LAST-ITERATE CONVERGENCE OF OPTIMISTIC GRADIENT DESCENT-ASCENT IN BEHAVIORAL FORM FOR EXTENSIVE-FORM GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper shows that a simple learning algorithm in the behavioral-form representation of extensive-form games (EFGs) achieves last-iterate convergence, without modifying the game or enforcing exploration. Existing last-iterate guarantees for EFGs are developed mainly in sequence form, while behavioral-form guarantees often rely on regularization or payoff perturbation. Our method applies simultaneous optimistic gradient descent-ascent (OGDA) with independent Euclidean projections at each information set, using ordinary counterfactual values. We show that, for every perfect-recall, two-player zero-sum EFG, any fixed step size below a positive threshold yields last-iterate convergence. Because this guarantee does not follow from the sequence-form analysis, our analysis instead compares the behavioral-form dynamics with sequence-form dilated OGDA and controls the additional correction term that ordinary counterfactual values introduce. We also provide a parameter-free variant that halves the step size until it is small enough and achieves last-iterate convergence for any initial step size. Experiments on benchmark games are consistent with last-iterate convergence.

1 INTRODUCTION

This paper shows that a simple learning algorithm achieves last-iterate convergence in the behavioral form of extensive-form games (EFGs), without modifying the game or enforcing exploration. EFGs model sequential interaction under imperfect information. To scale to large EFGs, practical solvers often use the *behavioral form*, in which a strategy specifies an action distribution at each information set. This local representation is compatible with function approximation, as demonstrated by methods such as Deep CFR (Brown et al., 2019). Counterfactual regret minimization (CFR) and its variants exploit this locality, updating each information set's distribution from counterfactual values (Zinkevich et al., 2007; Tammelin, 2014; Farina et al., 2019a).

These methods are no-regret, so in two-player zero-sum EFGs, the time average of the generated strategies approaches a Nash equilibrium (Zinkevich et al., 2007). The strategies played at individual iterations can nevertheless cycle (Bailey & Piliouras, 2018; Lee et al., 2021), and most regret-based methods therefore rely on iterate averaging to converge. Last-iterate convergence asks instead for the played strategy itself to approach an equilibrium. This guarantee matters when the final strategy is deployed directly, or when maintaining a separate average strategy is inconvenient.

Optimistic and extragradient methods provide guarantees of last-iterate convergence in two-player zero-sum normal-form games and monotone variational inequalities (Rakhlin & Sridharan, 2013; Daskalakis & Panageas, 2019; Mokhtari et al., 2020; Golowich et al., 2020; Cai et al., 2022; Gorbunov et al., 2022). For EFGs, however, the corresponding theory is primarily formulated in the *sequence form*, which converts the game into a bilinear saddle-point problem over tree-structured polytopes (von Stengel, 1996) and supports convergent optimistic methods (Wei et al., 2021; Lee et al., 2021). Our focus is instead the update of behavioral strategies from ordinary counterfactual values (Zinkevich et al., 2007), where each information set updates its own action distribution independently. Existing behavioral-form guarantees, however, rely on regularization or payoff perturbation (Kalogiannis & Farina, 2025; Meng et al., 2025), and the policy-gradient guarantees among them further require a lower bound on reach probabilities, that is, every information set must keep being reached with

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.

Table 1: Last-iterate guarantees for selected regularized behavioral-form methods and Behavioral OGDA in two-player zero-sum perfect-recall extensive-form games. The guarantee column specifies the equilibrium target and the convergence rate. The dagger marks a continuous-time guarantee.

Reference	Guarantee for last-iterate convergence	Requires regularization
Perolat et al. (2021)†	NE (asymptotic)	Yes
Sokota et al. (2023)	Regularized NE (linear)	Yes
Liu et al. (2023)	Regularized NE (asymptotic)	Yes
Kalogiannis & Farina (2025)	NE (polynomial)	Yes
Meng et al. (2025)	NE (asymptotic)	Yes
Yu et al. (2026)	NE (asymptotic)	Yes
This work	NE (asymptotic)	No

non-negligible probability (Kalogiannis & Farina, 2025). In summary, it remains open whether standard behavioral-form updates, applied independently at each information set using ordinary counterfactual values, can support last-iterate convergence under a fixed step size, without game regularization or a lower bound on current reach probabilities.

We answer this question affirmatively. We study simultaneous optimistic gradient descent-ascent (OGDA), applied directly at each information set. The update takes an optimistic gradient step on the ordinary counterfactual value and projects the result onto the information set's simplex in Euclidean distance. For every perfect-recall, two-player zero-sum EFG, we prove last-iterate convergence for every fixed step size at most a positive threshold, without regularizing the game or enforcing exploration.

This guarantee does not follow from the sequence-form analysis. In sequence form, each update depends on the already-updated strategies at its descendants. This dependence combines the local updates into a single potential that decreases toward an equilibrium. Our updates instead act on ordinary counterfactual values at the previous iterate. These independent updates no longer form that potential, which is the central difficulty. A further obstacle is that an information set can have zero own reach at an equilibrium while its strategy keeps moving. To establish convergence, we develop a descent analysis tailored to these independent updates, based on the reach probabilities at an equilibrium.

The fixed-step threshold below which last-iterate convergence holds depends on quantities associated with an equilibrium and is therefore unavailable to the algorithm before solving the game. We also provide a parameter-free variant that halves the step size until it is small enough and achieves last-iterate convergence, without knowledge of these equilibrium-dependent quantities. Experiments on five benchmark EFGs are consistent with last-iterate convergence for suitable fixed step sizes and with similar behavior for the adaptive variant without tuning the step size.

2 Related Literature

Last-iterate convergence in extensive-form games has been studied under two standard strategy representations, the sequence form and the behavioral form. The theory is most developed for the sequence form, which also provides the natural baseline for our analysis. A body of work develops learning methods in this representation, under both full and bandit feedback (Lee et al., 2021; Hoda et al., 2010; Farina et al., 2019c;b; 2021; Liu et al., 2023; Fiegel et al., 2023; Zhao et al., 2025). All of these methods update sequence-form strategies or apply the corresponding global proximal map. In contrast, our updates are independent Euclidean projections of ordinary counterfactual values onto each information set's simplex, an update rule these sequence-form guarantees do not cover.

Several behavioral-form methods obtain last-iterate guarantees by regularizing the game or perturbing its payoffs. The method of Sokota et al. (2023) converges only to an equilibrium of the regularized game, not to one of the original game. Liu et al. (2023) also study regularized local updates at information sets. For the original game, their guarantees are for best and average iterates. Their asymptotic last-iterate guarantee applies only to a perturbed EFG. The methods of Perolat et al. (2021), Meng et al. (2025), and Yu et al. (2026) target the original game's equilibrium by perturbing

the payoffs toward a progressively updated reference strategy. The policy-gradient method of Kalogiannis & Farina (2025) guarantees only approximate equilibrium under a lower bound on reach probabilities. Under bandit feedback, Fiegel et al. (2024) introduce LocalOMD, which updates behavioral policies locally at information sets using regularized losses. The method also admits a global OMD interpretation over the treeplex, while its sample-complexity guarantees concern time-averaged policies rather than the last iterate. A separate regularization-free approach is the reach-weighted Brown–von Neumann–Nash dynamics of Zhang & Stella (2026), whose EFG analysis gives last-iterate bounds for diminishing step sizes. Our method gives a fixed-step guarantee for independent projected OGDA updates, without requiring a lower bound on current reach probabilities. Table 1 compares Behavioral OGDA with selected regularized behavioral-form methods.

3 PRELIMINARIES

Extensive-form games. In this study, we consider a *two-player zero-sum extensive-form game*, described by a tuple $\langle \mathcal{H}, \mathcal{Z}, \mathcal{A}, P, \pi_c, \mathcal{I}, u \rangle$. Here, $\mathcal{H}$ is a finite set of histories that contains the empty history $\varnothing$, and $\mathcal{Z} \subset \mathcal{H}$ is the set of terminal histories. Each nonterminal history $h \in \mathcal{H} \setminus \mathcal{Z}$ is assigned a finite action set $\mathcal{A}(h)$, and appending an action $a \in \mathcal{A}(h)$ produces the successor history ha. We write $h \sqsubseteq h'$ when h is a prefix of h', that is, when h' is reached from h by appending zero or more actions. The player function $P : \mathcal{H} \setminus \mathcal{Z} \to \{1, 2, c\}$ specifies whether player 1, player 2, or chance acts at each nonterminal history, and $\pi_c(\cdot \mid h)$ is the fixed chance distribution over $\mathcal{A}(h)$. The pair $\mathcal{I} = (\mathcal{I}_1, \mathcal{I}_2)$ partitions each player's histories into *information sets*, and $u = (u_1, u_2)$ gives the terminal payoffs. Since the game is zero-sum, $u_1(z) = -u_2(z)$ for every terminal history $z \in \mathcal{Z}$.

An information set of player i collects the histories that player i cannot distinguish when choosing an action. All histories in the same information set therefore share a common action set, which we denote by $\mathcal{A}(I)$. Throughout the paper, we assume perfect recall (Kuhn, 1953). Under perfect recall, any two histories in the same information set of player i share the same earlier information sets and the same earlier actions of that player.

Behavioral strategies and reach probabilities. We represent strategies in *behavioral form*, which specifies one distribution over actions at each information set. For a finite nonempty set B, we denote its probability simplex by $\Delta(B) := \{p \in \mathbb{R}^B_{\geq 0} : \sum_{a \in B} p_a = 1\}$. A *behavioral strategy* of player i is a tuple $\pi_i = (\pi_{i,I})_{I \in \mathcal{I}_i}$ with $\pi_{i,I} \in \Delta(\mathcal{A}(I))$, and a *behavioral profile* is a pair $\pi = (\pi_1, \pi_2)$. We write $-i$ for the opponent of player i.

For histories $h \sqsubseteq z$, let $\rho^\pi(h, z)$ denote the product of all chance and behavioral action probabilities on the path strictly after h and up to z, with the convention $\rho^\pi(h, h) = 1$. The *reach probability* of a history is $\rho^\pi(h) := \rho^\pi(\varnothing, h)$. We split this product into the two factors used throughout the analysis. The *own reach probability* $\rho_i^{\pi_i}(h)$ is the product of player i's own action probabilities on the path to h, and the *counterfactual reach probability* $\rho_{-i,c}^{\pi}(h)$ is the product of the remaining chance and opponent action probabilities on that path. We define both factors directly as products along the path, rather than as quotients of the total reach probability. Under perfect recall, $\rho_i^{\pi_i}(h)$ takes a common value over all histories $h \in I$, which we denote by $\rho_i^{\pi_i}(I)$. We say that an information set I has zero own reach probability under π_i when $\rho_i^{\pi_i}(I) = 0$. For every information set $I \in \mathcal{I}_i$ and action $a \in \mathcal{A}(I)$, the *realization plan* $\mathcal{R}_i(\pi_i)_{Ia} := \rho_i^{\pi_i}(I)\, \pi_{i,I}(a)$ records the probability that player i's own actions reach I and then select a. We write $\mathcal{R}(\pi) := (\mathcal{R}_1(\pi_1), \mathcal{R}_2(\pi_2))$ for the realization pair induced by a behavioral profile π.

Nash equilibrium and exploitability. The expected utility of player i under a behavioral profile π is $U_i(\pi) = \sum_{z \in \mathcal{Z}} \rho^\pi(z) u_i(z)$. For a behavioral strategy π_i', we abbreviate $U_i(\pi_i', \pi_{-i}) := U_i((\pi_i', \pi_{-i}))$. The standard solution concept for two-player zero-sum games is the *Nash equilibrium*, a behavioral profile at which neither player can improve their own expected utility by deviating unilaterally. Formally, a behavioral profile π is a Nash equilibrium if

$$U_i(\pi_i, \pi_{-i}) \geq U_i(\pi_i', \pi_{-i}) \qquad \text{for every player } i \text{ and every behavioral strategy } \pi_i'.$$

We denote the set of behavioral Nash equilibria by $\Pi^\star$. To quantify how close a behavioral profile is to a Nash equilibrium, we use its *exploitability*

$$\mathrm{Expl}(\pi) := \sum_{i=1}^{2} \left[\max_{\pi_i'} U_i(\pi_i', \pi_{-i}) - U_i(\pi) \right].$$

The exploitability is non-negative, and it equals zero if and only if π is a Nash equilibrium.

Counterfactual values. Each information set is updated from a local oracle, the *counterfactual value* vector. For $I \in \mathcal{I}_i$ and $a \in \mathcal{A}(I)$, it is defined by

$$v_{i,I}(\pi)_a := \sum_{h \in I} \rho^{\pi}_{-i,c}(h) \sum_{\substack{z \in \mathcal{Z} \\ ha \sqsubseteq z}} \rho^{\pi}(ha, z) u_i(z). \tag{1}$$

The definition in (1) omits the factor $\rho_i^{\pi_i}(I)$ and fixes the action a at I, while all later behavioral probabilities, including player i's own, remain inside the continuation factor $\rho^\pi(ha, z)$. Hence, $v_{i,I}(\pi)$ is well defined even when I has zero own reach probability under the current profile.

Additional notation. For any integer $m \geq 1$ and any nonempty closed convex set $C \subseteq \mathbb{R}^m$, we let $\Pi_C(x) := \arg\min_{y \in C} \|y - x\|$ denote the Euclidean projection onto C. For any behavioral profile π, we let $\mathrm{dist}(\pi, \Pi^\star) := \inf_{\pi' \in \Pi^\star} \|\pi - \pi'\|$ denote its distance to the equilibrium set $\Pi^\star$.

4 Algorithm

This section presents *simultaneous optimistic behavioral gradient descent-ascent* (Behavioral OGDA), summarized in Algorithm 1. The method runs projected OGDA independently at each information set, and it treats the counterfactual value (1) as the local oracle at that information set.

Optimistic updates. OGDA reuses the previous oracle as a prediction of the current one. The method maintains two behavioral profiles π^t and $\bar{\pi}^t$, and updates both from a single counterfactual value evaluated at the previous profile π^{t-1}. At each iteration $t \geq 1$, the method first updates $\bar{\pi}^t$ and then updates π^t from that same value:

$$\boxed{\begin{aligned} \bar{\pi}^t_{i,I} &= \Pi_{\Delta(\mathcal{A}(I))} \left[\bar{\pi}^{t-1}_{i,I} + \eta v_{i,I}(\pi^{t-1}) \right], \\ \pi^t_{i,I} &= \Pi_{\Delta(\mathcal{A}(I))} \left[\bar{\pi}^{t}_{i,I} + \eta v_{i,I}(\pi^{t-1}) \right], \end{aligned}} \tag{2}$$

where $\eta > 0$ is the step size. Both sequences start from $\pi^0 = \bar{\pi}^0 = \pi^{\mathrm{init}}$.

Equivalent local form. Both updates in (2) evaluate the value vector $v_{i,I}(\pi^{t-1})$ at every information set. The definition in (1) is global, but the value vectors satisfy an equivalent recursion over the game tree, which is how Algorithm 1 evaluates them in a single bottom-up pass. For $I \in \mathcal{I}_i$ and $a \in \mathcal{A}(I)$, let $\mathcal{C}_i(Ia)$ be the next information sets of player i after action a, with no intervening information set of that player. For an opponent strategy τ_{-i}, let $\rho^{\tau_{-i}}_{-i,c}(z)$ denote the product of the opponent and chance action probabilities on the path to a terminal history z. Define the direct contribution at action a by

$$\ell_{i,Ia}(\tau_{-i}) := \sum_{\substack{h \in I,\ z \in \mathcal{Z},\ ha \sqsubseteq z, \\ P(g) \neq i \text{ for } ha \sqsubseteq g \sqsubset z}} \rho^{\tau_{-i}}_{-i,c}(z)\, u_i(z),$$

where the sum runs over the terminal histories reached from I through a before player i decides again. For an own strategy π_i, write $v_{i,I}(\pi_i, \tau_{-i}) := v_{i,I}((\pi_i, \tau_{-i}))$. The ordinary counterfactual value in (1) then admits the recursive representation

$$v_{i,I}(\pi_i, \tau_{-i})_a = \ell_{i,Ia}(\tau_{-i}) + \sum_{J \in \mathcal{C}_i(Ia)} \langle \pi_{i,J}, v_{i,J}(\pi_i, \tau_{-i}) \rangle. \tag{3}$$

At a leaf information set, the child set $\mathcal{C}_i(Ia)$ is empty, and the value at each action a equals its direct contribution $\ell_{i,Ia}(\tau_{-i})$. Evaluating (3) from the leaves to the root yields the value vectors used in both updates in (2). In (3), every continuation term uses the profile at which the value is evaluated, so no term depends on an updated strategy at a child information set.

Algorithm 1 Simultaneous optimistic behavioral gradient descent-ascent

Input: initial behavioral profile π^{init} and step size $\eta > 0$
Output: $\{\pi^t\}_{t\geq 0}$ and $\{\bar{\pi}^t\}_{t\geq 0}$
$\pi^0 \leftarrow \pi^{\text{init}}$ and $\bar{\pi}^0 \leftarrow \pi^{\text{init}}$
for $t = 1, 2, \ldots$ **do**
  $\{v_{i,I}(\pi^{t-1})\}_{i,I} \leftarrow \textsc{CounterfactualValues}(\pi^{t-1})$
  for all $i \in \{1, 2\}$ and $I \in \mathcal{I}_i$ **do**
    $\bar{\pi}^t_{i,I} \leftarrow \Pi_{\Delta(\mathcal{A}(I))}\left[\bar{\pi}^{t-1}_{i,I} + \eta v_{i,I}(\pi^{t-1})\right]$
  end for
  for all $i \in \{1, 2\}$ and $I \in \mathcal{I}_i$ **do**
    $\pi^t_{i,I} \leftarrow \Pi_{\Delta(\mathcal{A}(I))}\left[\bar{\pi}^t_{i,I} + \eta v_{i,I}(\pi^{t-1})\right]$
  end for
end for

5 Comparison with Sequence-Form Dilated OGDA

This section explains why the last-iterate analysis of Algorithm 1 cannot follow the potential argument of sequence-form dilated OGDA, even though the two methods share the same local Euclidean projection geometry. The two methods build their local updates differently, in the profile at which each evaluates its child values and in the presence of a child proximal cost. The absence of this proximal cost in Algorithm 1 is what blocks the potential argument.

5.1 How the Local Updates Differ

We first recall sequence-form dilated OGDA. Let Ω be the product of the two realization-plan polytopes, and let A be the sequence-form payoff matrix. For $Z = (x, y) \in \Omega$, define the saddle operator by $F(Z) := (-Ay, A^{\mathsf{T}}x)$. Let d be the dilated Euclidean distance-generating function obtained from the local function $w \mapsto \|w\|^2/2$ at every information set. For any two behavioral profiles r and $\bar{r}$, define the dilated Bregman divergence (Farina et al., 2019c) by

$$D_d\left(\mathcal{R}(r), \mathcal{R}(\bar{r})\right) = \frac{1}{2}\sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\rho_i^{r_i}(I)\left\|r_{i,I} - \bar{r}_{i,I}\right\|^2, \tag{4}$$

which weights each local movement by the current own reach. For every $t \geq 1$, sequence-form dilated OGDA computes the iterates (Farina et al., 2019c)

$$\begin{aligned} Q_t &:= \arg\min_{Z\in\Omega}\left\{\eta\left\langle F(P_{t-1}), Z\right\rangle + D_d(Z, Q_{t-1})\right\}, \\ P_t &:= \arg\min_{Z\in\Omega}\left\{\eta\left\langle F(P_{t-1}), Z\right\rangle + D_d(Z, Q_t)\right\}. \end{aligned} \tag{5}$$

Algorithm 1 computes neither minimization in (5). To read the sequence-form updates in behavioral coordinates, let σ^t and $\bar{\sigma}^t$ be behavioral representatives satisfying $\mathcal{R}(\sigma^t) = P_t$ and $\mathcal{R}(\bar{\sigma}^t) = Q_t$, respectively.

Like Algorithm 1, sequence-form dilated OGDA also decomposes exactly into one local proximal map per information set, but with a modified continuation value (Farina et al., 2019c). For an output policy π_i, proximal center $\bar{\pi}_i$, and opponent policy τ_{-i} in the linear term, the modified continuation value is defined by the recursion in (3), with each child value replaced by its own modified value and the child proximal cost subtracted:

$$\widehat{v}_{i,I}(\pi_i; \bar{\pi}_i, \tau_{-i})_a := \ell_{i,Ia}(\tau_{-i}) + \sum_{J\in\mathcal{C}_i(Ia)}\left[\left\langle \pi_{i,J}, \widehat{v}_{i,J}(\pi_i; \bar{\pi}_i, \tau_{-i})\right\rangle - \frac{1}{2\eta}\left\|\pi_{i,J} - \bar{\pi}_{i,J}\right\|^2\right]. \tag{6}$$

In the value-maximization convention of Algorithm 1, the updates in (5) are equivalent to the two local proximal maps

$$\begin{aligned} \bar{\sigma}^t_{i,I} &= \Pi_{\Delta(\mathcal{A}(I))}\left[\bar{\sigma}^{t-1}_{i,I} + \eta\widehat{v}_{i,I}(\bar{\sigma}^t_i; \bar{\sigma}^{t-1}_i, \sigma^{t-1}_{-i})\right], \\ \sigma^t_{i,I} &= \Pi_{\Delta(\mathcal{A}(I))}\left[\bar{\sigma}^t_{i,I} + \eta\widehat{v}_{i,I}(\sigma^t_i; \bar{\sigma}^t_i, \sigma^{t-1}_{-i})\right]. \end{aligned}$$

These updates differ from Algorithm 1 in two ways. First, they evaluate their child values at different profiles. Algorithm 1 evaluates (3) at a profile fixed before the local projections, namely π^{t-1} in the prediction step and π^t in the correction step, so a single bottom-up pass suffices. Sequence-form dilated OGDA instead evaluates every child value in (6) at the output own strategy $\bar{\sigma}_i^t$ or σ_i^t, so each child information set must be updated before its parent. Second, the modified continuation value (6) subtracts a child proximal cost $\frac{1}{2\eta}\left\|\pi_{i,J} - \bar{\pi}_{i,J}\right\|^2$ at each child information set, and the ordinary value (3) has no such term.

5.2 Why the Sequence-Form Analysis Does Not Transfer

Summing the local three-point identities and expanding (6) gives, for every $Z \in \Omega$, the global sequence-form inequality (Nemirovski, 2004; Farina et al., 2019c)

$$\begin{aligned} D_d(Z, Q_{t+1}) \le D_d(Z, Q_t) - \eta \left\langle F(P_t), P_t - Z \right\rangle \\ - D_d(P_t, Q_t) - D_d(Q_{t+1}, P_t) + r_t^{\mathrm{seq}}, \end{aligned} \tag{7}$$

where $r_t^{\mathrm{seq}} := \eta \left\langle F(P_t) - F(P_{t-1}), P_t - Q_{t+1} \right\rangle$. The two movement terms $D_d(P_t, Q_t)$ and $D_d(Q_{t+1}, P_t)$ are the crux of (7). Summing over t telescopes the potential $D_d(Z, Q_t)$ and leaves these two terms as the dissipation. For η small enough, their accumulated sum dominates the residual r_t^{seq}. The movement is then summable, and therefore vanishes. These two terms weight each local movement by the current own reach, as (4) shows. This weighting comes from the child proximal cost $\frac{1}{2\eta}\left\|\pi_{i,J} - \bar{\pi}_{i,J}\right\|^2$ that the modified continuation value (6) subtracts at each child information set. To see this, fix a behavioral comparator z_i and two behavioral strategies r_i and $\bar{r}_i$. At a child information set J, the local identity contributes the comparator weight $\rho_i^{z_i}(J)$. The child proximal cost in the parent identity contributes the remaining $\rho_i^{r_i}(J) - \rho_i^{z_i}(J)$, since it is multiplied by the updated action probabilities on the way to the root. Adding the two contributions gives the movement weighted by the current own reach,

$$\frac{1}{2}\rho_i^{z_i}(J)\left\|r_{i,J} - \bar{r}_{i,J}\right\|^2 + \frac{1}{2}\left(\rho_i^{r_i}(J) - \rho_i^{z_i}(J)\right)\left\|r_{i,J} - \bar{r}_{i,J}\right\|^2 = \frac{1}{2}\rho_i^{r_i}(J)\left\|r_{i,J} - \bar{r}_{i,J}\right\|^2.$$

Algorithm 1 instead uses the ordinary values in (3) at every local simplex, with no child proximal cost. Summing its local identities, each weighted by the own reach of a fixed comparator z, still yields an inequality of the same shape as (7). Its dissipation, however, is not the divergences $D_d(P_t, Q_t)$ and $D_d(Q_{t+1}, P_t)$. Each local movement is weighted only by the fixed comparator reach $\rho_i^{z_i}(I)$, never by the current own reach $\rho_i^{r_i}(I)$. The child proximal cost is exactly what would supply the missing weight $\rho_i^{r_i}(I) - \rho_i^{z_i}(I)$ and form these divergences. Without it, the dissipation controls the residual r_t^{seq} only weakly, since $\rho_i^{z_i}(I)$ can vanish at an information set that Algorithm 1 keeps updating. Thus, the two methods differ not in their local Euclidean projection geometry, but in how their local identities combine over the game tree.

6 Last-Iterate Convergence

In this section, we provide a last-iterate convergence result for Behavioral OGDA. We then introduce a variant with an adaptive step size that retains this guarantee without prior knowledge of the step-size threshold.

6.1 Main Theorem

As discussed in the previous section, the independent local updates of Behavioral OGDA do not assemble into a single potential that decreases toward an equilibrium. Theorem 6.1 shows that, perhaps surprisingly, these updates still enjoy last-iterate convergence to a Nash equilibrium of the original game, without regularizing the game or enforcing exploration.

Theorem 6.1 (Last-iterate convergence of Algorithm 1)**.** *There is a step-size threshold $\eta_0 > 0$ that depends only on the game. For every initial behavioral profile π^0 and every fixed $0 < \eta \le \eta_0$, there exists a Nash realization pair $Z^\star$ such that*

$$\lim_{t\to\infty} \operatorname{dist}(\pi^t, \Pi^\star) = \lim_{t\to\infty} \operatorname{dist}(\bar{\pi}^t, \Pi^\star) = 0,$$
$$\lim_{t\to\infty} \mathcal{R}(\pi^t) = \lim_{t\to\infty} \mathcal{R}(\bar{\pi}^t) = Z^\star.$$

The behavioral profiles approach the set of Nash equilibria, rather than a single equilibrium, because a realization plan does not determine the action probabilities at the information sets that the acting player reaches with probability zero.

6.2 Proof Sketch of Theorem 6.1

This section outlines the proof sketch of Theorem 6.1. The complete proofs of the theorem and the lemmas are provided in Appendix B. For Algorithm 1, set $Z^t := \mathcal{R}(\pi^t)$ and $\bar{Z}^t := \mathcal{R}(\bar{\pi}^t)$. We compare the iterates against a strictly complementary equilibrium of the sequence-form linear programs, which the Goldman–Tucker theorem provides (Goldman & Tucker, 1956). Let Z^{sc} be its realization pair, and let π^{sc} be a behavioral profile that realizes Z^{sc}. The own reach $\rho_i^{\pi_i^{\mathrm{sc}}}(I)$ of I under this comparator is the fixed comparator reach of Section 5.2. The comparator's dual slacks (Appendix B) are strictly positive exactly where the reach-weighted movement of Section 5.2 degenerates.

Define the reach-weighted energy and the duality gap

$$\begin{aligned} E_t &:= \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left(\left\| \pi_{i,I}^t - \bar{\pi}_{i,I}^t \right\|^2 + \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2 \right), \\ G_t &:= \left\langle F(Z^t), Z^t - Z^{\mathrm{sc}} \right\rangle. \end{aligned} \tag{8}$$

On the realization plans, the counterpart of r_t^{seq} in (7) is $r_t^{\mathrm{beh}} := \eta \left\langle F(Z^t) - F(Z^{t-1}), Z^t - \bar{Z}^{t+1} \right\rangle$. Weighting each local oracle variation by the fixed reach $\rho_i^{\pi_i^{\mathrm{sc}}}(I)$ does not reproduce r_t^{beh}. The correction records the difference:

$$\Delta_t^{\mathrm{beh}} := \eta \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\langle v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}), \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\rangle - r_t^{\mathrm{beh}}. \tag{9}$$

Summing the local three-point identities against Z^{sc} gives a descent inequality of the same shape as (7).

Lemma 6.2 (Descent inequality)**.** *Let $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ be generated by Algorithm 1 with a fixed step size $\eta > 0$. Then, for every $t \geq 1$,*

$$D_d(Z^{\mathrm{sc}}, \bar{Z}^{t+1}) \leq D_d(Z^{\mathrm{sc}}, \bar{Z}^t) - \eta G_t - \frac{1}{2} E_t + r_t^{\mathrm{beh}} + \Delta_t^{\mathrm{beh}}.$$

The residual and correction are controlled by a cumulative bound.

Lemma 6.3 (Cumulative error bound)**.** *Fix $0 < \bar{\eta} \leq 1$. There exist finite nonnegative constants C_0, C_E, C_G that depend only on the game tree, the maximum absolute payoff, and $\bar{\eta}$. For every $T \geq 2$ and every $0 < \eta \leq \bar{\eta}$,*

$$\sum_{t=1}^{T-1} \left| r_t^{\mathrm{beh}} + \Delta_t^{\mathrm{beh}} \right| \leq C_0 \eta^2 + \frac{C_E}{\rho_{\mathrm{sc}}} \eta \sum_{t=1}^{T-1} E_t + C_G \frac{\eta^3}{\delta} \sum_{t=1}^{T-1} G_t,$$

where ρ_{sc} is the minimum positive coordinate of Z^{sc} and δ is the minimum positive dual slack over the coordinates that vanish at Z^{sc} and have a terminal descendant. The term involving δ is omitted when no such coordinate exists.

Choose η_0 so that $\frac{C_E}{\rho_{\mathrm{sc}}} \eta \leq \frac{1}{4}$ and $C_G \frac{\eta^2}{\delta} \leq \frac{1}{2}$ for every $\eta \leq \eta_0$. Telescoping the descent inequality and inserting the cumulative bound then give, for every $T \geq 2$,

$$\frac{\eta}{2} \sum_{t=1}^{T-1} G_t + \frac{1}{4} \sum_{t=1}^{T-1} E_t \leq D_d(Z^{\mathrm{sc}}, \bar{Z}^1) + C_0 \eta^2.$$

The right-hand side is independent of T. Hence, $\sum_t E_t < \infty$ and $\sum_t G_t < \infty$.

The weight $\rho_i^{\pi_i^{\mathrm{sc}}}(I)$ in E_t can be zero at an information set that Algorithm 1 keeps updating, so $\sum_t E_t$ does not yet control the movement there. Induction over game-tree depth, driven by the dual slack δ, supplies the missing control. Every local movement then vanishes:

$$\left\| \pi^t - \bar{\pi}^t \right\| \to 0, \qquad \left\| \pi^t - \pi^{t-1} \right\| \to 0.$$

These limits make every accumulation point of $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ a behavioral Nash equilibrium. To pin down a single limit, we form a second potential relative to a fixed accumulation point in place of Z^{sc}. Its increments are summable, so the potential converges, and its limit along the subsequence converging to that point is zero. Both realization-plan sequences therefore converge to one common Nash pair $Z^\star$.

6.3 Adaptive Behavioral OGDA

Theorem 6.1 guarantees last-iterate convergence only when the step size is at most the threshold η_0. As shown in Appendix B, this threshold is determined by the minimum own reach and the minimum dual slack of a strictly complementary equilibrium. Both are properties of an equilibrium and are unavailable before solving the game. To remove this requirement, we introduce Adaptive Behavioral OGDA, a variant of Behavioral OGDA that requires no knowledge of either quantity. The variant monitors two normalized cumulative sums along the run. It halves the step size whenever their maximum exceeds a constant budget. Both sums are computable from quantities that Algorithm 1 already evaluates.

Set $\bar{\Phi} := |\mathcal{I}_1| + |\mathcal{I}_2|$ and $L := |\mathcal{Z}| \max_{z\in\mathcal{Z}} |u_1(z)|$. For a run of Algorithm 1 with step size η, the test after iteration T is

$$\max \left\{ \begin{array}{l} \dfrac{1}{\eta} \displaystyle\sum_{t=1}^{T} \sum_{i=1}^{2} \sum_{I\in\mathcal{I}_i} \left[\eta^2 \left\langle v_{i,I}(\pi^{t-1}), \pi_{i,I}^{t-1} - \bar{\pi}_{i,I}^{t-1} \right\rangle^2 + \left\| \bar{\pi}_{i,I}^{t} - \pi_{i,I}^{t-1} \right\|^2 \right], \\ 11L\eta \displaystyle\sum_{t=1}^{T} \sum_{i=1}^{2} \sum_{s} \left(\prod_{(I,a)\in s} \pi_{i,I}^{t-1}(a) - \prod_{(I,a)\in s} \bar{\pi}_{i,I}^{t-1}(a) \right)^2 \end{array} \right\} \le 5\bar{\Phi}, \tag{10}$$

where s ranges over the nonempty sequences of player i, and each product runs over the information set–action pairs along s. Every quantity in (10) is an explicit function of the iterates, the number of information sets, the number of terminal histories, and the maximum absolute payoff.

Adaptive Behavioral OGDA starts from an initial step size η_1. It evaluates one counterfactual-value oracle at every inner iteration. It runs the two projection updates of Algorithm 1 with the current step size. After each iteration, it applies the test (10). When the test fails, it discards both update candidates. It then halves the step size and restarts from the last profile evaluated by the oracle. We index the global outputs by a single counter n that advances by one at each update step across all restarts. At index n, the outputs π_{out}^n and $\bar{\pi}_{\mathrm{out}}^n$ are the updated iterate when the test passes, and the profile from which the algorithm restarts when the test fails. The complete pseudocode is Algorithm 2 in Appendix C.

Theorem 6.4 shows that this variant retains the last-iterate convergence guarantee of Theorem 6.1.

Theorem 6.4 (Convergence of Algorithm 2)**.** *For every initial behavioral profile π^{init} and every initial step size $\eta_1 > 0$, Algorithm 2 eventually stops halving the step size. There exists a Nash realization pair $Z^\star$ such that its global output sequences satisfy*

$$\lim_{n\to\infty} \mathrm{dist}(\pi_{\mathrm{out}}^n, \Pi^\star) = \lim_{n\to\infty} \mathrm{dist}(\bar{\pi}_{\mathrm{out}}^n, \Pi^\star) = 0,$$
$$\lim_{n\to\infty} \mathcal{R}(\pi_{\mathrm{out}}^n) = \lim_{n\to\infty} \mathcal{R}(\bar{\pi}_{\mathrm{out}}^n) = Z^\star.$$

The proof is provided in section C.

7 Experiments

In this section, we evaluate Behavioral OGDA and Adaptive Behavioral OGDA against the sequence-form baseline of dilated OGDA with the Euclidean local regularizer (Farina et al., 2019c; Lee et al., 2021). To isolate the effect of the optimistic prediction, we also include Behavioral GDA, which removes it from Behavioral OGDA. The three behavioral methods apply an independent Euclidean projection at each information set. We implement all four methods in LiteEFG (Liu et al., 2024) using the same counterfactual-value oracle.

Our experiments focus on five two-player zero-sum perfect-recall extensive-form games: Kuhn poker, Leduc poker, four- and five-card Goofspiel, and four-sided Liars' Dice. All runs start from the uniform

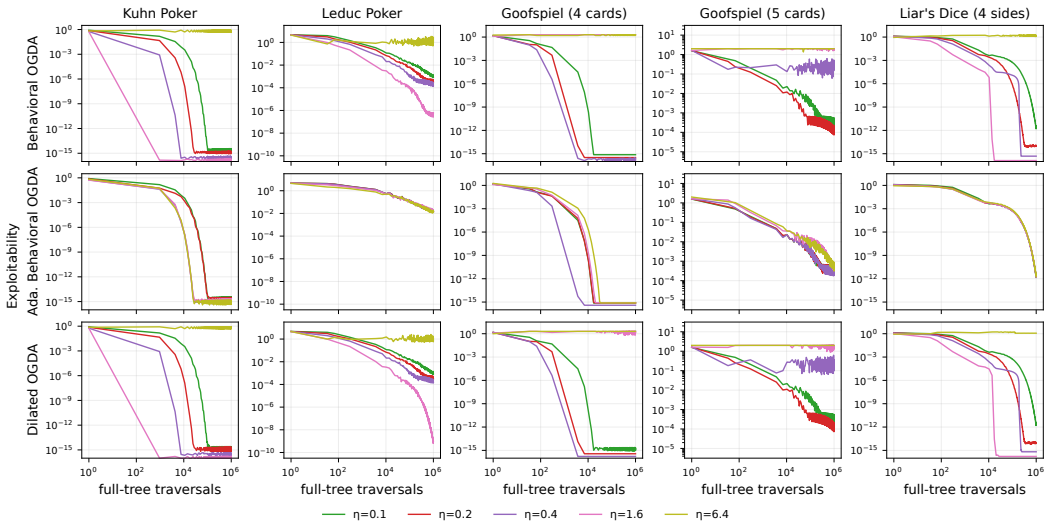

Figure 1: Exploitability of the played profile at the current iterate against the number of full-tree counterfactual-value evaluations. The rows show Behavioral OGDA, Adaptive Behavioral OGDA, and dilated OGDA, respectively. The five curves correspond to $\eta \in \{1/10, 1/5, 2/5, 8/5, 32/5\}$. For Adaptive Behavioral OGDA, η is the initial step size.

behavioral profile and evaluate counterfactual values over the entire game tree. Hence, each run is deterministic. Since every method performs one oracle evaluation per iteration, we use the number of oracle evaluations as the computational budget and run each method for 10^6 evaluations. We sweep nine step sizes from $\eta = 1/40$ to $\eta = 32/5$. At each checkpoint, we measure the exploitability $\mathrm{Expl}(\pi^t)$ of the played profile. Figure 1 reports five representative step sizes for the three optimistic methods, and Appendix D reports the full sweep together with the Behavioral GDA results.

As Figure 1 shows, Behavioral OGDA and dilated OGDA follow similar trajectories across the five games. With suitable step sizes, their exploitability appears to approach zero in Kuhn poker, four-card Goofspiel, and four-sided Liars' Dice. It also decreases in Leduc poker and five-card Goofspiel, though more slowly, with dilated OGDA slightly ahead. Adaptive Behavioral OGDA generally decreases exploitability more slowly than the fixed-step methods at favorable step sizes. However, its exploitability continues to decrease even from large initial step sizes for which the fixed-step methods stall. Behavioral GDA, without the optimistic prediction, does not show the same trend toward zero (Appendix D). Overall, the behavior of Behavioral OGDA is consistent with Theorem 6.1, though these finite runs do not determine a convergence rate.

8 Conclusion

We establish last-iterate convergence for simultaneous optimistic gradient descent-ascent applied directly to the behavioral simplexes of perfect-recall, two-player zero-sum extensive-form games. For a sufficiently small fixed step size, both behavioral sequences approach the set of Nash equilibria, and their realization plans converge to a single Nash realization pair. The adaptive variant eventually stops halving the step size without evaluating any equilibrium-dependent quantity, and its global outputs retain these guarantees. These guarantees require neither game regularization nor a lower bound on current reach probabilities. The analysis gives no convergence rate and does not cover alternating updates. Finite-time rates, alternating variants, and updates from sampled counterfactual values are natural directions for future work.

AI use statement

We used generative AI tools for two purposes. First, we used them to aid and polish the writing, improving the prose and the clarity of exposition in the statements and proofs. Second, we used them

to assist in developing and checking the proofs of our theoretical results. We did not use generative AI tools for research ideation, for writing the experiment code, or to draft sections of the paper. The authors verified all definitions, theorems, and proofs, and take full responsibility for their correctness. The experiment code was written and checked by the authors. We have reviewed all AI-assisted text and proofs. We take responsibility for the final content of this work, including all text, claims, and artifacts.

Reproducibility statement

The main text defines the game model, the algorithm, and the explicit step-size threshold under which the guarantees hold. Complete proofs of the fixed-step and adaptive results are given in sections B and C, together with the notation and constants that they use. Section 7 specifies the benchmark games, the per-iteration oracle budget, the step sizes, and the evaluation metric, and section D reports the full step-size sweep. Source code to reproduce all experiments is included in the supplementary material.

References

James P. Bailey and Georgios Piliouras. Multiplicative weights update in zero-sum games. In *EC*, pp. 321–338, 2018.

Noam Brown, Adam Lerer, Sam Gross, and Tuomas Sandholm. Deep counterfactual regret minimization. In *ICML*, pp. 793–802, 2019.

Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Tight last-iterate convergence of the extragradient method for constrained monotone variational inequalities. *arXiv preprint arXiv:2204.09228*, 2022.

Constantinos Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *ITCS*, pp. 27:1–27:18, 2019.

Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. Stable-predictive optimistic counterfactual regret minimization. In *ICML*, pp. 1853–1862, 2019a.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Online convex optimization for sequential decision processes and extensive-form games. In *AAAI*, pp. 1917–1925, 2019b.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. In *NeurIPS*, 2019c.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive Blackwell approachability: Connecting regret matching and mirror descent. In *AAAI*, pp. 5363–5371, 2021.

Côme Fiegel, Pierre Ménard, Tadashi Kozuno, Rémi Munos, Vianney Perchet, and Michal Valko. Adapting to game trees in zero-sum imperfect information games. In *ICML*, pp. 10093–10135, 2023.

Côme Fiegel, Pierre Ménard, Tadashi Kozuno, Rémi Munos, Vianney Perchet, and Michal Valko. Local and adaptive mirror descents in extensive-form games. In *NeurIPS*, 2024.

A. J. Goldman and A. W. Tucker. Theory of linear programming. In Harold W. Kuhn and Albert W. Tucker (eds.), *Linear Inequalities and Related Systems*, volume 38 of *Annals of Mathematics Studies*, pp. 53–97. Princeton University Press, 1956.

Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *NeurIPS*, pp. 20766–20778, 2020.

Eduard Gorbunov, Adrien Taylor, and Gauthier Gidel. Last-iterate convergence of optimistic gradient method for monotone variational inequalities. In *NeurIPS*, 2022.

Samid Hoda, Andrew Gilpin, Javier Peña, and Tuomas Sandholm. Smoothing techniques for computing Nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2): 494–512, 2010.

Fivos Kalogiannis and Gabriele Farina. Policy gradient methods converge globally in imperfect-information extensive-form games. In *NeurIPS*, 2025.

Harold W. Kuhn. Extensive games and the problem of information. In Harold W. Kuhn and Albert W. Tucker (eds.), *Contributions to the Theory of Games II*, volume 28 of *Annals of Mathematics Studies*. Princeton University Press, 1953.

Chung-Wei Lee, Christian Kroer, and Haipeng Luo. Last-iterate convergence in extensive-form games. In *NeurIPS*, 2021.

Mingyang Liu, Asuman Ozdaglar, Tiancheng Yu, and Kaiqing Zhang. The power of regularization in solving extensive-form games. In *ICLR*, 2023.

Mingyang Liu, Gabriele Farina, and Asuman Ozdaglar. LiteEFG: An efficient Python library for solving extensive-form games. *arXiv preprint arXiv:2407.20351*, 2024.

Linjian Meng, Tianpei Yang, Youzhi Zhang, Zhenxing Ge, Shangdong Yang, Tianyu Ding, Wenbin Li, Bo An, and Yang Gao. Efficient last-iterate convergence in solving extensive-form games. In *NeurIPS*, 2025.

Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *AISTATS*, pp. 1497–1507, 2020.

Arkadi Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex–concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

Julien Perolat, Remi Munos, Jean-Baptiste Lespiau, Shayegan Omidshafiei, Mark Rowland, Pedro Ortega, Neil Burch, Thomas Anthony, David Balduzzi, Bart De Vylder, Georgios Piliouras, Marc Lanctot, and Karl Tuyls. From Poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *ICML*, pp. 8525–8535, 2021.

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *NeurIPS*, 2013.

Samuel Sokota, Ryan D'Orazio, J. Zico Kolter, Nicolas Loizou, Marc Lanctot, Ioannis Mitliagkas, Noam Brown, and Christian Kroer. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. In *ICLR*, 2023.

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

Bernhard von Stengel. Efficient computation of behavior strategies. *Games and Economic Behavior*, 14(2):220–246, 1996.

Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *ICLR*, 2021.

Eason Yu, Tzu Hao Liu, Clément L. Canonne, Yunke Wang, Chang Xu, Nguyen H. Tran, and Stefano V. Albrecht. NashPG: A policy gradient method with iteratively refined regularization for finding Nash equilibria. *Transactions on Machine Learning Research*, 2026. URL https://arxiv.org/abs/2510.18183.

Tuo Zhang and Leonardo Stella. Teaching an old dynamics new tricks: Regularization-free last-iterate convergence in zero-sum games via BNN dynamics. In *AAMAS*, pp. 2327–2336, 2026.

Canzhe Zhao, Yutian Cheng, Jing Dong, Baoxiang Wang, and Shuai Li. Learning imperfect information extensive-form games with last-iterate convergence under bandit feedback. In *ICML*, pp. 77405–77440, 2025.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In *NeurIPS*, pp. 1729–1736, 2007.

A NOTATIONS

We collect the notation used throughout the proofs, grouped by role.

Sequences and realization plans. Throughout the appendix, we index coordinates by one player's sequences. For player i, let Σ_i consist of the empty sequence $\varnothing$ and every sequence ending with an action at one of player i's information sets. Perfect recall gives each $I \in \mathcal{I}_i$ a unique parent $p(I) \in \Sigma_i$. The sequence obtained by taking $a \in \mathcal{A}(I)$ is denoted Ia. Several information sets may share one parent. With this sequence notation, the child set introduced in (3) is

$$\mathcal{C}_i(s) := \{I \in \mathcal{I}_i : p(I) = s\}.$$

The realization plan $\mathcal{R}_i(r_i)$ satisfies the recursion

$$\mathcal{R}_{i,\varnothing}(r_i) = 1, \qquad \mathcal{R}_{i,Ia}(r_i) = \mathcal{R}_{i,p(I)}(r_i)\, r_{i,I}(a). \tag{11}$$

Hence, for every $I \in \mathcal{I}_i$, the parent coordinate is the own reach $\mathcal{R}_{i,p(I)}(r_i) = \rho_i^{r_i}(I)$. We write

$$Z_i^t := \mathcal{R}_i(\pi_i^t), \qquad \bar{Z}_i^t := \mathcal{R}_i(\bar{\pi}_i^t).$$

We also use sequence-form coordinates, with vectors $x \in \mathbb{R}^{\Sigma_1}$ and $y \in \mathbb{R}^{\Sigma_2}$ indexed by the sequences, including the empty ones.

Sequence-form linear programs and strict complementarity. These fix the comparator equilibrium and the strict-complementarity gap δ that the last-iterate argument relies on. Chance probabilities and terminal utilities are absorbed into a finite matrix A so that player 1's payoff is $x^\mathsf{T} A y$ and player 2 maximizes its negative. The corresponding maximin and minimax linear programs are

$$\begin{aligned} &\max_{x,\psi} \quad f^\mathsf{T}\psi \quad \text{subject to} \quad Ex = e, \quad x \geq 0, \quad F^\mathsf{T}\psi \leq A^\mathsf{T} x, \\ &\min_{y,u} \quad e^\mathsf{T} u \quad \text{subject to} \quad Fy = f, \quad y \geq 0, \quad E^\mathsf{T} u \geq Ay, \end{aligned} \tag{12}$$

where the flow potentials u and ψ are unrestricted, and redundant equality rows are allowed. Here, E and F are the sequence-form constraint matrices of players 1 and 2, and e and f are their right-hand sides. A nonnegative vector satisfies $Ex = e$ exactly when it is a realization plan of player 1, and $Fy = f$ plays the same role for player 2. Fix a strictly complementary primal–dual solution $(x^{\mathrm{sc}}, y^{\mathrm{sc}}, u^{\mathrm{sc}}, \psi^{\mathrm{sc}})$ of (12), and let π^{sc} be a behavioral equilibrium whose realization is its primal pair $(x^{\mathrm{sc}}, y^{\mathrm{sc}})$. We write $Z_i^{\mathrm{sc}} := \mathcal{R}_i(\pi_i^{\mathrm{sc}})$. Define the nonnegative dual slacks

$$s_x = E^\mathsf{T} u^{\mathrm{sc}} - Ay^{\mathrm{sc}} \geq 0, \qquad s_y = A^\mathsf{T} x^{\mathrm{sc}} - F^\mathsf{T}\psi^{\mathrm{sc}} \geq 0, \tag{13}$$

with

$$\langle s_x, x^{\mathrm{sc}} \rangle = 0, \qquad \langle s_y, y^{\mathrm{sc}} \rangle = 0.$$

By the Goldman–Tucker theorem, we can choose an optimal solution such that

$$x_s^{\mathrm{sc}} + (s_x)_s > 0, \qquad y_q^{\mathrm{sc}} + (s_y)_q > 0$$

for every nonempty coordinate (Goldman & Tucker, 1956). Empty coordinates are fixed at one and require no dual slack. Every nonempty sequence that is the parent of an information set has a terminal descendant. Define

$$\delta := \min \left\{ (s_x)_s, (s_y)_q : \begin{array}{l} \text{the corresponding equilibrium coordinate is zero,} \\ \text{and the sequence has a terminal descendant} \end{array} \right\} > 0,$$

with the convention that estimates involving δ below are omitted if the set is empty. The minimum positive coordinate of Z^{sc} is

$$\rho_{\mathrm{sc}} = \min\{Z_{i,s}^{\mathrm{sc}} : Z_{i,s}^{\mathrm{sc}} > 0\} > 0.$$

Behavioral updates, movements, and energy. These are the algorithm-side quantities the descent argument tracks. For every $t \geq 1$, shifting the first relation in (2) by one iteration writes the updates of Algorithm 1 as the predictor and corrector relations

$$\pi^t_{i,I} = \Pi_{\Delta(\mathcal{A}(I))}\left[\bar{\pi}^t_{i,I} + \eta v_{i,I}(\pi^{t-1})\right], \qquad \bar{\pi}^{t+1}_{i,I} = \Pi_{\Delta(\mathcal{A}(I))}\left[\bar{\pi}^t_{i,I} + \eta v_{i,I}(\pi^t)\right] \quad (t \geq 1). \tag{14}$$

For every $t \geq 0$ and player i, define the vector

$$b^t_i := (\bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I})_{I \in \mathcal{I}_i}. \tag{15}$$

Define the local inner products

$$K^t_{i,I} := \eta \left\langle v_{i,I}(\pi^{t-1}), \pi^t_{i,I} - \bar{\pi}^t_{i,I} \right\rangle, \qquad R^t_{i,I} := \eta \left\langle v_{i,I}(\pi^{t-1}), \bar{\pi}^t_{i,I} - \pi^{t-1}_{i,I} \right\rangle. \tag{16}$$

Define the reach-weighted energy

$$\Phi_t := \frac{1}{2} \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\| \bar{\pi}^t_{i,I} - \pi^{\mathrm{sc}}_{i,I} \right\|^2. \tag{17}$$

Lift operator. The lift carries a local vector to sequence-form coordinates. It bridges the local updates and the realization plans. For local vectors $h = (h_I)_I$, fixed ancestor weights $\mu = (\mu_I)_I$, and descendant behavior r_i, define $\mathcal{L}_i(h; \mu, r_i) \in \mathbb{R}^{\Sigma_i}$ recursively by

$$\mathcal{L}_{i,\varnothing} = 0, \qquad \mathcal{L}_{i,Ia} = \mu_I h_I(a) + \mathcal{L}_{i,p(I)} r_{i,I}(a). \tag{18}$$

If the own path to s passes through information sets $I_1, \ldots, I_k$ and uses actions $a_1, \ldots, a_k$, then repeated substitution gives the path expansion

$$\mathcal{L}_{i,s}(h; \mu, r_i) = \sum_{j=1}^{k} \mu_{I_j} h_{I_j}(a_j) \prod_{\ell=j+1}^{k} r_{i,I_\ell}(a_\ell). \tag{19}$$

Thus, (19) contains every product with one factor from h at an ancestor information set and behavioral probabilities at its descendants. Set

$$\Gamma^t_i := \mathcal{L}_i(b^t_i; (\rho_i^{\pi_i^{\mathrm{sc}}}(I))_{I \in \mathcal{I}_i}, \pi^t_i). \tag{20}$$

Game-tree quantities and the movement aggregate. These quantities are fixed by the game. The movement aggregate $\mathfrak{M}$ sums coordinate changes over the resulting depth grading. For an integer $k \geq 0$, let $\Sigma_i(k)$ collect the sequences in Σ_i with exactly k own actions, and let $\mathcal{I}_i(k) := \{I \in \mathcal{I}_i : p(I) \in \Sigma_i(k)\}$. Let

$$B_k := \max_i \max_{s \in \Sigma_i(k)} |\mathcal{C}_i(s)|, \qquad N_k := \sum_{i=1}^{2} |\mathcal{I}_i(k)|.$$

Empty maxima are zero. Let d be the largest number of own actions on any sequence. The payoff bound is

$$L := |\mathcal{Z}| \max_{z \in \mathcal{Z}} |u_1(z)|. \tag{21}$$

Every realization-plan difference has zero empty coordinate, as does Γ^t_i. The initialization $\bar{\pi}^0 = \pi^0$ gives $Z^0_i = \bar{Z}^0_i$, while $\bar{Z}^1_i - Z^0_i$ need not vanish. Define the sum of squared coordinate changes at depth k by

$$\mathfrak{M}_{t,k} := \sum_{i=1}^{2} \sum_{s \in \Sigma_i(k)} \left[(Z^t_{i,s} - \bar{Z}^t_{i,s})^2 + (\bar{Z}^{t+1}_{i,s} - Z^t_{i,s})^2 \right], \qquad \mathfrak{M}_t := \sum_{k=1}^{d} \mathfrak{M}_{t,k}.$$

B Proofs for Theorem 6.1

Let C_E and C_G be the constants of lemma 6.3 for $\bar{\eta} = 1$. Define the step-size threshold

$$\eta_0 := \min \left\{ 1, \frac{\rho_{\mathrm{sc}}}{4C_E}, \sqrt{\frac{\delta}{2C_G}} \right\} > 0. \tag{22}$$

An entry with a vanishing denominator is omitted from the minimum. The last entry is also omitted when the set defining δ is empty.

B.1 Proof of Theorem 6.1

Proof of Theorem 6.1. Recall from (8) the reach-weighted energy

$$E_t := \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left(\left\| \pi_{i,I}^t - \bar{\pi}_{i,I}^t \right\|^2 + \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2 \right)$$

and the duality gap $G_t := \langle F(Z^t), Z^t - Z^{\mathrm{sc}} \rangle$. Fix any initial behavioral profile π^0 and any $0 < \eta \leq \eta_0$. By lemma B.7, it suffices to verify the summability of the local predictor and corrector terms and the realization-plan movements.

First, lemma B.4 and the nonnegativity of E_t and G_t give

$$\sum_{t=1}^{\infty} E_t < \infty, \qquad \sum_{t=1}^{\infty} G_t < \infty.$$

Since $\eta \leq \eta_0 \leq 1$, lemma B.6 applies with $\bar{\eta} = 1$. Summing its bound over t gives $\sum_{t=1}^{\infty} \mathfrak{M}_t < \infty$. Consequently, lemma B.1 gives

$$\sum_{t=1}^{\infty} \sum_{i=1}^{2} \left\| Z_i^t - Z_i^{t-1} \right\|^2 < \infty.$$

Second, lemma B.5 applies with $\bar{\eta} = 1$. The bound and the summability of E_t and G_t give

$$\sum_{t=1}^{\infty} \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \left[(K_{i,I}^t)^2 + \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2 \right] < \infty.$$

Thus, both hypotheses of lemma B.7 hold. Therefore, lemma B.7 gives the two limits in theorem 6.1. □

Lemma B.1 (Realization-plan movement bound)**.** *For every $t \geq 1$, the realization-plan differences satisfy*

$$Z_i^t - Z_i^{t-1} = (Z_i^t - \bar{Z}_i^t) + (\bar{Z}_i^t - Z_i^{t-1}). \tag{23}$$

For every $I \in \mathcal{I}_i$ and $a \in \mathcal{A}(I)$, they satisfy the recursions

$$\begin{aligned} Z_{i,Ia}^t - \bar{Z}_{i,Ia}^t &= \bar{Z}_{i,p(I)}^t (\pi_{i,I}^t - \bar{\pi}_{i,I}^t)(a) + (Z_{i,p(I)}^t - \bar{Z}_{i,p(I)}^t) \pi_{i,I}^t(a), \\ \bar{Z}_{i,Ia}^{t+1} - Z_{i,Ia}^t &= Z_{i,p(I)}^t (\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t)(a) + (\bar{Z}_{i,p(I)}^{t+1} - Z_{i,p(I)}^t) \bar{\pi}_{i,I}^{t+1}(a). \end{aligned} \tag{24}$$

Consequently, for every $T \geq 2$,

$$\begin{aligned} \sum_{i=1}^{2} \left\| Z_i^t - Z_i^{t-1} \right\|^2 &\leq 2 \sum_{i=1}^{2} \left\| Z_i^t - \bar{Z}_i^t \right\|^2 + 2 \sum_{i=1}^{2} \left\| \bar{Z}_i^t - Z_i^{t-1} \right\|^2 \leq 2\mathfrak{M}_t + 2\mathfrak{M}_{t-1}, \\ \sum_{t=1}^{T-1} \sum_{i=1}^{2} \left\| Z_i^t - Z_i^{t-1} \right\|^2 &\leq 4 \sum_{t=1}^{T-1} \mathfrak{M}_t + 2\mathfrak{M}_0. \end{aligned} \tag{25}$$

Proof of Lemma B.1. The identity (23) telescopes. Direct product expansion of the realization coordinates gives the recursions (24). By (23) and $2uv \leq u^2 + v^2$, the pointwise bound in (25) holds. Each squared coordinate change is counted exactly once in $\mathfrak{M}_t$, which gives the summed bound. □

Lemma B.2 (Layer product estimate)**.** *If a family satisfies*

$$\zeta_{Ia} = \omega_I u_I(a) + \zeta_{p(I)} \theta_I(a)$$

with real numbers ω_I and $\theta_I \in \Delta(\mathcal{A}(I))$, then

$$\sum_{i} \sum_{Ia \in \Sigma_i(k+1)} \zeta_{i,Ia}^2 \leq 2 \sum_{i} \sum_{I \in \mathcal{I}_i(k)} \omega_{i,I}^2 \left\| u_{i,I} \right\|^2 + 2B_k \sum_{i} \sum_{s \in \Sigma_i(k)} \zeta_{i,s}^2. \tag{26}$$

Proof of Lemma B.2. The elementary bound $(\omega u + z\theta)^2 \le 2\omega^2 u^2 + 2z^2\theta^2$, the simplex bound $\|\theta_I\|^2 \le 1$, and the reuse of each parent coordinate by at most B_k child information sets give the claim. □

Lemma B.3 (Initial movement bounds)**.** *The initial local and realization-plan movements satisfy*

$$\mathfrak{B}_0 := \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \left\| \bar{\pi}^1_{i,I} - \pi^0_{i,I} \right\|^2 \le B_\partial \eta^2, \qquad \mathfrak{M}_0 \le M_\partial \eta^2, \tag{27}$$

where

$$B_\partial := \bar{\Phi} L^2, \qquad M_\partial := d\,(|\Sigma_1| + |\Sigma_2|)\, B_\partial.$$

Proof of Lemma B.3. The definition of the counterfactual values gives $\left\| v_{i,I}(\pi^0) \right\| \le L$. Hence, projection nonexpansiveness gives $\left\| \bar{\pi}^1_{i,I} - \pi^0_{i,I} \right\| \le \eta L$. For each nonempty sequence, telescoping its product gives a squared realization-coordinate difference of at most d times the sum of $\left\| \bar{\pi}^1_{i,I} - \pi^0_{i,I} \right\|^2$ along that sequence. Summing this bound over the sequence coordinates gives (27). □

Lemma B.4 (Summed descent inequality)**.** *Let $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ be generated by Algorithm 1 from any initial behavioral profile π^0 with a fixed $0 < \eta \le \eta_0$. Let C_0 be the constant of lemma 6.3 for $\bar{\eta} = 1$. Then, for every $T \ge 2$,*

$$\boxed{\frac{\eta}{2} \sum_{t=1}^{T-1} G_t + \frac{1}{4} \sum_{t=1}^{T-1} E_t \le \Phi_1 + C_0 \eta^2.}$$

B.2 Proof of Lemma B.4

Proof of Lemma B.4. Fix $T \ge 2$. By (22), every $0 < \eta \le \eta_0$ satisfies

$$\eta \le 1, \qquad \frac{C_E}{\rho_{\mathrm{sc}}} \eta \le \frac{1}{4}, \qquad C_G \frac{\eta^2}{\delta} \le \frac{1}{2},$$

where the last condition is omitted when the set defining δ is empty.

Since $\eta \le 1$, lemma 6.3 applies with $\bar{\eta} = 1$. Summing lemma 6.2 from $t = 1$ to $T - 1$ and applying lemma 6.3 give

$$\begin{aligned}
&\eta \sum_{t=1}^{T-1} G_t + \frac{1}{2} \sum_{t=1}^{T-1} E_t \\
&\quad \le \Phi_1 - \Phi_T + \sum_{t=1}^{T-1} \left| r^{\mathrm{beh}}_t + \Delta^{\mathrm{beh}}_t \right| \\
&\quad \le \Phi_1 - \Phi_T + C_0 \eta^2 + \frac{C_E}{\rho_{\mathrm{sc}}} \eta \sum_{t=1}^{T-1} E_t + C_G \frac{\eta^3}{\delta} \sum_{t=1}^{T-1} G_t.
\end{aligned}$$

The terms involving δ are omitted when the set defining δ is empty. Therefore, the step-size conditions above, the nonnegativity of E_t and G_t, and $\Phi_T \ge 0$ imply

$$\begin{aligned}
\frac{\eta}{2} \sum_{t=1}^{T-1} G_t + \frac{1}{4} \sum_{t=1}^{T-1} E_t &\le \left(\eta - C_G \frac{\eta^3}{\delta} \right) \sum_{t=1}^{T-1} G_t + \left(\frac{1}{2} - \frac{C_E}{\rho_{\mathrm{sc}}} \eta \right) \sum_{t=1}^{T-1} E_t \\
&\le \Phi_1 + C_0 \eta^2.
\end{aligned}$$

This proves Lemma B.4. □

Lemma B.5 (Summed predictor and corrector bound)**.** *Fix $0 < \bar{\eta} \le 1$. There exist finite nonnegative constants J_0, J_E, J_G that depend only on the game tree, the maximum absolute payoff, and $\bar{\eta}$. For every $T \ge 2$ and every $0 < \eta \le \bar{\eta}$,*

$$\sum_{t=1}^{T-1} \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \left[(K^t_{i,I})^2 + \left\| \bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I} \right\|^2 \right] \le J_0 \eta^3 + \frac{J_E}{\rho_{\mathrm{sc}}} \eta^2 \sum_{t=1}^{T-1} E_t + J_G \frac{\eta^4}{\delta} \sum_{t=1}^{T-1} G_t.$$

The term involving δ is omitted when the set defining δ is empty.

For information sets $I, J \in \mathcal{I}_i$, write $J \succ I$ when J is a strict own descendant of I, that is, when J is reached from I after one of I's own actions. For $I \in \mathcal{I}_i$, define its continuation height recursively by

$$h(I) = 0 \quad \text{if } I \text{ has no strict own descendant,}$$

$$h(I) = 1 + \max\{h(J) : J \text{ is an immediate own child below an action of } I\}$$

otherwise. Hence,

$$J \succ I \quad \Longrightarrow \quad h(J) < h(I). \tag{28}$$

Let $h_{\max} := 0$ when $\mathcal{I}_1 \cup \mathcal{I}_2 = \varnothing$. Otherwise, let $h_{\max} := \max_{i,I} h(I)$. For every horizon $T \geq 2$, define

$$P_{i,I}(T) := \sum_{t=1}^{T-1} (K_{i,I}^t)^2, \qquad Q_{i,I}(T) := \sum_{t=1}^{T-1} \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2.$$

For every player i and information set $I \in \mathcal{I}_i$, let

$$n_{i,I} := |\{J \in \mathcal{I}_i : J \succ I\}|.$$

B.3 Proof of Lemma B.5

Proof of Lemma B.5. Fix $T \geq 2$ and $0 < \eta \leq \bar{\eta}$. We first establish the local bounds that enter the height recursion. Projection nonexpansiveness, (2), and the oracle norm bound in Lemma B.17 imply

$$\left\| \pi_{i,I}^t - \bar{\pi}_{i,I}^t \right\| \leq L\eta, \qquad \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\| \leq 2L\eta. \tag{29}$$

The counterfactual value at an internal information set changes because the opponent changes and because the player's own descendants change. The latter change involves both $\pi_{i,I}^t - \bar{\pi}_{i,I}^t$ and $\bar{\pi}_{i,I}^t - \pi_{i,I}^{t-1}$. Therefore, we bound the sums of $(K_{i,I}^t)^2$ and $\left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2$ jointly. We next separate the oracle variation into opponent and own-strategy contributions. The opponent-variation bound in Lemma B.17 gives

$$\left\| v_{i,I}(\pi_i^t, \pi_{-i}^t) - v_{i,I}(\pi_i^t, \pi_{-i}^{t-1}) \right\| \leq L \left\| Z_{-i}^t - Z_{-i}^{t-1} \right\|.$$

For the own change, the local identity in Lemma B.18 gives, for each action of I, a sum over strict descendants with coefficients in $[0, 1]$. The Euclidean square of all action components is therefore bounded by $n_{i,I}$ times the sum of the squared descendant contributions. We have

$$\eta \left\langle v_{i,J}(\pi^{t-1}), \pi_{i,J}^t - \pi_{i,J}^{t-1} \right\rangle = K_{i,J}^t + R_{i,J}^t$$

and, by the oracle norm bound in Lemma B.17,

$$(R_{i,J}^t)^2 \leq \eta^2 L^2 \left\| \bar{\pi}_{i,J}^t - \pi_{i,J}^{t-1} \right\|^2, \tag{30}$$

Shifting the summation index gives

$$\sum_{t=1}^{T-1} \left\| \bar{\pi}_{i,J}^t - \pi_{i,J}^{t-1} \right\|^2 = \left\| \bar{\pi}_{i,J}^1 - \pi_{i,J}^0 \right\|^2 + \sum_{t=1}^{T-2} \left\| \bar{\pi}_{i,J}^{t+1} - \pi_{i,J}^t \right\|^2 \leq \left\| \bar{\pi}_{i,J}^1 - \pi_{i,J}^0 \right\|^2 + Q_{i,J}(T). \tag{31}$$

Define

$$\beta := 40 \left(\sqrt{2} + \bar{\eta} L \right)^2 + 1.$$

Combining lemmas B.8 and B.9 with (31) and (25) gives

$$\begin{aligned} P_{i,I}(T) + Q_{i,I}(T) \leq\ & 4\sqrt{2} L^3 \eta^3 + 8\beta L^2 \eta^2 \sum_{t=1}^{T-1} \mathfrak{M}_t \\ & + 4\beta n_{i,I} \sum_{J \succ I} \left[P_{i,J}(T) + L^2 \eta^2 Q_{i,J}(T) \right] \\ & + 4\beta L^2 \eta^2 \left(\mathfrak{M}_0 + n_{i,I} \sum_{J \succ I} \left\| \bar{\pi}_{i,J}^1 - \pi_{i,J}^0 \right\|^2 \right). \end{aligned} \tag{32}$$

We now aggregate (32) by continuation height. Let

$$\mathcal{H}_h := \{(i, I) : h(I) = h\},$$

and define the cumulative sum through height h by

$$S_h(T) := \sum_{(i,I):h(I)\le h} \big(P_{i,I}(T) + Q_{i,I}(T)\big), \qquad S_{-1}(T) := 0.$$

For $h \ge 0$, set

$$\mathfrak{a}_h := 4\sqrt{2}L^3|\mathcal{H}_h| + 4\beta\bar{\eta}L^2\left(|\mathcal{H}_h|M_\partial + L^2\sum_{(i,I)\in\mathcal{H}_h} n_{i,I}^2\right), \qquad \mathfrak{b}_h := 8\beta L^2|\mathcal{H}_h|,$$

$$\Lambda_h := 4\beta\left(1 + \bar{\eta}^2L^2\right)\sum_{(i,I)\in\mathcal{H}_h} n_{i,I}.$$

By (28), every $P_{i,J}$ and $Q_{i,J}$ on the right of (32) belongs to $S_{h-1}(T)$ when $h(I) = h$. By (27), the last line of (32) is at most

$$4\beta L^2\left(M_\partial + n_{i,I}^2L^2\right)\eta^4.$$

Since $\eta \le \bar{\eta}$, summing (32) over height h gives

$$\boxed{S_h(T) \le \mathfrak{a}_h\eta^3 + \mathfrak{b}_h\eta^2\sum_{t=1}^{T-1}\mathfrak{M}_t + (1 + \Lambda_h)S_{h-1}(T).}$$

No Q at height h occurs on the right. Therefore, no term at height h must be moved to the left-hand side.

Define horizon-independent recursive constants

$$\mathcal{X}_{-1} = \mathcal{Y}_{-1} = 0, \qquad \mathcal{X}_h = \mathfrak{a}_h + (1 + \Lambda_h)\mathcal{X}_{h-1}, \qquad \mathcal{Y}_h = \mathfrak{b}_h + (1 + \Lambda_h)\mathcal{Y}_{h-1}.$$

At height zero, the descendant sums in (32) are empty. Induction on h using the recurrence for $S_h(T)$ therefore gives

$$S_h(T) \le \mathcal{X}_h\eta^3 + \mathcal{Y}_h\eta^2\sum_{t=1}^{T-1}\mathfrak{M}_t.$$

Substituting the bound of lemma B.6 into the preceding bound at $h = h_{\max}$ gives

$$\boxed{\begin{aligned}\sum_{i,I}\sum_{t=1}^{T-1}\left[(K_{i,I}^t)^2 + \left\|\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t\right\|^2\right] \le{}& \mathcal{X}_{h_{\max}}\eta^3 \\ &+ \frac{\mathcal{Y}_{h_{\max}}\widehat{C}_d^M}{\rho_{\mathrm{sc}}}\eta^2\sum_{t=1}^{T-1}E_t \\ &+ \mathcal{Y}_{h_{\max}}\widehat{B}_d^M\frac{\eta^4}{\delta}\sum_{t=1}^{T-1}G_t.\end{aligned}}$$

Every constant is independent of the horizon T. Abbreviate

$$J_0 := \mathcal{X}_{h_{\max}}, \qquad J_E := \mathcal{Y}_{h_{\max}}\widehat{C}_d^M, \qquad J_G := \mathcal{Y}_{h_{\max}}\widehat{B}_d^M. \tag{33}$$

With these definitions, the displayed bound for the total predictor and corrector sum is precisely the assertion of Lemma B.5. □

Lemma B.6 (Movement bound). *Fix $0 < \bar{\eta} \le 1$. There exist finite nonnegative constants $\widehat{C}_d^M$ and $\widehat{B}_d^M$ that depend only on the game tree, the maximum absolute payoff, and $\bar{\eta}$. For every $t \ge 1$ and every $0 < \eta \le \bar{\eta}$,*

$$\mathfrak{M}_t \le \frac{\widehat{C}_d^M}{\rho_{\mathrm{sc}}}E_t + \widehat{B}_d^M\eta^2\frac{G_t}{\delta}.$$

The term involving δ is omitted when the set defining δ is empty.

B.4 Proof of Lemma B.6

Proof of Lemma B.6. Fix $t \geq 1$ and $0 < \eta \leq \bar{\eta}$. Define

$$C_0^M = B_0^M = 0, \qquad C_{k+1}^M = \alpha_k C_k^M + 2, \qquad B_{k+1}^M = \alpha_k B_k^M + \gamma_k.$$

Induction on k using lemma B.10 gives

$$\mathfrak{M}_{t,k} \leq \frac{C_k^M}{\rho_{\mathrm{sc}}} E_t + B_k^M \eta^2 \frac{G_t}{\delta}.$$

Set

$$\widehat{C}_d^M := \sum_{k=1}^{d} C_k^M, \qquad \widehat{B}_d^M := \sum_{k=1}^{d} B_k^M.$$

Summing the preceding inequality over $k = 1, \ldots, d$ proves the claimed pointwise bound. Both $\widehat{C}_d^M$ and $\widehat{B}_d^M$ are built from L, the counts B_k and N_k, and $\bar{\eta}$ by the recursions above. There is no negative power of η, no δ^{-2}, and no lower bound on a current reach. Finally, summing the pointwise bound from $t = 1$ to $T - 1$ gives

$$\sum_{t<T} \mathfrak{M}_t \leq \frac{\widehat{C}_d^M}{\rho_{\mathrm{sc}}} \sum_{t<T} E_t + \widehat{B}_d^M \frac{\eta^2}{\delta} \sum_{t<T} G_t.$$

□

Lemma B.7 (Convergence from summable movements)**.** *Let $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ be generated by Algorithm 1 from any initial behavioral profile with a fixed $\eta > 0$. Suppose that*

$$\sum_{t=1}^{\infty} \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \left[(K_{i,I}^t)^2 + \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2 \right] < \infty, \qquad \sum_{t=1}^{\infty} \sum_{i=1}^{2} \left\| Z_i^t - Z_i^{t-1} \right\|^2 < \infty.$$

Then, there exists a Nash realization pair $Z^\star$ such that

$$\lim_{t \to \infty} \operatorname{dist}(\pi^t, \Pi^\star) = \lim_{t \to \infty} \operatorname{dist}(\bar{\pi}^t, \Pi^\star) = 0,$$
$$\lim_{t \to \infty} \mathcal{R}(\pi^t) = \lim_{t \to \infty} \mathcal{R}(\bar{\pi}^t) = Z^\star.$$

B.5 Proof of Lemma B.7

Proof of Lemma B.7. We first show that every local predictor and corrector movement vanishes. For every $t \geq 1$, the predictor relation in (14) is $\pi_{i,I}^t = \Pi_{\Delta(\mathcal{A}(I))}[\bar{\pi}_{i,I}^t + \eta v_{i,I}(\pi^{t-1})]$. Projection optimality for this relation, tested against $\bar{\pi}_{i,I}^t \in \Delta(\mathcal{A}(I))$, gives

$$K_{i,I}^t = \eta \left\langle v_{i,I}(\pi^{t-1}), \pi_{i,I}^t - \bar{\pi}_{i,I}^t \right\rangle \geq \left\| \pi_{i,I}^t - \bar{\pi}_{i,I}^t \right\|^2 \geq 0.$$

By the assumed summability, $K_{i,I}^t \to 0$ and $\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \to 0$ for every fixed local simplex. The preceding projection bound then gives

$$\pi_{i,I}^t - \bar{\pi}_{i,I}^t \to 0, \qquad \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \to 0, \qquad \bar{\pi}_{i,I}^t - \pi_{i,I}^{t-1} \to 0.$$

Hence,

$$\pi_{i,I}^t - \pi_{i,I}^{t-1} = (\pi_{i,I}^t - \bar{\pi}_{i,I}^t) + (\bar{\pi}_{i,I}^t - \pi_{i,I}^{t-1}) \to 0$$

for every player and every information set, including those for which $\rho_i^{\pi_i^{\mathrm{sc}}}(I) = 0$.

These limits verify the hypotheses of lemma B.11. Therefore, every accumulation point of $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ is a behavioral Nash equilibrium. The product of behavioral simplexes is compact. If either sequence did not approach $\Pi^\star$, it would have a subsequence at a positive distance from $\Pi^\star$. Compactness would give that subsequence an accumulation point outside $\Pi^\star$, which would contradict lemma B.11. Hence, both sequences approach $\Pi^\star$.

We next prove error summability after replacing π^{sc} by another equilibrium. Let $\widetilde{\pi}$ be any behavioral Nash equilibrium and $\widetilde{Z}_i = \mathcal{R}_i(\widetilde{\pi}_i)$. Set $\widetilde{\Gamma}_i^t := \mathcal{L}_i(b_i^t; (\rho_i^{\widetilde{\pi}_i}(I))_{I\in\mathcal{I}_i}, \pi_i^t)$. Every realization coordinate lies in $[0,1]$. Hence, part (b) of lemma B.13 and the assumed summability give

$$\left\|\widetilde{\Gamma}^t\right\|^2 \le C_\Gamma \sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\left\|\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t\right\|^2, \qquad \sum_{t=1}^{\infty}\left\|\widetilde{\Gamma}^t\right\|^2 < \infty.$$

Define

$$\widetilde{\mathrm{Err}}_t := \eta \sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\rho_i^{\widetilde{\pi}_i}(I)\left\langle v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}), \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t\right\rangle.$$

Applying lemma B.12 with $w_{i,I} = \rho_i^{\widetilde{\pi}_i}(I)$ gives

$$\widetilde{\mathrm{Err}}_t = \sum_{i,J}\widetilde{\Gamma}_{i,p(J)}^t K_{i,J}^t + \sum_{i,J}\widetilde{\Gamma}_{i,p(J)}^t R_{i,J}^t + \widetilde{C}_t,$$

with

$$\widetilde{C}_t = \eta\left[(\widetilde{\Gamma}_1^t)^{\mathsf{T}} A(Z_2^t - Z_2^{t-1}) - (Z_1^t - Z_1^{t-1})^{\mathsf{T}} A\widetilde{\Gamma}_2^t\right].$$

We bound the three terms by separate $\ell^2\times\ell^2$ estimates. Let $\nu := \max_i \max_{s\in\Sigma_i}|\mathcal{C}_i(s)|$ be the largest number of information sets sharing one parent. First,

$$\sum_{t,J}|\widetilde{\Gamma}_{i,p(J)}^t K_{i,J}^t| \le \left(\nu\sum_t\left\|\widetilde{\Gamma}_i^t\right\|^2\right)^{1/2}\left(\sum_{t,J}(K_{i,J}^t)^2\right)^{1/2} < \infty.$$

Second, by (30), (31), and the assumed summability,

$$\sum_{i,J,t}(R_{i,J}^t)^2 \le \eta^2 L^2 \sum_{i,J,t}\left\|\bar{\pi}_{i,J}^t - \pi_{i,J}^{t-1}\right\|^2 < \infty,$$

and hence

$$\sum_{t,J}|\widetilde{\Gamma}_{i,p(J)}^t R_{i,J}^t| \le \left(\nu\sum_t\left\|\widetilde{\Gamma}_i^t\right\|^2\right)^{1/2}\left(\sum_{t,J}(R_{i,J}^t)^2\right)^{1/2} < \infty.$$

Third, the assumed square summability bounds $\sum_t\left\|Z_i^t - Z_i^{t-1}\right\|^2$, so

$$\sum_t|\widetilde{C}_t| \le \eta\|A\|_{\mathrm{op}}\Bigg[\left(\sum_t\left\|\widetilde{\Gamma}_1^t\right\|^2\right)^{1/2}\left(\sum_t\left\|Z_2^t - Z_2^{t-1}\right\|^2\right)^{1/2} + \left(\sum_t\left\|Z_1^t - Z_1^{t-1}\right\|^2\right)^{1/2}\left(\sum_t\left\|\widetilde{\Gamma}_2^t\right\|^2\right)^{1/2}\Bigg] < \infty.$$

The three bounds for the predictor, memory, and bilinear terms prove

$$\boxed{\sum_{t=1}^{\infty}|\widetilde{\mathrm{Err}}_t| < \infty.}$$

Define the potential relative to $\widetilde{\pi}$ by

$$\widetilde{\Phi}_t := \frac{1}{2}\sum_{i,I}\rho_i^{\widetilde{\pi}_i}(I)\left\|\bar{\pi}_{i,I}^t - \widetilde{\pi}_{i,I}\right\|^2. \tag{34}$$

Because $\widetilde{\pi}$ is Nash, its comparator saddle gap is nonnegative. The summed projection identity for this comparator therefore implies

$$\widetilde{\Phi}_{t+1} \le \widetilde{\Phi}_t + |\widetilde{\mathrm{Err}}_t|.$$

The sequence obtained by adding the tail $\sum_{r\ge t}|\widetilde{\mathrm{Err}}_r|$ is nonincreasing. It is also nonnegative. Therefore, $\widetilde{\Phi}_t$ converges.

Take any realization cluster point $\widetilde{Z}$ of $\{Z^t\}$. Along a subsequence, compactness permits $\pi^{t_k}\to\widetilde{\pi}$ with $\mathcal{R}(\widetilde{\pi})=\widetilde{Z}$. By lemma B.11, $\widetilde{\pi}$ is Nash. Since $\pi^{t_k}-\bar{\pi}^{t_k}\to 0$, $\bar{\pi}^{t_k}\to\widetilde{\pi}$ as well. Hence, $\widetilde{\Phi}_{t_k}\to 0$. Since the full potential has a limit, this subsequence limit implies

$$\widetilde{\Phi}_t\to 0.$$

We now use $\widetilde{\Phi}_t\to 0$ to prove convergence of every realization coordinate by induction on personal-history depth. The empty coordinate is constantly one. Assume $\bar{Z}^t_{i,p(I)}\to\rho_i^{\widetilde{\pi}_i}(I)$. If $\rho_i^{\widetilde{\pi}_i}(I)>0$, the coefficient of the local squared distance in (34) is positive. Therefore, $\bar{\pi}^t_{i,I}\to\widetilde{\pi}_{i,I}$ and

$$\bar{Z}^t_{i,Ia}=\bar{Z}^t_{i,p(I)}\bar{\pi}^t_{i,I}(a)\longrightarrow\rho_i^{\widetilde{\pi}_i}(I)\widetilde{\pi}_{i,I}(a)=\widetilde{Z}_{i,Ia}.$$

If $\rho_i^{\widetilde{\pi}_i}(I)=0$, nonnegativity gives $0\le\bar{Z}^t_{i,Ia}\le\bar{Z}^t_{i,p(I)}\to 0=\widetilde{Z}_{i,Ia}$. Thus,

$$\bar{Z}^t\to\widetilde{Z}.$$

Finally, the realization map is a polynomial on a finite product of simplexes. Hence, it is Lipschitz on that compact set. Since $\|\pi^t-\bar{\pi}^t\|\to 0$,

$$\left\|Z^t-\bar{Z}^t\right\|\to 0.$$

Therefore, $Z^t\to\widetilde{Z}$. The realization cluster point is unique, and both $\{Z^t\}$ and $\{\bar{Z}^t\}$ converge to it. Setting $Z^\star:=\widetilde{Z}$ gives the second conclusion. □

B.6 Proof of Lemma 6.3

Proof of Lemma 6.3. Fix $T\ge 2$ and $0<\eta\le\bar{\eta}$. Lemma B.14 gives the summed projection identity and $G_t\ge 0$ for every $t\ge 1$. We first expand the definition of Err_t in Lemma B.14 without using a Lipschitz bound. This expansion is necessary because the squared local distances in (17) have coefficient zero whenever the corresponding information set has zero own reach under π^{sc}. By (20), the instance $w_{i,I}=\rho_i^{\pi_i^{\mathrm{sc}}}(I)$ of lemma B.12 reads

$$\mathrm{Err}_t=A_t^{\mathrm{new}}+A_t^{\mathrm{mem}}+C_t, \tag{35}$$

where

$$A_t^{\mathrm{new}}:=\sum_{i=1}^{2}\sum_{J\in\mathcal{I}_i}\Gamma^t_{i,p(J)}K^t_{i,J},$$

$$A_t^{\mathrm{mem}}:=\sum_{i=1}^{2}\sum_{J\in\mathcal{I}_i}\Gamma^t_{i,p(J)}R^t_{i,J},$$

and

$$C_t:=\eta\left[(\Gamma_1^t)^{\mathsf{T}}A(Z_2^t-Z_2^{t-1})-(Z_1^t-Z_1^{t-1})^{\mathsf{T}}A\Gamma_2^t\right].$$

We abbreviate $\|\Gamma^t\|^2:=\|\Gamma_1^t\|^2+\|\Gamma_2^t\|^2$. Part (a) of lemma B.13 with $w_{i,I}=\rho_i^{\pi_i^{\mathrm{sc}}}(I)$ and (20) give the pointwise bound

$$\left\|\Gamma^t\right\|^2\le C_\Gamma E_t.$$

We now bound r_t^{beh} and Δ_t^{beh} separately. We use the constants J_0,J_E,J_G in (33), the bound of lemma B.5, and the bound of Lemma B.6 summed over $1\le t<T$. Since each parent coordinate is used by at most ν child information sets, Young's inequality gives

$$\sum_{t<T}|A_t^{\mathrm{new}}|\le\frac{\eta\nu}{2}\sum_{t<T}\left\|\Gamma^t\right\|^2+\frac{1}{2\eta}\sum_{i,I,t<T}(K^t_{i,I})^2.$$

Part (a) of Lemma B.13, with $w_{i,I} = \rho_i^{\pi_i^{\mathrm{sc}}}(I)$, and the bound in Lemma B.5 yield

$$\sum_{t<T} |A_t^{\mathrm{new}}| \le \frac{J_0}{2}\eta^2 + \left(\frac{\nu C_\Gamma}{2} + \frac{J_E}{2\rho_{\mathrm{sc}}}\right)\eta \sum_{t<T} E_t + \frac{J_G}{2}\frac{\eta^3}{\delta}\sum_{t<T} G_t.$$

Equation (16) and the oracle norm bound in Lemma B.17 imply $|R_{i,I}^t| \le \eta L \left\|\bar{\pi}_{i,I}^t - \pi_{i,I}^{t-1}\right\|$. Another application of Young's inequality gives

$$\sum_{t<T} |A_t^{\mathrm{mem}}| \le \frac{\eta L}{2}\left[\nu \sum_{t<T} \left\|\Gamma^t\right\|^2 + \sum_{i,I,t<T} \left\|\bar{\pi}_{i,I}^t - \pi_{i,I}^{t-1}\right\|^2\right].$$

By (31), the last sum is bounded by $\mathfrak{B}_0$ plus the $\left\|\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t\right\|^2$ contribution in the bound of Lemma B.5. Since $\eta \le \bar{\eta}$, part (a) of Lemma B.13, with $w_{i,I} = \rho_i^{\pi_i^{\mathrm{sc}}}(I)$, and the bound in Lemma B.5 give

$$\sum_{t<T} |A_t^{\mathrm{mem}}| \le \left(\frac{LJ_0\bar{\eta}^2}{2} + \frac{L\bar{\eta}B_\partial}{2}\right)\eta^2 + \left(\frac{L\nu C_\Gamma}{2} + \frac{LJ_E\bar{\eta}^2}{2\rho_{\mathrm{sc}}}\right)\eta\sum_{t<T} E_t + \frac{LJ_G\bar{\eta}^2}{2}\frac{\eta^3}{\delta}\sum_{t<T} G_t.$$

By the definition of F, r_t^{beh} has the expansion

$$r_t^{\mathrm{beh}} = \eta\left[(\bar{Z}_1^{t+1} - Z_1^t)^\mathsf{T} A (Z_2^t - Z_2^{t-1}) - (Z_1^t - Z_1^{t-1})^\mathsf{T} A (\bar{Z}_2^{t+1} - Z_2^t)\right].$$

Hence, $2uv \le u^2 + v^2$ and (25) give

$$|r_t^{\mathrm{beh}}| \le \frac{\|A\|_{\mathrm{op}}}{2}\eta\left(3\mathfrak{M}_t + 2\mathfrak{M}_{t-1}\right).$$

Summing the preceding pointwise bound over $1 \le t < T$ and applying Lemma B.6, together with $\|A\|_{\mathrm{op}} \le L$ from lemma B.17, gives

$$\sum_{t=1}^{T-1} |r_t^{\mathrm{beh}}| \le L\bar{\eta}M_\partial\eta^2 + \frac{5L\widehat{C}_d^M}{2\rho_{\mathrm{sc}}}\eta\sum_{t=1}^{T-1} E_t + \frac{5L}{2}\widehat{B}_d^M\frac{\eta^3}{\delta}\sum_{t=1}^{T-1} G_t.$$

Here, the term involving M_∂ accounts for $\mathfrak{M}_0$ in (25), and we used $\eta \le \bar{\eta}$.

The exact decomposition (35) and the displayed bilinear expansion of r_t^{beh} give

$$\begin{aligned}\Delta_t^{\mathrm{beh}} &= A_t^{\mathrm{new}} + A_t^{\mathrm{mem}} \\ &\quad + \eta\left[(\Gamma_1^t - (\bar{Z}_1^{t+1} - Z_1^t))^\mathsf{T} A (Z_2^t - Z_2^{t-1}) - (Z_1^t - Z_1^{t-1})^\mathsf{T} A (\Gamma_2^t - (\bar{Z}_2^{t+1} - Z_2^t))\right].\end{aligned}$$

The last line satisfies the pointwise bound

$$\left|C_t - r_t^{\mathrm{beh}}\right| \le \frac{\|A\|_{\mathrm{op}}}{2}\eta\left(2C_\Gamma E_t + 4\mathfrak{M}_t + 2\mathfrak{M}_{t-1}\right).$$

Indeed, $\left\|\Gamma_i^t - (\bar{Z}_i^{t+1} - Z_i^t)\right\|^2 \le 2\left\|\Gamma_i^t\right\|^2 + 2\left\|\bar{Z}_i^{t+1} - Z_i^t\right\|^2$, followed by $2uv \le u^2 + v^2$, part (a) of Lemma B.13 with $w_{i,I} = \rho_i^{\pi_i^{\mathrm{sc}}}(I)$, and (25), bounds the four squared coordinate-change terms by $2C_\Gamma E_t + 4\mathfrak{M}_t + 2\mathfrak{M}_{t-1}$. Therefore, summing the preceding bound over $1 \le t < T$, applying Lemma B.6, and using $\|A\|_{\mathrm{op}} \le L$ give

$$\sum_{t<T} |C_t - r_t^{\mathrm{beh}}| \le L\bar{\eta}M_\partial\eta^2 + L\left(C_\Gamma + \frac{3\widehat{C}_d^M}{\rho_{\mathrm{sc}}}\right)\eta\sum_{t<T} E_t + 3L\widehat{B}_d^M\frac{\eta^3}{\delta}\sum_{t<T} G_t.$$

Here again, the first term accounts for $\mathfrak{M}_0$ and uses $\eta \le \bar{\eta}$.

Define

$$C_0 := \frac{1 + L\bar{\eta}^2}{2}J_0 + \bar{\eta}\left(\frac{L}{2}B_\partial + 2LM_\partial\right),$$

$$C_E := \frac{(1+L)\nu C_\Gamma}{2} + \frac{1+L\bar{\eta}^2}{2} J_E + LC_\Gamma + \frac{11}{2} L\widehat{C}_d^M, \tag{36}$$

and

$$C_G := \frac{1+L\bar{\eta}^2}{2} J_G + \frac{11}{2} L\widehat{B}_d^M. \tag{37}$$

By the displayed expression for Δ_t^{beh} and the triangle inequality, $\sum_{t<T} |r_t^{\mathrm{beh}} + \Delta_t^{\mathrm{beh}}|$ is bounded by the sum of the bounds above for $\sum_{t<T} |r_t^{\mathrm{beh}}|$, $\sum_{t<T} |A_t^{\mathrm{new}}|$, $\sum_{t<T} |A_t^{\mathrm{mem}}|$, and $\sum_{t<T} |C_t - r_t^{\mathrm{beh}}|$. The three initial-boundary contributions in these estimates are absorbed into C_0 by its definition. In that sum, the coefficient of $\eta \sum_{t<T} E_t$ is

$$\frac{(1+L)\nu C_\Gamma}{2} + LC_\Gamma + \frac{1}{\rho_{\mathrm{sc}}} \left(\frac{1+L\bar{\eta}^2}{2} J_E + \frac{11}{2} L\widehat{C}_d^M \right) \le \frac{C_E}{\rho_{\mathrm{sc}}},$$

where the inequality uses $\rho_{\mathrm{sc}} \le 1$. Collecting the remaining coefficients proves the claimed bound with the displayed constants C_0, C_E, C_G. Every constant is finite and horizon independent. Every constant is also built from the payoff bound L in (21), the counting quantities B_k, N_k, $n_{i,I}$, $|\mathcal{H}_h|$, and ν of the game tree, and $\bar{\eta}$ by the displayed recursions, without reference to the primal–dual solution. No step divided by a current reach probability. □

B.7 Proof of Lemma 6.2

Proof of Lemma 6.2. Let $\widetilde{f}(s,x) := \|x\|^2/(2s)$ for $s > 0$, with value zero at $(0,0)$ and value $+\infty$ elsewhere on the boundary $s = 0$. For local distributions p and q, select the perspective subgradient $(-\|q\|^2/2, q)$ at $(0,0)$. Direct expansion gives

$$D_{\widetilde{f}}^q\big((s,sp),(r,rq)\big) = \frac{s}{2}\|p-q\|^2 \qquad (s,r \ge 0).$$

Summing this identity with $s = \rho_i^{\pi_i^{\mathrm{sc}}}(I)$, $p = \pi_{i,I}^{\mathrm{sc}}$, and $q = \bar{\pi}_{i,I}^t$ gives

$$D_d(Z^{\mathrm{sc}}, \bar{Z}^t) = \Phi_t.$$

This identity also holds when $\rho_i^{\pi_i^{\mathrm{sc}}}(I) = 0$.

Equation (9) gives $\Delta_t^{\mathrm{beh}} = \mathrm{Err}_t - r_t^{\mathrm{beh}}$. Lemma B.14 gives

$$\Phi_{t+1} = \Phi_t - \eta G_t - \frac{1}{2}E_t - S_t + r_t^{\mathrm{beh}} + \Delta_t^{\mathrm{beh}}, \qquad S_t \ge 0.$$

Since $S_t \ge 0$, the claimed descent inequality follows. □

Lemma B.8 (Local variation bounds)**.** *Fix $0 < \bar{\eta} \le 1$. Let $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ be generated by Algorithm 1 with a fixed $0 < \eta \le \bar{\eta}$. Then, for every player i, information set $I \in \mathcal{I}_i$, and horizon $T \ge 2$,*

$$P_{i,I}(T) \le 4\sqrt{2}L^3\eta^3 + 40\left(\sqrt{2} + \bar{\eta}L\right)^2 \eta^2 \sum_{t=1}^{T-1} \left\| v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}) \right\|^2.$$

Moreover,

$$Q_{i,I}(T) \le \eta^2 \sum_{t=1}^{T-1} \left\| v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}) \right\|^2.$$

B.8 Proof of Lemma B.8

Proof of Lemma B.8. Fix player i, information set $I \in \mathcal{I}_i$, and horizon $T \ge 2$. Apply lemma B.16 with horizon $T-1$ after shifting its index by one. For $\tau = 0, \dots, T-2$, set

$$r_\tau = \bar{\pi}_{i,I}^{\tau+1}, \quad \widehat{r}_\tau = \pi_{i,I}^{\tau+1}, \quad r_{\tau+1} = \bar{\pi}_{i,I}^{\tau+2},$$

$$c_\tau^0 = v_{i,I}(\pi^\tau), \quad c_\tau^1 = v_{i,I}(\pi^{\tau+1}), \quad c_{\tau+1}^0 = v_{i,I}(\pi^{\tau+1}).$$

The second oracle variation in lemma B.16 is zero. The quantity K_τ in that lemma is

$$\eta \left\langle v_{i,I}(\pi^\tau), \pi_{i,I}^{\tau+1} - \bar{\pi}_{i,I}^{\tau+1} \right\rangle = K_{i,I}^{\tau+1}.$$

The simplex $\Delta(\mathcal{A}(I))$ has Euclidean diameter at most $\sqrt{2}$, and the oracle norm bound in Lemma B.17 bounds every oracle vector by L. Hence, lemma B.16 applies with $D = \sqrt{2}$ and $M = L$. The bound for the squared predictor inner products and the initial-term estimate in Lemma B.16 then give the claimed bound for $P_{i,I}(T)$. Finally, the two projections defining $\pi^t_{i,I}$ and $\bar{\pi}^{t+1}_{i,I}$ start from the same point $\bar{\pi}^t_{i,I}$. Projection nonexpansiveness gives

$$\left\|\bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I}\right\| \le \eta \left\|v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1})\right\|,$$

which yields the claimed bound for $Q_{i,I}(T)$ after summation. □

Lemma B.9 (Variation of counterfactual values). *Let $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ be generated by Algorithm 1 with a fixed $\eta > 0$. Then, for every player i, information set $I \in \mathcal{I}_i$, and iteration $t \ge 1$,*

$$\begin{aligned}\eta^2 \left\|v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1})\right\|^2 \le{}& 2L^2\eta^2 \left\|Z^t_{-i} - Z^{t-1}_{-i}\right\|^2 \\ &+ 4n_{i,I} \sum_{J \succ I} (K^t_{i,J})^2 \\ &+ 4n_{i,I} L^2 \eta^2 \sum_{J \succ I} \left\|\bar{\pi}^t_{i,J} - \pi^{t-1}_{i,J}\right\|^2.\end{aligned}$$

B.9 Proof of Lemma B.9

Proof of Lemma B.9. Decompose the oracle change as

$$\begin{aligned}v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}) = {}& \left[v_{i,I}(\pi^t_i, \pi^t_{-i}) - v_{i,I}(\pi^t_i, \pi^{t-1}_{-i})\right] \\ &+ \left[v_{i,I}(\pi^t_i, \pi^{t-1}_{-i}) - v_{i,I}(\pi^{t-1}_i, \pi^{t-1}_{-i})\right].\end{aligned}$$

The squared norm of this sum is at most twice the sum of the two squared norms. The opponent square is bounded by $L^2 \left\|Z^t_{-i} - Z^{t-1}_{-i}\right\|^2$ by the opponent-variation bound in Lemma B.17, with the own strategy fixed at π^t_i. Multiplication by $2\eta^2$ gives the first line.

For the own-strategy term, the local identity in Lemma B.18 expresses every action component as

$$\frac{1}{\eta} \sum_{J \succ (I,a)} \rho^{\pi^t_i}_{i,(I,a) \rightsquigarrow p(J)} (K^t_{i,J} + R^t_{i,J}).$$

All coefficients lie in $[0, 1]$, and each descendant belongs below one action component of I. Cauchy–Schwarz gives a factor at most $n_{i,I}$. Combining this factor with the outer factor two, $(K + R)^2 \le 2K^2 + 2R^2$, and (30) gives the second and third lines. □

Lemma B.10 (Depth-wise recurrence for squared coordinate changes). *Fix $0 < \eta \le \bar{\eta} \le 1$. For every $k = 0, \dots, d-1$, define*

$$\alpha_k := 2B_k \left(1 + 2\bar{\eta}^2 L^2\right), \qquad \gamma_k := 12 N_k L^2.$$

Then, for every $t \ge 1$ and every $k = 0, \dots, d-1$,

$$\mathfrak{M}_{t,k+1} \le \alpha_k \mathfrak{M}_{t,k} + \frac{2}{\rho_{\mathrm{sc}}} E_t + \gamma_k \eta^2 \frac{G_t}{\delta}.$$

The term involving δ is omitted when the set defining δ is empty.

B.10 Proof of Lemma B.10

Proof of Lemma B.10. Fix $t \ge 1$ and $k \in \{0, \dots, d-1\}$. We bound separately the information sets whose parent sequences have positive and zero equilibrium reach. Lemmas B.15 and B.14, evaluated at (Z^t_1, Z^t_2), give

$$G_t = \left\langle s_x, Z^t_1 \right\rangle + \left\langle s_y, Z^t_2 \right\rangle.$$

Therefore, every relevant equilibrium-zero sequence s satisfies

$$0 \le Z^t_{i,s} \le \frac{G_t}{\delta}.$$

No current reach is divided by another reach.

We first consider positive equilibrium reach. The source coefficients $\bar{Z}^t_{i,p(I)}$ and $Z^t_{i,p(I)}$ are actual reaches and are at most one. Hence, when $\rho_i^{\pi_i^{\mathrm{sc}}}(I) > 0$,

$$(\bar{Z}^t_{i,p(I)})^2 \left\|\pi^t_{i,I} - \bar{\pi}^t_{i,I}\right\|^2 \le \frac{1}{\rho_{\mathrm{sc}}} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\|\pi^t_{i,I} - \bar{\pi}^t_{i,I}\right\|^2,$$
$$(Z^t_{i,p(I)})^2 \left\|\bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I}\right\|^2 \le \frac{1}{\rho_{\mathrm{sc}}} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\|\bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I}\right\|^2.$$

For U and V, respectively, the contributions with positive equilibrium reach are therefore bounded by

$$\frac{2}{\rho_{\mathrm{sc}}} \sum_{i, I \in \mathcal{I}_i(k)} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\|\pi^t_{i,I} - \bar{\pi}^t_{i,I}\right\|^2,$$
$$\frac{2}{\rho_{\mathrm{sc}}} \sum_{i, I \in \mathcal{I}_i(k)} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\|\bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I}\right\|^2.$$

The sum of the two positive-reach contributions above is bounded by $(2/\rho_{\mathrm{sc}})E_t$.

We next consider zero equilibrium reach. If $Z^{\mathrm{sc}}_{i,s} = 0$, then the zero-reach estimate above and $2uv \le u^2 + v^2$ applied to $\bar{Z}^t_{i,s} = Z^t_{i,s} - (Z^t_{i,s} - \bar{Z}^t_{i,s})$ give

$$(\bar{Z}^t_{i,s})^2 \le 2\frac{G_t}{\delta} + 2(Z^t_{i,s} - \bar{Z}^t_{i,s})^2.$$

For the two vectors, the first terms in (26) are therefore bounded as follows, using (29):

$$\text{(i)} \quad 2(\bar{Z}^t_{i,s})^2 \left\|\pi^t_{i,I} - \bar{\pi}^t_{i,I}\right\|^2 \le 4L^2\eta^2\frac{G_t}{\delta} + 4L^2\eta^2(Z^t_{i,s} - \bar{Z}^t_{i,s})^2,$$
$$\text{(ii)} \quad 2(Z^t_{i,s})^2 \left\|\bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I}\right\|^2 \le 8L^2\eta^2\frac{G_t}{\delta}.$$

After summing the two zero-reach bounds above over at most N_k information sets, their G_t/δ contribution is at most $\gamma_k\eta^2 G_t/\delta$. The extra term in bound (i) involves a coordinate at depth k and reuses each parent at most B_k times. Together with the second term from (26), its total contribution is at most

$$\left[2B_k + 4B_k\eta^2L^2\right]\mathfrak{M}_{t,k} \le \alpha_k\mathfrak{M}_{t,k}.$$

Combining the positive- and zero-reach bounds in (26) proves the claimed recurrence. □

Lemma B.11 (Behavioral accumulation points are Nash)**.** *Suppose that the iterates of Algorithm 1 satisfy*

$$\left\|\pi^t - \bar{\pi}^t\right\| \to 0, \qquad \left\|\pi^t - \pi^{t-1}\right\| \to 0.$$

Then, every accumulation point of either sequence is a behavioral Nash equilibrium.

B.11 Proof of Lemma B.11

Proof of Lemma B.11. The finite product of local simplexes is compact. Let $\pi^{t_k} \to \widetilde{\pi}$. By the hypotheses of the lemma,

$$\bar{\pi}^{t_k} \to \widetilde{\pi}, \qquad \pi^{t_k - 1} \to \widetilde{\pi}.$$

Take limits along $t = t_k$ in the predictor relation in (14). Since π^{t_k}, $\bar{\pi}^{t_k}$, and π^{t_k-1} all converge to $\widetilde{\pi}$, continuity of the counterfactual values and Euclidean projection gives, for every i, I,

$$\widetilde{\pi}_{i,I} = \Pi_{\Delta(\mathcal{A}(I))}\left[\widetilde{\pi}_{i,I} + \eta v_{i,I}(\widetilde{\pi})\right].$$

Projection optimality is equivalent to

$$\widetilde{\pi}_{i,I} \in \operatorname*{arg\,max}_{q \in \Delta(\mathcal{A}(I))} \left\langle v_{i,I}(\widetilde{\pi}), q\right\rangle.$$

For any behavioral deviation π'_i, the global identity in Lemma B.18 gives

$$U_i(\pi'_i, \widetilde{\pi}_{-i}) - U_i(\widetilde{\pi}) = \sum_{I \in \mathcal{I}_i} \rho_i^{\pi'_i}(I)\left\langle \pi'_{i,I} - \widetilde{\pi}_{i,I}, v_{i,I}(\widetilde{\pi})\right\rangle.$$

Every coefficient is nonnegative and every inner product is nonpositive by the local best-response property just established. Hence, the deviation cannot improve utility. Thus, $\widetilde{\pi}$ is Nash. The condition $\|\pi^t - \bar{\pi}^t\| \to 0$ implies that $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ have the same accumulation-point set. This proves the statement for $\{\bar{\pi}^t\}$ as well. □

Lemma B.12 (Decomposition of weighted oracle-change sums)**.** *Fix a real weight $w_{i,I}$ for each player i and each information set $I \in \mathcal{I}_i$. For every iteration $t \geq 1$, set $\Gamma_i^t[w] := \mathcal{L}_i(b_i^t; w_i, \pi_i^t)$. Then, for every iteration $t \geq 1$,*

$$\eta \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} w_{i,I} \left\langle v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}), \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\rangle = \sum_{i=1}^{2} \sum_{J \in \mathcal{I}_i} \Gamma_{i,p(J)}^t[w] \left(K_{i,J}^t + R_{i,J}^t\right) + \eta \left[(\Gamma_1^t[w])^\mathsf{T} A (Z_2^t - Z_2^{t-1}) - (Z_1^t - Z_1^{t-1})^\mathsf{T} A \Gamma_2^t[w]\right].$$

Root information sets contribute zero to the first sum because $\Gamma_{i,\varnothing}^t[w] = 0$.

B.12 Proof of Lemma B.12

Proof of Lemma B.12. For player i, split the oracle change in the order

$$v_i(\pi_i^t, \pi_{-i}^t) - v_i(\pi_i^{t-1}, \pi_{-i}^{t-1}) = \left[v_i(\pi_i^t, \pi_{-i}^{t-1}) - v_i(\pi_i^{t-1}, \pi_{-i}^{t-1})\right] + \left[v_i(\pi_i^t, \pi_{-i}^t) - v_i(\pi_i^t, \pi_{-i}^{t-1})\right].$$

For the own-change bracket, apply the local identity in Lemma B.18 with $\pi_i' = \pi_i^t$ and $\pi_i = \pi_i^{t-1}$. After pairing with $\eta w_{i,I}(\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t)$ and summing over ancestor information sets I, exchange the ancestor and descendant sums. For a fixed descendant J, the coefficient multiplying

$$\eta \left\langle \pi_{i,J}^t - \pi_{i,J}^{t-1}, v_{i,J}(\pi^{t-1}) \right\rangle$$

is the path sum (19) evaluated at the parent $p(J)$, namely $\Gamma_{i,p(J)}^t[w]$. Since

$$\pi_{i,J}^t - \pi_{i,J}^{t-1} = (\pi_{i,J}^t - \bar{\pi}_{i,J}^t) + (\bar{\pi}_{i,J}^t - \pi_{i,J}^{t-1}),$$

the own-change contribution is the first sum in the claimed decomposition.

For the opponent-change bracket, fix player 1's current descendants. Pairing the local opponent-linear change with the weighted differences $w_{1,I} b_{1,I}^t$ gives the recursively defined vector $\Gamma_1^t[w]$ and the opponent realization difference $Z_2^t - Z_2^{t-1}$. The resulting term is $(\Gamma_1^t[w])^\mathsf{T} A (Z_2^t - Z_2^{t-1})$. For player 2, whose utility is $-x^\mathsf{T} A y$, the corresponding contribution is $-(Z_1^t - Z_1^{t-1})^\mathsf{T} A \Gamma_2^t[w]$. Empty coordinates are included in A. Their differences are zero. Direct terminal coefficients independent of the changing opponent cancel. Every opponent-dependent direct coefficient is already an entry of the displayed bilinear pairing.

The recursion (18) assigns one term to each information set. If multiple information sets share one parent, they occur as distinct indices in the first sum of the claimed decomposition. The shared parent coordinate is reused only as a coefficient. No terminal or descendant contribution is omitted or counted twice. This proves Lemma B.12. □

Lemma B.13 (Lift bounds)**.** *Define*

$$c_0^\Gamma := 0, \qquad c_{k+1}^\Gamma := 2B_k c_k^\Gamma + 2 \quad (k = 0, \ldots, d-1), \qquad C_\Gamma := \sum_{k=1}^{d} c_k^\Gamma.$$

Fix a real weight $w_{i,I}$ for each player i and information set $I \in \mathcal{I}_i$, a behavioral profile $r = (r_1, r_2)$, and an iteration $t \geq 0$.

(a) *If $t \geq 1$ and $0 \leq w_{i,I} \leq \rho_i^{\pi_i^{\mathrm{sc}}}(I)$ for every i and I, then*

$$\sum_{i=1}^{2} \left\|\mathcal{L}_i(b_i^t; w_i, r_i)\right\|^2 \leq C_\Gamma E_t.$$

(b) *If $|w_{i,I}| \leq 1$ for every i and I, then*

$$\sum_{i=1}^{2} \left\|\mathcal{L}_i(b_i^t; w_i, r_i)\right\|^2 \leq C_\Gamma \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \left\|\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t\right\|^2.$$

B.13 Proof of Lemma B.13

Proof of Lemma B.13. Write $\Gamma_i := \mathcal{L}_i(b_i^t; w_i, r_i)$. By (18), the family Γ_i satisfies the recursion of (26) with $\omega_{i,I} = w_{i,I}$, $u_{i,I} = \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t$, and $\theta_I = r_{i,I}$. Hence, for every $k = 0, \dots, d-1$,

$$\sum_{i=1}^{2} \sum_{s \in \Sigma_i(k+1)} \Gamma_{i,s}^2 \le 2 \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i(k)} w_{i,I}^2 \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2 + 2B_k \sum_{i=1}^{2} \sum_{s \in \Sigma_i(k)} \Gamma_{i,s}^2.$$

Under the hypothesis of part (a), $w_{i,I}^2 \le w_{i,I} \le \rho_i^{\pi_i^{\mathrm{sc}}}(I)$. Hence, by (8), the source term in the preceding recurrence is at most $2E_t$. Under the hypothesis of part (b), $w_{i,I}^2 \le 1$. Hence, the source term in the preceding recurrence is at most $2 \sum_{i,I} \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2$. The empty coordinates satisfy $\Gamma_{i,\varnothing} = 0$. Therefore, induction over k gives, for $k = 0, \dots, d$,

$$\sum_{i=1}^{2} \sum_{s \in \Sigma_i(k)} \Gamma_{i,s}^2 \le c_k^{\Gamma} E_t \quad \text{in case (a)}, \qquad \sum_{i=1}^{2} \sum_{s \in \Sigma_i(k)} \Gamma_{i,s}^2 \le c_k^{\Gamma} \sum_{i,I} \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\|^2 \quad \text{in case (b)}.$$

Summing over $k = 1, \dots, d$ proves both parts of the lemma. □

Lemma B.14 (Saddle gap and summed projection identity)**.** *For every $t \ge 1$, the saddle gap satisfies*

$$\begin{aligned} G_t &= \left\langle F(Z^t), Z^t - Z^{\mathrm{sc}} \right\rangle \\ &= \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\langle v_{i,I}(\pi^t), \pi_{i,I}^{\mathrm{sc}} - \pi_{i,I}^t \right\rangle \\ &= U_1(\pi_1^{\mathrm{sc}}, \pi_2^t) - U_1(\pi_1^t, \pi_2^{\mathrm{sc}}) \ge 0. \end{aligned}$$

Define

$$\mathrm{Err}_t := \eta \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\langle v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}), \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t \right\rangle.$$

Moreover, there exists $S_t \ge 0$ such that

$$\eta G_t + \frac{1}{2} E_t + S_t = \Phi_t - \Phi_{t+1} + \mathrm{Err}_t.$$

B.14 Proof of Lemma B.14

Proof of Lemma B.14. Fix $t \ge 1$. For the equilibrium comparator π^{sc}, apply lemma B.18 to each player i with $\pi_i' = \pi_i^{\mathrm{sc}}$ and $\pi_i = \pi_i^t$. The definition of F and the identity $U_2 = -U_1$ give the asserted representation of G_t.

For every player i and information set $I \in \mathcal{I}_i$, apply lemma B.19 with

$$C = \Delta(\mathcal{A}(I)), \quad q_t = \bar{\pi}_{i,I}^t, \quad p_t = \pi_{i,I}^t, \quad q_{t+1} = \bar{\pi}_{i,I}^{t+1},$$

$$c_{t-1} = v_{i,I}(\pi^{t-1}), \qquad c_t = v_{i,I}(\pi^t), \qquad z = \pi_{i,I}^{\mathrm{sc}}.$$

Multiply the resulting identity by $\rho_i^{\pi_i^{\mathrm{sc}}}(I)$ and sum over i and I. Let S_t be the sum of the weighted normal-cone slacks in lemma B.19. Each weight and each slack is nonnegative. Hence, $S_t \ge 0$. The gap representation just established identifies the first inner products with G_t. The two squared-distance terms sum to $E_t/2$. The comparator-distance difference equals $\Phi_t - \Phi_{t+1}$. The oracle-difference terms equal Err_t by definition because $q_{t+1} - p_t = \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t$. This is the asserted summed identity. □

Lemma B.15 (Dual slacks and the saddle gap)**.** *Let $(x^{\mathrm{sc}}, y^{\mathrm{sc}})$ be the saddle pair and let s_x, s_y be the dual slacks defined in (13). For every $x, y \ge 0$ with $Ex = e$ and $Fy = f$,*

$$x^{\mathrm{sc}\mathsf{T}} A y - x^{\mathsf{T}} A y^{\mathrm{sc}} = \langle s_x, x \rangle + \langle s_y, y \rangle.$$

B.15 Proof of Lemma B.15

Proof of Lemma B.15. Feasibility and (13) give

$$x^{\mathrm{sc}\mathsf{T}} A y = \left\langle A^{\mathsf{T}} x^{\mathrm{sc}}, y \right\rangle = \left\langle F^{\mathsf{T}} \psi^{\mathrm{sc}} + s_y, y \right\rangle = \left\langle \psi^{\mathrm{sc}}, f \right\rangle + \left\langle s_y, y \right\rangle,$$

and

$$x^{\mathsf{T}} A y^{\mathrm{sc}} = \left\langle x, E^{\mathsf{T}} u^{\mathrm{sc}} - s_x \right\rangle = \left\langle u^{\mathrm{sc}}, e \right\rangle - \left\langle s_x, x \right\rangle.$$

Strong duality gives $\langle \psi^{\mathrm{sc}}, f \rangle = \langle u^{\mathrm{sc}}, e \rangle$, proving Lemma B.15. □

Lemma B.16 (Variation bound for projected optimistic updates)**.** *Let $C \subset \mathbb{R}^m$ be nonempty, compact, and convex. Suppose that its diameter is at most D for some $D \geq 0$. Fix a step size $\eta > 0$ and an initial point $r_0 \in C$. Let $\{c_t^0\}_{t\geq 0}$ and $\{c_t^1\}_{t\geq 0}$ be sequences in $\mathbb{R}^m$ such that, for some $M \geq 0$,*

$$\left\|c_t^0\right\|, \left\|c_t^1\right\| \leq M \qquad (t \geq 0).$$

For every $t \geq 0$, define

$$\widehat{r}_t = \Pi_C(r_t + \eta c_t^0), \qquad r_{t+1} = \Pi_C(r_t + \eta c_t^1).$$

Define

$$d_t := \widehat{r}_t - r_t, \qquad K_t := \eta \left\langle c_t^0, d_t \right\rangle,$$

$$H_t := \eta \left\langle c_t^0, d_t \right\rangle - \frac{1}{2}\left\|d_t\right\|^2, \qquad g_t := \max_{q\in C} \left\langle c_t^0, q - r_t \right\rangle,$$

$$v_t := \left\|c_t^1 - c_t^0\right\| + \left\|c_{t+1}^0 - c_t^1\right\|.$$

Then, for every $T \geq 1$,

$$\sum_{t=0}^{T-1} H_t^2 \leq 2\eta H_0 g_0 + 10\eta^2 (D + \eta M)^2 \sum_{t=0}^{T-1} v_t^2,$$

$$\sum_{t=0}^{T-1} K_t^2 \leq 8\eta H_0 g_0 + 40\eta^2 (D + \eta M)^2 \sum_{t=0}^{T-1} v_t^2.$$

Moreover,

$$8\eta H_0 g_0 \leq 4\eta^3 M^3 D.$$

The lemma requires no relation among the oracle vectors other than boundedness.

B.16 Proof of Lemma B.16

Proof of Lemma B.16. First, we establish the projection inequalities. Projection optimality gives

$$\left\langle r_t + \eta c_t^0 - \widehat{r}_t, q - \widehat{r}_t \right\rangle \leq 0 \qquad (q \in C).$$

The choice $q = r_t$ yields

$$K_t = \eta \left\langle c_t^0, d_t \right\rangle \geq \left\|d_t\right\|^2.$$

Therefore,

$$\frac{1}{2}\left\|d_t\right\|^2 \leq H_t \leq K_t \leq 2H_t. \tag{38}$$

We next establish the gap identity for a fixed oracle vector. Let

$$\widehat{g}_t := \max_{q\in C} \left\langle c_t^0, q - \widehat{r}_t \right\rangle.$$

Since $\widehat{r}_t = r_t + d_t$,

$$\widehat{g}_t = g_t - \left\langle c_t^0, d_t \right\rangle = g_t - \frac{K_t}{\eta}.$$

We now define H and prove its monotonicity for a fixed oracle vector. Define

$$H(r, c) := \max_{q\in C} \left\{ \eta \left\langle c, q - r \right\rangle - \frac{1}{2}\left\|q - r\right\|^2 \right\}. \tag{39}$$

Completing the square shows that the unique maximizer is $\Pi_C(r+\eta c)$. Hence, $H_t = H(r_t, c_t^0)$. Let $p = \Pi_C(r+\eta c)$, $d = p - r$, and $a = \eta c - d$. Then, $a \in N_C(p)$. For every $q \in C$,

$$\begin{aligned}\eta \langle c, q-p\rangle - \frac{1}{2}\|q-p\|^2 &= \langle a+d, q-p\rangle - \frac{1}{2}\|q-p\|^2 \\ &\le \langle d, q-p\rangle - \frac{1}{2}\|q-p\|^2 \le \frac{1}{2}\|d\|^2 .\end{aligned}$$

Thus, $H(p,c) \le \|d\|^2/2$. On the other hand,

$$H(r,c) = \langle a+d, d\rangle - \frac{1}{2}\|d\|^2 = \langle a, d\rangle + \frac{1}{2}\|d\|^2 .$$

Since $r = p - d \in C$ and $a \in N_C(p)$, $\langle a, -d\rangle \le 0$. Hence, $H(r,c) \ge \|d\|^2/2$. Therefore,

$$H(\Pi_C(r+\eta c), c) \le H(r,c).$$

We next prove that $\sqrt{H}$ is Lipschitz continuous in the oracle vector. The maximizer in (39) is unique. Hence, Danskin's theorem gives

$$\nabla_c H(r,c) = \eta(\Pi_C(r+\eta c) - r).$$

By (38),

$$\|\nabla_c H(r,c)\| \le \eta\sqrt{2H(r,c)}.$$

When $H(r,c) > 0$ this implies

$$\left\|\nabla_c \sqrt{H(r,c)}\right\| \le \frac{\eta}{\sqrt{2}}.$$

To include the nonsmooth boundary $H = 0$, let $c_s = c + s(c'-c)$ and $\phi_\varepsilon(s) = \sqrt{H(r,c_s)+\varepsilon}$. For $\varepsilon > 0$,

$$|\phi_\varepsilon'(s)| \le \frac{\eta}{\sqrt{2}}\|c'-c\| .$$

Integrating from 0 to 1 and sending $\varepsilon \downarrow 0$ gives

$$\left|\sqrt{H(r,c')} - \sqrt{H(r,c)}\right| \le \frac{\eta}{\sqrt{2}}\|c'-c\| .$$

Applying the Lipschitz bound for $\sqrt{H}$ before and after the fixed-oracle step $r_{t+1} = \Pi_C(r_t + \eta c_t^1)$, and using the fixed-oracle monotonicity of H established above, gives

$$\begin{aligned}\sqrt{H_{t+1}} &= \sqrt{H(r_{t+1}, c_{t+1}^0)} \\ &\le \sqrt{H(r_{t+1}, c_t^1)} + \frac{\eta}{\sqrt{2}}\left\|c_{t+1}^0 - c_t^1\right\| \\ &\le \sqrt{H(r_t, c_t^1)} + \frac{\eta}{\sqrt{2}}\left\|c_{t+1}^0 - c_t^1\right\| \\ &\le \sqrt{H_t} + \frac{\eta}{\sqrt{2}}v_t.\end{aligned}$$

Hence,

$$\left[\sqrt{H_{t+1}} - \sqrt{H_t}\right]_+ \le \frac{\eta}{\sqrt{2}}v_t.$$

If $H_{t+1} > H_t$, multiply by $\sqrt{H_{t+1}} + \sqrt{H_t} \le 2\sqrt{H_{t+1}}$ to obtain

$$[H_{t+1} - H_t]_+ \le \sqrt{2}\eta\sqrt{H_{t+1}}v_t.$$

The right-hand side necessarily uses the later index H_{t+1}. After shifting indices in Abel summation, it becomes H_t.

We next bound the change in the gap when the point and oracle vector change. Write

$$g(r,c) := \max_{q\in C}\langle c, q-r\rangle .$$

For $r, r' \in C$ and $c, c' \in \mathbb{R}^m$,

$$|g(r, c') - g(r, c)| \le D \|c' - c\|,$$
$$|g(r', c) - g(r, c)| \le \|c\| \|r' - r\|.$$

Projection nonexpansiveness gives

$$\|r_{t+1} - \widehat{r}_t\| \le \eta \left\|c_t^1 - c_t^0\right\|.$$

Set

$$\varepsilon_t := g_{t+1} - \widehat{g}_t.$$

Using the two gap-Lipschitz estimates above, the triangle inequality, and the oracle bound,

$$|\varepsilon_t| \le (D + \eta M) v_t.$$

Let q_t maximize g_t. Projection optimality gives

$$\left\langle \eta c_t^0 - d_t, q_t - \widehat{r}_t \right\rangle \le 0.$$

Consequently,

$$\begin{aligned} \eta g_t &= K_t + \eta \left\langle c_t^0, q_t - \widehat{r}_t \right\rangle \\ &\le K_t + \left\langle d_t, q_t - \widehat{r}_t \right\rangle \\ &\le (\eta M + D) \|d_t\| \le \sqrt{2}(D + \eta M)\sqrt{H_t}. \end{aligned}$$

Thus,

$$\eta g_t \le (D + \eta M) \|d_t\| \le \sqrt{2}(D + \eta M)\sqrt{H_t}.$$

Finally, we combine the preceding bounds by Abel summation. By the fixed-oracle gap identity above and the definition of ε_t,

$$g_t - g_{t+1} + \varepsilon_t = g_t - \widehat{g}_t = K_t/\eta.$$

Since $H_t \le K_t$,

$$H_t^2 \le \eta H_t (g_t - g_{t+1} + \varepsilon_t).$$

The exact Abel identity is

$$\sum_{t=0}^{T-1} H_t (g_t - g_{t+1}) = H_0 g_0 - H_{T-1} g_T + \sum_{t=1}^{T-1} (H_t - H_{t-1}) g_t.$$

Because $H_{T-1}, g_T \ge 0$, the second boundary term is nonpositive. Put $A_\eta = D + \eta M$ and $S_T = \sum_{t<T} H_t^2$. Summing the inequality for H_t^2 and using the displayed Abel identity, the bound on ε_t, the positive-increment bound with the index shifted by one, and the gap estimate in terms of $\sqrt{H_t}$ gives

$$S_T \le \eta H_0 g_0 + 2\eta A_\eta \sum_{t=1}^{T-1} H_t v_{t-1} + \eta A_\eta \sum_{t=0}^{T-1} H_t v_t.$$

Apply

$$2\eta A_\eta H_t v \le \frac{1}{4} H_t^2 + 4\eta^2 A_\eta^2 v^2, \qquad \eta A_\eta H_t v \le \frac{1}{4} H_t^2 + \eta^2 A_\eta^2 v^2.$$

After extending shortened sums by nonnegative terms,

$$S_T \le \eta H_0 g_0 + \frac{1}{2} S_T + 5\eta^2 A_\eta^2 \sum_{t<T} v_t^2.$$

This establishes the claimed bound for the sum of squared H_t terms. Since $K_t \le 2H_t$, multiplying by four gives the claimed bound for the sum of squared K_t terms.

Finally, $g_0 \le MD$. Completing the square in (39) gives $H_0 \le \eta^2 M^2/2$. Hence,

$$8\eta H_0 g_0 \le 4\eta^3 M^3 D,$$

which is the claimed initial-term estimate. □

Lemma B.17 (Oracle and operator bounds)**.** *Fix a player i and behavioral profiles π and π' such that $\pi_i = \pi_i'$. Then, for every $I \in \mathcal{I}_i$,*

$$\|v_{i,I}(\pi)\| \le L, \qquad \|v_{i,I}(\pi) - v_{i,I}(\pi')\| \le L \left\|\mathcal{R}_{-i}(\pi_{-i}) - \mathcal{R}_{-i}(\pi'_{-i})\right\|.$$

Moreover, $\|A\|_{\mathrm{op}} \le L$.

B.17 Proof of Lemma B.17

Proof of Lemma B.17. Fix i and $I \in \mathcal{I}_i$. By perfect recall, no history in I is a prefix of another history in I. Hence, every terminal history z has at most one prefix $h_z \in I$. For $a \in \mathcal{A}(I)$, let $\mathcal{Z}(I,a)$ be the set of terminal histories z such that h_z exists and $h_z a \sqsubseteq z$. The sets $\mathcal{Z}(I,a)$ are disjoint across the actions $a \in \mathcal{A}(I)$. By (1),

$$v_{i,I}(\pi)_a = \sum_{z \in \mathcal{Z}(I,a)} \rho^{\pi}_{-i,c}(h_z)\, \rho^{\pi}(h_z a, z)\, u_i(z). \tag{40}$$

Every probability product in (40) lies in $[0,1]$. Hence, by disjointness, $|u_i(z)| = |u_1(z)|$, and (21),

$$\|v_{i,I}(\pi)\| \le \sum_{a \in \mathcal{A}(I)} |v_{i,I}(\pi)_a| \le \sum_{z \in \mathcal{Z}} |u_1(z)| \le L.$$

This proves the first assertion.

For the second bound, group the probability factors in (40) by their owner. For $z \in \mathcal{Z}(I,a)$, the factors owned by chance and by player i form a product $c_z \in [0,1]$ that does not depend on π_{-i}. The factors owned by the opponent form the realization coordinate $\mathcal{R}_{-i,s(z)}(\pi_{-i})$, where $s(z)$ is the opponent sequence at z. Hence,

$$v_{i,I}(\pi)_a = \sum_{z \in \mathcal{Z}(I,a)} c_z u_i(z) \mathcal{R}_{-i,s(z)}(\pi_{-i}).$$

Since $\pi_i = \pi'_i$, the coefficients $c_z u_i(z)$ are the same for π and π'. Therefore, bounding every coordinate difference by the Euclidean norm and using (21),

$$\begin{aligned}\|v_{i,I}(\pi) - v_{i,I}(\pi')\| &\le \sum_{a \in \mathcal{A}(I)} \sum_{z \in \mathcal{Z}(I,a)} |u_1(z)| \left|\mathcal{R}_{-i,s(z)}(\pi_{-i}) - \mathcal{R}_{-i,s(z)}(\pi'_{-i})\right| \\ &\le L \left\|\mathcal{R}_{-i}(\pi_{-i}) - \mathcal{R}_{-i}(\pi'_{-i})\right\|.\end{aligned}$$

Finally, every entry of A is a sum of products of chance probabilities and $u_1(z)$ over a set of terminal histories. These sets are disjoint across entries. Since the operator norm is at most the Frobenius norm, (21) gives

$$\|A\|_{\mathrm{op}} \le \Big(\sum_{s,q} A_{s,q}^2\Big)^{1/2} \le \sum_{s,q} |A_{s,q}| \le \sum_{z \in \mathcal{Z}} |u_1(z)| \le L.$$

□

Lemma B.18 (Performance-difference identities)**.** *For any two behavioral strategies π'_i and π_i of player i and any fixed opponent strategy τ_{-i},*

$$U_i(\pi'_i, \tau_{-i}) - U_i(\pi_i, \tau_{-i}) = \sum_{I \in \mathcal{I}_i} \rho_i^{\pi'_i}(I) \left\langle \pi'_{i,I} - \pi_{i,I}, v_{i,I}(\pi_i, \tau_{-i}) \right\rangle.$$

Define $v_{i,I}(a; \pi_i, \tau_{-i}) := v_{i,I}(\pi_i, \tau_{-i})_a$. For an action a at I, write $J \succ (I,a)$ when J is a strict own descendant reached after action a. Let $\rho^{\pi'_i}_{i,(I,a)\rightsquigarrow p(J)}$ be the product of the π'_i probabilities from immediately after (I,a) through the parent sequence $p(J)$. Then, for every information-set action (I,a),

$$\begin{aligned}&v_{i,I}(a; \pi'_i, \tau_{-i}) - v_{i,I}(a; \pi_i, \tau_{-i}) \\ &\quad = \sum_{J \succ (I,a)} \rho^{\pi'_i}_{i,(I,a)\rightsquigarrow p(J)} \left\langle \pi'_{i,J} - \pi_{i,J}, v_{i,J}(\pi_i, \tau_{-i}) \right\rangle.\end{aligned}$$

B.18 Proof of Lemma B.18

Proof of Lemma B.18. Fix a player i, an opponent profile τ_{-i}, and an own profile r_i. The ordinary counterfactual value $v_{i,I}(r_i, \tau_{-i})$ satisfies (3). The coefficient $\ell_{i,Ia}$ permits terminal payoff entries at any sequence, including entries reached before another own decision. Equation (3) is the recursive

form of the history definition (1). We prove the local identity by induction on the number of remaining own information sets below (I, a). If there is no own descendant, the continuation contains only the direct coefficient $\ell_{i,Ia}(\tau_{-i})$. This coefficient is independent of π'_i and π_i. Hence, both sides of the claimed local identity are zero.

Suppose the statement holds below every immediate child $J \in \mathcal{C}_i(Ia)$. By (3), the difference at action (I, a) is the sum over these children of

$$\left\langle \pi'_{i,J}, v_{i,J}(\pi'_i, \tau_{-i}) \right\rangle - \left\langle \pi_{i,J}, v_{i,J}(\pi_i, \tau_{-i}) \right\rangle .$$

For one child J, add and subtract $\left\langle \pi'_{i,J}, v_{i,J}(\pi_i, \tau_{-i}) \right\rangle$ to obtain

$$\begin{aligned} &\left\langle \pi'_{i,J} - \pi_{i,J}, v_{i,J}(\pi_i, \tau_{-i}) \right\rangle \\ &\quad + \left\langle \pi'_{i,J}, v_{i,J}(\pi'_i, \tau_{-i}) - v_{i,J}(\pi_i, \tau_{-i}) \right\rangle . \end{aligned}$$

The first term of this decomposition is the term indexed by J. The componentwise induction hypothesis and the subsequent average with $\pi'_{i,J}$ multiply every deeper term by the appropriate action probability at J. Consequently, its coefficient is precisely the product of current probabilities from after (I, a) to the deeper parent. The sum over all immediate children gives the claimed local identity. The child sets are disjoint as information-set indices. Hence, shared parents create no duplicate descendant term.

At the artificial root of player i, the same decomposition compares the total expected utilities. The reach preceding a deviation at I is the product generated by π'_i, which gives the asserted global identity. □

Lemma B.19 (Optimistic projection identity)**.** *Let $C \subset \mathbb{R}^n$ be nonempty, compact, and convex. Fix $\eta > 0$, a point $q_t \in C$, and vectors $c_{t-1}, c_t \in \mathbb{R}^n$, and define*

$$p_t = \Pi_C(q_t + \eta c_{t-1}), \qquad q_{t+1} = \Pi_C(q_t + \eta c_t).$$

For any comparator $z \in C$, define

$$\xi_t^{\mathrm{p}} = q_t - p_t + \eta c_{t-1} \in N_C(p_t), \qquad \xi_t^{\mathrm{c}} = q_t - q_{t+1} + \eta c_t \in N_C(q_{t+1}),$$

where $\xi \in N_C(r)$ means $\langle \xi, u - r \rangle \le 0$ for all $u \in C$. Set

$$S_t^{\mathrm{p}} := -\langle \xi_t^{\mathrm{p}}, q_{t+1} - p_t \rangle \ge 0, \qquad S_t^{\mathrm{c}} := -\langle \xi_t^{\mathrm{c}}, z - q_{t+1} \rangle \ge 0.$$

Then,

$$\begin{aligned} &\eta \langle c_t, z - p_t \rangle + \frac{1}{2} \|p_t - q_t\|^2 + \frac{1}{2} \|q_{t+1} - p_t\|^2 + S_t^{\mathrm{p}} + S_t^{\mathrm{c}} \\ &\quad = \frac{1}{2} \|z - q_t\|^2 - \frac{1}{2} \|z - q_{t+1}\|^2 + \eta \langle c_t - c_{t-1}, q_{t+1} - p_t \rangle . \end{aligned}$$

B.19 Proof of Lemma B.19

Proof of Lemma B.19. The normal-cone inclusions follow from Euclidean projection optimality. The inclusion at q_{t+1} gives

$$\eta \langle c_t, z - q_{t+1} \rangle + S_t^{\mathrm{c}} = \langle q_{t+1} - q_t, z - q_{t+1} \rangle .$$

The three-point identity

$$2 \langle r - s, u - r \rangle = \|u - s\|^2 - \|u - r\|^2 - \|r - s\|^2$$

with $(r, s, u) = (q_{t+1}, q_t, z)$ gives

$$\eta \langle c_t, z - q_{t+1} \rangle + S_t^{\mathrm{c}} = \frac{1}{2} \big(\|z - q_t\|^2 - \|z - q_{t+1}\|^2 - \|q_{t+1} - q_t\|^2 \big).$$

Similarly, the inclusion at p_t gives

$$\eta \langle c_{t-1}, q_{t+1} - p_t \rangle + S_t^{\mathrm{p}} = \langle p_t - q_t, q_{t+1} - p_t \rangle ,$$

and the same identity yields

$$\eta \langle c_{t-1}, q_{t+1} - p_t \rangle + S_t^{\mathrm{p}} = \frac{1}{2} \big(\|q_{t+1} - q_t\|^2 - \|q_{t+1} - p_t\|^2 - \|p_t - q_t\|^2 \big).$$

Algorithm 2 Adaptive Behavioral OGDA with step-size halving

Input: initial behavioral profile π^{init} and initial step size $\eta_1 > 0$
Output: global output sequences $\{\pi^n_{\text{out}}\}_{n\geq 1}$ and $\{\bar{\pi}^n_{\text{out}}\}_{n\geq 1}$
$\sigma^1 \leftarrow \pi^{\text{init}}$ and $n \leftarrow 0$
for $r = 1, 2, \ldots$ **do**
  $\eta_r \leftarrow 2^{1-r}\eta_1$
  $\pi^{r,0} \leftarrow \sigma^r$ and $\bar{\pi}^{r,0} \leftarrow \sigma^r$
  $S_{r,0} \leftarrow 0$ and $W_{r,0} \leftarrow 0$
  for $t = 1, 2, \ldots$ **do**
    $\{v^{r,t-1}_{i,I}\}_{i,I} \leftarrow \textsc{CounterfactualValues}(\pi^{r,t-1})$
    from $\pi^{r,t-1}$, $\bar{\pi}^{r,t-1}$, and $\{v^{r,t-1}_{i,I}\}_{i,I}$, compute $\bar{\pi}^{r,t}$ and $\pi^{r,t}$ by the two projection updates of Algorithm 1 with step size η_r
    $S_{r,t} \leftarrow S_{r,t-1} + \sum_{i,I}\left[\eta_r^2\left\langle v^{r,t-1}_{i,I}, \pi^{r,t-1}_{i,I} - \bar{\pi}^{r,t-1}_{i,I}\right\rangle^2 + \left\|\bar{\pi}^{r,t}_{i,I} - \pi^{r,t-1}_{i,I}\right\|^2\right]$
    $W_{r,t} \leftarrow W_{r,t-1} + \sum_i \left\|\mathcal{R}_i(\pi^{r,t-1}_i) - \mathcal{R}_i(\bar{\pi}^{r,t-1}_i)\right\|^2$
    $n \leftarrow n + 1$
    if $\max\{S_{r,t}/\eta_r, 11L\eta_r W_{r,t}\} > 5\bar{\Phi}$ **then**
      $\sigma^{r+1} \leftarrow \pi^{r,t-1}$
      $\pi^n_{\text{out}} \leftarrow \sigma^{r+1}$ and $\bar{\pi}^n_{\text{out}} \leftarrow \sigma^{r+1}$
      break ▷ Discard both candidates at iteration t
    else
      $\pi^n_{\text{out}} \leftarrow \pi^{r,t}$ and $\bar{\pi}^n_{\text{out}} \leftarrow \bar{\pi}^{r,t}$
    end if
  end for
end for

Finally,

$$\langle c_t, z - p_t\rangle = \langle c_t, z - q_{t+1}\rangle + \langle c_{t-1}, q_{t+1} - p_t\rangle + \langle c_t - c_{t-1}, q_{t+1} - p_t\rangle .$$

Adding the corrector and predictor equalities and using the preceding decomposition proves the claimed projection identity. No inequality has been used except the displayed nonnegativity of the normal-cone slacks. □

C Proofs for Theorem 6.4

Throughout this section, constants refer to the explicit constants of section B for $\bar{\eta} = 1$. Each run started by Algorithm 2 follows Algorithm 1 from its initial behavioral profile with a fixed step size until any restart. We apply results from section B to the corresponding fixed-step continuation only when their hypotheses, including any step-size restrictions, are satisfied. For one run, omit the run index. For every $t \geq 0$, let

$$\widetilde{K}^t_{i,I} := \eta\left\langle v_{i,I}(\pi^t), \pi^t_{i,I} - \bar{\pi}^t_{i,I}\right\rangle \tag{41}$$

be the inner product evaluated by the adaptive test, and define

$$S_T := \sum_{t=0}^{T-1}\sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\left[(\widetilde{K}^t_{i,I})^2 + \left\|\bar{\pi}^{t+1}_{i,I} - \pi^t_{i,I}\right\|^2\right], \qquad W_T := \sum_{t=0}^{T-1}\sum_{i=1}^{2}\left\|Z^t_i - \bar{Z}^t_i\right\|^2 . \tag{42}$$

By (15), (41), and (11), the two terms inside the maximum in (10) are S_T/η and $11L\eta W_T$.

C.1 Proof of Theorem 6.4

Proof of Theorem 6.4. We first treat the cases $\bar{\Phi} = 0$ and $L = 0$. If $\bar{\Phi} = 0$, the players have no information sets. Hence, both terms in (10) vanish, and every conclusion holds trivially. If $L = 0$, then the oracle norm bound in Lemma B.17 gives $v_{i,I} \equiv 0$ for every i and I. Then, every projection

in (2) receives a point of its simplex and returns it unchanged. Hence, both sequences are constant. Moreover, (21) gives $u_1(z) = 0$ for every terminal history z. Hence, every expected utility vanishes. Consequently, every behavioral profile is a Nash equilibrium. Thus, the test never fails. The constant iterates satisfy both limits in theorem 6.4. We now assume that $\bar{\Phi} > 0$ and $L > 0$. Suppose, for contradiction, that Algorithm 2 halves the step size infinitely often. By lemma C.1, there is a threshold $\widehat{\eta}_0 > 0$ such that every fixed-step run with $0 < \eta \le \widehat{\eta}_0$ satisfies (10) at every horizon, uniformly over all initial behavioral profiles. Choose an integer $k \ge 0$ such that $\eta_1 2^{-k} \le \widehat{\eta}_0$. Under the contradiction assumption, run $r = k + 1$ is reached and has step size $\eta_r = \eta_1 2^{-k}$. The uniformity in lemma C.1 allows us to apply it to the initial behavioral profile of this run. Its fixed-step continuation therefore satisfies (10) for every $T \ge 1$. Since the actual run agrees with this continuation until any test failure, no such failure can occur, and this run cannot trigger another halving. This contradicts the assumption of infinitely many halvings. Therefore, Algorithm 2 eventually stops halving the step size.

Consequently, there is a final run. It starts from some behavioral profile and uses some fixed step size $\eta > 0$. The test (10) holds for every $T \ge 1$ on this run. By (42), the first term in the test gives

$$\sum_{t=0}^{\infty}\sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\left[(\widetilde{K}_{i,I}^{t})^2 + \left\|\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^{t}\right\|^2\right] \le 5\bar{\Phi}\eta < \infty. \tag{43}$$

Summing the inequality in Lemma C.2 and applying (43) give

$$\sum_{t=1}^{\infty}\sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\left(K_{i,I}^{t}\right)^2 \le 10\bar{\Phi}\eta\left(1+\eta^2L^2\right) < \infty.$$

By (24) and (18), $\bar{Z}_i^{t+1} - Z_i^t = \mathcal{L}_i(b_i^t; w_i, \bar{\pi}_i^{t+1})$ for the weights $w_{i,I} := Z_{i,p(I)}^t \in [0,1]$. Hence, part (b) of lemma B.13 and (43) give

$$\sum_{t=0}^{\infty}\sum_{i=1}^{2}\left\|\bar{Z}_i^{t+1} - Z_i^t\right\|^2 \le C_\Gamma \sum_{t=0}^{\infty}\sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\left\|\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^{t}\right\|^2 < \infty.$$

By (42), the second term in the test gives

$$\sum_{t=0}^{\infty}\sum_{i=1}^{2}\left\|Z_i^t - \bar{Z}_i^t\right\|^2 \le \frac{5\bar{\Phi}}{11L\eta} < \infty.$$

Therefore, (23) and the square summability of $\bar{Z}_i^{t+1} - Z_i^t$ established above give

$$\sum_{t=1}^{\infty}\sum_{i=1}^{2}\left\|Z_i^t - Z_i^{t-1}\right\|^2 \le 2\sum_{t=1}^{\infty}\sum_{i=1}^{2}\left\|Z_i^t - \bar{Z}_i^t\right\|^2 + 2\sum_{t=0}^{\infty}\sum_{i=1}^{2}\left\|\bar{Z}_i^{t+1} - Z_i^t\right\|^2 < \infty.$$

The square-summability bounds for $K_{i,I}^t$ and $Z_i^t - Z_i^{t-1}$, together with (43), imply the hypotheses of Lemma B.7. Hence, lemma B.7 gives the distance and realization-plan limits for the final run. The outputs before the final run form a finite prefix. Therefore, the two limits in theorem 6.4 hold for the global output sequences. □

Lemma C.1 (The test holds below a positive threshold)**.** *Suppose that $\bar{\Phi} > 0$ and $L > 0$. There exists a constant $\widehat{\eta}_0 > 0$ that depends only on the game such that the following claim holds. Let $\{\pi^t\}$ and $\{\bar{\pi}^t\}$ be generated by Algorithm 1 from any initial behavioral profile with a fixed $0 < \eta \le \widehat{\eta}_0$. Then, the test (10) holds for every $T \ge 1$.*

C.2 Proof of Lemma C.1

Proof of Lemma C.1. Fix $0 < \eta \le \eta_0$ and $T \ge 1$. Recall from (8) the reach-weighted energy

$$E_t := \sum_{i=1}^{2}\sum_{I\in\mathcal{I}_i}\rho_i^{\pi_i^{\mathrm{sc}}}(I)\left(\left\|\pi_{i,I}^t - \bar{\pi}_{i,I}^t\right\|^2 + \left\|\bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^t\right\|^2\right)$$

and the duality gap $G_t := \langle F(Z^t), Z^t - Z^{\mathrm{sc}} \rangle$. Since every local simplex has Euclidean diameter at most $\sqrt{2}$ and $\rho_i^{\pi_i^{\mathrm{sc}}}(I) \le 1$,

$$\Phi_1 = \frac{1}{2} \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \left\| \bar{\pi}_{i,I}^{1} - \pi_{i,I}^{\mathrm{sc}} \right\|^2 \le \sum_{i=1}^{2} \sum_{I \in \mathcal{I}_i} \rho_i^{\pi_i^{\mathrm{sc}}}(I) \le \bar{\Phi}.$$

Set $P_* := \bar{\Phi} + C_0$. Since $\eta \le \eta_0 \le 1$, lemma B.4 applied with horizon $T+1$ gives

$$\frac{\eta}{2} \sum_{t=1}^{T} G_t + \frac{1}{4} \sum_{t=1}^{T} E_t \le \bar{\Phi} + C_0 \eta^2 \le P_*.$$

Consequently,

$$\sum_{t=1}^{T} E_t \le 4P_*, \qquad \sum_{t=1}^{T} G_t \le \frac{2P_*}{\eta}. \tag{44}$$

Define

$$Q_* := J_0 + \left(\frac{4J_E}{\rho_{\mathrm{sc}}} + \frac{2J_G}{\delta} \right) P_*.$$

The term involving δ is omitted when the set defining δ is empty. By lemma B.5 with horizon $T+1$ and (44),

$$\begin{aligned} \sum_{t=1}^{T} \sum_{i,I} \left[(K_{i,I}^t)^2 + \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^{t} \right\|^2 \right] &\le J_0 \eta^3 + \frac{4J_E P_*}{\rho_{\mathrm{sc}}} \eta^2 + \frac{2J_G P_*}{\delta} \eta^3 \\ &\le Q_* \eta^2. \end{aligned} \tag{45}$$

Define

$$M_* := \left(\frac{4\widehat{C}_d^M}{\rho_{\mathrm{sc}}} + \frac{2\widehat{B}_d^M}{\delta} \right) P_*.$$

The term involving δ is again omitted when its defining set is empty. By Lemma B.6, summed over $1 \le t \le T$, and (44),

$$\sum_{t=1}^{T} \mathfrak{M}_t \le M_*. \tag{46}$$

Equations (25), (31), (27), (45), and (46) give

$$\sum_{t=1}^{T} \sum_{i} \left\| Z_i^t - Z_i^{t-1} \right\|^2 \le D_* := 4M_* + 2M_\partial, \qquad \sum_{t=1}^{T} \sum_{i,I} \left\| \bar{\pi}_{i,I}^{t} - \pi_{i,I}^{t-1} \right\|^2 \le (B_\partial + Q_*)\, \eta^2.$$

We next bound the inner products monitored by the test. Summing lemma B.9 over the information sets and using $n_{i,I} \le \bar{\Phi}$ gives

$$\begin{aligned} &\eta^2 \sum_{t=1}^{T} \sum_{i,I} \left\| v_{i,I}(\pi^t) - v_{i,I}(\pi^{t-1}) \right\|^2 \\ &\quad \le 2\bar{\Phi} L^2 \eta^2 \sum_{t=1}^{T} \sum_{i} \left\| Z_i^t - Z_i^{t-1} \right\|^2 + 4\bar{\Phi}^2 \sum_{t=1}^{T} \sum_{i,I} (K_{i,I}^t)^2 \\ &\qquad + 4\bar{\Phi}^2 L^2 \eta^2 \sum_{t=1}^{T} \sum_{i,I} \left\| \bar{\pi}_{i,I}^{t} - \pi_{i,I}^{t-1} \right\|^2 \\ &\quad \le \eta^2 V_*, \end{aligned}$$

where

$$V_* := 2\bar{\Phi} L^2 D_* + 4\bar{\Phi}^2 Q_* + 4\bar{\Phi}^2 L^2 \left(B_\partial + Q_* \right).$$

For every $t \geq 1$, (16) and (41) give

$$\widetilde{K}_{i,I}^{t} = K_{i,I}^{t} + \eta \left\langle v_{i,I}(\pi^{t}) - v_{i,I}(\pi^{t-1}), \pi_{i,I}^{t} - \bar{\pi}_{i,I}^{t} \right\rangle .$$

Projection nonexpansiveness gives $\left\| \pi_{i,I}^{t} - \bar{\pi}_{i,I}^{t} \right\| \leq \eta L$. Hence, (45) and the oracle-variation estimate above imply

$$\sum_{t=1}^{T-1} \sum_{i,I} (\widetilde{K}_{i,I}^{t})^2 \leq 2Q_* \eta^2 + 2L^2 V_* \eta^4.$$

The initialization gives $\widetilde{K}_{i,I}^{0} = 0$ and $Z_i^0 = \bar{Z}_i^0$. Therefore, (42), (27), (45), (46), and the bound just obtained for the sum of squared $\widetilde{K}_{i,I}^{t}$ terms give

$$S_T \leq C_* \eta^2, \qquad W_T \leq M_*, \qquad C_* := B_\partial + 3Q_* + 2L^2 V_*.$$

Define

$$\widehat{\eta}_0 := \min \left\{ \eta_0, \frac{5\bar{\Phi}}{C_*}, \frac{5\bar{\Phi}}{11 L M_*} \right\} > 0,$$

where an entry with a zero denominator is omitted. If $0 < \eta \leq \widehat{\eta}_0$, then the bounds $S_T \leq C_* \eta^2$ and $W_T \leq M_*$ give

$$\frac{S_T}{\eta} \leq 5\bar{\Phi}, \qquad 11 L \eta W_T \leq 5\bar{\Phi}.$$

Thus, the test holds for every $T \geq 1$ and every initial behavioral profile. □

Lemma C.2 (Comparison of the monitored and predictor inner products)**.** *For every $t \geq 0$, every player i, and every information set $I \in \mathcal{I}_i$,*

$$\left(K_{i,I}^{t+1} \right)^2 \leq 2 \left(\widetilde{K}_{i,I}^{t} \right)^2 + 2\eta^2 L^2 \left\| \bar{\pi}_{i,I}^{t+1} - \pi_{i,I}^{t} \right\|^2 .$$

C.3 Proof of Lemma C.2

Proof of Lemma C.2. Fix $t \geq 0$, a player i, and an information set $I \in \mathcal{I}_i$. Suppress the player and information-set indices, and set

$$q := \bar{\pi}^{t}, \quad p := \pi^{t}, \quad c := v(\pi^{t}), \quad q' := \bar{\pi}^{t+1}, \quad p' := \pi^{t+1}.$$

Also set $d := q' - q$, $e := p' - q'$, $a := p - q$, and $b := q' - p$. Then, $d = a + b$. The projection optimality condition for p' gives

$$K^{t+1} = \eta \langle c, e \rangle \geq \|e\|^2 \geq 0.$$

The projection optimality condition for q', applied with comparator p', gives $K^{t+1} \leq \langle d, e \rangle$. Projection nonexpansiveness gives $\|e\| \leq \|d\|$. Hence, $K^{t+1} \leq \|d\|^2$. The projection optimality condition for q', applied with comparator q, gives $\|d\|^2 \leq \eta \langle c, d \rangle$. Consequently,

$$\begin{aligned} 0 \leq K^{t+1} &\leq \eta \langle c, a + b \rangle \\ &= \widetilde{K}^{t} + \eta \langle c, b \rangle \\ &\leq \left| \widetilde{K}^{t} \right| + \eta L \|b\| . \end{aligned}$$

Squaring the last inequality proves the lemma. □

D Additional Experimental Results

Figure 2 shows the complete step-size sweep used in the experiments. The rows correspond to Behavioral OGDA, dilated OGDA, Behavioral GDA, and Adaptive Behavioral OGDA. Each column is one of the five benchmark games. The nine curves use $\eta \in \{1/40, 1/20, 1/10, 1/5, 2/5, 4/5, 8/5, 16/5, 32/5\}$. The same color denotes the same step size in every row.

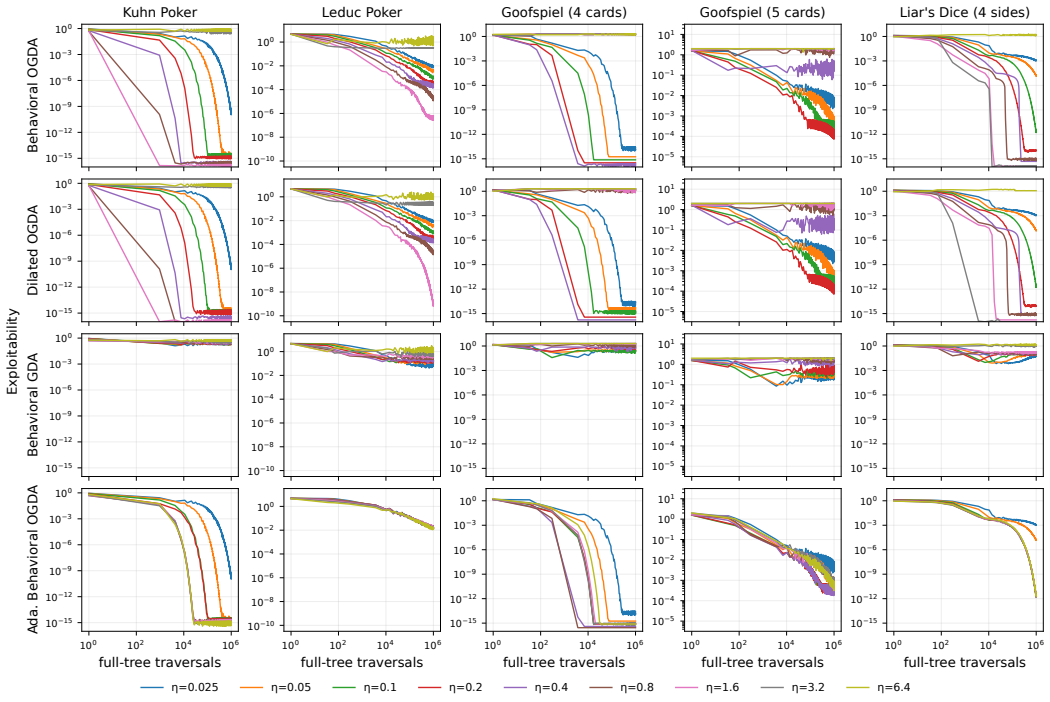

Figure 2: Complete sweep of exploitability for the played profile at the current iterate. The horizontal axis counts evaluations of counterfactual values over the full game tree.