CONSTANT SWAP REGRET IN GENERAL-SUM GAMES VIA TWO-SCALE HIGHER-ORDER OPTIMISM

Anonymous authors
Paper under double-blind review

ABSTRACT

We give deterministic and uncoupled learning dynamics for finite multiplayer general-sum games under full-information feedback that achieve constant individual swap regret in self-play, independent of the horizon T. With n players and at most m actions each, every player's individual swap regret is $O(\sqrt{n}\, m \log m \log^{5/2}(nm))$ at every finite horizon. The dynamics use the classical Blum–Mansour framework with optimism. Each player predicts the deviation gains, uses these predictions to update a row-stochastic transition matrix, and plays its stationary distribution. Our new ingredients include a tailored row normalization map and a two-scale higher-order predictor. An adversarially robust variant, obtained through a generic common-prefix switching wrapper, preserves the self-play bound up to a universal constant and guarantees individual swap regret at most $7\sqrt{mT \log m}$ in the adversarial setting.

1 INTRODUCTION

Online learning provides a general framework for sequential decision-making in an unknown and potentially changing environment: at each round, a learner acts, observes feedback, and updates its behavior. Its central performance measure is regret, which measures the learner's performance against a prescribed class of hindsight benchmarks. This viewpoint is especially natural in repeated games, where each player is an online learner and the feedback faced by one player is generated by the evolving behavior of the others. Correspondingly, no-regret guarantees translate into convergence toward game-theoretic equilibrium notions.

External regret compares with the best fixed action in hindsight, while swap regret allows a deviation map and compares against the best fixed deviation map in hindsight. If every player has small average swap regret, the time-averaged product distribution of play is an approximate correlated equilibrium (CE) (Aumann, 1974, Blum & Mansour, 2007). A constant individual swap-regret bound gives an $O(1/T)$ CE rate (see Appendix A.1).

In the adversarial setting, the Blum–Mansour reduction, which constructs a swap-regret learner from external-regret learners (Blum & Mansour, 2007), yields $O(\sqrt{mT \log m})$ swap regret. In the self-play setting, however, the players mutually determine each other's payoff vectors. This interaction allows better swap regret bounds (Chen & Peng, 2020, Anagnostides et al., 2022a,b, Tsuchiya, 2026). In light of the recent constant external-regret results (Liu et al., 2026, Abbadi et al., 2026), we ask whether swap-regret admits the same guarantee: *Can uncoupled learning dynamics achieve $O_{n,m}(1)$ individual swap regret in general-sum games, while retaining $O(\sqrt{mT \log m})$ individual swap regret in the adversarial setting?*

Results. Under full-information feedback, we give an affirmative answer. Let n be the number of players and m_i be the number of actions of player i. For $m_i \le m$, every player's individual swap regret is $O\Big(\sqrt{n}\, m \log m \log^{5/2}(nm)\Big)$ at every finite horizon. An adversarially robust variant, obtained through a generic common-prefix switching wrapper, preserves the self-play bound up to a universal constant and gives individual swap regret at most $7\sqrt{m_i T \log m_i}$ in the adversarial setting.

Technical Overview. We use the Blum–Mansour (BM) reduction for swap regret. For each source action a, a player maintains a distribution over the destination actions b. These distributions form the

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.

rows of a transition matrix, and the player uses its stationary distribution as its mixed strategy. The deviation gain assigned to changing action a to action b is $x_{i,a}(v_{i,b} - v_{i,a})$, where $\boldsymbol{x}_i$ and $\boldsymbol{v}_i$ denote the mixed strategy and the exact expected payoff of player i respectively. Therefore, controlling the best deviation in each row controls the player's swap regret. We update each row optimistically by combining cumulative gains with a prediction of the next gain into scores. A row normalization then converts these scores into a distribution over destination actions.

Let η_i be player i's learning rate. Let E^2 be an upper bound on each player's cumulative squared error in forecasting its row gains. Let $S_i \geq 0$ denote player i's cumulative row curvature. S_i is the local Hessian quadratic form of a row potential, evaluated in the direction of the observed row gain. It measures the cumulative sensitivity of the row probabilities to changes in their scores. For our choice of row normalization, a potential analysis gives

$$\eta_i \operatorname{SwapReg}_i(T) + S_i \leq 2m_i \log m_i + \eta_i^2 E^2. \tag{1.1}$$

The main remaining task is to bound E independently of T. A natural idea is to use a constant external-regret learner in every BM row. However, the constant external-regret results rely on the feedback generated when all players follow the prescribed dynamics in a fixed game (Liu et al., 2026, Abbadi et al., 2026). They do not guarantee constant regret for arbitrary feedback sequences.

In BM, each row learner chooses a distribution over destination actions, but this is not the strategy played in the game. The player instead uses the stationary distribution of the entire transition matrix. Therefore, changing one row can change the stationary probabilities and hence the feedback to other rows. Thus the row learners interact differently from players running the original external-regret dynamics. BM still converts the row learners' external regret into swap regret, but their regret on these new feedback sequences requires a separate analysis. Earlier BM analyses use the Markov-chain tree theorem (Anantharam & Tsoucas, 1989) to address this issue (Anagnostides et al., 2022a,b).

We build on this tree representation while extending the higher-order optimism of Liu et al. (2026), Abbadi et al. (2026) to BM. Our predictor's error is a filtered higher-order difference of the deviation gains, so bounding it requires understanding how these gains change as the rows are updated. The additional difficulty is that successive differences also act on the stationary distribution, producing interactions between different rows of the same player. To bound E independently of T, our analysis must relate these interactions to the row curvature supplied by the potential bound. At the same time, we must retain the desired dependence on n and m. Grouping the changes in the gains only by player is too coarse to capture the interactions between rows. On the other hand, working immediately with all individual transitions would incur the full action-dependent sensitivity of the stationary distribution and lead to poor dependence on m. We address this tension with a two-scale predictor, matched by a two-stage expansion analysis: first attenuate the terms at the player level, then use curvature to close the remaining transition-level expansion.

The Markov-chain tree theorem represents stationary play as an average over rooted directed trees. The first stage of our analysis keeps the change grouped by player and exploits centering across players. Concretely, differentiating the product-tree distribution yields centered player contributions that are independent across players, so an L^2 moment bound combines their coefficients in quadrature. Thus, the learning rates combine through $\left(\sum_i \eta_i^2\right)^{1/2}$, rather than $\sum_i \eta_i$, which preserves a final $\sqrt{n}$ dependence. After $\ell = \Theta(\log(nm))$ such backward-difference steps, the total weight of the terms that still require a transition-level sensitivity bound has been attenuated by a factor of order κ^ℓ, where $0 < \kappa < 1$ is a sufficiently small universal constant. The total weight is then small enough to apply the more expensive, action-dependent curvature bound without spoiling the dependence on m.

In the second stage, the proof records the individual transition indices (i, a, b) encountered as further differences are taken. If an index has appeared before, its new contribution can be paired with the earlier one, and this pair is bounded by the curvature energy. Thus only a newly encountered transition index needs to be tracked further. Since there are only $N = \sum_i m_i^2$ transition indices, there can be at most N fresh-index continuations. The rooted-tree representation also provides the sparsity needed to avoid extra factors of m: each sampled gain is associated with a single source action, and a directed tree contains at most one outgoing edge from that source action. The predictor's two commuting exponential moving average (EMA) cascades, of depths ℓ and N, mirror the player-level and transition-level stages, respectively.

To close the argument, let Y^2 denote a suitable weighted aggregate of the curvature energies S_i. Since swap regret is nonnegative, from (1.1), we can show that

$$Y^2 \le B + \alpha E^2,$$

where B and α may depend on the game dimensions but not on T. The analysis of the predictor gives a complementary bound

$$E \le C + \beta E + \tau_\ell Y,$$

where C and β are constants, and τ_ℓ is exponentially small in the depth $\ell = \Theta(\log(nm))$ of the first stage. Choosing the parameters so that $\tau_\ell\sqrt{B} = O(1)$ and $\beta + \tau_\ell\sqrt{\alpha} < 1$ gives $E = O(1)$. Substituting back in (1.1) gives the constant swap-regret bound.

Our design of the row normalization makes all parts of the argument fit together. Its associated potential has initialization cost $O(\log m_i)$ per row and supplies the curvature term in (1.1). The same map provides the derivative bounds needed in the first stage and the bounds in terms of curvature needed in the second stage, including when a transition repeats. Crucially, these bounds use the same curvature that the potential argument controls.

Contributions relative to prior work. Our analysis combines established techniques. The Blum–Mansour reduction provides the underlying framework (Blum & Mansour, 2007). The use of Markov chain tree theorem to analyze stationary play appears in Anagnostides et al. (2022a,b). In particular, Anagnostides et al. (2022a) use the tree representation to bound higher-order differences through the BM fixed point. Higher-order prediction for constant external regret and recursions that record player labels appear in Liu et al. (2026), Abbadi et al. (2026).

Our main technical contribution is to make these techniques fit together while controlling the dependence on the game dimensions. The row normalization supplies the derivative bounds used in both stages. The first stage groups the expansion by player and reduces the coefficient passed to the transition-level analysis. The second stage then tracks individual transitions and uses curvature to stop a branch when a transition repeats. This yields Lemma 4.2, in which the coefficient of the curvature energy is exponentially small in the first-stage depth.

Triggered switches to adversarial dynamics also appear in Anagnostides et al. (2022b), Tsuchiya (2026), Abbadi et al. (2026). Such an immediate switch retains the self-play budget additively. Our wrapper adds a common adversarial prefix before a fresh simultaneous start of the base dynamics, then permanently switches any player whose base-phase regret crosses its threshold to a freshly initialized adversarial fallback. The prefix absorbs the base budget into the all-horizon $O(\sqrt{m_i T \log m_i})$ envelope, while the anytime self-play guarantee prevents threshold crossings in self-play. The construction is black-box in the base dynamics and applies whenever the common-prefix absorption condition holds. Section 5 and Appendix I give the construction and proof.

Table 1 compares selected individual swap-regret guarantees under full-information feedback. Appendix A.5 provides broader related work and detailed comparisons.

Table 1: Comparisons of selected algorithms for individual swap regret. BM denotes Blum–Mansour, and SL denotes Stoltz–Lugosi. $\widetilde{O}$ suppresses logarithmic factors in m, n, T.

Reference	Algorithm	Self-play setting	Adversarial setting
Blum & Mansour (2007)	BM–MWU	$O(\sqrt{mT\log m})$	$O(\sqrt{mT\log m})$
Chen & Peng (2020)	BM–OMWU	$O(\sqrt{n}\,(m\log m)^{3/4}T^{1/4})$	$\widetilde{O}(\sqrt{mT} + \sqrt{n}\,m^{3/4}T^{1/4})$
Anagnostides et al. (2022a)	SL–OMWU / adaptive variant	$O(nm\log m\log^4 T)$	$O(nm\log m\log^4 T + m\sqrt{T\log m})$
Anagnostides et al. (2022b)	BM–OFTRL (log-barrier)	$O(nm^{5/2}\log T)$	$O(nm^{5/2}\log T + \sqrt{mT\log m})$
Tsuchiya (2026)	BM–OFTRL (hybrid)	$O(nm^2\sqrt{\log m\log T})$	$O(nm^2\sqrt{\log m\log T} + \sqrt{mT\log m})$
This work	BM–OFTRL with two-scale higher-order optimism	$O\Big(\sqrt{n}\,m\log m\log^{5/2}(nm)\Big)$	$O(\sqrt{mT\log m})$

2 MODEL AND MAIN RESULTS

Player $i \in [n]$ has $m_i \geq 2$ actions and a fixed payoff function $u_i : \prod_j [m_j] \to [0,1]$. Let $m = \max_i m_i$. At round t, players simultaneously choose $\boldsymbol{x}_i^{(t)} \in \Delta_{m_i}$, and player i observes every exact expected payoff $v_{i,a}^{(t)} = u_i(a, \boldsymbol{x}_{-i}^{(t)})$. Each strategy is a deterministic function of public action counts, the player's own state, and past payoff vectors. Thus the dynamics are uncoupled and require no horizon T. In the adversarial setting, $\boldsymbol{v}_i^{(t)}$ may be any vector in $[0,1]^{m_i}$.

Let $\mathcal{I}_i = \{(i,a,b) : a,b \in [m_i]\}$, $\mathcal{I} = \bigcup_i \mathcal{I}_i$, and $N = |\mathcal{I}| = \sum_i m_i^2$. Define

$$r_{i,a,b}^{(t)} = x_{i,a}^{(t)}(v_{i,b}^{(t)} - v_{i,a}^{(t)}),$$
$$\mathrm{SwapReg}_i(T) = \max_{\varphi:[m_i]\to[m_i]} \sum_{t=1}^{T} \sum_a r_{i,a,\varphi(a)}^{(t)} = \sum_a \max_b \sum_{t=1}^{T} r_{i,a,b}^{(t)} \geq 0. \tag{2.1}$$

The last inequality follows from using the identity function φ. Let

$$A_i = m_i \log m_i, \quad A = \sum_i A_i, \quad \ell = \ell_0 + \lceil \log(2+n+N) \rceil, \quad k = \ell + 1, \quad g = c/k^{5/2}, \tag{2.2}$$

with universal $c > 0$ sufficiently small and integer ℓ_0 sufficiently large. The default rates are

$$\eta_i = g\sqrt{A_i/A}, \qquad \sum_i \eta_i^2 = g^2. \tag{2.3}$$

Theorem 2.1 (Constant swap regret)**.** *There exist universal choices of c, ℓ_0 such that, in every finite game, simultaneous use of the base dynamics in Algorithm 1 gives every player i, at every $T \geq 1$,*

$$\mathrm{SwapReg}_i(T) \leq B_i := \frac{4}{c} k^{5/2} \sqrt{A_i A} = O\Big(\sqrt{n}\, m \log m \log^{5/2}(nm)\Big).$$

The dynamics are deterministic, uncoupled, and horizon-free.

For highly unequal action counts, an alternative rate allocation can improve the base dynamics' worst-player bound by a factor of $n^{1/4}$. Appendix A.4 gives the full argument. Next we state the adversarial robustness result.

Theorem 2.2 (Adversarial robustness)**.** *The common-prefix wrapper of Section 5, instantiated with the base dynamics and default rates, satisfies, for every player i and every $T \geq 1$,*

$$\mathrm{SwapReg}_i(T) \leq 7\sqrt{A_i T} \quad \textit{against arbitrary payoff vectors},$$
$$\mathrm{SwapReg}_i(T) \leq 4B_i \quad \textit{in self-play}.$$

3 TWO-SCALE OPTIMISTIC TRANSITION-MATRIX DYNAMICS

This section presents the base dynamics. In each round, player i combines a prediction of its next pairwise deviation-gain matrix with its cumulative gains into a score matrix. A smooth row-normalization function converts each score row into a probability distribution. These probability rows form a transition matrix whose stationary distribution is played. Sections 3.1 and 3.2 define the normalization function and predictor, respectively, before Algorithm 1 combines them.

3.1 ROW NORMALIZATION AND STATIONARY PLAY

First we explain the normalization process for a score row $\boldsymbol{z} \in \mathbb{R}^{m_i}$. The normalization map must satisfy certain properties to support both a potential argument and a higher-order prediction argument later in the proof. These properties are established in Lemmas B.1, B.2, and C.1. Let δ, β_i and d_i be public parameters defined in Appendix A.2. For $s \in \mathbb{R}$, let

$$u(s) = \frac{\sqrt{s^2+4} - s}{2}, \qquad t_+(s) = u(s)^{-1}, \qquad f_i(s) = \beta_i \frac{(1 + t_+(s)/d_i)^{d_i}}{1 + u(s) + (\delta/4)u(s)^2}. \tag{3.1}$$

The function f_i is smooth, positive, strictly increasing, and ranges from zero to infinity. A score row $\boldsymbol{z} \in \mathbb{R}^{m_i}$ determines a unique shift $\mu_i(\boldsymbol{z})$ through

$$q_{i,b}(\boldsymbol{z}) = f_i(z_b - \mu_i(\boldsymbol{z})), \qquad \sum_b f_i(z_b - \mu_i(\boldsymbol{z})) = 1. \tag{3.2}$$

3.2 TWO-SCALE PREDICTION AND IMPLEMENTATION

Let S be the backward shift and $\mathrm{D} = I - \mathrm{S}$, with zero prehistory. The predictor $\widehat{\boldsymbol{r}}$ is calculated as:

$$\rho_L = \frac{L}{L+1}, \quad \mathrm{R}_L = (I - \rho_L \mathrm{S})^{-1}, \quad \mathrm{F}_L = \mathrm{R}_L \mathrm{D}, \quad \mathrm{H} = \mathrm{D}^3 \mathrm{F}_\ell^\ell \mathrm{F}_N^N, \quad \widehat{\boldsymbol{r}} = (I - \mathrm{H})[\boldsymbol{r}]. \tag{3.3}$$

Every factor of H has current-input coefficient one, so the current input cancels in $I - \mathrm{H}$: the predictor uses only past inputs. The two cascades have depths ℓ and N, matching the two analysis stages. Their filters commute, so Algorithm 1 applies the N-stage cascade before the ℓ-stage cascade without changing the predictor. Appendix H.1 verifies the state-based implementation in Algorithm 1.

Algorithm 1 Base learning dynamics for player i

Input. Parameters from (2.2)–(2.3) and Appendix A.2.
State. Initialize $m_i \times m_i$ arrays $\boldsymbol{\theta}, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{N+\ell}, \boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ to zero. Set $\nu_j = (N+1)^{-1}$ for $j \le N$ and $\nu_j = (\ell+1)^{-1}$ otherwise.

for $t = 1, 2, \ldots$ **do**
  $\widehat{\boldsymbol{r}} \leftarrow \sum_{j=1}^{N+\ell} \boldsymbol{a}_j + 3\boldsymbol{c}_1 - 3\boldsymbol{c}_2 + \boldsymbol{c}_3$
  $\overline{\boldsymbol{z}} \leftarrow \boldsymbol{\theta} + \eta_i \widehat{\boldsymbol{r}}$
  Normalize each row by (3.2) or Appendix H and form $\boldsymbol{Q}$.
  Solve $\boldsymbol{Q}^\top \boldsymbol{x} = \boldsymbol{x}$, $\sum_a x_a = 1$, and play $\boldsymbol{x}$.
  Observe $\boldsymbol{v}$ and set $r_{a,b} \leftarrow x_a(v_b - v_a)$ for every a, b.
  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \eta_i \boldsymbol{r}$, then $\boldsymbol{u} \leftarrow \boldsymbol{r}$
  for $j = 1, \ldots, N + \ell$ **do**
    $\boldsymbol{u} \leftarrow \boldsymbol{u} - \boldsymbol{a}_j$, then $\boldsymbol{a}_j \leftarrow \boldsymbol{a}_j + \nu_j \boldsymbol{u}$
  end for
  $(\boldsymbol{c}_3, \boldsymbol{c}_2, \boldsymbol{c}_1) \leftarrow (\boldsymbol{c}_2, \boldsymbol{c}_1, \boldsymbol{u})$ simultaneously
end for

Score changes. For exact normalization, let $\boldsymbol{\xi}_i = \mathrm{H}[\boldsymbol{r}_i]$ and $\boldsymbol{\theta}_i^{(t)} = \eta_i \sum_{s \le t} \boldsymbol{r}_i^{(s)}$. Then

$$\boldsymbol{z}_i = \boldsymbol{\theta}_i - \eta_i \boldsymbol{\xi}_i, \qquad \mathrm{D}\boldsymbol{z}_i = \eta_i(\boldsymbol{r}_i - \mathrm{D}\boldsymbol{\xi}_i). \tag{3.4}$$

Thus each score change splits into a gain direction $\eta_i \boldsymbol{r}_i$ and an error direction $-\eta_i \mathrm{D}\boldsymbol{\xi}_i$. For the finite solver, the same identity holds with its effective scores and $\boldsymbol{\xi}_i = \mathrm{H}[\boldsymbol{r}_i] - \boldsymbol{\varepsilon}_i$. Appendix H.2 supplies the summable-error bound, covering both implementations.

4 ANALYSIS OF CONSTANT SWAP REGRET

Fix a finite horizon T and positive public rates with

$$\sum_i \eta_i^2 \le g^2. \tag{4.1}$$

For each player i and each $\boldsymbol{w}_i = (w_{i,a,b})_{a,b \in [m_i]}$, let

$$\|\boldsymbol{w}_i\|_{\infty,1} = \sum_a \max_b |w_{i,a,b}|, \qquad \|\boldsymbol{w}_i\|_{\infty,1;T}^2 = \sum_{t=1}^{T} \left\| \boldsymbol{w}_i^{(t)} \right\|_{\infty,1}^2.$$

The deviation gains satisfy $\left\| \boldsymbol{r}_i^{(t)} \right\|_{\infty,1} \le 1$. Let $E = \max_i \|\boldsymbol{\xi}_i\|_{\infty,1;T}$ be the cumulative prediction error. To define the curvature energy that will control it, for a normalized row $\boldsymbol{q}_i(\boldsymbol{z})$, define $h_i(f_i(s)) = f_i'(s)$ and

$$h_b = h_i(q_b), \qquad \sigma_b = \frac{h_b}{\sum_c h_c}, \qquad \widetilde{w}_b = w_b - \langle \boldsymbol{\sigma}, \boldsymbol{w} \rangle, \qquad \boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w}) = \sum_b h_b \widetilde{w}_b^2. \tag{4.2}$$

At the effective scores, set $b_{i,a}^{(t)} = (\eta_i^2/8) \boldsymbol{K}_{\boldsymbol{z}_{i,a}^{(t)}}(\boldsymbol{r}_{i,a}^{(t)})$ and $S_i = \sum_{t \le T, a} b_{i,a}^{(t)}$. With the universal C_ω fixed in Appendix A.3, write

$$\omega_i = C_\omega (m_i - 1)/\beta_i, \qquad Y^2 = \sum_i \omega_i S_i, \qquad V = \sum_i \omega_i A_i, \qquad \Lambda^2 = \sum_i \omega_i \eta_i^2.$$

The proof rests on two bounds relating E and Y. First, the row-potential argument gives

$$\eta_i \,\mathrm{SwapReg}_i(T) + S_i \le 2A_i + \eta_i^2 E^2, \qquad Y^2 \le 2V + \Lambda^2 E^2.$$

Second, the self-play prediction argument gives

$$E \le C_0 + C_1 g E + \tau_\ell Y, \qquad \kappa := C_E g k^{5/2}, \qquad \tau_\ell := C_P k^2 \kappa^\ell,$$

for universal constants C_E, C_0, C_1, C_P. Notice that the dimension-dependent quantity Y enters only after its coefficient has been reduced exponentially in the first-stage depth ℓ. The public parameter choices allow the two bounds to be absorbed, yielding a constant bound on E and hence the claimed swap-regret bound. Section 4.1 states the two bounds formally and carries out the absorption. Section 4.2 proves the row-potential bound. Sections 4.3–4.4 prove the self-play prediction bound.

4.1 The energy and prediction bounds imply the theorem

We first state the two bounds formally and show why they suffice. The potential bound holds for arbitrary payoff vectors and controls regret and curvature in terms of prediction error.

Lemma 4.1 (Row-potential bound)**.** *For sufficiently small universal c, every payoff-vector sequence in $[0,1]^{m_i}$ obeys*

$$\eta_i \,\mathrm{SwapReg}_i(T) + S_i \le 2A_i + \eta_i^2 E^2, \qquad Y^2 \le 2V + \Lambda^2 E^2, \qquad \Lambda^2 \le g^2 V. \tag{4.3}$$

For the default rates, $\Lambda^2 = g^2 V/A$.

The complementary bound uses self-play to bound prediction error by the same curvature, with an exponentially small coefficient offsetting the dimension-dependent weights in Y.

Lemma 4.2 (Two-scale prediction bound)**.** *There are universal C_E, C_0, C_1, C_P such that, under* (4.1) *and the public parameter bounds of Appendix A.2, if $\kappa = C_E g k^{5/2} \le 1/16$ and $\tau_\ell = C_P k^2 \kappa^\ell$, then in self-play*

$$E \le C_0 + C_1 g E + \tau_\ell Y. \tag{4.4}$$

Appendix A.3 chooses the universal constants so that

$$\kappa \le 1/16, \qquad \tau_\ell \sqrt{2V} \le 1, \qquad (C_1 + 1) g \le 1/2. \tag{4.5}$$

Using $Y \le \sqrt{2V} + \Lambda E$ and $\Lambda \le g\sqrt{V}$, we obtain

$$\begin{aligned} E &\le C_0 + C_1 g E + \tau_\ell\big(\sqrt{2V} + \Lambda E\big) \\ &\le C_0 + 1 + (C_1 + 1) g E. \end{aligned}$$

Since $(C_1 + 1)g \le 1/2$, moving the final term to the left gives $E \le E_0 := 2(C_0 + 1)$. The same choice ensures $\eta_i^2 E_0^2 \le 1 \le A_i$, and (4.3) gives

$$\mathrm{SwapReg}_i(T) \le 3A_i/\eta_i. \tag{4.6}$$

Substituting the default rates gives Theorem 2.1.

4.2 Stationarity and the row-potential bound

We prove Lemma 4.1 by summing the row-potential inequalities and using stationarity to cancel their linear terms. The row map is the gradient of a convex potential Φ_i with Hessian quadratic form $\boldsymbol{K}_{\boldsymbol{z}}$, coordinate domination $\Phi_i(\boldsymbol{z}) \ge \max_b z_b$, and initialization $0 \le \Phi_i(0) \le 2 \log m_i$ (Appendix B). Local curvature comparison and the optimistic score identity give

$$\Phi_i(\boldsymbol{\theta}_{i,a}^{(t)}) - \Phi_i(\boldsymbol{\theta}_{i,a}^{(t-1)}) \le \eta_i \left\langle \boldsymbol{q}_{i,a}^{(t)}, \boldsymbol{r}_{i,a}^{(t)} \right\rangle + \eta_i^2 \left\| \boldsymbol{\xi}_{i,a}^{(t)} \right\|_\infty^2 - b_{i,a}^{(t)}.$$

The linear terms cancel exactly after summing over source rows:

$$\sum_{a,b} q_{i,a,b}^{(t)} r_{i,a,b}^{(t)} = (\boldsymbol{x}_i^{(t)})^\top (\boldsymbol{Q}_i^{(t)} - I) \boldsymbol{v}_i^{(t)} = 0.$$

Telescoping and coordinate domination yield the first inequality in (4.3). Multiplying by ω_i, summing over players, and using $\mathrm{SwapReg}_i(T) \ge 0$ gives the second. The rate budget gives the third. This proves Lemma 4.1. Appendix B.2 gives the detailed calculation. The remaining prediction bound must exploit how the gains depend on all players' stationary strategies.

4.3 PLAYER-CENTERED DIFFERENCES AND EARLY-STAGE ATTENUATION

Appendix H.2 gives $\max_i \|\boldsymbol{\varepsilon}_i\|_{\infty,1;T} \le 1$. Since $\boldsymbol{\xi}_i = \mathrm{H}[\boldsymbol{r}_i] - \boldsymbol{\varepsilon}_i$, the triangle inequality yields $E \le 1 + \max_i \|\mathrm{H}[\boldsymbol{r}_i]\|_{\infty,1;T}$. We analyze the differences in H to bound error terms.

First step: separate error and gain terms. Let $\boldsymbol{T} \sim \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}} = \bigotimes_j \mathbb{P}^{\mathrm{tr}}_{j,\boldsymbol{y}}$ be the product of normalized tree laws, with player-j root $a_j(\boldsymbol{T})$ (Appendix D). Define $\boldsymbol{G}_i = (G_{i,a,b})_{a,b}$ by

$$G_{i,a,b}(\boldsymbol{T}) = \mathbf{1}_{\{a_i(\boldsymbol{T})=a\}}\big(u_i(b, \boldsymbol{a}_{-i}(\boldsymbol{T})) - u_i(a, \boldsymbol{a}_{-i}(\boldsymbol{T}))\big).$$

Only one source row is nonzero, so $\|\boldsymbol{G}_i\|_{\infty,1} \le 1$. The tree theorem gives $\boldsymbol{r}_i^{(t)} = \mathbb{E}_{\mathbb{P}^{\mathrm{tr}}_{\boldsymbol{z}^{(t)}}} \boldsymbol{G}_i$. For fixed samples, $\boldsymbol{G}_i$ is score-independent. For edge-inclusion indicators I_e (zero on diagonal indices),

$$\nabla_{\boldsymbol{w}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}} = \sum_e (I_e - \mathbb{E}I_e)\frac{h_e}{q_e}\widetilde{w}_e. \tag{4.7}$$

Centering and independence across players give (Lemma D.1)

$$\big\|\nabla_{\boldsymbol{w}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}}\big\|_{L^2} \le 2\Big(\sum_j \|\boldsymbol{w}_j\|^2_{\infty,1}\Big)^{1/2}, \tag{4.8}$$

so learning rates combine through $(\sum_j \eta_j^2)^{1/2} \le g$, not $\sum_j \eta_j$. For $t \ge 2$, interpolate $\boldsymbol{y}_\theta = (1-\theta)\boldsymbol{z}^{(t-1)} + \theta \boldsymbol{z}^{(t)}$ and let $\mathbb{E}_\theta$ denote expectation at $\boldsymbol{y}_\theta$. Suppressing t and writing $\eta\boldsymbol{w} = (\eta_j \boldsymbol{w}_j)_j$, equation (3.4) gives

$$\mathrm{D}\boldsymbol{r}_i = -\int_0^1 \mathbb{E}_\theta[\boldsymbol{G}_i \nabla_{\eta \mathrm{D}\boldsymbol{\xi}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}_\theta}]\, d\theta + \int_0^1 \mathbb{E}_\theta[\boldsymbol{G}_i \nabla_{\eta \boldsymbol{r}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}_\theta}]\, d\theta. \tag{4.9}$$

The first integral has cumulative norm at most CgE by (4.8) and $\|\mathrm{D}\boldsymbol{\xi}_j\|_{\infty,1;T} \le 2E$. The first-round contribution arising from zero prehistory has norm at most one. We retain these bounds and expand only the second integral. We call it the gain integral.

Repeat the step on the retained terms. Fix $t \ge 2$ and θ. Draw independent $T, T' \sim \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}_\theta^{(t)}}$ and $S \sim \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{z}^{(t)}}$, and let $\mathbb{E}^+_{t,\theta}$ average all three samples. In the tree-score sum (4.7), replace $I_e - \mathbb{E}I_e$ by $I_e(T) - I_e(T')$ and the direction $\boldsymbol{w}$ by $(\eta_j \boldsymbol{G}_j(S))_j$. Dividing this sum by $c_0 := 2Bg\sqrt{\ell}$ defines the normalized centered *atom* $Z^{(0)}_\theta$, where B is from Lemma C.1. Averaging over T' recovers the centered indicator, and averaging over S recovers $\boldsymbol{r}_j^{(t)} = \mathbb{E}_S \boldsymbol{G}_j(S)$. Hence

$$\boldsymbol{U}^{(t)}_\theta := \mathbb{E}^+_{t,\theta}[\boldsymbol{G}_i(T) Z^{(0)}_\theta] = \frac{1}{c_0}\mathbb{E}_\theta\Big[\boldsymbol{G}_i(T) \nabla_{\eta \boldsymbol{r}^{(t)}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}^{(t)}_\theta}(T)\Big].$$

Thus the gain integral is exactly $c_0 \int_0^1 \boldsymbol{U}^{(t)}_\theta\, d\theta$. For fixed samples, the atom depends on the scores through the local factors $(h_e/q_e)(v_b - \langle \boldsymbol{\sigma}_{j,a}, \boldsymbol{v}\rangle)$, where $e = (j,a,b)$. An order-r atom $Z^{(r)}$ uses normalized order-r score derivatives of these factors, with payoff samples held fixed (Appendix E).

We next compute $\boldsymbol{U}^{(t)}_\theta - \boldsymbol{U}^{(t-1)}_\theta$, holding θ fixed. For $t \ge 3$, interpolate $\boldsymbol{y}^{(t-1)}_\theta$ to $\boldsymbol{y}^{(t)}_\theta$ and $\boldsymbol{z}^{(t-1)}$ to $\boldsymbol{z}^{(t)}$. Integrating the derivative along this path gives the difference. Both the sampling distributions and the atom vary, while $\boldsymbol{G}_i(T)$ is fixed for fixed T. Splitting score changes into error and gain directions as before, the product rule gives two kinds of gain terms: a derivative of any sampling distribution adds an atom, giving $\boldsymbol{G}_i Z^{(0)} Z^{(0)}_{\mathrm{new}}$, and a derivative of the existing atom gives $\boldsymbol{G}_i Z^{(1)}$.

More generally, a retained depth-p expression has the form $\boldsymbol{U}^{(t)} = \mathbf{1}_{\{t>p\}}\mathbb{E}_t[\boldsymbol{G}\prod_{j=1}^h Z_j^{(r_j)}]$, with $\sum_j (r_j+1) \le p$, where $\mathbb{E}_t$ averages all samples and $\boldsymbol{G}$ is a fixed observable with $\|\boldsymbol{G}\|_{\infty,1} \le 1$. Each gain operation increases the total degree by one, producing the depth-$(p+1)$ expressions $\boldsymbol{U}_\nu$ below. Moreover, boundary and error-direction terms form $\boldsymbol{V}$. For every admissible $\boldsymbol{U}$ with $p < \ell$, Lemma E.1 gives time-independent coefficients $c(\nu)$ such that

$$\mathrm{D}\boldsymbol{U} = \boldsymbol{V} + \int c(\nu)\boldsymbol{U}_\nu\, d\nu,$$

$$\|\boldsymbol{V}\|_{\infty,1;T} \le 1 + C_E(p+1)gE, \qquad \int |c(\nu)|\, d\nu \le C_E g\sqrt{\ell}\,(p+1).$$

The normalization gives $\|Z^{(r)}\|_{L^{2\ell}} \le 1$, so Hölder bounds products of at most ℓ atoms in L^2.

Account for the remaining filters. Lemma G.1 gives $\mathrm{K}_0 = \mathrm{H}$, $\mathrm{K}_p = \mathrm{K}_{p+1}\mathrm{D}$, and $\|\mathrm{K}_{p+1}\|_{\mathrm{ker},1} \le Ck^p$. The kernel mass is the sum of absolute convolution coefficients and bounds the operator norm on $\|\cdot\|_{\infty,1;T}$. Since the coefficients $c(\nu)$ are time-independent,

$$\mathrm{K}_p[\boldsymbol{U}] = \mathrm{K}_{p+1}[\boldsymbol{V}] + \int c(\nu)\mathrm{K}_{p+1}[\boldsymbol{U}_\nu]\, d\nu.$$

We bound the first term using the filter norm (*stop*) and repeat the rule only on the second (*continue*). Let a_p be the total absolute coefficient mass and $M_p = k^p a_p$. Then

$$M_{p+1} \le C_E g k\sqrt{\ell}\,(p+1)M_p \le \kappa M_p, \qquad M_0 = 1, \qquad M_p \le \kappa^p. \tag{4.10}$$

The stopped costs sum to $C\sum_{p<\ell}\kappa^p[1 + C_E(p+1)gE] \le C'(1+gE)$. At depth ℓ, $\mathrm{K}_\ell = \mathrm{R}_\ell^\ell \mathrm{L}_0$, where $\mathrm{L}_0 = \mathrm{D}^3\mathrm{F}_N^N$. Its resolvent mass k^ℓ is already included in $M_\ell \le \kappa^\ell$, proving the stage output.

Proposition 4.3 (Early-stage attenuation)**.** *For $\kappa \le 1/16$, the early expansion generates admissible depth-ℓ expressions $\boldsymbol{U}_\nu$ such that*

$$\|\mathrm{H}[\boldsymbol{r}_i]\|_{\infty,1;T} \le C(1+gE) + \kappa^\ell \sup_\nu \|\mathrm{L}_0[\boldsymbol{U}_\nu]\|_{\infty,1;T}\,. \tag{4.11}$$

4.4 Repeated transitions and finite curvature closure

Fix an early depth-ℓ expression $\boldsymbol{U}$ from Proposition 4.3. We now bound $\|\mathrm{L}_0[\boldsymbol{U}]\|_{\infty,1;T}$ by continuing its expansion, starting at late depth $j = 0$.

Separate the curvature-controlled score from its residual. For $e = (i,a,b)$, Lemma B.1 gives

$$h_e/q_e = \gamma_e + \delta\alpha_e, \qquad 0 < \alpha_e \le 1, \qquad \gamma_e^2 \le C_\gamma h_e/\beta_i, \qquad \delta\alpha_e^2 \le C_\alpha\sqrt{h_e/\beta_i}. \tag{4.12}$$

The bound on γ_e^2 controls the γ-contribution in (4.7) by curvature (Lemma D.1). In contrast, the bound on $\delta\alpha_e^2$ requires two factors of α_e, but a distribution derivative initially supplies only one. We therefore record each fresh transition e and retain $\chi^\alpha_{e,\boldsymbol{v}}(\boldsymbol{y}) = \alpha_e(\boldsymbol{y})(v_b - \langle \boldsymbol{\sigma}_{i,a}(\boldsymbol{y}), \boldsymbol{v}\rangle)$ for a later repetition. At depth j, the j recorded indices are distinct wherever $\boldsymbol{G} \ne 0$.

A repeated transition supplies the missing factor. In direction $\eta\boldsymbol{r}^{(s)}$, the residual term in (4.7) contains $\delta\eta_i\alpha_e(\boldsymbol{y})\widetilde{r}_e^{(s)}(\boldsymbol{y})$. When e repeats, this pairs with the retained $\alpha_e(\boldsymbol{y}')$. For $\boldsymbol{y}' = \boldsymbol{y}$, the scalar bound gives

$$\delta\eta_i\alpha_e^2|\widetilde{r}_e^{(s)}| \le C\beta_i^{-1/2}\eta_i\sqrt{h_e}\,|\widetilde{r}_e^{(s)}| \le C\beta_i^{-1/2}\eta_i\sqrt{\boldsymbol{K}_{\boldsymbol{y}_{i,a}}(\boldsymbol{r}_{i,a}^{(s)})},$$

where scalar quantities are evaluated at $\boldsymbol{y}$ and the last inequality follows from (4.2). For the possibly different interpolated scores in the expansion, Lemma F.1 gives (F.2), bounding the pair uniformly in e by $C2^j(\sum_{i',a'}\omega_{i'}b_{i',a'}^{(s)})^{1/2}$ before introducing a new gain sample. Hölder and $|\chi^\alpha_{e,\boldsymbol{v}}| \le \alpha_e \le 1$ bound the remaining product in L^1 (Lemma D.2). Since $Y^2 = \sum_{s\le T,i,a}\omega_i b_{i,a}^{(s)}$, each repeated term has cumulative norm at most $C2^jY$ and requires no further expansion.

Score derivatives of existing Z and χ^α factors in direction $\eta\boldsymbol{r}^{(s)}$ can also be bounded by curvature (Lemma C.1). Boundary and error terms are bounded as before.

Only fresh transitions continue. After splitting by player and sign, an independent gain sample selects one source row, and $\sum_b I_{i,a,b} \le 1$ selects at most one outgoing edge. This avoids an extra m_i factor. Each continuation appends one fresh index, so none remains at depth $N = |\mathcal{I}|$.

The total cost is a geometric sum. Lemma F.2 gives the same exact decomposition $\mathrm{D}\boldsymbol{U} = \boldsymbol{V} + \int c(\nu)\boldsymbol{U}_\nu\, d\nu$, with depth-$(j+1)$ children, time-independent coefficients, and no children at $j = N$. The stopped bound is $C_L k^2 8^j(1 + gE + Y)$, and the continuation mass is at most a_{late}

below. The remaining filters satisfy $\mathrm{L}_j = \mathrm{L}_{j+1}\mathrm{D}$ and $\|\mathrm{L}_{j+1}\|_{\mathrm{ker},1} < 24(N+1)^j$ (Lemma G.1). With $\eta_{\max} = \max_i \eta_i$, let

$$a_{\mathrm{late}} = 2n\delta\eta_{\max}(1 + 2\ell + 2N), \qquad \rho = 8(N+1)a_{\mathrm{late}} \le 32\eta_{\max} < 1/25.$$

The public δ gives the last two inequalities (Appendix G). Multiplying coefficient mass, stopped cost, and filter norm at depth j gives

$$\|\mathrm{L}_0[\boldsymbol{U}]\|_{\infty,1;T} \le 24C_L k^2(1 + gE + Y)\sum_{j=0}^{N}\rho^j \le 25C_L k^2(1 + gE + Y).$$

The sum includes the terminal boundary impulse. Backward-only filters keep every term within the same horizon T and energies E, Y. This proves the late-stage output.

Proposition 4.4 (Late-stage curvature bound)**.** *Every admissible depth-ℓ early expression satisfies*

$$\|\mathrm{L}_0[\boldsymbol{U}]\|_{\infty,1;T} \le 25C_L k^2(1 + gE + Y). \tag{4.13}$$

Combining the stages. Substituting (4.13) into (4.11) and including the solver error gives

$$\begin{aligned} E &\le 1 + \max_i \|\mathrm{H}[\boldsymbol{r}_i]\|_{\infty,1;T} \\ &\le 1 + C(1 + gE) + 25C_L k^2\kappa^\ell(1 + gE + Y) \\ &\le C_0 + C_1 gE + C_P k^2\kappa^\ell Y. \end{aligned}$$

Since $k^2\kappa^\ell = O(1)$ for $k = \ell + 1$ and $\kappa \le 1/16$, this proves Lemma 4.2. Together with Lemma 4.1 and the absorption in Section 4.1, it completes Theorem 2.1.

5 Adversarial robustness via a common-prefix wrapper

We prove Theorem 2.2 using only the base dynamics' anytime guarantee. An immediate switch from base dynamics to an adversarial fallback retains the self-play budget additively. Instead, all players run a fallback for a common prefix of length W, then initialize the base dynamics afresh. Player i permanently switches to a new fallback copy after the first base round s with

$$R_{i,\mathrm{base}}(s) = \sum_a \max_b \sum_{u=1}^{s} r_{i,a,b}^{(W+u)} > B_i. \tag{5.1}$$

The anytime BM fallback has regret at most $4\sqrt{A_i s}$ over any s rounds. Algorithm 2 and its implementation appear in Appendix I.

Lemma 5.1 (Common-prefix wrapper)**.** *Suppose simultaneous base self-play satisfies $R_{i,\mathrm{base}}(s) \le B_i$ at every prefix. No threshold-triggered switch then occurs in self-play. For arbitrary payoff vectors and $T > W$,*

$$\mathrm{SwapReg}_i(T) \le 4\sqrt{A_i W} + (B_i + 1) + 4\sqrt{A_i(T - W)}. \tag{5.2}$$

If $B_i + 1 \le \gamma_i\sqrt{A_i W}$, then, for every T,

$$\mathrm{SwapReg}_i(T) \le \sqrt{(4 + \gamma_i)^2 + 16}\,\sqrt{A_i T}. \tag{5.3}$$

A first threshold crossing in self-play contradicts the guarantee of the simultaneously initialized base run. Against arbitrary feedback, the crossing round overshoots by at most one. Subadditivity over the three phases and Cauchy–Schwarz give the bounds. Appendix I.2 gives the proof and a nonconstant-threshold extension.

To apply this black-box lemma to our base dynamics, take

$$B_i = (4/c)k^{5/2}\sqrt{A_i A}, \qquad W = \left\lceil 9k^5 A/c^2 \right\rceil. \tag{5.4}$$

Equation (4.6) gives base regret at most $3B_i/4$. Appendix I.3 verifies

$$B_i + 1 \le (5/3)\sqrt{A_i W}, \qquad 4\sqrt{A_i W} + 3B_i/4 < 4B_i. \tag{5.5}$$

Thus Lemma 5.1 gives $\mathrm{SwapReg}_i(T) < 7\sqrt{A_i T}$ adversarially and $\mathrm{SwapReg}_i(T) < 4B_i$ in self-play, proving Theorem 2.2.

AI USE STATEMENT

In this work, generative AI tools were used to identify and summarize relevant literature, assist with completing proof arguments from author-provided high-level directions and quantitative targets, and help draft and edit the manuscript. The authors substantially steered, checked, and revised the AI-generated material and take responsibility for the final claims, proofs, citations, and presentation.

REPRODUCIBILITY STATEMENT

Section 2 states the feedback model and assumptions. Algorithm 1 specifies the base updates. Appendix A.2 and Appendix A.3 give the public parameters and the order of universal-constant selection. Appendices B–G provide the detailed theoretical arguments, Appendix H specifies the normalization solver and exact-real computation model, and Appendix I specifies and analyzes the robust wrapper. Appendix A maps the main claims to these supporting arguments.

REFERENCES

Omar Abbadi, Rida Laraki, and Panayotis Mertikopoulos. Constant regret in general games via higher-order optimism, 2026. URL https://arxiv.org/abs/2609.04113v1.

Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multiplayer general-sum games. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 736–749, 2022a. doi: 10.1145/3519935.3520031. URL https://arxiv.org/abs/2111.06008v3.

Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $O(\log T)$ swap regret in multiplayer games. In *Advances in Neural Information Processing Systems*, volume 35, pp. 3292–3304, 2022b. URL https://arxiv.org/abs/2204.11417v2.

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On last-iterate convergence beyond zero-sum games. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 536–581, 2022c. URL https://proceedings.mlr.press/v162/anagnostides22a.html.

Ioannis Anagnostides, Gabriele Farina, Maxwell Fishelson, Haipeng Luo, and Jon Schneider. Swap regret minimization through response-based approachability. In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 195–223, 2026. URL https://proceedings.mlr.press/v336/anagnostides26a.html.

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

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

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(47):1307–1324, 2007. URL https://www.jmlr.org/papers/v8/blum07a.html.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Advances in Neural Information Processing Systems*, volume 33, pp. 18990–18999, 2020. URL https://arxiv.org/abs/2006.04953v2.

Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. From external to swap regret 2.0: An efficient reduction for large action spaces. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pp. 1216–1222, 2024. doi: 10.1145/3618260.3649681. URL https://doi.org/10.1145/3618260.3649681.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. In *Advances in Neural Information Processing Systems*, volume 34, pp. 27604–27616, 2021. URL https://arxiv.org/abs/2108.06924v2.

Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning dynamics for general convex games. In *Advances in Neural Information Processing Systems*, volume 35, pp. 39076–39089, 2022. URL https://arxiv.org/abs/2206.08742.

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

Geoffrey J. Gordon, Amy Greenwald, and Casey Marks. No-regret learning in convex games. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 360–367, 2008. doi: 10.1145/1390156.1390202. URL https://www.cs.cmu.edu/~ggordon/gordon-greenwald-marks-icml-phi-regret.pdf.

Amy Greenwald and Amir Jafari. A general class of no-regret learning algorithms and game-theoretic equilibria. In *Learning Theory and Kernel Machines*, volume 2777 of *Lecture Notes in Computer Science*, pp. 2–12, 2003. doi: 10.1007/978-3-540-45167-9_2. URL https://doi.org/10.1007/978-3-540-45167-9_2.

Soumita Hait, Ping Li, Haipeng Luo, and Mengxiao Zhang. Comparator-adaptive Φ-regret: Improved bounds, simpler algorithms, and applications to games. In *Advances in Neural Information Processing Systems*, volume 38, pp. 102828–102867, 2025. doi: 10.52202/085713-3100. URL https://arxiv.org/abs/2505.17277v2.

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

Mingyang Liu, Gabriele Farina, and Asuman Ozdaglar. Constant individual regret in general games, 2026. URL https://arxiv.org/abs/2608.31166v1.

Zhou Lu, Y Jennifer Sun, and Zhiyu Zhang. Sparsity-based interpolation of external, internal and swap regret. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 3795–3828, 2025. URL https://proceedings.mlr.press/v291/lu25a.html.

Binghui Peng and Aviad Rubinstein. Fast swap regret minimization and applications to approximate correlated equilibria. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pp. 1223–1234, 2024. doi: 10.1145/3618260.3649691. URL https://doi.org/10.1145/3618260.3649691.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 993–1019, 2013. URL https://proceedings.mlr.press/v30/Rakhlin13.html.

Ashkan Soleymani, Georgios Piliouras, and Gabriele Farina. Cautious optimism: A meta-algorithm for near-constant regret in general games, 2025a. URL https://arxiv.org/abs/2506.05005v2. Appeared at EC 2025.

Ashkan Soleymani, Georgios Piliouras, and Gabriele Farina. Faster rates for no-regret learning in general games via Cautious Optimism, 2025b. URL https://arxiv.org/abs/2503.24340v1. Appeared at STOC 2025.

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

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, volume 28, pp. 2989–2997, 2015. URL https://arxiv.org/abs/1507.00407.

Taira Tsuchiya. Sublogarithmic swap regret in multiplayer general-sum games via hybrid regularization, 2026. URL https://arxiv.org/abs/2608.04149v1.

Taira Tsuchiya, Haipeng Luo, and Shinji Ito. Scale-invariant fast convergence in games, 2026. URL https://arxiv.org/abs/2602.11857v1.

A SUPPLEMENTARY BACKGROUND, PROOF ROADMAP, AND PARAMETERS

The main text states the two proof bounds and explains how they imply the guarantees. The table below locates the complete verification of each component. The subsections that follow record the standard CE implication, public parameters, and comparisons used by those arguments.

Component	Complete verification	Output used next
Public parameters and absorption	Appendices A.2–A.3	Conditions used in Section 4.1
Row normalization and curvature comparison	Appendix B	Bounds used in Sections 4.2 and 4.4
Uniform normalized derivatives	Appendix C	Derivative bounds used in both closures
Stationary-tree representation and moments	Appendix D	Inputs to the early and late closures
Early player-centered closure	Appendix E	Lemma E.1 and the κ^{ℓ} handoff
Late transition-index closure	Appendix F	Lemma F.2
Two-scale filters and assembly	Appendix G	Lemma 4.2
Normalization solver and implementation	Appendix H	Effective-score interface and complexity
Fallback and common-prefix wrapper	Appendix I	Lemma 5.1 and Theorem 2.2

Norm convention. In addition to $\|\boldsymbol{w}_i\|_{\infty,1} = \sum_a \max_b |w_{i,a,b}|$, the mixed row norm is $\|\boldsymbol{w}_i\|_{\infty,2}^2 = \sum_a \max_b |w_{i,a,b}|^2$. Then $\|\boldsymbol{w}_i\|_{\infty,2} \le \|\boldsymbol{w}_i\|_{\infty,1}$ and $\|\boldsymbol{w}_i\|_{\infty,1;T}^2 = \sum_{t=1}^{T} \left\|\boldsymbol{w}_i^{(t)}\right\|_{\infty,1}^2$.

A.1 CORRELATED-EQUILIBRIUM CONSEQUENCE

The following standard implication records the equilibrium consequence mentioned in the introduction. Let

$$\pi_T = \frac{1}{T}\sum_{t=1}^{T} \bigotimes_i \boldsymbol{x}_i^{(t)}$$

be the time-averaged product distribution in a fixed game.

Corollary A.1 (Correlated equilibrium)**.** *In self-play, π_T is an $(\max_i B_i)/T$-approximate CE for the base dynamics, and a $4(\max_i B_i)/T$-approximate CE for the robust dynamics, under the swap-deviation convention.*

Proof. For every player i and deviation map $\varphi : [m_i] \to [m_i]$, multilinearity gives

$$\mathbb{E}_{\boldsymbol{S}\sim\pi_T}[u_i(\varphi(S_i), \boldsymbol{S}_{-i}) - u_i(\boldsymbol{S})] = \frac{1}{T}\sum_{t=1}^{T}\sum_a x_{i,a}^{(t)}\left(v_{i,\varphi(a)}^{(t)} - v_{i,a}^{(t)}\right) \le \frac{\mathrm{SwapReg}_i(T)}{T}.$$

Apply Theorems 2.1 and 2.2 and maximize over players and deviation maps. □

This is a time-averaged product distribution, not a last iterate or the product of the players' average strategies (Blum & Mansour, 2007, Theorem 3).

For the single-action convention

$$g_i(a, b, \pi) = \mathbb{E}_{\boldsymbol{S}\sim\pi}\left[\mathbf{1}_{\{S_i=a\}}\big(u_i(b, \boldsymbol{S}_{-i}) - u_i(\boldsymbol{S})\big)\right],$$

one has $0 \le \max_{a,b} g_i(a, b, \pi) \le \sum_a \max_b g_i(a, b, \pi) \le m_i \max_{a,b} g_i(a, b, \pi)$. Thus positive tolerances may differ by a factor m_i, although the zero-tolerance conditions agree.

A.2 PUBLIC PARAMETERS

In addition to (2.2), set

$$\delta = \frac{1}{n(N+\ell+2)^2}, \qquad J = \lceil \log(1/\delta) \rceil, \qquad \beta_i = \frac{\log m_i}{2^{10} m_i J}.$$

Let d_i be the least power of two at least $\lceil \log(1/\beta_i) \rceil$. For the default allocation, use (2.3). With the proof-only constant C_ω fixed in Appendix A.3, define

$$\omega_i = C_\omega \frac{m_i - 1}{\beta_i}, \qquad V = \sum_i \omega_i A_i.$$

Then $\eta_{\max} \le g$, $m_i \beta_i \le 2^{-10}$, $0 < \delta < 1/2$, and $\log m_i \le J$. All parameters depend only on public action counts and universal constants.

A.3 Noncircular selection of universal constants

First fix integer upper bounds at least one for all universal constants in Appendices B–G, and set $C_\omega = \max\{1, C_\gamma\}$. Fix

$$c = \min\left\{ \frac{1}{16 C_E}, \frac{1}{4(C_1+1)}, \frac{1}{2(C_0+1)} \right\}.$$

Then $\kappa = C_E c \le 1/16$. To choose $\ell_0 \ge 1$, write

$$v = \lceil \log(2 + n + N) \rceil, \qquad k = \ell_0 + v + 1.$$

Every nontrivial instance has $N \ge 4$, so $v \ge 3$ and $k \ge 5$. Moreover,

$$N + \ell + 2 \le 2^v (k+2), \qquad J \le 1 + 3v + 2\log(k+2) \le 5k,$$
$$V = 2^{10} C_\omega J \sum_i m_i^2 (m_i - 1) \le 2^{13} C_\omega k N^{3/2}.$$

Choose ℓ_0 large enough that

$$C_P \sqrt{2^{14} C_\omega}\, (\ell_0 + 1)^{5/2} 16^{-\ell_0} \le 1.$$

Since $k \le (\ell_0 + 1)(v+1)$ and $N \le 2^v$,

$$\tau_\ell \sqrt{2V} \le C_P \sqrt{2^{14} C_\omega}\, (\ell_0+1)^{5/2} 16^{-\ell_0} (v+1)^{5/2} 2^{-13v/4} \le 1.$$

The final dimension factor equals one at $v = 0$ and has consecutive ratio at most $2^{5/2 - 13/4} < 1$, proving (4.5) without changing any previously fixed lemma constant. Section 4.1 then gives $E \le E_0 = 2(C_0 + 1)$ and (4.6). Because $cE_0 \le 1$, the default rates satisfy $\eta_i E_0 \le 1 \le \sqrt{A_i}$. This proves Theorem 2.1 with the slack used by the robust wrapper.

A.4 Optional unequal-action rate allocation

Remark A.2 (Unequal action counts). For the base dynamics, one may instead set $U = \sum_j A_j^2$ and $\eta_i = g A_i / \sqrt{U}$. The same proof gives $\max_i \mathrm{SwapReg}_i(T) \le \frac{4}{c} k^{5/2} \sqrt{U}$ for every prefix. This equals the default worst-player bound for equal action counts and can improve it by a factor $\Theta(n^{1/4})$ for highly unequal action counts. The details follow below.

Set

$$U = \sum_j A_j^2, \qquad \eta_i = \frac{g A_i}{\sqrt{U}}. \tag{A.1}$$

These rates satisfy (4.1) and $\eta_i \le g$, so (4.6) gives

$$\max_i \mathrm{SwapReg}_i(T) \le \frac{4}{c} k^{5/2} \sqrt{U}.$$

The allocation minimizes $\max_i A_i/\eta_i$ subject to (4.1): if this maximum is M, then $\eta_i \ge A_i/M$ and hence $g^2 \ge U/M^2$. Relative to the default worst-player bound, the improvement factor is

$$\sqrt{\frac{A_{\max} A}{U}} \le \sqrt{\frac{\sqrt{n}+1}{2}}, \qquad A_{\max} = \max_i A_i.$$

The inequality follows from Cauchy–Schwarz and is tight in order: $n-1$ binary players and one player with $m_i \log m_i = \Theta(\sqrt{n})$ give a factor $\Theta(n^{1/4})$. Equal action counts recover the default rates. For the common-prefix wrapper, the relevant objective is $\max_i A_i/\eta_i^2$. Its minimum under (4.1) is A/g^2, attained by the default square-root rates. Thus the optional allocation improves the base worst-player bound but not the present wrapper's worst-player prefix order.

A.5 TECHNICAL COMPARISON WITH RELATED WORK

This subsection collects the broader related work supporting the introduction and explains the assumptions, mechanisms, and tradeoffs behind the selected bounds in Table 1. It also discusses related algorithms not included in that table.

Regret and equilibrium learning. Correlated equilibrium originates with Aumann (1974). The learning-to-CE connection was developed through calibrated learning and regret-based adaptive procedures by Foster & Vohra (1997), Hart & Mas-Colell (2000). The classical reduction of Blum & Mansour (2007) converts external-regret guarantees into swap regret by maintaining one learner per source action and playing the stationary distribution of its transition matrix. More generally, Greenwald & Jafari (2003) study regret with respect to a prescribed deviation family, and Gordon et al. (2008) connect this Φ-regret viewpoint to external-regret learning over transformations and fixed-point computation.

External and swap regret in self-play. Optimistic learning exploits predictable payoff sequences (Rakhlin & Sridharan, 2013). The analysis of Syrgkanis et al. (2015) balances payoff variation against strategy movement and provides a central framework for predictable feedback in multiplayer games. For external regret, this line yields polylogarithmic bounds in finite games (Daskalakis et al., 2021) and logarithmic bounds in general convex games (Farina et al., 2022). Soleymani et al. (2025b) introduce Cautious Optimism through adaptive, non-monotone learning-rate control for OMWU, achieving logarithmic individual external regret in self-play. In a broader treatment, Soleymani et al. (2025a) develop a meta-algorithm that paces an underlying optimistic FTRL learner and establishes the same logarithmic dependence on T under suitable regularizer assumptions. More recent results obtain constant external regret through higher-order optimism (Liu et al., 2026, Abbadi et al., 2026).

For swap regret in the present full-information setting, Chen & Peng (2020) obtain $T^{1/4}$ horizon dependence with BM–OMWU and, in Appendix D, with an optimistic fixed-point construction over all swap maps. Anagnostides et al. (2022a) give polylogarithmic bounds through BM–OMWU and SL–OMWU, the latter using the internal-regret reduction of Stoltz & Lugosi (2005). Anagnostides et al. (2022b) obtain a logarithmic bound with BM–OFTRL. Tsuchiya (2026) further improves the BM–OFTRL horizon dependence to a sublogarithmic bound. Table 1 reports selected guarantees from this line of work. Our result is a constant swap regret bound independent of T. At the same time, we achieve the best polynomial dependence on both n and m over the previous $o_{n,m}(\log T)$ results.

Comparator-adaptive guarantees and alternative reductions. Lu et al. (2025) interpolate external, internal, and swap regret through comparator sparsity. Building on this line, Hait et al. (2025) improve comparator-adaptive Φ-regret guarantees and obtain constant regret in a class of general-sum games satisfying *nonnegative social external regret*, namely $\sum_i \mathrm{ExtReg}_i(T) \geq 0$ for every play sequence and horizon under the unclipped external-regret convention. This class was studied by Anagnostides et al. (2022c). Our constant swap-regret guarantee holds in arbitrary finite general-sum games. Beyond the classical per-action BM reduction, Dagan et al. (2024), Peng & Rubinstein (2024) achieve any fixed target level of average swap regret with round complexity polylogarithmic in the number of actions. These action–accuracy tradeoffs are distinct from a constant cumulative swap-regret guarantee in self-play. In convex domains, Anagnostides et al. (2026) use response-based approachability to minimize linear and profile swap regret and extend their results to polynomial-dimensional deviation families. These broader comparator classes lie outside the finite normal-form swap-regret setting studied here. Finally, Tsuchiya et al. (2026) obtain scale-free and scale-invariant swap regret $O(U_{\max} n^{3/2} m^{5/2} \log T)$ in general-sum games.

Adversarial robustness. Adversarial robustness in this line is obtained through closely related certificate-triggered constructions. Earlier work describes these as a wrapper or an adaptive choice of learning rate (Chen & Peng, 2020, Daskalakis et al., 2021, Anagnostides et al., 2022a), while the closest recent constructions use a switching rule that switches a player permanently to adversarial dynamics after a self-play certificate fails (Anagnostides et al., 2022b, Tsuchiya, 2026, Abbadi et al., 2026). Such an immediate switch retains the self-play budget additively. We add a common phase before a fresh simultaneous start of the base dynamics, allowing that budget to be absorbed into the all-horizon $O(\sqrt{mT \log m})$ adversarial guarantee. Our approach is black-box in the underlying

self-play dynamics. Therefore, it can strengthen existing results with anytime self-play regret bounds by removing their additive self-play budget whenever a common-prefix absorption condition holds.

Detailed technical comparison. The following discussion focuses on the additional bounds needed to combine BM, tree representations, optimistic potential bounds, higher-order prediction, and finite-label recursions to obtain constant swap regret with the stated dependence on n and m. Table 1 compares selected bounds under full-information feedback, not every algorithm in the cited papers or methods with equal computational budgets.

Prediction, movement, and adaptive pacing. The potential argument follows the work on predictable sequences and regret bounded by variation in utilities (RVU) (Rakhlin & Sridharan, 2013, Syrgkanis et al., 2015). It balances prediction error against a negative movement or curvature term supplied by the regularizer. Farina et al. (2022) obtain logarithmic external regret in convex games by lifting the action space and using self-concordant regularization. Soleymani et al. (2025b) instead control the learning rate of OMWU through Cautious Optimism. Their broader treatment gives a meta-algorithm that adjusts the pace of an FTRL learner (Soleymani et al., 2025a). These external-regret methods do not directly establish a prediction bound for BM row gains, which depend on the player's entire stationary distribution.

Higher-order smoothness through the Blum–Mansour fixed point. The closest earlier analysis of the same fixed-point difficulty is the BM–OMWU result of Anagnostides et al. (2022a). Their Theorem 2.4 states the Markov-chain tree formula, and Theorem 3.1 represents SL–OMWU using directed trees. For BM–OMWU, Section 4 and Lemmas D.5–D.6 combine this formula with multivariate Cauchy bounds in logarithmic transition coordinates. Building on Daskalakis et al. (2021), they bound higher-order finite differences through the stationary-distribution operation and obtain polylogarithmic swap regret. Thus higher-order smoothness through the fixed point is an existing ingredient, distinct from an explicit higher-order predictor, since the cited dynamics use OMWU.

Appendix C uses related ideas from complex analysis, but applies them to local centered factors rather than proving one global smoothness bound for the full stationary map. Normalizing the weights of rooted trees and differentiating their distribution gives the centered edge score in (4.7). Neither this normalization nor the centered-score identity is a separate novelty. Lemma C.1 additionally gives every normalized derivative both an ordinary bound and a bound in terms of the same real row curvature used by the potential argument. The first bound allows the early stage to raise derivative orders without using the energy weights that depend on the game dimensions. The second allows the late stage to bound those derivatives directly by curvature.

Movement-based control of the stationary strategy. Anagnostides et al. (2022b) combine Blum–Mansour with optimistic FTRL and a log-barrier. The barrier supplies a negative movement term. Their Lemma 4.2 uses the tree theorem to transfer multiplicative row stability to a bound on movement of the stationary strategy. Their Theorem 4.4 uses nonnegative swap regret to bound second-order path length. This gives $O(\log T)$ swap regret. Tsuchiya (2026) uses a hybrid entropy/log-barrier regularizer: entropy controls optimistic prediction error, the log-barrier controls row movement, and a global sensitivity theorem for rooted trees transfers this control to the stationary distribution without requiring local closeness.

Our proof uses a different interface rather than a sharper sensitivity inequality. The row potential has Hessian $\boldsymbol{K}_z$, and stationarity cancels the linear term after summing over source rows, giving the nonnegative curvature energy in (4.3). We spend this energy inside derivatives of the tree distribution instead of first bounding $\left\|x_i^{(t)} - x_i^{(t-1)}\right\|$. This distinction is essential for the action dependence: applying a bound on stationary sensitivity or tree variance at every high-order step would introduce the weights $\omega_i = \Theta((m_i - 1)/\beta_i)$ from the beginning. Our analysis delays this weighted curvature bound until the total absolute weight of the continuing terms, including the early filter cost, is exponentially small.

Constant external regret and finite label recursions. The horizon-independent prediction mechanism is closest to the concurrent external-regret results of Liu et al. (2026), Abbadi et al. (2026).

ECHO-OFTRL of Liu et al. (2026) preconditions a high-order difference by an EMA-resolvent cascade and recursively records exposed player labels. HOOD of Abbadi et al. (2026) uses a discounted finite recurrence and an explicit expansion in which a repeated player is charged to movement while only a fresh player continues. Our filters are closest algorithmically to ECHO, and our exact stop/continue bookkeeping is closest to HOOD.

The BM reduction prevents a direct application of these self-play theorems. The original payoff vector is multiaffine in the opponents' strategies. Centered external gains may also depend on the player's own strategy, as in Liu et al. (2026), while remaining multiaffine in the original player strategies. In contrast, the row gains

$$r_{i,a,b} = x_{i,a}(v_{i,b} - v_{i,a})$$

combine this payoff structure with the stationary-distribution map of the entire transition matrix. All rows of the same player are therefore coupled. BM transfers regret guarantees for the actual row feedback, but does not ensure that it satisfies the prediction bounds of an external-regret self-play theorem.

For a concrete illustration, let

$$Q = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}, \qquad 0 < p, q < 1.$$

Its stationary mass is $x_1 = q/(p+q)$, so for a fixed payoff vector the row gain is $r_{1,2} = q(v_2 - v_1)/(p+q)$. This is already rational, rather than affine, in either row parameter. For fixed q and $v_2 \neq v_1$, its derivatives with respect to p need not vanish at any finite order. Thus treating BM rows as additional players does not preserve the original finite-game payoff structure. This identifies a missing hypothesis for a direct transfer, not an impossibility result for other reductions.

Our two closures separate these effects. During the first $\ell = O(\log(nm))$ differences, the differentiated product-tree distribution yields centered sums independent across players. Moment bounds combine the rates through $(\sum_i \eta_i^2)^{1/2}$ and keep the action-weighted energy out of the recursion. After the resolvent mass is included, the total absolute weight of the continuing terms is at most κ^ℓ. Only then does the proof refine its bookkeeping from players to transition indices $e = (i, a, b)$. The split

$$\frac{h(q)}{q} = \gamma(q) + \delta\alpha(q)$$

separates a component paid collectively by curvature from a residual carrying an individual transition index. If the transition index repeats, the old α factor supplies the second factor needed for (4.12), so the branch is charged to the energy and stops. Only a fresh index continues. Since there are $N = \sum_i m_i^2$ indices, the late expansion terminates exactly. The root selects a single source row, and a tree has at most one outgoing edge from that row. These properties avoid additional factors from the action counts. The two commuting EMA cascades implement these two depths; their implementation order need not match the order of the proof stages.

Why the row normalization is not interchangeable. The construction of f_i must support the low initial potential in Lemma B.2, the two derivative bounds in Lemma C.1, and the residual split in Lemma B.1. For example, the exponential response $f(s) = e^s$ has $h(q) = q$ and $h(q)/q = 1$. With the present $0 < \delta < 1/2$ and $0 < \alpha \leq 1$, a split $1 = \gamma + \delta\alpha$ forces $\gamma > 1/2$. For fixed $\beta > 0$, this cannot satisfy $\gamma^2 \leq C_\gamma q/\beta$ uniformly as $q \downarrow 0$. Ordinary entropy normalization therefore cannot simply replace the constructed map in this proof. This observation concerns the specific curvature split, not the possibility of other entropy-based constant-regret algorithms.

Technical contribution and scope. The main technical contribution is the prediction bound in Lemma 4.2, which combines the derivative bounds, the bound for repeated transitions, and the delayed use of curvature to yield constant swap regret. Relative to the hybrid BM bound $O(nm^2\sqrt{\log m \log T})$ of Tsuchiya (2026) displayed in Table 1, our bound removes the horizon dependence and improves the simultaneous polynomial dependence on n and m. This is not a uniform comparison over every regime: Chen & Peng (2020) has a smaller exponent of m, while Hait et al. (2025) obtain horizon-independent, comparator-adaptive Φ-regret in games with nonnegative social external regret. Our arbitrary-game guarantee does not include their comparator-complexity refinement. The constant-regret results of Liu et al. (2026), Abbadi et al. (2026) have polylogarithmic action dependence but concern the weaker external-regret/CCE benchmark.

B ROW NORMALIZATION, POTENTIAL, AND CURVATURE COMPARISON

This appendix proves the scalar residual bounds and the potential/curvature facts used in Sections 4.2 and 4.4. Fix a player i and suppress its index on the scalar functions, while retaining the local dimension m_i. Write $\beta = \beta_i$ and $d = d_i$. The public parameters satisfy $m_i \geq 2$, $0 < \delta \leq 1/2$, $m_i\beta \leq 1/(4e)$, and $d \geq \log(1/\beta)$. In particular $d \geq 5$.

Lemma B.1 (Scalar bounds)**.** *For $q = f(s)$, define*

$$h(q) = f'(s), \qquad \lambda(s) = \frac{f'(s)}{f(s)}, \qquad \alpha(q) = \frac{1}{1+\delta u(s)/2}, \qquad \gamma(q) = \frac{h(q)}{q} - \delta\alpha(q).$$

There are universal constants C_γ, C_α such that, for $0 < q \leq 1$,

$$0 < h(q) \leq q, \qquad |h'(q)| \leq 2, \qquad 0 < \alpha(q) \leq 1,$$
$$\gamma(q)^2 \leq C_\gamma h(q)/\beta, \qquad \delta\alpha(q)^2 \leq C_\alpha\sqrt{h(q)/\beta}.$$

Moreover,

$$0 \leq \frac{d}{ds}\alpha(f(s)) \leq \frac{\delta}{2}\alpha(f(s))^2, \qquad \sum_b h(q_b) \geq \frac{3}{10} \tag{B.1}$$

for every normalized row. If $s \geq 0$ and $f(s) \leq 1$, then $h(f(s)) \geq (2/5)f(s)$.

Proof. Write $u = u(s) > 0$ and $B(u) = 1 + u + \delta u^2/4$. Since $s = u^{-1} - u$,

$$u' = -\frac{u^2}{1+u^2}, \qquad B(u) = (1+au)(1+bu), \qquad a = \frac{1+\sqrt{1-\delta}}{2}, \quad b = \frac{1-\sqrt{1-\delta}}{2}.$$

Direct logarithmic differentiation of (3.1) gives

$$\begin{aligned}\lambda(s) &= \frac{du}{(1+u^2)(1+du)} + \frac{au^2}{(1+u^2)(1+au)} + \frac{bu^2}{(1+u^2)(1+bu)} \\ &= \frac{1}{1+u^2}\frac{du}{1+du} + \frac{u^2}{1+u^2}\frac{1+\delta u/2}{B(u)}.\end{aligned} \tag{B.2}$$

The last line is a convex combination of numbers in $(0, 1]$. Hence $0 < \lambda \leq 1$, f is strictly increasing, and $0 < h(q) = q\lambda \leq q$.

Each of the three positive terms in the first line of (B.2) has logarithmic derivative of magnitude at most one. Indeed $(\ln u)' = -u/(1+u^2)$. For the first term, the remaining bracket is

$$1 - \frac{2u^2}{1+u^2} - \frac{du}{1+du},$$

for the terms with $c = a, b$, it is

$$\frac{2}{1+u^2} - \frac{cu}{1+cu}.$$

Both brackets have magnitude at most two, while $u/(1+u^2) \leq 1/2$. Thus $|\lambda'/\lambda| \leq 1$, and

$$|h'(f(s))| = \left|\frac{f''(s)}{f'(s)}\right| = \left|\lambda + \frac{\lambda'}{\lambda}\right| \leq 2.$$

Positive-score region. For $u \leq 1$, Bernoulli's inequality gives $(1+1/(du))^d \geq 1 + 1/u$. Retaining the first term of (B.2),

$$\frac{h(f(s))}{\beta} \geq \frac{d(1+u)}{(1+du)(1+u^2)B(u)} \geq \frac{2}{9}. \tag{B.3}$$

If also $f(s) \leq 1$, then

$$1 + \frac{1}{du} \leq \left(\frac{B(u)}{\beta}\right)^{1/d} \leq 2(9/4)^{1/d} \leq \frac{5}{2}.$$

Therefore $du/(1+du) \geq 2/5$, while $(1+\delta u/2)/B(u) \geq 1/2$, and the second line of (B.2) gives $\lambda \geq 2/5$. For $u \geq 1$, $f(s) \leq e\beta$. At most $em_i\beta \leq 1/4$ of a normalized row's probability mass lies in this region. On the remaining mass, $h(q) \geq (2/5)q$, so $\sum_b h(q_b) \geq (3/4)(2/5) = 3/10$.

Negative-score region and residual split. For $u \geq 1$, direct bounds in (B.2) give universal two-sided comparisons

$$\frac{f(s)}{\beta} \asymp \frac{1}{u(1+\delta u)}, \qquad \lambda(s) \asymp \frac{1}{u}, \qquad \frac{h(f(s))}{\beta} \asymp \frac{1}{u^2(1+\delta u)}. \tag{B.4}$$

Indeed, $1 \leq (1+1/(du))^d \leq e$, $u(1+\delta u)/4 \leq B(u) \leq 2u(1+\delta u)$, and $a \geq 1/2$. The cancellation

$$\frac{1+\delta u/2}{B(u)} - \frac{\delta}{1+\delta u/2} = \frac{1-\delta}{B(u)(1+\delta u/2)}$$

implies

$$|\gamma(f(s))| \leq C\big[u^{-2} + u^{-1}(1+\delta u)^{-2}\big].$$

Its square is at most $Ch(f(s))/\beta$ by (B.4). Also

$$\frac{\delta\alpha^2}{\sqrt{h/\beta}} \leq C\frac{\delta u}{(1+\delta u)^{3/2}} \leq C.$$

For $u \leq 1$, both weighted bounds follow from (B.3), $\lambda \leq 1$, and $\delta \leq 1/2$. Finally,

$$\frac{d}{ds}\alpha(f(s)) = \frac{\delta}{2}\alpha(f(s))^2\frac{u^2}{1+u^2},$$

which proves (B.1). □

B.1 Potential and local comparison

Define

$$\begin{aligned} P_i(s) &= \int_{-\infty}^{s} f_i(v)\,dv, \qquad c_{\Phi,i} = \min_s\{P_i(s) - s\}, \\ \Phi_i(\boldsymbol{z}) &= \mu_i(\boldsymbol{z}) + \sum_b P_i(z_b - \mu_i(\boldsymbol{z})) - c_{\Phi,i}. \end{aligned} \tag{B.5}$$

Suppressing the player index, the primitive is finite because $f(s)$ decays quadratically as $s \to -\infty$. The minimum in c_Φ occurs where $f(s) = 1$.

Lemma B.2 (Potential and comparison)**.** *The potential is smooth and convex, and*

$$\nabla\Phi(\boldsymbol{z}) = \boldsymbol{q}(\boldsymbol{z}), \qquad \Phi(\boldsymbol{z}) \geq \max_b z_b, \qquad 0 \leq \Phi(0) \leq 2\log m_i.$$

For the quantities in (4.2)*,*

$$\nabla_{\boldsymbol{w}} q_b = h_b\widetilde{w}_b, \qquad \boldsymbol{w}^\top\nabla^2\Phi(\boldsymbol{z})\boldsymbol{w} = \boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w}) = \min_{c\in\mathbb{R}}\sum_b h_b(w_b - c)^2 \leq \|\boldsymbol{w}\|_\infty^2. \tag{B.6}$$

If $\|\boldsymbol{y} - \boldsymbol{z}\|_\infty \leq \varepsilon$*, then*

$$e^{-4\varepsilon}h_b(\boldsymbol{z}) \leq h_b(\boldsymbol{y}) \leq e^{4\varepsilon}h_b(\boldsymbol{z}), \qquad e^{-\delta\varepsilon}\alpha_b(\boldsymbol{z}) \leq \alpha_b(\boldsymbol{y}) \leq e^{\delta\varepsilon}\alpha_b(\boldsymbol{z}), \tag{B.7}$$

$$e^{-4\varepsilon}\boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w}) \leq \boldsymbol{K}_{\boldsymbol{y}}(\boldsymbol{w}) \leq e^{4\varepsilon}\boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w}). \tag{B.8}$$

Proof. The lower bound $\sum_b h_b \geq 3/10$ and the implicit-function theorem give a smooth normalizer. Differentiating $\sum_b f(z_b - \mu) = 1$ gives

$$\nabla_{\boldsymbol{w}}\mu = \langle\boldsymbol{\sigma}, \boldsymbol{w}\rangle.$$

The terms involving $\nabla_{\boldsymbol{w}}\mu$ in $\nabla_{\boldsymbol{w}}\Phi$ cancel because $\sum_b q_b = 1$, proving the gradient and Hessian formulas. Completing the square gives the minimum representation. Choosing $c = 0$ and using $h_b \leq q_b$ proves the upper bound. For $s_b = z_b - \mu$, coordinate domination follows from

$$\Phi(\boldsymbol{z}) - z_b = P(s_b) - s_b - c_\Phi + \sum_{c\neq b} P(s_c) \geq 0.$$

Constant row shifts leave $\boldsymbol{q}$ unchanged and shift Φ by the same constant, so $\sqrt{\boldsymbol{K}}$ is a seminorm.

For comparison, Lemma B.1 and $|\widetilde{w}_b| \le 2\|\boldsymbol{w}\|_\infty$ imply

$$|\nabla_{\boldsymbol{w}} \ln h_b| = |h'(q_b)\widetilde{w}_b| \le 4\|\boldsymbol{w}\|_\infty, \qquad |\nabla_{\boldsymbol{w}} \ln \alpha_b| \le \delta\|\boldsymbol{w}\|_\infty.$$

Integrating along the segment proves (B.7). Compare the weighted sums for every fixed centering constant and then minimize to obtain (B.8).

It remains to bound $\Phi(0)$. Let $s_{m_i} = f^{-1}(1/m_i)$ and $s_1 = f^{-1}(1)$. Changing variables and integrating by parts gives

$$\Phi(0) = (m_i - 1)\int_0^{1/m_i} \frac{q}{h(q)}\,dq + \int_{1/m_i}^1 \frac{1-q}{h(q)}\,dq.$$

Set $q_0 = f(0) \le e\beta/2 < 1/m_i$. Since $|ds/du| = 1 + u^{-2}$ on $u \ge 1$,

$$P(0) \le 2e\beta\int_1^\infty \frac{du}{u(1+\delta u/4)} = 2e\beta\ln(1+4/\delta).$$

For $q \in [q_0, 1]$, the positive-region argument gives $h(q) \ge (2/5)q$. Splitting the first integral at q_0 and integrating the remaining elementary terms yields

$$\Phi(0) \le m_i P(0) + \frac{5}{2}\ln m_i.$$

For the public parameters, $m_i\beta = (\log m_i)/(2^{10}J)$ and $J \ge 2$. Since $1 + 4\cdot 2^J \le 5^J$, $\ln(1+4/\delta) \le J\ln 5$, and therefore

$$0 \le \Phi(0) \le \left(\frac{5}{2}\ln 2 + \frac{2e\ln 5}{2^{10}}\right)\log m_i < 2\log m_i.$$

□

B.2 Complete proof of the row-potential bound

This subsection proves Lemma 4.1 using the potential in (B.5). The construction in Appendix B gives

$$\begin{aligned} \nabla\Phi_i(\boldsymbol{z}) &= \boldsymbol{q}_i(\boldsymbol{z}), \qquad \Phi_i(\boldsymbol{z}) \ge \max_b z_b, \\ 0 \le \Phi_i(0) &\le 2\log m_i, \qquad \boldsymbol{w}^\top\nabla^2\Phi_i(\boldsymbol{z})\boldsymbol{w} = \boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w}), \end{aligned} \tag{B.9}$$

together with the local curvature comparison used below. Write

$$D_\Phi(\boldsymbol{y}, \boldsymbol{z}) = \Phi_i(\boldsymbol{y}) - \Phi_i(\boldsymbol{z}) - \langle \boldsymbol{q}_i(\boldsymbol{z}), \boldsymbol{y} - \boldsymbol{z}\rangle.$$

Fix a row and a round, abbreviate $\boldsymbol{z} = \boldsymbol{z}_{i,a}^{(t)}$, and set

$$\boldsymbol{d}^+ = \boldsymbol{\theta}_{i,a}^{(t)} - \boldsymbol{z} = \eta_i\boldsymbol{\xi}_{i,a}^{(t)}, \qquad \boldsymbol{d}^- = \boldsymbol{\theta}_{i,a}^{(t-1)} - \boldsymbol{z} = \eta_i(\boldsymbol{\xi}_{i,a}^{(t)} - \boldsymbol{r}_{i,a}^{(t)}).$$

By (H.5), both displacements have sup norm at most $98\eta_i$. The universal choice of c gives $98\eta_i \le 98c/k^{5/2} < 1/8$, so the local comparison applies on both segments. Taylor's integral formula gives

$$D_\Phi(\boldsymbol{z} + \boldsymbol{d}^+, \boldsymbol{z}) + D_\Phi(\boldsymbol{z} + \boldsymbol{d}^-, \boldsymbol{z}) \ge \frac{1}{4}\big(\boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{d}^+) + \boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{d}^-)\big) \ge b_{i,a}^{(t)}.$$

The second inequality uses $Q(x) + Q(y) \ge Q(x-y)/2$ for every positive-semidefinite quadratic form Q, together with $\boldsymbol{d}^+ - \boldsymbol{d}^- = \eta_i\boldsymbol{r}_{i,a}^{(t)}$. Also, by (B.6),

$$2D_\Phi(\boldsymbol{z} + \boldsymbol{d}^+, \boldsymbol{z}) \le \big\|\boldsymbol{d}^+\big\|_\infty^2.$$

Subtracting the two Taylor expansions yields

$$\Phi_i(\boldsymbol{\theta}_{i,a}^{(t)}) - \Phi_i(\boldsymbol{\theta}_{i,a}^{(t-1)}) \le \eta_i\left\langle \boldsymbol{q}_{i,a}^{(t)}, \boldsymbol{r}_{i,a}^{(t)}\right\rangle + \eta_i^2\left\|\boldsymbol{\xi}_{i,a}^{(t)}\right\|_\infty^2 - b_{i,a}^{(t)}.$$

The linear terms vanish after summing the source rows, because stationarity gives

$$\sum_{a,b} q_{i,a,b}^{(t)} r_{i,a,b}^{(t)} = (\boldsymbol{x}_i^{(t)})^\top(\boldsymbol{Q}_i^{(t)} - I)\boldsymbol{v}_i^{(t)} = 0.$$

Telescoping over time and using (B.9) gives

$$\sum_a \Phi_i(\boldsymbol{\theta}_{i,a}^{(T)}) \geq \eta_i \sum_a \max_b \sum_{t=1}^{T} r_{i,a,b}^{(t)} = \eta_i \operatorname{SwapReg}_i(T), \qquad \sum_a \Phi_i(0) \leq 2A_i.$$

Moreover,

$$\sum_{t=1}^{T} \sum_a \left\| \boldsymbol{\xi}_{i,a}^{(t)} \right\|_\infty^2 \leq \sum_{t=1}^{T} \left(\sum_a \left\| \boldsymbol{\xi}_{i,a}^{(t)} \right\|_\infty \right)^2 = \sum_{t=1}^{T} \left\| \boldsymbol{\xi}_i^{(t)} \right\|_{\infty,1}^2 \leq E^2.$$

Therefore

$$\eta_i \operatorname{SwapReg}_i(T) + S_i \leq 2A_i + \eta_i^2 E^2.$$

Multiplying by ω_i, summing over players, and using $\operatorname{SwapReg}_i(T) \geq 0$ gives

$$Y^2 \leq 2V + \Lambda^2 E^2.$$

Finally, $\Lambda^2 \leq (\max_i \omega_i) \sum_i \eta_i^2 \leq g^2 V$, while the default rates give $\Lambda^2 = g^2 V/A$. This proves Lemma 4.1.

C Uniform bounds for normalized derivatives

This appendix proves Lemma C.1, which supplies the local derivative bounds used in both prediction stages. The ordinary bound controls error-direction derivatives, and the order-raising identity represents gain-direction derivatives in the early stage. In the late stage, the curvature-weighted bound controls derivatives of existing atoms and residual factors directly: their score dependence is still differentiated, but their orders no longer increase.

For $e = (i, a, b)$ and $\boldsymbol{v} \in [0,1]^{m_i}$, define

$$\chi_{e,\boldsymbol{v}}^{\lambda}(\boldsymbol{y}) = \frac{h_e(\boldsymbol{y})}{q_e(\boldsymbol{y})}\big(v_b - \langle \boldsymbol{\sigma}_{i,a}(\boldsymbol{y}), \boldsymbol{v} \rangle\big), \qquad \chi_{e,\boldsymbol{v}}^{\alpha}(\boldsymbol{y}) = \alpha_e(\boldsymbol{y})\big(v_b - \langle \boldsymbol{\sigma}_{i,a}(\boldsymbol{y}), \boldsymbol{v} \rangle\big).$$

They depend only on the selected row and are invariant under a constant shift of that row.

Lemma C.1 (Uniform normalized derivatives)**.** *There is a universal integer $B \geq 1$. Set $B_r = B^{r+1} r!$, and for $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_r \in [0,1]^{m_i}$ in row (i,a) define*

$$\mathcal{J}_{e,\boldsymbol{v}_0,\ldots,\boldsymbol{v}_r}^{(r)}(\boldsymbol{y}) = B_r^{-1} \nabla_{\boldsymbol{v}_r} \cdots \nabla_{\boldsymbol{v}_1} \chi_{e,\boldsymbol{v}_0}^{\lambda}(\boldsymbol{y}).$$

Uniformly in $r \geq 0$,

$$|\mathcal{J}^{(r)}| \leq 1, \qquad |\nabla_{\boldsymbol{w}} \mathcal{J}^{(r)}| \leq C \min\left\{ \|\boldsymbol{w}_{i,a}\|_\infty, \beta_i^{-1/2} \sqrt{\boldsymbol{K}_{\boldsymbol{y}_{i,a}}(\boldsymbol{w}_{i,a})} \right\}, \tag{C.1}$$

$$\nabla_{\boldsymbol{v}_{r+1}} \mathcal{J}^{(r)} = B(r+1) \mathcal{J}^{(r+1)}.$$

Each factor is multilinear in its payoff directions and row-shift invariant. Also $|\chi^{\alpha}| \leq \alpha_e \leq 1$, and its first derivative obeys the same minimum bound as in (C.1)*.*

We first prove the χ^{α} statement. Suppress row indices and put $h_{\mathrm{total}} = \sum_b h_b$. By Lemma B.1, $h_{\mathrm{total}} \geq 3/10$. Differentiating $\sigma_b = h_b / h_{\mathrm{total}}$ gives

$$\nabla_{\boldsymbol{w}} \langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle = \operatorname{Cov}_{b \sim \boldsymbol{\sigma}}(v_b, h'(q_b) \widetilde{w}_b), \qquad |\nabla_{\boldsymbol{w}} \langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle| \leq \sqrt{\boldsymbol{K}_{\boldsymbol{y}}(\boldsymbol{w})/h_{\mathrm{total}}} \leq C \sqrt{\boldsymbol{K}_{\boldsymbol{y}}(\boldsymbol{w})}.$$

Moreover, Lemma B.1 gives

$$|\nabla_{\boldsymbol{w}} \alpha_e| \leq \frac{\delta}{2} \alpha_e^2 |\widetilde{w}_e| \leq \frac{\delta \alpha_e^2}{2\sqrt{h_e}} \sqrt{\boldsymbol{K}_{\boldsymbol{y}}(\boldsymbol{w})} \leq C \beta_i^{-1/2} \sqrt{\boldsymbol{K}_{\boldsymbol{y}}(\boldsymbol{w})}.$$

The product rule proves the curvature-weighted bound. Using $|\widetilde{w}_b| \leq 2 \|\boldsymbol{w}\|_\infty$ instead proves the ordinary bound.

It remains to prove the χ^{λ} statement. The argument has four steps.

C.1 SCALAR COMPLEX-ANALYTIC BOUNDS

Fix real s_0 and write $u_0 = u(s_0) > 0$. For $|\zeta| \le 1/8$,

$$|(s_0+\zeta)^2+4| \ge s_0^2+4-2|s_0||\zeta|-|\zeta|^2 \ge \frac{3}{4}(s_0^2+4).$$

Choose the square-root branch positive at s_0. The corresponding u is nonzero, because $(\sqrt{s^2+4}-s)(\sqrt{s^2+4}+s)=4$. The local logarithm satisfies $(\ln u)' = -1/\sqrt{s^2+4}$, hence

$$\left|\ln \frac{u(s_0+\zeta)}{u_0}\right| \le |\zeta|. \tag{C.2}$$

For every $c \ge 0$, the relative error of $1+cu$ from $1+cu_0$ is at most $e^{|\zeta|}-1$. For $1+u^2$ it is at most $e^{2|\zeta|}-1$. The three terms in (B.2) are therefore holomorphic, nonzero, and have uniformly bounded logarithmic derivatives on a smaller universal disk. Their real base values are positive, so

$$|\lambda(s_0+\zeta)-\lambda(s_0)| \le C|\zeta|\lambda(s_0), \qquad |\lambda(s_0+\zeta)| \le 2\lambda(s_0) \le 2.$$

Integrating $f'/f = \lambda$ and using $f' = f\lambda$, on a sufficiently small universal radius $r_s > 0$,

$$|f(s_0+\zeta)-f(s_0)| \le C|\zeta|f(s_0), \qquad |f'(s_0+\zeta)-f'(s_0)| \le C|\zeta|f'(s_0), \tag{C.3}$$

$$\left|\frac{f''(s_0+\zeta)}{f'(s_0+\zeta)}\right| \le C, \qquad |\lambda'(s_0+\zeta)| \le C\sqrt{f'(s_0)/\beta}.$$

No factor d appears: it cancels in the logarithmic derivative. For the last bound, if $u_0 \le 1$, use (B.3) and the uniform bound on λ'. If $u_0 \ge 1$, the first term of (B.2) is bounded by Cu_0^{-2}, the other terms by Cu_0^{-1}, and their logarithmic derivatives by Cu_0^{-1} uniformly by (C.2), after enlarging the universal constant C. Thus $|\lambda'| \le C/u_0^2$. Equation (B.4) gives

$$\sqrt{f'(s_0)/\beta} \ge \frac{c}{u_0\sqrt{1+\delta u_0}},$$

which dominates u_0^{-2} because $\sqrt{1+\delta u_0} \le \sqrt{2u_0}$.

C.2 A DIMENSION-UNIFORM NORMALIZATION RADIUS

Fix a real normalized row, write $s_b = z_b - \mu_0$, $h_b^0 = f'(s_b)$, and $h_{\text{total}}^0 = \sum_b h_b^0 \ge 3/10$. For a complex score displacement $\boldsymbol{\zeta}$, define

$$F(\boldsymbol{\zeta},\nu) = \sum_b f(s_b+\zeta_b-\nu) - 1, \qquad \Psi_{\boldsymbol{\zeta}}(\nu) = \nu + F(\boldsymbol{\zeta},\nu)/h_{\text{total}}^0.$$

Choose $r_0 > 0$ so that $3r_0 < r_s$ and the relative f' errors in (C.3) are at most $1/10$ on radius $3r_0$. For $\|\boldsymbol{\zeta}\|_\infty < r_0$ and $|\nu| \le 2r_0$,

$$|\partial_\nu \Psi_{\boldsymbol{\zeta}}| = \left|1 - \frac{\sum_b f'(s_b+\zeta_b-\nu)}{h_{\text{total}}^0}\right| \le \frac{1}{10}, \qquad |\Psi_{\boldsymbol{\zeta}}(0)| \le \frac{11}{10}r_0.$$

Thus $\Psi_{\boldsymbol{\zeta}}$ is a contraction of the closed disk $|\nu| \le 2r_0$ into itself. Iteration from zero converges uniformly, and its iterates are holomorphic. The limit $\nu(\boldsymbol{\zeta})$ is holomorphic and satisfies $F(\boldsymbol{\zeta},\nu(\boldsymbol{\zeta})) = 0$. It extends the real normalizer by $\mu(\boldsymbol{z}+\boldsymbol{\zeta}) = \mu_0 + \nu(\boldsymbol{\zeta})$.

At these complex scores, let $h_b = f'(s_b+\zeta_b-\nu)$, $h_{\text{total}} = \sum_b h_b$, and $\sigma_b = h_b/h_{\text{total}}$. Relative bounds give

$$|h_b| \le 2h_b^0, \qquad |h_{\text{total}}| \ge h_{\text{total}}^0/2, \qquad \sum_b |\sigma_b| \le 4. \tag{C.4}$$

The radius is independent of the row dimension and of the smallest probability.

C.3 Retaining one direction at real curvature

For fixed real $\boldsymbol{v} \in [0,1]^{m_i}$, write

$$\chi(\boldsymbol{z}+\boldsymbol{\zeta}) = \lambda(s_b + \zeta_b - \nu)\big(v_b - \langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle\big).$$

We claim that every real direction $\boldsymbol{w}$ satisfies

$$\sup_{\|\boldsymbol{\zeta}\|_\infty < r_0} |\nabla_{\boldsymbol{w}} \chi(\boldsymbol{z}+\boldsymbol{\zeta})| \le C \min\left\{ \|\boldsymbol{w}\|_\infty, \beta^{-1/2}\sqrt{\boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w})} \right\}. \tag{C.5}$$

Complex normalization and χ are invariant under a common score shift. For the weighted bound, subtract the real mean $\sum_b h_b^0 w_b / h_{\mathrm{total}}^0$, so $\sum_b h_b^0 w_b = 0$, and set $Q_0 = \sum_b h_b^0 w_b^2 = \boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w})$. Differentiating normalization gives $\nabla_{\boldsymbol{w}}\mu = \sum_b h_b w_b / h_{\mathrm{total}}$. By (C.4) and Cauchy–Schwarz, for $\widetilde{w}_b = w_b - \nabla_{\boldsymbol{w}}\mu$,

$$|\nabla_{\boldsymbol{w}}\mu| \le C\sqrt{Q_0/h_{\mathrm{total}}^0}, \qquad |\widetilde{w}_b| \le C\sqrt{Q_0/h_b^0}, \qquad \sum_b h_b^0 |\widetilde{w}_b|^2 \le C Q_0.$$

Writing $\varkappa_b = f''/f'$ at the complex scalar arguments,

$$\nabla_{\boldsymbol{w}} \langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle = \frac{1}{h_{\mathrm{total}}} \sum_b h_b \varkappa_b \widetilde{w}_b \big(v_b - \langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle\big), \qquad |\nabla_{\boldsymbol{w}} \langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle| \le C\sqrt{Q_0}.$$

The last bound uses $|\varkappa_b| \le C$, (C.4), and $h_{\mathrm{total}}^0 \ge 3/10$. The product rule gives

$$\nabla_{\boldsymbol{w}} \chi = \lambda' \widetilde{w}_b \big(v_b - \langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle\big) - \lambda \nabla_{\boldsymbol{w}} \langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle .$$

The first term is bounded by $\sqrt{h_b^0/\beta}\sqrt{Q_0/h_b^0} = \beta^{-1/2}\sqrt{Q_0}$, and the second by $C\sqrt{Q_0}$. For the ordinary bound, do not center $\boldsymbol{w}$ at the real point. The same formulas give $|\nabla_{\boldsymbol{w}}\mu| \le 4\|\boldsymbol{w}\|_\infty$ and $|\widetilde{w}_b| \le 5\|\boldsymbol{w}\|_\infty$. This proves (C.5). The same bounds also give $|\chi| \le 10$.

C.4 Cauchy normalization

For $r \ge 1$ and directions of sup norm at most one, the map

$$(t_1, \dots, t_r) \longmapsto \chi\left(\boldsymbol{z} + \sum_j t_j \boldsymbol{v}_j\right)$$

is holomorphic on $|t_j| < r_0/(2r)$. Repeated Cauchy bounds give

$$|\nabla_{\boldsymbol{v}_r} \cdots \nabla_{\boldsymbol{v}_1} \chi| \le 10(2r/r_0)^r \le 10(2e/r_0)^r r!.$$

Apply the same bounds to the holomorphic function $\nabla_{\boldsymbol{w}}\chi(\boldsymbol{z}+\boldsymbol{\zeta})$ using (C.5). Only r further derivatives are taken, so

$$|\nabla_{\boldsymbol{v}_r} \cdots \nabla_{\boldsymbol{v}_1} \nabla_{\boldsymbol{w}} \chi| \le C(2e/r_0)^r r! \min\left\{ \|\boldsymbol{w}\|_\infty, \beta^{-1/2}\sqrt{\boldsymbol{K}_{\boldsymbol{z}}(\boldsymbol{w})} \right\}.$$

Choose a universal integer B at least $2e/r_0$, 10, and the finite constants in these bounds. Divide by $B_r = B^{r+1} r!$. This proves Lemma C.1, including $r = 0$, multilinearity, row-shift invariance, and the recursion $B_{r+1}/B_r = B(r+1)$. The weighted quadratic form remains at the fixed real row throughout. No complex curvature comparison or minimum-probability bound is used.

D Stationary trees, centered gain identities, and moment bounds

This appendix supplies the probabilistic inputs to both prediction stages. The first subsection represents each deviation gain under the stationary-tree distribution. The second proves an unweighted player-level tree-score bound for the early stage and a curvature-weighted tree-score bound for the late stage. The third proves the moment bound that permits products of overlapping centered atoms.

D.1 THE MARKOV CHAIN TREE THEOREM

Let Q be a strictly positive row-stochastic matrix. For a root a, set

$$w_a = \sum_{\tau \text{ directed toward } a} \prod_{(c,d)\in\tau} q_{c,d}.$$

Every w_a is positive. For a fixed b, take a tree rooted at $a \neq b$, add the edge $a \to b$, and delete the outgoing edge $b \to c$ on the resulting cycle. This gives a weight-preserving bijection with a tree rooted at b together with a distinguished edge $b \to c$. Hence

$$\sum_{a\neq b} w_a q_{a,b} = w_b \sum_{c\neq b} q_{b,c}.$$

Therefore $x_a = w_a / \sum_c w_c$ satisfies $\boldsymbol{x}^\top (Q - I) = 0$. If $(Q - I)\boldsymbol{v} = 0$ and v_a is maximal, then

$$0 = \sum_{b\neq a} q_{a,b}(v_b - v_a).$$

Every summand is nonpositive and every coefficient is positive, so every $v_b = v_a$. The right kernel, and hence the left kernel, is one-dimensional. Thus the normalized vector above is the unique stationary distribution. Replacing one redundant stationarity equation by $\sum_a x_a = 1$ gives the nonsingular linear system used by the exact stationary-distribution solve in the learning dynamics.

For each player and score array $\boldsymbol{y}$, let $\mathbb{P}^{\mathrm{tr}}_{i,\boldsymbol{y}}$ be the corresponding normalized distribution on rooted directed trees, and let $\mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}} = \bigotimes_i \mathbb{P}^{\mathrm{tr}}_{i,\boldsymbol{y}}$. Write I_e for edge inclusion, with diagonal indicators zero, and $a_i(\boldsymbol{T})$ for the root of player i's tree. A tree has $m_i - 1$ edges and at most one outgoing edge from each source row.

Define

$$\boldsymbol{v}_i(\boldsymbol{T}) = \big(u_i(b, \boldsymbol{a}_{-i}(\boldsymbol{T}))\big)_{b\in[m_i]}, \qquad G_{i,a,b}(\boldsymbol{T}) = \mathbf{1}_{\{a_i(\boldsymbol{T})=a\}}\big(u_i(b, \boldsymbol{a}_{-i}(\boldsymbol{T})) - u_i(a, \boldsymbol{a}_{-i}(\boldsymbol{T}))\big).$$

Then $\|\boldsymbol{G}_i(\boldsymbol{T})\|_{\infty,1} \leq 1$ and

$$\boldsymbol{r}_i^{(s)} = \mathbb{E}_{\boldsymbol{T}\sim\mathbb{P}^{\mathrm{tr}}_{\boldsymbol{z}^{(s)}}} \boldsymbol{G}_i(\boldsymbol{T}). \tag{D.1}$$

At trial scores $\boldsymbol{y}$, define the centered represented gain

$$\widetilde{r}^{(s)}_{i,a,b}(\boldsymbol{y}) = x^{(s)}_{i,a}\left(v^{(s)}_{i,b} - \left\langle \boldsymbol{\sigma}_{i,a}(\boldsymbol{y}), \boldsymbol{v}_i^{(s)} \right\rangle\right).$$

For either $\zeta = \alpha$ or $\zeta = h/q$,

$$\zeta_e(\boldsymbol{y})\widetilde{r}^{(s)}_e(\boldsymbol{y}) = \mathbb{E}_{\boldsymbol{T}\sim\mathbb{P}^{\mathrm{tr}}_{\boldsymbol{z}^{(s)}}}\left[\mathbf{1}_{\{a_i(\boldsymbol{T})=a\}}\chi^{\zeta}_{e,\boldsymbol{v}_i(\boldsymbol{T})}(\boldsymbol{y})\right], \tag{D.2}$$

where $\chi^{h/q} = \chi^{\lambda}$. The identity follows from independence of player i's root and the opponents' roots. A fresh independent sample makes it valid inside an expectation involving old samples.

D.2 TREE SCORE BOUNDS

Lemma D.1 (Tree score bounds)**.** *For a deterministic score direction $\boldsymbol{w}$,*

$$\nabla_{\boldsymbol{w}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}} = \sum_e (I_e - \mathbb{E}I_e)\frac{h_e}{q_e}\widetilde{w}_e = \sum_e (I_e - \mathbb{E}I_e)(\gamma_e + \delta\alpha_e)\widetilde{w}_e. \tag{D.3}$$

Moreover,

$$\left\|\nabla_{\boldsymbol{w}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}}\right\|_{L^2} \leq 2\left(\sum_i \|\boldsymbol{w}_i\|^2_{\infty,1}\right)^{1/2}, \tag{D.4}$$

$$\mathbb{E}\left|\sum_e (I_e - \mathbb{E}I_e)\gamma_e\widetilde{w}_e\right|^2 \leq \sum_{i,a} \omega_i \boldsymbol{K}_{\boldsymbol{y}_{i,a}}(\boldsymbol{w}_{i,a}). \tag{D.5}$$

If a normed-vector integrand $\boldsymbol{F}$ satisfies $\mathbb{E}\|\boldsymbol{F}\|^2 \leq 1$, these inequalities also bound the norm of the expectation of $\boldsymbol{F}$ times the corresponding centered score.

Proof. A tree's log-weight derivative is $\sum_e I_e(h_e/q_e)\widetilde{w}_e$, by normalized row differentiation. Differentiating the tree normalizer subtracts its mean, proving (D.3). The row center is essential even before the tree mean is subtracted.

Since $h_e/q_e \le 1$, player i's uncentered tree score has absolute value at most $2\,\|\boldsymbol{w}_i\|_{\infty,1}$. Independence across players proves (D.4). For deterministic coefficients c_e, Cauchy–Schwarz on the $m_i - 1$ tree edges gives

$$\operatorname{Var}\left(\sum_e I_e c_e\right) \le \sum_i (m_i - 1) \sum_{e \in \mathcal{I}_i} c_e^2.$$

Take $c_e = \gamma_e \widetilde{w}_e$, use $\gamma_e^2 \le C_\gamma h_e/\beta_i$, and recall $\omega_i = C_\omega(m_i - 1)/\beta_i$ with $C_\omega \ge C_\gamma$. This proves (D.5). Finally,

$$\|\mathbb{E}(\boldsymbol{F}X)\| \le (\mathbb{E}\,\|\boldsymbol{F}\|^2)^{1/2}(\mathbb{E}|X|^2)^{1/2}.$$

Here $\boldsymbol{F}$ need not be independent of the centered score. □

Bound (D.4) is used in the early closure, while (D.5) is used in the late closure.

D.3 A MOMENT BOUND FOR OVERLAPPING CENTERED FACTORS

Lemma D.2 (Centered sums and products)**.** *Conditional on an auxiliary object D_0, suppose the pairs $(\boldsymbol{T}_i, \boldsymbol{T}_i')$ are independent across players and iid within each pair. Let $F_i(\cdot, D_0) \in [-1, 1]$, and let fixed real weights satisfy $\sum_i a_i^2 \le 1$. For integer $\ell \ge 1$, define*

$$Z = \frac{1}{2\sqrt{\ell}} \sum_i a_i \big(F_i(\boldsymbol{T}_i, D_0) - F_i(\boldsymbol{T}_i', D_0)\big).$$

Then $\|Z\|_{L^{2\ell}} \le 1$. If at most ℓ factors each have this marginal bound, their product has L^2 norm at most one, regardless of dependencies between factors.

Proof. Conditional Hoeffding bounds give

$$\mathbb{E}\exp\left(t \sum_i a_i(F_i - F_i')\right) \le e^{t^2}.$$

The resulting two-sided tail $2e^{-u^2/4}$ implies

$$\mathbb{E}\left|\sum_i a_i(F_i - F_i')\right|^{2\ell} \le 2\,4^\ell \ell!.$$

For $\ell \ge 2$, $2\ell! \le \ell^\ell$. For $\ell = 1$, variance at most two suffices. Remove the conditioning. For $h \le \ell$, Hölder gives

$$\mathbb{E}\prod_{j=1}^h |Z_j|^2 \le \prod_{j=1}^h (\mathbb{E}|Z_j|^{2h})^{1/h} \le 1.$$

□

This moment bound supplies the L^2 product bound used in both expression classes.

E THE EARLY PLAYER-CENTERED CLOSURE

This appendix verifies the early expression class and one-difference closure used in Section 4.3. Combining Lemma E.1 with the coefficient bound (E.5) supplies the early-stage attenuation in Proposition 4.3. Throughout, we fix the finite deterministic effective-score history. Integrals may include finite sums and compact parameter intervals.

E.1 SCORE TYPES, SLOTS, AND ATOMS

A *score type* is a sequence

$$\boldsymbol{y}_\tau^{(t)} = \sum_{a=0}^{p} c_{\tau,a} \boldsymbol{z}^{(t-a)}, \qquad c_{\tau,a} \geq 0, \qquad \sum_a c_{\tau,a} = 1, \tag{E.1}$$

with coefficients independent of time and common to all players. A *sample slot* q is an independent product-tree draw $\boldsymbol{T}^q \sim \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}^{(t)}_{\tau(q)}}$. Different slots may have the same type without being the same sample. Write $\Omega = (\boldsymbol{T}^1, \ldots, \boldsymbol{T}^M)$ for their tuple and $\mathbb{E}_t$ for expectation under their time-t product distribution.

An order-r atom uses two distinct inclusion slots q, q' of one type τ, and pairwise distinct direction slots $d_0, \ldots, d_r$ disjoint from q, q'. Put

$$a_i^0 = a_i(\boldsymbol{T}^{d_0}), \qquad \boldsymbol{v}_i^j = \boldsymbol{v}_i(\boldsymbol{T}^{d_j}),$$

and

$$F_i(\boldsymbol{T}^q) = \left(\prod_{j=1}^{r} \mathbf{1}_{\{a_i(\boldsymbol{T}^{d_j}) = a_i^0\}} \right) \sum_b I_{i,a_i^0,b}(\boldsymbol{T}_i^q) \mathcal{J}^{(r)}_{i,a_i^0,b,\boldsymbol{v}_i^0,\ldots,\boldsymbol{v}_i^r}(\boldsymbol{y}_\tau).$$

For fixed weights $\sum_i a_i^2 \leq 1$, define

$$Z^{(r)}(\Omega, \boldsymbol{y}_\tau^{(t)}) = \frac{1}{2\sqrt{\ell}} \sum_i a_i \big(F_i(\boldsymbol{T}^q) - F_i(\boldsymbol{T}^{q'}) \big).$$

We suppress the sample and score arguments when unambiguous. Each selected row has at most one outgoing tree edge, so $|F_i| \leq 1$. Conditioning on this atom's direction slots leaves its inclusion pair iid and independent across players. Hence Lemma D.2 gives $\big\| Z^{(r)} \big\|_{L^{2\ell}} \leq 1$. Different atoms may share slots, even in different roles. We apply the marginal bound separately and use unconditional Hölder for products.

A depth-p early expression, $0 \leq p \leq \ell$, has the form

$$\boldsymbol{U}^{(t)} = \mathbf{1}_{\{t>p\}} \mathbb{E}_t \left[\boldsymbol{G}(\Omega) \prod_{\nu=1}^{h} Z_\nu^{(r_\nu)}(\Omega, \boldsymbol{y}_{\tau_\nu}^{(t)}) \right], \qquad \sum_{\nu=1}^{h} (r_\nu + 1) \leq p. \tag{E.2}$$

There are at most $1 + 2p$ independent slots, and all types have the window (E.1). The observable is a fixed matrix-valued function of the sample tuple with $\|\boldsymbol{G}(\Omega)\|_{\infty,1} \leq 1$. Initially it is $\boldsymbol{G}_i(\boldsymbol{T}^1)$ from (D.1). The observable, slot assignments, atom weights, and type coefficients are independent of time. Holding the samples fixed, all score dependence lies in the normalized factors. The expectation also depends on the scores through the sampling distributions. By Lemma D.2, the integrand has squared norm expectation at most one, so

$$\big\| \boldsymbol{U}^{(t)} \big\|_{\infty,1} \leq 1.$$

The gain representation (D.1) is a depth-zero expression. We call an index s attached to a gain direction $\boldsymbol{r}^{(s)}$ introduced by this representation a *represented gain time*. The leading indicator $\mathbf{1}_{\{t>p\}}$ is the *gate*. It ensures that every represented gain time introduced later is positive.

E.2 ONE-DIFFERENCE CLOSURE

Lemma E.1 (Early one-difference closure)**.** *Every depth-p early expression, $p < \ell$, has an exact decomposition*

$$\mathrm{D}\boldsymbol{U} = \boldsymbol{V} + \int c(\nu) \boldsymbol{U}_\nu \, d\nu,$$

where $\boldsymbol{V}$ consists of the boundary impulse and error-direction terms, every $\boldsymbol{U}_\nu$ has early depth $p+1$, and the coefficients are time-independent. Moreover,

$$\|\boldsymbol{V}\|_{\infty,1;T} \leq 1 + C_E(p+1)gE, \qquad \int |c(\nu)| \, d\nu \leq C_E g \sqrt{\ell} \, (p+1).$$

Proof of Lemma E.1. The gate contributes one impulse at $t = p + 1$, of source norm at most one. For $t \geq p + 2$, interpolate every score type simultaneously between its value at $t - 1$ and t. For independent slots with distributions $P_q(\theta)$, finite differentiation is exactly

$$\frac{d}{d\theta}\mathbb{E}_\theta F_\theta = \mathbb{E}_\theta \frac{dF_\theta}{d\theta} + \sum_q \mathbb{E}_\theta \left[F_\theta \frac{d}{d\theta} \ln P_q(\theta) \right]. \tag{E.3}$$

Equal types move together, but each independent slot density contributes separately. By (3.4), every type direction is a convex combination of $\eta_i \boldsymbol{r}_i^{(s)} - \eta_i \mathrm{D}\boldsymbol{\xi}_i^{(s)}$ at positive represented times. Treat one shift at a time.

Error directions stop. For a distribution derivative, use (D.4) with $\boldsymbol{w}_i = -\eta_i \mathrm{D}\boldsymbol{\xi}_i$. Since backward differences have kernel mass two,

$$\sum_i \eta_i^2 \sum_{t=1}^{T} \left\| (\mathrm{D}\boldsymbol{\xi}_i)^{(t)} \right\|_{\infty,1}^2 \leq 4g^2E^2.$$

The sequence cost is therefore at most CgE.

For an atom, freeze its samples. At most two outgoing edges in its inclusion pair contribute per player. By (C.1) and Cauchy–Schwarz over the deterministic atom weights,

$$|\nabla_{\boldsymbol{w}} Z^{(r)}| \leq \frac{C}{\sqrt{\ell}} \left(\sum_i \|\boldsymbol{w}_i\|_{\infty,1}^2 \right)^{1/2}. \tag{E.4}$$

After removing the differentiated atom, the remaining product has L^1 norm at most one. Its error contribution is again at most CgE. There are at most $1 + 2p$ distribution densities and at most p atoms, so, after enlarging a universal constant, all stopped error terms cost at most $C_E(p + 1)gE$.

Gain-direction distribution derivatives continue. When a slot q is differentiated, introduce an independent ghost q' of the same type. Inside the expectation, replace $I_e(\boldsymbol{T}^q) - \mathbb{E}I_e$ by $I_e(\boldsymbol{T}^q) - I_e(\boldsymbol{T}^{q'})$. The ghost is independent of the complete old integrand. Introduce a separate fresh gain slot d_0 at the represented time s, using (D.2) with $\zeta = h/q$. The multiplier is

$$\sum_i \eta_i \sum_{a,b} \left[I_{i,a,b}(\boldsymbol{T}_i^q) - I_{i,a,b}(\boldsymbol{T}_i^{q'}) \right] \mathbf{1}_{\{a_i(\boldsymbol{T}^{d_0}) = a\}} \chi^\lambda_{i,a,b,\boldsymbol{v}_i(\boldsymbol{T}^{d_0})}(\boldsymbol{y}_\tau).$$

Since $\chi^\lambda = B\mathcal{J}^{(0)}$, this is $2Bg\sqrt{\ell}$ times a new order-zero atom with weights $a_i = \eta_i/g$. It adds two slots and one degree unit.

Gain-direction atom derivatives raise the order. Row-shift invariance removes the constant source payoff. For an order-r normalized factor,

$$\nabla_{\eta_i \boldsymbol{r}_{i,a}^{(s)}} \mathcal{J}^{(r)} = \eta_i x_{i,a}^{(s)} B(r+1) \mathcal{J}^{(r+1)}_{\ldots, \boldsymbol{v}_i^{(s)}}.$$

A fresh gain slot d_{r+1} represents both the source probability and the additional payoff argument, by multilinearity and root independence. Factor out $\eta_{\max} B(r + 1)$ and replace the atom weights by $a_i\eta_i/\eta_{\max}$. Their squared sum remains at most one. One slot and one degree unit are added.

These are all product-rule terms. Distribution-density continuations have total coefficient mass at most $C(1 + 2p)g\sqrt{\ell}$. Explicit-factor continuations have mass at most $Bg\sum_\nu (r_\nu + 1) \leq Bgp$. This proves the stated coefficient bound after fixing C_E. Old types interpolate into the enlarged window, ghosts inherit a type, and new gain slots have a single-time type. Every coefficient remains independent of time, and the children satisfy all slot, disjointness, and degree conditions. □

Early coefficient mass. Starting from a depth-zero gain, let a_p bound the total absolute coefficient mass at depth p. The closure gives

$$a_p \leq (C_E g\sqrt{\ell})^p p!.$$

For $p \leq \ell$, $p! \leq p^p \leq k^p$, and therefore

$$a_p k^p \leq (C_E g k^{5/2})^p = \kappa^p. \tag{E.5}$$

E.3 WORKED FIRST AND SECOND DIFFERENCES

By the solver guarantee (H.3),

$$E \le 1 + \max_i \|\mathrm{H}[\boldsymbol{r}_i]\|_{\infty,1;T}\,.$$

A pointwise gain bound alone would give a bound growing as $\sqrt{T}$. Instead, we use the differences already present in H and expand only the terms not yet controlled.

The filter supplies a backward difference. Since $\mathrm{F}_L = \mathrm{R}_L\mathrm{D}$ for every cascade depth $L \ge 1$ and all factors commute,

$$\mathrm{H} = \mathrm{R}_\ell^\ell \mathrm{R}_N^N \mathrm{D}^{N+\ell+3}.$$

For $0 \le p \le \ell$, define

$$\mathrm{K}_p := \mathrm{R}_\ell^\ell \mathrm{R}_N^N \mathrm{D}^{N+\ell+3-p}.$$

Then $\mathrm{K}_p = \mathrm{K}_{p+1}\mathrm{D}$ for $p < \ell$, and the first step is

$$\mathrm{H}[\boldsymbol{r}_i] = \mathrm{K}_0[\boldsymbol{r}_i] = \mathrm{K}_1[\mathrm{D}\boldsymbol{r}_i].$$

We now follow the first two differences of the gain to see what can be bounded and what must be expanded further.

The first difference separates error and gain directions. Fix $t \ge 2$ and interpolate $\boldsymbol{y}_\theta^{(t)} = (1-\theta)\boldsymbol{z}^{(t-1)} + \theta\boldsymbol{z}^{(t)}$. In the following calculation, suppress the superscript (t) and write $\mathbb{E}_\theta$ for expectation at $\boldsymbol{y}_\theta$. For a product direction, use the shorthand $\eta\boldsymbol{w} = (\eta_j \boldsymbol{w}_j)_{j=1}^n$. Applying (4.9) and using $\mathrm{D}\boldsymbol{z} = \eta\boldsymbol{r} - \eta\mathrm{D}\boldsymbol{\xi}$, we obtain

$$\mathrm{D}\boldsymbol{r}_i = \underbrace{-\int_0^1 \mathbb{E}_\theta\big[\boldsymbol{G}_i \nabla_{\eta\mathrm{D}\boldsymbol{\xi}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}_\theta}\big]\, d\theta}_{\text{error direction: stop and bound using } E} \quad + \quad \underbrace{\int_0^1 \mathbb{E}_\theta\big[\boldsymbol{G}_i \nabla_{\eta\boldsymbol{r}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}_\theta}\big]\, d\theta}_{\text{gain direction: continue taking derivative}}\,.$$

The error contribution has cumulative norm at most CgE by (4.8) and

$$\sum_j \eta_j^2 \,\|\mathrm{D}\boldsymbol{\xi}_j\|^2_{\infty,1;T} \le 4g^2E^2.$$

At $t = 1$, zero prehistory contributes the boundary impulse $\boldsymbol{r}_i^{(1)}$, of source norm at most one. We call these contributions *stopped*: they are included in the final bound without taking further differences.

A gain-direction distribution derivative creates an atom. For $e = (j,a,b)$ and $\boldsymbol{v} \in [0,1]^{m_j}$, define the local centered factor

$$\chi^\lambda_{e,\boldsymbol{v}}(\boldsymbol{y}) = \frac{h_e(\boldsymbol{y})}{q_e(\boldsymbol{y})}\big(v_b - \langle \boldsymbol{\sigma}_{j,a}(\boldsymbol{y}), \boldsymbol{v}\rangle\big).$$

The gain-direction score in (4.7) is a centered sum of these factors over players, weighted by their learning rates and source probabilities. Fix $t \ge 2$ and $\theta \in [0,1]$. Introduce an independent copy of the original tree sample and a separate gain sample:

$$\boldsymbol{T}, \boldsymbol{T}' \sim \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}_\theta^{(t)}}, \qquad \boldsymbol{S} \sim \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{z}^{(t)}},$$

with all three samples independent. Write $\mathbb{E}^+_{t,\theta}$ for expectation under this enlarged distribution. The construction in Appendix E gives an order-zero *atom* $Z_\theta^{(0)}$, with player weights η_j/g, such that

$$\mathbb{E}_\theta\Big[\boldsymbol{G}_i(\boldsymbol{T})\nabla_{\eta\boldsymbol{r}^{(t)}} \ln \mathbb{P}^{\mathrm{tr}}_{\boldsymbol{y}_\theta^{(t)}}(\boldsymbol{T})\Big] = c_0 \mathbb{E}^+_{t,\theta}\Big[\boldsymbol{G}_i(\boldsymbol{T}) Z_\theta^{(0)}\Big], \qquad c_0 := 2Bg\sqrt{\ell}.$$

Here B is the universal constant in Lemma C.1. Indeed, averaging over $\boldsymbol{T}'$ replaces $I_e(\boldsymbol{T}) - I_e(\boldsymbol{T}')$ by $I_e(\boldsymbol{T}) - \mathbb{E}I_e$, and averaging over $\boldsymbol{S}$ represents the gain direction by (D.2). Thus the equality holds after averaging over the auxiliary samples, not pointwise between the original score and the atom. We suppress the sample and score arguments of the atom.

Define the continuing sequence, for each fixed θ, by

$$\boldsymbol{U}_\theta^{(t)} := \begin{cases} \mathbb{E}^+_{t,\theta}[\boldsymbol{G}_i(\boldsymbol{T}) Z_\theta^{(0)}], & t \geq 2, \\ 0, & t \leq 1. \end{cases}$$

The first difference is therefore the exact sequence identity

$$\mathrm{D}\boldsymbol{r}_i = \boldsymbol{V} + c_0 \int_0^1 \boldsymbol{U}_\theta \, d\theta.$$

Here $\boldsymbol{V}$ contains the boundary impulse at $t = 1$ and the error-direction integral above for $t \geq 2$.

The second difference adds a factor or raises its order. Fix one continuing sequence $\boldsymbol{U} = \boldsymbol{U}_\theta$, whose integrand has the form $\boldsymbol{G} Z^{(0)}$, with $\boldsymbol{G} = \boldsymbol{G}_i(\boldsymbol{T})$. For $t \geq 3$, interpolate all its score arguments between times $t-1$ and t and apply the product rule. Every sampling distribution contributes a derivative, including the distributions of the independent copy and the gain sample. A gain-direction distribution derivative creates a new order-zero atom. A gain derivative of the existing atom creates an order-one atom. Collecting the error directions into $\boldsymbol{V}$ gives the exact identity

$$(\mathrm{D}\boldsymbol{U})^{(t)} = \boldsymbol{V}^{(t)} \quad + \quad \underbrace{\int c_1(\nu)\, \mathbb{E}_{t,\nu,1}\left[\boldsymbol{G} Z^{(0)}_{\nu,1} Z^{(0)}_{\nu,2}\right] d\nu}_{\text{differentiate a sampling distribution}} \quad + \quad \underbrace{\int c_2(\nu)\, \mathbb{E}_{t,\nu,2}\left[\boldsymbol{G} Z^{(1)}_{\nu}\right] d\nu}_{\text{differentiate the existing atom}}.$$

The two integrals include all interpolation parameters and finite branching choices for their respective derivative families. Each $\mathbb{E}_{t,\nu,k}$ uses the complete auxiliary-sample distribution of that branch, with the old observable $\boldsymbol{G}$ viewed as a function of the enlarged sample tuple. The coefficients $c_k(\nu)$ are independent of time. Sample and score arguments of the atoms are suppressed. At $t = 2$, set $\boldsymbol{V}^{(2)} = \boldsymbol{U}^{(2)}$ and all continuing terms to zero. All terms are zero for $t \leq 1$. As throughout the recursion, $\boldsymbol{V}$ denotes the stopped sequence for the expression currently being differentiated.

More generally, a gain-direction distribution derivative adds $Z^{(0)}$, whereas a gain derivative of $Z^{(r)}$ raises its order to $r+1$. An order-r atom contains normalized order-r derivatives of the local factor, together with auxiliary payoff samples. Lemma C.1 and the exact constructions in Appendix E justify these two operations.

The form preserved by the recursion. Every continuing branch at depth p therefore has the form

$$\boldsymbol{U}^{(t)} = \mathbf{1}_{\{t>p\}} \mathbb{E}_t \left[\boldsymbol{G}(\Omega) \prod_{\nu=1}^{h} Z_\nu^{(r_\nu)}(\Omega, \boldsymbol{y}_\nu^{(t)}) \right], \qquad \sum_{\nu=1}^{h} (r_\nu + 1) \leq p.$$

Here $\mathbb{E}_t$ averages the auxiliary product-tree samples Ω under their time-t distributions, and $\boldsymbol{G}$ is a fixed observable with $\|\boldsymbol{G}\|_{\infty,1} \leq 1$. Each continuing difference adds one unit of degree. The indicator ensures positive represented gain times and produces a boundary impulse at the next difference. Initially there are no atoms, so the gain is a depth-zero expression. Throughout, an *early expression* means the admissible construction of Appendix E, including its sample-independence conditions and fixed convex-combination rules for the score arguments. Section 4.3 bounds this recursion and proves the early-stage handoff.

F THE LATE TRANSITION-INDEX CLOSURE

This appendix verifies the late expression class and proves Lemma F.2, the unfiltered bound used in Section 4.4. It records the distinct transition indices encountered along each branch: curvature-controlled terms and repeated indices stop, while only a fresh index continues. Appendix G combines this closure with the remaining filters to obtain Proposition 4.4 and Lemma 4.2.

F.1 LATE EXPRESSIONS AND THEIR SCORE WINDOW

A depth-j late expression, $0 \leq j \leq N$, has the form

$$\boldsymbol{U}^{(t)} = \mathbf{1}_{\{t>\ell+j\}} \mathbb{E}_t \left[\boldsymbol{G}(\Omega) \prod_{\nu=1}^{h} Z_\nu^{(r_\nu)}(\Omega, \boldsymbol{y}_{\tau_\nu}^{(t)}) \prod_{\mu=1}^{j} \chi^{\alpha}_{e_\mu(\Omega), \boldsymbol{v}_\mu(\Omega)}(\boldsymbol{y}_{\upsilon_\mu}^{(t)}) \right].$$

There are at most $1+2\ell+2j$ independent slots. Score types are fixed convex combinations in the depth-$(\ell+j)$ window. All early-atom conditions hold with total degree at most ℓ. The observable, transition indices, and payoff directions depend only on samples and fixed parameters, with $\|\boldsymbol{G}\|_{\infty,1} \le 1$. Whenever $e_\mu(\Omega) = (i,a,b)$, the associated payoff vector satisfies $\boldsymbol{v}_\mu(\Omega) \in [0,1]^{m_i}$. The j transition indices are distinct wherever $\boldsymbol{G} \neq 0$, and every scalar residual factor has a fixed score type even when its transition index is sample-dependent. By Lemma D.2 and $|\chi^\alpha| \le 1$, the integrand has squared norm expectation at most one. An early depth-ℓ expression is a late depth-zero expression.

We first record the consequence of the energy normalization used throughout this appendix. If $s \ge 1$ and $\boldsymbol{y}$ lies within distance ε of $\boldsymbol{z}_{i,a}^{(s)}$, then (B.8) gives

$$\eta_i \sqrt{\boldsymbol{K}_{\boldsymbol{y}}(\boldsymbol{r}_{i,a}^{(s)})} \le \sqrt{8}\, e^{2\varepsilon} \sqrt{b_{i,a}^{(s)}}.$$

Keeping one summand in the quadratic form gives the same bound for $\eta_i \sqrt{h_e(\boldsymbol{y})}\, |\widetilde{r}_e^{(s)}(\boldsymbol{y})|$.

Lemma F.1 (Comparison on the late score window)**.** *In a depth-j difference, let $\boldsymbol{y}, \boldsymbol{y}'$ lie in the convex hull of $\boldsymbol{z}^{(t)}, \dots, \boldsymbol{z}^{(t-\ell-j-1)}$, and let $s \ge 1$ be a represented gain time in this window. For $e = (i,a,b)$,*

$$\eta_i \sqrt{\boldsymbol{K}_{\boldsymbol{y}_{i,a}}(\boldsymbol{r}_{i,a}^{(s)})} \le C\, 2^j \sqrt{b_{i,a}^{(s)}}, \tag{F.1}$$

$$\delta \eta_i \alpha_e(\boldsymbol{y}) \alpha_e(\boldsymbol{y}') |\widetilde{r}_e^{(s)}(\boldsymbol{y})| \le C\, 2^j \left(\sum_{i',a'} \omega_{i'} b_{i',a'}^{(s)} \right)^{1/2}. \tag{F.2}$$

Proof. From $\boldsymbol{z} = \boldsymbol{\theta} - \eta \boldsymbol{\xi}$, the gain bound and (H.5), the player-i hull diameter is at most $\eta_i(\ell+j+195)$. The comparisons in (B.7)–(B.8) therefore cost at most

$$C \exp\big((2+\delta)\eta_{\max}(\ell+j+195)\big) \le C\, 2^j,$$

where the last inequality uses $c \le 1/16$, $k \ge 5$, and $\delta \le 1/2$. This proves (F.1). For (F.2), compare the two α values, use $\delta \alpha_e^2 \le C\sqrt{h_e/\beta_i}$, retain the corresponding summand in $\boldsymbol{K}$, and use $\omega_i \ge 1/\beta_i$. This is a deterministic, index-uniform bound before any new gain sample is introduced. □

F.2 One-difference closure

Lemma F.2 (Late one-difference closure)**.** *Every depth-j late expression, $0 \le j \le N$, has an exact decomposition*

$$\mathrm{D}\boldsymbol{U} = \boldsymbol{V} + \int c(\nu) \boldsymbol{U}_\nu \, d\nu,$$

with time-independent coefficients. The stopped sequence $\boldsymbol{V}$ contains the boundary impulse, error directions, the γ-score, derivatives of existing factors, and repeated transition indices. For a universal C_L,

$$\|\boldsymbol{V}\|_{\infty,1;T} \le C_L k^2 8^j (1 + gE + Y), \qquad \int |c(\nu)|\, d\nu \le 2n\delta\eta_{\max}(1 + 2\ell + 2j).$$

If $j < N$, every continuing expression has late depth $j+1$. At $j = N$, the continuing term is absent.

Proof of Lemma F.2. The gate contributes one impulse. At later times, interpolate equal score types together and apply (E.3). Directions are convex combinations of $\eta_i \boldsymbol{r}_i^{(s)} - \eta_i \mathrm{D}\boldsymbol{\xi}_i^{(s)}$ at positive represented times.

Error directions stop. Use Lemma D.1 for a distribution, (E.4) for an early atom, and the ordinary derivative bound in Lemma C.1 for a scalar residual factor. Each has sequence cost at most CgE, even when a selected transition index is sample-dependent.

Curvature-controlled gain derivatives stop. For a distribution, split its centered score as in (D.3). Keep the entire γ score intact and use (D.5) with the complete integrand. Its L^2 bound and Lemma F.1 give cost $C\,2^jY$.

For an existing atom, freeze the samples and use the curvature-weighted branch of (C.1), as in the proof of (E.4), to obtain

$$|\nabla_{\eta \boldsymbol{r}^{(s)}} Z^{(r)}| \le \frac{C}{\sqrt{\ell}} \left(\sum_{i,a} \beta_i^{-1} \eta_i^2 \boldsymbol{K}_{\boldsymbol{y}_{i,a}}(\boldsymbol{r}_{i,a}^{(s)}) \right)^{1/2}.$$

Remove that atom and bound the remaining product in L^1. Since $\omega_i \ge 1/\beta_i$, Lemma F.1 gives cost $C\,2^jY$. Existing scalar residual factors stop by the same weighted derivative bound. Thus no old normalized derivative order increases in the late stage.

Repeated transition indices stop. If the $\delta\alpha$ part of a distribution derivative selects an already present transition index e, its old factor supplies $\alpha_e(\boldsymbol{y}')$, while its centered payoff has magnitude at most one. Apply (F.2) while the gain direction is still deterministic. The centered edge indicator has magnitude at most one. Removing the old factor that pays for the repetition leaves an L^1-bounded product. At most j transition indices can repeat on any sample outcome, so this costs at most $Cj2^jY$ per distribution.

There are at most $1 + 3\ell + 3j$ distribution densities and explicit factors. Summing these payments and the gate gives

$$\begin{aligned} \|\boldsymbol{V}\|_{\infty,1;T} &\le 1 + C(\ell + j + 1)gE + C(\ell + j + 1)^2 2^j Y \\ &\le C_L k^2 8^j (1 + gE + Y). \end{aligned}$$

For every fixed time shift, the squared curvature sequence sums to at most Y^2. Convex shift weights and interpolation do not enlarge this finite-prefix bound. The last inequality uses $\ell + j + 1 \le k(j+1)$ and $(j+1)^2 \le 4^j$.

Only fresh transition indices continue. Fix a candidate transition index e and a differentiated slot q. Let $M_e(\Omega)$ be the indicator that e is absent from the old transition indices. For a new independent ghost q' of the same distribution,

$$\mathbb{E}\big[FM_e(I_e(\boldsymbol{T}^q) - \mathbb{E}I_e)\big] = \mathbb{E}\big[FM_e(I_e(\boldsymbol{T}^q) - I_e(\boldsymbol{T}^{q'}))\big]. \tag{F.3}$$

The mask may depend on $\boldsymbol{T}^q$. Only independence of the new ghost is needed. Introduce a separate fresh gain slot at time s using (D.2) with $\zeta = \alpha$. Split the indicator difference by sign and player. In each branch, the gain root selects a source row, and the indicated tree has at most one outgoing edge from that row. Hence the complete sum over candidate transition indices selects at most one edge.

If the selected edge is fresh, append its χ^α factor at the differentiated distribution's score type, and put the freshness mask and $\eta_i/\eta_{\max}$ into the bounded observable. Otherwise set the observable to zero and use an arbitrary fixed dummy transition index for that player. Its value is irrelevant on this zero-observable branch. This selection occurs after the fixed-candidate identity (F.3). It neither conditions nor renormalizes a tree distribution.

Each player/sign/distribution branch has coefficient $\delta\eta_{\max}$. There are at most $2n(1 + 2\ell + 2j)$ branches, proving the continuation bound. Two slots and one distinct transition index are added. Old atoms remain unchanged, so their marginal moment bounds remain valid. A bounded sample-dependent mask cannot increase the integrand's L^2 norm. At depth $N = |\mathcal{I}|$, no fresh transition index remains wherever the observable is nonzero. □

G TWO-SCALE FILTER BOUNDS AND PROOF OF THE PREDICTION BOUND

This appendix supplies the filter bounds used to assemble the two stages. After bounding the remaining kernels, it applies the early closure from Appendix E and the late closure from Appendix F to complete Lemma 4.2, with all bounds on the same finite horizon.

G.1 FILTER BOUNDS

For a convolution $\mathrm{K} = \sum_{j\ge 0} c_j \mathrm{S}^j$, recall $\|\mathrm{K}\|_{\mathrm{ker},1} = \sum_j |c_j|$. Backward shifts contract the finite-prefix squared sum, so

$$\sum_{t=1}^{T} \left\| (\mathrm{K}[\boldsymbol{w}])^{(t)} \right\|_{\infty,1}^2 \le (\|\mathrm{K}\|_{\mathrm{ker},1})^2 \sum_{t=1}^{T} \left\| \boldsymbol{w}^{(t)} \right\|_{\infty,1}^2 .$$

Lemma G.1 (Two-scale filter bounds)**.** *For $0 \le a \le \ell$ and $0 \le b \le N$,*

$$\left\| \mathrm{D}^2 \mathrm{F}_\ell^a \mathrm{F}_N^b \right\|_{\mathrm{ker},1} < 48, \qquad \|\mathrm{H}\|_{\mathrm{ker},1} < 96.$$

Define

$$\mathrm{K}_p = \mathrm{R}_\ell^\ell \mathrm{R}_N^N \mathrm{D}^{N+\ell+3-p} \quad (0 \le p \le \ell+1), \qquad \mathrm{L}_j = \mathrm{R}_N^N \mathrm{D}^{N+3-j} \quad (0 \le j \le N+1).$$

Then $\mathrm{K}_0 = \mathrm{H}$, $\mathrm{K}_p = \mathrm{K}_{p+1}\mathrm{D}$, $\mathrm{L}_j = \mathrm{L}_{j+1}\mathrm{D}$, *and*

$$\begin{gathered} \|\mathrm{K}_{p+1}\|_{\mathrm{ker},1} < 48k^p \quad (0 \le p < \ell), \qquad \mathrm{K}_\ell = \mathrm{R}_\ell^\ell \mathrm{L}_0, \\ \|\mathrm{L}_{j+1}\|_{\mathrm{ker},1} < 24(N+1)^j \quad (0 \le j \le N). \end{gathered} \tag{G.1}$$

Proof. If $F(z) = \sum_{j\ge 0} a_j z^j$ is analytic on a neighborhood of the closed unit disk, coefficient Cauchy–Schwarz and normalized circle orthogonality give

$$\sum_{j\ge 0} |a_j| \le \sqrt{1+\pi/2} \left(\|F\|_{L^2(\mathbb{T})}^2 + \|F'\|_{L^2(\mathbb{T})}^2 \right)^{1/2}. \tag{G.2}$$

Indeed, $\sum_{j\ge 0}(1+j^2)^{-1} \le 1 + \int_0^\infty (1+x^2)^{-1} dx = 1 + \pi/2$. This scalar calculation does not require the sequence output space to be Hilbert.

For an integer cascade depth $L \ge 1$, the scalar symbol of F_L satisfies, on $|z| = 1$,

$$F_L(z) = \frac{1-z}{1 - Lz/(L+1)}, \qquad |F_L(z)| \le \frac{2(L+1)}{2L+1}, \qquad |F_L(z)|^j < 2 \quad (0 \le j \le L+1).$$

For $F_{a,b}(z) = (1-z)^2 F_\ell(z)^a F_N(z)^b$, differentiation gives

$$F'_{a,b} = -2(1-z) F_\ell^a F_N^b - \frac{a}{\ell+1} F_\ell^{a+1} F_N^b - \frac{b}{N+1} F_\ell^a F_N^{b+1}.$$

Thus $|F_{a,b}| \le 16$ and $|F'_{a,b}| \le 24$ in the stated ranges. Equation (G.2) gives kernel norm below 48. One additional difference costs at most two, proving $\|\mathrm{H}\|_{\mathrm{ker},1} < 96$.

For the single-scale symbol $(1-z)^2 F_N(z)^b$, the analogous bounds are 8 and 10, giving kernel norm below 24. Finally,

$$\mathrm{K}_{p+1} = \mathrm{R}_\ell^p \mathrm{D}^2 \mathrm{F}_\ell^{\ell-p} \mathrm{F}_N^N, \qquad \mathrm{L}_{j+1} = \mathrm{R}_N^j \mathrm{D}^2 \mathrm{F}_N^{N-j}.$$

The nonnegative resolvent kernels have masses k^p and $(N+1)^j$, respectively, which proves (G.1). These factorizations retain the difference operators. In particular, the bare resolvent R_N^N has mass $(N+1)^N$ and is not being bounded by a universal constant. □

G.2 PROOF OF THE PREDICTION BOUND

Proof of Lemma 4.2. Starting from a depth-zero gain, let a_p be the total early coefficient mass. By (E.5), $a_p k^p \le \kappa^p$ for $p \le \ell$. Coefficients are time-independent, so filter every exact early decomposition using $\mathrm{K}_p = \mathrm{K}_{p+1}\mathrm{D}$. By Lemma G.1, all early stopped contributions sum to at most

$$48 \sum_{p=0}^{\ell-1} \kappa^p [1 + C_E(p+1)gE] \le C + CgE.$$

At the handoff, $\mathrm{K}_\ell = \mathrm{R}_\ell^\ell \mathrm{L}_0$. The kernel mass of R_ℓ^ℓ is k^ℓ, already paid by $a_\ell k^\ell \le \kappa^\ell$.

For a late depth-zero expression, set

$$a_{\mathrm{late}} = 2n\delta\eta_{\max}(1 + 2\ell + 2N), \qquad B_{\mathrm{late}} = 1 + gE + Y.$$

Substitute the late closure repeatedly and retain the stopped term at depth N. Since $\mathrm{L}_j = \mathrm{L}_{j+1}\mathrm{D}$,

$$\left(\sum_{t=1}^{T} \left\| (\mathrm{L}_0[\boldsymbol{U}])^{(t)} \right\|_{\infty,1}^2 \right)^{1/2} \le 24 C_L k^2 B_{\mathrm{late}} \sum_{j=0}^{N} (8(N+1)a_{\mathrm{late}})^j \le 25 C_L k^2 B_{\mathrm{late}}.$$

Indeed,

$$8(N+1)a_{\mathrm{late}} \le 32n\delta\eta_{\max}(N + \ell + 2)^2 = 32\eta_{\max} \le \frac{2}{5^{5/2}} < \frac{1}{25}.$$

This is a finite sum of $N+1$ terms, including all boundary impulses. There is no terminal remainder or infinite-depth limit. All filters depend only on current and earlier inputs, so only the same finite-prefix energies appear.

Multiply the late bound by the handoff coefficient κ^ℓ. Since $k^2\kappa^\ell$ is uniformly bounded when $\kappa \le 1/16$, adding the unit solver-error norm and collecting the constant and gE terms gives

$$E \le C_0 + C_1 gE + C_P k^2 \kappa^\ell Y,$$

which is (4.4). □

H Bracket-and-mix row normalization and implementation

This appendix gives the finite normalization solver used in Algorithm 1, proves its summable effective-score error, and states the computation model. Sections H.1–H.2 then verify the predictor and score identities used by the main proof.

Fix player i, one raw score row $\overline{\boldsymbol{z}} \in \mathbb{R}^{m_i}$, and local time t. If all coordinates coincide, return the uniform row, which is exact by symmetry. Otherwise initialize

$$\mu^- = \min_b \overline{z}_b - \frac{3}{m_i\beta_i}, \qquad \mu^+ = \max_b \overline{z}_b.$$

At μ^+, every scalar argument is nonpositive, and the row sum is at most $m_i f_i(0) \le e m_i\beta_i/2 < 1$. At μ^-, every argument is at least $s = 3/(m_i\beta_i) > 0$. Since $t_+(s) \ge s$ and $u(s) \le 1/s$, Bernoulli's inequality gives

$$f_i(s) \ge \beta_i \frac{1+s}{1 + s^{-1} + \delta/(4s^2)} > \frac{1}{m_i}.$$

Thus the endpoints bracket the unique normalization shift.

Bisect until an exactly normalized midpoint is found or the bracket width is at most

$$e_{i,t} = \frac{\eta_i}{m_i(t+1)^2}. \tag{H.1}$$

An exact root is returned immediately. Otherwise put

$$y_b^- = f_i(\overline{z}_b - \mu^-), \qquad y_b^+ = f_i(\overline{z}_b - \mu^+), \qquad S^\pm = \sum_b y_b^\pm, \qquad \tau = \frac{1 - S^+}{S^- - S^+},$$

$$\widehat{q}_b = \tau y_b^- + (1-\tau) y_b^+.$$

Since $S^- > 1 > S^+$, the returned row is strictly positive and exactly normalized.

Lemma H.1 (Effective-score representation)**.** *The returned row is the exact normalized row at $\overline{\boldsymbol{z}} + \boldsymbol{\zeta}$ for some vector satisfying $0 \le \zeta_b \le \mu^+ - \mu^-$ in every coordinate. Across all source rows, this proves* (H.3)*.*

Proof. Monotonicity places $\widehat{q}_b$ between $f_i(\overline{z}_b - \mu^+)$ and $f_i(\overline{z}_b - \mu^-)$. Define, only for the proof,

$$\zeta_b = f_i^{-1}(\widehat{q}_b) - (\overline{z}_b - \mu^+).$$

Then $0 \le \zeta_b \le \mu^+ - \mu^-$ and

$$\widehat{q}_b = f_i(\overline{z}_b + \zeta_b - \mu^+), \qquad \sum_b \widehat{q}_b = 1.$$

Uniqueness of the normalizing shift identifies the returned row with $\boldsymbol{q}_i(\overline{\boldsymbol{z}} + \boldsymbol{\zeta})$. Exact roots and uniform rows use $\boldsymbol{\zeta} = 0$. Set $\boldsymbol{\varepsilon} = \boldsymbol{\zeta}/\eta_i$. For every source row,

$$\left\|\boldsymbol{\varepsilon}_{i,a}^{(t)}\right\|_\infty \le \frac{1}{m_i(t+1)^2}.$$

Summing over the m_i source rows gives

$$\left\|\boldsymbol{\varepsilon}_i^{(t)}\right\|_{\infty,1} \le (t+1)^{-2}, \qquad \sum_{t=1}^{T} \left\|\boldsymbol{\varepsilon}_i^{(t)}\right\|_{\infty,1}^2 \le \sum_{t\ge 1}(t+1)^{-4} < 1.$$

□

Work, storage, and computation model. The operation counts below assume that the public parameters are supplied as exact real constants. They count exact arithmetic, comparison, and square-root operations during play. Public initialization, including the exact logarithms used to define these parameters, is not included in these per-round counts. No finite-bit complexity claim is made. At local round t, player i uses

$$O\big(m_i^2\,[N + \log(nm)\log(nmt)] + m_i^3\big) \text{ operations and } O\big((N+\ell)m_i^2\big) \text{ storage.} \tag{H.2}$$

Since $N \le nm^2$, this is at most $O\big(m^2[nm^2 + \log(nm)\log(nmt)]\big)$ operations and $O(nm^4)$ storage.

The public bounds imply $d_i \le C\log(nm)$. Evaluating f_i uses arithmetic, one square root, and $O(\log(nm))$ repeated-squaring operations. The implementation evaluates neither f_i^{-1}, the potential, higher derivatives, nor trees. By (H.5), the raw-score range at round t is at most $2\eta_i(t+96)$, so the initial bracket width divided by (H.1) is at most

$$2m_i(t+96)(t+1)^2 + \frac{3(t+1)^2}{\beta_i\eta_i}.$$

For the default rates, $1/(\beta_i\eta_i)$ is polynomial in nm by $J \le 5k$, $A \le N \le nm^2$, and $A_i \ge 2$. For the optional rates, (A.1) gives $\eta_i^{-1} \le N/(2g)$. Hence either allocation requires $O(\log(nmt))$ bisections and yields the normalization term in (H.2).

The $N+\ell$ cascade stages use $O((N+\ell)m_i^2)$ work and storage. The exact stationary-distribution system is nonsingular by Appendix D. Gaussian elimination costs $O(m_i^3)$ operations and $O(m_i^2)$ storage. These are per-round counts under the computation model specified above. Inexact stationarity, sampled-action feedback, numerical conditioning, finite-bit complexity, and floating-point behavior are outside the present guarantees.

H.1 Cascade implementation identity

Write $\boldsymbol{a}_j^{(t)}$ for the state entering round t and $\boldsymbol{u}_j^{(t)}$ for the residual produced after cascade stage j in that round, with $\boldsymbol{u}_0^{(t)} = \boldsymbol{r}^{(t)}$ and $\boldsymbol{c}^{(t)} = \boldsymbol{u}_{N+\ell}^{(t)}$. The update gives $[I - (1-\nu_j)\mathrm{S}]\boldsymbol{u}_j = \mathrm{D}\boldsymbol{u}_{j-1}$, where $\nu_j = (N+1)^{-1}$ in the first cascade and $\nu_j = (\ell+1)^{-1}$ in the second. Hence

$$\boldsymbol{c} = \mathrm{F}_\ell^\ell \mathrm{F}_N^N \boldsymbol{r}, \qquad \sum_j \boldsymbol{a}_j^{(t)} = \boldsymbol{r}^{(t)} - \boldsymbol{c}^{(t)}, \qquad \boldsymbol{r} - \widehat{\boldsymbol{r}} = \mathrm{D}^3\boldsymbol{c} = \mathrm{H}[\boldsymbol{r}].$$

Appendix G proves the depth-uniform filter bounds.

H.2 Effective scores and summable normalization error

We define the cumulative score and the raw score passed to row normalization, respectively:

$$\boldsymbol{\theta}_i^{(t)} := \eta_i \sum_{s\le t} \boldsymbol{r}_i^{(s)}, \qquad \bar{\boldsymbol{z}}_i^{(t)} := \boldsymbol{\theta}_i^{(t-1)} + \eta_i \widehat{\boldsymbol{r}}_i^{(t)}$$

Lemma H.1 shows that the finite row normalization solver returns an effective score

$$\boldsymbol{z}_i^{(t)} = \overline{\boldsymbol{z}}_i^{(t)} + \eta_i \boldsymbol{\varepsilon}_i^{(t)}, \qquad \left\|\boldsymbol{\varepsilon}_i^{(t)}\right\|_{\infty,1} \le (t+1)^{-2}, \qquad \sum_{s=1}^{T} \left\|\boldsymbol{\varepsilon}_i^{(s)}\right\|_{\infty,1}^2 \le 1. \tag{H.3}$$

Exact normalization corresponds to $\varepsilon_i^{(t)} = 0$. Using $\widehat{\boldsymbol{r}}_i = (I - \mathrm{H})[\boldsymbol{r}_i]$ and $\boldsymbol{\theta}_i^{(t)} = \boldsymbol{\theta}_i^{(t-1)} + \eta_i \boldsymbol{r}_i^{(t)}$, the effective score can be rewritten as

$$\boldsymbol{z}_i^{(t)} = \boldsymbol{\theta}_i^{(t)} - \eta_i \left(\mathrm{H}[\boldsymbol{r}_i]^{(t)} - \boldsymbol{\varepsilon}_i^{(t)}\right).$$

We therefore define the effective prediction error

$$\boldsymbol{\xi}_i := \mathrm{H}[\boldsymbol{r}_i] - \boldsymbol{\varepsilon}_i.$$

Consequently,

$$\boldsymbol{z}_i = \boldsymbol{\theta}_i - \eta_i \boldsymbol{\xi}_i, \qquad \mathrm{D}\boldsymbol{z}_i = \eta_i \big(\boldsymbol{r}_i - \mathrm{D}\boldsymbol{\xi}_i\big). \tag{H.4}$$

The filter bounds of Appendix G and (H.3) imply, for arbitrary payoff-vector sequences,

$$\left\|\widehat{\boldsymbol{r}}_i^{(t)}\right\|_{\infty,1} \le 97, \qquad \left\|\boldsymbol{\xi}_i^{(t)}\right\|_{\infty,1} \le 97. \tag{H.5}$$

H.3 Feedback-loop schematic

Figure 1 summarizes the feedback loop in Algorithm 1. The prediction and row normalization precede stationary play. After observing the payoff vector, the player updates the cumulative scores and predictor state.

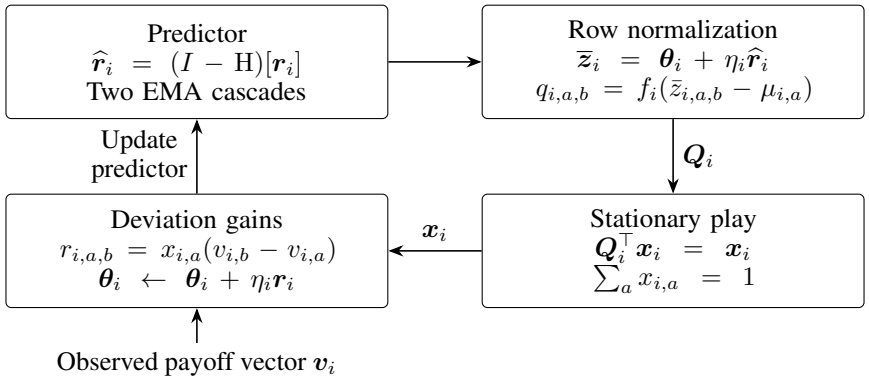


Figure 1: Optimistic transition-matrix dynamics. Each row uses its own normalization shift. The variable $\boldsymbol{\theta}_i$ is the cumulative score entering the round. The diagram depicts exact normalization. Appendix H gives the finite normalization solver and its effective-score representation.

I Anytime fallback and the common-prefix wrapper

This appendix completes the robustness argument in Section 5. We first verify the anytime fallback guarantee, then prove the common-prefix wrapper, and finally substitute the base bound and public prefix length to obtain Theorem 2.2.

I.1 An anytime adversarial fallback

We define the fallback FB used in Section 5. For player i, use $A_i = m_i \log m_i$ and doubling epochs with planned lengths $M = 1, 2, 4, \ldots$. Initialize each epoch's row distributions uniformly and set

$$\eta_{\mathrm{fb}} = \sqrt{A_i / M}.$$

Algorithm 2 Common-prefix robust wrapper for player i

Input. Base dynamics, a threshold $B_i \geq 0$, and a common prefix length $W \geq 1$.
Anytime fallback FB_i**.** Use epochs of lengths $M = 1, 2, 4, \ldots$. At each epoch start, initialize $q_{a,b} = 1/m_i$ and set $\eta_{\mathrm{fb}} = \sqrt{A_i/M}$. In each round, play the stationary distribution $\boldsymbol{x}$ of $\boldsymbol{Q}$, observe $\boldsymbol{v}$, set $g_{a,b} = x_a v_b$, and update

$$q_{a,b} \leftarrow \frac{q_{a,b}(1 + \eta_{\mathrm{fb}} g_{a,b})}{1 + \eta_{\mathrm{fb}} \langle \boldsymbol{q}_a, \boldsymbol{g}_a \rangle}.$$

After M rounds, double M and reset the rows uniformly. This fallback may be stopped after any s rounds and satisfies $\mathrm{SwapReg}_{i,\mathsf{FB}}(s) \leq 4\sqrt{A_i s}$.
Switching rule.

1: Run a new copy of FB_i for rounds $1, \ldots, W$.
2: Discard it and initialize the base dynamics from its prescribed initial state.
3: After each base round s, compute (5.1). At the first s with $R_{i,\mathrm{base}}(s) > B_i$, discard the base state and run a new copy of FB_i forever from the next round.

Play the stationary distribution $\boldsymbol{x}$ of the strictly positive row-stochastic matrix. After observing $\boldsymbol{v}$, set $g_{a,b} = x_a v_b$ and update

$$q^+_{a,b} = \frac{q_{a,b}(1 + \eta_{\mathrm{fb}} g_{a,b})}{1 + \eta_{\mathrm{fb}} \langle \boldsymbol{q}_a, \boldsymbol{g}_a \rangle}.$$

Rows remain strictly positive and exactly normalized for every rate, because $g_{a,b} \geq 0$. This is the Blum–Mansour reduction (Blum & Mansour, 2007) with a linear multiplicative-weights update.

Comparing the unnormalized weight of a fixed destination with the row-weight sum gives, for an epoch prefix,

$$\sum_t \ln(1 + \eta_{\mathrm{fb}} g^{(t)}_{a,b}) - \ln m_i \leq \sum_t \ln\left(1 + \eta_{\mathrm{fb}} \left\langle \boldsymbol{q}^{(t)}_a, \boldsymbol{g}^{(t)}_a \right\rangle\right).$$

For $z \geq 0$, $z - z^2/2 \leq \ln(1 + z) \leq z$. No upper rate cap is needed. Thus the cumulative row comparison is at most

$$\frac{\ln m_i}{\eta_{\mathrm{fb}}} + \frac{\eta_{\mathrm{fb}}}{2} \sum_t (g^{(t)}_{a,b})^2.$$

Sum over source rows with a fixed deviation map. Stationarity identifies the linear comparison with the map's actual gain, while the squared comparator gains satisfy

$$\sum_a x_a^2 v^2_{\varphi(a)} \leq \sum_a x_a^2 \leq 1$$

per round. A prefix of $\tau \leq M$ rounds in the epoch therefore satisfies

$$\mathrm{SwapReg}_{\mathrm{epoch}}(\tau) \leq \sqrt{A_i M}\left(\ln 2 + \frac{\tau}{2M}\right).$$

Let the current epoch have planned length M and $1 \leq \tau \leq M$ completed rounds. Total time is $s = M - 1 + \tau$. All earlier epochs are complete. Put $d = \ln 2 + 1/2$. Their total cost is at most

$$d\sqrt{A_i}\frac{\sqrt{M} - 1}{\sqrt{2} - 1}.$$

With $x = \tau/M$ and $B_M = d(1 - M^{-1/2})/(\sqrt{2} - 1)$, divide the total by $\sqrt{A_i M}$. The ratio

$$\frac{B_M + \ln 2 + x/2}{\sqrt{1 - M^{-1} + x}}$$

decreases in x: its derivative has the sign of $1 - M^{-1} + x/2 - (B_M + \ln 2)$, which is nonpositive for $M \geq 2$ and $x \leq 1$. The maximum is at $x = 1/M$, where $s = M$ and

$$B_M + \ln 2 + \frac{1}{2M} < 2.9 + 0.7 + 0.25 < 4.$$

For $M = 1$, the bound is immediate. Hence $\mathrm{SwapReg}(s) \leq 4\sqrt{A_i s}$ for every $s \geq 0$, proving the fallback guarantee stated in Algorithm 2. The exact stationary-distribution solve costs $O(m_i^3)$ operations and $O(m_i^2)$ storage. Row updates cost $O(m_i^2)$. These bounds use the same exact-real-arithmetic convention as the implementation analysis in Appendix H.

I.2 Proof of the common-prefix wrapper

Starting immediately with the base dynamics and switching after the first violation of the anytime self-play bound would give $\mathrm{SwapReg}_i(T) \le B_i + 1 + 4\sqrt{A_i T}$, retaining the base budget additively. The common prefix instead makes this budget absorbable once $B_i + 1 = O(\sqrt{A_i W})$, while its simultaneous reset makes the subsequent play a fresh self-play trajectory.

Proof of Lemma 5.1. Swap regret is subadditive over consecutive intervals, since each fixed deviation map's total gain is the sum of its interval gains. Moreover, $R_{i,\mathrm{base}}(s) \le R_{i,\mathrm{base}}(s-1) + 1$, because the one-round gain of every deviation map is at most one.

If a threshold were crossed in self-play, consider the first crossing among all players. Through that round every player has followed the simultaneously initialized fresh base dynamics, contradicting the assumed anytime self-play bound. Hence no threshold-triggered switch occurs.

For an arbitrary payoff sequence, split the horizon into the initial fallback, base, and terminal fallback intervals. If a crossing occurs, the base interval contributes at most $B_i + 1$ by the one-round bound. If none occurs, it contributes at most B_i. Applying the fallback guarantee stated in Algorithm 2 to the two newly initialized fallback intervals and enlarging the terminal interval to length $T - W$ proves (5.2). Under $B_i + 1 \le \gamma_i \sqrt{A_i W}$, for $T > W$,

$$\mathrm{SwapReg}_i(T) \le (4 + \gamma_i)\sqrt{A_i W} + 4\sqrt{A_i(T - W)} \le \sqrt{(4 + \gamma_i)^2 + 16}\,\sqrt{A_i T},$$

by Cauchy–Schwarz. For $T \le W$, the same fallback guarantee gives the conclusion. □

For the stated extension to a nondecreasing anytime bound, let $b_i(s)$ be nondecreasing. At a first crossing time τ, the one-round bound gives $R_{i,\mathrm{base}}(\tau) \le b_i(\tau) + 1$. Hence, under $b_i(s) + 1 \le \gamma_i \sqrt{A_i(W + s)}$, the switched case is bounded by

$$4\sqrt{A_i W} + \gamma_i \sqrt{A_i(W + \tau)} + 4\sqrt{A_i(T - W - \tau)} \le \sqrt{(4 + \gamma_i)^2 + 16}\,\sqrt{A_i T},$$

where we used $W \le W + \tau$ before Cauchy–Schwarz. Without a crossing, $4\sqrt{A_i W} + b_i(T - W)$ is bounded directly by the same envelope. The same first-crossing argument prevents switching in self-play.

I.3 Specialization to Algorithm 1

Put

$$W_0 = \frac{9k^5 A}{c^2}.$$

Then $B_i = \frac{4}{3}\sqrt{A_i W_0}$ and $W = \lceil W_0 \rceil$. Since $c \le 1/16$, $k \ge 5$, and $A \ge 2$, we have $W_0 \ge 6$, hence $W \le (7/6)W_0$.

By (4.6), the anytime self-play regret of the base dynamics is at most $3B_i/4$, so Lemma 5.1 prevents switching. The initial fallback satisfies

$$4\sqrt{A_i W} \le 3\sqrt{7/6}\, B_i < \frac{13}{4} B_i.$$

Adding the base continuation proves the self-play inequality in (5.5), including horizons within the common prefix.

Also $B_i \ge 4$, so

$$B_i + 1 \le \frac{5}{4} B_i \le \frac{5}{3}\sqrt{A_i W}.$$

Thus Lemma 5.1 applies with $\gamma_i = 5/3$, and

$$\sqrt{(4 + 5/3)^2 + 16} = \sqrt{433/9} < 7.$$

This proves the adversarial part of Theorem 2.2. The preceding paragraph proves its self-play part.

Maintaining (5.1) costs $O(m_i^2)$ work and storage per round and does not change the stated asymptotic implementation bounds. All states are discarded at phase transitions, and the threshold is evaluated only after the crossing round's feedback has been observed. Every round, including the common prefix and crossing round, has been counted.