObserveUtility*($\boldsymbol{\ell}^{(t)}$)
5 $\quad \boldsymbol{\mathcal{L}}^{(t)} = \mathbf{0} \in \mathbb{R}^{|\mathbb{T}^n|} = \mathbb{R}^{(n^{n-1})}$
6 $\quad$ **for** $\mathcal{T} \in \mathbb{T}^n$ **do**
7 $\quad\quad \boldsymbol{\mathcal{L}}^{(t)}[\mathcal{T}] \leftarrow \sum_{(j,k) \in E(\mathcal{T})} \boldsymbol{x}^{(t)}[j](\boldsymbol{\ell}^{(t)}[k] - \boldsymbol{\ell}^{(t)}[j])$
8 $\quad \mathcal{R}_\Delta.\text{ObserveUtility}(\boldsymbol{\mathcal{L}}^{(t)})$

We now show that Condition 2 of Lemma 3.2 holds at time $t+1$ as well. To do so, we analyze the (unique) stationary distribution of the matrix $\mathbf{M}^{(t+1)}$ defined on Line 3 of Algorithm 1. First, we claim that for any $j \neq k \in [[n]]$, $\mathbf{M}^{(t+1)}[j,k] = \boldsymbol{p}^{(t+1)}[j \to k]$. Indeed, for any $j \neq k \in [[n]]$ the unique non-zero non-diagonal entry of the matrix $\mathbf{E}_{j \to k}$ appears as $\mathbf{E}_{j \to k}[k,j] = 1$ (recall (2)). Thus, our claim follows directly by the definition of the matrix $\mathbf{M}^{(t+1)}$ in Line 3 of Algorithm 1. As a result, the j-th coordinate of the fixed point $\boldsymbol{x}^{(t+1)}$ of $\mathbf{M}^{(t+1)}$ can be expressed using the Markov chain tree theorem (Theorem 2.4) as

$$\boldsymbol{x}^{(t+1)}[j] = \frac{\sum_{\mathcal{T} \in \mathbb{T}_j} \prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t+1)}[a \to b]}{\sum_{j=1}^n \sum_{\mathcal{T} \in \mathbb{T}_j^n} \prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t+1)}[a \to b]}. \tag{13}$$

Using (11) together with the fact that $\boldsymbol{X}^{(t+1)} \in \Delta^{(n^{n-1})}$, and therefore $\sum_{\mathcal{T} \in \mathbb{T}^n} \boldsymbol{X}^{(t+1)}[\mathcal{T}] = 1$, the denominator of (13) satisfies

$$\sum_{j=1}^n \sum_{\mathcal{T} \in \mathbb{T}_j^n} \prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t+1)}[a \to b] = \sum_{\mathcal{T} \in \mathbb{T}^n} N^{(t+1)} \boldsymbol{X}^{(t+1)}[\mathcal{T}] = N^{(t+1)}. \tag{14}$$

Similarly, using Equation (11) together with (12), the numerator of (13) can be expressed as

$$\sum_{\mathcal{T} \in \mathbb{T}_j^n} \prod_{(a,b) \in E(\mathcal{T})} \boldsymbol{p}^{(t+1)}[a \to b] = N^{(t+1)} \sum_{\mathcal{T} \in \mathbb{T}_j^n} \boldsymbol{X}^{(t+1)}[\mathcal{T}]. \tag{15}$$

[6] The matrix $\mathbf{M}^{(t)}$ has strictly positive entries at all times t, since the iterates $\boldsymbol{p}^{(t)}$ produced by (OMWU) lie in the relative interior of the simplex. So, each $\mathbf{M}^{(t)}$ admits a *unique* fixed point (stationary distribution).

Finally, plugging (14) and (15) into Equation (13) we can conclude that $\boldsymbol{x}^{(t+1)}[j] = \sum_{\mathcal{T} \in \mathbb{T}_j^n} \boldsymbol{X}^{(t+1)}[\mathcal{T}]$, which is exactly the j-th coordinate of the iterate produced by Algorithm 2 at time $t+1$ (Line 3). Thus, the strategies of Algorithms 1 and 2 at time $t+1$ are the same, completing the inductive proof. □

3.2 Bounding the Internal Regret

Here we explain how to leverage the techniques in [13] to bound the external regret of $\mathcal{R}_\Delta$ employed in Algorithm 1, which also bounds the internal regret of SL-OMWU [38]. In particular, the crux in the analysis lies in showing that the sequence of observed losses of $\mathcal{R}_\Delta$ exhibits *higher-order smoothness*. Let us first recall the notion of finite differences.

Definition 3.3. Consider a sequence of vectors $\boldsymbol{z} = (\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(T)})$. For an integer $h \geq 0$, the *h-order finite difference* for the sequence $\boldsymbol{z}$, denoted by $D_h \boldsymbol{z}$, is the sequence

$$D_h \boldsymbol{z} := ((D_h \boldsymbol{z})^{(1)}, \dots, (D_h \boldsymbol{z})^{(T-h)})$$

defined recursively as $(D_0 \boldsymbol{z})^{(t)} := \boldsymbol{z}^{(t)}$, for $1 \leq t \leq T$, and

$$(D_h \boldsymbol{z})^{(t)} := (D_{h-1} \boldsymbol{z})^{(t+1)} - (D_{h-1} \boldsymbol{z})^{(t)}, \tag{16}$$

for $h \geq 1$, and $1 \leq t \leq T - h$.

To establish higher-order smoothness, we use Theorem 3.1 to "lift" the analysis to the regret minimizer over the arborescences (Algorithm 2). Then, we leverage the particular structure of the losses in SL-OMWU to adapt the argument in [13], leading to the following guarantee.

Lemma 3.4. *Consider a parameter $\alpha \leq 1/(H+3)$. If all players employ SL-OMWU with learning rate $\eta \leq \frac{\alpha}{36m}$, then for any player $i \in [[m]]$, $0 \leq h \leq H$ and $t \in [[T-h]]$ it holds that*

$$\|(D_h \boldsymbol{L}_i)^{(t)}\|_\infty \leq \alpha^h h^{3h+1}.$$

Recall that $\boldsymbol{L}_i^{(t)}$ is the loss observed by algorithm $\mathcal{R}_\Delta$ at time t (see Algorithm 1). Armed with this crucial lemma, we are also able to extend the other technical ingredients used in [13], as we formally show in Appendix B.

Adversarial Bound. The learning algorithm can also be slightly modified in order to guarantee robustness when faced against adversarial losses. Indeed, the following corollary implies near-optimal internal regret in both regimes.

Corollary 3.5. *There exists a learning algorithm such that, if employed by all players, it guarantees that the internal regret of each player $i \in [[m]]$ is bounded by $O(m \log n_i \log^4 T)$. Moreover, under adversarial losses the algorithm ensures internal regret bounded by $O(m \log n_i \log^4 T + \sqrt{\log n_i T})$.*

The idea is to use an adaptive choice of learning rate depending on whether the bound predicted in (48) has been violated in some repetition of the game, analogously to [13, Corollary D.1].

4 ANALYSIS OF BLUM-MANSOUR

In this section we give an overview of our analysis for the no-swap-regret algorithm of Blum and Mansour [6] (BM). Unlike the no-internal-regret algorithm of Stoltz and Lugosi [38], BM maintains n independent no-external-regret algorithms $\mathcal{R}_{\Delta,1}, \ldots, \mathcal{R}_{\Delta,n}$, operating over Δ^n. For this reason, our previous approach appears to be no longer applicable. Instead, we develop more robust techniques that delve into the inner-workings of the higher-order smoothness argument in [13]. In particular, we substantially generalize their approach by demonstrating how higher-order smoothness bounds can be established even under the additional complexity of fixed point operations.

First, to keep the exposition reasonably self-contained, let us briefly recall the BM algorithm. At each time $t \geq 1$, every algorithm $\mathcal{R}_{\Delta,g}$ produces an iterate $\mathbf{Q}^{(t)}[g,\cdot] = (\mathbf{Q}^{(t)}[g,1], \ldots, \mathbf{Q}^{(t)}[g,n]) \in \Delta^n$, for all $g \in [[n]]$. Then, the algorithm computes a stationary distribution $\Delta^n \ni \boldsymbol{x}^{(t)} = (\mathbf{Q}^{(t)})^\top \boldsymbol{x}^{(t)}$ of the transition matrix $\mathbf{Q}^{(t)}$. Moreover, upon receiving a loss vector $\boldsymbol{\ell}^{(t)} \in \mathbb{R}^n$, the BM algorithm distributes the loss vector $\boldsymbol{x}^{(t)}[g]\boldsymbol{\ell}^{(t)}$ to each regret minimizer $\mathcal{R}_{\Delta,g}$ for $g \in [[n]]$. In what follows, we will be concerned with the particular case where each regret minimizer is set to OMWU.

Our primary technical contribution in the analysis of BM-OMWU is to show that the losses observed by each individual regret minimizer exhibit high-order smoothness. Namely, we show the following lemma.

Lemma 4.1. *Fix a parameter $\alpha \in \left(0, \frac{1}{H+3}\right)$. There exists a sufficiently large universal constant C such that, if all players follow BM-OMWU with learning rate $\eta \leq \frac{1}{CH^2 n_i^3}$, then for any player $i \in [[m]]$, integer h satisfying $0 \leq h \leq H$, time step $t \in [[T-h]]$, and $g \in [[n_i]]$, it holds that*

$$\| \left(\mathrm{D}_h \left(\boldsymbol{x}_i[g] \cdot \boldsymbol{\ell}_i\right)\right)^{(t)} \|_\infty \leq \alpha^h h^{3h+1}.$$

At a high level, the proof of this higher-order smoothness lemma hinges on the cyclic relationship between the losses incurred by the players $\boldsymbol{\ell}_i^{(t)}$, the iterates produced by the copies of the OMWU algorithm $\mathbf{Q}_i^{(t)}$, and the final strategies $\boldsymbol{x}_i^{(t)}$ output by BM-OMWU. In particular, the iterates $\mathbf{Q}_i^{(t)}[g,\cdot]$ are determined based on the overall history of loss vectors $\sum_{t'<t} \boldsymbol{x}_i^{(t')}[g]\boldsymbol{\ell}_i^{(t')}$ weighted exponentially in terms of the *softmax* function; the strategies $\boldsymbol{x}_i^{(t)}$ are determined by applying the Markov Chain Tree Theorem (Theorem 2.4) to the stochastic matrix $\mathbf{Q}_i^{(t)}$ formed by the previous iterates; and, the losses $\boldsymbol{\ell}_i^{(t)}$ are determined as a function of the strategies of all the other players $\boldsymbol{x}_{-i}^{(t)}$. Therefore, using the "Boundedness Chain Rule for Finite Differences" (Lemma A.1), one of the main technical tools shown in [13], we can demonstrate that bounds on the h-th order finite differences of $\boldsymbol{x}_i^{(t)}$ and $\boldsymbol{\ell}_i^{(t)}$ imply a bound on the $(h+1)$-th order finite differences of $\mathbf{Q}^{(t)}$.[7] This, in turn, implies a bound on the $(h+1)$-th order finite differences of $\boldsymbol{x}_i^{(t)}$, which then gives a bound on the $(h+1)$-th order finite differences of $\boldsymbol{\ell}_i^{(t)}$, as long as the Taylor coefficients of the softmax function and the Markov Chain Tree Theorem are sufficiently well-behaved. While this is not quite the case, we make a slight modification to this cycle of implication, bounding the finite differences of $\log \mathbf{Q}_i^{(t)}$ instead. This cycle of implication enables an inductive argument that proves Lemma 4.1 as long as the log of the softmax function and an exponential version of the Markov Chain Tree Theorem exhibit bounded Taylor coefficients. These bounds are proved in Lemma C.6 and Lemma C.3 respectively. The proof of Lemma C.6 follows an explicit, combinatorial framework similar to that presented in [13]. On the other hand, Lemma C.3 introduces a novel technique for proving Taylor coefficient bounds, applying the Cauchy Integral Formula. This approach is far more general, and could help establish the necessary preconditions of Boundedness Chain Rule Lemma in much broader settings wherein combinatorial approaches are insufficient.

5 CONCLUSIONS AND OPEN PROBLEMS

In conclusion, we have extended the recent result of Daskalakis, Fishelson and Golowich [13] from external to internal and swap regret. As a corollary, we obtained the first near-optimal—within the no-regret framework—rates of convergence for correlated equilibrium. To do so, we developed several new techniques that allowed us to establish higher-order smoothness for no-internal and no-swap learning dynamics.

Finally, we identify several possible avenues for future research related to our results.

- Although our internal-regret bounds are near-optimal in terms of the dependency on the number of actions n_i of each player i, for swap regret our bounds depend polynomially on n_i. While a polynomial dependence on n_i is necessary in the adversarial setting [6, 24], we are not aware of any lower bounds for the setting of smooth, predictable sequences of losses within which our paper operates.
- Can our results be extended beyond OMWU, for example to other instances of the general optimistic FTRL algorithm [40]?
- Finally, our equivalence theorem (Theorem 3.1) was only established with respect to the set Φ_i^{int} of transformations corresponding to internal regret. Is it possible to extend our results beyond

[7] This is due to the fact that $D_{h+1}\mathbf{Q}^{(t)}[g,\cdot] = D_{h+1}\sum_{t'<t} \boldsymbol{x}_i^{(t')}[g]\boldsymbol{\ell}_i^{(t')} = D_h \boldsymbol{x}_i^{(t)}[g]\boldsymbol{\ell}_i^{(t)}$.

such transformations (*e.g.*, see [19]) via closed-form formulas for the associated fixed points, analogous to the Markov chain tree theorem? Exploring such connections further constitutes a promising direction for the future.

ACKNOWLEDGMENTS

C.D. is supported by NSF Awards CCF-1901292, DMS-2022448 and DMS-2134108, by a Simons Investigator Award, by the Simons Collaboration on the Theory of Algorithmic Fairness, by a DSTA grant, and by the DOE PhILMs project (No. DE-AC05-76RL01830). N.G. is supported by a Fannie & John Hertz Foundation Fellowship and an NSF Graduate Fellowship. T.S. is supported by NSF grants IIS-1718457, IIS-1901403, and CCF-1733556, and ARO award W911NF2010081.

A USEFUL TECHNICAL TOOLS

Most of the following technical ingredients were shown in [13], but we include them to keep the exposition reasonably self-contained. First, it will be useful to express h-order finite differences, as introduced in Definition 3.3, in the following form ([13, Remark 4.3]):

$$(D_h z)^{(t)} = \sum_{s=0}^{h} \binom{h}{s} (-1)^{h-s} z^{(t+s)}. \tag{17}$$

Next, we state the "boundedness chain rule" for finite differences [13]. Specifically, let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a real function analytic in a neighborhood of the origin. For real numbers $Q, R > 0$, we will say that ϕ is (Q, R)-*bounded* if the Taylor series of ϕ with respect to the origin, denoted with $P_\phi(z_1, \dots, z_n) = \sum_{\gamma \in \mathbb{Z}^n_{\geq 0}} \alpha_\gamma z^\gamma$ is such that

$$\sum_{\gamma \in \mathbb{Z}^n_{\geq 0} : |\gamma| = k} |\alpha_\gamma| \leq Q R^k, \tag{18}$$

for any integer $k \geq 0$. The following result [13, Lemma 4.5] is one of the central technical ingredients developed in [13].

Lemma A.1 ([13]). *Consider a (Q, R)-bounded analytic function $\phi \in \mathbb{R}^n \to \mathbb{R}$ so that the radius of convergence of its power series with respect to the origin is greater than ν, for some $\nu > 0$. Moreover, consider a sequence of vectors $(z^{(1)}, \dots, z^{(T)})$ such that $\|z^{(t)}\|_\infty \leq \nu$ for all $t \in [\![T]\!]$. Finally, suppose that for some parameter $\alpha \in (0, 1)$, and for each $0 \leq h' \leq h$ and $t \in [\![T - h']\!]$, it holds that[8] $\|(D_{h'} z)^{(t)}\|_\infty \leq \frac{1}{B_1} \alpha^{h'} (h')^{B_0 h'}$, where $B_1 \geq 2e^2 R$ and $B_0 \geq 3$. Then, for all $t \in [\![T - h]\!]$ it holds that*

$$\left| (D_h(\phi \circ z))^{(t)} \right| \leq \frac{12RQe^2}{B_1} \alpha^h h^{B_0 h + 1}.$$

In the statement of this lemma the notation $\circ$ represents the composition of functions. Further, we remark that while technically the statement of Lemma A.1 in [13] is only stated for the special case $\nu = 1$, the lemma readily extends for general ν.

We additionally make use of the following lemma from [13] that enables us to bound the variance of the loss sequences arising from OMWU as a result of the smoothness of the sequences. In the following lemma, we let $\mathrm{Var}_{\boldsymbol{q}}(z) \coloneqq \sum_{j=1}^n \boldsymbol{q}(j) \left(z(j) - \sum_{k=1}^n \boldsymbol{q}(k) z(k) \right)^2$.

[8] Here it is assumed that $0^0 = 1$.

Lemma A.2. *For any integers $n \geq 2$ and $T \geq 4$, we set $H \coloneqq \lceil \log T \rceil$, $\alpha = 1/(4H)$, and $\alpha_0 = \frac{\sqrt{\alpha/8}}{H^3}$. Suppose that $Z^{(1)}, \dots, Z^{(T)} \in [0, 1]^n$ and $\boldsymbol{P}^{(1)}, \dots, \boldsymbol{P}^{(T)} \in \Delta^n$ satisfy the following*

1. *For each $0 \leq h \leq H$ and $1 \leq t \leq T - h$, it holds that $\left\| (\mathrm{D}_h Z)^{(t)} \right\|_\infty \leq H \cdot \left(\alpha_0 H^3 \right)^h$;*
2. *The sequence $\boldsymbol{P}^{(1)}, \dots, \boldsymbol{P}^{(T)}$ is ζ-consecutively close for some $\zeta \in [1/(2T), \alpha^4/8256]$.*

Then,

$$\sum_{t=1}^{T} \mathrm{Var}_{\boldsymbol{P}^{(t)}} \left(Z^{(t)} - Z^{(t-1)} \right) \leq 2\alpha \sum_{t=1}^{T} \mathrm{Var}_{\boldsymbol{P}^{(t)}} \left(Z^{(t-1)} \right) + 165120(1 + \zeta) H^5 + 2,$$

where we say that a sequence $\boldsymbol{P}^{(1)}, \dots, \boldsymbol{P}^{(T)}$ is ζ-consecutively close if for all t,

$$\max \left\{ \left\| \frac{\boldsymbol{P}^{(t)}}{\boldsymbol{P}^{(t+1)}} \right\|_\infty, \left\| \frac{\boldsymbol{P}^{(t+1)}}{\boldsymbol{P}^{(t)}} \right\|_\infty \right\} \leq 1 + \zeta.$$

B ANALYSIS OF STOLTZ-LUGOSI

In this section we provide all of the technical ingredients required for the analysis of SL-OMWU, and subsequently for the proof of Theorem 1.1. First, let us cast the refined bound under adversarial losses [13, Lemma 4.1] in our setting.

Lemma B.1 ([13]). *Consider some player $i \in [\![m]\!]$ employing SL-OMWU with learning rate $\eta < 1/C$, where C is a sufficiently large universal constant. Then, under any sequence of losses $\boldsymbol{L}_i^{(1)}, \dots, \boldsymbol{L}_i^{(T)}$, the external regret of $\mathcal{R}_\Delta$ can be bounded as*

$$\mathrm{Reg}_i^T \leq 2\frac{\log n_i}{\eta} + \sum_{t=1}^{T} \left(\frac{\eta}{2} + C\eta^2 \right) \mathrm{Var}_{\boldsymbol{p}_i^{(t)}} \left(\boldsymbol{L}_i^{(t)} - \boldsymbol{L}_i^{(t-1)} \right) - \sum_{t=1}^{T} \frac{(1 - C\eta)\eta}{2} \mathrm{Var}_{\boldsymbol{p}_i^{(t)}} \left(\boldsymbol{L}_i^{(t-1)} \right). \tag{19}$$

Recall from Algorithm 1 that $\boldsymbol{L}_i^{(t)}$ is the loss observed by the regret minimizer $\mathcal{R}_\Delta$ employing (OMWU). Next, we continue with the proof of Lemma 3.4. For the convenience of the reader, the statement of the lemma is included below.

Lemma 3.4. *Consider a parameter $\alpha \leq 1/(H + 3)$. If all players employ SL-OMWU with learning rate $\eta \leq \frac{\alpha}{36m}$, then for any player $i \in [\![m]\!]$, $0 \leq h \leq H$ and $t \in [\![T - h]\!]$ it holds that*

$$\|(D_h \boldsymbol{L}_i)^{(t)}\|_\infty \leq \alpha^h h^{3h+1}.$$

Proof. First of all, we know that

$$\boldsymbol{\ell}_i^{(t)}[a_i] = \sum_{a_{i'} \in [\![n_{i'}]\!], i' \neq i} \Lambda(a_1, \dots, a_i, \dots, a_m) \prod_{i' \neq i} \boldsymbol{x}_{i'}^{(t)}[a_{i'}], \tag{20}$$

where recall that by assumption $\Lambda(\cdot) \in [0, 1]$. In particular, given that $\boldsymbol{L}_i^{(t)}[j \to k] = \boldsymbol{x}_i^{(t)}[j](\boldsymbol{\ell}_i^{(t)}[k] - \boldsymbol{\ell}_i^{(t)}[j])$, we may conclude that

$$\boldsymbol{L}_i^{(t)}[j \to k] = \sum_{\mathcal{T}_{i'} \in \mathbb{T}_{i'}} \Lambda'(\mathcal{T}_1, \dots, \mathcal{T}_m) \prod_{i' \in [\![m]\!]} \boldsymbol{X}_{i'}^{(t)}[\mathcal{T}_{i'}], \tag{21}$$

for some function $\Lambda' : \mathbb{T}_1 \times \cdots \times \mathbb{T}_m \to [-1,1]$, where we used that $\boldsymbol{x}_{i'}^{(t)}[a_{i'}] = \sum_{\mathcal{T}_{i'} \in \mathbb{T}_{i',a_{i'}}} \boldsymbol{X}_{i'}^{(t)}[\mathcal{T}_{i'}]$ (by Theorem 3.1), as well as the fact that the sets of directed trees with different roots are disjoint. As a result, we have that $\left|(D_h \boldsymbol{L}_i)^{(t)}[j \to k]\right|$ is equal to

$$\left|\sum_{s=0}^{h}\binom{h}{s}(-1)^{h-s}\boldsymbol{L}_i^{(t+s)}[j\to k]\right| \tag{22}$$

$$= \left|\sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\Lambda'(\mathcal{T}_1,\dots,\mathcal{T}_m)\sum_{s=0}^{h}\binom{h}{s}(-1)^{h-s}\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t+s)}[\mathcal{T}_{i'}]\right| \tag{23}$$

$$\le \sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\left|\sum_{s=0}^{h}\binom{h}{s}(-1)^{h-s}\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t+s)}[\mathcal{T}_{i'}]\right| \tag{24}$$

$$= \sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\left|\left(D_h\left(\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]\right)\right)^{(t)}\right|, \tag{25}$$

where (22) uses the equivalent formulation of (17) for h-order finite differences; (23) follows from (21); (24) follows from the triangle inequality and the fact that $|\Lambda'(\cdot)| \in [0,1]$; and the final line uses again the equivalent formulation of finite differences of (17), with the convention that $\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]$ refers to the sequence $\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(1)}[\mathcal{T}_{i'}],\dots,\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(T)}[\mathcal{T}_{i'}]$. Next, it will be convenient to assume that $\boldsymbol{X}_i^{(0)}$ is the uniform distribution over the simplex $\Delta^{|\mathbb{T}_i|}$, as well as $\boldsymbol{\mathcal{L}}_i^{(0)} = \boldsymbol{\mathcal{L}}_i^{(-1)} = \boldsymbol{0}$. By construction, for every player $i \in [[m]]$ the vector $\boldsymbol{X}_i^{(t)}$ is updated using (OMWU) under the sequence of losses $\boldsymbol{\mathcal{L}}^{(1)},\dots,\boldsymbol{\mathcal{L}}^{(T)}$, implying that for each $\mathcal{T} \in \mathbb{T}_i$, $\boldsymbol{X}_i^{(t_0+t+1)}[\mathcal{T}]$ is proportional to

$$\exp\left\{\eta\left(\boldsymbol{\mathcal{L}}_i^{(t_0-1)}[\mathcal{T}] - \sum_{s=0}^{t}\boldsymbol{\mathcal{L}}_i^{(t_0+s)}[\mathcal{T}] - \boldsymbol{\mathcal{L}}_i^{(t_0+t)}[\mathcal{T}]\right)\right\}\boldsymbol{X}_i^{(t_0)}[\mathcal{T}]. \tag{26}$$

For notational convenience, we let $\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}^{(t)} = \boldsymbol{\mathcal{L}}_i^{(t_0-1)} - \sum_{s=0}^{t-1}\boldsymbol{\mathcal{L}}_i^{(t_0+s)} - \boldsymbol{\mathcal{L}}_i^{(t_0+t-1)}$, for $0 \le t_0 \le T$ and $t \ge 0$. Moreover, for a vector $\boldsymbol{z} \in \mathbb{R}^{|\mathbb{T}_i|}$ we define the following function:

$$\phi_{t_0,\mathcal{T}}(\boldsymbol{z}) = \frac{\exp\{\boldsymbol{z}[\mathcal{T}]\}}{\sum_{\mathcal{T}'\in\mathbb{T}_i}\boldsymbol{X}_i^{(t_0)}[\mathcal{T}']\exp\{\boldsymbol{z}[\mathcal{T}']\}}. \tag{27}$$

Equipped with this notation, we can equivalently write (26) as $\boldsymbol{X}_i^{(t_0+t)}[\mathcal{T}] = \boldsymbol{X}_i^{(t_0)}[\mathcal{T}]\phi_{t_0,\mathcal{T}}\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}^{(t)}\right)$ for $t \ge 1$. In particular, this implies that for any $i' \in [[m]]$ and $\mathcal{T}_{i'} \in \mathbb{T}_{i'}$,

$$\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t_0+t)}[\mathcal{T}_{i'}] = \prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t_0)}[\mathcal{T}_{i'}]\phi_{t_0,\mathcal{T}_{i'}}\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{i',t_0}^{(t)}\right). \tag{28}$$

Before we proceed with the analysis, let us introduce the *shift operator*. Specifically, for a sequence of vectors $\boldsymbol{z} = (\boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(T)})$, the s-shifted sequence, denoted with $E_s\boldsymbol{z}$, is such that $(E_s\boldsymbol{z})^{(t)} = \boldsymbol{z}^{(t+s)}$, for $1 \le t \le T-s$. With this notation, we observe that

$$\left(D_1\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}\right)^{(t)} = \boldsymbol{\mathcal{L}}_i^{(t_0+t-1)} - 2\boldsymbol{\mathcal{L}}_i^{(t_0+t)} = \boldsymbol{\mathcal{L}}_i^{(t_0+t-1)} - 2(E_1\boldsymbol{\mathcal{L}}_i)^{(t_0+t-1)},$$

which in particular implies that for any $h' \ge 1$,

$$\left(D_{h'}\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}\right)^{(t)} = (D_{h'-1}\boldsymbol{\mathcal{L}}_i)^{(t_0+t-1)} - 2(E_1 D_{h'-1}\boldsymbol{\mathcal{L}}_i)^{(t_0+t-1)}. \tag{29}$$

Next we proceed with bounding $\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}^{(t)}$, for any $0 \le t_0 \le T$ and $t \ge 0$. In particular, for a fixed $\mathcal{T} \in \mathbb{T}_i$, we know that $\boldsymbol{\mathcal{L}}_i^{(t)}[\mathcal{T}] = \sum_{(j,k)\in E(\mathcal{T})}\boldsymbol{x}_i^{(t)}[j](\boldsymbol{\ell}_i^{(t)}[k] - \boldsymbol{\ell}_i^{(t)}[j])$. Thus, given that $\boldsymbol{\ell}_i^{(t)}[j] \in [0,1]$, for any $j \in [[n_i]]$ and $t \ge 0$, as follows from (20), we can conclude that there exists $\mathcal{T} \in \mathbb{T}_i$ such that

$$\begin{aligned}\|\boldsymbol{\mathcal{L}}_i^{(t)}\|_\infty &= \left|\sum_{(j,k)\in E(\mathcal{T})}\boldsymbol{x}_i^{(t)}[j](\boldsymbol{\ell}_i^{(t)}[k] - \boldsymbol{\ell}_i^{(t)}[j])\right| \\ &\le \sum_{(j,k)\in E(\mathcal{T})}\boldsymbol{x}_i^{(t)}[j]\left|\boldsymbol{\ell}_i^{(t)}[k] - \boldsymbol{\ell}_i^{(t)}[j]\right| \le 1,\end{aligned} \tag{30}$$

where we used that $\left|\boldsymbol{\ell}_i^{(t)}[k] - \boldsymbol{\ell}_i^{(t)}[j]\right| \le 1$, as well as the fact that $\mathcal{T}$ is a directed tree, implying that every node has at most one outgoing (directed) edge (recall Definition 2.3). Along with the triangle inequality, this implies that

$$\|\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}^{(t)}\|_\infty = \left\|\boldsymbol{\mathcal{L}}_i^{(t_0-1)} - \sum_{s=0}^{t-1}\boldsymbol{\mathcal{L}}_i^{(t_0+s)} - \boldsymbol{\mathcal{L}}_i^{(t_0+t-1)}\right\|_\infty \le (t+2). \tag{31}$$

Before we proceed with the next claim, it will be useful to introduce the sequence $\mathfrak{L}^{(t)}$, defined as

$$\mathfrak{L}^{(t)} := \left(\eta\bar{\boldsymbol{\mathcal{L}}}_{1,t_0}^{(t)}, \eta\bar{\boldsymbol{\mathcal{L}}}_{2,t_0}^{(t)},\dots,\eta\bar{\boldsymbol{\mathcal{L}}}_{m,t_0}^{(t)}\right).$$

Lemma B.2. *Let $\alpha \in (0,1)$ be such that $\left\|\left(D_{h'}\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}\right)\right)^{(t)}\right\|_\infty \le \frac{1}{B_1}\alpha^{h'}(h')^{B_0 h'}$, for all $i \in [[m]]$, $0 \le h' \le h$, and $t \in [[h+1-h']]$, where $B_1 = 12e^5 m$ and $B_0 \ge 3$. Then, for any $0 \le t_0 \in T-h-1$ and $\mathcal{T}_1 \in \mathbb{T}_1,\dots,\mathcal{T}_m \in \mathbb{T}_m$, it holds that*

$$\left|\left(D_h\left(\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]\right)\right)^{(t_0+1)}\right| \le \alpha^h h^{B_0 h+1}\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t_0)}(\mathcal{T}_{i'}). \tag{32}$$

Proof. We will apply Lemma A.1 with $n := \sum_{i'\in[[m]]}|\mathbb{T}_{i'}|$, time horizon $h+1$, the sequence $\mathfrak{L}^{(t)}$, and the function ϕ that maps the concatenation of $\boldsymbol{z}_{i'} \in \mathbb{R}^{|\mathbb{T}_{i'}|}$, for all $i' \in [[m]]$, to the function $\prod_{i'}\phi_{t_0,\mathcal{T}_{i'}}(\boldsymbol{z}_{i'})$, where $\phi_{t_0,\mathcal{T}_{i'}}$ was defined in (27); that is,

$$\phi_{t_0}(\boldsymbol{z}_1,\dots,\boldsymbol{z}_m) := \prod_{i'\in[[m]]}\phi_{t_0,\mathcal{T}_{i'}}(\boldsymbol{z}_{i'}). \tag{33}$$

Let us verify the conditions of Lemma A.1. First, [13, Lemma B.6] implies that the function ϕ_{t_0} is $(1, e^3 m)$-bounded (in the sense of Lemma A.1), and $B_1 = 12e^5 m \ge 2e^2(e^3 m)$. Moreover, [13, Lemma B.7] implies that each function $\phi_{t_0,\mathcal{T}_{i'}}$ has radius of convergence—with respect to the origin $\boldsymbol{0}$—greater than 1, and hence, the radius of convergence of ϕ_{t_0} is also greater than 1. We also know, by assumption, that $\left\|(D_{h'}\mathfrak{L})^{(t)}\right\|_\infty \le \frac{1}{B_1}\alpha^{h'}(h')^{B_0 h'}$, for all $0 \le h' \le h$ and $t \in [[h+1-h']]$. As a result, Lemma A.1 implies that

$$\left|(D_h(\phi\circ\mathfrak{L}))^{(1)}\right| \le \frac{12e^5 m}{B_1}\alpha^h h^{B_0 h+1}. \tag{34}$$

Finally, we have that

$$\frac{1}{\prod_{i'\in[[m]]} \boldsymbol{X}_{i'}^{(t_0)}[\mathcal{T}_{i'}]}\left|\left(D_h\left(\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]\right)\right)^{(t_0+1)}\right|$$

$$=\left|\left(D_h\left(\prod_{i'\in[[m]]}\left(\phi_{t_0,\mathcal{T}_{i'}}\circ\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{i',t_0}\right)\right)\right)\right)^{(1)}\right| \tag{35}$$

$$=\left|\left(D_h\left(\phi_{t_0}\circ\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{1,t_0},\ldots,\eta\bar{\boldsymbol{\mathcal{L}}}_{m,t_0}\right)\right)\right)^{(1)}\right| \tag{36}$$

$$=\left|D_h\left(\phi_{t_0}\circ\mathfrak{L}\right)^{(1)}\right|$$

$$\le\frac{12e^5m}{B_1}\alpha^h h^{B_0h+1}=\alpha^h h^{B_0h+1}, \tag{37}$$

where (35) follows from Equation (28); (36) simply uses the definition of ϕ_{t_0} in (33); and (37) follows from (34) (which in turn is a consequence of Lemma A.1), as well as the fact that $B_1=12e^5m$. □

Lemma B.3. *Fix some $1\le h\le H$, a parameter $\alpha\in(0,1)$, and assume that the learning rate η is such that $\eta\le\min\left\{\frac{\alpha}{36e^5m},\frac{1}{12e^5(H+3)m}\right\}$. Moreover, assume that for all $0\le h'<h$, $t\le T-h'$, and $i\in[[m]]$, it holds that $\|(D_{h'}\boldsymbol{\mathcal{L}}_i)^{(t)}\|_\infty\le\alpha^{h'}(h'+1)^{B_0(h'+1)}$. Then, for all $t\in[[T-h]]$ and $i\in[[m]]$,*

$$\left\|(D_h\boldsymbol{\mathcal{L}}_i)^{(t)}\right\|_\infty\le\alpha^h h^{B_0h+1}. \tag{38}$$

Proof. Let us set $B_1:=12e^5m$, so that $\eta\le\min\left\{\frac{\alpha}{3B_1},\frac{1}{B_1(H+3)}\right\}$. We know from (31) that for $t+2\le h+3$ it follows that

$$\left\|D_0\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}\right)^{(t)}\right\|_\infty=\eta\left\|\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}^{(t)}\right\|\le\eta(t+2)\le\frac{1}{B_1}, \tag{39}$$

where we used the fact that $\eta\le1/(B_1(H+3))$. Next, for $1\le h'\le h$,

$$\left\|\left(D_{h'}\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}\right)\right)^{(t)}\right\|_\infty\le\eta\left\|(D_{h'-1}\boldsymbol{\mathcal{L}}_i)^{(t_0+t-1)}\right\|_\infty$$
$$+2\eta\left\|(D_{h'-1}\boldsymbol{\mathcal{L}}_i)^{(t_0+t)}\right\|_\infty \tag{40}$$

$$\le3\eta\alpha^{h'-1}(h')^{B_0h'} \tag{41}$$

$$\le\frac{1}{B_1}\alpha^{h'}(h')^{B_0h'}, \tag{42}$$

where (40) follows from (29) and the triangle inequality; (41) is a consequence of the assumption in the claim; and (42) uses the fact that $\eta\le\frac{\alpha}{3B_1}$. Next, given that $\boldsymbol{\mathcal{L}}_i^{(t)}[\mathcal{T}]=\sum_{(j,k)\in E(\mathcal{T})}\boldsymbol{x}_i^{(t)}[j](\boldsymbol{\ell}_i^{(t)}[k]-\boldsymbol{\ell}_i^{(t)}[j])$, we can infer that

$$\boldsymbol{\mathcal{L}}_i^{(t)}=\sum_{a_{i'}\in[[n_{i'}]]}\hat{\Lambda}(a_1,\ldots,a_m)\prod_{i'\in[[m]]}\boldsymbol{x}_{i'}[a_{i'}]$$
$$=\sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\tilde{\Lambda}(\mathcal{T}_1,\ldots,\mathcal{T}_m)\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}], \tag{43}$$

where $\hat{\Lambda}$ and $\tilde{\Lambda}$ are functions such that $\hat{\Lambda}(\cdot)\in[-1,1]$ and $\tilde{\Lambda}(\cdot)\in[-1,1]$. More precisely, (43) is obtained from Theorem 3.1. As a result, similarly to (25) we may conclude that

$$\left|(D_h\boldsymbol{\mathcal{L}}_i)^{(t)}[\mathcal{T}]\right|\le\sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\left|\left(D_h\left(\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]\right)\right)^{(t)}\right|. \tag{44}$$

The next step is to invoke Lemma B.2 in order to bound the induced term. Specifically, by (39) and (42) we see that its conditions are met, from which we can conclude that for $t\in[[T-h]]$

$$\left\|(D_h\boldsymbol{\mathcal{L}}_i)^{(t)}\right\|_\infty\le\sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\left|\left(D_h\left(\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]\right)\right)^{(t)}\right| \tag{45}$$

$$\le\alpha^h h^{B_0h+1}\sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t-1)}[\mathcal{T}_{i'}] \tag{46}$$

$$=\alpha^h h^{B_0h+1}, \tag{47}$$

where (45) follows from (43); (46) is an immediate application of Lemma B.2; and (47) follows from the fact that each $\boldsymbol{X}_{i'}$ is a probability distribution over the space of directed trees $\mathbb{T}_{i'}$, for all $i'\in[[m]]$, and as such the induced product distribution normalizes to 1. □

Finally, it follows from (30) that $\left\|(D_0\boldsymbol{\mathcal{L}}_i)^{(t)}\right\|_\infty\le1=\alpha^0 1^{B_0}1$, for all $i\in[[m]]$. Thus, we can inductively invoke Lemma B.3 for $B_0=3$ to infer that for all $0\le h\le H$, $i\in[[m]]$, and $t\in[[T-h]]$ that $\left\|(D_h\boldsymbol{\mathcal{L}}_i)^{(t)}\right\|_\infty\le\alpha^h h^{3h+1}$, as long as $\eta\le\frac{\alpha}{36e^5m}$ and $\alpha\le1/(H+3)$. Finally, following the argument given in the proof of Lemma B.3, we obtain that for any $0\le h'\le h$,

$$\left\|\left(D_{h'}\left(\eta\bar{\boldsymbol{\mathcal{L}}}_{i,t_0}\right)\right)^{(t)}\right\|_\infty\le\frac{1}{B_1}\alpha^{h'}(h')^{3h'}.$$

Thus, we can invoke Lemma B.2 to conclude that for any $0\le t_0\le T-h-1$ and $\mathcal{T}_1\in\mathbb{T}_1,\ldots,\mathcal{T}_m\in\mathbb{T}_m$,

$$\left|\left(D_h\left(\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]\right)\right)^{(t_0+1)}\right|\le\alpha^h h^{3h+1}\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t_0)}(\mathcal{T}_{i'}).$$

As a result, plugging-in this bound into (25) we obtain that for $t\in[[T-h]]$,

$$\left\|(D_h\boldsymbol{L}_i)^{(t)}\right\|_\infty\le\sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\left|\left(D_h\left(\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}[\mathcal{T}_{i'}]\right)\right)^{(t)}\right|$$
$$\le\alpha^h h^{3h+1}\sum_{\mathcal{T}_{i'}\in\mathbb{T}_{i'}}\prod_{i'\in[[m]]}\boldsymbol{X}_{i'}^{(t-1)}[\mathcal{T}_{i'}],$$
$$=\alpha^h h^{3h+1}.$$

□

The final technical ingredient is the following lemma, which can be shown by applying Lemma A.2 using the smoothness of the losses established by Lemma 3.4.

Lemma B.4. *There are universal constants $C,C'\ge1$ so that for a time horizon $T\ge4$ and $H:=\lceil\log T\rceil$, if all players employ SL-OMWU with learning rate η such that $1/T\le\eta\le\frac{1}{CmH^4}$, then,*

$$\sum_{t=1}^{T}\mathrm{Var}_{\boldsymbol{p}_i^{(t)}}\left(\boldsymbol{L}_i^{(t)}-\boldsymbol{L}_i^{(t-1)}\right)\le\frac{1}{2}\mathrm{Var}_{\boldsymbol{p}_i^{(t)}}\left(\boldsymbol{L}_i^{(t-1)}\right)+C'H^5, \tag{48}$$

for any $i\in[[m]]$.

C ANALYSIS OF BLUM-MANSOUR

To prove Theorem 1.3, we start with a regret bound analogous to that of [13] for the swap regret setting. The BM algorithm is composed of n copies of a no-external-regret algorithm (such as OMWU), and thus we can achieve the following regret bound using similar techniques as [13]. For any swap function $\phi : [[n]] \to [[n]]$,

$$\sum_{t=1}^{T} \langle \boldsymbol{x}^{(t)}, \boldsymbol{\ell}^{(t)} \rangle - \sum_{t=1}^{T} \sum_{g=1}^{n} \boldsymbol{x}^{(t)}[g] \cdot \boldsymbol{\ell}^{(t)}[\phi(g)]$$
$$\leq \sum_{t=1}^{T} \sum_{g=1}^{n} \left(\frac{\eta}{2} + C\eta^2 \right) \mathrm{Var}_{\mathbf{Q}^{(t)}[g,\cdot]} \left(\boldsymbol{x}^{(t)}[g] \cdot \boldsymbol{\ell}^{(t)} - \boldsymbol{x}^{(t-1)}[g] \cdot \boldsymbol{\ell}^{(t-1)} \right)$$
$$- \sum_{t=1}^{T} \sum_{g=1}^{n} \frac{(1 - C\eta)\eta}{2} \cdot \mathrm{Var}_{\mathbf{Q}^{(t)}[g,\cdot]} \left(\boldsymbol{x}^{(t-1)}[g] \cdot \boldsymbol{\ell}^{(t-1)} \right) + \frac{n \log n}{\eta}.$$

Thus, proving Theorem 1.3 boils down to the following lemma, which will be proved in the sequel of the appendix.

Lemma C.1. *Suppose all players play according to BM-OMWU with step size η satifying $1/T \leq \eta \leq \frac{1}{C \cdot mn^3 \log^4(T)}$ for a sufficiently large constant C. Then for any player $i \in [[m]]$ and any $g \in [[n_i]]$, the overall losses for player i: $\boldsymbol{\ell}_i^{(1)}, \ldots, \boldsymbol{\ell}_i^{(T)} \in \mathbb{R}^n$, the probability player i places on action g: $\boldsymbol{x}^{(1)}[g], \ldots, \boldsymbol{x}^{(T)}[g] \in \mathbb{R}$, and the strategy vectors output by player i's g^{th} instance of OMWU $\mathcal{R}_{\Delta,i,g}$: $\mathbf{Q}_i^{(1)}[g,\cdot], \ldots, \mathbf{Q}_i^{(T)}[g,\cdot] \in \Delta^n$ satisfy*

$$\sum_{t=1}^{T} \mathrm{Var}_{\mathbf{Q}^{(t)}[g,\cdot]} \left(\boldsymbol{x}^{(t)}[g] \cdot \boldsymbol{\ell}^{(t)} - \boldsymbol{x}^{(t-1)}[g] \cdot \boldsymbol{\ell}^{(t-1)} \right)$$
$$\leq \frac{1}{2} \cdot \sum_{t=1}^{T} \mathrm{Var}_{\mathbf{Q}^{(t)}[g,\cdot]} \left(\boldsymbol{x}^{(t-1)}[g] \cdot \boldsymbol{\ell}^{(t-1)} \right) + O\left(\log^5(T) \right).$$

By combining the previous inequality with Lemma C.1, for $\eta \in \left[\frac{1}{T}, \frac{1}{C \cdot mn^3 \log^4(T)} \right]$, we obtain that

$$\mathrm{SwapReg}_i^T \leq \frac{n \log n}{\eta} + \frac{2n\eta}{3} O\left(\log^5(T) \right).$$

Hence, by setting $\eta = \frac{1}{C \cdot mn^3 \log^4(T)}$ we recover the bound stated in Theorem 1.3.

The rest of the appendix is devoted to the proof of Lemma C.1. There, the main technical tool is Lemma 4.1, which establishes higher-order smoothness for the iterates of BM-OMWU.

C.1 Technical Lemmas for BM-OMWU

We first reiterate the Markov chain tree theorem under a slightly different formulation, catering to the proof techniques of this section. We let $\mathbb{R}^{n \times n}$ (respectively, $\mathbb{C}^{n \times n}$) denote the space of real-valued (respectively, complex-valued) $n \times n$ matrices. For each $j \in [[n]]$, we introduce the following functions $\Phi_{\mathrm{MCT},j} : \mathbb{C}^{n \times n} \to \mathbb{C}$:

$$\Phi_{\mathrm{MCT},j}(Z) := \frac{\sum_{\mathcal{T} \in \mathbb{T}_j^n} \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{ab} \right)}{\sum_{j=1}^{n} \sum_{\mathcal{T} \in \mathbb{T}_j^n} \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{ab} \right)}.$$

Theorem C.2 (Markov chain tree theorem). *Let $\mathbf{Q} \in \mathbb{R}^{n \times n}$ be the transition matrix of a Markov chain so that $\mathbf{Q}[i,j] > 0$ for all $i, j \in [[n]]$. Then, the stationary distribution of $\mathbf{Q}$ is given by the vector $(\Phi_{\mathrm{MCT},1}(\ln \mathbf{Q}), \ldots, \Phi_{\mathrm{MCT},1}(\ln \mathbf{Q}))$, where $\ln \mathbf{Q}$ denotes the matrix whose (i,j) entry is $\ln \mathbf{Q}[i,j]$.*

Lemma C.3. *Fix any $Z^{(0)} \in \mathbb{R}^{n \times n}$. Consider any function of the form $\phi(Z) := \frac{\Phi_{\mathrm{MCT},j}(Z^{(0)} + Z)}{\Phi_{\mathrm{MCT},j}(Z^{(0)})}$. Then the sum of the Taylor series coefficients of order $k \geq 0$ of ϕ at $\mathbf{0}$ is bounded above by $30 \cdot (2n^3)^k$, and has radius of convergence greater than $1/n$.*

Proof. Fix any multi-index $\gamma \in \mathbb{Z}_{\geq 0}^{n \times n}$; we will bound $\frac{d^\gamma}{dZ^\gamma} \phi(\mathbf{0})$. Fix any $Z \in \mathbb{C}^{n \times n}$ so that $\|Z\|_\infty \leq \pi/(3n)$. Set $\zeta = Z^{(0)} + Z$. Note that, for any $\mathcal{T} \in \mathbb{T}_j^n$,

$$\left| \exp\left(\sum_{(a,b) \in E(\mathcal{T})} \zeta_{ab} \right) \right| = \left| \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{ab}^{(0)} + \sum_{(a,b) \in E(\mathcal{T})} (\zeta_{ab} - Z_{ab}^{(0)}) \right) \right|$$
$$\leq \exp(\pi/3) \cdot \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{ab}^{(0)} \right),$$

where we used the fact that $\left| \sum_{(a,b) \in E(\mathcal{T})} (\zeta_{ab} - Z_{ab}^{(0)}) \right| \leq \pi/3$ and that $Z^{(0)}$ is real-valued. Further, for any $a \in \mathbb{R}$ and $\zeta \in \mathbb{C}$ with $|\zeta| \leq \pi/3$, we have

$$\Re(\exp(a + \zeta)) = \exp(a) \cdot \Re(\zeta)$$
$$\geq \exp(a) \cdot \cos(\pi/3) \cdot \exp(-\pi/3) > \exp(a)/10.$$

Thus, for any $j' \in [[n]]$ and $\mathcal{T} \in \mathbb{T}_{j'}^n$, it holds that

$$\Re\left(\exp\left(\sum_{(a,b) \in E(\mathcal{T})} \zeta_{ab} \right) \right) \geq \frac{1}{10} \cdot \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{ab}^{(0)} \right)$$

Thus, since $Z^{(0)}$ is real-valued,

$$|\phi(Z)| = \frac{|\Phi_{\mathrm{MCT},j}(\zeta)|}{\Phi_{\mathrm{MCT},j}(Z^{(0)})}$$
$$\leq \frac{\sum_{j'=1}^{n} \sum_{\mathcal{T} \in \mathbb{T}_{j'}^n} \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{a,b}^{(0)} \right)}{\sum_{\mathcal{T} \in \mathbb{T}_j^n} \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{a,b}^{(0)} \right)} \cdot \frac{\sum_{\mathcal{T} \in \mathbb{T}_j^n} \left| \exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{ab} \right) \right|}{\sum_{j'=1}^{n} \sum_{\mathcal{T} \in \mathbb{T}_{j'}^n} \Re\left(\exp\left(\sum_{(a,b) \in E(\mathcal{T})} Z_{ab} \right) \right)}$$
$$= \exp(\pi/3) \cdot 10 < 30.$$

By the multivariate version of Cauchy's integral formula,

$$\left| \frac{d^\gamma}{dZ^\gamma} \phi(\mathbf{0}) \right| \left| \frac{\gamma!}{(2\pi i)^{n^2}} \int_{|Z_{11}| = \pi/(3n)} \cdots \int_{|Z_{nn}| = \pi/(3n)} \frac{\phi(Z)}{\prod_{j_1, j_2 \in [[n]]} (Z_{j_1 j_2})^{\gamma_{j_1 j_2} + 1}} d\zeta_{11} \cdots d\zeta_{nn} \right|$$
$$\leq 30 \cdot \gamma! \cdot (3n/\pi)^{|\gamma|}. \tag{49}$$

For any integer $k \geq 0$, the number of tuples $\gamma \in \mathbb{Z}^{n^2}_{\geq 0}$ with $|\gamma| = k$ is $\binom{k+n^2-1}{n^2-1} \leq (2n^2)^k$. Thus, if we write $\phi(Z) = \sum_{\gamma \in \mathbb{Z}^{n^2}_{\geq 0}} a_\gamma \cdot z^\gamma$, it follows that $a_\gamma = \frac{1}{\gamma!} \cdot \frac{d^\gamma}{dZ^\gamma}\phi(\mathbf{0})$, and so for each $k \geq 0$, $\sum_{\gamma:|\gamma|=k} |a_\gamma| \leq 30 \cdot (2n^2)^k \cdot (3n/\pi)^k \leq 30 \cdot (2n^3)^k$.

To see the lower bound on the radius of convergence, note that (49) gives that for each $\gamma \in \mathbb{Z}^{n^2}_{\geq 0}$, letting $k := |\gamma|$, we have $|a_\gamma|^{1/k} \leq 30^{1/k} \cdot (3n/\pi)$, which tends to $3n/\pi < n$ as $k \to \infty$. Thus, by the multivariate version of the Cauchy-Hadamard theorem, the radius of convergence of the power series of ϕ at $\mathbf{0}$ is greater than $1/n$. □

In the statement of Lemma C.4 below, the quantity 0^0 is interpreted as 1 (in particular, $(h')^{B_0 h'} = 1$ for $h' = 0$).

Lemma C.4. *Fix any $B_1 \geq 4e^2n^3$, $B_0 \geq 3$. Consider a sequence $\mathbf{Q}^{(0)}, \dots, \mathbf{Q}^{(h)} \in \mathbb{R}^{n \times n}$ of ergodic Markov chains, so that $\| \log \mathbf{Q}^{(0)} - \log \mathbf{Q}^{(t)} \|_\infty \leq \frac{1}{B_1}$ for $0 \leq t \leq h$. Suppose that for some $\alpha \in (0,1)$, for each $1 \leq h' \leq h$ and $0 \leq t \leq h - h'$, it holds that $\| (\mathrm{D}_{h'} (\ln \mathbf{Q}))^{(t)} \|_\infty \leq \frac{1}{B_1} \cdot \alpha^{h'} \cdot (h')^{B_0 h'}$. Then, if $p^{(0)}, \dots, p^{(h)} \in \Delta^n$ denotes the sequence of stationary distributions for $\mathbf{Q}^{(0)}, \dots, \mathbf{Q}^{(h)}$, it holds that, for any $j \in [[n]]$,*

$$\frac{\left|(\mathrm{D}_h\, p[j])^{(t)}\right|}{p^{(0)}[j]} \leq \frac{720n^3e^2}{B_1} \cdot \alpha^h \cdot h^{B_0 h+1}.$$

Proof. By the Markov chain tree theorem we have that, for each $j \in [[n]]$, $p^{(t)}[j] = \Phi_{\mathrm{MCT},j}(\ln \mathbf{Q}^{(t)})$. We now apply Lemma A.1 with $T = h+1$, $\mathbf{z}^{(t)} = \ln \mathbf{Q}^{(t-1)} - \ln \mathbf{Q}^{(0)}$ for $1 \leq t \leq h+1$, $R_1 = 1$, $R_2 = 2n^3$, and the values of B_0, B_1 given in the hypothesis of this lemma (Lemma C.4). Moreover, we set $\phi(Z) = \frac{\Phi_{\mathrm{MCT},j}((\ln \mathbf{Q}^{(0)}) + Z)}{\Phi_{\mathrm{MCT},j}(\ln \mathbf{Q}^{(0)})} = \frac{\Phi_{\mathrm{MCT},j}((\ln \mathbf{Q}^{(0)}) + Z)}{p^{(0)}[j]}$. Lemma C.3 gives that the function ϕ is $(30, 2n^3)$-bounded, and has radius of convergence greater than $1/n$ at the point $Z = \mathbf{0}$. Then the hypotheses of Lemma C.4 imply those of Lemma A.1, and Lemma A.1 gives that

$$\frac{\left|(\mathrm{D}_h\, p[j])^{(t)}\right|}{p^{(0)}[j]} = \frac{\left|\left(\mathrm{D}_h\, (\phi \circ (\ln \mathbf{Q} - \ln \mathbf{Q}^{(0)}))\right)^{(0)}\right|}{p^{(0)}[j]} \leq \frac{720n^3e^2}{B_1} \cdot \alpha^h \cdot h^{B_0 h+1}.$$

(Here we have used that $\phi(\ln \mathbf{Q}^{(t)} - \ln \mathbf{Q}^{(0)}) = \frac{\Phi_{\mathrm{MCT},j}(\ln \mathbf{Q}^{(t)})}{p^{(0)}[j]} = p^{(t)}[j]/p^{(0)}[j]$.) □

Lemma C.5. *For $n \in \mathbb{N}$, let $\xi_1, \dots, \xi_n \geq 0$ so that $\xi_1 + \cdots + \xi_n = 1$. For any $j \in [[n]]$, the function*

$$\phi_j((z_1, \dots, z_n)) = \frac{\exp(z_j)}{\sum_{k=1}^n \xi_k \cdot \exp(z_k)}$$

satisfies, for any $z \in \mathbb{R}^n$ with $\|z\|_\infty \leq 1/4$,

$$|\log \phi_j(z)| \leq \|z\|_\infty \leq 6\|z\|_\infty.$$

Proof. For $0 \leq x \leq 1$, we have $1 + x \leq \exp(x) \leq 1 + 2x$. Then, for $\|z\|_\infty \leq 1/2$,

$$\phi_j(z) \leq \frac{1 + 2z_j}{\sum_{k=1}^n \xi_k \cdot (1 + z_k)} \leq \frac{1 + 2\|z\|_\infty}{1 - \|z\|_\infty} \leq (1 + 2\|z\|_\infty)^2$$

and

$$\phi_j(z) \geq \frac{1 + z_j}{\sum_{k=1}^n \xi_k \cdot (1 + 2z_k)} \geq \frac{1 - \|z\|_\infty}{1 + 2\|z\|_\infty} \geq (1 - \|z\|_\infty)(1 - 2\|z\|_\infty).$$

Thus, for $\|z\|_\infty \leq 1/4$,

$$-6\|z\|_\infty \leq \log(1 - \|z\|_\infty) + \log(1 - 2\|z\|_\infty) \leq \log \phi_j(z) \leq 2\log(1 + 2\|z\|_\infty) \leq 4\|z\|_\infty.$$ □

Lemma C.6. *For $n \in \mathbb{N}$, let $\xi_1, \dots, \xi_n \geq 0$ such that $\xi_1 + \cdots + \xi_n = 1$. For each $j \in [[n]]$, define $\phi_j : \mathbb{R}^n \to \mathbb{R}$ to be the function*

$$\phi_j((z_1, \dots, z_n)) = \frac{\xi_j \exp(z_j)}{\sum_{k=1}^n \xi_k \cdot \exp(z_k)}$$

and let $P_{\log \phi_j}(z) = \sum_{\gamma \in \mathbb{Z}^n_{\geq 0}} a_{j,\gamma} \cdot z^\gamma$ denote the Taylor series of $\log \phi_j$. Then for any $j \in [[n]]$ and any integer $k \geq 1$,

$$\sum_{\gamma \in \mathbb{Z}^n_{\geq 0}:\ |\gamma| = k} |a_{j,\gamma}| \leq e^k/k.$$

Proof. We have that

$$\frac{\partial \log \phi_j}{\partial z_t} = \frac{1}{\phi_j} \cdot \frac{\partial \phi_j}{\partial z_t} = \begin{cases} -\phi_t & \text{if } t \neq j \\ 1 - \phi_j & \text{if } t = j \end{cases}$$

and so,

$$\sum_{\gamma \in \mathbb{Z}^n_{\geq 0}:\ |\gamma| = k} |a_{j,\gamma}| = \frac{1}{k!} \sum_{t \in [[n]]^k} \left| \frac{\partial^k \log \phi_j(\mathbf{0})}{\partial z_{t_1} \partial z_{t_2} \cdots \partial z_{t_k}} \right|$$
$$= \frac{1}{k!} \sum_{t_1 \in [[n]]} \sum_{t_{-1} \in [[n]]^{k-1}} \left| \frac{\partial^{k-1}}{\partial z_{t_2} \cdots \partial z_{t_k}} \left(\phi_{t_1} - \mathbf{1}[t_1 = j]\right)(\mathbf{0}) \right|.$$

For $k = 1$,

$$\frac{1}{k!} \sum_{t_1 \in [[n]]} \left|\phi_{t_1}(\mathbf{0}) - \mathbf{1}[t_1 = j]\right| \leq 1 + \sum_{t_1 \in [[n]]} \xi_{t_1} = 2.$$

For $k \geq 2$, $\mathbf{1}[t_1 = j]$ will be removed by the derivative; hence,

$$\frac{1}{k!} \sum_{t_1 \in [[n]]} \sum_{t_{-1} \in [[n]]^{k-1}} \left| \frac{\partial^{k-1} \phi_{t_1}(\mathbf{0})}{\partial z_{t_2} \cdots \partial z_{t_k}} \right| \leq \frac{1}{k!} \sum_{t_1 \in [[n]]} \left((k-1)! \cdot \xi_{t_1} e^k\right) \tag{50}$$
$$= \frac{1}{k!} \cdot (k-1)! \cdot e^k \cdot \sum_{t_1 \in [[n]]} \xi_{t_1} = e^k/k,$$

where (50) comes from the following lemma due to [13].

Lemma C.7 ([13]). *For $n \in \mathbb{N}$, let $\xi_1, \dots, \xi_n \geq 0$ such that $\xi_1 + \cdots + \xi_n = 1$. For each $j \in [[n]]$, define $\phi_j : \mathbb{R}^n \to \mathbb{R}$ to be the function*

$$\phi_j((z_1, \dots, z_n)) = \frac{\xi_j \exp(z_j)}{\sum_{k=1}^n \xi_k \cdot \exp(z_k)}$$

and let $P_{\phi_j}(z) = \sum_{\gamma \in \mathbb{Z}^n_{\geq 0}} a_{j,\gamma} \cdot z^\gamma$ denote the Taylor series of ϕ_j. Then for any $j \in [[n]]$ and any integer $k \geq 1$,

$$\sum_{\gamma \in \mathbb{Z}^n_{\geq 0}:\ |\gamma| = k} |a_{j,\gamma}| \leq \xi_j e^{k+1}.$$

□

C.2 Proof of Higher-Order Smoothness

Lemma 4.1 (Detailed). *Fix a parameter $\alpha \in \left(0, \frac{1}{H+3}\right)$. If all players follow BM-OMWU updates with step size $\eta \le \frac{1}{311040e^7H^2n_i^3}$, then for any player $i \in [[m]]$, integer h satisfying $0 \le h \le H$, time step $t \in [[T-h]]$, and $g \in [[n_i]]$, it holds that*

$$\| \left(\mathrm{D}_h \left(\boldsymbol{x}_i[g] \cdot \boldsymbol{\ell}_i\right)\right)^{(t)} \|_\infty \le \alpha^h \cdot h^{3h+1}.$$

Proof. We prove this inductively, showing that, for all $i \in [[m]], 0 \le h \le H, t \in [[T-h]], g \in [[n_i]]$, and $B_0 \ge 3$

$$\left\|\left(\mathrm{D}_h\, \boldsymbol{\ell}_i\right)^{(t)}\right\|_\infty \le \alpha^h \cdot h^{B_0h+1}; \tag{51}$$

$$\left\|\left(\mathrm{D}_h\, \boldsymbol{x}_i[g] \cdot \boldsymbol{\ell}_i\right)^{(t)}\right\|_\infty \le \alpha^h \cdot h^{B_0h+1}; \tag{52}$$

$$\left\|\left(\mathrm{D}_h\, \boldsymbol{x}_i\right)^{(t)}\right\|_\infty \le \alpha^h \cdot h^{B_0h+1}. \tag{53}$$

The base case of $h = 0$ is evident from the fact that losses and strategy probabilities are in $[0,1]$, and therefore $\|\boldsymbol{\ell}_i^{(t)}\|_\infty, \|\boldsymbol{x}_i^{(t)}\|_\infty \le 1$. So, proving the following inductive statement is sufficient to prove the lemma. Assume (51), (52), (53) hold for all h' satisfying $1 \le h' < h$. Then, they hold for h as well.

First, notice that for any agent $i \in [[m]]$, any OMWU instance $\mathcal{R}_{\Delta,i,g}$ of agent i with $g \in [n_i]$, any $t_0 \in \{0,1,\dots,T\}$, and any $t \ge 0$, by the definition (OMWU) of the OMWU updates, it holds that, for each $j \in [n_i]$,

$$\mathbf{Q}_i^{(t_0+t+1)}[g,j] = \frac{\mathbf{Q}_i^{(t_0)}[g,j] \cdot \exp\left(\eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,j]\right)}{\sum_{k=1}^{n_i} \mathbf{Q}_i^{(t_0)}[g,k] \cdot \exp\left(\eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,k]\right)},$$

where $\mathbf{Q}_i^t[g,j]$ denotes the weight placed on action j by algorithm $\mathcal{R}_{\Delta,i,g}$ at time t, and $\mathbf{L}_{i,t_0}^{(t)}[g,j]$ is defined

$$\mathbf{L}_{i,t_0}^{(t)}[g,j] := \boldsymbol{x}_i^{(t_0-1)}[g]\boldsymbol{\ell}_i^{(t_0-1)}[j] - \sum_{s=0}^{t-1} \boldsymbol{x}_i^{(t_0+s)}[g]\boldsymbol{\ell}_i^{(t_0+s)}[j] - \boldsymbol{x}_i^{(t_0+t-1)}[g]\boldsymbol{\ell}_i^{(t_0+t-1)}[j].$$

We can define edge values $\boldsymbol{\ell}_i^{(0)}, \boldsymbol{\ell}_i^{(-1)}, \mathbf{Q}_i^{(0)}$ to ensure that the above equation holds even for $t_0 \in \{0,1\}$. Now, for any $g, j \in [[n_i]]$, any integer t_0 satisfying $0 \le t_0 \le T$; and any integer $t \ge 0$, let us define Also, for a vector $z = (z[1],\dots,z[n_i]) \in \mathbb{R}^{n_i}$ and indices $g, j \in [[n_i]]$, define

$$\phi_{t_0,g,j}(z) := \frac{\exp(z[j])}{\sum_{k=1}^{n_i} \mathbf{Q}_i^{(t_0)}[g,k] \cdot \exp(z[k])}, \tag{54}$$

so that

$$\mathbf{Q}_i^{(t_0+t)}[g,j] = \mathbf{Q}_i^{(t_0)}[g,j] \cdot \phi_{t_0,g,j}(\eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,\cdot]) \tag{55}$$

for $t \ge 1$, where $\mathbf{L}_{i,t_0}^{(t)}[g,\cdot]$ denotes the vector $(\mathbf{L}_{i,t_0}^{(t)}[g,1],\dots,\mathbf{L}_{i,t_0}^{(t)}[g,n_i])$.

Next, note that, for all $g, j \in [[n_i]]$,

$$\left(\mathrm{D}_1\, \mathbf{L}_{i,t_0}[g,j]\right)^{(t)} = \boldsymbol{x}_i^{(t_0+t-1)}[g]\boldsymbol{\ell}_i^{(t_0+t-1)}[j] - 2\boldsymbol{x}_i^{(t_0+t)}[g]\boldsymbol{\ell}_i^{(t_0+t)}[j]$$

and so, for any $1 \le h' \le h$,

$$\left|\left(\mathrm{D}_{h'}\left(\eta \cdot \mathbf{L}_{i,t_0}[g,j]\right)\right)^{(t)}\right| \le 3\eta \cdot \alpha^{h'-1} \cdot (h'-1)^{B_0h'}, \tag{56}$$

where (56) follows from the inductive hypothesis. Additionally, since $\left|\mathbf{L}_{i,t_0}[g,j]\right| \le t+2$ for all t_0, i, g, j, we have $\left|\eta \cdot \mathbf{L}_{i,t_0}[g,j]\right| \le \eta \cdot (t+2) \le 1$ for $\eta \le \frac{1}{t+2}$. By Lemma C.6, the function $z \mapsto \log \phi_{t_0,g,j}(z)$ is $(1,e)$-bounded, and so for each $1 \le h' \le h$ we may apply Lemma A.1 with $h = h'$, $z^{(t)} = \eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,\cdot]$ and $B_1 = \frac{1}{\eta} \cdot \min\left\{\frac{\alpha}{3}, \frac{1}{H+3}\right\} = \frac{\alpha}{3\eta}$. Thus, we can conclude, for all $1 \le h' \le h$, $t \le h+1$, and $g \in [[n_i]]$,

$$\left(\mathrm{D}_{h'} \log\left(\phi_{t_0,g,j}(\eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,\cdot])\right)\right)^{(t)} \le 36e^3H\eta \cdot \alpha^{h'} \cdot (h')^{B_0h'}. \tag{57}$$

Taking the logarithm on both sides of (55), we have

$$\log\left(\mathbf{Q}_i^{(t_0+t)}[g,j]\right) = \log \mathbf{Q}_i^{(t_0)}[g,j] + \log\left(\phi_{t_0,g,j}(\eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,\cdot])\right). \tag{58}$$

Let $\mathbf{Q}_i^{(t)} \in \mathbb{R}^{n_i \times n_i}$ be the matrix with entries $\mathbf{Q}_i^{(t)}[g,j]$ and $P_{i,t_0}^{(t)}$ be the matrix with entries $\phi_{t_0,g,j}\left(\eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,\cdot]\right)$, for $g, j \in [[n_i]]$. Then, for all $g \in [[n_i]]$, all $t_0 \in [[T-h]]$, and $t \in [[h]]$, we have

$$\begin{aligned}\boldsymbol{x}_i^{(t_0+t)}[g] &= \Phi_{\mathrm{MCT},g}\left(\log \mathbf{Q}_i^{(t_0+t)}\right)\\ &= \Phi_{\mathrm{MCT},g}\left(\log \mathbf{Q}_i^{(t_0)} + \log\left(P_{i,t_0}^{(t)}\right)\right).\end{aligned}$$

Next note that, by Lemma C.5 and $t \le H+1$,

$$\begin{aligned}\left\|\log \mathbf{Q}_i^{(t_0+t)} - \log \mathbf{Q}_i^{(t_0)}\right\|_\infty &\le \max_{g,j\in[n_i]} \left|\log \phi_{t_0,g,j}(\eta \cdot \mathbf{L}_{i,t_0}^{(t)}[g,\cdot])\right|\\ &\le 6\eta \cdot (t+2) \le 36e^3H\eta.\end{aligned} \tag{59}$$

We now apply Lemma C.4 with $\mathbf{Q}^{(t)} = P_{i,t_0}^{(t)}$, for $0 \le t \le h$, and $1/B_1 = 36e^3H\eta$. The preconditions of the lemma hold from (57). Then Lemma C.4 gives that for all $g \in [n_i]$,

$$\left|\left(\mathrm{D}_h\, \boldsymbol{x}_i[g]\right)^{(t_0+1)}\right| \le |\boldsymbol{x}_i^{(t_0)}[g]| \cdot 720n_i^3e^2 \cdot 36e^3H\eta \cdot \alpha^{h'} \cdot (h)^{B_0h+1}.$$

verifying the first of three desired inductive conclusions (53) as long as $\eta \le 1/(25920e^5Hn_i^3)$.

As a result, for $t \in [[T]]$,

$$\begin{aligned}\left|\left(\mathrm{D}_h\, \boldsymbol{\ell}_i\right)^{(t)}[a_i]\right| &= \left|\sum_{s=0}^{h}\binom{h}{s}(-1)^{h-s}\boldsymbol{\ell}_i^{(t+s)}[a_i]\right|\\ &= \left|\sum_{a_{i'}\in[n_{i'}],\ \forall i'\neq i} \Lambda_i(a_1,\dots,a_m)\sum_{s=0}^{h}\binom{h}{s}(-1)^{h-s}\cdot\prod_{i'\neq i}\boldsymbol{x}_{i'}^{(t+s)}[a_{i'}]\right|\\ &\le \sum_{a_{i'}\in[n_{i'}],\ \forall i'\neq i}\left|\sum_{s=0}^{h}\binom{h}{s}(-1)^{h-s}\cdot\prod_{i'\neq i}\boldsymbol{x}_{i'}^{(t+s)}[a_{i'}]\right|\\ &= \sum_{a_{i'}\in[n_{i'}],\ \forall i'\neq i}\left|\left(\mathrm{D}_h\left(\prod_{i'\neq i}\boldsymbol{x}_{i'}[a_{i'}]\right)\right)^{(t)}\right|\\ &\le 25920e^5Hn_i^3\eta\cdot\alpha^h\cdot(h)^{B_0h+1}\sum_{a_{i'}\in[n_{i'}],\ \forall i'\neq i}\left(\prod_{i'\neq i}\boldsymbol{x}_i^{(t)}[a_i']\right)\\ &\le 25920e^5Hn_i^3\eta\cdot\alpha^h\cdot(h)^{B_0h+1},\end{aligned}$$

verifying the second of three desired inductive conclusions (51) as long as $\eta \le 1/(25920e^5 H n_i^3)$.

Lastly, for $\eta \le \frac{1}{12e^2 H} \cdot \frac{1}{25920e^5 H n_i^3} = \frac{1}{311040e^7 H^2 n_i^3}$, we have now verified the inductive hypotheses:

$$\| \left(\mathrm{D}_{h'} \boldsymbol{\ell}_i\right)^{(t)} \|_\infty \le \frac{1}{2e^2} \alpha^{h'} \cdot (h')^{B_0 h'};$$

$$\| \left(\mathrm{D}_{h'} \boldsymbol{x}_i\right)^{(t)} \|_\infty \le \frac{1}{2e^2} \alpha^{h'} \cdot (h')^{B_0 h'},$$

for all h' up to and including h. Thus, we can apply Lemma A.1 with $n = 2$, $\phi(a, b) = ab$ (which is $(1, 1)$-bounded), $B_1 = 12e^2$, and the sequence $\boldsymbol{z}^{(t)} = (\boldsymbol{x}_i^{(t)}[g] \cdot \boldsymbol{\ell}_i^{(t)}[j])$. Therefore, for the product sequence $\boldsymbol{x}_i[g] \cdot \boldsymbol{\ell}_i[j]$, we have, for all $t \in [\![T - h + 1]\!]$,

$$\left| \left(\mathrm{D}_h \boldsymbol{x}_i[g] \cdot \boldsymbol{\ell}_i[j]\right)^{(t)} \right| \le \alpha^h \cdot (h)^{B_0 \cdot h + 1}, \tag{60}$$

verifying the final inductive conclusion (52). □

Finally, Lemma C.1 follows by applying Lemma A.2 using the smoothness of the sequence $\boldsymbol{x}_i[g] \cdot \boldsymbol{\ell}_i$ implied by Lemma 4.1.

REFERENCES

[1] Jacob Abernethy, Peter L Bartlett, and Elad Hazan. 2011. Blackwell Approachability and No-Regret Learning are Equivalent.. In *COLT*. 27–46.

[2] V. Anantharam and P. Tsoucas. 1989. A proof of the Markov chain tree theorem. *Statistics & Probability Letters* 8, 2 (1989), 189–192. https://doi.org/10.1016/0167-7152(89)90016-3

[3] Robert Aumann. 1974. Subjectivity and Correlation in Randomized Strategies. *Journal of Mathematical Economics* 1 (1974), 67–96. https://doi.org/10.1016/0304-4068(74)90037-8

[4] Yakov Babichenko and Aviad Rubinstein. 2017. Communication complexity of approximate Nash equilibria. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017*. ACM, 878–889. https://doi.org/10.1145/3055399.3055407

[5] David Blackwell. 1956. An analog of the minmax theorem for vector payoffs. *Pacific J. Math.* 6 (1956), 1–8. https://doi.org/pjm/1103044235

[6] Avrim Blum and Yishay Mansour. 2007. From External to Internal Regret. *J. Mach. Learn. Res.* 8 (2007), 1307–1324.

[7] Noam Brown and Tuomas Sandholm. 2017. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science* (Dec. 2017). https://doi.org/10.1126/science.aao1733

[8] Arthur Cayley. 1889. A theorem on trees. *Quart. J. Math.* 23 (1889), 376–378.

[9] Nicolo Cesa-Bianchi and Gabor Lugosi. 2006. *Prediction, learning, and games*. Cambridge University Press. https://doi.org/10.1017/CBO9780511546921

[10] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. 2009. Settling the Complexity of Computing Two-Player Nash Equilibria. *J. ACM* (2009). https://doi.org/10.1145/1516512.1516516

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

[12] Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. 2011. Near-optimal no-regret algorithms for zero-sum games. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. https://doi.org/10.1137/1.9781611973082.21

[13] Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. 2021. Near-Optimal No-Regret Learning in General Games. In *Advances in Neural Information Processing Systems*, Vol. 34. Curran Associates, Inc., 27604–27616.

[14] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. 2009. The complexity of computing a Nash equilibrium. *SIAM J. Comput.* 39, 1 (2009). https://doi.org/10.1137/070699652

[15] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. 2018. Training GANs with Optimism. In *6th International Conference on Learning Representations, ICLR 2018*. OpenReview.net.

[16] Constantinos Daskalakis and Ioannis Panageas. 2018. The Limit Points of (Optimistic) Gradient Descent in Min-Max Optimization. In *NeurIPS 2018*. 9256–9266.

[17] Constantinos Daskalakis and Ioannis Panageas. 2019. Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization. In *10th Innovations in Theoretical Computer Science Conference, ITCS 2019 (LIPIcs, Vol. 124)*, Avrim Blum (Ed.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 27:1–27:18. https://doi.org/10.4230/LIPIcs.ITCS.2019.27

[18] Kousha Etessami and Mihalis Yannakakis. 2007. On the Complexity of Nash Equilibria and Other Fixed Points (Extended Abstract). In *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS)*. 113–123. https://doi.org/10.1109/FOCS.2007.48

[19] Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. 2021. Simple Uncoupled No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. *CoRR* abs/2104.01520 (2021).

[20] Dean Foster and Rakesh Vohra. 1997. Calibrated Learning and Correlated Equilibrium. *Games and Economic Behavior* 21 (1997), 40–55. https://doi.org/10.1006/game.1997.0595

[21] Geoffrey J Gordon, Amy Greenwald, and Casey Marks. 2008. No-regret learning in convex games. In *Proceedings of the 25th international conference on Machine learning*. ACM, 360–367. https://doi.org/10.1145/1390156.1390202

[22] Amy Greenwald and Amir Jafari. 2003. A General Class of No-Regret Learning Algorithms and Game-Theoretic Equilibria. In *Conference on Learning Theory (COLT)*. Washington, D.C. https://doi.org/10.1007/978-3-540-45167-9_2

[23] Sergiu Hart and Andreu Mas-Colell. 2000. A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica* 68 (2000), 1127–1150. https://doi.org/10.1111/1468-0262.00153

[24] Shinji Ito. 2020. A Tight Lower Bound and Efficient Reduction for Swap Regret. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*.

[25] Pravesh K. Kothari and Ruta Mehta. 2018. Sum-of-Squares Meets Nash: Lower Bounds for Finding Any Equilibrium. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2018)*. Association for Computing Machinery, New York, NY, USA, 1241–1248. https://doi.org/10.1145/3188745.3188892

[26] Alex Kruckman, Amy Greenwald, and John R. Wicks. 2010. *An elementary proof of the Markov chain tree theorem*. Technical Report 10-04. Brown University.

[27] Nick Littlestone and M. K. Warmuth. 1994. The Weighted Majority Algorithm. *Information and Computation* 108, 2 (1994), 212–261. https://doi.org/10.1006/inco.1994.1009

[28] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. 2017. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* (May 2017). https://doi.org/10.1126/science.aam6960

[29] John Nash. 1950. Equilibrium points in N-person games. *Proceedings of the National Academy of Sciences* 36 (1950), 48–49. https://doi.org/10.1073/pnas.36.1.48

[30] Yurii Nesterov. 2005. Smooth Minimization of Non-Smooth Functions. *Mathematical Programming* 103 (2005). https://doi.org/10.1007/s10107-004-0552-5

[31] Christos H. Papadimitriou and Tim Roughgarden. 2008. Computing correlated equilibria in multi-player games. *J. ACM* 55, 3 (2008), 14:1–14:29. https://doi.org/10.1145/1379759.1379762

[32] Alexander Rakhlin and Karthik Sridharan. 2013. Online Learning with Predictable Sequences. In *Conference on Learning Theory*. 993–1019.

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

[34] Julia Robinson. 1951. An iterative method of solving a game. *Annals of Mathematics* 54 (1951), 296–301. https://doi.org/10.2307/1969530

[35] Tim Roughgarden. 2015. Intrinsic Robustness of the Price of Anarchy. *J. ACM* 62, 5 (2015), 32:1–32:42. https://doi.org/10.1145/2806883

[36] Aviad Rubinstein. 2016. Settling the Complexity of Computing Approximate Two-Player Nash Equilibria. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS*, Irit Dinur (Ed.). IEEE Computer Society, 258–265. https://doi.org/10.1109/FOCS.2016.35

[37] Shai Shalev-Shwartz. 2012. Online Learning and Online Convex Optimization. *Foundations and Trends in Machine Learning* 4, 2 (2012). https://doi.org/10.1561/2200000018

[38] Gilles Stoltz and Gábor Lugosi. 2005. Internal regret in on-line portfolio selection. *Machine Learning* 59, 1-2 (2005), 125–159. https://doi.org/10.1007/s10994-005-0465-4

[39] Gilles Stoltz and Gábor Lugosi. 2007. Learning correlated equilibria in games with compact sets of strategies. *Games Econ. Behav.* 59, 1 (2007), 187–208. https://doi.org/10.1016/j.geb.2006.04.007

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

[41] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. 2021. Linear Last-iterate Convergence in Constrained Saddle-point Optimization. In *ICLR*. OpenReview.net.