A LOGARITHMIC REGRET BOUND FOR OPTIMISTIC HEDGE IN GENERAL-SUM GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Can simple no-regret dynamics attain smaller regret in self-play than against worst adversaries? In n-player general-sum games with d_i actions for each player i, Daskalakis et al. (2021) proved an $O(n \log d_i \log^4 T)$ individual regret bound for Optimistic Hedge, which improves upon classical $O(\sqrt{T})$ adversarial regret bound. We show that Optimistic Hedge with a constant step size achieves $O(\sqrt{n} \log d_i \log T)$ individual external regret under expected loss-vector feedback. The time-averaged play consequently enjoys a coarse correlated equilibrium gap $O(\sqrt{n} \log d \log T/T)$, where $d = \max_i d_i$. The improvement comes from a larger admissible step size $\eta = \Theta(1/(\sqrt{n} \log T))$. Our analysis proves factorial bounds on high-order differences of probability-weighted pairwise loss gaps, then applies finite-difference interpolation in a fixed Euclidean norm. These estimates sharpen the analysis of Daskalakis et al. (2021) and yield a logarithmic regret bound.

1 INTRODUCTION

Game theory studies strategic interaction (von Neumann & Morgenstern, 1944; Nash, 1951). How can independent players learn to approach equilibrium (Fudenberg & Levine, 1998)? The same question arises in poker self-play (Brown & Sandholm, 2019), adversarial training (Goodfellow et al., 2014), and language-model alignment (Munos et al., 2024; Wu et al., 2026).

No-regret learning is a decentralized route. Players compete with their best fixed action in hindsight (Blackwell, 1956; Hannan, 1957). Sublinear individual regret drives average play to coarse correlated equilibrium (CCE) (Hart & Mas-Colell, 2001; Cesa-Bianchi & Lugosi, 2006). While adversarial minimax regret is tightly bounded as $\Theta(\sqrt{T})$ (Freund & Schapire, 1997; Cesa-Bianchi et al., 1997), self-play offers an additional structure: each player's losses evolve with the same update rule.

Optimistic methods exploit such structure and predict future losses from the past (Rakhlin & Sridharan, 2013a; Syrgkanis et al., 2015). For Optimistic Hedge, Daskalakis et al. (2021) proved $O(\log^4 T)$ individual regret, and subsequent work achieved logarithmic or constant regret by modifying the dynamics (Farina et al., 2022; Soleymani et al., 2025a;b; Abbadi et al., 2026). This line of recent progress leaves us with the following natural question:

Can plain Optimistic Hedge achieve logarithmic individual regret in general-sum games?

Contributions. We answer the question affirmatively. Our contributions are two-fold.

(i) **An improved high-order smoothness analysis.** Building on the high-order smoothness framework of Daskalakis et al. (2021), we construct a centered-logit recurrence map that transfers derivative bounds to sharper factorial bounds on high-order weighted-gap differences. Then, a direct finite-difference interpolation produces a larger admissible step size. Our smoothness analysis based on a recurrence map may be of independent interest for regret analysis in games.

(ii) **Improved regret bound.** For n players with action dimension d_i, we obtain $O(\sqrt{n} \log d_i \log T)$ individual regret; this improves upon the $O(n \log d_i \log^4 T)$ bound of Daskalakis et al. (2021). Our result suggests that Optimistic Hedge with a constant step size already suffices: neither modified dynamics nor an adaptive learning rate is necessary.

Sections 2 and 3 review prior work and the setup; Sections 4 and 5 present the results and analysis.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2 Related work

No-regret learning and games. No-regret learning connects individual performance guarantees to equilibrium behavior. Its classical foundations compare a learner's cumulative loss with that of the best fixed action in hindsight (Blackwell, 1956; Hannan, 1957), and algorithms such as Weighted Majority and Hedge achieve this guarantee against arbitrary loss sequences (Littlestone & Warmuth, 1994; Freund & Schapire, 1997). When all players learn in this way, their individual guarantees combine: averaging yields approximate Nash equilibria in two-player zero-sum games and coarse correlated equilibria in general-sum games (Freund & Schapire, 1999; Moulin & Vial, 1978; Hart & Mas-Colell, 2001). Controlling deviations that depend on a mediator's recommendation gives the stronger notion of correlated equilibrium, motivating connections with calibration, regret matching, and swap regret (Aumann, 1974; Foster & Vohra, 1997; Hart & Mas-Colell, 2000; Blum & Mansour, 2007). These connections make regret analysis a key quantitative tool for equilibrium computation: better regret bounds translate directly into faster equilibrium approximation.

Accelerated rates under self-play. Self-play offers a crucial structure that admits improved rates: each player's losses evolve with the same shared update rule. Optimistic methods exploit this structure by using past losses to predict future ones (Rakhlin & Sridharan, 2013a;b). Early logarithmic-regret guarantees in two-player zero-sum games demonstrated that faster learning is indeed possible under self-play (Daskalakis et al., 2011), and subsequent work developed increasingly sharp guarantees for optimistic dynamics in general-sum games (Syrgkanis et al., 2015; Chen & Peng, 2020). One of the key advances was the high-order smoothness analysis of Daskalakis et al. (2021), which established $O(n \log d_i \log^4 T)$ individual regret for Optimistic Hedge. A complementary direction sought faster rates by changing the learning rule. Lifting and adaptive learning speeds yield logarithmic external regret (Farina et al., 2022; Soleymani et al., 2025a;b), while implicit updates and higher-order predictions yield constant regret at every horizon (Liu et al., 2026; Abbadi et al., 2026). These results establish strong guarantees for modified dynamics, but leave open how far plain Optimistic Hedge can go. We show that Optimistic Hedge already can achieve logarithmic regret without any modifications.

The high-order smoothness framework. Daskalakis et al. (2021) pioneered a framework that analyzes the high-order smoothness of loss sequences under self-play and established the first polylogarithmic individual regret bound for Optimistic Hedge in general-sum games. Their analysis first bounds high-order finite differences of losses, obtained by repeatedly subtracting consecutive terms. They then transfer these bounds one order at a time to first differences—the algorithm's prediction errors—while accounting for changing action probabilities. We build on this framework and analyze probability-weighted differences between losses. Our sharper high-order difference bound and direct comparison of high- and first-order differences together yield logarithmic regret.

3 Preliminaries

We study repeated play in finite general-sum games within the classical framework of no-regret learning (Blackwell, 1956; Hannan, 1957; Fudenberg & Levine, 1998; Cesa-Bianchi & Lugosi, 2006).

Notation. For an integer $d \geq 1$, define $[d] := \{1, \dots, d\}$ and the simplex $\Delta_d := \{x \in \mathbb{R}^d_+ : \sum_a x_a = 1\}$. An n-player game has action sets $A_i = [d_i]$, joint action space $A := \prod_{j\in[n]} A_j$, opponents' action space $A_{-i} := \prod_{j\neq i} A_j$, and losses $\ell_i : A \to [0,1]$. Strategies $x_i \in \Delta_{d_i}$ form the profile $x = (x_j)_{j\in[n]}$, opponents' profile $x_{-i} = (x_j)_{j\neq i}$ and product distribution $\bigotimes_j x_j$ on A.

We also use $\ell_i : \prod_{j\neq i} \Delta_{d_j} \to [0,1]^{d_i}$ for the expected loss map. Unlike the scalar loss $\ell_i(a)$ at a pure action profile, the vector $\ell_i(x_{-i})$ is multilinear in the opponents' mixed strategies:

$$[\ell_i(x_{-i})]_{a_i} := \sum_{a_{-i}\in A_{-i}} \ell_i(a_i, a_{-i}) \prod_{j\neq i} x_{j,a_j}. \tag{1}$$

For a horizon $T \geq 1$, players simultaneously choose a strategy profile x^t at each round $t \in [T]$. Player i incurs expected loss $\langle x_i^t, \ell_i^t \rangle$ and observes the exact loss vector $\ell_i^t := \ell_i(x_{-i}^t)$, using only its own observed vectors to update. Feedback is noiseless, and no additional game structure is assumed.

Table 1: Individual external regret under simultaneous self-play and exact expected loss-vector feedback; $d = \max_i d_i$. The result of Chen & Peng (2020) assumes two players; the others allow n players. *Single loop*: closed-form updates without inner optimization or root finding, excluding feedback cost. The OFTRL/OMD check refers to their entropy-regularized Optimistic Hedge instantiation.

Reference	Algorithm	Single loop	Regret bound
Syrgkanis et al. (2015)	OFTRL / OMD	✓	$O\big(\sqrt{n}\log d\, T^{1/4}\big)$
Chen & Peng (2020)	Optimistic Hedge	✓	$O\Big(\log^{5/6} d\, T^{1/6}\Big)$
Daskalakis et al. (2021)	Optimistic Hedge	✓	$O\big(n\log d\log^4 T\big)$
Farina et al. (2022)	LRL-OFTRL	✗	$O(nd\log T)$
Soleymani et al. (2025a;b)	COMWU	✗	$O\big(n\log^2 d\log T\big)$
Liu et al. (2026)	ECHO-OFTRL	✗	$O\big(n^{21}\log^4 d\big)$
Abbadi et al. (2026)	HOOD	✗	$O\big(n^3\log^2 d\big)$
This work	Optimistic Hedge	✓	$O(\sqrt{n}\log d\log T)$

Definition 3.1 (External regret)**.** *Player i's external regret against the best fixed action in hindsight is*

$$\mathrm{Reg}_i(T) := \sum_{t=1}^{T} \langle x_i^t, \ell_i^t \rangle - \min_{a_i \in A_i} \sum_{t=1}^{T} \ell_{i,a_i}^t. \tag{2}$$

Regret $\mathrm{Reg}_i(T)$ may be negative, since adaptive play can outperform every fixed action. We seek bounds on each $\mathrm{Reg}_i(T)$, rather than only their sum.

While Nash equilibrium is one of the most widely used solution concepts in games (Nash, 1951), computing one is PPAD-complete even for two-player general-sum games (Daskalakis et al., 2009; Chen et al., 2009). We therefore consider a standard relaxation, coarse correlated equilibrium (CCE) (Moulin & Vial, 1978), which permits correlated plays.

Definition 3.2 (Coarse correlated equilibrium)**.** *For $\varepsilon \geq 0$, a distribution μ over A is an ε-coarse correlated equilibrium (ε-CCE) if*

$$\mathrm{Gap}_{\mathrm{CCE}}(\mu) := \max_{i\in[n]} \max_{a_i' \in A_i} \mathbb{E}_{a\sim\mu}[\ell_i(a) - \ell_i(a_i', a_{-i})] \leq \varepsilon. \tag{3}$$

No-regret learning provides a direct route to CCE: if every player has sublinear external regret, average play approaches the CCE set (Hart & Mas-Colell, 2001; Cesa-Bianchi & Lugosi, 2006). Specifically, define the time-averaged joint distribution $\widehat{\mu}_T := T^{-1}\sum_{t=1}^{T} \bigotimes_{j\in[n]} x_j^t$. It selects a common round uniformly, then samples from that round's product distribution. By multilinearity,

$$\mathrm{Gap}_{\mathrm{CCE}}(\widehat{\mu}_T) = T^{-1} \max_i \mathrm{Reg}_i(T). \tag{4}$$

The joint distribution $\widehat{\mu}_T$ averages products, not the individual strategies x_i^t.

Optimistic Hedge. Hedge is an instance of the follow-the-regularized-leader (FTRL) algorithm with negative entropy regularization (Freund & Schapire, 1997; 1999). Optimistic Hedge is a variant of Hedge that uses the most recent loss vector ℓ_i^t to predict ℓ_i^{t+1} (Rakhlin & Sridharan, 2013a;b).

Starting from the uniform strategy $x_i^1 = \mathbf{1}_{d_i}/d_i$ and initial prediction $\ell_i^0 = 0$, all players use a common constant step size $\eta > 0$, chosen before play. For $t \geq 1$, the variational update is

$$x_i^{t+1} = \arg\min_{x_i \in \Delta_{d_i}} \left\{ \left\langle x_i, \sum_{s=1}^{t} \ell_i^s + \ell_i^t \right\rangle + \frac{1}{\eta} \sum_{a\in A_i} x_{i,a} \log x_{i,a} \right\}. \tag{5}$$

The entropy regularizer in Equation (5) admits an explicit solution. Eliminating cumulative losses between consecutive rounds gives

$$x_{i,a}^{t+1} = \frac{x_{i,a}^t \exp\{-\eta(2\ell_{i,a}^t - \ell_{i,a}^{t-1})\}}{\sum_{b\in A_i} x_{i,b}^t \exp\{-\eta(2\ell_{i,b}^t - \ell_{i,b}^{t-1})\}}, \qquad a \in A_i. \tag{6}$$

With suitable step sizes, Optimistic Hedge improves the adversarial $O(\sqrt{T})$ regret bound to $O(T^{1/4})$ (Syrgkanis et al., 2015) and $O(\log^4 T)$ (Daskalakis et al., 2021). Our main result in the next section sharpens these guarantees to $O(\log T)$.

4 Main result

The following result shows that plain Optimistic Hedge with a horizon-tuned constant step size can achieve a logarithmic individual regret bound in finite general-sum games.

Theorem 4.1 (Parameterized individual-regret bound)**.** *Under the setup of Section 3, there exist universal constants $c_0, c_1, C > 0$ such that, for every player i, integer horizon $T \geq 1$, and $0 < \eta \leq c_0/\sqrt{n}$,*

$$\mathrm{Reg}_i(T) \leq \frac{\log d_i}{\eta} + C + C\eta T \exp\left(-\frac{c_1}{\sqrt{n}\eta}\right). \tag{7}$$

The usual entropy bound leaves an $O(\eta T)$ term for bounded adversarial losses (Daskalakis et al., 2021, Lemma 4.1). Theorem 4.1 replaces this by $O(\eta T \exp\{-c_1/(\sqrt{n}\eta)\})$, up to a bounded additive term. Hence, choosing $1/(\sqrt{n}\eta)$ to be a sufficiently large multiple of $\log(T+2)$ keeps the remainder bounded and gives the following logarithmic regret bound.

Corollary 4.2 (Dimension-explicit logarithmic regret)**.** *There are universal constants $\bar{c}, C > 0$ such that, for every player i and integer horizon $T \geq 1$, choosing the common constant step size $\eta_T = \bar{c}/[\sqrt{n}\log(T+2)]$ before round 1 gives*

$$\mathrm{Reg}_i(T) \leq C\sqrt{n}\,\log d_i\,\log(T+2). \tag{8}$$

Corollary 4.2 improves the $O(n \log d_i \log^4 T)$ bound of Daskalakis et al. (2021) in two respects: the horizon dependence decreases from $\log^4 T$ to $\log T$, and the player dependence from n to $\sqrt{n}$. The action dependence remains logarithmic. This improvement comes from the larger admissible step size $\eta = \Theta(1/[\sqrt{n}\log T])$, rather than $\Theta(1/[n\log^4 T])$ in Daskalakis et al. (2021).

Recent work obtains logarithmic or constant external regret and sublogarithmic swap regret through modified regularization, adaptive step sizes, or higher-order predictions (Farina et al., 2022; Soleymani et al., 2025a;b; Tsuchiya, 2026; Liu et al., 2026; Abbadi et al., 2026). The lifted and cautious methods in Table 1, however, require inner optimization or scalar root finding. Corollary 4.2, on the other hand, attains a logarithmic regret bound with the closed-form update rule in Equation (6).

The following corollary shows that the improved individual regret bound yields a faster CCE approximation. It follows directly from Corollary 4.2 and the regret-to-CCE identity Equation (4).

Corollary 4.3 (Coarse-correlated-equilibrium convergence)**.** *Let $d := \max_i d_i$. Under the step-size choice of Corollary 4.2, there is a universal constant $C > 0$ such that, for every integer $T \geq 1$,*

$$\mathrm{Gap}_{\mathrm{CCE}}(\widehat{\mu}_T) \leq \frac{C\sqrt{n}\log d \log(T+2)}{T}. \tag{9}$$

Corollary 4.3 improves the CCE approximation rate of Optimistic Hedge from $O(n \log d \log^4 T/T)$ to $O(\sqrt{n}\log d \log T/T)$ (Daskalakis et al., 2021).

5 Proof Overview

We follow the approach of Daskalakis et al. (2021) in bounding regret through high-order smoothness of self-play. A centered-logit recurrence gives sharper difference bounds, which together with fixed-norm finite-difference interpolation produce a larger admissible step size and logarithmic regret.

5.1 Reduction to a First-Order Difference Estimate

For a mixed strategy $x \in \Delta_d$ and a vector of action losses $v \in \mathbb{R}^d$, define the weighted variance $\mathrm{Var}_x(v) := \sum_{a=1}^d x_a(v_a - \langle x, v\rangle)^2$. Optimistic regret bounds charge prediction error and subtract

stability (Rakhlin & Sridharan, 2013a;b), as formalized by the regret bounded by variation in utilities (RVU) framework (Syrgkanis et al., 2015). We use the variance refinement of Daskalakis et al. (2021, Lemma 4.1), specialized to a small constant step size.

Lemma 5.1 (Variance-based RVU bound). *There is a universal $\bar{\eta} > 0$ such that, for every player i, integer horizon $T \geq 1$, and $0 < \eta \leq \bar{\eta}$,*

$$\mathrm{Reg}_i(T) \leq \frac{\log d_i}{\eta} + \frac{2\eta}{3} \sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^t - \ell_i^{t-1}) - \frac{\eta}{3} \sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^{t-1}). \tag{10}$$

Lemma 5.1 directs us to bound the prediction-error variance sum by half the stability variance sum, up to a small remainder. A sup-norm bound, however, discards the weights x_i^t, while retaining them inside the variance leaves a weighting that changes across rounds. The identity $\mathrm{Var}_x(v) = \sum_{a<b} x_a x_b (v_a - v_b)^2$ suggests absorbing $\sqrt{x_a x_b}$ into each pairwise loss difference to work in a fixed Euclidean norm. This motivates us to define the weighted-gap vector $h_i^t \in \mathbb{R}^{\binom{d_i}{2}}$, for $a < b$, by

$$[h_i^t]_{ab} := \sqrt{x_{i,a}^t x_{i,b}^t}\big(\ell_{i,a}^{t-1} - \ell_{i,b}^{t-1}\big). \tag{11}$$

The squared norm $\|h_i^t\|_2^2$ already equals the stability variance. To control prediction-error variance using the same sequence, we must relate it to $\left\|h_i^{t+1} - h_i^t\right\|_2^2$. Prediction error compares consecutive losses under the same weights x_i^t, whereas $h_i^{t+1} - h_i^t$ changes both the losses and their weights. The following identity isolates this weight change. For each $a < b$, write $w^t := \sqrt{x_{i,a}^t x_{i,b}^t}$. Then

$$w^t\big[(\ell_{i,a}^t - \ell_{i,b}^t) - (\ell_{i,a}^{t-1} - \ell_{i,b}^{t-1})\big] = [h_i^{t+1} - h_i^t]_{ab} + \left(\frac{w^t}{w^{t+1}} - 1\right)[h_i^{t+1}]_{ab}.$$

The second term corrects for the changing weights. Optimistic Hedge gives $w^t/w^{t+1} - 1 = O(\eta)$, so the squared correction, summed over action pairs, is at most $O(\eta^2)\left\|h_i^{t+1}\right\|_2^2$. We square this identity, bound the cross term, and sum over action pairs and rounds to obtain the following reduction.

Lemma 5.2 (Weighted-gap reduction). *There are universal constants $C > 0$ and $\eta_0 > 0$ such that, for every player i, integer horizon $T \geq 1$, and $0 < \eta \leq \eta_0$,*

$$\sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^{t-1}) = \sum_{t=1}^{T} \left\|h_i^t\right\|_2^2, \tag{12}$$

$$\sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^t - \ell_i^{t-1}) \leq \frac{3}{2} \sum_{t=1}^{T} \left\|h_i^{t+1} - h_i^t\right\|_2^2 + C\eta^2 \sum_{t=1}^{T} \left\|h_i^t\right\|_2^2 + C\eta^2. \tag{13}$$

Lemmas 5.1 and 5.2 reduce bounding regret to comparing the two sums $\sum_{t=1}^{T} \left\|h_i^{t+1} - h_i^t\right\|_2^2$ and $\sum_{t=1}^{T+1} \|h_i^t\|_2^2$; the following lemma provides that key comparison.

Lemma 5.3 (First-order difference estimate). *There are universal constants $c, C > 0$ such that, for every player i, integer horizon $T \geq 1$, and $0 < \eta \leq c/\sqrt{n}$,*

$$\sum_{t=1}^{T} \left\|h_i^{t+1} - h_i^t\right\|_2^2 \leq \frac{1}{4} \sum_{t=1}^{T+1} \left\|h_i^t\right\|_2^2 + C\left(\frac{1}{\sqrt{n}\,\eta} + Te^{-c/(\sqrt{n}\eta)}\right). \tag{14}$$

Together with Lemma 5.2, this estimate controls the variance terms in Equation (10) and proves Theorem 4.1. We derive this key lemma in the following subsections.

5.2 Reduction to a High-Order Difference

For a sequence $y = (y^t)_{t\in\mathbb{Z}}$, define finite differences by $\Delta^0 y^t := y^t$, $\Delta y^t := y^{t+1} - y^t$, and $\Delta^{k+1} y^t := \Delta(\Delta^k y^t)$, with squared sequence norm $\|y\|_{\ell_2}^2 := \sum_t \|y^t\|^2$ for finite support. Lemma 5.2 motivates bounding the squared first differences $\|\Delta h_i^t\|_2^2$ by the squared norms $\|h_i^t\|_2^2$. The following lemma provides this comparison with an additional high-order difference term.

Lemma 5.4 (Finite-difference interpolation)**.** *For every finitely supported sequence y in a finite-dimensional Hilbert space, integer $m \geq 1$, and $\theta \in (0,1)$,*

$$\|\Delta y\|_{\ell_2}^2 \leq \theta \|y\|_{\ell_2}^2 + \theta^{1-m} \|\Delta^m y\|_{\ell_2}^2 . \tag{15}$$

The proof applies Parseval's identity and splits frequencies $\omega \in [-\pi, \pi]$ according to whether the condition $|e^{i\omega} - 1|^2 \leq \theta$ holds or not. The first term $\theta \|y\|_{\ell_2}^2$ bounds the low-frequency contribution to $\|\Delta y\|_{\ell_2}^2$, and the high-order term $\theta^{1-m} \|\Delta^m y\|_{\ell_2}^2$ bounds the contribution from the remaining frequencies. Hence, Lemma 5.4 suggests proving Lemma 5.3 by deriving a sharp bound on $\|\Delta^m h_i^t\|_2$.

5.3 The Centered-Logit Recurrence and High-Order Difference Estimate

To bound the high-order differences $\Delta^k h_i^t$ of the weighted gaps, we express Optimistic Hedge through a centered-logit state z^t. Derivative bounds for the centered loss map F and weighted-gap map H_i, combined with the centered-logit recurrence and a chain rule, then give the required estimates.

Centered-logits as a state. Since each x_i^t satisfies $x_{i,a}^t > 0$ for all $a \in A_i$, we can define the *centered logits* $z_i^t := P_i \log x_i^t$, where the centering map $P_i v := v - d_i^{-1}(\sum_b v_b)\mathbf{1}_{d_i}$ subtracts the coordinate mean. The logits z_i^t encode relative probabilities: $z_{i,a}^t - z_{i,b}^t = \log(x_{i,a}^t / x_{i,b}^t)$, and the softmax map $[\mathrm{softmax}(v)]_a := e^{v_a} / \sum_b e^{v_b}$ recovers the strategy $x_i^t = \mathrm{softmax}(z_i^t)$. We write $\mathrm{osc}(v) := \max_a v_a - \min_a v_a$ for the coordinate span and define the centered-logit spaces and product norm by

$$E_i := \{z_i \in \mathbb{R}^{d_i} : \mathbf{1}_{d_i}^\top z_i = 0\}, \qquad E := \prod_{i=1}^n E_i, \qquad \|z\|_E := \left(\sum_{i=1}^n \mathrm{osc}(z_i)^2\right)^{1/2}. \tag{16}$$

Centered-logit recurrence. Define the centered loss map $F : E \to E$, which maps centered logits to negated centered loss vectors, and the weighted loss-gap map $H_i : E \times E \to \mathbb{R}^{\binom{d_i}{2}}$, which maps two centered-logit profiles to weighted gaps, by

$$F_i(z) := [F(z)]_i = -P_i \ell_i\big((\mathrm{softmax}(z_j))_{j \neq i}\big), \tag{17}$$

$$[H_i(z, z')]_{ab} := \sqrt{[\mathrm{softmax}(z_i)]_a [\mathrm{softmax}(z_i)]_b}\big([F(z')]_{i,b} - [F(z')]_{i,a}\big), \qquad a < b. \tag{18}$$

The map $H_i(z, z')$ uses z for the probability weights and z' for the losses. For the joint centered-logit state $z^t := (z_i^t)_{i=1}^n$, Equation (6) gives $z^1 = 0$, $z^2 = 2\eta F(z^1)$, and, for $t \geq 2$,

$$z^{t+1} = z^t + \eta\big(2F(z^t) - F(z^{t-1})\big), \qquad h_i^t = H_i(z^t, z^{t-1}). \tag{19}$$

High-order difference. Applying $k-1$ finite differences to the increment Δz^t in the centered-logit recurrence gives, for $k \geq 1$ and $t \geq 2$,

$$\Delta^k z^t = \eta\big(2\Delta^{k-1} F(z^t) - \Delta^{k-1} F(z^{t-1})\big). \tag{20}$$

Hence loss differences of order $k-1$ control centered-logit differences of order k, with a factor η. We will transfer this control through H_i and prove the following estimate:

Lemma 5.5 (High-order difference estimate of the weighted gaps)**.** *There are universal constants $C, R > 0$ such that, for every player i and integer $k \geq 1$,*

$$\sup_{t \geq k+2} \left\|\Delta^k h_i^t\right\|_2 \leq C(R\sqrt{n}\,\eta)^k k!. \tag{21}$$

Notably, the weighted-gap estimate in Equation (21) becomes exponentially small at a suitable difference order. To see this, write $\delta := \sqrt{n}\eta$ and choose the difference order $m = \lfloor a/\delta \rfloor$, with $a > 0$, $Ra < 1$, and $\delta \leq a/2$. Then $m \geq a/(2\delta)$ and

$$\sup_{t \geq m+2} \left\|\Delta^m h_i^t\right\|_2 \leq C(R\delta)^m m! \leq C(R\delta m)^m \leq C(Ra)^m \leq Ce^{-c/\delta}.$$

Using $m! \leq m^m$ and choosing a small enough, this decay offsets the interpolation factor θ^{1-m}. Together with the finite-difference interpolation in Lemma 5.4, Lemma 5.5 yields Lemma 5.3. The remaining task is therefore to prove the weighted-gap bound in Lemma 5.5.

5.4 Proof of the High-Order Difference Estimate

We first bound the high-order differences $\left\|\Delta^k z^t\right\|_E$ of the centered logits using derivative bounds and a finite-difference chain rule. We then transfer these bounds through the weighted-gap map H_i to prove Lemma 5.5. For a smooth map Ψ, write $D^r\Psi(x)$ for its rth Fréchet derivative at x and $\|D^r\Psi(x)\|$ for its induced operator norm; see Appendix C.1 for the rigorous definition. For a real formal power series $U(s)$, let $[s^k]U$ denote the coefficient of s^k.

Step 1: control high-order derivatives $\|D^r F\|_E$. We bound the derivatives $D^r F$ through coordinate differences of the centered loss map F_i. Write $F_{i,a}$ for its action-a coordinate. For $z \in E$ and directions $v^1, \dots, v^r \in E$, the product norm $\|\cdot\|_E$ from Equation (16) gives

$$\left\|D^r F(z)[v^1, \dots, v^r]\right\|_E \le \sqrt{n} \max_i \max_{a,b \in A_i} \left|D^r(F_{i,a} - F_{i,b})(z)[v^1, \dots, v^r]\right|.$$

Now suppose we draw the opponents' actions A_j independently from $\mathrm{softmax}(z_j)$ for $j \neq i$. Their joint action A_{-i} determines the loss gap $g := \ell_i(b, \mathsf{A}_{-i}) - \ell_i(a, \mathsf{A}_{-i})$. For each direction v^q, let $v_j^q \in E_j$ be the perturbation to player j's logits; evaluating these perturbations at the same opponent actions and centering their sum gives $S_q := \sum_{j \neq i}\big(v_j^q(\mathsf{A}_j) - \mathbb{E}v_j^q(\mathsf{A}_j)\big)$, and the softmax in F_i gives the following joint-cumulant representation

$$D^r(F_{i,a} - F_{i,b})(z)[v^1, \dots, v^r] = \kappa(g, S_1, \dots, S_r).$$

Here $\kappa(g, S_1, \dots, S_r)$ is the joint cumulant of $g, S_1, \dots, S_r$, defined by

$$\kappa(Y_1, \dots, Y_m) := \sum_{\pi \in \Pi_m} (-1)^{|\pi|-1}(|\pi| - 1)! \prod_{B \in \pi} \mathbb{E} \prod_{j \in B} Y_j,$$

for real random variables $Y_1, \dots, Y_m$ with finite mixed moments, where Π_m is the set of partitions of $[m]$ into nonempty blocks (Leonov & Shiryaev, 1959).

The joint-cumulant bound in Lemma C.6, boundedness of g, and moment bounds for the centered sums S_q in terms of $\|v^q\|_E$ yield the following derivative estimate; Appendix C.4 gives the proof.

Lemma 5.6 (Factorial derivatives of the centered loss map)**.** *There are universal constants $A, R > 0$ such that every integer $r \ge 1$, player i, point $z \in E$, and directions $v^1, \dots, v^r \in E$ satisfy*

$$\mathrm{osc}\big(D^r F_i(z)[v^1, \dots, v^r]\big) \le A R^r r! \prod_{q=1}^{r} \|v^q\|_E, \tag{22}$$

$$\left\|D^r F(z)[v^1, \dots, v^r]\right\|_E \le A\sqrt{n}\, R^r r! \prod_{q=1}^{r} \|v^q\|_E. \tag{23}$$

Step 2: transfer to high-order differences $\left\|\Delta^k z^t\right\|_E$. By the high-order difference identity given by Equation (20), bounding the finite difference $\Delta^k z^t$ reduces to bounding order-$(k-1)$ differences of $F(z^t)$. The following chain rule uses derivatives of F and lower-order logit differences to control these terms, extending the chain rule of Daskalakis et al. (2021, Lemma 4.5) to normed spaces.

Lemma 5.7 (Finite-difference chain rule)**.** *Let X, Y be finite-dimensional normed spaces, $(u^t)_{t \ge 1}$ a sequence in X, and $\Psi : X \to Y$ smooth. Suppose $a_r := \sup_{t \ge r+1} \|\Delta^r u^t\| / r!$ and $c_p := \sup_{x \in X} \|D^p\Psi(x)\| / p!$ are finite for all integers $r, p \ge 1$. Then, for every $k \ge 1$,*

$$\sup_{t \ge k+1} \frac{\left\|\Delta^k \Psi(u^t)\right\|}{k!} \le [s^k] \sum_{p=1}^{k} c_p \left(\sum_{r=1}^{k} a_r s^r\right)^p. \tag{24}$$

Like the usual chain rule, Lemma 5.7 controls finite differences of the composition $\Psi(u^t)$ using derivatives of Ψ and finite differences of u^t. Here a_r bounds the rth input difference normalized by $r!$, and c_p bounds the pth derivative normalized by $p!$. Intuitively, repeated differentiation acts on the outer map and transfers to its input derivatives; Equation (24) bounds this chained contribution.

Together with the high-order difference identity in Equation (20) and the derivative bounds in Lemma 5.6, the chain rule with $\Psi = F$ and $u^t = z^t$ at order $k-1$ bounds order-k centered-logit differences $\Delta^k z^t$ using orders $1, \dots, k-1$. Starting from $\|\Delta z^t\|_E \le 3\sqrt{n}\eta$, the following lemma controls how these bounds grow with k.

Lemma 5.8 (High-order differences of the centered-logit state)**.** *For the trajectory z^t generated by the centered-logit recurrence, there are universal constants $C_z, R_z > 0$ such that, for every integer $k \geq 1$,*

$$\sup_{t \geq k+1} \left\| \Delta^k z^t \right\|_E \leq C_z (R_z \sqrt{n}\, \eta)^k k!. \tag{25}$$

The proof works with $a_k := \sup_{t \geq k+1} \left\| \Delta^k z^t \right\|_E / k!$ and seeks a geometric bound in k. The derivative estimates and the chain rule turn the logit recurrence into a polynomial upper bound for a_k in terms of $a_1, \ldots, a_{k-1}$ for $k \geq 2$. As the chain rule involves only sums and products of norm bounds, this polynomial upper bound has nonnegative coefficients. As a result, replacing each a_j in the polynomial by a number $w_j \geq a_j$, for $1 \leq j < k$, gives another upper bound on a_k. Hence, we can start from the initial upper bound $w_1 = C\sqrt{n}\eta \geq a_1$ for some constant C, and define the upper bound w_k by evaluating the polynomial at previous bounds $w_1, \ldots, w_{k-1}$. Computing the power series $\sum_{k \geq 1} w_k s^k$ explicitly gives $a_k \leq w_k \leq C_z (R_z \sqrt{n}\, \eta)^k$; restoring $k!$ proves the lemma.

The bound has one factor of $\sqrt{n}\eta$ per difference order: η comes from the update and $\sqrt{n}$ from combining playerwise derivative bounds in the product norm. After division by $k!$, the remaining growth is geometric with constants independent of the order. This permits a suitable choice of finite-difference order $m \asymp (\sqrt{n}\eta)^{-1}$ that yields an exponentially small bound on high-order differences.

Step 3: transfer to high-order differences $\left\| \Delta^k h_i^t \right\|_2$. Recall the identity $h_i^{t+1} = H_i(z^{t+1}, z^t)$; hence, we can transfer the estimate of centered-logits $\Delta^k z^t$ in Lemma 5.8 to an estimate of weighted gaps $\Delta^k h_i^t$ through the map H_i. Lemma 5.9 supplies a derivative bound of H_i for this argument.

Lemma 5.9 (Factorial derivatives of the weighted loss-gap map)**.** *For the weighted-gap map $H_i : E \times E \to \mathbb{R}^{\binom{d_i}{2}}$, equip the input space $E \times E$ with $\|(p,q)\| = (\|p\|_E^2 + \|q\|_E^2)^{1/2}$ and the output space with the Euclidean norm on unordered action pairs. Then, there are universal constants $A_H, R_H > 0$ such that, for every player i and integer $r \geq 0$,*

$$\sup_{z, z' \in E} \|D^r H_i(z, z')\| \leq A_H R_H^r r!. \tag{26}$$

For a mixed strategy $x_i = \mathrm{softmax}(z_i)$, differentiating the exponential formula for $\sqrt{x_{i,a}}$ leaves $\sqrt{x_{i,a}}$ as a multiplier. Cumulant bounds in Lemma C.6 and Lemma 5.6 then give factorial derivative bounds for H_i weighted by $\sqrt{x_{i,a} x_{i,b}}$, which can be bounded as $\sum_{a<b} x_{i,a} x_{i,b} \leq 1/2$.

Now, we are ready to prove Lemma 5.5 by combining Lemma 5.7, Lemma 5.8, and Lemma 5.9.

Proof of Lemma 5.5. Write $\delta := \sqrt{n}\eta$. For the paired centered-logit state (z^{t+1}, z^t), Lemma 5.8 and the product norm in Lemma 5.9 give $\sup_{t \geq k+1} \left\| \Delta^k (z^{t+1}, z^t) \right\| / k! \leq C_z (R_z \delta)^k$, after enlarging C_z by $\sqrt{2}$. Applying Lemmas 5.7 and 5.9 to $h_i^{t+1} = H_i(z^{t+1}, z^t)$ yields

$$\sup_{t \geq k+1} \frac{\left\| \Delta^k h_i^{t+1} \right\|_2}{k!} \leq [s^k] A_H \frac{R_H C_z R_z \delta s}{1 - (1 + R_H C_z) R_z \delta s} \leq C (R\delta)^k,$$

The chain rule sums the geometric series of input difference bounds. Multiplying by $k!$ and shifting to $t + 1 \geq k + 2$ proves Lemma 5.5. □

5.5 Proof of the First-Order Difference Estimate

Lemma 5.5 successfully controls the high-order difference of weighted gaps h_i^t. However, the finite-difference interpolation in Lemma 5.4 requires a finitely supported sequence. The following cutoff lemma transforms a given sequence u^t into a finitely supported one y^t while preserving the smoothness of the original sequence u^t within a small error. We defer the proof to Appendix D.

Lemma 5.10 (Finite-support cutoff)**.** *Fix $B, C_0, R_0, \rho > 0$ and integers $m \geq 1$, $L \geq 2$. Let $(u^t)_{t \geq 1}$ be a sequence in a finite-dimensional Euclidean space satisfying*

$$\sup_{t \geq 1} \left\| u^t \right\| \leq B, \qquad \sup_{t \geq k+2} \left\| \Delta^k u^t \right\| \leq C_0 (R_0 \rho)^k k! \quad (1 \leq k \leq m). \tag{27}$$

There is a constant $C_L > 0$, depending only on B, C_0, L, such that every integer horizon $T \geq 1$ admits a finitely supported sequence $y = (y^t)_{t\in\mathbb{Z}}$ with

$$\|\Delta y\|_{\ell_2}^2 \geq \sum_{t=1}^{T} \left\|\Delta u^t\right\|^2 - C_L m, \tag{28}$$

$$\|y\|_{\ell_2}^2 \leq \sum_{t=1}^{T+1} \left\|u^t\right\|^2 + C_L m, \tag{29}$$

$$\|\Delta^m y\|_{\ell_2}^2 \leq C_L(T+m)(2/L + R_0\rho m)^{2m}. \tag{30}$$

Intuitively, Lemma 5.10 constructs a tapered, finitely supported version y of the weighted gaps h_i^t. Now, we can apply finite-difference interpolation (Lemma 5.4) to y, then transfer the comparison back to h_i^t, and prove Lemma 5.3. With a sufficiently large L and a suitable $m \asymp (\sqrt{n}\eta)^{-1}$, this gives a $O(m)$ boundary error and a small $O((T+m)e^{-cm})$ contribution from the high-order differences.

Proof of Lemma 5.3. Set $\delta := \sqrt{n}\eta$ and $\theta := 1/4$, and let C, R be the constants in Lemma 5.5. Choose a fixed integer $L \geq 2$ sufficiently large and $a > 0$ sufficiently small that $\gamma := 2/L + Ra$ satisfies $\gamma^2/\theta \leq 1/2$. For $\delta \leq a/2$, set $m := \lfloor a/\delta \rfloor$, so $a/(2\delta) \leq m \leq a/\delta$.

By $\|h_i^t\|_2 \leq 1/2$ and Lemma 5.5, Lemma 5.10 applies with $u^t = h_i^t$ and $(B, C_0, R_0, \rho) = (1/2, C, R, \delta)$. It gives $\|\Delta^m y\|_{\ell_2}^2 \leq C_L(T+m)\gamma^{2m}$ since $R\delta m \leq Ra$. Its two squared-norm comparisons and finite-difference interpolation (Lemma 5.4) yield

$$\begin{aligned}\sum_{t=1}^{T} \left\|\Delta h_i^t\right\|_2^2 &\leq \theta \sum_{t=1}^{T+1} \left\|h_i^t\right\|_2^2 + Cm + C(T+m)\theta^{1-m}\gamma^{2m} \\ &\leq \frac{1}{4}\sum_{t=1}^{T+1} \left\|h_i^t\right\|_2^2 + C\left(\frac{1}{\delta} + Te^{-c'/\delta}\right).\end{aligned}$$

The last line uses $\theta^{1-m}\gamma^{2m} \leq 2^{-m}$, $m \leq a/\delta$, and $2^{-m} \leq e^{-c'/\delta}$ for $c' := a\log 2/2$. Since L is fixed, the constants are universal. Taking the stated $c \leq \min\{a/2, c'\}$ proves Lemma 5.3. □

Completing the regret bound. Combining Lemmas 5.1 to 5.3 proves Theorem 4.1; choosing $\eta_T = \bar{c}/[\sqrt{n}\log(T+2)]$ gives Corollary 4.2. Appendix A supplies the complete regret calculation.

6 Discussion

Implications. A notable recent line of work obtains logarithmic or constant regret bounds through modified learning dynamics (Farina et al., 2022; Abbadi et al., 2026). Our result shows that such modifications are unnecessary for logarithmic individual regret in general-sum games: under expected loss-vector feedback, Optimistic Hedge with a horizon-tuned constant step size already achieves this guarantee.

Improvement upon Daskalakis et al. (2021). Our larger constant step size $\eta = \Theta(1/[\sqrt{n}\log T])$, compared to the step size $\eta = \Theta(1/[n\log^4 T])$ in Daskalakis et al. (2021), comes from two improvements that work together. The factorial estimate in Lemma 5.5 controls high-order differences with an exponentially small bound at a suitably chosen order $m \asymp (\sqrt{n}\eta)^{-1}$. Then, the weighted-gap representation lets finite-difference interpolation transfer this control directly to first differences in a fixed Euclidean norm, avoiding the repeated comparisons under changing weights in Daskalakis et al. (2021). The $\sqrt{n}$ factor, on the other hand, originates from combining playerwise derivative bounds in the product norm. Appendix E details the estimates underlying these conditions.

7 Conclusion

In this work, we established the first logarithmic $O(\sqrt{n}\log d_i \log T)$ individual regret bound for Optimistic Hedge in general-sum games under expected loss-vector feedback. Our analysis transfers derivative bounds through a centered-logit recurrence to obtain a high-order difference bound for weighted loss gaps, and finite-difference interpolation yields a larger admissible step size. Whether Optimistic Hedge can achieve horizon-independent regret remains an open question.

AI USE STATEMENT

Generative AI tools assisted with mathematical exploration, including formulating and refining claims and hypotheses, developing the analytical framework and proof ingredients, drafting and revising proofs, and interpreting the resulting bounds. They also assisted with literature search and comparison, manuscript organization and exposition, LaTeX editing, and reference formatting. No datasets or empirical experiments are reported in this theoretical work. The author takes responsibility for the final claims, proofs, citations, and text, including material developed with AI assistance.

REPRODUCIBILITY STATEMENT

Sections 3 and 4 specify the setup, algorithm, and assumptions. Proofs of the theoretical results are provided in Section 5 and Appendices A–D.

REFERENCES

Omar Abbadi, Rida Laraki, and Panayotis Mertikopoulos. Constant regret in general games via higher-order optimism. *arXiv preprint arXiv:2609.04113*, 2026.

Robert J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, 1974. doi: 10.1016/0304-4068(74)90037-8.

David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956. doi: 10.2140/pjm.1956.6.1.

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

Noam Brown and Tuomas Sandholm. Superhuman AI for multiplayer poker. *Science*, 365(6456): 885–890, 2019. doi: 10.1126/science.aay2400.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997. doi: 10.1145/258128.258179.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Advances in Neural Information Processing Systems*, 2020.

Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 56(3):14:1–14:57, 2009. doi: 10.1145/1516512.1516516.

Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009. doi: 10.1137/070699652.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 235–254, 2011. doi: 10.1137/1.9781611973082.21.

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

Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning dynamics for general convex games. In *Advances in Neural Information Processing Systems*, 2022.

Dean P. Foster and Rakesh V. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1–2):40–55, 1997. doi: 10.1006/game.1997.0595.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997. doi: 10.1006/jcss.1997.1504.

Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1–2):79–103, 1999. doi: 10.1006/game.1999.0738.

Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. MIT Press, 1998.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pp. 2672–2680, 2014.

James Hannan. Approximation to Bayes risk in repeated play. In Melvin Dresher, Albert W. Tucker, and Philip Wolfe (eds.), *Contributions to the Theory of Games, Volume III*, volume 39 of *Annals of Mathematics Studies*, pp. 97–139. Princeton University Press, 1957.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000. doi: 10.1111/1468-0262.00153.

Sergiu Hart and Andreu Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98(1):26–54, 2001. doi: 10.1006/jeth.2000.2746.

V. P. Leonov and A. N. Shiryaev. On a method of calculation of semi-invariants. *Theory of Probability & Its Applications*, 4(3):319–329, 1959. doi: 10.1137/1104031.

Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994. doi: 10.1006/inco.1994.1009.

Mingyang Liu, Gabriele Farina, and Asuman Ozdaglar. Constant individual regret in general games. *arXiv preprint arXiv:2608.31166*, 2026.

Hervé Moulin and Jean-Philippe Vial. Strategically zero-sum games: The class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory*, 7 (3–4):201–221, 1978. doi: 10.1007/BF01769190.

Remi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J. Mankowitz, Doina Precup, and Bilal Piot. Nash learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 36743–36768, 2024.

John Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951. doi: 10.2307/1969529.

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

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 993–1019, 2013b.

Ashkan Soleymani, Georgios Piliouras, and Gabriele Farina. Cautious optimism: A meta-algorithm for near-constant regret in general games. In *Proceedings of the 26th ACM Conference on Economics and Computation*, 2025a. doi: 10.1145/3736252.3742639. URL https://arxiv.org/abs/2506.05005.

Ashkan Soleymani, Georgios Piliouras, and Gabriele Farina. Faster rates for no-regret learning in general games via cautious optimism. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pp. 518–529, 2025b. doi: 10.1145/3717823.3718242.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, volume 28, pp. 2989–2997, 2015.

Taira Tsuchiya. Sublogarithmic swap regret in multiplayer general-sum games via hybrid regularization. *arXiv preprint arXiv:2608.04149*, 2026.

John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1944.

Fang Wu, Xu Huang, Weihao Xuan, Zhiwei Zhang, Yijia Xiao, Guancheng Wan, Xiaomin Li, Bing Hu, Peng Xia, Jure Leskovec, and Yejin Choi. Multiplayer Nash preference optimization. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=x7aLhLMVn1.

APPENDIX CONTENTS

A Proof of the main result **14**

B Preliminary lemmas **15**

B.1 The variance-based RVU bound 15

B.2 The weighted pairwise loss gaps 16

C High-order finite-difference bounds **17**

C.1 Fréchet derivatives and operator norms 17

C.2 Proof of finite-difference interpolation 18

C.3 A joint-cumulant estimate 18

C.4 Proof of the centered-loss derivative bound 20

C.5 Proof of the finite-difference chain rule 21

C.6 Proof of the centered-logit difference bound 22

C.7 Proof of the weighted-gap derivative bound 24

D Finite-support cutoff construction **25**

D.1 Proof of the finite-support cutoff lemma 25

E Detailed Comparison with Daskalakis et al. (2021) **27**

E.1 The common objective 27

E.2 The analysis of Daskalakis et al. (2021) 27

E.3 Our approach: sharper differences in a fixed norm 28

A PROOF OF THE MAIN RESULT

We prove the regret bound for a constant step size, its logarithmic specialization, and the CCE guarantee in order.

Theorem 4.1 (Parameterized individual-regret bound)**.** *Under the setup of Section 3, there exist universal constants $c_0, c_1, C > 0$ such that, for every player i, integer horizon $T \geq 1$, and $0 < \eta \leq c_0/\sqrt{n}$,*

$$\mathrm{Reg}_i(T) \leq \frac{\log d_i}{\eta} + C + C\eta T \exp\left(-\frac{c_1}{\sqrt{n}\eta}\right). \tag{7}$$

Proof. Choose $c_0 \leq 1$ no larger than the universal thresholds in Lemmas 5.1 to 5.3 and small enough that $3/8 + C\eta^2 \leq 1/2$ whenever $\eta \leq c_0/\sqrt{n}$. Since $\left\|h_i^{T+1}\right\|_2^2 = \mathrm{Var}_{x_i^{T+1}}(\ell_i^T) \leq 1/4$ for losses in $[0,1]$, Lemmas 5.2 and 5.3 give

$$\begin{aligned}\sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^t - \ell_i^{t-1}) &\leq \left(\frac{3}{8} + C\eta^2\right) \sum_{t=1}^{T} \left\|h_i^t\right\|_2^2 + C\left(1 + \frac{1}{\sqrt{n}\eta} + Te^{-c/(\sqrt{n}\eta)}\right) \\ &\leq \frac{1}{2}\sum_{t=1}^{T} \left\|h_i^t\right\|_2^2 + C\left(\frac{1}{\sqrt{n}\eta} + Te^{-c/(\sqrt{n}\eta)}\right).\end{aligned}$$

The second line also uses $\sqrt{n}\eta \leq c_0 \leq 1$, so $1 \leq 1/(\sqrt{n}\eta)$. Substitution into Equation (10) yields

$$\begin{aligned}\mathrm{Reg}_i(T) \leq \frac{\log d_i}{\eta} + \frac{2\eta}{3}\left[\frac{1}{2}\sum_{t=1}^{T} \left\|h_i^t\right\|_2^2 + C\left(\frac{1}{\sqrt{n}\eta} + Te^{-c/(\sqrt{n}\eta)}\right)\right] \\ - \frac{\eta}{3}\sum_{t=1}^{T} \left\|h_i^t\right\|_2^2 \leq \frac{\log d_i}{\eta} + C + C\eta T e^{-c/(\sqrt{n}\eta)}.\end{aligned}$$

Taking c_1 to be the exponential constant supplied by Lemma 5.3 proves Equation (7). □

Theorem 4.1 leaves only $C\eta T e^{-c_1/(\sqrt{n}\eta)}$ to control. The horizon-tuned constant step size η_T makes this remainder bounded; the entropy term $(\log d_i)/\eta_T$ then gives Equation (8).

Corollary 4.2 (Dimension-explicit logarithmic regret)**.** *There are universal constants $\bar{c}, C > 0$ such that, for every player i and integer horizon $T \geq 1$, choosing the common constant step size $\eta_T = \bar{c}/[\sqrt{n}\log(T+2)]$ before round 1 gives*

$$\mathrm{Reg}_i(T) \leq C\sqrt{n}\,\log d_i\,\log(T+2). \tag{8}$$

Proof. Choose $\bar{c} > 0$ so that $\bar{c} \leq c_0 \log 3$ and $c_1/\bar{c} \geq 2$. Then $\eta_T \leq c_0/\sqrt{n}$ for every $T \geq 1$, and

$$\exp\left(-\frac{c_1}{\sqrt{n}\eta_T}\right) = (T+2)^{-c_1/\bar{c}} \leq (T+2)^{-2}.$$

Substituting $\eta_T = \bar{c}/[\sqrt{n}\log(T+2)]$ into Equation (7) gives

$$\mathrm{Reg}_i(T) \leq \frac{\sqrt{n}\,\log d_i\,\log(T+2)}{\bar{c}} + C + \frac{C\bar{c}T}{\sqrt{n}\log(T+2)(T+2)^2}.$$

For $d_i = 1$, regret is zero. For $d_i \geq 2$, $\sqrt{n}\log d_i \log(T+2) \geq \log 2 \log 3$, so the bounded terms are absorbed by enlarging the universal constant. This proves Equation (8). □

The identity in Equation (4) concerns $\widehat{\mu}_T$, the time-average of the joint distributions $\bigotimes_j x_j^t$, and is exact under the expected-vector model.

Corollary 4.3 (Coarse-correlated-equilibrium convergence)**.** *Let $d := \max_i d_i$. Under the step-size choice of Corollary 4.2, there is a universal constant $C > 0$ such that, for every integer $T \geq 1$,*

$$\mathrm{Gap}_{\mathrm{CCE}}(\widehat{\mu}_T) \leq \frac{C\sqrt{n}\log d \log(T+2)}{T}. \tag{9}$$

Proof. For any player i and fixed deviation a_i', multilinearity gives

$$\mathbb{E}_{a\sim\widehat{\mu}_T}\ell_i(a) - \mathbb{E}_{a\sim\widehat{\mu}_T}\ell_i(a_i', a_{-i}) = \frac{1}{T}\sum_{t=1}^{T}\big(\langle x_i^t, \ell_i^t\rangle - \ell_{i,a_i'}^t\big).$$

Maximizing over a_i' and then over i proves Equation (4). Combining this identity with Corollary 4.2 and $d_i \le d$ proves Equation (9). □

B Preliminary lemmas

We recall the variance-based RVU bound for Optimistic Hedge, then prove the weighted-gap representation used to compare its two variance sums.

B.1 The variance-based RVU bound

Soft minimum and Bregman divergence. Fix a player i, suppress its index, and fix $\eta > 0$. For loss-score vectors $q, q' \in \mathbb{R}^{d_i}$, define the soft minimum $\Phi : \mathbb{R}^{d_i} \to \mathbb{R}$ and the Bregman divergence of $-\Phi$, denoted by $D_\Phi : \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} \to \mathbb{R}_+$, by

$$\Phi(q) := -\frac{1}{\eta}\log\sum_{a=1}^{d_i} e^{-\eta q_a}, \qquad D_\Phi(q', q) := \Phi(q) + \langle \nabla\Phi(q), q' - q\rangle - \Phi(q').$$

Here $\nabla\Phi(q) = \mathrm{softmax}(-\eta q)$.

We recover the variance refinement of the RVU bound in Daskalakis et al. (2021, Lemma 4.1), following optimistic prediction-error analysis (Rakhlin & Sridharan, 2013b;a) and the RVU framework (Syrgkanis et al., 2015). Regret contains a difference of divergence sums, so we need the following comparison with local variance from both above and below.

Lemma B.1 (Soft-min divergence and local variance)**.** *Let $p = \nabla\Phi(q)$ and $\mathrm{osc}(v) \le R$. Then*

$$\frac{\eta}{2}e^{-\eta R}\,\mathrm{Var}_p(v) \le D_\Phi(q+v, q) \le \frac{\eta}{2}e^{\eta R}\,\mathrm{Var}_p(v). \tag{31}$$

Proof. We compare the weighted variance along the segment $q + sv$ with the weighted variance under p. Set $p_s := \nabla\Phi(q + sv)$ for $0 \le s \le 1$. The Hessian identity and Taylor's integral remainder give

$$\begin{aligned} -\nabla^2\Phi(q+sv) &= \eta\big(\mathrm{Diag}(p_s) - p_s p_s^\top\big), \\ D_\Phi(q+v, q) &= \eta\int_0^1 (1-s)\,\mathrm{Var}_{p_s}(v)\,\mathrm{d}s. \end{aligned} \tag{32}$$

Thus $D_\Phi(q+v, q)$ averages $\mathrm{Var}_{p_s}(v)$ along the segment. To compare with $\mathrm{Var}_p(v)$, use $\mathrm{Var}_x(v) = \min_{c\in\mathbb{R}}\sum_a x_a(v_a - c)^2$ for a mixed strategy $x \in \Delta_{d_i}$. Then

$$\begin{aligned} e^{-\eta s R} \le \frac{(p_s)_a}{p_a} = \frac{e^{-\eta s v_a}}{\sum_b p_b e^{-\eta s v_b}} \le e^{\eta s R}, \\ e^{-\eta R}\,\mathrm{Var}_p(v) \le \mathrm{Var}_{p_s}(v) \le e^{\eta R}\,\mathrm{Var}_p(v). \end{aligned}$$

The first line uses $\mathrm{osc}(v) \le R$. For the second, multiply the coordinatewise probability bounds by $(v_a - c)^2$, sum, and minimize over c, using $s \le 1$. Substituting into Equation (32) and using $\int_0^1 (1-s)\,\mathrm{d}s = 1/2$ proves the claim. □

We now derive the variance-based RVU bound.

Lemma 5.1 (Variance-based RVU bound)**.** *There is a universal $\bar{\eta} > 0$ such that, for every player i, integer horizon $T \ge 1$, and $0 < \eta \le \bar{\eta}$,*

$$\mathrm{Reg}_i(T) \le \frac{\log d_i}{\eta} + \frac{2\eta}{3}\sum_{t=1}^{T}\mathrm{Var}_{x_i^t}(\ell_i^t - \ell_i^{t-1}) - \frac{\eta}{3}\sum_{t=1}^{T}\mathrm{Var}_{x_i^t}(\ell_i^{t-1}). \tag{10}$$

Proof. To telescope the soft-min potential Φ, distinguish cumulative loss L^t from the cumulative loss plus prediction q^t defining x^t. Let

$$L^t := \sum_{\tau=1}^{t} \ell^\tau, \qquad q^t := L^{t-1} + \ell^{t-1}.$$

Here $L^0 = 0$, and q^t adds the prediction ℓ^{t-1} to past losses L^{t-1}. Unrolling Optimistic Hedge gives $x^t = \mathrm{softmax}(-\eta q^t) = \nabla\Phi(q^t)$. The divergence difference separates the incurred loss $\langle x^t, \ell^t\rangle$ from the change in Φ:

$$D_\Phi(L^t, q^t) - D_\Phi(L^{t-1}, q^t) = \langle x^t, \ell^t\rangle - \big[\Phi(L^t) - \Phi(L^{t-1})\big].$$

Summing cancels the intermediate potential values. Using $\Phi(L^T) \le \min_a L_a^T$ and $\Phi(0) = -(\log d_i)/\eta$ then gives

$$\mathrm{Reg}(T) \le \frac{\log d_i}{\eta} + \sum_{t=1}^{T} \big[D_\Phi(L^t, q^t) - D_\Phi(L^{t-1}, q^t)\big]. \tag{33}$$

We bound $D_\Phi(L^t, q^t)$ above and $D_\Phi(L^{t-1}, q^t)$ below. Their displacements are the prediction error $\ell^t - \ell^{t-1}$ and the negative previous loss $-\ell^{t-1}$.

The prediction-error displacement $L^t - q^t = \ell^t - \ell^{t-1}$ has coordinate span at most 2, whereas $L^{t-1} - q^t = -\ell^{t-1}$ has span at most 1. Applying Lemma B.1 to the positive and negative divergences in Equation (33) gives

$$\mathrm{Reg}(T) \le \frac{\log d_i}{\eta} + \frac{\eta e^{2\eta}}{2} \sum_{t=1}^{T} \mathrm{Var}_{x^t}(\ell^t - \ell^{t-1}) - \frac{\eta e^{-\eta}}{2} \sum_{t=1}^{T} \mathrm{Var}_{x^t}(\ell^{t-1}). \tag{34}$$

Choose a universal $\bar\eta > 0$ such that $e^{2\eta}/2 \le 2/3$ and $e^{-\eta}/2 \ge 1/3$ for $0 < \eta \le \bar\eta$. This proves Lemma 5.1. □

B.2 The weighted pairwise loss gaps

The identity $\mathrm{Var}_{x_i^t}(\ell_i^{t-1}) = \|h_i^t\|_2^2$ represents stability exactly. Prediction error is more delicate: differencing the weighted gaps h_i^t changes both the losses and their probability weights. We show that this additional weight change contributes only $O(\eta^2)$ times the squared weighted-gap magnitude, plus a bounded terminal correction.

Lemma 5.2 (Weighted-gap reduction)**.** *There are universal constants $C > 0$ and $\eta_0 > 0$ such that, for every player i, integer horizon $T \ge 1$, and $0 < \eta \le \eta_0$,*

$$\sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^{t-1}) = \sum_{t=1}^{T} \big\|h_i^t\big\|_2^2, \tag{12}$$

$$\sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^t - \ell_i^{t-1}) \le \frac{3}{2} \sum_{t=1}^{T} \big\|h_i^{t+1} - h_i^t\big\|_2^2 + C\eta^2 \sum_{t=1}^{T} \big\|h_i^t\big\|_2^2 + C\eta^2. \tag{13}$$

Proof. For a mixed strategy $x \in \Delta_d$ and an action-loss vector $v \in \mathbb{R}^d$, the pairwise variance identity is

$$\mathrm{Var}_x(v) = \sum_{a<b} x_a x_b (v_a - v_b)^2. \tag{35}$$

This identifies $\mathrm{Var}_{x_i^t}(\ell_i^{t-1}) = \|h_i^t\|_2^2$ and gives Equation (12). To verify it, expand $\frac{1}{2}\sum_{a,b} x_a x_b (v_a - v_b)^2$; the diagonal terms vanish.

To compare the consecutive weighted gaps h_i^{t+1} and h_i^t, isolate their probability weights. Fix player i and pair $a < b$, and write

$$w_{ab}^t := \sqrt{x_{i,a}^t x_{i,b}^t}.$$

The weight w_{ab}^t scales the pairwise gap $\ell_{i,a}^{t-1} - \ell_{i,b}^{t-1}$. To track its update, let $m_i^t := 2\ell_i^t - \ell_i^{t-1}$ and $Z_i^t := \sum_c x_{i,c}^t e^{-\eta m_{i,c}^t}$; these are the loss vector in the exponent and the normalizing constant in the next update. The Optimistic Hedge update gives

$$\frac{w_{ab}^{t+1}}{w_{ab}^t} = \frac{\exp\{-\eta(m_{i,a}^t + m_{i,b}^t)/2\}}{Z_i^t}, \qquad e^{-\eta \max_c m_{i,c}^t} \le Z_i^t \le e^{-\eta \min_c m_{i,c}^t}.$$

Since $\mathrm{osc}(m_i^t) \le 3$,

$$e^{-3\eta} \le \frac{w_{ab}^{t+1}}{w_{ab}^t} \le e^{3\eta}, \qquad \left| \frac{w_{ab}^t}{w_{ab}^{t+1}} - 1 \right| \le C\eta \tag{36}$$

for a universal C and small universal η. Thus the relative weight change is $O(\eta)$, whose square produces the $O(\eta^2)$ correction.

Since $[h_i^t]_{ab} = w_{ab}^t(\ell_{i,a}^{t-1} - \ell_{i,b}^{t-1})$,

$$w_{ab}^t\big[(\ell_{i,a}^t - \ell_{i,b}^t) - (\ell_{i,a}^{t-1} - \ell_{i,b}^{t-1})\big] = [h_i^{t+1} - h_i^t]_{ab} + \left(\frac{w_{ab}^t}{w_{ab}^{t+1}} - 1\right)[h_i^{t+1}]_{ab}. \tag{37}$$

In Equation (37), $[h_i^{t+1} - h_i^t]_{ab}$ is the weighted-gap change; the remaining term accounts for the weight ratio w_{ab}^t/w_{ab}^{t+1}. Apply $(a+b)^2 \le \frac{3}{2}a^2 + 3b^2$ to Equation (37), then sum over t and $a < b$. Using Equations (35) and (36),

$$\begin{aligned}
\sum_{t=1}^T \mathrm{Var}_{x_i^t}(\ell_i^t - \ell_i^{t-1}) &= \sum_{t=1}^T \sum_{a<b} \Big(w_{ab}^t\big[(\ell_{i,a}^t - \ell_{i,b}^t) - (\ell_{i,a}^{t-1} - \ell_{i,b}^{t-1})\big]\Big)^2 \\
&\le \frac{3}{2}\sum_{t=1}^T \big\|h_i^{t+1} - h_i^t\big\|_2^2 + C\eta^2 \sum_{t=1}^T \big\|h_i^{t+1}\big\|_2^2 \\
&\le \frac{3}{2}\sum_{t=1}^T \big\|h_i^{t+1} - h_i^t\big\|_2^2 + C\eta^2 \sum_{t=1}^T \big\|h_i^t\big\|_2^2 + C\eta^2.
\end{aligned}$$

The last inequality uses $h_i^1 = 0$ and

$$\sum_{t=1}^T \big\|h_i^{t+1}\big\|_2^2 = \sum_{t=1}^T \big\|h_i^t\big\|_2^2 + \big\|h_i^{T+1}\big\|_2^2 \le \sum_{t=1}^T \big\|h_i^t\big\|_2^2 + \frac{1}{4}.$$

The terminal variance $\big\|h_i^{T+1}\big\|_2^2 \le 1/4$ supplies the additive correction. The weight estimate controls the ratio w_{ab}^{t+1}/w_{ab}^t, so no uniform lower bound on the action probabilities is needed. □

C HIGH-ORDER FINITE-DIFFERENCE BOUNDS

Proof strategy. The target is the factorial finite-difference bound of Lemma 5.5. We first reduce the derivative norm of F to derivatives of expected loss gaps, then bound their joint-cumulant representations to prove Lemma 5.6. The finite-difference chain rule and scalar comparison then prove the centered-logit difference bound in Lemma 5.8 from the centered-logit recurrence. Finally, Lemmas 5.7 to 5.9 transfer that bound through H_i to prove Lemma 5.5.

Notation. We use moment norms for random quantities and operator norms for derivatives. For $p \ge 1$, the moment norm is $\|Y\|_{L^p} := (\mathbb{E}|Y|^p)^{1/p}$. To track all temporal orders together, we use formal power series. For a formal power series $P(s) = \sum_{k\ge 0} p_k s^k$, write $[s^k]P := p_k$, with the same notation for other indeterminates. Write $\mathbb{R}[[s]]$ for real formal power series and $U \preceq V$ for coefficientwise domination. This bounds each order separately without requiring series convergence.

C.1 FRÉCHET DERIVATIVES AND OPERATOR NORMS

The derivatives below measure how the centered loss map F and weighted-gap map H_i respond to perturbations of their centered-logit inputs. We first define the derivatives, then the operator norms used to bound them.

Definition C.1 (Fréchet derivative)**.** *Let X, Y be finite-dimensional normed spaces, let $U \subseteq X$ be open, and let $\Psi : U \to Y$. The Fréchet derivative at $x \in U$ is the unique continuous linear map $D\Psi(x) : X \to Y$, when it exists, satisfying*

$$\lim_{\substack{h \to 0 \\ h \neq 0}} \frac{\|\Psi(x+h) - \Psi(x) - D\Psi(x)[h]\|_Y}{\|h\|_X} = 0.$$

Definition C.2 (Higher-order Fréchet derivatives)**.** *Let X, Y be finite-dimensional normed spaces, let $U \subseteq X$ be open, and let $\Psi : U \to Y$. For an integer $r \geq 2$, suppose $D^{r-1}\Psi$ is defined near $x \in U$ and is Fréchet differentiable at x as a map into the continuous $(r-1)$-linear maps, equipped with the operator norm defined below. The rth Fréchet derivative $D^r\Psi(x) : X^r \to Y$ is defined by*

$$D^r\Psi(x)[v^1, \ldots, v^r] := \big(D(D^{r-1}\Psi)(x)[v^r]\big)[v^1, \ldots, v^{r-1}].$$

We set $D^0\Psi(x) := \Psi(x)$ and call Ψ smooth when these derivatives exist and are continuous at every order.

For smooth Ψ, the derivative $D^r\Psi(x)$ is symmetric. Its evaluation has a direct interpretation: perturb x independently in the directions $v^1, \ldots, v^r$ and differentiate once in each perturbation amplitude:

$$D^r\Psi(x)[v^1, \ldots, v^r] = \partial_{s_1} \cdots \partial_{s_r} \Psi\left(x + \sum_{q=1}^{r} s_q v^q\right)\Bigg|_{s_1 = \cdots = s_r = 0}.$$

The operator norm makes this response uniform over directions of norm at most one.

Definition C.3 (Multilinear operator norm)**.** *Let X, Y be finite-dimensional normed spaces and $r \geq 1$. The induced operator norm of a continuous r-linear map $L : X^r \to Y$ is*

$$\|L\| := \sup_{\substack{\|v^q\|_X \leq 1 \\ q=1,\ldots,r}} \big\|L[v^1, \ldots, v^r]\big\|_Y.$$

Notation. We use this operator norm for $L = D^r\Psi(x)$, and set $\big\|D^0\Psi(x)\big\| := \|\Psi(x)\|_Y$ at order zero. For $F : E \to E$, both spaces carry the centered-logit product norm $\|\cdot\|_E$, and we write $\|D^r F(z)\|_E$ for the resulting operator norm. For $H_i : E \times E \to \mathbb{R}^{\binom{d_i}{2}}$, the input norm is $\|(p,q)\| = (\|p\|_E^2 + \|q\|_E^2)^{1/2}$ and the output norm is Euclidean, as in Lemma 5.9.

C.2 Proof of finite-difference interpolation

Lemma 5.4 (Finite-difference interpolation)**.** *For every finitely supported sequence y in a finite-dimensional Hilbert space, integer $m \geq 1$, and $\theta \in (0,1)$,*

$$\|\Delta y\|_{\ell_2}^2 \leq \theta \, \|y\|_{\ell_2}^2 + \theta^{1-m} \, \|\Delta^m y\|_{\ell_2}^2. \tag{15}$$

Proof. We compare the sequence norms in an orthonormal basis of the finite-dimensional Hilbert space. Define the Fourier transform $\widehat{y}(\omega) := \sum_{t \in \mathbb{Z}} y^t e^{-it\omega}$ for $\omega \in [-\pi, \pi]$, using the complex Euclidean norm on these coordinates. The sum is finite, and shifting the time index gives $\widehat{\Delta^r y}(\omega) = (e^{i\omega} - 1)^r \widehat{y}(\omega)$. Parseval's identity therefore yields, for every integer $r \geq 0$,

$$\|\Delta^r y\|_{\ell_2}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |e^{i\omega} - 1|^{2r} \, \|\widehat{y}(\omega)\|^2 \, \mathrm{d}\omega.$$

For $\lambda := |e^{i\omega} - 1|^2$, the inequality $\lambda \leq \theta + \theta^{1-m}\lambda^m$ holds at every frequency: the first term suffices when $\lambda \leq \theta$, and the second when $\lambda > \theta$. Multiplying by $\|\widehat{y}(\omega)\|^2 / (2\pi)$ and integrating gives Equation (15). □

C.3 A joint-cumulant estimate

To prove Lemma 5.6, we express derivatives of expectations under softmax distributions as joint cumulants. We first define these quantities and prove their derivative representation. We then bound them using moments, separating growth in the derivative order from the number of players and actions.

Definition C.4 (Joint cumulant)**.** *Let $m \geq 1$, and let $Y_1, \ldots, Y_m$ be real random variables on a common probability space with $\mathbb{E}\prod_{j\in B}|Y_j| < \infty$ for every nonempty $B \subseteq [m]$. Their joint cumulant is the real number*

$$\kappa(Y_1, \ldots, Y_m) := \sum_{\pi\in\Pi_m} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{B\in\pi} \mathbb{E}\prod_{j\in B} Y_j, \tag{38}$$

where Π_m is the set of partitions of $[m]$ into nonempty blocks.

For one variable, $\kappa(Y_1) = \mathbb{E}Y_1$; for two, $\kappa(Y_1, Y_2) = \mathrm{Cov}(Y_1, Y_2)$. The next lemma identifies every joint cumulant with a mixed derivative of the logarithm of the joint moment-generating function. Its exponential-moment assumption holds automatically for the finite-support random variables in our softmax calculations.

Lemma C.5 (Derivative representation of joint cumulants)**.** *Let the joint moment-generating function $M(s) := \mathbb{E}\exp\{\sum_{j=1}^m s_j Y_j\}$ be finite in a neighborhood of $s = 0 \in \mathbb{R}^m$. Then*

$$\kappa(Y_1, \ldots, Y_m) = \partial_{s_1} \cdots \partial_{s_m} \log M(s)|_{s=0}\,.$$

Proof. The exponential-moment assumption permits Taylor expansion and differentiation under the expectation near zero. Let $[s_1 \cdots s_m]$ denote the coefficient of the monomial in which each variable appears once. Since $M(0) = 1$, only the first m terms of $\log M = \sum_{k\geq 1}(-1)^{k-1}(M-1)^k/k$ contribute to this coefficient. Each contribution assigns a nonempty subset of $[m]$ to each of the k factors. Thus

$$\begin{aligned}
\partial_{s_1} \cdots \partial_{s_m} \log M(s)|_{s=0} &= [s_1 \cdots s_m] \log M(s) \\
&= \sum_{k=1}^m \frac{(-1)^{k-1}}{k} \sum_{(B_1,\ldots,B_k)} \prod_{q=1}^k \mathbb{E}\prod_{j\in B_q} Y_j \\
&= \sum_{\pi\in\Pi_m} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{B\in\pi} \mathbb{E}\prod_{j\in B} Y_j.
\end{aligned}$$

The inner sum is over ordered partitions of $[m]$ into k nonempty blocks. Each unordered partition has $k!$ orderings, giving the factor $k!/k = (k-1)!$ in the last line. This is Equation (38). □

The following lemma bounds a joint cumulant by the product of its arguments' moment scales, with a single factorial in the order. It does not require independence among the arguments.

Lemma C.6 (Joint-cumulant bound)**.** *Let $m \geq 2$, and let $Y_1, \ldots, Y_m$ be real random variables on the same probability space. Suppose that numbers $K_1, \ldots, K_m \geq 0$ satisfy*

$$\|Y_j\|_{L^p} \leq K_j\sqrt{p} \qquad \text{for every } p \geq 1 \text{ and } j \in [m].$$

Then

$$|\kappa(Y_1, \ldots, Y_m)| \leq (2e)^m (m-1)! \prod_{j=1}^m K_j. \tag{39}$$

Moreover, a joint cumulant of order $m \geq 2$ is unchanged when any one of its arguments is shifted by a constant.

Proof. We bound the moment for each block of a partition, then evaluate the weighted partition sum exactly. For a block B of size b, generalized Hölder gives

$$\left|\mathbb{E}\prod_{j\in B} Y_j\right| \leq \prod_{j\in B} \|Y_j\|_{L^b} \leq b^{b/2} \prod_{j\in B} K_j \leq e^b b! \prod_{j\in B} K_j.$$

The last step uses $b! \geq (b/e)^b$. For k ordered blocks, there are $\binom{m-1}{k-1}$ positive tuples of block sizes summing to m. For each tuple, the multinomial count $m!/(b_1! \cdots b_k!)$ cancels the product of block

factorials. Removing the $k!$ block orderings and multiplying by the cumulant weight $(k-1)!$ leaves the factor $1/k$. Hence

$$\sum_{\pi\in\Pi_m}(|\pi|-1)!\prod_{B\in\pi}|B|! = m!\sum_{k=1}^{m}\frac{1}{k}\binom{m-1}{k-1}$$
$$= (m-1)!\sum_{k=1}^{m}\binom{m}{k} = (m-1)!(2^m-1).$$

Consequently,

$$|\kappa(Y_1,\dots,Y_m)| \le e^m(m-1)!(2^m-1)\prod_{j=1}^{m}K_j \le (2e)^m(m-1)!\prod_{j=1}^{m}K_j.$$

For shift invariance, Equation (38) is multilinear, so it suffices to show that a cumulant with one constant argument is zero. Fix a partition of the other $m-1$ indices into k blocks. Making the constant a singleton has coefficient $(-1)^k k!$; adjoining it to one of the k blocks has total coefficient $k(-1)^{k-1}(k-1)!$. These contributions have the same moment product and cancel. □

C.4 Proof of the centered-loss derivative bound

For the derivative bound below, the cumulant arguments are a bounded loss gap g and centered opponent sums $S_1,\dots,S_r$, defined in the proof. We apply Lemma C.6 with $m=r+1$, $Y_1=g$, and $Y_{q+1}=S_q$ for $q\in[r]$. The moment bounds $\|g\|_{L^p}\le 1$ and $\|S_q\|_{L^p}\le C\sqrt{p}\,\|v^q\|_E$ give $K_1=1$ and $K_{q+1}=C\,\|v^q\|_E$, yielding the required $r!$ derivative bound without independence among the S_q.

Lemma 5.6 (Factorial derivatives of the centered loss map)**.** *There are universal constants $A,R>0$ such that every integer $r\ge 1$, player i, point $z\in E$, and directions $v^1,\dots,v^r\in E$ satisfy*

$$\operatorname{osc}\big(D^rF_i(z)[v^1,\dots,v^r]\big) \le AR^r r!\prod_{q=1}^{r}\|v^q\|_E\,, \tag{22}$$

$$\big\|D^rF(z)[v^1,\dots,v^r]\big\|_E \le A\sqrt{n}\,R^r r!\prod_{q=1}^{r}\|v^q\|_E\,. \tag{23}$$

Proof. The proof strategy is to bound each scalar derivative $D^r(F_{i,a}-F_{i,b})$ before combining the player blocks. Fix $r\ge 1$, a centered-logit state $z\in E$, and perturbation directions $v^1,\dots,v^r\in E$. Write $v^q=(v_1^q,\dots,v_n^q)$, where $v_j^q\in E_j$ is the player-j block of direction q, and $v_j^q(c)$ is its coordinate for action $c\in A_j$. Write $F_{i,a}$ for action a's coordinate of F_i. The product norm gives

$$\big\|D^rF(z)[v^1,\dots,v^r]\big\|_E = \left(\sum_{i=1}^{n}\operatorname{osc}\big(D^rF_i(z)[v^1,\dots,v^r]\big)^2\right)^{1/2}$$
$$\le \sqrt{n}\max_i\max_{a,b\in A_i}\big|D^r(F_{i,a}-F_{i,b})(z)[v^1,\dots,v^r]\big|\,.$$

The inequality uses that a vector's oscillation is its largest absolute coordinate difference. It therefore suffices to bound every scalar derivative in the last line by $AR^r r!\prod_{q=1}^{r}\|v^q\|_E$, uniformly over i,a,b.

Fix such a player i and actions $a,b\in A_i$. Introduce independent opponent actions $\mathsf{A}_j\sim\operatorname{softmax}(z_j)$ for $j\ne i$, and write $\mathsf{A}_{-i}:=(\mathsf{A}_j)_{j\ne i}$. Define the loss gap g and centered directional sums S_q by

$$g:=\ell_i(b,\mathsf{A}_{-i})-\ell_i(a,\mathsf{A}_{-i}),\qquad S_q:=\sum_{j\ne i}\big(v_j^q(\mathsf{A}_j)-\mathbb{E}v_j^q(\mathsf{A}_j)\big),\qquad q\in[r].$$

The random variable g takes values in $[-1,1]$. Each S_q is the directional derivative at z of the opponents' joint log-probability along v^q, evaluated at their random action profile. All expectations and centering means below use this fixed law at z. Since P_i subtracts the same value from every coordinate, $(F_{i,a}-F_{i,b})(z)=\mathbb{E}g$.

We next express derivatives of this expectation as cumulants. Let $s_1, \dots, s_r \in \mathbb{R}$ specify the perturbation amplitudes. Softmax multiplies the opponents' joint probability weights by $\exp\{\sum_q s_q \sum_{j \neq i} v_j^q(\mathsf{A}_j)\}$ and then normalizes. Subtracting the means in S_q changes every weight by a common factor that cancels on normalization. Introducing an auxiliary scalar $u \in \mathbb{R}$ to extract the loss gap by differentiation gives

$$(F_{i,a} - F_{i,b})\left(z + \sum_{q=1}^{r} s_q v^q\right) = \frac{\mathbb{E}\left[g \exp\{\sum_{q=1}^r s_q S_q\}\right]}{\mathbb{E}\exp\{\sum_{q=1}^r s_q S_q\}}$$
$$= \partial_u \log \mathbb{E}\exp\left\{ug + \sum_{q=1}^{r} s_q S_q\right\}\Bigg|_{u=0}.$$

The auxiliary scalar u inserts the loss gap g through differentiation, while the logarithm supplies the normalizing denominator. Differentiating once in each s_q and setting the perturbation amplitudes to zero gives, by Lemma C.5,

$$D^r(F_{i,a} - F_{i,b})(z)[v^1, \dots, v^r] = \kappa(g, S_1, \dots, S_r). \tag{40}$$

All arguments have finite support, which justifies the differentiations and the cumulant representation.

It remains to bound this cumulant. Since $|g| \leq 1$, Lemma C.6 reduces the task to controlling the moments of each S_q. Its summands are independent across opponents, centered, and have range lengths $\operatorname{osc}(v_j^q)$. Hoeffding's lemma and the resulting subgaussian moment bound give

$$\mathbb{E}e^{\lambda S_q} \leq \exp\left\{\frac{\lambda^2}{8}\sum_{j\neq i}\operatorname{osc}(v_j^q)^2\right\} \leq \exp\left\{\frac{\lambda^2}{8}\|v^q\|_E^2\right\}, \qquad \lambda \in \mathbb{R},$$
$$\|S_q\|_{L^p} \leq C\sqrt{p}\,\|v^q\|_E, \qquad p \geq 1.$$

The second line follows by integrating the subgaussian tail implied by the first; C is universal. The sums S_q may depend on one another, but Lemma C.6 requires only their individual moment bounds. Applying it to the $r+1$ arguments in Equation (40), using $\|g\|_{L^p} \leq 1$, yields

$$\left|D^r(F_{i,a} - F_{i,b})(z)[v^1, \dots, v^r]\right| \leq (2e)^{r+1} C^r r! \prod_{q=1}^{r} \|v^q\|_E$$
$$\leq A R^r r! \prod_{q=1}^{r} \|v^q\|_E$$

for universal $A, R > 0$. Maximizing over a, b proves Equation (22); substituting into the opening norm inequality proves Equation (23). All constants are uniform over z, i, a, b, and the only factor $\sqrt{n}$ comes from combining the player blocks. □

C.5 Proof of the finite-difference chain rule

We now convert bounds on the derivatives $D^p\Psi$ and input differences $\Delta^r u^t$ into bounds on the output differences $\Delta^k\Psi(u^t)$. The key issue is combinatorial: a separate bound on the number of partitions would introduce an additional factorial, preventing the difference bound used in the interpolation argument.

Lemma 5.7 (Finite-difference chain rule)**.** *Let X, Y be finite-dimensional normed spaces, $(u^t)_{t\geq 1}$ a sequence in X, and $\Psi : X \to Y$ smooth. Suppose $a_r := \sup_{t\geq r+1} \|\Delta^r u^t\|/r!$ and $c_p := \sup_{x\in X} \|D^p\Psi(x)\|/p!$ are finite for all integers $r, p \geq 1$. Then, for every $k \geq 1$,*

$$\sup_{t\geq k+1} \frac{\left\|\Delta^k\Psi(u^t)\right\|}{k!} \leq [s^k]\sum_{p=1}^{k} c_p \left(\sum_{r=1}^{k} a_r s^r\right)^p. \tag{24}$$

Proof. The proof strategy is to express $\Delta^k\Psi(u^t)$ as an integral of a mixed derivative over $[0,1]^k$. Multilinear interpolation makes the inner derivatives averages of input differences, and counting

the resulting partitions gives the coefficient formula in Equation (24). Fix $t \geq k+1$ and define the multilinear interpolation $U : [0,1]^k \to X$ by

$$U(\theta) := \sum_{\varepsilon \in \{0,1\}^k} u^{t+|\varepsilon|} \prod_{j=1}^{k} \theta_j^{\varepsilon_j} (1-\theta_j)^{1-\varepsilon_j}.$$

This interpolant agrees with the trajectory values at the cube's vertices: $U(\varepsilon) = u^{t+|\varepsilon|}$. Repeated use of the fundamental theorem of calculus gives

$$\Delta^k \Psi(u^t) = \int_{[0,1]^k} \partial_1 \cdots \partial_k (\Psi \circ U)(\theta)\, \mathrm{d}\theta.$$

To track the input derivative associated with a block, write $\partial_B := \prod_{j \in B} \partial_j$ for nonempty $B \subseteq [k]$. Differentiating U shows that $\partial_B U(\theta)$ is a convex combination of vectors $\Delta^{|B|} u^{t+j}$ with $0 \leq j \leq k - |B|$. Since $t \geq k+1 \geq |B|+1$, all these indices lie in the range defining $a_{|B|}$, and hence

$$\|\partial_B U(\theta)\| \leq |B|! a_{|B|}.$$

The mixed Faà di Bruno formula is indexed by the partitions of $[k]$:

$$\partial_1 \cdots \partial_k (\Psi \circ U) = \sum_{\pi \in \Pi_k} D^{|\pi|} \Psi(U)[\partial_B U : B \in \pi].$$

Each block supplies an input derivative, while the number of blocks sets the outer derivative order. Consequently,

$$\frac{\left\|\Delta^k \Psi(u^t)\right\|}{k!} \leq \frac{1}{k!} \sum_{p=1}^{k} p! c_p \sum_{\substack{\pi \in \Pi_k \\ |\pi| = p}} \prod_{B \in \pi} |B|! a_{|B|}.$$

Multiplying the sum over unordered p-block partitions by $p!$ is the same as summing over ordered blocks. For a fixed ordered size tuple $(r_1, \ldots, r_p)$ with every $r_j \geq 1$ and $\sum_j r_j = k$, there are exactly $k!/(r_1! \cdots r_p!)$ ordered partitions of those sizes. After division by $k!$, their contribution is $c_p \prod_{j=1}^{p} a_{r_j}$. Summing over the ordered size tuples gives exactly the coefficient on the right-hand side of Equation (24) in Lemma 5.7. Taking the supremum over $t \geq k+1$ proves the claim. □

C.6 Proof of the centered-logit difference bound

To prove Lemma 5.8, we divide each order-k centered-logit difference by $k!$ and bound the resulting coefficients geometrically. The following scalar comparison will control the recursion obtained from the chain rule and centered-logit recurrence.

Lemma C.7 (Scalar coefficient comparison)**.** *Let $K, R > 0$, and let $A(s) = \sum_{k \geq 1} a_k s^k$ be a formal power series with nonnegative coefficients and $A(0) = 0$. If $A'(s) \preceq K/(1 - RA(s))$, then $a_k \leq C_R (C_R K)^k$ for a constant $C_R \geq 1$ depending only on R.*

Proof. The proof strategy is coefficientwise comparison with a scalar equality. Since each new coefficient depends only on earlier coefficients, induction suffices without assuming convergence of the formal series. Let $W(s) = \sum_{k \geq 1} w_k s^k$ be the unique formal solution of

$$W'(s) = \frac{K}{1 - RW(s)}, \qquad W(0) = 0.$$

The coefficient recursion of this equality matches the coefficientwise differential inequality in Lemma C.7 with equality, making W a candidate upper bound for A. Solving it gives

$$W(s) = \frac{1 - \sqrt{1 - 2KRs}}{R}.$$

We prove $A \preceq W$ coefficientwise. Suppose inductively that $a_j \leq w_j$ for $1 \leq j < k$. The coefficient of degree $k-1$ in

$$(1 - RA(s))^{-1} = \sum_{p \geq 0} (RA(s))^p$$

is a polynomial with nonnegative coefficients in $a_1, \dots, a_{k-1}$; it does not involve a_k because $A(0) = 0$. The same polynomial evaluated at $w_1, \dots, w_{k-1}$ is the degree-$(k-1)$ coefficient of $(1 - RW)^{-1}$. Therefore

$$ka_k = [s^{k-1}]A'(s) \le K[s^{k-1}](1 - RA)^{-1} \le K[s^{k-1}](1 - RW)^{-1} = kw_k.$$

The induction starts at $k = 1$, so $A \preceq W$. Finally, the binomial expansion of $\sqrt{1 - 2KRs}$ gives a constant $C_R \ge 1$, depending only on R, such that $w_k \le C_R(C_R K)^k$. This proves the lemma. □

We now apply the scalar comparison to the centered-logit differences.

Lemma 5.8 (High-order differences of the centered-logit state)**.** *For the trajectory z^t generated by the centered-logit recurrence, there are universal constants $C_z, R_z > 0$ such that, for every integer $k \ge 1$,*

$$\sup_{t \ge k+1} \left\|\Delta^k z^t\right\|_E \le C_z(R_z\sqrt{n}\,\eta)^k k!. \tag{25}$$

Proof. The proof strategy is to combine the derivative bounds for F, the finite-difference chain rule, and the centered-logit recurrence into the coefficient inequality of Lemma C.7. Define the factorially normalized centered-logit and loss differences, and their formal series, by

$$a_k := \sup_{t \ge k+1} \frac{\left\|\Delta^k z^t\right\|_E}{k!}, \qquad b_k := \sup_{t \ge k+1} \frac{\left\|\Delta^k F(z^t)\right\|_E}{k!},$$
$$A(s) := \sum_{k \ge 1} a_k s^k, \qquad B(s) := \sum_{k \ge 1} b_k s^k.$$

These coefficients are finite before any high-order estimate is used: the centered-logit recurrence gives $\|\Delta z^t\|_E \le 3\sqrt{n}\,\eta$, so $a_k \le 3\sqrt{n}\,\eta\, 2^{k-1}/k!$, while $\|F(z)\|_E \le \sqrt{n}$ gives $b_k \le 2^k\sqrt{n}/k!$. The series track factorially normalized differences, have nonnegative coefficients, and satisfy $A(0) = B(0) = 0$. By Lemmas 5.6 and 5.7, for universal $A_0, R > 0$,

$$B(s) \preceq A_0\sqrt{n}\,\frac{RA(s)}{1 - RA(s)}. \tag{41}$$

Indeed, for every $k \ge 1$,

$$b_k \le [s^k] \sum_{p=1}^{k} A_0\sqrt{n}\,R^p A(s)^p \le [s^k] A_0\sqrt{n} \sum_{p \ge 1} (RA(s))^p.$$

Since $\|F(z)\|_E \le \sqrt{n}$, the centered-logit recurrence gives $a_1 \le 3\sqrt{n}\,\eta$. For $k \ge 2$, applying Δ^{k-1} to the increment Δz^t gives

$$\Delta^k z^t = \eta\left(2\Delta^{k-1}F(z^t) - \Delta^{k-1}F(z^{t-1})\right).$$

For $t \ge k+1$, both starting indices t and $t-1$ lie in the range $s \ge k$ defining b_{k-1}. Dividing by $(k-1)!$ and taking the supremum therefore gives

$$ka_k \le 3\eta b_{k-1}, \qquad k \ge 2. \tag{42}$$

Let $M := \max\{1, A_0\}$. Coefficientwise,

$$1 + A_0\frac{RA}{1 - RA} = 1 + A_0 \sum_{p \ge 1} (RA)^p \preceq M \sum_{p \ge 0} (RA)^p = \frac{M}{1 - RA}.$$

Combining this bound with Equations (41) and (42) gives

$$A'(s) = a_1 + \sum_{k \ge 2} k a_k s^{k-1} \preceq 3\sqrt{n}\,\eta + 3\eta B(s)$$
$$\preceq 3\sqrt{n}\,\eta\left(1 + A_0\frac{RA(s)}{1 - RA(s)}\right) \tag{43}$$
$$\preceq \frac{3M\sqrt{n}\,\eta}{1 - RA(s)}. \tag{44}$$

Thus Equation (44) has the form required by Lemma C.7:

$$A'(s) \preceq \frac{K}{1 - RA(s)}, \qquad K := 3M\sqrt{n}\,\eta. \tag{45}$$

The series A has nonnegative coefficients, $A(0) = 0$, and $1 - RA(s)$ has constant coefficient one. Since R and M are universal, Lemma C.7 gives $a_k \le C_z(R_z\sqrt{n}\,\eta)^k$ for universal constants $C_z, R_z > 0$. Multiplying by $k!$ proves Equation (25) and completes the proof of Lemma 5.8. □

C.7 Proof of the weighted-gap derivative bound

The output $H_i(z, z')$ has $\binom{d_i}{2}$ coordinates. To retain logarithmic dependence on d_i in Equation (8), we sum squared derivatives using probability normalization rather than count these pairs.

Lemma 5.9 (Factorial derivatives of the weighted loss-gap map)**.** *For the weighted-gap map $H_i : E \times E \to \mathbb{R}^{\binom{d_i}{2}}$, equip the input space $E \times E$ with $\|(p, q)\| = (\|p\|_E^2 + \|q\|_E^2)^{1/2}$ and the output space with the Euclidean norm on unordered action pairs. Then, there are universal constants $A_H, R_H > 0$ such that, for every player i and integer $r \ge 0$,*

$$\sup_{z,z' \in E} \|D^r H_i(z, z')\| \le A_H R_H^r r!. \tag{26}$$

Proof. We first show that each derivative of a square-root probability is bounded by that square-root probability times a factorial bound. The trilinear map in Equation (48) then combines those factors with the loss gaps in F_i. Define the square-root probability map $\rho_i : E_i \to \mathbb{R}^{d_i}$ by $\rho_i(z_i) := \sqrt{\mathrm{softmax}(z_i)}$ coordinatewise, and write $\rho_{i,a} = e^{\psi_a}$, where the logarithm ψ_a is

$$\psi_a(z_i) = \frac{1}{2} z_{i,a} - \frac{1}{2} \log \sum_b e^{z_{i,b}}.$$

The logarithm separates the action coordinate from the common normalizer, whose derivatives are cumulants. For every direction p,

$$D\psi_a(z_i)[p] = \frac{1}{2}\left(p_a - \mathbb{E}_{\mathsf{A} \sim \mathrm{softmax}(z_i)} p_{\mathsf{A}}\right).$$

For $r \ge 2$, $D^r\psi_a[p^1, \dots, p^r]$ is $-1/2$ times the joint cumulant of $p^1_{\mathsf{A}}, \dots, p^r_{\mathsf{A}}$. Centering the arguments, applying Hoeffding's lemma, and then Lemma C.6 give universal constants C_ψ, R_ψ such that

$$|D^r\psi_a(z_i)[p^1, \dots, p^r]| \le C_\psi R_\psi^r r! \prod_{q=1}^r \mathrm{osc}(p^q), \qquad r \ge 1.$$

For $r \ge 1$, the exponential Faà di Bruno formula gives

$$D^r\rho_{i,a}(z_i)[p^1, \dots, p^r] = \rho_{i,a}(z_i) \sum_{\pi \in \Pi_r} \prod_{B \in \pi} D^{|B|}\psi_a(z_i)[p^q : q \in B].$$

Every term retains the same prefactor $\rho_{i,a}(z_i)$. Combining the derivative bound for ψ_a with

$$\sum_{\pi \in \Pi_r} \prod_{B \in \pi} |B|! \le r! 2^{r-1}$$

gives a universal constant $C_\rho > 0$ such that

$$|D^r\rho_{i,a}(z_i)[p^1, \dots, p^r]| \le \rho_{i,a}(z_i) C_\rho^r r! \prod_{q=1}^r \mathrm{osc}(p^q), \qquad r \ge 0. \tag{46}$$

The case $r = 0$ is immediate. Since $\sum_a \rho_{i,a}^2 = 1$, summing the squared coordinate bounds introduces no factor depending on d_i:

$$\left\|(D^r\rho_{i,a}[p^1, \dots, p^r])_a\right\|_2 \le C_\rho^r r! \prod_{q=1}^r \mathrm{osc}(p^q). \tag{47}$$

To control all pairs together, separate the two probability factors from the loss gap. Define the trilinear map $B_i : (\mathbb{R}^{d_i})^3 \to \mathbb{R}^{\binom{d_i}{2}}$ by

$$[B_i(r,s,w)]_{ab} := r_a s_b (w_b - w_a), \qquad a < b.$$

The map B_i forms weighted coordinate differences and satisfies a bound independent of d_i:

$$\|B_i(r,s,w)\|_2 \le \|r\|_2 \|s\|_2 \operatorname{osc}(w), \tag{48}$$

because the sum over $a < b$ is bounded by the sum over all ordered pairs. Moreover,

$$H_i(z,z') = B_i\big(\rho_i(z_i), \rho_i(z_i), F_i(z')\big).$$

Since $\operatorname{osc}(F_i(z')) \le 1$, the playerwise centered-loss derivative bound also holds at order zero after enlarging A. For labeled directions $\xi^\ell = (p^\ell, q^\ell)$, apply the multilinear Leibniz rule and group the directions assigned to the two ρ_i factors and the F_i factor. Combining Equations (47), (48), and (22) gives

$$\begin{aligned}\left\|D^r H_i(z,z')[\xi^1,\dots,\xi^r]\right\|_2 &\le \sum_{p+j+k=r} \frac{r!}{p!j!k!}(C_\rho^p p!)(C_\rho^j j!)(AR^k k!) \prod_{\ell=1}^r \left\|\xi^\ell\right\| \\ &\le A(2C_\rho + R)^r r! \prod_{\ell=1}^r \left\|\xi^\ell\right\|.\end{aligned}$$

This proves Equation (26) in Lemma 5.9. The identity $\|\rho_i\|_2 = 1$ and the bound in Equation (48) make the resulting constants independent of d_i. □

The proof of Lemma 5.5 is given in Section 5.4.

D FINITE-SUPPORT CUTOFF CONSTRUCTION

We prove Lemma 5.10 by multiplying the forward trajectory $(u^t)_{t\ge1}$ by a cutoff and setting the result to zero outside a finite interval. The sequence y is defined on all integer times, but need not equal u^t at every $1 \le t \le T$: the cutoff tapers the endpoints and can use the continuation beyond $T+1$. Its squared-norm errors are controlled by the lemma, so finite-difference interpolation (Lemma 5.4) can be applied.

D.1 PROOF OF THE FINITE-SUPPORT CUTOFF LEMMA

Lemma 5.10 (Finite-support cutoff). *Fix $B, C_0, R_0, \rho > 0$ and integers $m \ge 1$, $L \ge 2$. Let $(u^t)_{t\ge1}$ be a sequence in a finite-dimensional Euclidean space satisfying*

$$\sup_{t\ge1} \left\|u^t\right\| \le B, \qquad \sup_{t\ge k+2} \left\|\Delta^k u^t\right\| \le C_0 (R_0\rho)^k k! \quad (1 \le k \le m). \tag{27}$$

There is a constant $C_L > 0$, depending only on B, C_0, L, such that every integer horizon $T \ge 1$ admits a finitely supported sequence $y = (y^t)_{t\in\mathbb{Z}}$ with

$$\|\Delta y\|_{\ell_2}^2 \ge \textstyle\sum_{t=1}^T \left\|\Delta u^t\right\|^2 - C_L m, \tag{28}$$

$$\|y\|_{\ell_2}^2 \le \textstyle\sum_{t=1}^{T+1} \left\|u^t\right\|^2 + C_L m, \tag{29}$$

$$\|\Delta^m y\|_{\ell_2}^2 \le C_L (T+m)(2/L + R_0\rho m)^{2m}. \tag{30}$$

Proof. The proof strategy is to construct a cutoff whose jth differences are bounded by $(2/L)^j$ for $j \le m$, with a fixed averaging length $L \ge 2$. Repeated averaging gives these bounds and a transition over $O(Lm)$ indices, while placing the cutoff after initialization avoids indices where the high-order assumption does not apply. We implement this construction by convolution and bound the ℓ_1 norms of the resulting kernel differences. For finitely supported scalar sequences f, g on $\mathbb{Z}$, define

$$(f * g)^t := \sum_{s\in\mathbb{Z}} f^{t-s} g^s, \qquad \|f\|_{\ell_1} := \sum_{t\in\mathbb{Z}} |f^t|.$$

Convolution averages shifted copies when one factor has nonnegative entries summing to one; the ℓ_1 norm controls the size of kernel differences.

We begin with uniform averaging over an interval. For $S \subseteq \mathbb{Z}$, let $\mathbf{1}_S$ denote its indicator sequence, and define

$$b^t := \frac{1}{L}\mathbf{1}_{\{0,\dots,L-1\}}(t).$$

The sequence b averages over L consecutive indices. Let $\nu = b^{*m}$ be its m-fold convolution, with total mass one and support in $\{0, \dots, m(L-1)\}$. Since $\Delta(f * g) = (\Delta f) * g = f * (\Delta g)$, we can assign one difference to each of j distinct factors; the other factors retain unit ℓ_1 norm. Thus

$$\left\|\Delta^j \nu\right\|_{\ell_1} \le \|\Delta b\|_{\ell_1}^j \, \|b\|_{\ell_1}^{m-j} = \left(\frac{2}{L}\right)^j, \qquad 0 \le j \le m. \tag{49}$$

The inequality uses $\|f * g\|_{\ell_1} \le \|f\|_{\ell_1} \|g\|_{\ell_1}$. Each additional difference multiplies the bound by $2/L$, with no combinatorial coefficient; the support interval of ν has length $m(L-1)+1$.

First consider a horizon satisfying

$$T + 1 \ge 2m + 2 + m(L-1). \tag{50}$$

This leaves an interval where the cutoff equals one. Use the given values of u^t for $t \ge 1$ and extend arbitrarily to nonpositive indices; those values will be irrelevant.

We smooth the interval indicator after initialization and multiply the trajectory by the resulting cutoff. Define

$$\chi := \mathbf{1}_{\{2m+2,\dots,T+1\}} * \nu,$$
$$y^t := \begin{cases} \chi^t u^t, & 2m+2 \le t \le T+1+m(L-1), \\ 0, & \text{otherwise.} \end{cases}$$

The sequence y agrees with u wherever $\chi = 1$. In particular, $0 \le \chi^t \le 1$, y is finitely supported, and

$$\chi^t = 0, \qquad t \notin \{2m+2, \dots, T+1+m(L-1)\}, \tag{51}$$
$$\chi^t = 1, \qquad 2m+2+m(L-1) \le t \le T+1, \tag{52}$$
$$\sup_t \left|\Delta^j \chi^t\right| \le \left\|\Delta^j \nu\right\|_{\ell_1} \le (2/L)^j, \qquad 0 \le j \le m. \tag{53}$$

These properties identify the interval where $y^t = u^t$ and control $\Delta^j \chi^t$. To bound $\Delta^m y^t$, split differences between the cutoff χ^t and the trajectory u^t:

$$\Delta^m y^t = \sum_{j=0}^{m} \binom{m}{j} (\Delta^j \chi)^{t+m-j} \Delta^{m-j} u^t. \tag{54}$$

Each term assigns j differences to the cutoff and $m-j$ to the trajectory. We must check that each nonzero term uses $\Delta^r u^t$ only at indices $t \ge r+2$, where the assumed bound applies. Fix a nonzero summand with $j < m$ and set $r = m - j \ge 1$. The factor $(\Delta^j \chi)^{t+r}$ depends on indices $t+r, \dots, t+m$, at least one of which must lie in the cutoff's support. Since $\chi^s = 0$ for $s < 2m+2$, we obtain $t \ge m+2 \ge r+2$, as required by Equation (27). For $j = m$, a nonzero $(\Delta^m \chi)^t$ likewise forces $t \ge m+2$, so u^t belongs to the original sequence and is bounded by B.

Let $M := \max\{B, C_0\}$. Since $r! \le m^r$ for $0 \le r \le m$, Equation (27) in Lemma 5.10, together with Equations (53) and (54), gives

$$\sup_t \left\|\Delta^m y^t\right\| \le M \sum_{j=0}^{m} \binom{m}{j} (2/L)^j (R_0 \rho m)^{m-j} = M\left(\frac{2}{L} + R_0 \rho m\right)^m. \tag{55}$$

The factor $2/L$ comes from the differences of χ, and $R_0 \rho m$ comes from the differences of u. Since $\Delta^m y$ is nonzero at no more than $T + 1 + mL$ indices, enlarging a constant depending only on B, C_0, L gives

$$\|\Delta^m y\|_{\ell_2}^2 \le C_L (T+m)\left(\frac{2}{L} + R_0 \rho m\right)^{2m}. \tag{56}$$

Whenever both t and $t+1$ lie in the central interval in Equation (52), one has $\Delta y^t = \Delta u^t$. At most $2m + m(L-1) + 1 = O_L(m)$ target indices $t \in \{1, \dots, T\}$ are excluded. The additional right-boundary interval also contains $O_L(m)$ indices. Since $\|u^t\| \le B$ and $\|\Delta u^t\| \le 2B$,

$$\sum_{t=1}^{T} \left\|\Delta u^t\right\|^2 \le \|\Delta y\|_{\ell_2}^2 + C_L m, \tag{57}$$

$$\|y\|_{\ell_2}^2 \le \sum_{t=1}^{T+1} \left\|u^t\right\|^2 + C_L m. \tag{58}$$

The two comparisons involve at most $O_L(m)$ additional or excluded indices, giving an additive bound of $C_L m$ in each inequality. These are the first two conclusions of Lemma 5.10; Equation (56) supplies the third.

If the preceding horizon condition fails, then $T = O_L(m)$, so the bound $\|\Delta u^t\| \le 2B$ controls the entire difference sum. When Equation (50) fails, set $y^t = 0$ for all t. This discards all target differences, but $T \le (2+L)m + 1$, so

$$\sum_{t=1}^{T} \left\|\Delta u^t\right\|^2 \le 4B^2 T \le C_L m,$$

while the other two conclusions are immediate. This completes the proof. □

E Detailed Comparison with Daskalakis et al. (2021)

Both analyses reduce individual regret to high-order smoothness of self-play. We compare the estimates that permit a larger step size, using our notation n for the number of players and suppressing universal numerical constants.

E.1 The common objective

By Lemma 5.1, bounding the prediction-error variance sum by half the stability variance sum allows their contributions to regret to cancel. Regret is then bounded by $(\log d_i)/\eta$ plus $(2\eta/3)$ times the comparison error. The objective is a relative bound, not separate estimates of the two sums.

The central idea of Daskalakis et al. (2021, Section 4) is to obtain this relative comparison from the high-order smoothness of self-play, rather than from a separate bound on each round's prediction error. Their proof first bounds high-order loss differences and then returns to the first-order quantity in the RVU bound. Our proof follows the same strategy, but controls the weighted gaps h_i^t through the centered-logit recurrence and compares their finite differences in one fixed norm.

E.2 The analysis of Daskalakis et al. (2021)

Let $H := \lceil \log_2 T \rceil$ be the largest difference order. The scale comparison below concerns the regime $T \ge 4$ and $\eta \ge 1/T$ used in the absorption argument of Daskalakis et al. (2021); their proof treats smaller horizons and step sizes separately.

Obtaining a sufficiently small high-order term. The upward induction of Daskalakis et al. (2021) alternates between loss differences and strategy differences, using a boundedness chain rule for softmax-type maps. Their detailed estimate has the form

$$\left\|\Delta^k \ell_i^t\right\|_\infty \le (Cn\eta)^k k^{3k+1} \le H(Cn\eta H^3)^k, \qquad 1 \le k \le H, \quad 1 \le t \le T-k,$$

for a universal $C > 0$ when η is sufficiently small relative to $1/(nH)$ (Daskalakis et al., 2021, Lemma 4.4, Appendix B.4). The factor k^{3k+1} matters because the estimate must remain useful as k grows with $\log T$. To start and maintain the subsequent downward comparison, Daskalakis et al. (2021) choose $Cn\eta H^3$ at most a sufficiently small multiple of $H^{-1/2}$. This gives the sufficient high-order restriction $\eta \lesssim (nH^{7/2})^{-1}$ (Daskalakis et al., 2021, Appendix C.3, Equation (96)).

Returning from high orders to first differences. For one player, write $E_k := \sum_{t=1}^{T-k} \mathrm{Var}_{x_i^t}(\Delta^k \ell_i^t)$ for the weighted variance sum at order k, where $0 \le k \le H$. Daskalakis et al. (2021) propagate comparisons between E_{k+1} and E_k downward from $k = H-1$ to $k = 0$, keeping the comparison coefficient of order $1/H$. The high-order estimate starts this induction; the difficult part is that the weights x_i^t change with time. Even a fixed loss vector is therefore measured by different variance forms across rounds.

Daskalakis et al. (2021) handle this by comparing the moving weights with fixed weights on short time intervals, where Fourier analysis applies. If β denotes the adjacent-order comparison coefficient, their argument uses intervals of length $S \asymp \beta^{-3}$. Optimistic Hedge changes consecutive probability ratios by $1 + O(\eta)$, so the relative change across such an interval is $O(\eta S)$ when ηS is small. Requiring this error to be of order β gives

$$S \asymp \beta^{-3}, \qquad \eta S \lesssim \beta \quad \Longrightarrow \quad \eta \lesssim \beta^4 \asymp H^{-4}.$$

This is the second sufficient restriction, coming from the changing weights rather than the high-order loss estimate (Daskalakis et al., 2021, Lemma 4.7, Appendix C.2). The downward induction, including the initial and terminal weight adjustments, yields the variance comparison (Daskalakis et al., 2021, Lemma 4.2, Appendix C.3)

$$\sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^t - \ell_i^{t-1}) \le \frac{1}{2} \sum_{t=1}^{T} \mathrm{Var}_{x_i^t}(\ell_i^{t-1}) + O(H^5)$$

Their common choice $\eta \asymp (nH^4)^{-1}$ satisfies both restrictions. The variance error contributes $O(\eta H^5) = O(H/n)$ to regret, while the entropy term gives $O(n \log d_i H^4)$ for $d_i \ge 2$; a player with one action has zero regret. These are restrictions of the proof, not lower bounds on the performance of Optimistic Hedge.

E.3 Our approach: sharper differences in a fixed norm

Controlling the trajectory through its recurrence. Instead of closing separate induction estimates for losses and strategies, we use the centered-logit state z^t and centered loss map F. The finite-difference recurrence Equation (20) expresses $\Delta^k z^t$ as η times two order-$(k-1)$ differences of $F(z^t)$. Factorial derivative bounds for F, combined with the normed-space chain rule, turn this relation into recursive bounds on $\left\|\Delta^k z^t\right\|_E / k!$. The scalar coefficient comparison controls their growth geometrically; Lemma 5.8 states the resulting centered-logit estimate.

For $t \ge 2$, the weighted gaps satisfy $h_i^t = H_i(z^t, z^{t-1})$. Applying the same chain rule with the factorial derivative bound for H_i gives

$$\sup_{t \ge k+2} \left\|\Delta^k h_i^t\right\|_2 \le C(R\sqrt{n}\,\eta)^k k! \le C(R\sqrt{n}\,\eta k)^k.$$

Thus the useful difference order is $m \asymp (\sqrt{n}\eta)^{-1}$: choosing its proportionality constant small enough makes the bound exponentially small in m. The single factorial is essential to this scale. The factor $\sqrt{n}$ comes from combining player blocks in the product norm, while probability normalization in H_i avoids an action-count factor.

Comparing high and first differences directly. The choice of h_i^t also addresses the moving-weight difficulty. Its squared Euclidean norm equals $\mathrm{Var}_{x_i^t}(\ell_i^{t-1})$, and its first difference controls prediction error up to the $O(\eta^2)$ weight correction of Lemma 5.2. The probabilities are part of the vector being differentiated, rather than part of a norm that changes with time.

After Lemma 5.10 constructs a tapered, finitely supported sequence y from h_i^t, finite-difference interpolation (Lemma 5.4) gives

$$\|\Delta y\|_{\ell_2}^2 \le \theta \|y\|_{\ell_2}^2 + \theta^{1-m} \|\Delta^m y\|_{\ell_2}^2 .$$

Here θ can remain fixed, independently of m. There is no need to compare adjacent orders under changing weights or to freeze those weights on time intervals. The cutoff contributes $O(m)$ boundary error, and the high-order bound controls the interpolation term by $O((T+m)e^{-cm})$ after choosing the constants as in Section 5.5. This yields Lemma 5.3.

Why the admissible step size is larger. The two changes work together: factorial temporal bounds permit $m \asymp (\sqrt{n}\eta)^{-1}$, and fixed-norm interpolation transfers that control directly to first differences. Taking m of order $\log(T+2)$ permits the horizon-tuned constant step size $\eta_T \asymp [\sqrt{n}\log(T+2)]^{-1}$. The entropy term then gives the logarithmic regret bound in Corollary 4.2. The improvement is therefore not a new RVU inequality or the use of Fourier analysis alone, but sharper high-order estimates in a representation that preserves the relevant probability weights. The chain rule and scalar comparison isolate a reusable question about discrete recurrences; applying them elsewhere would require corresponding uniform derivative bounds.