SELF-BOUNDING REGRET MATCHING+ IN POTENTIAL GAMES AND PRODUCT-SIMPLEX OPTIMIZATION

Pahan Dewasurendra
Johns Hopkins University

Subhashini Jayawardhana
San Diego Miramar College

ABSTRACT

Regret matching+ (RM+) is parameter free, scale invariant, and central to large game solving, but its only general individual-regret guarantee grows as $\sqrt{T}$. A recent ICLR result used this envelope to prove that RM+ reaches an ϵ-stationary point of a smooth objective over a product of simplices in $O(\epsilon^{-4})$ iterations, or $O(\epsilon^{-8})$ from the standard zero initialization. We give an exact one-step conservation law for RM+. It states that forward utility gain pays for both squared state motion and growth of the regret-state norm. Norm growth is at most $\sqrt{m-1}$ times forward gain for m actions, and the coefficient is sharp. This yields four results for unmodified RM+. Its regret on any utility path is controlled by centered temporal variation. Its regret is uniformly bounded under alternating play in every finite exact potential game, resolving an open question and making squared activation gaps summable. Both certified lazy and ordinary cyclic play attain an ϵ^{-2} exponent. On any smooth, possibly nonconcave simplex objective, RM+ finds an ϵ-KKT point in $O(\epsilon^{-2})$ iterations. Most broadly, for a smooth objective over an arbitrary product of simplices, cyclic block RM+ attains the same $O(\epsilon^{-2})$ exponent from arbitrary initialization, with an explicit trajectory-dependent constant. The proof controls the finite objective loss caused by low-state blocks and then self-bounds every block state and the total squared path length. Complete proofs cover zero states, sharpness, common-profile stationarity, and robust gain dominance. Oracle-normalized diagnostics compare RM+ with predictive and smooth extra-gradient variants on graphical potential games and dense nonconvex objectives.

1 INTRODUCTION

Regret matching is a foundational procedure for online learning and adaptive play (Hart & Mas-Colell, 2000). Its positive-truncation variant, regret matching+ (RM+), is a basic component of counterfactual regret minimization systems that solve very large imperfect-information games (Tammelin, 2014; Zinkevich et al., 2007). RM+ is unusually simple. It keeps a nonnegative regret vector, normalizes it to obtain the next mixed strategy, and introduces no learning-rate parameter. Its scale invariance and practical speed have motivated connections to Blackwell approachability and mirror descent, predictive variants, and stabilization schemes (Farina et al., 2021; 2023; Cai et al., 2025; Zhang et al., 2025; Meng et al., 2026).

Beyond two-player zero-sum games, however, even qualitative convergence was unresolved until recently. Anagnostides et al. (2026b) proved that RM+ reaches approximate KKT points for smooth optimization over products of simplices. Their result implies convergence to Nash equilibria in potential games, a question open for two decades (Marden et al., 2007; Kleinberg et al., 2009). The analysis gives an $O(\epsilon^{-4})$ rate when every block is initialized above a prescribed threshold and only $O(\epsilon^{-8})$ from standard zero initialization. Its key denominator is the accumulated regret state. If that state were uniformly bounded, their framework would instead give the canonical $O(\epsilon^{-2})$ rate. They explicitly leave open whether RM+ can incur $\Omega(\sqrt{T})$ regret in a potential game. The broader optimization benchmarks of Tewolde et al. (2026) show strong empirical RM+ performance. No proof or superconstant example was known.

We prove that constant regret is forced by the RM+ update whenever its forward gains have finite budget. Let r and r' be consecutive nonnegative regret states, let x and x' be their normalizations,

and let u be the current utility vector. Define the forward gain $D = \langle x' - x, u \rangle$. Our starting point is the exact identity

$$2\|r'\|_1 D = \|r'\|_2^2 - \|r\|_2^2 + \|r' - r\|_2^2. \tag{1}$$

The squared-motion term underlies the earlier one-step improvement analysis. The first term on the right was discarded there. Retaining it closes the loop because regret itself is bounded by the state norm. In fact, we prove the sharp charge

$$\|r'\|_2 - \|r\|_2 \le \sqrt{m-1}\, D. \tag{2}$$

The factor $\sqrt{m-1}$ cannot be improved.

Equation equation 2 gives a pathwise online-learning theorem. For every utility sequence and arbitrary nonnegative initialization,

$$\mathrm{Reg}_T \le \|r_T\|_2 \le \|r_0\|_2 + \sqrt{m-1} \sum_{t=1}^{T} \langle x_{t+1} - x_t, u_t \rangle. \tag{3}$$

This is a self-bounding guarantee. It does not require the utilities to be stochastic, smooth, or chosen obliviously. Summation by parts immediately bounds the last sum by centered temporal variation. Thus vanilla RM+, without predictions, attains the better of its adversarial square-root guarantee and a first-order variation guarantee whose accompanying state-norm constant is sharp.

The same theorem has stronger consequences when the utility path is generated by optimization or a game. In an exact potential game under sequential unilateral updates, D_t is exactly the increase in the mixed extension of the potential. It is nonnegative, and its sum never exceeds the potential range. Therefore every player's regret is bounded by a horizon-independent constant. This answers the open question for alternating RM+ and improves the ϵ dependence of both lazy and ordinary cyclic play from four to two. The ordinary bound has a trajectory-dependent δ^{-2} constant, while a phased lazy variant avoids this constant, is target-free, and always retains a certified approximate Nash profile.

For a smooth objective, the argument is less direct because forward linearized gain need not equal objective improvement. Above an explicit state-norm threshold, smoothness ensures that the objective pays at least half of D_t. On one simplex, the finite objective range therefore bounds all subsequent gain, and equation 2 prevents further state growth. This gives

$$\min_{1 \le t \le T} \mathrm{Gap}(x_t) = O(T^{-1/2}), \tag{4}$$

where Gap is the simplex KKT gap. This is an $O(\epsilon^{-2})$ iteration bound for the original parameter-free trajectory.

The main obstacle on a product of simplices is that low-state block updates may decrease the objective and erase high-state ascent. We show that this loss is finite. Before a block crosses its threshold, state monotonicity bounds its squared state motion. A trajectory-dependent lower bound δ on positive state norms converts this into a finite squared strategy-motion budget, hence a finite smoothness loss. The objective range then bounds all high-state gain. The sharp charge closes a simultaneous system of bounds on every block state and on the global squared path. Finally, smoothness transfers each block's activation-time gap to a common cycle endpoint. For cyclic block RM+ on any smooth product-simplex objective,

$$\min_{1 \le k \le K} \mathrm{Gap}_{\Pi}(x^k) = O(K^{-1/2}), \tag{5}$$

with all constants explicit. Thus the ϵ exponent improves from four to two under prescribed large initialization and from eight to two under zero initialization.

Table 1 summarizes the changes. Simultaneous block updates remain outside our general product theorem because their linearized gains need not decompose into sequential objective changes.

Contributions. We establish the exact identity equation 1, the sharp charge equation 2, and a matching construction. We derive a variation-adaptive regret theorem for vanilla RM+. We prove constant regret and summable activation gaps in alternating exact potential games, a certified $O(\epsilon^{-2})$ lazy rate, and an $O(\delta^{-2}\epsilon^{-2})$ ordinary cyclic rate. For smooth nonconvex optimization, we prove $O(\epsilon^{-2})$ KKT rates both on one simplex and, through a new low-state loss budget, on arbitrary products under cyclic block updates. The proofs are deterministic and preserve the parameter-free, scale-invariant update.

Setting	Previous guarantee	This work
Favorable online utilities	$O(\sqrt{mT})$	$O(\sqrt{m}\,V_T)$
Alternating exact potential game	$O(\sqrt{mT})$ regret	$O(\sqrt{m}\,\Delta_\Phi)$ regret
ϵ-lazy potential play	$O(\epsilon^{-4})$ sweeps	$O(\epsilon^{-2})$ sweeps
Ordinary cyclic potential play	$O(\epsilon^{-4})$ sweeps	$O(\delta^{-2}\epsilon^{-2})$ sweeps
Smooth objective, one simplex	$O(\epsilon^{-4})$ iterations	$O(\epsilon^{-2})$ iterations
Smooth product, large initialization	$O(\epsilon^{-4})$ cycles	$O(\epsilon^{-2})$ cycles
Smooth product, zero initialization	$O(\epsilon^{-8})$ cycles	$O_\delta(\epsilon^{-2})$ cycles

Table 1: Consequences for unmodified RM+. Here V_T is centered temporal variation, Δ_Φ is potential range, and δ is the first-positive-state quantity defined below. The displayed rates suppress dimension, range, smoothness, and initialization constants. Previous optimization and potential-game rates are from Anagnostides et al. (2026b).

Related work. RM and RM+ arise from Blackwell approachability and can be viewed as FTRL and Euclidean OMD on the positive orthant (Hart & Mas-Colell, 2000; Farina et al., 2021). Predictive RM+ can be unstable, motivating restarted, clipped, smooth, and extra-gradient variants (Farina et al., 2023; Cai et al., 2025; Meng et al., 2025). Zhang et al. (2025) obtain constant regret under a weighted Minty condition for their modified extra-gradient IREG-PRM+ algorithm. Meng et al. (2026) modify smooth predictive RM+ through adaptive regret domains. In contrast, our results concern vanilla nonpredictive RM+ and use favorable realized forward gain rather than an RVU inequality. D'Orazio & Huang (2021) give prediction-sensitive Lagrangian-hedging bounds, including a path-length result for an optimistic RM+ variant on a fixed smooth loss. Variation-dependent regret more broadly has a long history in online learning (Hazan & Kale, 2010; Chiang et al., 2012). Those results design regularized-leader or modified gradient and multiplicative-weight methods around gradual variation. Our centered temporal-variation result changes neither RM+ nor its feedback.

Near-constant or constant regret is known for other dynamics under different structures. These include paced optimistic FTRL in general games, adaptive scale-invariant FTRL in zero-sum games, extrapolated FTRL in harmonic games, and optimistic fictitious play in two-action zero-sum games (Soleymani et al., 2025; Tsuchiya et al., 2026; Legacci et al., 2024; Lazarsfeld et al., 2025). These algorithms use optimism, regularization, adaptive pacing, or best responses. Conversely, Anagnostides et al. (2026a) prove exponential convergence lower bounds for broad permutation-invariant FTRL dynamics in two-player potential games, and doubly exponential bounds for fictitious play with many players. RM+'s orthant truncation creates the different self-bounding mechanism studied here. Recent fast statistical rates for KL-regularized α-potential games concern offline estimation under reference-anchored coverage, not iteration complexity of unmodified RM+ (Chen & Zhang, 2026). Discounted regret matching deliberately bounds old regret through forgetting (Brown & Sandholm, 2019). We show that no forgetting is needed in exact potential games because forward gain itself pays for state growth. Earlier potential-game dynamics either change regret aggregation or analyze different learning rules (Marden et al., 2007; Kleinberg et al., 2009). The closest work is Anagnostides et al. (2026b), whose one-step inequality and ascent region we sharpen and close.

2 RM+ AND ITS EXACT CONSERVATION LAW

For $m \geq 2$, write $\Delta_m = \{x \in \mathbb{R}^m_+ : \langle x, \mathbf{1}\rangle = 1\}$. RM+ begins with $r_0 \in \mathbb{R}^m_+$. If $r_{t-1} \neq 0$, it plays $x_t = r_{t-1}/\|r_{t-1}\|_1$. At the origin it plays an arbitrary fallback. After observing $u_t \in \mathbb{R}^m$, it sets

$$g_t = u_t - \langle x_t, u_t\rangle \mathbf{1}, \qquad r_t = [r_{t-1} + g_t]_+ . \tag{6}$$

The next strategy is $x_{t+1} = r_t/\|r_t\|_1$ if $r_t \neq 0$. If $r_t = 0$, retain $x_{t+1} = x_t$. Let $s_t = \|r_t\|_1$, $R_t = \|r_t\|_2$, $q_t = r_t - r_{t-1}$, and

$$D_t = \langle x_{t+1} - x_t, u_t\rangle = \langle x_{t+1}, g_t\rangle. \tag{7}$$

The second equality uses $\langle x_t, g_t\rangle = 0$. Adding an arbitrary multiple of $\mathbf{1}$ to u_t changes none of these quantities.

Lemma 1 (Exact accounting)**.** *Every RM+ update satisfies*

$$2s_t D_t = R_t^2 - R_{t-1}^2 + \|q_t\|_2^2. \tag{8}$$

Moreover, $R_t \geq R_{t-1}$*,* $D_t \geq 0$*, and*

$$D_t \geq \frac{\|q_t\|_2^2}{s_t} \tag{9}$$

whenever $s_t > 0$*.*

Proof. On a coordinate where r_t is positive, $q_t = g_t$. A coordinate where r_t is zero contributes nothing to either $\langle r_t, q_t\rangle$ or $\langle r_t, g_t\rangle$. Hence $s_t D_t = \langle r_t, g_t\rangle = \langle r_t, q_t\rangle$. Expanding $r_{t-1} = r_t - q_t$ proves equation 8. Also $\langle r_{t-1}, q_t\rangle \geq \langle r_{t-1}, g_t\rangle = 0$, where truncation can only make the first inner product larger. Thus $R_t^2 - R_{t-1}^2 = 2\langle r_{t-1}, q_t\rangle + \|q_t\|_2^2 \geq \|q_t\|_2^2$. The remaining claims follow from equation 8. □

The earlier one-step analysis keeps equation 9. The equality reveals that forward gain also pays for norm growth. A direct reverse-triangle argument gives $R_t - R_{t-1} \leq (s_t/R_t)D_t \leq \sqrt{m}D_t$. The next result improves the dimension factor and is sharp.

Theorem 2 (Sharp forward-gain charge)**.** *For every RM+ update,*

$$R_t - R_{t-1} \leq \sqrt{m-1}\, D_t. \tag{10}$$

Consequently, for every horizon T*,*

$$\mathrm{Reg}_T := \max_{a\in[m]} \sum_{t=1}^{T} g_t[a] \leq \|r_T\|_\infty \leq R_T \leq R_0 + \sqrt{m-1}\sum_{t=1}^{T} D_t. \tag{11}$$

The constant $\sqrt{m-1}$ *is optimal.*

Proof. Only equation 10 needs proof. Suppose $R_t > R_{t-1}$ and normalize $y = r_t/R_t$, $z = r_{t-1}/R_t$, and $\rho = \|z\|_2 < 1$. Since $s_t D_t = \langle r_t, r_t - r_{t-1}\rangle$,

$$\frac{R_t - R_{t-1}}{D_t} = \frac{\|y\|_1(1-\rho)}{1 - \langle y, z\rangle}. \tag{12}$$

If y has a zero coordinate, then $\langle y, z\rangle \leq \rho$ and $|\operatorname{supp}(y)| \leq m-1$, so the ratio is at most $\|y\|_1 \leq \sqrt{m-1}$. If y is strictly positive, no coordinate was clipped. Thus $q_t = g_t$ and $\langle z, y - z\rangle = 0$, which gives $\langle y, z\rangle = \rho^2$. With $a = \min_j y[j]$, positivity gives $\rho \geq a$, while Cauchy–Schwarz gives $\|y\|_1 \leq a + \sqrt{m-1}\sqrt{1-a^2} \leq \sqrt{m-1}(1+a)$. The ratio in equation 12 is therefore $\|y\|_1/(1+\rho) \leq \sqrt{m-1}$. The zero-state edge case is detailed in Appendix A.

Clipping implies $r_T[a] \geq r_0[a] + \sum_{t=1}^{T} g_t[a]$, which proves the regret inequalities after summing equation 10. For sharpness, take $r_0 = 0$, fallback $x_1 = e_1$, and $u_1 = (0, 1, \ldots, 1)$. Then $r_1 = (0, 1, \ldots, 1)$, $D_1 = 1$, and $R_1 - R_0 = \sqrt{m-1}$. □

Theorem 2 is useful whenever a scalar process controls $\sum_t D_t$. We first record a purely online consequence.

Corollary 3 (Centered temporal variation)**.** *Let* $v_t = u_t - c_t\mathbf{1}$*, where* $c_t = (\max_a u_t[a] + \min_a u_t[a])/2$*, and define*

$$V_T = \|v_1\|_\infty + \|v_T\|_\infty + \sum_{t=2}^{T} \|v_t - v_{t-1}\|_\infty. \tag{13}$$

With $r_0 = 0$*, vanilla RM+ satisfies*

$$\mathrm{Reg}_T \leq R_T \leq \sqrt{m-1}\, V_T. \tag{14}$$

This holds simultaneously with the standard adversarial bound $\mathrm{Reg}_T \leq (\sum_t \|g_t\|_2^2)^{1/2}$*. The state-norm coefficient in equation 14 is sharp, even at* $T = 1$*.*

Proof. Shift invariance and summation by parts give

$$\sum_{t=1}^{T} D_t = \sum_{t=1}^{T} \langle x_{t+1} - x_t, v_t \rangle$$

$$= \langle x_{T+1}, v_T \rangle - \langle x_1, v_1 \rangle + \sum_{t=2}^{T} \langle x_t, v_{t-1} - v_t \rangle \le V_T. \tag{15}$$

Apply Theorem 2. The sharp one-step example in its proof has $V_1 = 1$ and makes the state-norm inequality in equation 14 an equality. □

3 FAST SMOOTH OPTIMIZATION OVER SIMPLICES

3.1 ONE SIMPLEX: A δ-FREE BOUND

Let $f : \Delta_m \to \mathbb{R}$ be differentiable and L-smooth in Euclidean norm. Define its range $\Delta_f = \max_x f(x) - \min_x f(x)$ and the simplex KKT gap

$$\mathrm{Gap}(x) = \max_{y \in \Delta_m} \langle y - x, \nabla f(x) \rangle = \max_a \nabla_a f(x) - \langle x, \nabla f(x) \rangle. \tag{16}$$

Assume the gradient coordinate range is at most G:

$$\max_{x \in \Delta_m} \max_{a,b} |\nabla_a f(x) - \nabla_b f(x)| \le G. \tag{17}$$

Run RM+ from $r_0 = 0$ with utilities $u_t = \nabla f(x_t)$. There is no step size. Scaling f by a positive constant scales the state but leaves every strategy unchanged.

Set

$$H = \max\{2mG, 9mL\}, \qquad M = H + 2\sqrt{m-1}\,\Delta_f, \qquad C = H^2 + 2\sqrt{m}\,M\Delta_f. \tag{18}$$

Theorem 4 (Uniform regret and an ϵ^{-2} KKT rate)**.** *For every $T \ge 1$, parameter-free RM+ satisfies*

$$\mathrm{Reg}_T \le R_T \le M \tag{19}$$

and

$$\sum_{t=1}^{T} \mathrm{Gap}(x_t)^2 \le C. \tag{20}$$

In particular,

$$\min_{1 \le t \le T} \mathrm{Gap}(x_t) \le \sqrt{C/T}, \qquad T \ge C/\epsilon^2 \tag{21}$$

suffices to reach an ϵ-KKT point.

Proof. We isolate the step that closes the previous analysis. If the new state satisfies $R_t \ge H$, then

$$f(x_{t+1}) - f(x_t) \ge \frac{1}{2} D_t. \tag{22}$$

Indeed, smoothness gives a lower bound by $D_t - (L/2)\|x_{t+1} - x_t\|_2^2$. Since $\|q_t\|_\infty \le G$, the conditions $s_t \ge 2mG$ and $s_t \ge 9mL$ imply

$$\frac{L}{2}\|x_{t+1} - x_t\|_2^2 \le \frac{\|q_t\|_2^2}{2s_t} \le \frac{1}{2} D_t. \tag{23}$$

Appendix B gives the normalization calculation.

The state norm is nondecreasing by Lemma 1. If it never reaches H, then equation 19 is immediate. Otherwise let τ be the first index with $R_\tau \ge H$. Equation equation 22 holds from the crossing update onward, so $\sum_{t=\tau}^{T} D_t \le 2\Delta_f$. The sharp charge gives

$$R_T \le R_{\tau-1} + \sqrt{m-1} \sum_{t=\tau}^{T} D_t < H + 2\sqrt{m-1}\Delta_f = M. \tag{24}$$

Before the crossing, the standard norm monotonicity inequality gives $R_t^2 - R_{t-1}^2 \geq \mathrm{Gap}(x_t)^2$. Hence the pre-threshold sum of squared gaps is below H^2. From the crossing onward, Lemma 1 and equation 23 also give

$$f(x_{t+1}) - f(x_t) \geq \frac{\|q_t\|_2^2}{2s_t} \geq \frac{\mathrm{Gap}(x_t)^2}{2\sqrt{m}M}. \tag{25}$$

Summing and using the range of f proves equation 20. Equation equation 21 follows by averaging. □

The ϵ^{-2} dependence matches the classical first-order scale for general smooth nonconvex optimization (Carmon et al., 2021). It improves the ϵ^{-4} rate of Anagnostides et al. (2026b) for the identical parameter-free trajectory. Their reduction also shows that Theorem 4 improves simultaneous RM+ in symmetric potential games when all players share the same initialization. We do not extend this corollary to arbitrary simultaneous play.

3.2 Products of simplices: cyclic block RM+

Let $\mathcal{X} = \prod_{i=1}^n \Delta_{m_i}$, let $m = \max_i m_i$, and let $f : \mathcal{X} \to \mathbb{R}$ be L-smooth in Euclidean norm, with range Δ_f. Assume every block-gradient coordinate range is at most G:

$$\max_{x \in \mathcal{X}} \max_{i,a,b} |\nabla_i f(x)[a] - \nabla_i f(x)[b]| \leq G. \tag{26}$$

Cyclic block RM+ visits $i = 1, \ldots, n$ and gives block i the current utility $\nabla_i f(x)$. Other blocks remain fixed. Let x^k be the end of cycle k, and define the product KKT gap

$$\mathrm{Gap}_\Pi(x) = \max_{y \in \mathcal{X}} \langle y - x, \nabla f(x) \rangle = \sum_{i=1}^n \left(\max_a \nabla_i f(x)[a] - \langle x_i, \nabla_i f(x) \rangle \right). \tag{27}$$

Allow arbitrary nonnegative initial states, with $x_i^0 = r_i^0 / \|r_i^0\|_1$ whenever $r_i^0 \neq 0$. Define δ as the minimum of one, every positive initial state norm, and the first positive state norm of each zero-initialized block that ever leaves the origin. Blocks that remain at the origin are omitted. Put

$$H_i = \max\{2m_i G, 9m_i L\}, \qquad A = \sum_i m_i H_i^2, \qquad E = \frac{2LA}{\delta^2}, \tag{28}$$

$$J_i = \max\{\|r_i^0\|_2, H_i\}, \qquad M_i = J_i + 2\sqrt{m_i - 1}(\Delta_f + E), \tag{29}$$

$$W = \max_i \sqrt{m_i} M_i, \qquad Q = \sum_i H_i^2 + 2W(\Delta_f + E), \qquad C_\Pi = \sqrt{n} + \frac{2n\sqrt{2m}L}{\delta}. \tag{30}$$

Theorem 5 (Cyclic product-simplex optimization)**.** *At every activation, block i has $\|r_i\|_2 \leq M_i$. If $q_{i,k}$ is its state increment in cycle k and $S_k = \sum_i \|q_{i,k}\|_2^2$, then*

$$\sum_{k \geq 1} S_k \leq Q, \qquad \min_{1 \leq k \leq K} \mathrm{Gap}_\Pi(x^k) \leq C_\Pi \sqrt{Q/K}. \tag{31}$$

Consequently $K \geq C_\Pi^2 Q / \epsilon^2$ suffices for an ϵ-KKT cycle endpoint, and $\mathrm{Gap}_\Pi(x^k) \to 0$ along the actual cycle-end sequence.

The proof resolves the interaction between blocks rather than assuming it away. Call an update of block i low if its new state norm is below H_i. Norm monotonicity makes low updates an initial prefix and gives $\sum_{\text{low}} \|q_i\|_2^2 \leq H_i^2$. The normalization map and the definition of δ give $\|x_i^+ - x_i\|_2 \leq 2\sqrt{m_i}\|q_i\|_2/\delta$, including transitions from the origin. Smoothness therefore limits total objective decrease over all low updates to E. Every remaining high update increases f by at least $D/2$, so its total forward gain is at most $2(\Delta_f + E)$. The sharp charge bounds every state by M_i, and exact accounting gives $\|q_i\|_2^2 \leq \|r_i^+\|_1 D$. This proves the squared-path budget Q.

It remains to certify all blocks at one profile. Immediately after block i updates, its gap against the gradient used for that update is at most $\|q_{i,k}\|_2$. Later blocks in the same cycle move the profile by

at most $2\sqrt{mS_k}/\delta$. Smoothness and the $\sqrt{2}$ Lipschitz constant of a simplex gap with respect to its utility vector yield

$$\mathrm{Gap}_\Pi(x^k) \le C_\Pi \sqrt{S_k}. \tag{32}$$

Averaging the squared-path budget proves equation 31. Summability gives convergence. Appendix C supplies every step and the origin cases. Unlike earlier product bounds, Theorem 5 neither uses lazy updates nor prescribes a large initialization. The trajectory-dependent δ affects constants but not the algorithm or the ϵ exponent.

4 Constant regret and fast Nash convergence in potential games

Consider a finite n-player exact potential game (Monderer & Shapley, 1996). Player i has $m_i \ge 2$ actions, expected utility U_i, and utility vector $u_i(x_{-i}) \in \mathbb{R}^{m_i}$. The multilinear extension Φ of the potential satisfies, for every unilateral mixed deviation,

$$\Phi(x_i', x_{-i}) - \Phi(x_i, x_{-i}) = \langle x_i' - x_i, u_i(x_{-i}) \rangle. \tag{33}$$

Let $\Delta_\Phi = \max_x \Phi(x) - \min_x \Phi(x)$. Under alternating RM+, players activate sequentially in a fixed cyclic order and observe their current full-information utility vector.

Theorem 6 (Constant regret in exact potential games)**.** *For any nonnegative initial states and every horizon, alternating RM+ satisfies for each player i*

$$\mathrm{Reg}_{i,T} \le \|r_{i,T}\|_2 \le \|r_{i,0}\|_2 + \sqrt{m_i - 1}\,\Delta_\Phi, \tag{34}$$

$$\|r_{i,T}\|_1 \le \sqrt{m_i}\|r_{i,0}\|_2 + \sqrt{m_i(m_i-1)}\,\Delta_\Phi =: B_i. \tag{35}$$

Thus the regret of standard zero-initialized RM+ is uniformly bounded by $\sqrt{m_i - 1}\Delta_\Phi$.

Proof. At every activation, equation 33 identifies the player's D_t with the realized potential increase. Lemma 1 makes every such increase nonnegative. Across all players and all cycles, these increases telescope to at most Δ_Φ. Apply Theorem 2 to the subsequence of activations of player i, then use $\|r\|_1 \le \sqrt{m_i}\|r\|_2$. □

This rules out $\Omega(\sqrt{T})$ regret for alternating RM+ and resolves the open question posed by Anagnostides et al. (2026b). The bound holds pathwise and at every horizon. It also converts their regret-parameterized Nash analysis into the faster rate suggested there. For a profile x, let $\mathrm{BRGap}_i(x) = \max_{y_i \in \Delta_{m_i}} U_i(y_i, x_{-i}) - U_i(x)$. The same accounting gives a finite gap budget along ordinary play.

Corollary 7 (Summable activation gaps)**.** *At activation t, let i_t be the active player, let x^t be the profile then observed, and set $\gamma_t = \mathrm{BRGap}_{i_t}(x^t)$. Alternating RM+ satisfies*

$$\sum_{1 \le t \le T:\, B_{i_t} > 0} \frac{\gamma_t^2}{B_{i_t}} \le \Delta_\Phi. \tag{36}$$

Players with $B_i = 0$ have zero activation gap. Thus $\sum_t \gamma_t^2 \le B\Delta_\Phi$, the best of the first T activation gaps is at most $\sqrt{B\Delta_\Phi/T}$, and $\gamma_t \to 0$.

Proof. The coordinate of q_t attaining the best-response gap equals γ_t, so $\|q_t\|_2 \ge \gamma_t$. When $B_{i_t} > 0$, Lemma 1 and $\|r_{i_t,t}\|_1 \le B_{i_t}$ give $D_t \ge \gamma_t^2 / B_{i_t}$. If $B_i = 0$, then the state stays zero and the utility vector is constant across that player's actions, so $\gamma_t = 0$. Sum and telescope the potential increases. □

The gaps in Corollary 7 are evaluated at different asynchronous profiles. They do not by themselves certify one joint approximate Nash profile. Lazy play supplies such a certificate without an instance-specific constant.

Fix $\epsilon > 0$. In ϵ-lazy alternating RM+, a player updates only when its current best-response gap exceeds ϵ. The procedure stops after a full cycle with no update.

Theorem 8 (Fast lazy Nash computation)**.** *Let $B = \max_i B_i$. An ϵ-lazy alternating RM+ run reaches an ϵ-Nash equilibrium within*

$$1 + \frac{B\Delta_\Phi}{\epsilon^2} \tag{37}$$

full cycles. With zero initialization and $m = \max_i m_i$, this is at most

$$1 + \frac{\sqrt{m(m-1)}\Delta_\Phi^2}{\epsilon^2}. \tag{38}$$

Proof. At an actual update, the positive coordinate of g_t attaining the best-response gap is unchanged by clipping, so $\|q_t\|_2 \geq \mathrm{BRGap}_i(x)$. Lemma 1 and equation 35 give

$$\Phi(x^+) - \Phi(x) = D_t \geq \frac{\|q_t\|_2^2}{\|r_{i,t}\|_1} > \frac{\epsilon^2}{B}. \tag{39}$$

Every nonterminal cycle contains such an update. More than $B\Delta_\Phi/\epsilon^2$ nonterminal cycles would exceed the potential range. A no-update cycle certifies all players' gaps at the same profile. □

The threshold in Theorem 8 must be supplied. A target-free certified version uses phases $\epsilon_k = 2^{-k}$. Run lazy cycles until one full cycle has no update, save that profile, halve the threshold, and continue without resetting the states. Through the first ϵ_k certificate, the total number of nonterminal cycles is at most

$$B\Delta_\Phi \sum_{j=0}^{k} \epsilon_j^{-2} \leq \frac{4B\Delta_\Phi}{3\epsilon_k^2}, \tag{40}$$

plus one certification cycle per phase. The algorithm never needs the eventual target and always retains its last certified profile.

Ordinary non-lazy cyclic RM+ is already target-free. For the explicit rate below, scale all utilities and the exact potential together so that the range across a player's actions is at most one at every pure opponent-action profile. Substituting the constant state bound into the two-consecutive-small-movements argument of Anagnostides et al. (2026b) changes its exponent from four to two. Appendix D proves the following explicit statement.

Theorem 9 (Ordinary cyclic RM+)**.** *Suppose the range across own actions is at most one at every pure opponent-action profile, and let $0 < \epsilon \leq 1$. Let $m = \max_i m_i$ and $B = \max_i B_i$. Define δ as the minimum of one, every positive initial state norm, and every first positive state norm of a player initialized at zero, ignoring players whose states remain zero forever. Ordinary cyclic RM+ reaches an ϵ-Nash profile within*

$$2\left\lceil \frac{25B\Delta_\Phi n^2 m}{\delta^2\epsilon^2} \right\rceil \tag{41}$$

cycles. Moreover, the Nash gap of the actual cycle-end iterates tends to zero.

The δ dependence comes only from converting state motion into strategy motion. The phased lazy procedure avoids it and has an observable stopping certificate. The ordinary result instead analyzes the unchanged always-update dynamics.

5 Discussion and limitations

The conservation law explains why the adversarial regret envelope can be loose on endogenous paths. Forward gain both measures normalized-strategy improvement and finances unnormalized-state growth. A bounded objective therefore supplies an implicit lower bound on RM+'s effective learning rate.

The product theorem requires cyclic updates because simultaneous gains need not decompose into objective changes. Its constant depends on the smallest first-positive state norm δ, which can be arbitrarily small, although the one-simplex theorem is δ-free. RM without per-round truncation also lacks the orthant identity. These boundaries agree with known instability outside potential structure (Farina et al., 2023; Cai et al., 2025; Anagnostides et al., 2026b).

Appendix E extends the charge to approximate potentials and perturbed ascent paths. Stochastic and bandit feedback remain open for RM+ because noise can make realized forward charge signed. Calibrated stochastic Frank–Wolfe methods provide complementary equilibrium and regret guarantees under bandit feedback (Dong et al., 2024).

REFERENCES

Ioannis Anagnostides, Ioannis Panageas, Nikolas Patris, and Tuomas Sandholm. (Doubly) exponential lower bounds for follow the regularized leader in potential games. *arXiv preprint arXiv:2601.23248*, 2026a. URL https://arxiv.org/abs/2601.23248.

Ioannis Anagnostides, Emanuel Tewolde, Brian Hu Zhang, Ioannis Panageas, Vincent Conitzer, and Tuomas Sandholm. Convergence of regret matching in potential games and constrained optimization. In *International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=EOV1q1U23N.

Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1829–1836, 2019. doi: 10.1609/aaai.v33i01.33011829.

Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Weiqiang Zheng. Last-iterate convergence properties of regret-matching algorithms in games. In *International Conference on Learning Representations*, 2025.

Yair Carmon, John C. Duchi, Oliver Hinder, and Aaron Sidford. Lower bounds for finding stationary points II: First-order methods. *Mathematical Programming*, 185:315–355, 2021. doi: 10.1007/s10107-019-01431-x.

Claire Chen and Yuheng Zhang. Fast rates in α-potential games via regularized mirror descent. *arXiv preprint arXiv:2605.00268*, 2026. URL https://arxiv.org/abs/2605.00268.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pp. 6.1–6.20, 2012. URL https://proceedings.mlr.press/v23/chiang12.html.

Jing Dong, Baoxiang Wang, and Yaoliang Yu. Convergence to Nash equilibrium and no-regret guarantee in (Markov) potential games. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2044–2052, 2024. URL https://proceedings.mlr.press/v238/dong24a.html.

Ryan D'Orazio and Ruitong Huang. Optimistic and adaptive lagrangian hedging. *arXiv preprint arXiv:2101.09603*, 2021.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive Blackwell approachability: Connecting regret matching and mirror descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5363–5371, 2021. doi: 10.1609/aaai.v35i6.16676.

Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, and Haipeng Luo. Regret matching+: (in)stability and fast convergence in games. In *Advances in Neural Information Processing Systems*, volume 36, pp. 61546–61572, 2023.

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.

Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine Learning*, 80(2–3):165–188, 2010. doi: 10.1007/s10994-010-5175-x.

Robert Kleinberg, Georgios Piliouras, and Éva Tardos. Multiplicative updates outperform generic no-regret learning in congestion games. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, pp. 533–542, 2009. doi: 10.1145/1536414.1536487.

John Lazarsfeld, Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Optimism without regularization: Constant regret in zero-sum games. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Davide Legacci, Panayotis Mertikopoulos, Christos Papadimitriou, Georgios Piliouras, and Bary Pradelski. No-regret learning in harmonic games: Extrapolation in the face of conflicting interests. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3931.

Jason R. Marden, Gürdal Arslan, and Jeff S. Shamma. Regret based dynamics: Convergence in weakly acyclic games. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 194–201, 2007. doi: 10.1145/1329125.1329175.

Linjian Meng, Youzhi Zhang, Zhenxing Ge, Tianyu Ding, Shangdong Yang, Zheng Xu, Wenbin Li, and Yang Gao. Last-iterate convergence of smooth regret matching+ variants in learning nash equilibria. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-0644.

Linjian Meng, Youzhi Zhang, Shangdong Yang, Wenbin Li, Tianyu Ding, and Yang Gao. A faster parameter-free regret matching algorithm. In *International Conference on Learning Representations*, 2026. URL `https://openreview.net/forum?id=JLllvi7dsg`.

Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1): 124–143, 1996. doi: 10.1006/game.1996.0044.

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

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

Emanuel Tewolde, Brian Hu Zhang, Ioannis Anagnostides, Tuomas Sandholm, and Vincent Conitzer. Decision making under imperfect recall: Algorithms and benchmarks. *arXiv preprint arXiv:2602.15252*, 2026.

Taira Tsuchiya, Haipeng Luo, and Shinji Ito. Scale-invariant fast convergence in games. *arXiv preprint arXiv:2602.11857*, 2026.

Brian Hu Zhang, Ioannis Anagnostides, and Tuomas Sandholm. Scale-invariant regret matching and online learning with optimal convergence: Bridging theory and practice in zero-sum games. *arXiv preprint arXiv:2510.04407*, 2025.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

A Full details for the conservation law

We first justify all zero-state cases in Lemma 1. If $r_{t-1} \neq 0$, then $\langle r_{t-1}, q_t \rangle \geq 0$, so $R_t \geq R_{t-1} > 0$. Thus a transition into the origin is impossible. If both states are zero, the two sides of equation 8 vanish under the convention $x_{t+1} = x_t$. A transition from the origin to a nonzero state has $r_t = [g_t]_+$. If r_t had full support, then every coordinate of g_t would be positive, contradicting $\langle x_t, g_t \rangle = 0$. Therefore its support has size at most $m - 1$, which is precisely the first case in the proof of Theorem 2.

For completeness, we supply the elementary inequality used in the full-support case. Let $y \in \mathbb{R}^m_+$ satisfy $\|y\|_2 = 1$ and put $a = \min_j y[j]$. Choose a coordinate attaining a. Then

$$\|y\|_1 \leq a + \sqrt{m-1}\sqrt{1-a^2} \leq a + \sqrt{m-1} \leq \sqrt{m-1}(1+a). \tag{42}$$

If y is the normalized new state and z is the old state divided by the new norm, the no-clipping relation $\langle z, y - z\rangle = 0$ yields $\langle y, z\rangle = \|z\|_2^2 = \rho^2$. Since all coordinates of y are at least a,

$$\rho^2 = \langle y, z\rangle \geq a\|z\|_1 \geq a\rho. \tag{43}$$

When $\rho > 0$, this gives $\rho \geq a$. The case $\rho = 0$ cannot have full support, as shown above. This completes the proof without genericity assumptions.

The standard adversarial inequality quoted in Corollary 3 follows from projection onto the positive orthant:

$$R_t^2 = \| [r_{t-1} + g_t]_+ \|_2^2 \leq \|r_{t-1} + g_t\|_2^2 = R_{t-1}^2 + \|g_t\|_2^2, \tag{44}$$

where $\langle r_{t-1}, g_t\rangle = 0$. Telescoping and $\mathrm{Reg}_T \leq R_T$ give the claim.

B Auxiliary smooth-optimization bounds

We prove equation 23. For nonzero $r, r' \in \mathbb{R}_+^m$ with sums s, s', normalizing and adding an intermediate term gives

$$\left\| \frac{r}{s} - \frac{r'}{s'} \right\|_1 \leq \|r - r'\|_1 \left(\frac{1}{s} + \frac{1}{s'} \right). \tag{45}$$

Under equation 17, every coordinate of g_t lies in $[-G, G]$. A clipped decrement $q_t[a] = -r_{t-1}[a]$ can occur only when $r_{t-1}[a] \leq G$, so $\|q_t\|_\infty \leq G$. Hence $|s_t - s_{t-1}| \leq mG$. A transition from the origin has $R_t = \| [g_t]_+ \|_2 \leq \sqrt{m}G < H$, so any update with $R_t \geq H$ has $s_{t-1} > 0$. If $s_t \geq 2mG$, then $s_{t-1} \geq s_t/2$. Applying equation 45, $\|q_t\|_1 \leq \sqrt{m}\|q_t\|_2$, and $\|\cdot\|_2 \leq \|\cdot\|_1$ gives

$$\|x_{t+1} - x_t\|_2^2 \leq \frac{9m}{s_t^2}\|q_t\|_2^2. \tag{46}$$

If also $s_t \geq 9mL$, then

$$\frac{L}{2}\|x_{t+1} - x_t\|_2^2 \leq \frac{\|q_t\|_2^2}{2s_t} \leq \frac{D_t}{2}, \tag{47}$$

where the last inequality is Lemma 1. Since $s_t \geq R_t$, the condition $R_t \geq H$ implies both required lower bounds.

It remains to justify the pre-threshold KKT charge. The positive coordinate of g_t that attains equation 16 is never clipped, so $\|q_t\|_2 \geq \mathrm{Gap}(x_t)$. From the proof of Lemma 1,

$$R_t^2 - R_{t-1}^2 \geq \|q_t\|_2^2 \geq \mathrm{Gap}(x_t)^2. \tag{48}$$

Summing over indices before the first crossing of H gives a total below H^2.

C Proof of the product-simplex theorem

We prove Theorem 5 for every finite prefix. All bounds then pass to the infinite path. Write $R_i = \|r_i\|_2$ and $s_i = \|r_i\|_1$ immediately before an activation, and add a superscript $+$ for the post-update quantities. Lemma 1 applies blockwise and gives

$$(R_i^+)^2 - R_i^2 \geq \|q_i\|_2^2, \qquad 2s_i^+ D_i = (R_i^+)^2 - R_i^2 + \|q_i\|_2^2. \tag{49}$$

In particular, each R_i is nondecreasing, $D_i \geq 0$, and $\|q_i\|_2^2 \leq s_i^+ D_i$.

We first account for low updates. For a fixed block, they form an initial prefix because R_i is nondecreasing. Telescoping the first inequality in equation 49 over this prefix gives

$$\sum_{\text{low updates of } i} \|q_i\|_2^2 \leq H_i^2. \tag{50}$$

Every update, low or high, also satisfies

$$\|x_i^+ - x_i\|_2 \leq \frac{2\sqrt{m_i}}{\delta}\|q_i\|_2. \tag{51}$$

If both states are positive, this follows from equation 45, $s_i, s_i^+ \geq \delta$, and $\|q_i\|_1 \leq \sqrt{m_i}\|q_i\|_2$. If an update leaves the origin, then $q_i = r_i^+$, $\|q_i\|_2 \geq \delta$, and the Euclidean diameter of the simplex is $\sqrt{2} \leq 2\sqrt{m_i}$. An update that remains at the origin does not move its retained fallback strategy. Combining equation 50 and equation 51,

$$\sum_{\text{low}} \|x_i^+ - x_i\|_2^2 \leq \frac{4}{\delta^2} \sum_i m_i H_i^2 = \frac{4A}{\delta^2}. \tag{52}$$

For any block update, smoothness gives

$$f(x^+) - f(x) \geq D_i - \frac{L}{2}\|x_i^+ - x_i\|_2^2. \tag{53}$$

Thus the sum of objective changes over low updates is at least $-E$. If an update is high, the block-wise calculation in Appendix B applies with m_i and the common bounds G, L. A transition from the origin cannot be high because $\|[g_i]_+\|_2 \leq \sqrt{m_i}G < H_i$. Therefore

$$f(x^+) - f(x) \geq D_i/2 \tag{54}$$

on every high update. Telescoping all objective changes and using the range of f yields

$$\sum_{\text{high}} D_i \leq 2(\Delta_f + E). \tag{55}$$

If block i never becomes high, its norm remains below $H_i \leq M_i$. Otherwise, immediately before its first high update its norm is at most J_i. All of its later updates are high. Applying Theorem 2 to that suffix and then equation 55 gives

$$R_i \leq J_i + \sqrt{m_i - 1} \sum_{\text{high updates of } i} D_i \leq M_i. \tag{56}$$

For high updates, equation 49 and equation 56 imply

$$\|q_i\|_2^2 \leq s_i^+ D_i \leq \sqrt{m_i} M_i D_i \leq W D_i. \tag{57}$$

Combining this with equation 50 and equation 55 proves

$$\sum_{k \geq 1} S_k = \sum_{i,k} \|q_{i,k}\|_2^2 \leq \sum_i H_i^2 + 2W(\Delta_f + E) = Q. \tag{58}$$

We finally transfer activation gaps to a cycle endpoint. Let $z_{i,k}$ and $y_{i,k}$ be the profiles immediately before and after block i updates in cycle k. Its activation gap $\gamma_{i,k}$ is at most $\|q_{i,k}\|_2$: a coordinate attaining the nonnegative maximum of g_i is not clipped and has state increment equal to $\gamma_{i,k}$. Moreover,

$$\max_a \nabla_i f(z_{i,k})[a] - \langle (y_{i,k})_i, \nabla_i f(z_{i,k}) \rangle = \gamma_{i,k} - D_{i,k} \leq \|q_{i,k}\|_2. \tag{59}$$

For fixed $p \in \Delta_{m_i}$, the function $u \mapsto \max_a u[a] - \langle p, u \rangle$ is $\sqrt{2}$-Lipschitz in Euclidean norm because $\|e_a - p\|_2 \leq \sqrt{2}$. Block i does not move again before the endpoint x^k. By equation 51 and the disjointness of block coordinates,

$$\|x^k - y_{i,k}\|_2^2 \leq \frac{4m}{\delta^2} \sum_{j > i} \|q_{j,k}\|_2^2 \leq \frac{4m}{\delta^2} S_k. \tag{60}$$

Smoothness and equation 59 therefore show that the block-i endpoint gap is at most

$$\|q_{i,k}\|_2 + \frac{2\sqrt{2m}L}{\delta}\sqrt{S_k}. \tag{61}$$

Summing over blocks and using $\sum_i \|q_{i,k}\|_2 \leq \sqrt{nS_k}$ proves equation 32. Among the first K cycles, equation 58 supplies one with $S_k \leq Q/K$, proving the finite-time claim. It also implies $S_k \to 0$, which proves convergence of the actual cycle-end gaps.

D ORDINARY ALTERNATING RM+

We prove Theorem 9 by making the substitution into the argument of Anagnostides et al. (2026b) explicit. Index full cycles by t. Let $q_{i,t}$ be player i's state increment at its activation during cycle t. Lemma 1, equation 33, and $\|r_{i,t}\|_1 \le B$ give

$$\sum_{t\ge 1}\sum_{i=1}^{n}\|q_{i,t}\|_2^2 \le B\sum_{t,i} D_{i,t} \le B\Delta_\Phi. \tag{62}$$

Among the first $2K$ cycles, one disjoint adjacent pair $2k-1, 2k$ therefore satisfies

$$\sum_i \left(\|q_{i,2k-1}\|_2^2 + \|q_{i,2k}\|_2^2\right) \le \frac{B\Delta_\Phi}{K}. \tag{63}$$

Choose $K = \lceil 25B\Delta_\Phi n^2 m/(\delta^2\epsilon^2)\rceil$. Every state increment in the pair then has norm at most

$$\eta = \frac{\delta\epsilon}{5n\sqrt{m}}. \tag{64}$$

At a player's activation, its best-response gap against the utility vector observed then is at most $\|q_{i,t}\|_\infty \le \eta$. A positive state norm is at least δ, so equation 45 yields a within-activation strategy movement at most $2\sqrt{m}\eta/\delta$ in ℓ_1. A player at the origin either has $q_{i,t} = 0$ and does not move, or its first positive update has norm at least δ. The latter is impossible in the selected pair because $\eta < \delta$. Thus the same movement bound covers every activation in the pair.

The player's own update cannot increase its gap against the opponents observed at activation because $D_{i,t} \ge 0$. For a fixed own mixed action, the best-response gap changes by at most twice the ℓ_1 change in the opponents' product distribution because every pure payoff difference lies in $[-1, 1]$. The opponents move by at most $2n\sqrt{m}\eta/\delta$ in total during a cycle. Multiplying this movement by the gap's factor two gives the transfer cost $4n\sqrt{m}\eta/\delta$. Therefore every player's gap at the end of the second selected cycle is at most

$$\eta + \frac{4n\sqrt{m}\eta}{\delta} \le \frac{5n\sqrt{m}\eta}{\delta} = \epsilon, \tag{65}$$

where $\delta \le 1$ by definition. This proves equation 41. The case with no positive state is already stationary.

For the asymptotic statement, equation 62 implies $q_{i,t} \to 0$ for every player. Equation equation 45 and the positive state lower bound imply that every within-cycle strategy movement tends to zero. A state that remains at the origin does not move. Each activation-time best-response gap is at most $\|q_{i,t}\|_\infty$ and also tends to zero. Multilinearity and continuity transfer these gaps to the end-of-cycle profile, so its Nash gap converges to zero.

E GAIN DOMINANCE AND APPROXIMATE POTENTIALS

The game proof uses only an upper bound on cumulative forward gain. The following form records the robust principle.

Proposition 10 (Approximate gain dominance)**.** *Suppose an RM+ utility path admits a scalar sequence Ψ_t and errors $e_t \ge 0$ such that*

$$D_t \le \Psi_t - \Psi_{t-1} + e_t \tag{66}$$

for every t. Then

$$\mathrm{Reg}_T \le R_0 + \sqrt{m-1}\left(\Psi_T - \Psi_0 + \sum_{t=1}^{T} e_t\right). \tag{67}$$

Proof. Sum the assumed inequalities and apply Theorem 2. □

For example, if unilateral utility improvement differs from the increment of an approximate potential by at most ζ_t at activation t, take $e_t = \zeta_t$. Bounded cumulative approximation error preserves constant regret. A uniform per-update error ζ yields $O(1 + \zeta T)$ regret, so average regret approaches an $O(\zeta)$ floor. This statement deliberately measures realized mixed deviations. Translating a particular pure-deviation definition of approximate potential into ζ_t may introduce an additional constant.