FROM VERIFICATION TO DISCOVERY: THE QUERY COMPLEXITY OF EQUILIBRIUM CERTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Self-play can learn approximate Nash equilibria from noisy interactions, but does finding an equilibrium require more information than checking a supplied candidate? We study this gap between discovery and verification and how it depends on payoff structure. For bounded $k \times k$ zero-sum games with unit-variance Gaussian payoff queries, verification under a constant-factor accuracy margin uses $O(k)$ queries at fixed small accuracy and confidence, while discovery can require $\Omega(k \log \log k)$. The lower bound applies to arbitrary approximate equilibria because every valid mixed output reveals hidden action labels. By contrast, every fixed-rank class has minimax discovery complexity $\Theta(k)$ in the same regime. Our fixed-rank upper bounds use response compression to learn strategic responses directly from noisy payoffs without supplied factors or anchors. For nontrivial $n \times m$ rank-one games, we obtain the exact minimax law $\Theta((n+m)\varepsilon^{-2}\log(1/\delta))$. Rank two retains linear dependence on $n+m$ and the optimal ε^{-2} accuracy exponent up to logarithmic factors. Finally, combining structured search, generic self-play, and independent verification retains these gains without knowing the rank while preserving correctness on every bounded game.

1 INTRODUCTION

Approximate Nash equilibria are pairs of mixed strategies for which neither player can gain substantially from a unilateral deviation. Motivated by self-play and policy evaluation through noisy simulated matchups (Lanctot et al., 2017; Rowland et al., 2019), we study two-player zero-sum games in which each matchup yields a noisy observation of a single unknown payoff-matrix entry (O'Donoghue et al., 2021; Maiti et al., 2023; Foster et al., 2023). This setting naturally separates verification from discovery. In *verification*, a candidate pair is supplied and the learner tests whether its deviation gains exceed the target tolerance. We use *equilibrium certification* for the end-to-end discovery task in which no candidate is supplied and the learner must return an approximate equilibrium from noisy payoff observations with prescribed confidence. Correctness requires controlling deviations over the full action sets, including actions assigned zero probability by the returned strategies.

For $k \times k$ games at fixed accuracy and confidence, we show that a supplied candidate can be verified with $O(k)$ queries under a constant-factor accuracy margin (Lemma 2.1), by combining standard adaptive best-response search (Mannor & Tsitsiklis, 2004; Even-Dar et al., 2006) with fresh payoff estimation. By contrast, existing high-probability bandit self-play gives an $O(k \log k)$ query upper bound for discovery (Bai et al., 2022). This difference in known upper bounds does not by itself imply an intrinsic information gap between finding and verifying an equilibrium. *Does discovering an approximate equilibrium necessarily require more noisy observations than verifying a supplied candidate, and when can low-rank structure eliminate this additional cost?*

We show that discovery does incur an intrinsic information cost. For bounded $k \times k$ games with Gaussian payoff queries, the worst-case expected cost is $\Omega(k \log \log k)$, even at fixed small error tolerance and fixed confidence. By contrast, the class of bounded matrices of rank at most any fixed $r \geq 1$ has minimax complexity $\Theta(k)$ in the same regime. Establishing the lower bound requires accommodating any approximate equilibrium, rather than a designated equilibrium or support, an answer-selection issue also central to pure exploration (Degenne & Koolen, 2019; Poiani et al., 2026). Our construction makes every strategy pair meeting the required guarantee reveal a hidden subset of action labels. Choosing another equilibrium cannot avoid this information cost (Figure 1).

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.

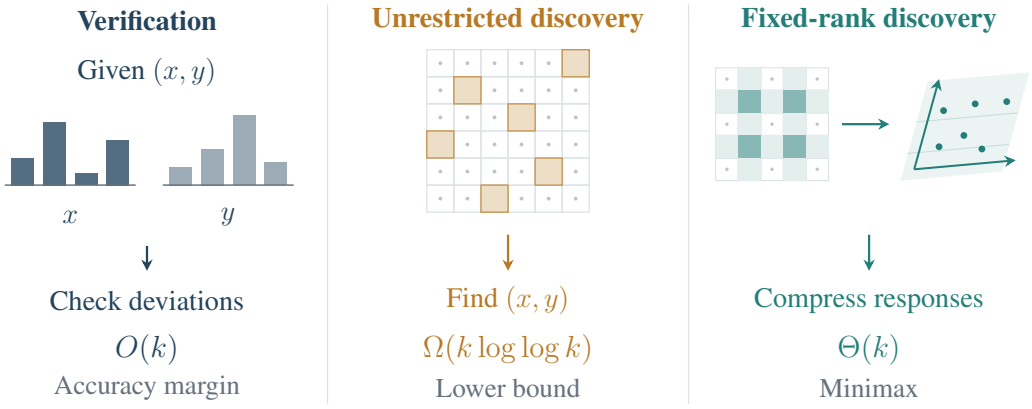

Figure 1: **Fixed rank removes the discovery overhead.** Query costs for bounded $k \times k$ games with unit-variance Gaussian payoff queries, at fixed small accuracy and confidence. Verification allows an accuracy margin. The middle rate is a worst-case expected lower bound. The right rate is minimax for each fixed rank bound. Response representations are learned without supplied factors or anchors. All drawings are schematic.

Low-rank structure has been exploited in bilinear bandits and multiagent evaluation (Jun et al., 2019; Du et al., 2021). In our setting, it restricts the diversity of strategic responses. Although payoff profiles lie in an r-dimensional space, the learner is given neither factors nor reference actions and must discover a useful representation directly from noisy payoffs. The general fixed-rank method alternates between solving a current low-dimensional surrogate and checking the resulting candidate in the original game. A reliable rejection exposes a response direction missing from the current representation. Rank bounds the number of such expansions, while bounded interpolation transfers control from the selected responses to all actions without uncontrolled amplification of estimation error. The resulting procedure uses a number of response searches controlled by rank and accuracy, each with linear dependence on the number of actions. Specialized rank-one and rank-two procedures exploit their scalar and planar geometry to obtain sharper accuracy dependence.

Main contributions. For bounded $n \times m$ matrix games with unit-variance Gaussian payoff queries, we study the worst-case expected number of observations needed to return an ε-Nash equilibrium with probability at least $1 - \delta$.

- **An intrinsic separation between discovery and verification.** For unrestricted $k \times k$ games at fixed small accuracy and fixed confidence, we prove an $\Omega(k \log \log k)$ discovery lower bound. A supplied candidate can be verified in $O(k)$ queries under a constant-factor accuracy margin, establishing an unavoidable information cost of discovery (Corollary 3.2). This bound covers every valid mixed output, regardless of which equilibrium or support the learner selects. We also identify high-confidence and balanced high-resolution regimes where the upper and lower bounds match.
- **Low rank removes the discovery overhead.** For every fixed rank bound $r \geq 1$ and fixed small accuracy ε, the minimax complexity is $\Theta_{r,\varepsilon}((n + m) \log(1/\delta))$ outside the trivial 1×1 case, matching the verification scale (Theorem 4.1). We sharpen the accuracy dependence to the exact rate $\Theta((n + m)\varepsilon^{-2} \log(1/\delta))$ at rank one, and retain linear action dependence with the optimal ε^{-2} exponent up to logarithmic factors at rank two (Theorems 4.3 and 4.5). We discover stable response representations directly from noisy payoffs, without supplied factors or anchors and without conditioning or spikiness assumptions. Reliable rejections expand the representation, while bounded interpolation controls every strategic response.
- **Globally valid adaptation to unknown structure.** A single procedure combines structured proposals, generic self-play, and independent verification without knowing the rank. It stops almost surely and returns a valid equilibrium with the prescribed confidence on every bounded game, while preserving the generic expected query bound and attaining the stated structural gains (Corollary 5.1). The structural assumption accelerates search. Correctness remains valid even when that assumption fails. It also attains the stated robustness guarantees near rank-one and shifted skew-rank-two games without a supplied perturbation radius.

Connection to the COLT 2025 open problem. Maiti (2025) asks for the optimal instance-dependent sample complexity of finding approximate Nash equilibria from noisy payoff queries. **We settle several sharp full-box minimax counterparts of this question.** Our upper and lower bounds match in the confidence-dominated regime, for every fixed smaller action-set size at fixed accuracy, and in the balanced fine-accuracy regime. At fixed accuracy and confidence, we further prove a strict separation between unrestricted and fixed-rank discovery, showing that structural assumptions can fundamentally change the information required to find an equilibrium. These results establish worst-case rates and structural phenomena that any complete instance-dependent theory must recover.

2 Problem Setup and Verification

2.1 Observation Model and Certification Objective

Let Δ_d be the probability simplex on d actions. For an unknown $A \in [-1,1]^{n\times m}$, the row player maximizes and the column player minimizes $x^\top Ay$ over $(x,y) \in \Delta_n \times \Delta_m$, with game value $V_A^\star = \max_x \min_y x^\top Ay$. Given history $\mathcal{F}_{t-1}$, the learner selects (I_t, J_t) and receives a fresh observation $X_t \mid \mathcal{F}_{t-1}, I_t = i, J_t = j \sim \mathcal{N}(A_{ij}, 1)$. Each observation, including a repeated query to the same cell, counts once.[1] For $z = (x,y)$, the two unilateral-deviation regrets, with standard basis vectors e_i, are

$$R_A(z) = \max_{i\in[n]} e_i^\top Ay - x^\top Ay, \qquad C_A(z) = x^\top Ay - \min_{j\in[m]} x^\top Ae_j. \tag{1}$$

Each is a player's largest possible gain and is nonnegative. Their sum is the duality gap $\mathrm{Gap}_A(z)$. An ε-Nash equilibrium requires both regrets to be at most ε (Maiti et al., 2023), defining $\mathcal{C}_\varepsilon^{\mathrm{ind}}(A) = \{z \in \Delta_n \times \Delta_m : R_A(z) \le \varepsilon,\ C_A(z) \le \varepsilon\}$. A gap of at most ε suffices, while a valid pair may have gap 2ε. A policy π chooses adaptive queries, stops after τ queries, and returns $\widehat{z}$. It is probably approximately correct, or (ε,δ)-PAC-Nash, on $\mathcal{F}$ if, for every $A \in \mathcal{F}$, it stops almost surely and $\mathbb{P}_A^\pi(\widehat{z} \in \mathcal{C}_\varepsilon^{\mathrm{ind}}(A)) \ge 1-\delta$. Requiring correctness on the full payoff box gives the minimax query complexity

$$\mathcal{M}_{n,m}^{\mathrm{G}}(\varepsilon,\delta) = \inf_{\pi\ \mathrm{PAC\text{-}Nash}} \sup_{A\in[-1,1]^{n\times m}} \mathbb{E}_A^\pi[\tau]. \tag{2}$$

For $\mathcal{F}_r(n,m) = \{A \in [-1,1]^{n\times m} : \mathrm{rank}(A) \le r\}$, restricting both correctness and the supremum to $\mathcal{F}_r(n,m)$ defines $\mathcal{M}_{\le r}^{\mathrm{G}}(n,m,\varepsilon,\delta)$. The policy knows r, but not factors or equilibrium supports.[2] Costs count observations. Arithmetic costs are stated separately. Unlike the expectation in Equation (2), a deterministic cap bounds τ on every transcript. For a 1×1 game, the unique strategy pair is an exact equilibrium without observations, so $\mathcal{M}_{1,1}^{\mathrm{G}}(\varepsilon,\delta) = 0$.

2.2 Fixed-Candidate Verification

For a fixed candidate, verification compares its payoff with each player's best response. The following lemma quantifies the cost of checking these deviation gains using fresh observations.

Lemma 2.1 (Fixed-candidate verification)**.** *For every $0 < \varepsilon \le 1/4$ and $0 < \alpha < 1/4$, there is a verifier with the following guarantee for every $A \in [-1,1]^{n\times m}$ and candidate $z = (x,y) \in \Delta_n \times \Delta_m$ determined by an arbitrary observation history. Holding z fixed, it uses at most*

$$C(n+m)\varepsilon^{-2}\log(c/\alpha)$$

fresh queries, for universal constants $c, C > 0$. Conditional on the history, it accepts with probability at most α if $z \notin \mathcal{C}_\varepsilon^{\mathrm{ind}}(A)$, and at least $1-\alpha$ if $\mathrm{Gap}_A(z) \le \varepsilon/4$. The query cap is deterministic.

The accuracy margin leaves an intermediate region where either decision is allowed: sufficiently accurate candidates are accepted, and candidates violating the target regret guarantee are rejected, with the stated confidence. This separation makes verification possible under noise. At fixed accuracy

[1] For binary feedback, $W \in \{-1,+1\}$ has mean A_{ij}. Equivalently, $X \sim \mathrm{Bernoulli}((1+A_{ij})/2)$ and $W = 2X - 1$. All payoff, regret, and rank conventions refer to the centered matrix A.

[2] Rank means $\mathrm{rank}(A)$, rather than the bimatrix convention $\mathrm{rank}(A+B)$ for payoff matrices (A,B), under which every zero-sum game has rank zero (Patris & Panageas, 2024).

and confidence, a constant-factor change in accuracy does not remove discovery's additional $\log\log k$ dependence on the number of actions.

To sample the response of row i, draw $J \sim y$ and query (i, J). For column j, draw $I \sim x$ and query (I, j). Drawing both actions independently samples the candidate payoff. PAC search finds near-best responses without an action-set logarithm (Lemma C.5). The verifier then accurately estimates only the selected responses and the candidate payoff, using disjoint fresh samples. This avoids estimating all responses simultaneously or learning a uniformly accurate payoff table. Lemma D.1 gives the full proof. The next section asks whether finding a valid candidate through the same observations requires more information.

3 Unrestricted Games: The Cost of Discovery

A supplied candidate can be verified with linear action dependence under the accuracy margin in Lemma 2.1. We now show that discovering one can require strictly more information, even when the learner may return any mixed-strategy pair satisfying the target regret guarantees.

Theorem 3.1 (Discovery lower bounds for unrestricted games)**.** *There is a universal constant $c > 0$ such that the following statements hold for $0 < \delta \le 1/4$.*

(i) Balanced games. *For every $k \ge 2$ and $0 < \varepsilon \le 1/16$,*

$$\mathcal{M}^{\mathrm{G}}_{k,k}(\varepsilon, \delta) \ge \frac{ck}{\varepsilon^2}\left[\log\log(ek) + \log\frac{1}{\delta}\right]. \tag{3}$$

(ii) Rectangular games. *For $n, m \ge 1$, $(n, m) \ne (1, 1)$, and $0 < \varepsilon \le 1/8$, define the effective discovery widths*

$$u_{n|m} = \min\left\{n, \varepsilon^{-1}, \frac{m}{\log(en)}\right\}, \qquad u_{m|n} = \min\left\{m, \varepsilon^{-1}, \frac{n}{\log(em)}\right\}.$$

Then

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon, \delta) \ge \frac{c}{\varepsilon^2}\left[(n + m - 2)\log\frac{1}{\delta} + n\log(e + u_{n|m}) + m\log(e + u_{m|n})\right]. \tag{4}$$

The first term in Equation (4) is the confidence cost. The discovery widths describe how many hidden labels fit within the two action sets and the accuracy budget. Part (i) uses a different construction to retain a $\log\log(ek)$ discovery factor even at fixed accuracy.

Corollary 3.2 (A separation between discovery and verification)**.** *For fixed $0 < \varepsilon \le 1/16$ and $0 < \delta \le 1/4$, discovery requires $\mathcal{M}^{\mathrm{G}}_{k,k}(\varepsilon, \delta) = \Omega_{\varepsilon,\delta}(k\log\log(ek))$. In contrast, Lemma 2.1 verifies a supplied candidate in $O_{\varepsilon,\delta}(k)$ queries under its stated accuracy margin, with error probability at most δ. The additional information cost is therefore intrinsic to discovering a valid candidate.*

For comparison, high-probability bandit self-play (Bai et al., 2022) gives, for the parameters in Theorem 3.1(ii) and a universal $C > 0$,

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon, \delta) \le \frac{C}{\varepsilon^2}\left[n\log(en) + m\log(em) + (n + m)\log\frac{1}{\delta}\right]. \tag{5}$$

Lemmas E.1 and E.2 transfer this guarantee to Gaussian observations by converting them into bounded unbiased feedback, using a constant expected number of queries per converted observation.

Corollary 3.3 (Regimes of matching query complexity)**.** *Under the parameters of Theorem 3.1(ii), the following rates hold.* ***(a) High confidence.*** *If $\log(1/\delta) \ge \log(e(n + m))$, then $\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon, \delta) = \Theta((n + m)\varepsilon^{-2}\log(1/\delta))$.* ***(b) Balanced high resolution.*** *For fixed δ and $\varepsilon \le \log(ek)/k$, the rate is $\mathcal{M}^{\mathrm{G}}_{k,k}(\varepsilon, \delta) = \Theta_\delta(k\varepsilon^{-2}\log(ek))$. The same rate also holds for $\varepsilon = k^{-a}$ with any fixed $a > 0$ as $k \to \infty$, with constants depending on a and δ.*

These rates follow from Theorem 3.1 and Equation (5). Separately, response compression in Appendix R gives $\Theta_{r,\varepsilon}((n + m)\log(1/\delta))$ for each fixed smaller side $r = \min\{n, m\} \ge 1$ and fixed $\varepsilon \in (0, 1/8]$ in nontrivial dimensions, including fixed-accuracy $2 \times k$ games. At fixed $0 < \varepsilon \le 1/16$ and $0 < \delta \le 1/4$, the bounds remain $\Omega_{\varepsilon,\delta}(k\log\log(ek)) \le \mathcal{M}^{\mathrm{G}}_{k,k}(\varepsilon, \delta) \le O_{\varepsilon,\delta}(k\log(ek))$. The strict separation from verification is established, but the exact balanced-game discovery rate remains open.

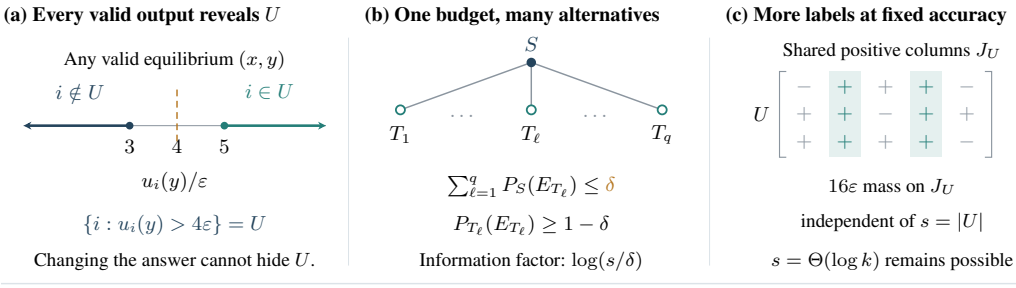

Figure 2: **Why discovery costs more information.** (a) Every valid output reveals U. (b) $q = \Theta(s)$ deletion neighbors share one error budget, and change of measure and averaging supply $\log(s/\delta)$. (c) Shared positive columns permit $s = \Theta(\log k)$ at fixed accuracy.

Proof idea. **Step 1: every valid output reveals the hidden label.** The hard games force every valid output to reveal a hidden subset $U \subseteq [N]$. For $\varepsilon \le 1/16$, public scores $a_i(j) \in [-1/2, 1/2]$ define paired rows $A^U_{(i,\sigma),j} = \sigma(a_i(j) - 8\varepsilon \mathbf{1}\{i \in U\})$, $\sigma \in \{-1, 1\}$. A witness ensures $0 \le V^\star_{A^U} \le \varepsilon$ for each *good* label. Write $u_i(y) = \sum_j a_i(j) y_j$ and $F_U(y) = \max_i |u_i(y) - 8\varepsilon \mathbf{1}\{i \in U\}|$. Since $F_U(y)$ is the largest row response, every valid pair satisfies $F_U(y) \le V^\star_{A^U} + R_{A^U}(x, y) + C_{A^U}(x, y) \le 3\varepsilon$. Thus $u_i(y) \ge 5\varepsilon$ on U and $u_i(y) \le 3\varepsilon$ outside it, so $\mathcal{D}(y) = \{i : u_i(y) > 4\varepsilon\} = U$. Choosing another equilibrium or support cannot avoid decoding. Both regrets matter: row regret controls the largest response, while column regret controls the candidate payoff's excess over the value.

Step 2: a shared error budget forces discovery information. The gain over pairwise testing comes from sharing one failure budget across many neighboring labels. A constant fraction of size-s labels S are good, with $\Theta(s)$ good deletion neighbors $T = S \setminus \{i\}$. For the stopped-transcript laws P_U, decoding events $E_T = \{\mathcal{D}(\widehat{y}) = T\}$ satisfy

$$\sum_T P_S(E_T) \le \delta, \qquad P_T(E_T) \ge 1 - \delta.$$

These disjoint errors under S share a total budget δ. The extra logarithm comes from keeping many distinct wrong answers jointly unlikely, even though the learner is free to choose its output. Each deletion changes one row pair by 8ε, so stopped change of measure relates $\mathrm{KL}(P_T \| P_S)$ to queries on that pair under T. Convexity and averaging over the good labels give $\Omega(N\varepsilon^{-2} \log(s/\delta))$ for $N \ge 2s \ge 4$. Exact output decoding thus converts equilibrium discovery into label identification, and the shared confidence budget supplies the extra $\log s$ factor (Figure 2).

Remark 3.4 (**Why fixed accuracy needs a different witness**). The additive construction uses the mass bound $16s\varepsilon$ and requires $s \le 1/(16\varepsilon)$. The intersection witness instead averages columns positive on every row in U: the same 16ε mass serves all selected labels, with the rest on the zero column. For good labels, unselected-row scores remain small. An $O(2^s \log N)$-column dictionary permits $N = \Theta(k)$ and $s = \Theta(\log k)$ at fixed accuracy. The shared error budget then contributes $\log(s/\delta) = \Theta(\log\log(ek) + \log(1/\delta))$, yielding Theorem 3.1(i).

The confidence term comes from a rank-two star with $n + m - 2$ alternatives. Each requires $\Omega(\varepsilon^{-2} \log(1/\delta))$ observations under the base instance on its changed row or column, with each query charged at most twice (Appendices H to J). The lower bounds also transfer to centered binary feedback at sufficiently small accuracy (Appendices J and N).

4 Low-Rank Games: Response Compression

At fixed accuracy, fixed-rank structure eliminates the discovery overhead. The learner is given only the rank bound and must identify suitable reference rows and columns directly from noisy payoff observations, even when the informative entries are highly localized and no useful anchors or low-dimensional representation are supplied in advance. The challenge is therefore not merely to exploit low rank, but to discover a stable response representation that remains valid across the full action sets.

Table 1: **Low rank removes the discovery overhead.** Square games with $k \geq 2$.

(a) Fixed accuracy and confidence

Task	Query complexity	Guarantee
Candidate verification	$O(k)$	Accuracy margin
Unrestricted discovery	$\Omega(k \log\log k)$ to $O(k \log k)$	Exact rate open
Rank $\leq r$ discovery	$\Theta_{r,\varepsilon,\delta}(k)$	Minimax optimal

(b) Accuracy and confidence dependence

Discovery class	Query upper bound	Optimality
Rank ≤ 1	$O\big(k\varepsilon^{-2}\log(1/\delta)\big)$	All parameters
Rank ≤ 2	$O\big(k\varepsilon^{-2}L^3\log(L/\delta)\big)$	ε exponent, up to logs
Skew rank ≤ 2	$O\big(k\varepsilon^{-2}[\ell+\log(1/\delta)]\big)$	k, δ at fixed ε

Throughout, $0 < \delta \leq 1/4$. In (a), $\varepsilon \in (0, 1/16]$, δ, and $r \geq 1$ are fixed. Discovery costs are minimax expected costs. In (b), $0 < \varepsilon \leq 1/8$ and the upper bounds are deterministic caps. Here $L = \log(2/\varepsilon)$ and $\ell = \log(2 + \min\{k, \varepsilon^{-1}\})$.

Theorem 4.1 (Fixed-rank dimension-confidence law)**.** *For each fixed integer $r \geq 1$ and $\varepsilon \in (0, 1/8]$, uniformly over $(n, m) \neq (1, 1)$ and $0 < \delta \leq 1/4$,*

$$\mathcal{M}^{\mathrm{G}}_{\leq r}(n, m, \varepsilon, \delta) = \Theta_{r,\varepsilon}\left((n+m)\log\frac{1}{\delta}\right).$$

A uniform Gaussian-query policy realizes the upper bound with deterministic cap $C_r(n+m)\varepsilon^{-(4r+6)}\log(c_r/\delta)$, where $c_r, C_r > 0$ depend only on r.

Remark 4.2 (**Why fixed rank suffices**). The general construction expands its response representation after each reliable rejection. An invertible $s \times s$ cross defines a low-rank surrogate, and rejection of a surrogate equilibrium in the original game exposes a nonzero Schur residual and hence a missing response direction. Each such rejection increases the represented dimension, so at most r expansions occur. Bounded interpolation controls all strategic responses from the selected directions, while determinant margins stabilize empirical inversion under noise. Since the number of representation changes, response directions, and required precision depend only on r and ε, and each response search has linear action dependence, the resulting query complexity is linear in $n + m$ for fixed r, ε. Appendix M gives the full construction and analysis.

4.1 Exact and Robust Rank-One Certification

Theorem 4.3 (Exact and robust rank-one certification)**.** *For universal $c, C > 0$, $(n, m) \neq (1, 1)$, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$,*

$$c(n+m)\varepsilon^{-2}\log\frac{1}{\delta} \leq \mathcal{M}^{\mathrm{G}}_{\leq 1}(n, m, \varepsilon, \delta) \leq C(n+m)\varepsilon^{-2}\log\frac{c}{\delta}.$$

A single Gaussian-query policy attains the upper bound with a deterministic query cap. For $\mathcal{F}_{1,\rho} = \{A \in [-1,1]^{n\times m} : \mathrm{dist}_\infty(A, \{B : \mathrm{rank}(B) \leq 1\}) \leq \rho\}$ and $0 \leq \rho \leq \varepsilon/16$, the same policy, without knowing ρ, is uniformly (ε, δ)-PAC-Nash on $\mathcal{F}_{1,\rho}$, with the same minimax order and deterministic cap.

Rank-one discovery matches the verification scale without a supplied informative entry. A one-sided rank-one family gives the matching lower bound. The same law and deterministic cap hold for Rademacher and centered Bernoulli feedback (Corollary L.2). For square games at $\varepsilon = k^{-a}$, fixed $a > 0$ and fixed confidence, Corollary 3.3 and Theorem 4.3 give the exact ratio $\mathcal{M}^{\mathrm{G}}_{k,k}/\mathcal{M}^{\mathrm{G}}_{\leq 1} = \Theta_{a,\delta}(\log k)$.

Remark 4.4 (**Extreme responses, query cost, and robustness**). When the required estimates are accurate, either an initial pure pair is certified or its rejection reveals an $\Omega(\varepsilon)$ payoff signal. This leads to a nonzero pivot A_{ab} with uniformly bounded row and column ratios. The identity $A_{ij} = A_{ib}A_{aj}/A_{ab}$ reduces response control to two scalar intervals. Approximating their endpoints controls response errors to $O(\varepsilon)$ against every opposing mixture. The selected rows and columns

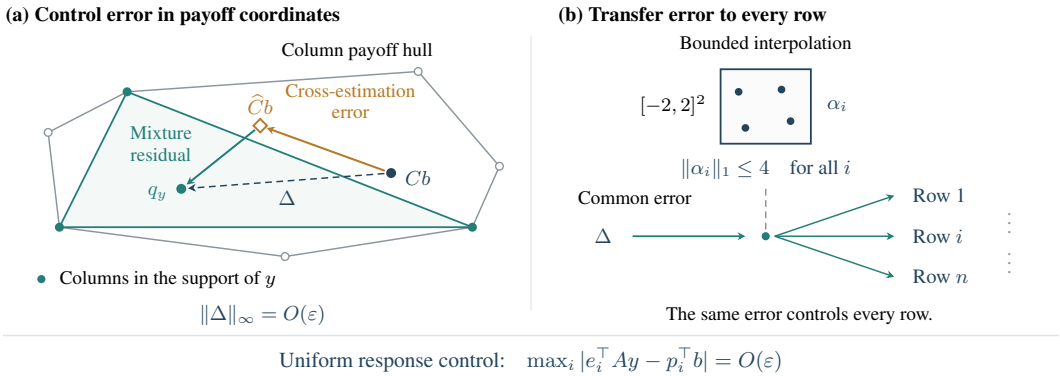

Figure 3: **Small payoff errors control every row response.** Here C is the true selected 2×2 payoff cross, $\widehat{C}$ its estimate, and q_y the payoff vector of y on the selected rows. (a) The green triangle uses columns in the support of y. With $O(\varepsilon)$ mixture residual and entrywise cross-estimation error, and $\|b\|_\infty \le 4$, the combined error $\Delta = q_y - Cb$ is $O(\varepsilon)$ in infinity norm. (b) Bounded interpolation transfers it to every row. The coefficients are analytical, not estimated. The exact bound is Equation (6). The full equilibrium guarantee also needs column-side control and PAC response comparisons.

define a constant-size game. Its payoffs are estimated directly at their intersections, avoiding noisy quotients. Endpoint transfer lifts its equilibrium guarantee to the full game.

A constant number of PAC extremum searches over the two action sets and fresh estimates of a constant number of selected payoffs, at accuracy proportional to ε, give $O((n+m)\varepsilon^{-2}\log(c/\delta))$ queries. No simultaneous estimation of all responses is needed, avoiding an action-count logarithm. For perturbations, the identity applies to the latent rank-one matrix. Entrywise errors alter best responses and candidate payoffs by at most ρ, and the available margins absorb them for unknown $\rho \le \varepsilon/16$. Appendix L contains the full construction and the $\varepsilon + O(\rho)$ extension for a supplied larger perturbation bound. This response-based guarantee also covers matrices concentrated on one unknown cell, whose full recovery can require inspecting the table (Krishnamurthy & Singh, 2013).

4.2 Rank-Two Certification at Payoff-Scale Precision

Rank two replaces scalar endpoint control with planar mixture geometry and includes nontransitive skew-symmetric games. In normalized rock–paper–scissors, every distribution on at most two actions has best-response payoff at least $1/3$, while the uniform mixture gives zero (Proposition P.8).

Theorem 4.5 (Rank-two response compression)**.** *For $0 < \varepsilon \le 1/8$ and $0 < \delta \le 1/4$, Gaussian-query policies give the following guarantees with probability at least $1 - \delta$, for universal $c, C > 0$. Write $L_\varepsilon = \log(2/\varepsilon)$ and $\ell_{n,\varepsilon} = \log(2 + \min\{n, \varepsilon^{-1}\})$.*

1. *For every $A \in [-1,1]^{n\times m}$ of rank at most two, a pair with both regrets at most ε uses at most $C(n+m)\varepsilon^{-2}L_\varepsilon^3 \log(cL_\varepsilon/\delta)$ queries.*
2. *For every $A = -A^\top \in [-1,1]^{n\times n}$ of rank at most two, a strategy q with $\max_i (Aq)_i \le \varepsilon$ uses at most $Cn\varepsilon^{-2}[\ell_{n,\varepsilon} + \log(1/\delta)]$ queries.*

Both query caps are deterministic. Each output distribution has support at most three, and both policies have polynomial real-arithmetic cost.

Remark 4.6 (**Response control and query complexity**). The rank-two bound keeps the same ε^{-2} sampling cost while allowing three-action mixtures. On its guarantee event, the rectangular construction either returns an ε-equilibrium or finds a nonsingular 2×2 submatrix $C = A_{I,J}$, called a *cross*, with bounded interpolation on both sides: $A_{i,:} = \alpha_i^\top A_{I,:}$ and $A_{:,j} = A_{:,J}\beta_j$, with $\|\alpha_i\|_\infty, \|\beta_j\|_\infty \le 2$. Thus each action's response is a bounded linear combination of the selected responses. These coefficients are analytical objects, not separately estimated features. Finding the cross is included in the query cost. The nonsingular branch occurs only at rank two. Lower-rank inputs use the equilibrium branch.

Let $p_i^\top = A_{i,J}$, $q_y = A_{I,:}y$, and let $\widehat{C}$ have entrywise error at most h. For $\|b\|_\infty \le 4$, the identity $e_i^\top Ay - p_i^\top b = \alpha_i^\top (q_y - Cb)$ gives

$$\max_{i\in[n]} \left|e_i^\top Ay - p_i^\top b\right| \le 4\left(\|q_y - \widehat{C}b\|_\infty + 8h\right). \tag{6}$$

Thus $O(\varepsilon)$ residual and cross-estimation error control every row response (Figure 3). Column interpolation ensures bounded feasible coordinates. PAC comparisons and the argument for $-A^\top$ give both saddle inequalities. A screened empirical inverse locates a feasible region. The output-error bound does not require accurate estimation of the true inverse.

A support query approximately maximizes a linear function of an action's two selected payoff coordinates. Planar membership, feasibility, and value search use $O(L_\varepsilon^3)$ such queries, each costing $O((n+m)\varepsilon^{-2}\log(cL_\varepsilon/\delta))$ observations. Cross discovery fits within the same order. Thus nested searches contribute precision logarithms while preserving the ε^{-2} exponent (Appendix P.4).

Skew symmetry parameterizes the relevant directions by one scalar. Shared coordinate samples control all directions on one event, including directions chosen adaptively from those data. The retained set has $O(\min\{n, \varepsilon^{-1}\})$ actions, giving $\ell_{n,\varepsilon}$ (Appendix P.3). The transitive rank-two subclass gives the optimal accuracy exponent. The remaining logarithms are not known necessary.

5 Globally Valid Adaptation to Unknown Structure

The learner need not identify or test the underlying structure before choosing an algorithm. We interleave structure-aware proposal routines with generic self-play and verify every candidate using fresh observations from the original game. Because verification is independent of the proposal mechanism, whichever candidate is accepted first can safely supply the final output.

Corollary 5.1 (Globally valid structural adaptation)**.** *Fix $0 < \varepsilon \le 1/8$ and $0 < \delta \le 1/4$. There exists a single policy, given only $n, m, \varepsilon, \delta$, that is globally (ε, δ)-PAC-Nash on $[-1,1]^{n\times m}$ and stops almost surely. Writing $D = n + m$, its expected query cost satisfies*

$$\mathbb{E}_A \tau \le C\varepsilon^{-2}\big[n\log(en) + m\log(em) + D\log(1/\delta)\big].$$

The same policy automatically attains sharper bounds when additional structure is present. On $\mathcal{F}_{1,\varepsilon/32}$, it satisfies $\mathbb{E}_A\tau \le CD\varepsilon^{-2}\log(c/\delta)$. For every matrix of rank at most two,

$$\mathbb{E}_A \tau \le CD\varepsilon^{-2}\big[L_\varepsilon^3 \log(cL_\varepsilon) + \log(c/\delta)\big].$$

For skew-symmetric matrices of rank at most two, this improves to $Cn\varepsilon^{-2}[\ell_{n,\varepsilon} + \log(c/\delta)]$. Here L_ε and $\ell_{n,\varepsilon}$ are as in Theorem 4.5, and all constants are universal.

Remark 5.2 (**Why structural adaptation remains globally valid**). Each candidate is fixed before its verification data are drawn, and a common error budget across verification attempts controls whichever candidate is ultimately accepted. The generic self-play branch guarantees almost-sure stopping and the unrestricted expected query bound on every bounded game.

When a structural assumption holds, the corresponding structured branch produces a verifier-acceptable candidate with conditional probability bounded away from zero on every attempt, so the number of attempts has a geometric tail even when proposal cost and proposal quality are dependent. Since structured candidates need only be generated at constant confidence, the dependence on the final confidence level enters through the additive verification cost $D\varepsilon^{-2}\log(c/\delta)$. Appendix Q gives the higher-rank guarantees and precise completeness margins.

The same policy also adapts to shifted skew-symmetric rank-two structure without knowing either the shift or the approximation radius. At entrywise distance at most $\varepsilon/32$, its expected query cost is $Cn\varepsilon^{-2}[L_\varepsilon\log(cL_\varepsilon) + \log(c/\delta)]$, while at distance at most $\varepsilon/2048$, it improves to $Cn\varepsilon^{-2}[\ell_{n,\varepsilon} + \log(c/\delta)]$. The corresponding shifted-skew neighborhoods are specified in Appendices P.5, P.6 and Q.

6 Related Work

We compare our results with closely related work. Further literature is discussed in Appendix B.

Noisy equilibrium learning and discovery. High-probability self-play (Bai et al., 2022) is the direct source of our generic query upper bound. Noisy matrix-game identification (Maiti et al., 2023; 2024; Maiti, 2025) asks for finer local complexity. The fast precision dependence of Jiang & Zhang (2026) concerns the expected strategy, whereas our certificate concerns the realized output with prescribed probability. Condorcet identification already separates locating informative comparisons from certifying them (Saad et al., 2026). Our lower bound covers arbitrary mixed answers without a pure-winner assumption. Infinite-answer pure exploration requires careful answer selection (Poiani et al., 2026). Our fixed-answer appendix is not a claim of end-to-end instance optimality. Finite-confidence subset selection also incurs discovery costs (Kalyanakrishnan et al., 2012). The additional issue here is that an equilibrium need not identify a prescribed subset or support. The dense construction supplies the missing reduction: every valid mixed answer decodes a label, so multiplicity cannot be avoided by selecting a different equilibrium.

Structural and oracle methods. Rank-one bandits optimize one payoff rather than two-sided saddle inequalities (Katariya et al., 2017). Leveraged matrix estimation selects its own anchors without supplied rank or conditioning bounds, while its recovery guarantee still depends on spikiness and conditioning (Stojanovic et al., 2024). Our query bounds avoid these instance parameters by controlling responses. CUR matrix approximation and max-volume pivoting are established tools (Goreinov et al., 1997). Our contribution is how failed certificates supply stable pivots and finite-confidence progress. Exact matrix-vector methods achieve $\widetilde{O}(\varepsilon^{-2/3})$ queries (Karmarkar et al., 2026), but each reveals a full exact response vector. With exact cell-query access, an entry is permanently revealed. A noisy query supplies one sample of its mean. Our bounds therefore account for repeated observations and prescribed error probability. An exact-query upper bound alone does not establish the stated finite-confidence cost (Proposition T.1).

Response compression and noisy feasibility. PAC arm elimination and vector-valued pure exploration provide established sampling tools. Approximate Pareto identification under a polyhedral order (Ararat & Tekin, 2023) returns a set of designs. Our compression contract instead retains a finite family's directional maxima and then combines this guarantee with stable pivot discovery and planar search to return a certified mixture. Its query bound pays for the number of directions that must be preserved. Noisy convex-hull membership is also an established sampling problem (Niss et al., 2022). The rectangular result needs a particular weak interface: an approximate mixture on selected action labels or a quantitative separator. PAC support queries and selected-coordinate estimates implement that interface after discovering a doubly dominant cross. The contribution is the complete reduction from noisy cell observations to equilibrium certification at $O(\varepsilon)$ physical precision. Here physical precision means error in actual payoffs, avoiding dependence on the arbitrary scaling of latent factors. The low-dimensional features are not supplied to the learner.

7 Conclusion

In this work, we establish an information-theoretic separation between discovering and verifying approximate Nash equilibria in bounded zero-sum games. Under Gaussian payoff queries at fixed accuracy and confidence, unrestricted discovery requires $\Omega(k \log \log k)$ observations, whereas a supplied candidate can be verified with $O(k)$ observations under a constant-factor accuracy margin. Fixed rank removes this overhead through stable response representations learned directly from noisy payoffs, with linear dependence on the number of actions. Independent verification retains these gains without knowing the rank and preserves correctness on every bounded game, even when the proposal mechanism's structural assumptions fail.

These results identify an intrinsic discovery cost and a structural mechanism that removes it. The lower bound permits any approximate equilibrium, without prescribing a particular equilibrium, support, or answer. Our sharp full-box minimax benchmarks in several regimes specify worst-case rates that any complete instance-dependent theory must recover. Future directions include obtaining the full instance-dependent characterization posed by Maiti (2025), determining the exact unrestricted discovery rate in the balanced fixed-accuracy regime, and characterizing the optimal joint dependence on rank and accuracy.

AI USE STATEMENT

Generative AI tools were used to assist with language revision, manuscript organization, presentation refinement, and LaTeX preparation. The research questions, discovery and verification formulations, lower-bound constructions, response-compression ideas, algorithmic designs, theorem statements, proof strategies, and complete proofs presented in this work were independently conceived and developed by the authors. After the relevant proofs had been completed by the authors, generative AI tools were used solely as supplementary verification aids to stress-test the existing proofs, check derivations and constants, and examine whether the arguments contained potential logical gaps, overly compressed steps, unstated assumptions, edge cases, or inconsistencies involving quantifiers, parameter regimes, adaptive sampling conditions, or requirements concerning the use of fresh independent observations. Any issues identified through this auxiliary verification process, together with any resulting revisions or clarifications, were independently evaluated, implemented, and verified by the authors. The authors take full responsibility for the correctness, originality, and content of the final manuscript.

ETHICS STATEMENT

This theoretical study does not involve personal data or human participants and makes no empirical claims about deployed systems.

REPRODUCIBILITY STATEMENT

The observation model, certification objective, and query complexity conventions are specified in Section 2, with additional assumptions and parameter restrictions stated alongside the corresponding results. Complete proofs, explicit lower-bound constructions, and details of the sampling and verification procedures are provided in the appendices.

REFERENCES

Cagin Ararat and Cem Tekin. Vector optimization with stochastic bandit feedback. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 2165–2190, 2023. URL https://proceedings.mlr.press/v206/ararat23a.html.

Yu Bai, Chi Jin, Song Mei, and Tiancheng Yu. Near-optimal learning of extensive-form games with imperfect information. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 1337–1382. PMLR, 2022.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009. doi: 10.1007/s10208-009-9045-5. URL https://doi.org/10.1007/s10208-009-9045-5.

Remy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*, volume 32, pp. 14591–14600, 2019.

Yali Du, Xue Yan, Xu Chen, Jun Wang, and Haifeng Zhang. Estimating α-Rank from a few entries with low rank matrix completion. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 2870–2879. PMLR, 2021. URL https://proceedings.mlr.press/v139/du21e.html.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(39):1079–1105, 2006.

John Fearnley and Rahul Savani. Finding approximate Nash equilibria of bimatrix games via payoff queries. *ACM Transactions on Economics and Computation*, 4(4):25:1–25:19, 2016. doi: 10.1145/2956579. URL https://doi.org/10.1145/2956579.

John Fearnley, Martin Gairing, Paul W. Goldberg, and Rahul Savani. Learning equilibria of games via payoff queries. *Journal of Machine Learning Research*, 16:1305–1344, 2015. URL https://jmlr.org/papers/v16/fearnley15a.html.

Côme Fiegel, Pierre Menard, Tadashi Kozuno, Michal Valko, and Vianney Perchet. The harder path: Last iterate convergence for uncoupled learning in zero-sum games with bandit feedback. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 17131–17152. PMLR, 2025. URL https://proceedings.mlr.press/v267/fiegel25a.html.

Dean Foster, Dylan J. Foster, Noah Golowich, and Alexander Rakhlin. On the complexity of multi-agent decision making: From learning in games to partial monitoring. In Gergely Neu and Lorenzo Rosasco (eds.), *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 2678–2792. PMLR, 12–15 Jul 2023. URL https://proceedings.mlr.press/v195/foster23a.html.

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

Aurelien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 998–1027. PMLR, 2016.

Paul W. Goldberg and Aaron Roth. Bounds for the query complexity of approximate equilibria. *ACM Transactions on Economics and Computation*, 4(4):24:1–24:25, 2016. doi: 10.1145/2956582. URL https://eccc.weizmann.ac.il/report/2013/136/.

Sergei A. Goreinov, Nikolai L. Zamarashkin, and Eugene E. Tyrtyshnikov. Pseudo-skeleton approximations by matrices of maximal volume. *Mathematical Notes*, 62:515–519, 1997. doi: 10.1007/BF02358985.

Michael D. Grigoriadis and Leonid G. Khachiyan. A sublinear-time randomized approximation algorithm for matrix games. *Operations Research Letters*, 18(2):53–58, 1995. doi: 10.1016/0167-6377(95)00032-0. URL https://doi.org/10.1016/0167-6377(95)00032-0.

Hédi Hadiji, Sarah Sachs, Tim van Erven, and Wouter M. Koolen. Towards characterizing the first-order query complexity of learning (approximate) Nash equilibria in zero-sum matrix games. In *Advances in Neural Information Processing Systems*, volume 36, pp. 13356–13373, 2023. URL https://arxiv.org/abs/2304.12768.

Shinji Ito, Haipeng Luo, Taira Tsuchiya, and Yue Wu. Instance-dependent regret bounds for learning two-player zero-sum games with bandit feedback. In *Proceedings of the 38th Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 2858–2892. PMLR, 2025. URL https://proceedings.mlr.press/v291/ito25a.html.

Jiashuo Jiang and Mengxiao Zhang. An LP-based approach for bilinear saddle point problem with instance-dependent guarantee and noisy feedback. *arXiv preprint arXiv:2602.12513*, 2026. URL https://arxiv.org/abs/2602.12513.

Kwang-Sung Jun, Rebecca Willett, Stephen Wright, and Robert Nowak. Bilinear bandits with low-rank structure. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3163–3172. PMLR, 2019. URL https://proceedings.mlr.press/v97/jun19a.html.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning*, 2012. URL https://icml.cc/2012/papers/359.pdf.

Ishani Karmarkar, Liam O'Carroll, and Aaron Sidford. Solving matrix games with near-optimal matvec complexity. *arXiv preprint arXiv:2601.02347*, 2026. URL https://arxiv.org/abs/2601.02347. Version 3, June 2026.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvari, Claire Vernade, and Zheng Wen. Stochastic Rank-1 bandits. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 392–401. PMLR, 2017. URL https://proceedings.mlr.press/v54/katariya17a.html.

Emilie Kaufmann, Olivier Cappe, and Aurelien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016.

Akshay Krishnamurthy and Aarti Singh. Low-rank matrix and tensor completion via adaptive sampling. In *Advances in Neural Information Processing Systems*, volume 26, pp. 836–844, 2013. URL https://proceedings.neurips.cc/paper/2013/hash/2050e03ca119580f74cca14cc6e97462-Abstract.html.

Marc Lanctot, Vinicius Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Perolat, David Silver, and Thore Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4190–4203, 2017.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Low-rank generalized linear bandit problems. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 460–468. PMLR, 2021. URL https://proceedings.mlr.press/v130/lu21a.html.

Arnab Maiti. Open problem: Optimal instance-dependent sample complexity for finding Nash equilibrium in two player zero-sum matrix games. In *Proceedings of the Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 6230–6234. PMLR, 2025. URL https://proceedings.mlr.press/v291/maiti25b.html.

Arnab Maiti, Kevin Jamieson, and Lillian Ratliff. Instance-dependent sample complexity bounds for zero-sum matrix games. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 9429–9469. PMLR, 2023.

Arnab Maiti, Ross Boczar, Kevin Jamieson, and Lillian Ratliff. Near-optimal pure exploration in matrix games: A generalization of stochastic bandits & dueling bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2602–2610. PMLR, 2024.

Arnab Maiti, Claire Jie Zhang, Kevin Jamieson, Jamie Heather Morgenstern, Ioannis Panageas, and Lillian J. Ratliff. Efficient uncoupled learning dynamics with $\widetilde{O}(T^{-1/4})$ last-iterate convergence in bilinear saddle-point problems over convex sets under bandit feedback. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 2431–2439. PMLR, 2026. URL https://proceedings.mlr.press/v300/maiti26a.html.

Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:623–648, 2004. URL https://jmlr.org/papers/v5/mannor04b.html.

Subhojyoti Mukherjee, Qiaomin Xie, Josiah P. Hanna, and Robert Nowak. Multi-task representation learning for pure exploration in bilinear bandits. In *Advances in Neural Information Processing Systems*, volume 36, pp. 47816–47827, 2023. doi: 10.52202/075280-2073. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/95464e2e49103dc560091ed2c64a5b12-Abstract-Conference.html.

Şerban Nacu and Yuval Peres. Fast simulation of new coins from old. *The Annals of Applied Probability*, 15(1A):93–115, 2005. doi: 10.1214/105051604000000549. URL https://doi.org/10.1214/105051604000000549.

Anupam Nayak, Tong Yang, Osman Yagan, Gauri Joshi, and Yuejie Chi. Achieving logarithmic regret in KL-regularized zero-sum Markov games. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2510.13060. arXiv:2510.13060.

Gergely Neu. Explore no more: Improved high-probability regret bounds for non-stochastic bandits. In *Advances in Neural Information Processing Systems*, volume 28, pp. 3168–3176, 2015.

Laura Niss, Yuekai Sun, and Ambuj Tewari. Achieving representative data via convex hull feasibility sampling algorithms. *arXiv preprint arXiv:2204.06664*, 2022. URL https://arxiv.org/abs/2204.06664.

Brendan O'Donoghue, Tor Lattimore, and Ian Osband. Matrix games with bandit feedback. In Cassio de Campos and Marloes H. Maathuis (eds.), *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 279–289. PMLR, 27–30 Jul 2021. URL https://proceedings.mlr.press/v161/o-donoghue21a.html.

Nikolas Patris and Ioannis Panageas. Learning Nash equilibria in Rank-1 games. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=8utTlmhw8v.

Riccardo Poiani, Martino Bernasconi, and Andrea Celli. Pure exploration with infinite answers. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 2989–2997. PMLR, 2026. URL https://proceedings.mlr.press/v300/poiani26a.html.

Mark Rowland, Shayegan Omidshafiei, Karl Tuyls, Julien Perolat, Michal Valko, Georgios Piliouras, and Remi Munos. Multiagent evaluation under incomplete information. In *Advances in Neural Information Processing Systems*, volume 32, pp. 12270–12282, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/510f2318f324cf07fce24c3a4b89c771-Abstract.html.

El Mehdi Saad, Victor Thuot, and Nicolas Verzelen. The sampling complexity of Condorcet winner identification in dueling bandits. *arXiv preprint arXiv:2603.15189*, 2026. URL https://arxiv.org/abs/2603.15189.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Proceedings of the 30th Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pp. 1794–1834. PMLR, 2017.

Stefan Stojanovic, Yassir Jedra, and Alexandre Proutiere. Model-free low-rank reinforcement learning via leveraged entry-wise matrix estimation. In *Advances in Neural Information Processing Systems*, volume 37, pp. 30886–30924, 2024. doi: 10.52202/079017-0972. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/371713c3e5314dff9483c62c5abb98a8-Abstract-Conference.html.

David Yang, Yuan Gao, Tianyi Lin, and Christian Kroer. A direct second-order method for solving two-player zero-sum games. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2512.12910. arXiv:2512.12910.

Brian Hu Zhang and Tuomas Sandholm. Small Nash equilibrium certificates in very large games. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/4fbe073f17f161810fdf3dab1307b30f-Abstract.html.

Yichi Zhou, Jialian Li, and Jun Zhu. Identify the Nash equilibrium in static games with random payoffs. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 4160–4169. PMLR, 2017.

Contents of Appendix

A Notation Reference **16**

B Additional Related Work **17**

C Probability, observation, and perturbation tools **18**

D Expected-cost Propose–Verify–Restart **19**

E Gaussian upper bound without clipping **20**
E.1 An exact bounded unbiased observation 20
E.2 Gaussian proposal with uniformly bounded expected cost 21
E.3 Fresh verification and restart 22

F Robust subgaussian upper bound by clipping **24**
F.1 Bounded bandit self-play 24
F.2 Correctness and expected cost 25

G Sequential Gaussian change of measure **25**

H Strengthened finite-confidence rank-two star lower bound **26**

I Dense-good-layer Gaussian lower bound **27**
I.1 Dimension-optimized form 29

J A Superlinear Lower Bound at Fixed Accuracy **30**

K Low-Rank Probability Primitives **32**

L Exact and Robust Rank-One Certification **32**
L.1 Statement 33
L.2 Sampling procedure 33
L.3 Proof of the upper bound 33
L.4 Matching lower bound 35
L.5 Entrywise approximate-rank-one robustness 36

Contents of Appendix (continued)

M Fixed-Rank Certification **36**
M.1 Main theorem . . . 36
M.2 Exact CUR residual identities . . . 38
M.3 Dominant-row purification . . . 39
M.4 A low-dimensional candidate for the current CUR surrogate . . . 40
M.5 Fresh verifier with a progress witness . . . 42
M.6 Verifier rejection produces a Schur pivot . . . 42
M.7 Three distinct invariants and their dependencies . . . 44
M.8 Filtration, risk budget, and totality . . . 45
M.9 Proof of the fixed-rank theorem . . . 45

N Bounded Rollouts and Additional Structural Consequences **46**
N.1 Rademacher and Bernoulli phase transfer . . . 46
N.2 Skew-symmetric single-policy self-play . . . 47
N.3 Why a universal strong-gap primitive is impossible . . . 47

O Formal Fixed-Candidate Characteristic Definitions **48**

P Rank-two response compression **49**
P.1 Stable skew crosses and a direct solver . . . 49
P.2 Planar support geometry and three-action output . . . 52
P.3 Shared physical sampling for the skew bound . . . 56
P.4 Rectangular rank two at physical precision . . . 58
P.5 Perturbing the planar support primitive . . . 64
P.6 A wider neighborhood of shifted skew games . . . 67
P.7 Compressed robustness and its safe interface . . . 70
P.8 Binary feedback implementations . . . 72

Q Globally Valid Structure-Adaptive Certification **72**
Q.1 The verifier's completeness margin . . . 72
Q.2 A physical-query scheduling lemma . . . 73
Q.3 A larger unknown-radius rank-one neighborhood . . . 74
Q.4 A single policy with no rank or radius input . . . 75

R Response compression with a fixed smaller action set **76**

S Individual-Regret and Duality-Gap Characteristics Can Differ Linearly **80**
S.1 Construction . . . 80
S.2 Individual-regret certificate . . . 80
S.3 Duality-gap lower characteristic . . . 81

T Relation to the COLT 2025 Open Problem and the Remaining Balanced Regime **82**
T.1 Full-box minimax and instance-dependent quantifiers . . . 82
T.2 What fixed-accuracy discovery now requires . . . 82
T.3 Boundary results and barriers . . . 83

A Notation Reference

The table collects the main notation and selected appendix quantities. The last column links to their definitions. Symbols used only within an individual proof retain their local meanings.

Table 2: Notation reference.

Notation	Meaning	First introduced
Game, observations, and certification		
n, m, k	Row and column action counts, with $n = m = k$ for balanced games.	Section 2
$A, A_{ij}, \|A\|_\infty$	Payoff matrix, cell mean, and entrywise maximum absolute value.	Section 2
Δ_d, x, y, z	Probability simplex on d actions, mixed strategies, and their pair $z = (x, y)$.	Section 2
$V_A^\star$	Game value $\max_x \min_y x^\top A y$.	Section 2
ε, δ	Target individual-regret tolerance and failure probability.	Section 2
$\mathcal{F}_{t-1}, I_t, J_t, X_t$	Past history, queried row and column, and fresh noisy payoff at time t.	Section 2
$\pi, \tau, \widehat{z}$	Query policy, its stopping time, and returned strategy pair.	Section 2
$R_A(z), C_A(z)$	Row and column unilateral-deviation regrets.	Equation (1)
$\mathrm{Gap}_A(z)$	Duality gap $R_A(z) + C_A(z)$.	Section 2
$\mathcal{C}_\varepsilon^{\mathrm{ind}}(A)$	Pairs whose two individual regrets are each at most ε.	Section 2
$\mathcal{F}_r(n, m)$	Bounded payoff matrices of rank at most r.	Section 2
$\mathcal{M}_{n,m}^{\mathrm{G}}, \mathcal{M}_{\le r}^{\mathrm{G}}$	Minimax expected Gaussian-query costs for unrestricted and rank-at-most-r games.	Section 2
α	Conditional error allowance for one fresh verification attempt.	Lemma 2.1
Discovery lower bounds		
$u_{n\|m}, u_{m\|n}$	Effective discovery widths in the rectangular lower bound.	Theorem 3.1
$U \subseteq [N], s$	Hidden subset label in the hard family, with label size s.	Section 3
$a_i(j), u_i(y), F_U(y)$	Public scores, their mixture averages, and the largest row response in the hard game.	Section 3
$\mathcal{D}(y)$	Decoder recovering the hidden label from a valid output mixture.	Section 3
Structured representations and robustness		
$\rho, \mathcal{F}_{1,\rho}$	Entrywise perturbation radius and the class within that distance of rank one.	Theorem 4.3
A_{ab}	Nonzero rank-one pivot used in $A_{ij} = A_{ib}A_{aj}/A_{ab}$.	Remark 4.4
$I, J, C = A_{I,J}$	Selected row and column labels and their nonsingular cross.	Remark 4.6
α_i, β_j	Bounded row and column interpolation coefficients.	Remark 4.6
p_i, q_y	Selected payoff coordinates $p_i^\top = A_{i,J}$ and mixture coordinates $q_y = A_{I,:}y$.	Remark 4.6
$\widehat{C}, h, b$	Estimated cross, its entrywise error bound, and a bounded response-coordinate vector.	Remark 4.6
$L_\varepsilon, \ell_{n,\varepsilon}$	Precision factors $\log(2/\varepsilon)$ and $\log(2 + \min\{n, \varepsilon^{-1}\})$.	Theorem 4.5
q (skew games)	One mixed strategy satisfying $\max_i (Aq)_i \le \varepsilon$.	Theorem 4.5
B, E	Cross surrogate $B = A_{:,J}C^{-1}A_{I,:}$ and residual $E = A - B$.	Lemma M.3
Fixed-candidate information quantities		
$\mathcal{C}_\varepsilon^{\mathrm{DG}}(A)$	Pairs with duality gap at most ε.	Appendix O
$\mathrm{Alt}_{\mathsf{C}}(z), w$	Matrices invalidating z under criterion C, and a cell-sampling allocation.	Appendix O
$D_{\mathsf{C}}, D_{\mathsf{C}}^\star, T_{\mathsf{C}}^\star$	Fixed-answer information rate, its answer/allocation optimum, and its reciprocal.	Appendix O

B Additional Related Work

Equilibrium computation and payoff access. Randomized approximation and multiplicative-weights methods established early connections between repeated play and matrix-game computation (Grigoriadis & Khachiyan, 1995; Freund & Schapire, 1999). Payoff-query complexity isolates the information needed to find equilibria when the payoff table is not supplied (Fearnley et al., 2015; Fearnley & Savani, 2016; Goldberg & Roth, 2016). First-order oracle models instead reveal response vectors (Hadiji et al., 2023). These perspectives motivate separating observations from arithmetic cost. Our observation unit is a fresh noisy scalar payoff, so learning an entry to a given precision can require repeated queries. The comparison between discovery and verification uses this same observation model for both tasks.

Noisy games and policy evaluation. Pure-strategy equilibrium identification from random payoffs (Zhou et al., 2017) and regret minimization with bandit feedback (O'Donoghue et al., 2021; Ito et al., 2025) study how strategically useful information can be acquired without a known matrix. Policy-space response methods organize reinforcement learning around a population of policies (Lanctot et al., 2017). Multiagent evaluation also studies adaptive sampling of noisy matchups, with ranking as the output (Rowland et al., 2019). These works motivate our finite-policy setting. Our certificate controls every unilateral deviation against the returned mixtures, including unused actions. This objective differs from ranking policies or controlling cumulative regret against the opponents encountered during learning.

Pure exploration and the choice of an answer. PAC arm selection provides linear action dependence for finding an approximately best response (Mannor & Tsitsiklis, 2004; Even-Dar et al., 2006). Change-of-measure lower bounds and optimal sampling allocations explain fixed-confidence best-arm identification (Kaufmann et al., 2016; Garivier & Kaufmann, 2016). Multiple-answer and infinite-answer formulations make the selection of a statistically favorable answer part of the learning problem (Degenne & Koolen, 2019; Poiani et al., 2026). Moderate-confidence analysis also exposes exploration costs that are not captured by a leading high-confidence term (Simchowitz et al., 2017). Our unrestricted lower bound addresses both issues. It holds at fixed confidence and forces every valid mixed equilibrium output to reveal hidden information, without prescribing which equilibrium or support must be returned.

Low-rank learning and matrix recovery. Matrix completion reconstructs missing entries from low-rank structure (Candès & Recht, 2009; Krishnamurthy & Singh, 2013), and this approach has been used for multiagent ranking (Du et al., 2021). Bilinear bandits exploit supplied action features and an unknown low-rank parameter (Jun et al., 2019). Related work studies generalized linear rewards (Lu et al., 2021) and representation learning for pure exploration across tasks (Mukherjee et al., 2023). Our learner receives action labels and noisy payoffs, rather than low-dimensional response features. The representation is learned only to the extent needed to control strategic responses. This permits certification without requiring uniform recovery of all matrix entries.

Other equilibrium guarantees. Small certificates in extensive-form games exploit partial game-tree structure (Zhang & Sandholm, 2020). Last-iterate convergence under bandit feedback (Fiegel et al., 2025; Maiti et al., 2026), learning in KL-regularized Markov games (Nayak et al., 2026), and second-order methods for solving zero-sum games (Yang et al., 2026) address other output requirements or computational resources. Our focus is the number of noisy payoff observations needed to return a statistically valid equilibrium. Independent verification makes the correctness guarantee apply even when the structural model used to propose a candidate is inaccurate.

C Probability, observation, and perturbation tools

Write $[d] = \{1, \dots, d\}$ and $\Delta_d = \{p \in \mathbb{R}^d : p \geq 0, \sum_i p_i = 1\}$. We use $\mathrm{Supp}(x) = \{i : x_i > 0\}$ for the support and Δ_S for probability distributions on a finite label set S. Let $N_{i,j}(t) = \sum_{s=1}^{t} \mathbf{1}\{(I_s, J_s) = (i, j)\}$ count queries to cell (i, j). A bar, as in $\bar{A}_{ij}$, denotes the empirical mean of the specified fresh sample block. Hats also denote general estimates and outputs.

A real random variable Z is σ-subgaussian when $\mathbb{E}\exp(\lambda(Z - \mathbb{E}Z)) \leq \exp(\lambda^2\sigma^2/2)$ for every $\lambda \in \mathbb{R}$. All robust upper bounds allow different fixed laws at different cells. Conditional on the complete past and the queried cell, the next observation must be a fresh independent draw from that cell's fixed law. We use only the stated centered subgaussian condition. Gaussian lower bounds use fresh independent $\mathcal{N}(A_{ij}, 1)$ observations.

Lemma C.1 (Subgaussian mean concentration)**.** *Let $Z_1, \dots, Z_N$ be independent with common mean μ and centered σ-subgaussian noise. Then*

$$\mathbb{P}\left(\left|N^{-1}\sum_{t=1}^{N} Z_t - \mu\right| > a\right) \leq 2\exp\left(-\frac{Na^2}{2\sigma^2}\right).$$

The same bound holds conditionally when the law is selected by a sigma-field independent of the fresh observations.

Proof. The exponential-moment bound and Markov's inequality give the upper tail $\exp(-\lambda N a + N\lambda^2\sigma^2/2)$. Optimize at $\lambda = a/\sigma^2$, repeat for $-Z_t$, and add. Conditional use fixes the selecting sigma-field first. □

Corollary C.2 (Event form of the stopped Gaussian change of measure)**.** *Let $A, B \in [-1, 1]^{n\times m}$ and let a sequential policy stop at τ. If $\mathbb{E}_A\tau < \infty$, then for every stopped-transcript event E,*

$$\frac{1}{2}\sum_{i,j} \mathbb{E}_A N_{i,j}(\tau)(A_{ij} - B_{ij})^2 \geq \mathrm{kl}(\mathbb{P}_A(E), \mathbb{P}_B(E)).$$

If $\mathbb{E}_A\tau = \infty$, every finite lower bound is immediate.

Proof. This is the event-level consequence of the stopped-transcript identity in Lemma G.1, whose proof in Appendix G handles the passage from deterministic truncations to the full stopped transcript. Apply data processing to the map that records $\mathbf{1}_E$. □

Lemma C.3 (Regret perturbation)**.** *For every $z = (x, y)$ and every two payoff matrices A, B,*

$$|R_A(z) - R_B(z)| \leq 2\|A - B\|_\infty, \qquad |C_A(z) - C_B(z)| \leq 2\|A - B\|_\infty.$$

Proof. For each row deviation, $|(e_i - x)^\top(A - B)y| \leq \|e_i - x\|_1\|A - B\|_\infty\|y\|_1 \leq 2\|A - B\|_\infty$. Taking a maximum preserves the bound. The column side is symmetric. □

Lemma C.4 (Uniform clipping bias)**.** *Let $X = \mu + \xi$ with $\mu \in [-1, 1]$ and centered 1-subgaussian ξ. For $B > 1$ let $[X]_{[-B,B]} = \max\{-B, \min\{B, X\}\}$. Then*

$$\left|\mathbb{E}[X]_{[-B,B]} - \mu\right| \leq \frac{2}{B-1}\exp\left(-\frac{(B-1)^2}{2}\right).$$

Proof. The clipping error is at most $(|X| - B)_+$. Integrate $\mathbb{P}(|X| > t) \leq 2\exp(-(t-1)^2/2)$ for $t > B$ and use the Gaussian-tail inequality. □

Lemma C.5 (PAC extremum interface)**.** *For $d \geq 1$, $0 < q \leq 1$, and $0 < \alpha \leq 1/2$, let the d arms have fixed means and centered constant-subgaussian observations. Median elimination returns a q-near maximizer with probability at least $1 - \alpha$ after at most $Cdq^{-2}\log(c/\alpha)$ pulls. The pull count is deterministic and has no $\log d$ factor. Negating observations gives the minimum-arm version. For $d = 1$ no sample is required.*

Proof. This is the median-elimination guarantee of Even-Dar et al. (2006), after rescaling the universal subgaussian proxy. □

Lemma C.6 (Restart accounting with random cycle cost). *Let $0 < p \le 1$, $0 < \delta \le 1/2$, and $a, b \ge 0$. Suppose the probability of correct acceptance in every reached cycle is at least p conditional on the past. If the conditional expected cost of cycle r is at most $a + b\log(r/\delta)$, then the procedure stops almost surely and*

$$\mathbb{E}\sum_{r=1}^{R}\left(a + b\log\frac{r}{\delta}\right) \le C_p\left(a + b\log\frac{1}{\delta} + b\right).$$

If conditional incorrect-acceptance risk in cycle r is at most α_r, total error is at most $\sum_r \alpha_r$.

Proof. The probability of reaching cycle r is at most $(1-p)^{r-1}$. Apply Tonelli's theorem after conditioning at cycle entry and sum the geometric series, including $\sum_r (1-p)^{r-1}\log r < \infty$. The error claim is a conditional union bound. □

D Expected-cost Propose–Verify–Restart

This section separates candidate generation from correctness. Fresh verification controls the error of each accepted candidate. Restarting after rejection converts a constant probability of proposing a good candidate into an expected query bound.

Fix $A \in [-1,1]^{n\times m}$, $0 < \varepsilon \le 1/4$, and $0 < \delta \le 1/4$. A proposal invocation may be randomized, adaptive, and measurable with respect to its own fresh data. Malformed outputs are rejected by a deterministic simplex-membership check.

Lemma D.1 (Fresh two-regret verifier). *Let $Z = (x, y)$ be measurable with respect to an arbitrary proposal transcript $\mathcal{H}$, and use only observations independent of $\mathcal{H}$. For every $\alpha \in (0, 1/4)$ there is a verifier which, conditionally on every realized transcript,*

(i) accepts an incorrect candidate with probability at most α.

(ii) accepts with probability at least $1 - \alpha$ if $\mathrm{Gap}_A(Z) \le \varepsilon/4$.

(iii) uses at most $C(n+m)\varepsilon^{-2}\log(c/\alpha)$ fresh observations.

(iv) on acceptance, reports $\widehat{i}, \widehat{j}$ and $U_R, U_C < \varepsilon$ that simultaneously upper-bound $R_A(Z)$ and $C_A(Z)$ except on an event of probability at most α.

The result is uniform over fixed per-cell centered 1-subgaussian laws with independent repeated pulls.

Proof. Condition on $\mathcal{H}$. For row i, draw $J \sim y$ and query (i, J). The resulting arm has mean $r_i = e_i^\top Ay$ and a universal subgaussian proxy: the random mean is bounded and the cell noise is subgaussian. Use Lemma C.5 with accuracy $q = \varepsilon/32$ and risk $\alpha/4$ to find $\widehat{i}$ with $r_{\widehat{i}} \ge \max_i r_i - q$. Symmetrically find $\widehat{j}$ with $c_{\widehat{j}} \le \min_j c_j + q$, where $c_j = x^\top Ae_j$.

Using disjoint fresh samples, estimate $r_{\widehat{i}}$, $c_{\widehat{j}}$, and $v = x^\top Ay$ to accuracy $a = \varepsilon/32$ with total error at most $\alpha/2$. Accept when $\widehat{r} - \widehat{v} \le \varepsilon/2$ and $\widehat{v} - \widehat{c} \le \varepsilon/2$. On the simultaneous good event, an actual regret larger than ε produces an estimated regret at least $\varepsilon - q - 2a = 29\varepsilon/32 > \varepsilon/2$. If $\mathrm{Gap}_A(Z) \le \varepsilon/4$, each nonnegative regret is at most $\varepsilon/4$ and each estimated regret is at most $\varepsilon/4 + 2a < \varepsilon/2$. The two PAC calls and three scalar estimates give the stated deterministic cap.

On acceptance, report

$$U_R = \max\{0, \widehat{r} - \widehat{v} + q + 2a\}, \qquad U_C = \max\{0, \widehat{v} - \widehat{c} + q + 2a\}.$$

On the same simultaneous event, the PAC inequalities and scalar confidence bounds give $R_A(Z) \le U_R$ and $C_A(Z) \le U_C$. Acceptance also gives $U_R, U_C \le \varepsilon/2 + 3\varepsilon/32 < \varepsilon$. □

Theorem D.2 (Expected-cost PVR wrapper). *Assume that conditional on every history before an invocation, a proposal terminates almost surely, has conditional expected physical cost at most $\overline{T}_{\mathrm{prop}}$, and returns $Z \in \Delta_n \times \Delta_m$ with $\mathrm{Gap}_A(Z) \le \varepsilon/4$ with probability at least $p > 0$. In cycle r, verify*

with $\alpha_r = 6\delta/(\pi^2 r^2)$ and restart after rejection. The resulting policy is globally (ε, δ)-PAC-Nash, stops almost surely, and

$$\mathbb{E}_A \tau \le C_p \left[\overline{T}_{\mathrm{prop}} + \frac{n+m}{\varepsilon^2} \left(1 + \log \frac{1}{\delta} \right) \right].$$

With probability at least $1-\delta$, the final verifier simultaneously reports deviation labels and bounds $R_A(Z) \le U_R < \varepsilon$ and $C_A(Z) \le U_C < \varepsilon$.

Proof. Condition on every reached proposal transcript. By Lemma D.1, an incorrect candidate is accepted with probability at most α_r, so the total error is at most $\sum_r \alpha_r = \delta$. A good proposal occurs with probability at least p and is accepted with probability at least $1 - \alpha_r \ge 3/4$. Hence each reached cycle has correct-acceptance probability at least $3p/4$ and cycle reach has a geometric tail.

The verifier cap is at most

$$C(n+m)\varepsilon^{-2} \left(1 + \log \frac{1}{\delta} + 2 \log r \right).$$

The proposal need not have a deterministic physical-sample cap. Conditional on entry into cycle r, its expected cost is at most $\overline{T}_{\mathrm{prop}}$. Tonelli's theorem and Lemma C.6 therefore give the displayed expected total cost and almost-sure stopping. The union bound over verifier failures also makes every reported final confidence bound valid on the same probability-$1-\delta$ event. □

E Gaussian upper bound without clipping

E.1 An exact bounded unbiased observation

The self-play bound below requires bounded observations, whereas Gaussian observations are unbounded. The following randomized conversion produces bounded feedback with exactly the same mean and constant expected sampling cost. This exact conversion is specific to the Gaussian model. The clipping argument in Appendix F handles general subgaussian laws.

Lemma E.1 (Fast bounded Gaussian mean factory)**.** *There is a fixed randomized procedure* BG *with the following property. Given access to independent $\mathcal{N}(\mu, 1)$ observations for an unknown $\mu \in [-1, 1]$, it returns $W \in \{-2, 2\}$ such that*

$$\mathbb{E}_\mu W = \mu.$$

If L is the number of Gaussian inputs used, then universal constants $C_F < \infty$ and $\rho_F \in (0,1)$ satisfy

$$\sup_{\mu \in [-1,1]} \mathbb{P}_\mu(L > \ell) \le C_F \rho_F^\ell, \qquad \ell \ge 0.$$

In particular, $\kappa_F := \sup_{\mu \in [-1,1]} \mathbb{E}_\mu L < \infty$.

Proof. For one Gaussian input X, the bit $U = \mathbf{1}\{X \ge 0\}$ is a p-coin with $p = \Phi(\mu)$. Hence p lies in the compact interval $I = [\Phi(-1), \Phi(1)] \subset (0,1)$. Define

$$f(p) = \frac{1}{2} + \frac{1}{4} \Phi^{-1}(p).$$

The function f is real analytic on a neighborhood of I and $f(I) = [1/4, 3/4] \subset (0,1)$. The fast Bernoulli-factory theorem of Nacu & Peres (2005, Theorem 2 and Proposition 21) therefore supplies one decision tree which, uniformly over $p \in I$, consumes a number L of independent p-coins with exponential tails and outputs $B \in \{0,1\}$ with $\mathbb{P}(B=1) = f(p)$. Each input coin costs one fresh Gaussian sample. Set $W = 4B - 2$. Then

$$\mathbb{E}W = 4 \left(\frac{1}{2} + \frac{\mu}{4} \right) - 2 = \mu,$$

and the tail statement is the uniform tail conclusion on the closed interval I. Because the sufficient part of the cited theorem is constructive and f, I are fixed and known, fix its resulting decision tree once and for all. At a node requesting a p-coin, query one fresh Gaussian input and pass the bit

$\mathbf{1}\{X \geq 0\}$. At a node requesting a fair bit, use private randomization. At a leaf, return its label. This is the implementation of BG, and it requires neither μ nor any matrix-dependent input. The constants C_F, ρ_F, κ_F are used only in analysis and are not algorithm parameters. Finally, the exponential tail implies $\kappa_F < \infty$. □

The analyticity theorem is used as an exact simulation theorem, not as an approximation. External fair bits used by the factory are policy randomization and do not count as cell observations.

E.2 Gaussian proposal with uniformly bounded expected cost

We use the following imported bounded-feedback result, in the exact form needed here.

Lemma E.2 (Bounded bandit self-play)**.** *Let $B \geq 1$, $0 < \eta \leq 1$, and $0 < \rho \leq 1/2$. Suppose every queried payoff lies in $[-B, B]$, has conditional mean A_{ij}, and is fresh conditional on the pre-round history and queried cell. There is a matrix-independent simultaneous bandit self-play routine that uses one bounded reward sample per round and, for the deterministic horizon*

$$T = \left\lceil C_0 B^2 \eta^{-2} \left[n \log \frac{c_0 n}{\rho} + m \log \frac{c_0 m}{\rho} \right] \right\rceil,$$

returns $(\bar{x}, \bar{y})$ satisfying $\mathrm{Gap}_A(\bar{x}, \bar{y}) \leq \eta$ with probability at least $1 - \rho$.

Proof. Put

$$M_{ij} = \frac{A_{ij} + B}{2B} \quad \text{and} \quad R_t = \frac{W_t + B}{2B}.$$

Then $R_t \in [0, 1]$ and, conditional on the pre-round history and on the sampled cell (I_t, J_t), its mean is $M_{I_t J_t}$.

We now spell out the specialization of Bai et al. (2022, Theorem 4). Take their extensive-form horizon to be $H = 1$, with one information set for each player. Thus their information-set parameters are $X = Y = 1$, while their two action counts are n and m. There are no transitions to estimate. In round t, the current policies x_t and y_t sample one row and one column, and the resulting scalar R_t is the shared reward. The row player runs the max-player version of Balanced OMD on reward R_t, whose loss input at $H = 1$ is $1 - R_t$. The column player runs a max-player copy on the complementary reward $1 - R_t$, so its loss input is $1 - (1 - R_t) = R_t$. Maximizing the complementary payoff is exactly minimizing M. In particular, the regret of this second copy is

$$\max_{y^\dagger \in \Delta_m} \sum_{t=1}^{T} \left(x_t^\top M y_t - x_t^\top M y^\dagger \right),$$

which is the min-player regret in the original normalized game.

Apply Bai et al. (2022, Theorem 4) to the two copies with failure probability $\rho/2$ each. The assumptions are preserved because each observed reward is fresh conditional on the history and sampled cell. A union bound, which does not require independence of the two regret events, shows that with probability at least $1 - \rho$,

$$\mathcal{R}_T^{\max} \leq C \sqrt{nT \log \frac{cn}{\rho}}, \qquad \mathcal{R}_T^{\min} \leq C \sqrt{mT \log \frac{cm}{\rho}}.$$

At $H = 1$, sequence-form averaging is ordinary coordinate-wise averaging. Hence Bai et al. (2022, Proposition 1) gives the exact online-to-batch identity

$$\mathrm{Gap}_M(\bar{x}, \bar{y}) = \frac{\mathcal{R}_T^{\max} + \mathcal{R}_T^{\min}}{T} \leq C \sqrt{\frac{n \log(cn/\rho) + m \log(cm/\rho)}{T}},$$

where the last step also absorbs the factor from adding the two square roots. This is the same simultaneous-self-play conversion recorded in their Corollary 5, with separate action-count logarithms retained here.

Finally, the affine shift does not affect either best-response difference and scaling back gives $\mathrm{Gap}_A(x, y) = 2B\,\mathrm{Gap}_M(x, y)$. Increasing the universal constant C_0 in the displayed choice of T therefore makes the last quantity at most η. A player with one action is implemented trivially and has zero regret. In this $H = 1$ specialization the routine is precisely the explicit importance-weighted update used below and consumes one shared scalar observation per round. □

Corollary E.3 (Bounded-feedback matrix-free certification). *Let $n, m \geq 1$, $0 < \varepsilon \leq 1/4$, $0 < \delta \leq 1/4$, and $B \geq 1$. Suppose each cell has a fixed payoff law P_{ij} supported on $[-B, B]$, with mean $A_{ij} \in [-1, 1]$, and every query is conditionally a fresh draw from that law. There is a globally (ε, δ)-PAC-Nash policy satisfying, uniformly over every such A and every fixed collection $P = (P_{ij})$,*

$$\sup_{A,P} \mathbb{E}_{A,P}\tau \leq \frac{CB^2}{\varepsilon^2}\left[n\log(en) + m\log(em) + (n+m)\log\frac{1}{\delta}\right].$$

Proof. Run Lemma E.2 with constant failure probability to propose a candidate of duality gap at most $\varepsilon/4$. Scale the observations and mean game to X/B and A/B, then apply Theorem D.2 with target accuracy ε/B. The proposal has scaled duality gap at most $\varepsilon/(4B)$, as required. The wrapper returns scaled regrets at most ε/B, which become at most ε in the original game. Both proposal and verification costs acquire the displayed B^2 factor. □

Algorithm E.1 (Gaussian constant-confidence proposal $\mathsf{GProp}(\varepsilon)$).

1. Set $\eta = \varepsilon/4$, $\rho = 1/16$, $B = 2$, and let T be the horizon in Lemma E.2.
2. Initialize x_1 and y_1 uniformly. If $n > 1$, set
$$\lambda_x = \sqrt{n\log(n)/T}, \qquad \gamma_x = \sqrt{n\log(6n/\rho)/T}.$$
If $m > 1$, define λ_y, γ_y by replacing n with m.
3. For $t = 1, \ldots, T$, independently conditional on the joint past draw $I_t \sim x_t$, $J_t \sim y_t$. Run BG on cell (I_t, J_t), supplying each requested input bit from one fresh $\mathcal{N}(A_{I_tJ_t}, 1)$ query. If its leaf is $H_t \in \{0, 1\}$, put $R_t = H_t$ (equivalently $W_t = 4H_t - 2$).
4. If $n > 1$, for every $i \in [n]$ form
$$\widehat{\ell}_t^x(i) = \frac{\mathbf{1}\{I_t = i\}(1 - R_t)}{x_t(i) + \gamma_x/n}, \qquad x_{t+1}(i) = \frac{x_t(i)\exp[-(\lambda_x/n)\widehat{\ell}_t^x(i)]}{\sum_{k=1}^n x_t(k)\exp[-(\lambda_x/n)\widehat{\ell}_t^x(k)]}.$$
If $n = 1$, leave $x_{t+1} = x_t$. If $m > 1$, for every $j \in [m]$ form
$$\widehat{\ell}_t^y(j) = \frac{\mathbf{1}\{J_t = j\}R_t}{y_t(j) + \gamma_y/m}, \qquad y_{t+1}(j) = \frac{y_t(j)\exp[-(\lambda_y/m)\widehat{\ell}_t^y(j)]}{\sum_{k=1}^m y_t(k)\exp[-(\lambda_y/m)\widehat{\ell}_t^y(k)]},$$
and leave the one-action column player unchanged. These are exactly the $H = 1$ Balanced OMD and minimizing-player updates used in Lemma E.2.
5. Return $(\bar{x}, \bar{y}) = (T^{-1}\sum_{t=1}^T x_t, T^{-1}\sum_{t=1}^T y_t)$.

Lemma E.4 (Proposal validity and expected physical cost). *For every $0 < \varepsilon \leq 1/4$ and every $A \in [-1, 1]^{n\times m}$, $\mathsf{GProp}(\varepsilon)$ terminates almost surely, its expected number of physical Gaussian cell observations is at most*

$$T_{\mathrm{prop}} \leq \frac{C}{\varepsilon^2}\{n\log(en) + m\log(em)\}$$

and it returns a pair of duality gap at most $\varepsilon/4$ with probability at least $15/16$.

Proof. Conditional on the pre-call history and requested cell, each factory output is fresh, is exactly unbiased for that cell mean, and lies in $[-2, 2]$. Hence Lemma E.2 fails with probability at most $1/16$. Each factory call terminates almost surely. Because T is deterministic and finite, all T calls do too. If S_T is their total number of Gaussian inputs, conditional expectation and the tower property give $\mathbb{E}S_T \leq \kappa_F T$, even though the queried cells are adaptive. Substitution of fixed ρ, $B = 2$, and $\eta = \varepsilon/4$, with the universal factor κ_F absorbed into C, gives the displayed expected-cost bound. □

E.3 Fresh verification and restart

Lemma E.5 (Deterministically capped PAC extremum). *Let $K \geq 2$, $\sigma^2 \geq 1$, $0 < q \leq 1$, and $0 < \alpha \leq 1/2$. From independent observations of K arms with fixed means and common subgaussian proxy at most σ^2, one can return a q-near maximizer with probability at least $1 - \alpha$, using at most*

$$C\sigma^2 K q^{-2}\log(c/\alpha)$$

pulls. The pull count is deterministic. Negating observations gives the corresponding minimizer. For $K = 1$, return the sole arm without sampling.

Proof. This is the median-elimination construction of Even-Dar et al. (2006), recalled to fix every schedule. Set $S_1 = [K]$, $q_1 = q/4$, and $\alpha_1 = \alpha/2$. In round r, pull every arm in S_r

$$N_r = \left\lceil \frac{8\sigma^2}{q_r^2} \log \frac{3}{\alpha_r} \right\rceil$$

fresh times, and retain the $\lceil |S_r|/2 \rceil$ arms with largest empirical means. Put $q_{r+1} = 3q_r/4$, $\alpha_{r+1} = \alpha_r/2$, and stop when one arm remains.

Let $\mu_r^\star$ be the largest true mean in S_r. The best arm's empirical mean is below $\mu_r^\star - q_r/2$ with probability at most $\alpha_r/3$. Each arm of mean below $\mu_r^\star - q_r$ exceeds its own mean by $q_r/2$ with probability at most $\alpha_r/3$. Markov's inequality shows that at least half of all arms do so with probability at most $2\alpha_r/3$. Off these two events, some retained arm has mean at least $\mu_r^\star - q_r$. Thus, with total failure probability at most $\sum_r \alpha_r = \alpha$, the returned mean is at least $\mu_1^\star - \sum_r q_r = \mu_1^\star - q$.

While the routine is active, $|S_r| \le 2K\, 2^{1-r}$. Since $q_r^{-2} = 16q^{-2}(16/9)^{r-1}$ and $\log(3/\alpha_r) = \log(6/\alpha) + (r-1)\log 2$, its total cost is bounded by a universal multiple of

$$\sigma^2 K q^{-2} \sum_{r \ge 1} (8/9)^{r-1} \left(\log(6/\alpha) + r\right),$$

which proves the deterministic cap. □

For a fixed candidate $z = (x, y)$, a row arm i is sampled by drawing $J \sim y$ and querying (i, J). Its mean is $e_i^\top A y$. A column arm j is sampled by drawing $I \sim x$ and querying (I, j). Its mean is $x^\top A e_j$. The randomized cell mean lies in $[-1, 1]$, so together with unit Gaussian noise these arms have a universal subgaussian proxy. Apply Lemma E.5 to locate an $\varepsilon/32$-near-best row and an $\varepsilon/32$-near-best (minimum) column, with error $\alpha/4$ each. Using disjoint fresh samples, estimate the selected row mean r, selected column mean c, and baseline $v = x^\top A y$, each to absolute error $\varepsilon/32$ with error $\alpha/6$. A baseline sample draws $I \sim x$ and $J \sim y$ independently and queries (I, J). Form $\widehat{R} = \widehat{r} - \widehat{v}$ and $\widehat{C} = \widehat{v} - \widehat{c}$, and accept exactly when both are at most $\varepsilon/2$.

The simultaneous event just described has probability at least $1 - \alpha$. On it, if $R_A(z) > \varepsilon$, then $\widehat{R} > \varepsilon - 3\varepsilon/32 > \varepsilon/2$, and the analogous conclusion holds when $C_A(z) > \varepsilon$. Conversely, if $\mathrm{Gap}_A(z) \le \varepsilon/4$, each individual regret is at most $\varepsilon/4$, so $\widehat{R}, \widehat{C} \le \varepsilon/4 + 2\varepsilon/32 < \varepsilon/2$. Fixed-sample subgaussian estimates and Lemma E.5 give the deterministic cost

$$V(\alpha) \le C(n+m)\varepsilon^{-2} \log(c/\alpha).$$

Algorithm E.2 (Gaussian Propose–Verify–Restart)**.** For cycles $r = 1, 2, \ldots$:

1. Run $Z \leftarrow \mathsf{GProp}(\varepsilon)$ with fresh observations.
2. If Z is a valid pair of simplex vectors, run the fresh verifier above with $\alpha_r = 6\delta/(\pi^2 r^2)$.
3. If the verifier accepts, stop and return Z. Otherwise continue to cycle $r + 1$.

Theorem E.6 (Gaussian upper bound)**.** *For all $n, m \ge 1$, $0 < \varepsilon \le 1/4$, and $0 < \delta \le 1/4$, algorithm E.2 is matrix-independent, globally (ε, δ)-PAC-Nash, stops almost surely, and satisfies*

$$\sup_{A \in [-1,1]^{n \times m}} \mathbb{E}_A \tau \le \frac{C}{\varepsilon^2} \left[n \log(en) + m \log(em) + (n+m) \log \frac{1}{\delta} \right].$$

For $n = m = 1$, return the unique strategy without sampling.

Proof. Condition on every reached cycle and its complete proposal transcript. Freshness makes the conditional probability of accepting an incorrect candidate at most α_r. Hence the global error is at most $\sum_r \alpha_r = \delta$. By Lemma E.4, every reached cycle produces a good candidate with conditional probability at least $15/16$. Given that event, the verifier accepts with probability at least $1 - \alpha_r \ge 3/4$. Thus every reached cycle stops correctly with probability at least $45/64$. The probability of reaching

cycle r is at most $(19/64)^{r-1}$. Every proposal and verifier invocation itself terminates almost surely, so the whole policy does too.

Let S_r be the random physical cost of the proposal in cycle r, and let V_r be the deterministic verifier cap. Conditional on reaching the cycle and its entering history, $\mathbb{E}[S_r \mid \text{entering history}] \le T_{\text{prop}}$, while

$$V_r \le C(n+m)\varepsilon^{-2}\left(1 + \log\frac{1}{\delta} + 2\log r\right).$$

Tonelli's theorem and conditioning at cycle entry therefore give

$$\mathbb{E}_A\tau \le \sum_{r\ge 1}\left(\frac{19}{64}\right)^{r-1}(T_{\text{prop}} + V_r).$$

The geometrically weighted $\log r$ sum is finite, which proves the claim. Every counted unit is one actual Gaussian cell observation: factory inputs and verifier observations are counted explicitly. □

F Robust subgaussian upper bound by clipping

This appendix records the law-robust result under the following observation model. Each cell (i, j) has a fixed law P_{ij} with mean A_{ij} and centered 1-subgaussian noise. At every round, conditional on the complete interaction history and on querying cell (i, j), the observation is a fresh draw from P_{ij} independent of that history. The laws may differ across cells. This result is separate from the sharper Gaussian factory theorem in Appendix E.

F.1 Bounded bandit self-play

Lemma F.1 (Bounded bandit self-play)**.** *Let $d_1, d_2 \ge 1$, $B \ge 1$, $0 < \eta \le 1$, and $0 < \rho \le 1/2$. Suppose cell (i, j) has a fixed law supported on $[-B, B]$ with mean C_{ij}. Conditional on the complete past and on querying (i, j), the observation is a fresh draw from that law independent of the past. There is one algorithm, independent of C, that uses exactly one queried entry per round and after at most*

$$C_0 B^2 \eta^{-2}(d_1 + d_2)\log\frac{c_0(d_1+d_2)}{\rho}$$

queries returns $(\bar{x}, \bar{y})$ satisfying

$$\max_i e_i^\top C\bar{y} - \min_j \bar{x}^\top C e_j \le \eta$$

with probability at least $1 - \rho$.

Proof. Affinely rescale observations to $[0, 1]$. Apply Bai et al. (2022, Corollary 5) with horizon $H = 1$, one information set for each player, and action counts d_1, d_2. In each episode, draw actions from the current policies, query the resulting cell once, feed the rescaled payoff with opposite signs to the two learners, and update. The cited high-probability guarantee applies to the conditional mean game. With $H = 1$, its size parameter is $d_1 + d_2$ and its two regret bounds imply

$$\max_i e_i^\top C\bar{y} - \min_j \bar{x}^\top C e_j \le CB\sqrt{\frac{d_1\log(cd_1/\rho) + d_2\log(cd_2/\rho)}{T}}.$$

Choose the displayed horizon. EXP3-IX supplies an alternative high-probability bandit interface (Neu, 2015). □

Set $B_\varepsilon = 1 + \sqrt{2\log(128/\varepsilon)}$ and clip each proposal observation to $[-B_\varepsilon, B_\varepsilon]$. Let C be the matrix of clipped means. By Lemma C.4, $\|C - A\|_\infty \le \varepsilon/64$. Run Lemma F.1 with $\eta = \varepsilon/8$ and $\rho = 1/16$. On its success event,

$$\begin{aligned}\mathrm{Gap}_A(\bar{x}, \bar{y}) &\le \max_i e_i^\top C\bar{y} - \min_j \bar{x}^\top C e_j + 2\|A - C\|_\infty \\ &< \varepsilon/4.\end{aligned}$$

Since $B_\varepsilon^2 \le C\log(e/\varepsilon)$, the deterministic proposal horizon is

$$T_{\mathrm{prop}} \le C(n+m)\varepsilon^{-2}\log(e(n+m))\log(e/\varepsilon). \tag{7}$$

Thus clipped self-play returns a candidate of duality gap at most $\varepsilon/4$ with probability $15/16$ and has the displayed deterministic proposal horizon.

F.2 Correctness and expected cost

Theorem F.2 (Universal subgaussian upper bound)**.** *For every $n, m \ge 1$, $0 < \varepsilon \le 1/4$, and $0 < \delta \le 1/4$, PVR-SP is a matrix-independent policy that stops almost surely, is globally (ε, δ)-PAC-Nash, and satisfies*

$$\mathbb{E}_A\tau \le \frac{C}{\varepsilon^2}\left[(n+m)\log(e(n+m))\log\frac{e}{\varepsilon} + (n+m)\log\frac{1}{\delta}\right]$$

uniformly over the full payoff box and every fixed collection of per-cell laws satisfying the observation model above. In particular, conditional on the complete past and the queried cell, each observation is a fresh independent draw from that cell's fixed law.

Proof. Apply Theorem D.2 with $p = 15/16$ and the cap in Equation (7). The wrapper proves conditional safety on every proposal branch, including incorrect or malformed branches, and its geometric reach bound gives almost-sure termination. Substituting the proposal cap into its expected-cost inequality yields the display. The proof does not select an equilibrium or assume uniqueness, nondegeneracy, full support, or a positive equilibrium gap. □

The decomposition is additive. Constant-confidence self-play pays the dimension and clipping logarithms. The fresh verifier pays $\log(1/\delta)$ without multiplying those proposal logarithms, because it locates near-best responses rather than estimating all nm means uniformly.

G Sequential Gaussian change of measure

The lower bounds use the following stopped-transcript identity.

Lemma G.1 (Stopped Gaussian KL and data processing)**.** *Let A, B be two matrices, and let $N_{i,j}(\tau)$ be the number of queries to cell (i,j). For any policy that stops almost surely under both instances and has $\mathbb{E}_A\tau < \infty$, the laws P_A, P_B of the stopped transcript (including the output) obey*

$$\mathrm{KL}(P_A\|P_B) = \frac{1}{2}\sum_{i,j}\mathbb{E}_A N_{i,j}(\tau)(A_{ij}-B_{ij})^2.$$

Consequently, for every stopped-transcript event E,

$$\frac{1}{2}\sum_{i,j}\mathbb{E}_A N_{i,j}(\tau)(A_{ij}-B_{ij})^2 \ge \mathrm{kl}(P_A(E), P_B(E)).$$

Proof. Pad the interaction after stopping with a common dummy symbol and first truncate at deterministic horizon T. Conditional on the past, the policy uses the same action kernel under A and B, so action-selection and private-randomization likelihood ratios cancel. A queried $\mathcal{N}(A_{ij}, 1)$ observation contributes KL $(A_{ij}-B_{ij})^2/2$. The finite-horizon conditional chain rule gives the equality with $N_{i,j}(\tau \wedge T)$.

The sigma-fields of these padded prefixes increase to the full stopped-history sigma-field. By the variational characterization of relative entropy, their relative entropies increase to the relative entropy of the stopped history. On the count side, $N_{i,j}(\tau \wedge T) \uparrow N_{i,j}(\tau)$, so monotone convergence gives the displayed equality. The reported output is drawn from the same conditional Markov kernel of the stopped history under both instances. The conditional chain rule therefore adds zero KL. Applying the measurable map $\mathbf{1}_E$ and the data-processing inequality gives the binary-KL bound. If $\mathbb{E}_A\tau = \infty$, every lower bound below is automatic. □

H Strengthened finite-confidence rank-two star lower bound

Put $D = n + m - 2$ and $\kappa = \mathbf{1}\{n > 1\} + \mathbf{1}\{m > 1\}$.

Theorem H.1 (Disjoint-answer star)**.** *Let $n, m \geq 1$, $0 < \varepsilon \leq 1/8$, and $0 < \delta < 1/2$. If $D = 0$, then $\mathcal{M}^{\mathrm{G}}_{1,1}(\varepsilon, \delta) = 0$. If $D \geq 1$, then*

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon, \delta) \geq \frac{D}{18\kappa\varepsilon^2} \operatorname{kl}\left(\frac{\delta}{D}, 1 - \delta\right).$$

All matrices in the nontrivial construction have rank at most two.

Proof. If $D = 0$, both strategy simplices are singletons, so the unique pair is an exact equilibrium without observations. Now fix $D \geq 1$, an arbitrary globally (ε, δ)-PAC-Nash policy π, and $M_\pi = \sup_A \mathbb{E}^\pi_A \tau$. If $M_\pi = \infty$, the claim is automatic.

For an additive matrix $A = r\mathbf{1}^\top + \mathbf{1}c^\top$,

$$R_A(x, y) = \max_i r_i - x^\top r, \qquad C_A(x, y) = c^\top y - \min_j c_j,$$

and $\operatorname{rank}(A) \leq 2$. Let $r_1 = c_1 = 0$, $r_i = -4\varepsilon$ for $i > 1$, and $c_j = 4\varepsilon$ for $j > 1$. The base matrix is $A^0_{ij} = r_i + c_j$. Its unique strict equilibrium is (e_1, e_1), and the two displayed regret identities show that every base-correct output obeys $x_1, y_1 \geq 3/4$.

For every $i > 1$, define a row alternative A^i by changing r_i from -4ε to 2ε. Its row-regret inequality is

$$2\varepsilon - \left(2\varepsilon x_i - 4\varepsilon \sum_{k \notin \{1,i\}} x_k\right) \leq \varepsilon,$$

so correctness under A^i implies

$$x_i \geq \frac{1}{2} + 2 \sum_{k \notin \{1,i\}} x_k, \qquad y_1 \geq \frac{3}{4}.$$

In particular $x_1 \leq 1/2$. For every $j > 1$, define a column alternative by changing c_j from 4ε to -2ε. The symmetric calculation gives

$$y_j \geq \frac{1}{2} + 2 \sum_{\ell \notin \{1,j\}} y_\ell, \qquad x_1 \geq \frac{3}{4},$$

and $y_1 \leq 1/2$.

All entries lie in $[-6\varepsilon, 6\varepsilon] \subset [-1, 1]$, every alternative remains additive, and hence every hard matrix has rank at most two. Let $C_h = \mathcal{C}^{\mathrm{ind}}_\varepsilon(A^h)$ be the correct-output region of alternative h. Two distinct row regions cannot both satisfy their displayed mass inequalities. The same holds for two column regions. A row region has $x_1 \leq 1/2, y_1 \geq 3/4$, whereas a column region has $x_1 \geq 3/4, y_1 \leq 1/2$. Thus the D regions C_h are pairwise disjoint, and each is disjoint from $\mathcal{C}^{\mathrm{ind}}_\varepsilon(A^0)$.

Write P_h for the stopped law under A^h and $p_h = P_0(C_h)$. Global correctness gives $P_h(C_h) \geq 1 - \delta$, while base correctness and disjointness give $\sum_h p_h \leq \delta$. Alternative h differs from the base by 6ε on one full row or column. If N_h counts queries to that row or column, Lemma G.1 and binary data processing first give

$$18\varepsilon^2 \mathbb{E}_0 N_h \geq \operatorname{kl}(p_h, P_h(C_h)) \geq \operatorname{kl}(p_h, 1 - \delta).$$

The second inequality uses $p_h \leq \delta < 1 - \delta \leq P_h(C_h)$ and monotonicity in the second argument. The function $p \mapsto \operatorname{kl}(p, 1 - \delta)$ is convex and decreasing on $[0, \delta]$. Therefore Jensen's inequality and $\sum_h p_h \leq \delta$ yield

$$18\varepsilon^2 \sum_h \mathbb{E}_0 N_h \geq D \operatorname{kl}(\delta/D, 1 - \delta).$$

Every cell is in at most one shifted row and one shifted column, so $\sum_h N_h \leq \kappa\tau$. Hence

$$M_\pi \geq \mathbb{E}_0 \tau \geq \frac{D}{18\kappa\varepsilon^2} \operatorname{kl}(\delta/D, 1 - \delta).$$

Infimizing over all globally correct π completes the proof. □

Appendix table of contents

For $\delta \le 1/4$, this implies $\mathcal{M}^{\mathrm{G}}_{n,m} \ge cD\varepsilon^{-2}\log(1/\delta)$. Unlike $\mathrm{kl}(1-\delta,\delta)$, the displayed expression also records how disjoint alternative answer regions share the base error budget.

I Dense-good-layer Gaussian lower bound

The next theorem is the main new obstruction. It uses a code only to make two adjacent layers simultaneously decodable. The hard matrices themselves may be degenerate and need not have unique equilibria.

Theorem I.1 (Dense sign code on adjacent Johnson layers)**.** *Let integers N, s, d satisfy*

$$N \ge 2s \ge 4, \qquad d = \lceil 128 s \log(128N) \rceil .$$

Suppose $n \ge 2N$, $m \ge 2d+1$, $0 < \varepsilon \le \min\{1/16, 1/(16s)\}$, and $0 < \delta < 1/2$. Put $r = \lfloor s/2 \rfloor$. Then

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta) \ge \frac{15}{512\varepsilon^2}\frac{(N-s+1)r}{s}\,\mathrm{kl}\left(1-\delta,\frac{\delta}{r}\right) \ge \frac{5(N-s+1)}{512\varepsilon^2}\,\mathrm{kl}\left(1-\delta,\frac{\delta}{r}\right).$$

The same statement holds after swapping n, m.

Proof. Fix an arbitrary globally (ε,δ)-PAC-Nash policy π, and put $M_\pi = \sup_A \mathbb{E}^\pi_A \tau$. If $M_\pi = \infty$, there is nothing to prove.

A deterministic code with two dense good layers. Draw $c_1, \dots, c_N$ independently and uniformly from $\{-1,1\}^d$. For $U \subset [N]$ of size $s-1$ or s, define

$$Z_i(U) = \frac{1}{d}\sum_{\ell \in U\setminus\{i\}} \langle c_i, c_\ell \rangle,$$

where, if $i \notin U$, the sum is over all $\ell \in U$. Call U good when $\max_i |Z_i(U)| \le 1/8$. Conditional on c_i, the variable $dZ_i(U)$ is a sum of at most ds independent Rademacher signs. Hoeffding's inequality and a union bound give

$$\mathbb{P}\{U \text{ is bad}\} \le 2N\exp\left(-\frac{d}{128s}\right) \le \frac{1}{64}.$$

Therefore some fixed realization of the code has bad-vertex fractions B_{s-1}, B_s satisfying $B_{s-1} + B_s \le 1/32$.

Consider the bipartite inclusion graph between $\mathcal{L} = \binom{[N]}{s-1}$ and $\mathcal{R} = \binom{[N]}{s}$. Its edge count is

$$E = |\mathcal{L}|(N-s+1) = |\mathcal{R}|s.$$

Edges touching bad vertices are at most $(B_{s-1}+B_s)E$, so at least $31E/32$ edges are good–good. It follows that at least $15|\mathcal{R}|/16$ good right vertices have at least $s/2$ good deletion neighbors: otherwise the good–good edge count would be less than $|\mathcal{R}|s/2 + (15|\mathcal{R}|/16)s/2 = 31E/32$. For each such right vertex S, select $r = \lfloor s/2 \rfloor$ distinct good neighbors $T = S \setminus \{i\}$.

Games and a low-value witness. Let $h = 8\varepsilon$. Core rows are (i,σ), $i \in [N]$, $\sigma \in \{-1,1\}$. Core columns are a reservoir 0 and (q,b), $q \in [d]$, $b \in \{-1,1\}$. Define

$$a_i(0) = 0, \qquad a_i(q,b) = \frac{b c_{iq}}{2}, \qquad A^U_{(i,\sigma),j} = \sigma\big(a_i(j) - h\mathbf{1}\{i \in U\}\big).$$

Pad extra rows by zero and extra columns by copies of the reservoir column. Since $|a_i(j)| \le 1/2$ and $h \le 1/2$,

$$|a_i(j) - h\mathbf{1}\{i \in U\}| \le 1.$$

Thus every core entry is in $[-1,1]$, and reservoir/zero padding preserves the box.

For a column strategy y, set

$$u_i(y) = \sum_j a_i(j) y_j, \qquad F_U(y) = \max_i |u_i(y) - h\mathbf{1}\{i \in U\}|.$$

Appendix table of contents

Because rows occur in opposite pairs and extra rows are zero, $F_U(y) = \max_k e_k^\top A^U y$. In particular the minimax theorem gives $V^\star_{A^U} = \min_y F_U(y)$, since the strategy spaces are finite simplexes. If U is good, let

$$z_q = \frac{2h}{d} \sum_{\ell \in U} c_{\ell q}.$$

The bound $\sum_q |z_q| \le 2h|U| \le 2hs \le 1$ makes z realizable by setting $y_{q,+} = z_q^+$, $y_{q,-} = z_q^-$, and putting the residual mass on the reservoir. Then

$$u_i(y) = h\mathbf{1}\{i \in U\} + hZ_i(U),$$

so $F_U(y) \le h/8 = \varepsilon$. As $F_U \ge 0$, the game value $V^\star_{A^U}$ lies in $[0, \varepsilon]$.

Every correct output decodes the layer vertex. For arbitrary x, let $g_U(x) = \min_j x^\top A^U e_j$. Weak duality gives $g_U(x) \le V^\star_{A^U}$. If (x, y) has both individual regrets at most ε, then

$$F_U(y) - g_U(x) = R_{A^U}(x, y) + C_{A^U}(x, y) \le 2\varepsilon,$$

and hence $F_U(y) \le V^\star_{A^U} + 2\varepsilon \le 3\varepsilon$. Thus

$$i \in U \Rightarrow u_i(y) \ge h - 3\varepsilon = 5\varepsilon, \qquad i \notin U \Rightarrow u_i(y) \le 3\varepsilon.$$

The deterministic decoder

$$\mathcal{D}(y) = \{i : u_i(y) > 4\varepsilon\}$$

therefore equals U on every correct output. This argument uses no selected equilibrium and remains valid for nonunique or degenerate equilibria.

Change of measure across selected edges. For each selected left vertex T, let $E_T = \{\mathcal{D}(\widehat{y}) = T\}$. Correctness under A^T gives $P_T(E_T) \ge 1 - \delta$. Fix a high-degree S and write $p_T = P_S(E_T)$ over its selected deletion neighbors. These events are disjoint and are all errors under S, so $\sum_T p_T \le \delta$.

If $T = S \setminus \{i\}$, the two games differ by absolute mean h only on the row pair $(i, +), (i, -)$, in every column. Let N_i count queries to that row pair. By Lemma G.1, data processing, and monotonicity of binary KL,

$$\frac{h^2}{2} \mathbb{E}_T N_i = \mathrm{KL}(P_T \| P_S) \ge \mathrm{kl}(1 - \delta, p_T).$$

Convexity and $\sum_T p_T \le \delta$ imply

$$\frac{h^2}{2} \sum_{T \text{ selected below } S} \mathbb{E}_T N_i \ge r\, \mathrm{kl}(1 - \delta, \delta/r).$$

Sum this inequality over the at least $15|\mathcal{R}|/16$ high-degree S's. Reindexing the selected edges by their left endpoints gives

$$\sum_{\text{selected } (T, S = T \cup \{i\})} \mathbb{E}_T N_i = \sum_{T \in \mathcal{L}} \mathbb{E}_T \left[\sum_{i : (T, T \cup \{i\}) \text{ selected}} N_i \right] \le \sum_{T \in \mathcal{L}} \mathbb{E}_T \tau \le |\mathcal{L}| M_\pi,$$

because the selected additions to a fixed T use distinct row pairs. Therefore

$$\frac{h^2}{2} |\mathcal{L}| M_\pi \ge \frac{15}{16} |\mathcal{R}| r\, \mathrm{kl}(1 - \delta, \delta/r).$$

Use $|\mathcal{R}|/|\mathcal{L}| = (N - s + 1)/s$ and $h^2 = 64\varepsilon^2$. Since $r/s \ge 1/3$, both displayed bounds follow for M_π. Infimize over globally correct π.

Finally set $B = -A^\top$. A query to B_{ji} is simulated by querying A_{ij} and negating the returned observation, which preserves independent unit Gaussian noise. If (x, y) is an output for A, then (y, x) for B satisfies

$$R_B(y, x) = C_A(x, y), \qquad C_B(y, x) = R_A(x, y).$$

Applying the simulation in both directions gives $\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon, \delta) = \mathcal{M}^{\mathrm{G}}_{m,n}(\varepsilon, \delta)$, proving the transposed statement. □

I.1 Dimension-optimized form

Define

$$u_{n|m} = \min\left\{n, \varepsilon^{-1}, \frac{m}{\log(en)}\right\}, \qquad u_{m|n} = \min\left\{m, \varepsilon^{-1}, \frac{n}{\log(em)}\right\}.$$

Corollary I.2 (Optimized full-box lower bound)**.** *There is a universal $c > 0$ such that, for all $n, m \geq 1$ not both one, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$,*

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon, \delta) \geq \frac{c}{\varepsilon^2}\left[(n + m - 2)\log\frac{1}{\delta} + n\log(e + u_{n|m}) + m\log(e + u_{m|n})\right].$$

Proof. Put

$$B_0 = \varepsilon^{-2}(n + m - 2)\log(1/\delta),$$
$$B_n = \varepsilon^{-2} n \log(e + u_{n|m}),$$
$$B_m = \varepsilon^{-2} m \log(e + u_{m|n}).$$

For $D \geq 1$ and $\delta \leq 1/4$, monotonicity of binary KL and direct expansion give

$$\mathrm{kl}(\delta/D, 1 - \delta) \geq \mathrm{kl}(\delta, 1 - \delta) = (1 - 2\delta)\log\frac{1 - \delta}{\delta} \geq \frac{1}{4}\log\frac{1}{\delta}.$$

Thus Theorem H.1 gives $\mathcal{M}^{\mathrm{G}}_{n,m} \geq cB_0$.

For one orientation, let $N = \lfloor n/2 \rfloor$ and $s = \lfloor c_0 u_{n|m} \rfloor$, with $c_0 = 1/8192$. Suppose first that $s \geq 2$. Then $s \leq c_0 n$, and $s \geq 2$ implies $n \geq 2/c_0$. Consequently

$$N \geq n/2 - 1 \geq 2s.$$

Also $s \leq c_0/\varepsilon \leq 1/(16\varepsilon)$, and $s \geq 2$ itself forces $\varepsilon < 1/16$. Since $\log(128N) \leq 5\log(en)$ and $s \leq c_0 m/\log(en)$,

$$2\lceil 128 s \log(128N)\rceil + 1 \leq 1280 c_0 m + 3 = \frac{5}{32}m + 3 \leq m.$$

The last inequality holds because $s \geq 2$, $u_{n|m} \leq m/\log(en)$, and hence $m \geq 2/c_0$. Finally

$$N - s + 1 \geq (1/2 - c_0)n \geq n/3.$$

All hypotheses of Theorem I.1 are therefore met. Put $r = \lfloor s/2 \rfloor$ and $L = \log(r/\delta)$. Let $H(a) = -a\log a - (1 - a)\log(1 - a)$ denote binary entropy. Since $r \geq 1$, $\delta \leq 1/4$, and $H(\delta) \leq \log 2$,

$$\mathrm{kl}(1 - \delta, \delta/r) \geq (1 - \delta)\log(r/\delta) - H(\delta) \geq \frac{3}{4}L - \log 2 \geq \frac{1}{4}L.$$

Moreover $s = \lfloor c_0 u_{n|m} \rfloor$ implies

$$u_{n|m} < \frac{s + 1}{c_0} \leq \frac{4r}{c_0},$$

and therefore $\log(e + u_{n|m}) \leq C\log(r/\delta)$. The second bound in Theorem I.1 now yields $\mathcal{M}^{\mathrm{G}}_{n,m} \geq cB_n$.

If $s < 2$, then $u_{n|m} < 2/c_0$, so $\log(e + u_{n|m})$ is bounded by a universal constant. For every $(n, m) \neq (1, 1)$, $D = n + m - 2 \geq n/2$. Since $\log(1/\delta) \geq \log 4$, the already proved star estimate gives $\mathcal{M}^{\mathrm{G}}_{n,m} \geq cB_n$ in this case as well. Applying the same argument to $-A^\top$ proves $\mathcal{M}^{\mathrm{G}}_{n,m} \geq cB_m$.

Thus the minimax quantity is at least a universal constant times each of B_0, B_n, B_m, and

$$\mathcal{M}^{\mathrm{G}}_{n,m} \geq c\max\{B_0, B_n, B_m\} \geq \frac{c}{3}(B_0 + B_n + B_m).$$

This gives the additive display without an invalid direct-sum construction. □

J A Superlinear Lower Bound at Fixed Accuracy

The additive witness in Theorem I.1 spends probability mass proportional to the layer size. A different witness removes that restriction: average columns whose signs are positive on every selected row. This uses more columns per layer, but permits a growing layer at fixed accuracy. The stopped change-of-measure argument still decodes every correct answer.

Theorem J.1 (Intersection witnesses on adjacent layers)**.** *Let $N \geq 2s \geq 4$ be integers and put*

$$d = \lceil 256\, 2^s \log(256N) \rceil, \qquad r = \lfloor s/2 \rfloor.$$

If $n \geq 2N$, $m \geq d+1$, $0 < \varepsilon \leq 1/16$, and $0 < \delta < 1/2$, then

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta) \geq \frac{15}{512\varepsilon^2} \frac{(N-s+1)r}{s} \operatorname{kl}\left(1-\delta, \frac{\delta}{r}\right) \geq \frac{5(N-s+1)}{512\varepsilon^2} \operatorname{kl}\left(1-\delta, \frac{\delta}{r}\right).$$

The transposed statement also holds. Unlike Theorem I.1, this theorem places no $O(1/\varepsilon)$ upper bound on s.

Proof. **A fixed public dictionary with dense good layers.** Draw independent signs $c_{ij} \in \{-1,1\}$ for $i \in [N]$, $j \in [d]$. For $U \subset [N]$ of size $s-1$ or s, let

$$J_U = \{j : c_{ij} = 1 \text{ for every } i \in U\}, \qquad K_U = |J_U|, \qquad \mu_U = d\, 2^{-|U|}.$$

Call U good when $K_U \geq \mu_U/2$ and

$$\max_{i \notin U} \left| \frac{1}{K_U} \sum_{j \in J_U} c_{ij} \right| \leq \frac{1}{8}.$$

Here K_U is binomial and $\mu_U \geq 256 \log(256N)$. The multiplicative Chernoff bound gives $\mathbb{P}\{K_U < \mu_U/2\} \leq \exp(-\mu_U/8) \leq (256N)^{-32}$. Conditionally on all signs on rows in U, the signs on each row outside U remain independent fair signs, including those indexed by J_U. On $K_U \geq \mu_U/2$, Hoeffding's inequality and a union bound therefore give

$$\mathbb{P}\{U \text{ is bad}\} \leq (256N)^{-32} + 2N \exp(-\mu_U/256) \leq (256N)^{-32} + \frac{1}{128} \leq \frac{1}{64}.$$

Let B_{s-1} and B_s be the bad fractions in the two complete layers. Their expected sum is at most $1/32$, so a deterministic realization of the dictionary has $B_{s-1} + B_s \leq 1/32$. Fix this realization before choosing the unknown game or the learner's random observations. The dictionary may be supplied to the learner. Hiding it is not needed.

Games and witnesses of small value. Put $h = 8\varepsilon$. Use rows (i,σ) with $i \in [N]$ and $\sigma \in \{-1,1\}$, and columns $0, 1, \ldots, d$. Define public scores $a_i(0) = 0$, $a_i(j) = c_{ij}/2$ for $j \geq 1$, and games

$$A^U_{(i,\sigma),j} = \sigma\big(a_i(j) - h\mathbf{1}\{i \in U\}\big).$$

Pad unused rows by zero and unused columns by copies of column 0. Since $h \leq 1/2$, all entries lie in $[-1,1]$. For a good U, assign total mass $2h$ uniformly to J_U and mass $1-2h$ to column 0. This is a probability distribution, regardless of s. Its public scores satisfy

$$u_i(y) := \sum_j a_i(j) y_j = h \quad (i \in U), \qquad |u_i(y)| \leq h/8 = \varepsilon \quad (i \notin U).$$

Thus, with $F_U(y) = \max_i |u_i(y) - h\mathbf{1}\{i \in U\}|$, the value $V^\star_{A^U} = \min_y F_U(y)$ lies in $[0,\varepsilon]$.

For *any* pair (x,y) satisfying both regret inequalities, weak duality gives

$$F_U(y) \leq V^\star_{A^U} + R_{A^U}(x,y) + C_{A^U}(x,y) \leq 3\varepsilon.$$

Consequently, $u_i(y) \geq 5\varepsilon$ for $i \in U$ and $u_i(y) \leq 3\varepsilon$ for $i \notin U$. The public decoder $\mathcal{D}(y) = \{i : u_i(y) > 4\varepsilon\}$ equals U on every correct output, including outputs outside the witness support.

Stopped information across the two layers. Let $\mathcal{L} = \binom{[N]}{s-1}$ and $\mathcal{R} = \binom{[N]}{s}$. The inclusion graph has $E = |\mathcal{L}|(N-s+1) = |\mathcal{R}|s$ edges. At most $E/32$ touch a bad vertex. Hence at least $15|\mathcal{R}|/16$

Appendix table of contents

right vertices are good and have at least $s/2$ good deletion neighbors. For each such S, select r good neighbors $T = S \setminus \{i\}$.

Fix a globally correct policy with finite worst-case expected cost M_π. Write $E_T = \{\mathcal{D}(\widehat{y}) = T\}$. Under A^T, $P_T(E_T) \geq 1 - \delta$. For the selected neighbors of a fixed S, the events E_T are disjoint errors under A^S, so $\sum_T p_T \leq \delta$, where $p_T = P_S(E_T)$. Adjacent matrices differ by magnitude h precisely on row pair i. If N_i counts observations on that pair, Lemma G.1 yields

$$\frac{h^2}{2}\mathbb{E}_T N_i \geq \mathrm{kl}(1-\delta, p_T).$$

Convexity in p_T and the shared error budget imply

$$\frac{h^2}{2}\sum_{T \text{ selected below } S} \mathbb{E}_T N_i \geq r\ \mathrm{kl}(1-\delta, \delta/r).$$

Summing over the selected right vertices and reindexing by T, each observation is charged at most once at a given left vertex. Therefore

$$\frac{h^2}{2}|\mathcal{L}|M_\pi \geq \frac{15}{16}|\mathcal{R}|r\ \mathrm{kl}(1-\delta, \delta/r).$$

Substitute $h^2 = 64\varepsilon^2$ and the layer-cardinality ratio. Since $r/s \geq 1/3$, both claimed inequalities follow. Infinite expected cost needs no argument. Infimizing over policies proves the result. Transposition with sign reversal is the same exact observation-preserving map used in Theorem I.1. □

Corollary J.2 (Fixed-accuracy discovery and class separation)**.** *For universal $c, C > 0$, every integer $k \geq 2$, $0 < \varepsilon \leq 1/16$, and $0 < \delta \leq 1/4$ satisfy*

$$ck\varepsilon^{-2}\left[\log\log(ek) + \log\frac{1}{\delta}\right] \leq \mathcal{M}^{\mathrm{G}}_{k,k}(\varepsilon, \delta) \leq Ck\varepsilon^{-2}\left[\log(ek) + \log\frac{1}{\delta}\right].$$

In particular, for each fixed $r \geq 1$, $\varepsilon_0 \in (0, 1/16]$, and $\delta_0 \in (0, 1/4]$,

$$\Omega_{r,\varepsilon_0,\delta_0}(\log\log k) \leq \frac{\mathcal{M}^{\mathrm{G}}_{k,k}(\varepsilon_0, \delta_0)}{\mathcal{M}^{\mathrm{G}}_{\leq r}(k, k, \varepsilon_0, \delta_0)} \leq O_{r,\varepsilon_0,\delta_0}(\log k) \qquad (k \to \infty).$$

Thus the optimal class complexities are strictly separated at fixed accuracy and confidence, including for rank two. The unrestricted rate itself is not yet matched.

Proof. For $k \geq 2^{40}$, take $N = \lfloor k/2 \rfloor$ and $s = \lfloor (\log_2 k)/2 \rfloor$. Then $N \geq 2s$, $2^s \leq \sqrt{k}$, and

$$d + 1 \leq 256\sqrt{k}\log(128k) + 2 \leq k.$$

The last inequality holds at 2^{40} and thereafter because $\log(128k)/\sqrt{k}$ decreases. Also $N - s + 1 \geq k/3$, $r = \Theta(\log k)$, and, for $\delta \leq 1/4$,

$$\mathrm{kl}(1-\delta, \delta/r) \geq \tfrac{3}{4}\log(r/\delta) - \log 2 \geq \tfrac{1}{4}\log(r/\delta).$$

Apply Theorem J.1. For $2 \leq k < 2^{40}$, $\log\log(ek)$ is bounded by a universal constant, so Theorem H.1 gives the same lower bound after reducing its universal constant. The upper bound follows from Equation (5). For fixed $(r, \varepsilon_0, \delta_0)$, Theorem 4.1 and the rank-one lower bound give $\mathcal{M}^{\mathrm{G}}_{\leq r}(k, k, \varepsilon_0, \delta_0) = \Theta_{r,\varepsilon_0,\delta_0}(k)$. Dividing proves both comparisons. □

For binary feedback, the same lower-bound order holds when $0 < \varepsilon \leq 1/32$. All construction means then lie in $[-3/4, 3/4]$. If adjacent means differ by h, their Bernoulli parameters lie in $[1/8, 7/8]$ and differ by $h/2$, so

$$\mathrm{kl}(p, q) \leq \frac{(p-q)^2}{q(1-q)} \leq \frac{16}{7}h^2.$$

Replacing $h^2/2$ by $(16/7)h^2$ in the stopped information bound changes only its universal constant. This argument uses the binary likelihood directly and does not infer a binary lower bound from a Gaussian simulator.

K Low-Rank Probability Primitives

The structural algorithms repeatedly need a near-best action, rather than accurate estimates for all actions. PAC extremum search performs this task with linear action dependence and no additional arm-count logarithm.

Lemma K.1 (PAC maximum and minimum)**.** *Let d arms have fixed means $\mu_1, \dots, \mu_d$ and centered observations with subgaussian proxy at most σ^2. For $0 < q \le 1$, $\sigma^2 \ge 1$, and $\alpha \in (0, 1/4)$, there is a deterministically capped routine that returns $\widehat{i}$ such that*

$$\mu_{\widehat{i}} \ge \max_i \mu_i - q$$

with probability at least $1 - \alpha$, using at most

$$C\sigma^2 d q^{-2} \log \frac{c}{\alpha}$$

pulls. Negating observations gives the minimum version.

Proof. Median elimination gives the claim after rescaling by the subgaussian proxy. The deterministic cap follows from fixing the elimination schedule in advance. □

The following absolute-extremum lemma is required both by the rank-one and fixed-rank arguments.

Lemma K.2 (PAC absolute maximum)**.** *Under the assumptions of Lemma K.1, there is a deterministically capped routine that returns $\widehat{i}$ satisfying*

$$\left|\mu_{\widehat{i}}\right| \ge \max_i |\mu_i| - q$$

with probability at least $1 - \alpha$, using at most

$$C\sigma^2 d q^{-2} \log \frac{c}{\alpha}$$

pulls.

Proof. Run a $q/4$-near maximum routine and a $q/4$-near minimum routine, each with risk $\alpha/4$. Estimate the two returned means to accuracy $q/8$ with total risk $\alpha/2$. Return the arm with larger estimated absolute mean. On the simultaneous event, one returned arm has absolute mean at least $\max_i |\mu_i| - q/4$, and the final comparison loses at most $q/4$. The displayed q guarantee follows with slack. The scalar estimates are absorbed by the extremum cost. □

We will repeatedly sample linear scores from selected columns or rows.

Lemma K.3 (One-query linear score)**.** *Fix distinct column labels $J = (j_1, \dots, j_s)$ and a vector $v \in \mathbb{R}^s$. For row i, put $p_i = A_{i,J}$. There is a one-cell random observation $Z_i(v)$ with mean $p_i^\top v$ and centered subgaussian proxy at most $2\|v\|_1^2$. The symmetric statement holds for column scores from selected rows.*

Proof. If $v = 0$, return zero. Otherwise sample a with probability $|v_a| / \|v\|_1$, query (i, j_a), and return $\|v\|_1 \operatorname{sign}(v_a) X$. The conditional Gaussian term has proxy $\|v\|_1^2$. The randomized conditional mean lies in an interval of length $2\|v\|_1$, so Hoeffding's lemma contributes another proxy $\|v\|_1^2$. □

L Exact and Robust Rank-One Certification

The algorithm first tests a pure strategy pair. If that pair is rejected, the observed deviation identifies an informative pivot. Rank-one structure then reduces the remaining response search to scalar endpoints. We give the algorithm, prove the upper and lower bounds, and finally handle perturbations.

L.1 Statement

Theorem L.1 (Optimal anchor-free rank-one certification)**.** *There are universal constants $c, C > 0$ such that, for every $(n, m) \neq (1, 1)$, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$,*

$$c(n+m)\varepsilon^{-2} \log \frac{1}{\delta} \leq \mathcal{M}^{\mathrm{G}}_{\leq 1}(n, m, \varepsilon, \delta) \leq C(n+m)\varepsilon^{-2} \log \frac{c}{\delta}.$$

The upper-bound policy has a deterministic query cap. It is not given a factorization $A = uv^\top$, an equilibrium support, a nonzero entry, an anchor row or column, an incoherence parameter, or a conditioning parameter.

The exceptional 1×1 game has zero query complexity.

L.2 Sampling procedure

Put $t = \varepsilon/64$ and fix the initial pure pair $(i_0, j_0) = (1, 1)$. The procedure has fewer than sixteen grouped estimation events. Each receives risk at most $\delta/16$. Every event following an adaptively selected label uses observations that are fresh conditional on the complete transcript that selected that label. In particular, the three initial scalar estimates, the two absolute scans, and the final four-cell estimates use disjoint streams.

Algorithm L.1 (Rank-one certification without structural side information)**.**

1. *Use a t-PAC maximum on column j_0 to obtain i_R, and a t-PAC minimum on row i_0 to obtain j_C.*
2. *On fresh streams, estimate $A_{i_R j_0}$, $A_{i_0 j_C}$, and $A_{i_0 j_0}$ to accuracy t, writing $\bar{A}_{ij}$ for each fresh empirical mean. Form $\widehat{R} = \bar{A}_{i_R j_0} - \bar{A}_{i_0 j_0}$ and $\widehat{C} = \bar{A}_{i_0 j_0} - \bar{A}_{i_0 j_C}$. If both are at most $3\varepsilon/4$, return (e_{i_0}, e_{j_0}).*
3. *From a rejected comparison, select the larger-absolute-value endpoint of its two fresh estimates and call it the seed (a_0, b_0).*
4. *On mutually fresh streams, run a t-PAC absolute maximum down column b_0 to obtain a, then across row a to obtain b.*
5. *On pivot column b, find t-near maximum and minimum rows i_+, i_-. On pivot row a, find t-near maximum and minimum columns j_+, j_-.*
6. *Set $S = \{i_+, i_-\}$ and $T = \{j_+, j_-\}$. After all four labels are fixed, estimate the four cells of $A_{S,T}$ on new streams to accuracy t. Solve this estimated game exactly and embed its saddle strategies into the full simplexes.*

L.3 Proof of the upper bound

Proof of the upper bound in Theorem L.1. Condition on the simultaneous event that every PAC call and every scalar estimate satisfies its guarantee. A union bound makes this event have probability at least $1 - \delta$.

The initial branch either certifies or creates a seed. For the pure pair (e_{i_0}, e_{j_0}),

$$R_A = \max_i A_{ij_0} - A_{i_0 j_0}.$$

The selected row is t-near optimal and the two scalar estimates lose at most $2t$, so acceptance implies

$$R_A \leq \widehat{R} + 3t \leq \frac{3\varepsilon}{4} + \frac{3\varepsilon}{64} < \varepsilon.$$

The column inequality is symmetric.

Suppose instead that $\widehat{R} > 3\varepsilon/4$. Then

$$A_{i_R j_0} - A_{i_0 j_0} > \frac{3\varepsilon}{4} - 2t = \frac{46\varepsilon}{64}.$$

At least one of the two cells has magnitude larger than $23\varepsilon/64$. Choosing the larger estimated magnitude loses at most $2t$, so the selected seed obeys

$$|A_{a_0 b_0}| > \frac{21\varepsilon}{64}. \tag{8}$$

The same conclusion follows from a rejected column inequality.

Two scans manufacture a well-scaled pivot. If the structural branch is reached, the matrix is nonzero. Write $A = uv^\top$, and set

$$U = \max_i |u_i|, \qquad V = \max_j |v_j|.$$

Let $M_0 = \max_i |A_{ib_0}| = U\,|v_{b_0}|$. By Equation (8), $M_0 > 21t$. The first absolute scan gives

$$|A_{ab_0}| \geq M_0 - t,$$

which implies

$$\frac{|u_a|}{U} \geq 1 - \frac{t}{M_0} > \frac{20}{21}. \tag{9}$$

Let $Q = \max_j |A_{aj}| = |u_a|\,V$. Since $Q \geq |A_{ab_0}| > 20t$, the second scan returns b such that, with $c = A_{ab}$,

$$\frac{Q}{|c|} \leq \frac{20}{19}, \qquad \frac{\max_i |A_{ib}|}{|c|} = \frac{U}{|u_a|} \leq \frac{21}{20}. \tag{10}$$

Thus the pivot is automatically well scaled relative to both its full row and its full column. No lower bound on a pre-existing entry was assumed.

Rank-one scalarization and endpoint transfer. Define

$$p_i = A_{ib}, \qquad q_j = A_{aj}, \qquad c = A_{ab}.$$

Every 2×2 minor of a rank-one matrix vanishes, hence

$$A_{ij} = \frac{p_i q_j}{c}. \tag{11}$$

Let y be supported on T and put $\bar{q} = \sum_j y_j q_j$. If $\bar{q}/c \geq 0$, the t-near maximum endpoint on the pivot column gives

$$\max_i e_i^\top A y \leq \max_{s \in S} e_s^\top A y + t\,|\bar{q}/c|.$$

If $\bar{q}/c < 0$, use the near minimum endpoint. By Equation (10),

$$\max_i e_i^\top A y \leq \max_{s \in S} e_s^\top A y + \frac{20}{19} t. \tag{12}$$

The symmetric argument gives, for every x supported on S,

$$\min_j x^\top A e_j \geq \min_{t' \in T} x^\top A e_{t'} - \frac{21}{20} t. \tag{13}$$

Restricted solution. An exact saddle point of an entrywise t-accurate estimate of $A_{S,T}$ has true restricted gap at most $2t$. Combining this with Equations (12) and (13) yields

$$\mathrm{Gap}_A(\widehat{x}, \widehat{y}) \leq 2t + \frac{20}{19} t + \frac{21}{20} t < 5t < \varepsilon.$$

Both individual regrets are nonnegative and sum to the gap, so each is at most ε.

Queries. The procedure makes a constant number of PAC extrema over n or m Gaussian arms, each at accuracy $t = \Theta(\varepsilon)$, together with a constant number of scalar estimates. Lemmas K.1 and K.2 give the deterministic cap

$$C(n+m)\varepsilon^{-2} \log \frac{c}{\delta}.$$

□

L.4 Matching lower bound

Proof of the lower bound in Theorem L.1. Let $d = \max\{n, m\} \geq 2$. Suppose first that $d = n$. Consider matrices of the form $A = u\mathbf{1}_m^\top$, which have rank at most one. Under the base instance, put

$$u_1 = 0, \qquad u_i = -4\varepsilon \quad (i > 1).$$

For each $h > 1$, define an alternative by changing only u_h to 2ε. Let $\mathcal{C}_h$ be the set of outputs whose row regret is at most ε under alternative h. Correctness under the base forces $x_1 \geq 3/4$. Correctness under alternative h forces

$$x_h \geq \frac{1}{2} + 2 \sum_{\ell \notin \{1,h\}} x_\ell.$$

Hence the $d - 1$ regions $\mathcal{C}_h$ are pairwise disjoint and disjoint from the base-correct region.

Let P_0 and P_h denote the stopped laws and put $p_h = P_0(\mathcal{C}_h)$. Global correctness gives $P_h(\mathcal{C}_h) \geq 1-\delta$ and $\sum_{h>1} p_h \leq \delta$. The two instances differ by 6ε on row h. If N_h is the number of queries to that row, stopped Gaussian change of measure gives

$$18\varepsilon^2 \mathbb{E}_0 N_h \geq \mathrm{kl}(p_h, 1 - \delta).$$

The function $p \mapsto \mathrm{kl}(p, 1 - \delta)$ is convex and decreasing on $[0, \delta]$. Jensen's inequality therefore yields

$$18\varepsilon^2 \mathbb{E}_0 \tau \geq (d-1)\,\mathrm{kl}\left(\frac{\delta}{d-1}, 1 - \delta\right) \geq \frac{d-1}{4} \log \frac{1}{\delta}.$$

The last inequality holds for $\delta \leq 1/4$.

If $d = m$, the max–min roles prevent a literal transposition of the row family. Use instead $A = \mathbf{1}_n v^\top$. At the base instance, $v_1 = 0$ and $v_j = 4\varepsilon$ for $j > 1$. Alternative $h > 1$ changes only v_h to -2ε. Base correctness forces $y_1 \geq 3/4$, while correctness under alternative h forces

$$y_h \geq \frac{1}{2} + 2 \sum_{\ell \notin \{1,h\}} y_\ell.$$

These $d - 1$ alternative-correct regions are pairwise disjoint and disjoint from the base-correct region. The two instances differ by 6ε on column h, so the same stopped-change-of-measure and error-sharing calculation, with N_h now counting queries to that column, gives the displayed lower bound. Since $d - 1 \geq (n + m)/4$ for every nontrivial dimension pair, the claim follows. □

Corollary L.2 (Binary-rollout rank-one law)**.** *Define $\mathcal{M}_{\leq 1}^{\mathrm{Rad}}$ analogously to $\mathcal{M}_{\leq 1}^{\mathrm{G}}$ when a query to (i, j) returns $W \in \{-1, +1\}$ with mean A_{ij}. For every $(n, m) \neq (1, 1)$, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$,*

$$\mathcal{M}_{\leq 1}^{\mathrm{Rad}}(n, m, \varepsilon, \delta) = \Theta\left((n + m)\varepsilon^{-2} \log \frac{1}{\delta}\right).$$

The upper-bound policy is algorithm L.1, has a deterministic query cap, and uses no structural side information. The same conclusion holds for a Bernoulli rollout $X \in \{0, 1\}$ with $\mathbb{P}_A(X = 1 \mid i, j) = (1 + A_{ij})/2$.

Proof. Rademacher observations are centered subgaussian with a universal proxy. Every upper-bound step in algorithm L.1 is a PAC extremum or a fresh scalar estimate, so the proof of Theorem L.1 and its deterministic cap apply unchanged up to universal constants.

For the lower bound, put $d = \max\{n, m\}$. If $d = n$, use the exact-rank-one family $A = u\mathbf{1}_m^\top$. At the base instance take $u_1 = 0$ and $u_i = -3\varepsilon$ for $i > 1$. Alternative $h > 1$ changes only u_h to 3ε. Base correctness forces $x_1 \geq 2/3$, whereas correctness under alternative h forces

$$x_h \geq \frac{2}{3} + \sum_{\ell \notin \{1,h\}} x_\ell.$$

Thus the alternative-correct regions are pairwise disjoint and disjoint from the base-correct region. If p_h is the base probability of region h, then $\sum_{h>1} p_h \leq \delta$, while its alternative probability is at least

$1-\delta$. All means lie in $[-3/8, 3/8]$, and the changed row differs by 6ε. Hence Lemmas N.1 and N.2 give

$$72\varepsilon^2 \mathbb{E}_0 N_h \geq \mathrm{kl}(p_h, 1-\delta).$$

Summing and applying the same convexity and error-sharing calculation as in the Gaussian proof yields $\mathbb{E}_0\tau \geq c(d-1)\varepsilon^{-2}\log(1/\delta)$. If $d = m$, use $A = \mathbf{1}_n v^\top$ with base values $v_1 = 0$, $v_j = 3\varepsilon$ for $j > 1$, and change v_h to -3ε. The identical column argument applies. Since $d - 1 \geq (n+m)/4$, this proves the lower bound. Finally, $W = 2X - 1$ is a deterministic bijection between the stated Bernoulli and Rademacher observations, so it changes neither the transcript information nor the number of physical queries. □

L.5 Entrywise approximate-rank-one robustness

Theorem L.3 (Entrywise approximate-rank-one robustness)**.** *Let $n, m \geq 1$, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$. Let $A \in [-1,1]^{n\times m}$ and suppose that*

$$\inf_{\mathrm{rank}(B)\leq 1} \|A - B\|_\infty \leq \rho.$$

Given ρ (or any certified upper bound on it), there is a matrix-independent policy with query cap

$$C(n+m)\varepsilon^{-2}\log\frac{c}{\delta}$$

that returns $(\widehat{x}, \widehat{y})$ satisfying

$$R_A(\widehat{x}, \widehat{y}),\ C_A(\widehat{x}, \widehat{y}) \leq \varepsilon + 256\rho$$

with probability at least $1-\delta$, uniformly over the stated class. In particular, when $\rho \leq \varepsilon/16$, the policy is exactly algorithm L.1 run at target ε, does not use ρ, and returns a genuine ε-equilibrium of A. For larger ρ, the displayed additive guarantee uses the supplied upper bound on ρ to retune the internal accuracy.

Proof. For $\rho \leq \varepsilon/16$, Lemma Q.3 proves the stronger bound $\max\{R_A, C_A\} \leq 51\varepsilon/64$ for the original algorithm, without knowing ρ. Its proof tracks the magnitudes of the actual scanned cells and uses $|\,\mathrm{Gap}_A - \mathrm{Gap}_B\,| \leq 2\rho$.

For larger supplied $\rho < 1/128$, run that algorithm at $\xi = \max\{\varepsilon, 16\rho\} \leq 1/8$. The same lemma gives $R_A, C_A \leq 51\xi/64 \leq \varepsilon + 256\rho$. If $\rho \geq 1/128$, any pair has both regrets at most $2 \leq 256\rho$. As $\xi \geq \varepsilon$, the cap can only decrease. These are sufficient constants, not claimed thresholds for failure of the algorithm. □

Corollary L.4 (Unknown-radius rank-one tube)**.** *Let*

$$\mathcal{F}_{1,\rho} = \left\{A \in [-1,1]^{n\times m} : \mathrm{dist}_\infty(A, \{B : \mathrm{rank}(B) \leq 1\}) \leq \rho\right\}.$$

If $\rho \leq \varepsilon/16$, then the rank-one policy, without knowing ρ, is globally (ε, δ)-PAC-Nash over $\mathcal{F}_{1,\rho}$ and has the same query cap as in Theorem L.1. Since the exact rank-one class is a subclass of every such tube, its restricted minimax complexity has the same $\Theta((n+m)\varepsilon^{-2}\log(1/\delta))$ order.

M Fixed-Rank Certification

The proof tracks three properties of a selected submatrix: its dimension, its interpolation coefficients, and its determinant. A rejected candidate increases the dimension. Row replacement bounds the coefficients. Determinant control makes noisy inversion reliable. The sections below establish these properties before combining them with fresh verification and a common risk budget.

M.1 Main theorem

Theorem M.1 (Fixed-rank query bound without structural side information)**.** *Fix an integer $r \geq 1$. There are constants $C_r, c_r > 0$ and a matrix-independent, deterministically capped policy such that, for every $n, m \geq 1$, $0 < \varepsilon \leq 1/8$, $0 < \delta \leq 1/4$, and every $A \in [-1,1]^{n\times m}$ with $\mathrm{rank}(A) \leq r$, the policy returns $(\widehat{x}, \widehat{y})$ with*

$$R_A(\widehat{x}, \widehat{y}) \leq \varepsilon, \qquad C_A(\widehat{x}, \widehat{y}) \leq \varepsilon$$

Appendix table of contents

with probability at least $1-\delta$*. Its query cap obeys*

$$C_r(n+m)\varepsilon^{-(4r+6)}\log\frac{c_r}{\delta}. \tag{14}$$

A more informative uncompressed bound is given in Equation (18) below.

Corollary M.2 (Fixed-rank dimension-confidence law)**.** *For every fixed* $r \geq 1$ *and fixed* $\varepsilon \in (0, 1/8]$*, uniformly over all nontrivial dimension pairs and* $0 < \delta \leq 1/4$*,*

$$\mathcal{M}^{\mathrm{G}}_{\leq r}(n, m, \varepsilon, \delta) = \Theta_{r,\varepsilon}\left((n+m)\log\frac{1}{\delta}\right).$$

Proof. The upper bound is Theorem M.1. The rank-one one-sided family in the proof of Theorem L.1 belongs to every rank-at-most-r class and gives the matching lower dependence on $n+m$ and $\log(1/\delta)$. □

Algorithm M.1 (Verifier-driven dominant skeleton for rank at most r)**.** *1. Initialize* $I = J = \varnothing$*.*

2. *While no answer has been returned, set* $s = |I| = |J|$*. If* $s \geq 1$*, purify the current rows at fixed* J *until the skeleton is* 2*-row-dominant.*
3. *Build an* $\varepsilon/16$*-gap equilibrium* z *of the current CUR surrogate* B*, then run the fresh certify-or-witness verifier. If it accepts, return* z*.*
4. *If the verifier rejects and* $s = r$*, return a fixed pair. This branch is impossible on the simultaneous good event.*
5. *Otherwise convert the deviation witness into a screened residual pivot* (i^+, j^+)*. Append* i^+ *to* I *and* j^+ *to* J*, then return to Step 2. The value of* s *increases exactly once because it is always defined by the two tuple lengths.*

Every empirical inversion and residual append is guarded. A singular-value screen failure, inverse-norm or feature-norm screen failure, duplicate-label or singular-append event, pivot-screen failure, request for a (K_s+1)st row swap, or exhaustion of a public call cap causes immediate return of a fixed pair. These branches make the policy well-defined and deterministically capped on every transcript. The proof shows that none occurs on the simultaneous good event.

Public parameter card. The following list consolidates the numerical choices used by the four steps. Every PAC or fixed-sample call is instantiated with the deterministic cap from its stated lemma, and a request beyond the listed finite call schedule triggers the fixed-pair fallback.

- At size $s \geq 1$, set $L_s = s^{s-1}(16/\varepsilon)^s$, $N_s = (1 + 4096s/\varepsilon)^s$, and

$$K_s = \left\lceil \frac{\log\left(s^{s/2}(16/\varepsilon)^s\right)}{\log(5/4)} \right\rceil.$$

  Each of purification, candidate construction, verification, and residual pivoting receives risk $\alpha_\star = \delta/[4(r+1)]$ per stage.
- Purification uses inverse accuracy $1/(64\sqrt{s})$, PAC and scalar accuracy $1/32$, swap threshold $3/2$, and at most K_s executed swaps.
- Candidate construction uses sphere-net radius $\varepsilon/(2048s)$, directional accuracy $\varepsilon/(256\sqrt{s})$, feature accuracies $\varepsilon/(2^{14}sL_s)$ and $\varepsilon/(2^{14}s)$, and inverse accuracy $\varepsilon/(2^{14}s)$.
- The verifier uses PAC and scalar accuracy $\varepsilon/32$ and accepts only when both estimated regrets are at most $3\varepsilon/4$. Residual conversion uses feature accuracy $\varepsilon/(2^{14}sL_s)$, inverse accuracy $\varepsilon/(2^{14}s)$, PAC accuracy $\varepsilon/128$, final scalar accuracy $\varepsilon/128$, and pivot threshold $\varepsilon/8$. The $s = 0$ route uses the same PAC and scalar thresholds without inversions.

The subroutine lemmas below specify the sampling distributions, internal risk splits, norm screens, and freshness order. Thus algorithm M.1 together with this card is a finite composition of fully capped routines, even though its arithmetic and net enumeration are not claimed efficient.

M.2 Exact CUR residual identities

Let $I = (i_1, \dots, i_s)$ and $J = (j_1, \dots, j_s)$ be ordered distinct labels such that

$$C = A_{I,J}$$

is invertible. Define

$$p_i = A_{i,J} \in \mathbb{R}^s, \qquad q_j = A_{I,j} \in \mathbb{R}^s, \qquad B_{ij} = p_i^\top C^{-1} q_j, \qquad E = A - B.$$

Lemma M.3 (Skeleton residual and rank drop)**.** *The residual satisfies $E_{I,:} = 0$ and $E_{:,J} = 0$. Moreover,*

$$\operatorname{rank}(A) = s + \operatorname{rank}(E).$$

If $i \notin I$ and $j \notin J$, then the enlarged block

$$C^+ = \begin{pmatrix} C & q_j \\ p_i^\top & A_{ij} \end{pmatrix}$$

satisfies

$$\det(C^+) = \det(C) E_{ij}.$$

Proof. The interpolation identities on selected rows and columns are immediate. Block Gaussian elimination gives

$$\begin{pmatrix} I & 0 \\ -PC^{-1} & I \end{pmatrix} A \begin{pmatrix} I & -C^{-1}Q \\ 0 & I \end{pmatrix} = \begin{pmatrix} C & 0 \\ 0 & E_{I^c,J^c} \end{pmatrix},$$

which proves the rank identity. The determinant formula is the Schur complement identity. □

A sequence of residual pivots of size at least $\varepsilon/16$ protects the inverse.

Lemma M.4 (Determinant and inverse bound)**.** *Suppose an $s \times s$ skeleton is obtained from the empty skeleton by s Schur expansions, each with residual pivot magnitude at least $\varepsilon/16$. Any number of determinant-nondecreasing row swaps may be interleaved between successive expansions and may also follow the final expansion. Then*

$$|\det C| \geq \left(\frac{\varepsilon}{16}\right)^s, \qquad \left\|C^{-1}\right\|_2 \leq L_s := s^{s-1} \left(\frac{16}{\varepsilon}\right)^s.$$

Proof. Each Schur expansion multiplies the current determinant magnitude by its residual pivot, by Lemma M.3, while every interleaved row swap can only increase it. Induction over the expansion–swap sequence gives the displayed determinant lower bound. Hadamard's inequality gives $\|C\|_2 \leq \|C\|_F \leq s$. Since the product of singular values equals $|\det C|$,

$$\sigma_{\min}(C) \geq \frac{|\det C|}{\|C\|_2^{s-1}},$$

which gives the inverse bound. □

Lemma M.5 (Guarded inverse estimation)**.** *Let $C \in [-1,1]^{s \times s}$ be invertible with $\left\|C^{-1}\right\|_2 \leq L$ for $L \geq 1$, and let $0 < \zeta \leq 1$. By estimating every entry of C from fresh Gaussian samples, one can construct a matrix $\widehat{K}$ with the following guarded guarantee. With probability at least $1 - \alpha$,*

$$\left\|\widehat{K} - C^{-1}\right\|_2 \leq \zeta, \qquad \left\|\widehat{K}\right\|_2 \leq 2L.$$

The number of cell observations is at most

$$C s^4 L^4 \zeta^{-2} \log \frac{cs}{\alpha}.$$

The public screen is explicit. Let $\widehat{C}$ be the empirical block. The routine returns a failure symbol unless

$$\sigma_{\min}(\widehat{C}) \geq \frac{1}{2L}.$$

Only after this test does it form $\widehat{K} = \widehat{C}^{-1}$, and it returns failure unless $\left\|\widehat{K}\right\|_2 \leq 2L$. Thus the routine is total and never attempts an inversion after a failed singular-value screen.

Proof. Estimate each entry of C to accuracy

$$\gamma = \frac{\zeta}{8sL^2}.$$

On the simultaneous event, the spectral perturbation is at most $s\gamma = \zeta/(8L^2)$. Since $L \geq 1$ and $\zeta \leq 1$, this is at most $1/(8L)$. Weyl's inequality and $\sigma_{\min}(C) \geq 1/L$ give $\sigma_{\min}(\widehat{C}) \geq 7/(8L) > 1/(2L)$, so the first screen passes. The empirical matrix is invertible and $\left\|\widehat{C}^{-1}\right\|_2 \leq (1 - L\left\|\widehat{C} - C\right\|_2)^{-1}L < 2L$, so the second screen also passes. The resolvent identity gives

$$\left\|\widehat{C}^{-1} - C^{-1}\right\|_2 \leq \left\|C^{-1}\right\|_2 \left\|\widehat{C} - C\right\|_2 \left\|\widehat{C}^{-1}\right\|_2 \leq \frac{\zeta}{4} < \zeta.$$

Fixed-sample Gaussian concentration over the s^2 entries gives the displayed cost. Both screens therefore change behavior only outside the simultaneous event. □

M.3 Dominant-row purification

The main conditioning device is a noisy version of determinant-improving max-volume row swaps.

Definition M.6 (Row dominance)**.** For an invertible skeleton $C = A_{I,J}$, define

$$\alpha_i^\top = p_i^\top C^{-1}.$$

The skeleton is Λ-row-dominant when

$$\max_{i \in [n]} \|\alpha_i\|_\infty \leq \Lambda.$$

Lemma M.7 (Row replacement identity)**.** *Replacing the ath selected row of C by row i multiplies the determinant magnitude by $|\alpha_{i,a}|$.*

Proof. Since $p_i^\top = \alpha_i^\top C$, multilinearity of the determinant leaves only the coefficient of the replaced basis row. □

Lemma M.8 (Noisy dominant-row purification)**.** *Assume the determinant and inverse bounds of Lemma M.4 for a fixed column tuple J. There is a deterministically capped procedure that, with conditional probability at least $1 - \alpha$, returns a row tuple I' such that*

$$|\det A_{I',J}| \geq |\det A_{I,J}|, \qquad \max_i \left\|A_{i,J} A_{I',J}^{-1}\right\|_\infty \leq 2.$$

The number of row swaps is at most

$$K_s \leq Cs \log \frac{16s}{\varepsilon}.$$

Its query cost is bounded by

$$CK_s \left[s^5 L_s^4 + ns^2 L_s^2\right] \log \frac{cK_s s}{\alpha}.$$

Proof. At the beginning of each pass, invoke Lemma M.5 with $L = L_s$ and $\zeta = 1/(64\sqrt{s})$. Thus, on its good event, the inverse estimate $\widehat{K}$ satisfies

$$\left\|\widehat{K} - C^{-1}\right\|_2 \leq \frac{1}{64\sqrt{s}}, \qquad \left\|\widehat{K}\right\|_2 \leq 2L_s.$$

A failed numerical guard aborts the whole fixed-rank routine and returns a fixed pair. Freeze $\widehat{K}$ before every score scan in the pass. All score observations and the selected-score estimate are fresh conditional on this frozen inverse and the current skeleton.

For each coordinate $a \in [s]$, use Lemmas K.2 and K.3 to find an approximate absolute maximizer of

$$p_i^\top \widehat{K} e_a.$$

Use accuracy $1/32$, and estimate the returned score to accuracy $1/32$. If one estimated absolute score exceeds $3/2$, replace selected row a by the returned row, discard the inverse and every score estimate from that pass, and restart with a fresh inverse. The inverse bias is at most $1/64$, so every executed swap has

$$|\alpha_{i,a}| > \frac{5}{4}.$$

By Lemma M.7, each swap increases the determinant magnitude by at least $5/4$.

If a full pass executes no swap, the PAC and scalar errors imply

$$\max_i |\alpha_{i,a}| < 2$$

for every coordinate a. This is the desired dominance.

Every $s \times s$ block with entries in $[-1, 1]$ has determinant magnitude at most $s^{s/2}$. The initial determinant is at least $(\varepsilon/16)^s$. The implementation allows the public cap K_s of executed swaps and aborts only if a $(K_s + 1)$st swap is requested. On the simultaneous good event, more than

$$\frac{\log\left(s^{s/2}(16/\varepsilon)^s\right)}{\log(5/4)}$$

swaps are impossible on the simultaneous good event. This is $O(s \log(16s/\varepsilon))$.

The inverse-estimation stage contributes the $s^5 L_s^4$ term per pass. A score has proxy $O(sL_s^2)$ and constant target accuracy. Scanning s coordinates over n rows contributes $O(ns^2 L_s^2)$ per pass. The routine splits its input risk α uniformly across the at most $K_s + 1$ inverse calls, the $s(K_s + 1)$ PAC scans, and their scalar screens. This proves the conditional guarantee and the displayed cap. □

M.4 A low-dimensional candidate for the current CUR surrogate

Assume now that C is 2-row-dominant. Then

$$B_{ij} = \alpha_i^\top q_j, \qquad \|\alpha_i\|_2 \le 2\sqrt{s}, \qquad \|q_j\|_2 \le \sqrt{s}.$$

The conditioning appears only in the variance of row-score queries, not in the covering number.

Lemma M.9 (Dominant-skeleton surrogate equilibrium)**.** *Let $1 \le s \le r$ and suppose $C = A_{I,J}$ is invertible, satisfies Lemma M.4, and is 2-row-dominant. There is a deterministically capped procedure that returns $z = (x, y)$ with*

$$\mathrm{Gap}_B(z) \le \frac{\varepsilon}{16}$$

with conditional probability at least $1 - \alpha$. Put

$$N_s = \left(1 + \frac{4096s}{\varepsilon}\right)^s.$$

Its query cost is at most

$$C\left[(nL_s^2 + m)N_s s^4 \varepsilon^{-2} + N_s s^5 L_s^2 \varepsilon^{-2} + s^6 L_s^4 \varepsilon^{-2}\right] \log \frac{cN_s s}{\alpha}.$$

For $s = 0$, the zero surrogate is solved by any fixed pure pair without sampling.

Proof. Let $\mathcal{U}$ be a Euclidean θ-net of the unit sphere in $\mathbb{R}^s$, where

$$\theta = \frac{\varepsilon}{2048s}, \qquad |\mathcal{U}| \le N_s.$$

For every $u \in \mathcal{U}$, use separate PAC maximum and minimum calls to collect near extrema of $\alpha_i^\top u$ over rows and of $q_j^\top u$ over columns. The target true directional accuracy is

$$\eta = \frac{\varepsilon}{256\sqrt{s}}.$$

Invoke Lemma M.5 with target spectral error $\eta/(4\sqrt{s})$. Freeze the resulting $\widehat{K}$ before all directional scans. Row scores are then sampled as $p_i^\top \widehat{K} u$ through Lemma K.3. Since $\|p_i\|_2 \le \sqrt{s}$, the plug-in bias relative to $\alpha_i^\top u$ is at most $\eta/4$, uniformly over all rows and net directions. The row-score proxy is $O(sL_s^2)$. Column scores use selected-row entries directly and have proxy $O(s)$. The PAC target is reduced by a constant factor so that, after plug-in bias, the true directional loss is at most η. Let S, T be the collected labels. Then $|S|, |T| \le 2N_s$.

For every $v \in \mathbb{R}^s$, the net argument gives

$$\max_i \alpha_i^\top v \le \max_{a\in S} \alpha_a^\top v + \|v\|_2 \left(\eta + 4\sqrt{s}\,\theta\right),$$
$$\min_j q_j^\top v \ge \min_{b\in T} q_b^\top v - \|v\|_2 \left(\eta + 2\sqrt{s}\,\theta\right).$$

Indeed, compare an exact extremizer in direction v with the returned near extremizer in a net direction within θ of $v/\|v\|_2$.

If y is supported on T, then $\bar{q} = \sum_j y_j q_j$ has norm at most $\sqrt{s}$. Hence

$$\max_i e_i^\top B y \le \max_{a\in S} e_a^\top B y + \frac{\varepsilon}{64}.$$

If x is supported on S, then $\bar{\alpha} = \sum_i x_i \alpha_i$ has norm at most $2\sqrt{s}$, and

$$\min_j x^\top B e_j \ge \min_{b\in T} x^\top B e_b - \frac{\varepsilon}{64}.$$

The remaining step constructs the restricted matrix from only $O(N_s s)$ fresh feature estimates, rather than sampling its $O(N_s^2)$ entries. Set

$$\gamma_p = \frac{\varepsilon}{2^{14} s L_s}, \qquad \gamma_q = \frac{\varepsilon}{2^{14} s}, \qquad \zeta_B = \frac{\varepsilon}{2^{14} s}.$$

Conditionally on the selected label sets, estimate every coordinate of $p_a = A_{a,J}$, $a \in S$, to accuracy γ_p, every coordinate of $q_b = A_{I,b}$, $b \in T$, to accuracy γ_q, and invoke Lemma M.5 with target error ζ_B. Use fresh streams for all three collections. Denote the resulting objects by $\widetilde{p}_a$, $\widetilde{q}_b$, and $\widetilde{K}$, and form

$$\widetilde{B}_{ab} = \widetilde{p}_a^\top \widetilde{K} \widetilde{q}_b.$$

The public screens enforce $\left\|\widetilde{K}\right\|_2 \le 2L_s$ and $\|\widetilde{q}_b\|_2 \le 2\sqrt{s}$. On the simultaneous estimation event, row dominance and the decomposition

$$\widetilde{p}_a^\top \widetilde{K} \widetilde{q}_b - p_a^\top C^{-1} q_b = (\widetilde{p}_a - p_a)^\top \widetilde{K} \widetilde{q}_b + p_a^\top (\widetilde{K} - C^{-1}) \widetilde{q}_b + \alpha_a^\top (\widetilde{q}_b - q_b)$$

give

$$\left|\widetilde{B}_{ab} - B_{ab}\right| \le 4 s L_s \gamma_p + 2 s \zeta_B + 2 s \gamma_q$$
$$< \frac{\varepsilon}{1024}$$

uniformly over $S \times T$. Here the final term uses $\|\alpha_a\|_\infty \le 2$. Thus an exact saddle point of $\widetilde{B}_{S,T}$ has true restricted B-gap below $\varepsilon/512$. Adding the two envelope errors from the previous displays gives full surrogate gap below $\varepsilon/16$.

There are at most $2N_s s$ selected feature coordinates. Their fixed-sample costs are bounded by $CN_s s^3 L_s^2 \varepsilon^{-2} \log(cN_s s/\alpha)$, and the inverse estimate costs $Cs^6 L_s^4 \varepsilon^{-2} \log(cs/\alpha)$. These are dominated by the displayed bound. All feature estimates are made after the labels are selected and before the restricted saddle point is computed, so no post-selection data are reused. The inverse is frozen before all directional scans. The routine divides its input risk uniformly over the at most $4N_s$ PAC calls, the grouped feature estimates, and the guarded inverse estimates. All schedules are deterministic once the public parameters are fixed. □

M.5 Fresh verifier with a progress witness

Lemma M.10 (Certify or return a large true deviation)**.** *For any data-dependent candidate $z = (x, y)$ fixed before a fresh verification transcript, there is a deterministically capped verifier with the following simultaneous guarantee with probability at least $1 - \alpha$.*

(i) If it accepts, then $R_A(z) \le \varepsilon$ and $C_A(z) \le \varepsilon$.

(ii) If it rejects on the row side, it returns a row i satisfying

$$e_i^\top Ay - x^\top Ay \ge \frac{11\varepsilon}{16}.$$

(iii) If it rejects on the column side, it returns a column j satisfying

$$x^\top Ay - x^\top Ae_j \ge \frac{11\varepsilon}{16}.$$

The cost is

$$C(n+m)\varepsilon^{-2}\log\frac{c}{\alpha}.$$

Proof. Use $\varepsilon/32$-PAC maximum and minimum calls for the two deviation families. Estimate the selected row payoff, selected column payoff, and baseline payoff on disjoint streams to accuracy $\varepsilon/32$. Accept when both estimated regrets are at most $3\varepsilon/4$. Acceptance implies each true regret is at most $3\varepsilon/4 + 3\varepsilon/32 < \varepsilon$. A rejected estimated regret exceeds $3\varepsilon/4$, while its selected true deviation loses at most $2\varepsilon/32$, which gives $11\varepsilon/16$. □

M.6 Verifier rejection produces a Schur pivot

Lemma M.11 (Witness-to-residual-pivot)**.** *Assume C is a 2-row-dominant skeleton, B is its CUR surrogate, and $z = (x, y)$ satisfies $\mathrm{Gap}_B(z) \le \varepsilon/16$. Given either witness in Lemma M.10, there is a fresh, deterministically capped procedure that returns (i^+, j^+) such that*

$$\left|E_{i^+j^+}\right| \ge \frac{\varepsilon}{16}$$

with conditional probability at least $1 - \alpha$. For $s = 0$, its cost is at most

$$C(n+m)\varepsilon^{-2}\log\frac{c}{\alpha}.$$

For $s \ge 1$, its cost is at most

$$C(n+m)s^4L_s^2\varepsilon^{-2}\log\frac{cs}{\alpha} + Cs^6L_s^4\varepsilon^{-2}\log\frac{cs}{\alpha}.$$

Proof. When $s = 0$, the surrogate and residual are $B = 0$ and $E = A$. For a row witness, the numbers $a_i = e_i^\top Ay$ satisfy $a_{i_0} - \sum_i x_i a_i \ge 11\varepsilon/16$, hence $\max_i |a_i| \ge 11\varepsilon/32$. An $\varepsilon/128$-PAC absolute scan of these mixed-payoff arms, followed by an independent $\varepsilon/128$-PAC absolute scan of the cells in the selected row, finds a cell of magnitude greater than $\varepsilon/4$. Estimate the selected cell on a new stream to accuracy $\varepsilon/128$ and retain it only when the estimated magnitude exceeds $\varepsilon/8$. The screen passes on the simultaneous good event, and whenever it passes the true cell magnitude exceeds $\varepsilon/8 - \varepsilon/128 > \varepsilon/16$.

For a column witness, set $b_j = x^\top Ae_j$ and $\bar{b} = \sum_j y_j b_j$. Then $\bar{b} - b_{j_0} \ge 11\varepsilon/16$, which implies $\max_j |b_j| \ge 11\varepsilon/32$. First scan the column mixture arms $x^\top Ae_j$ in absolute value. After a column is selected, scan its cells over rows on a new stream, again in absolute value, and apply the same fresh $\varepsilon/128$ scalar estimate and $\varepsilon/8$ screen. The same thresholds yield an $\varepsilon/16$ pivot. The two routes use $O((n+m)\varepsilon^{-2}\log(c/\alpha))$ observations.

Assume henceforth that $s \ge 1$ and consider a row witness i_0. Let

$$r_i = e_i^\top Ey, \qquad \bar{r} = x^\top Ey.$$

Since the corresponding deviation in B is at most $R_B(z) \le \varepsilon/16$,

$$r_{i_0} - \bar{r} \ge \frac{11\varepsilon}{16} - \frac{\varepsilon}{16} = \frac{5\varepsilon}{8}.$$

If every $|r_i|$ were below $5\varepsilon/16$, then their difference from the convex average $\bar{r}$ would be below $5\varepsilon/8$. Therefore

$$\max_i |r_i| \ge \frac{5\varepsilon}{16}. \tag{15}$$

We first manufacture uniformly accurate row-residual arms. Put

$$q_y = \sum_j y_j q_j, \qquad v = C^{-1} q_y, \qquad \gamma_y = \frac{\varepsilon}{2^{14} s L_s}, \qquad \zeta_y = \frac{\varepsilon}{2^{14} s}.$$

For each selected row in I, estimate its payoff against the mixture y to accuracy γ_y, producing $\widetilde{q}_y$. Independently invoke Lemma M.5 with target error ζ_y, and put $\widetilde{v} = \widetilde{K}\widetilde{q}_y$. The public screens enforce $\|\widetilde{q}_y\|_2 \le 2\sqrt{s}$ and $\left\|\widetilde{K}\right\|_2 \le 2L_s$. On the simultaneous event,

$$\|\widetilde{v} - v\|_2 \le 2L_s\sqrt{s}\,\gamma_y + \zeta_y\sqrt{s} < \frac{\varepsilon}{4096\sqrt{s}}.$$

For row i, subtract an independent linear-score observation of $p_i^\top \widetilde{v}$ from a fresh sample of $e_i^\top A y$. The resulting arm has mean μ_i satisfying

$$|\mu_i - r_i| \le \|p_i\|_2 \, \|\widetilde{v} - v\|_2 < \frac{\varepsilon}{4096}$$

and has subgaussian proxy $O(s^2 L_s^2)$. An $\varepsilon/128$-PAC absolute maximum, together with Equation (15), returns a row i^+ such that

$$\left|e_{i^+}^\top E y\right| > \frac{19\varepsilon}{64}. \tag{16}$$

Next estimate the bounded interpolation vector $\alpha_{i^+} = p_{i^+}^\top C^{-1}$. Estimate the coordinates of p_{i^+} to accuracy

$$\gamma_\alpha = \frac{\varepsilon}{2^{14} s L_s}$$

and use a fresh inverse estimate with error $\zeta_\alpha = \varepsilon/(2^{14}s)$. Put $\widetilde{\alpha} = \widetilde{p}_{i^+}^\top \widetilde{K}$ and screen it to satisfy $\|\widetilde{\alpha}\|_\infty \le 3$. On the simultaneous event,

$$\|\widetilde{\alpha} - \alpha_{i^+}\|_1 \le 2sL_s\gamma_\alpha + s\zeta_\alpha < \frac{\varepsilon}{4096}.$$

For column j, subtract an independent linear-score observation of $\widetilde{\alpha}^\top q_j$ from one fresh observation of $A_{i^+ j}$. This arm has mean ν_j with

$$\left|\nu_j - E_{i^+ j}\right| < \frac{\varepsilon}{4096}$$

and proxy $O(s^2)$. Since an average cannot exceed the largest absolute summand, Equation (16) implies $\max_j |E_{i^+ j}| > 19\varepsilon/64$. A second $\varepsilon/128$-PAC absolute maximum therefore returns j^+ with

$$\left|E_{i^+ j^+}\right| > \frac{9\varepsilon}{32}.$$

Finally estimate ν_{j^+} on a new stream to accuracy $\varepsilon/128$ and keep the pivot only when the estimate has magnitude above $\varepsilon/8$. On the simultaneous event the screen necessarily passes. Conversely, whenever it passes on that event,

$$\left|E_{i^+ j^+}\right| > \frac{\varepsilon}{8} - \frac{\varepsilon}{128} - \frac{\varepsilon}{4096} > \frac{\varepsilon}{16}.$$

Thus the returned append is always a genuine Schur pivot on the good event.

For a column witness, write

$$c_j = x^\top E e_j, \qquad \bar{c} = x^\top E y.$$

The witness gives $\bar{c} - c_{j_0} \geq 5\varepsilon/8$, so

$$\max_j |c_j| \geq 5\varepsilon/16. \tag{17}$$

We now give the two scans explicitly because row dominance makes their variances different from those in the row-witness route.

Put

$$p_x^\top = \sum_i x_i p_i^\top = x^\top A_{:,J}, \qquad \bar{\alpha}^\top = p_x^\top C^{-1} = \sum_i x_i \alpha_i^\top,$$

and set $\gamma_x = \varepsilon/(2^{14} s L_s)$ and $\zeta_x = \varepsilon/(2^{14} s)$. Estimate every coordinate of p_x from fresh mixture samples. Independently estimate the inverse with error ζ_x, then freeze $\widetilde{\alpha}^\top = \widetilde{p}_x^\top \widetilde{K}$. Screen it to satisfy $\|\widetilde{\alpha}\|_\infty \leq 3$. Row dominance gives $\|\bar{\alpha}\|_\infty \leq 2$, and on the simultaneous event

$$\|\widetilde{\alpha} - \bar{\alpha}\|_1 \leq 2 s L_s \gamma_x + s\zeta_x < \frac{\varepsilon}{4096}.$$

For column j, draw a fresh row from x to sample $x^\top A e_j$ and subtract an independent linear-score observation of $\widetilde{\alpha}^\top q_j$. The resulting arm has mean $\widetilde{\nu}_j$ with

$$|\widetilde{\nu}_j - c_j| < \frac{\varepsilon}{4096}$$

and subgaussian proxy $O(s^2)$. An $\varepsilon/128$-PAC absolute maximum, together with Equation (17), returns j^+ with

$$\left|x^\top E e_{j^+}\right| > \frac{19\varepsilon}{64}.$$

After j^+ is fixed, estimate every coordinate of q_{j^+} on a new stream to accuracy $\gamma_v = \varepsilon/(2^{14} s L_s)$. Use a separate inverse estimate with error $\zeta_v = \varepsilon/(2^{14} s)$ and freeze $\widetilde{v} = \widetilde{K}\widetilde{q}_{j^+}$. The public screens enforce $\|\widetilde{q}_{j^+}\|_2 \leq 2\sqrt{s}$ and $\|\widetilde{K}\|_2 \leq 2L_s$. Hence

$$\|\widetilde{v} - C^{-1} q_{j^+}\|_2 \leq 2L_s\sqrt{s}\,\gamma_v + \zeta_v\sqrt{s} < \frac{\varepsilon}{4096\sqrt{s}}.$$

For row i, subtract an independent linear-score observation of $p_i^\top \widetilde{v}$ from a fresh observation of A_{ij^+}. This arm has mean within $\varepsilon/4096$ of E_{ij^+} and proxy $O(s^2 L_s^2)$. A second $\varepsilon/128$-PAC absolute maximum returns i^+ with $\left|E_{i^+ j^+}\right| > 9\varepsilon/32$. Finally, estimate the selected residual-arm mean on a new stream to accuracy $\varepsilon/128$ and retain the pivot only when its estimated magnitude exceeds $\varepsilon/8$. On the simultaneous event the screen passes, and whenever it passes the true residual magnitude exceeds $\varepsilon/16$.

Each route uses two fresh inverse estimates. They cost at most $C s^6 L_s^4 \varepsilon^{-2} \log(cs/\alpha)$. The mixture-feature and selected-feature estimates are smaller. The two PAC scans cost at most $C(n+m) s^4 L_s^2 \varepsilon^{-2} \log(cs/\alpha)$ after a conservative polynomial enlargement in s. Within the routine, the input risk α is split uniformly across the two inverse calls, feature groups, PAC scans, and final screen. This proves the displayed cap for both witness types. □

M.7 Three distinct invariants and their dependencies

Mechanism	Certified invariant	Role in the proof
Residual pivot	$\lvert E_{ij}\rvert \geq \varepsilon/16$ and the Schur identity	Increases skeleton rank. At most r expansions.
Row purification	$\lVert p_i^\top C^{-1}\rVert_\infty \leq 2$ for all rows	Controls response directions without a condition-number-dependent net.
Determinant margin	$\lvert \det C\rvert \geq (\varepsilon/16)^s$	Protects empirical inversion via a public perturbation screen.

These guarantees are established by Lemmas M.11, M.8, and M.4, respectively. The first is algebraic progress, the second geometric control, and the third numerical stability. None can be replaced by either of the others. The risk table below specifies the fresh estimation events and their caps.

M.8 Filtration, risk budget, and totality

At the start of each stage, condition on the complete transcript and the current ordered skeleton. Every grouped routine below is then run on observations independent of that transcript. Within a routine, an inverse or selected label is first frozen. Every scan, scalar screen, and feature estimate that follows uses a new stream conditional on the enlarged transcript. A row swap ends its purification pass and invalidates every estimate made in that pass.

There are at most $r + 1$ stages. Give each of the following grouped calls risk

$$\alpha_\star = \frac{\delta}{4(r+1)}.$$

Grouped routine	Maximum calls per stage	Internal allocation
Row purification	one for $s \geq 1$	inverse calls, scans, screens over $K_s + 1$ passes
Surrogate candidate	one	net directions, features, guarded inverses
Fresh verifier	one	two PAC extrema and three scalar estimates
Residual pivot	one after rejection	features, two scans, inverses, final screen

Each lemma divides its input risk across its deterministic public list of subcalls. Unused calls spend no risk. A union bound therefore makes all guarantees simultaneous with probability at least $1 - \delta$. On this event, every guard passes, every accepted answer is correct, and every rejection at a skeleton size strictly below $\mathrm{rank}(A)$ produces a valid new pivot. Outside this event, every guard or cap branch returns a fixed pair, so the policy remains total and deterministically capped.

M.9 Proof of the fixed-rank theorem

Proof of Theorem M.1. The algorithm maintains an invertible skeleton of size s, starting with $s = 0$. At every size $s \geq 1$, it first runs dominant-row purification. It then constructs an $\varepsilon/16$-gap equilibrium of the current CUR surrogate by Lemma M.9. Fresh verification either accepts or returns a large true deviation. On rejection, Lemma M.11 returns a Schur residual pivot of magnitude at least $\varepsilon/16$. Append its row and column labels to the skeleton.

By Lemma M.3, every appended pivot increases the skeleton rank by one. Hence at most $\mathrm{rank}(A) \leq r$ expansions occur. The determinant recursion and determinant-increasing row swaps maintain the invariant of Lemma M.4.

Every primitive call has a deterministic cap. The risk table above and the conditional-freshness rule make every guarantee valid after the adaptive history. On the simultaneous event, every numerical guard passes, every acceptance is correct, and every rejection below the true rank makes rank progress. If $s = \mathrm{rank}(A)$, Lemma M.3 gives $E = 0$ and therefore $B = A$. The surrogate candidate then has true gap at most $\varepsilon/16$, so the fresh verifier must accept on its good event. Hence the $s = r$ rejection branch and every request for an $(r + 1)$st pivot lie outside the simultaneous event. On any guard failure or cap exhaustion, the policy returns a fixed pair immediately. It is therefore total and deterministically capped on every transcript, while the probability of an incorrect output is at most δ.

For a compact query bound, define

$$L = r^{r-1}\left(\frac{16}{\varepsilon}\right)^r, \qquad N = \left(1 + \frac{4096r}{\varepsilon}\right)^r, \qquad K = 1 + Cr^2 \log\frac{16r}{\varepsilon}.$$

Summing Lemmas M.8 to M.11 over at most $r + 1$ stages gives

$$C_r K\left[L^4\varepsilon^{-2} + (n+m)NL^2\varepsilon^{-2}\right]\log\frac{c_r KN}{\delta}. \tag{18}$$

Polynomial factors in r are absorbed into C_r. Since $N = \varepsilon^{-r}$ and $L = \varepsilon^{-r}$ up to r-dependent constants, the two principal accuracy exponents are $4r + 2$ and $3r + 2$. Finally, for $\varepsilon \leq 1/8$, the factor $K \log(c_r KN/\delta)$ can be absorbed into $C_r\varepsilon^{-4}\log(c_r/\delta)$. This gives Equation (14). □

Remark M.12 (Why no conditioning assumption remains). A small singular value is not ignored. It is paid for through the conservative factor $L_s = \varepsilon^{-O(s)}$. The crucial point is that it no longer appears inside the directional covering number after dominant-row purification. This is what keeps the overall dependence at $\varepsilon^{-O(r)}$ rather than $\varepsilon^{-O(r^2)}$.

N Bounded Rollouts and Additional Structural Consequences

N.1 Rademacher and Bernoulli phase transfer

Define the Rademacher observation model by

$$W \in \{-1, +1\}, \qquad \mathbb{P}_A(W = +1 \mid i, j) = \frac{1 + A_{ij}}{2}.$$

Let $\mathcal{M}^{\mathrm{Rad}}_{n,m}$ denote the corresponding full-box minimax complexity.

Lemma N.1 (Local Rademacher KL)**.** *For $a, b \in [-3/8, 3/8]$,*

$$\mathrm{kl}\left(\frac{1+a}{2}, \frac{1+b}{2}\right) \le 2(a-b)^2.$$

Proof. For Bernoulli parameters p, q, $\mathrm{kl}(p, q) \le (p-q)^2/[q(1-q)]$. Here $q \in [5/16, 11/16]$, so $q(1-q) \ge 55/256$. Since $p - q = (a-b)/2$, the result follows with room to spare. □

Lemma N.2 (Stopped adaptive Rademacher change of measure)**.** *Let $A, B \in [-3/8, 3/8]^{n\times m}$, and let a sequential policy stop at τ almost surely under both Rademacher instances. If $\mathbb{E}_A\tau < \infty$, the stopped-transcript laws, including the terminal output, satisfy*

$$\begin{aligned}\mathrm{KL}(P_A^\tau \| P_B^\tau) &= \sum_{i,j} \mathbb{E}_A N_{i,j}(\tau)\, \mathrm{kl}\left(\frac{1+A_{ij}}{2}, \frac{1+B_{ij}}{2}\right) \\ &\le 2\sum_{i,j} \mathbb{E}_A N_{i,j}(\tau)(A_{ij} - B_{ij})^2.\end{aligned}$$

Consequently, for every stopped-transcript event E, the first line is at least $\mathrm{kl}(P_A^\tau(E), P_B^\tau(E))$.

Proof. Pad the interaction after stopping with a common dummy symbol and truncate at a deterministic horizon T. Conditional on the past, the policy uses the same action kernel under both instances. Hence action-selection and private randomization likelihood ratios cancel, while a query to (i, j) contributes the displayed Bernoulli KL. The conditional chain rule gives the identity with $N_{i,j}(\tau \wedge T)$. The prefix sigma-fields increase to the complete stopped history. The terminal output is drawn from the same conditional kernel given that history under both instances, so it contributes zero additional conditional KL. The variational characterization of relative entropy and monotone convergence give the first line as $T \to \infty$. The local bound is Lemma N.1, and event-level data processing gives the final statement. □

Theorem N.3 (Bounded-rollout sandwich and matched regimes)**.** *There are universal constants $c, C, c_0 > 0$ such that, for every $(n, m) \ne (1, 1)$, $0 < \varepsilon \le c_0$, and $0 < \delta \le 1/4$,*

$$\begin{aligned}\mathcal{M}^{\mathrm{Rad}}_{n,m}(\varepsilon, \delta) &\ge \frac{c}{\varepsilon^2}\left[(n+m-2)\log\frac{1}{\delta} + n\log\big(e + u_{n|m}\big) + m\log\big(e + u_{m|n}\big)\right], \\ \mathcal{M}^{\mathrm{Rad}}_{n,m}(\varepsilon, \delta) &\le \frac{C}{\varepsilon^2}\left[n\log(en) + m\log(em) + (n+m)\log\frac{1}{\delta}\right],\end{aligned}$$

where

$$u_{n|m} = \min\left\{n, \varepsilon^{-1}, \frac{m}{\log(en)}\right\}, \qquad u_{m|n} = \min\left\{m, \varepsilon^{-1}, \frac{n}{\log(em)}\right\}.$$

The confidence-dominated, fixed-small-side, and balanced fine-accuracy conclusions therefore remain unchanged up to universal constants. The same statement holds when the observed rollout is $X \in \{0, 1\}$ with $\mathbb{P}_A(X = 1 \mid i, j) = (1 + A_{ij})/2$, after the affine transformation $W = 2X - 1$ and the corresponding constant rescaling of accuracy.

Proof. The upper bound is simpler than in the Gaussian model. Bounded-feedback self-play consumes the Rademacher observation directly. The fresh verifier and fixed-small-side compression use only subgaussian concentration, so their proofs are unchanged.

For the high-confidence lower bound, use the additive star family from the Gaussian proof with $\varepsilon \le 1/16$. Every hard mean then lies in $[-3/8, 3/8]$. The stopped identity in Lemma N.2 replaces the Gaussian coefficient by another universal constant.

For the discovery-entropy lower bound, use the same sign code and Johnson-layer argument with the following rescaling. Put $h = 8\varepsilon$ and

$$a_i(q, b) = \frac{bc_{iq}}{8}.$$

Restrict the layer size to $s \le 1/(64\varepsilon)$. For a good set U, define

$$z_q = \frac{8h}{d} \sum_{\ell \in U} c_{\ell q}.$$

Then $\sum_q |z_q| \le 8hs \le 1$, so z is realizable by the signed column pairs and a reservoir column. The induced response values are exactly

$$u_i(y) = h\mathbf{1}\{i \in U\} + hZ_i(U).$$

The value is at most $h/8 = \varepsilon$, and every individual-regret ε-answer decodes U by the same 4ε threshold. All hard means lie in $[-3/8, 3/8]$. Neighboring layer instances differ by h on one row pair, and Lemma N.2 bounds the stopped KL by $2h^2$ times the expected number of queries to that pair. The error-sharing and layer-counting proof is otherwise identical. Optimizing N, s, d gives the displayed lower bound.

For Bernoulli observations, center by $W = 2X - 1$. Affine shifts do not affect regrets and scaling changes the target accuracy only by a factor of two. □

N.2 Skew-symmetric single-policy self-play

Proposition N.4 (One-policy symmetrization)**.** *Let $A = -A^\top \in [-1, 1]^{k \times k}$. Suppose an algorithm returns (x, y) with*

$$R_A(x, y) \le \varepsilon, \qquad C_A(x, y) \le \varepsilon.$$

Then

$$q = \frac{x + y}{2}$$

satisfies

$$R_A(q, q) = C_A(q, q) \le \varepsilon.$$

Thus every two-policy certification theorem immediately yields a single certified policy for symmetric self-play at no additional query cost.

Proof. Define $f(z) = \max_i e_i^\top Az$. Since A is skew-symmetric,

$$\mathrm{Gap}_A(x, y) = f(y) + f(x).$$

The function f is convex, so

$$f(q) \le \frac{f(x) + f(y)}{2} = \frac{\mathrm{Gap}_A(x, y)}{2} \le \varepsilon.$$

Also $q^\top Aq = 0$ and $-\min_j q^\top Ae_j = f(q)$, proving the two equal regrets. □

N.3 Why a universal strong-gap primitive is impossible

A tempting route to the remaining balanced regime would assume a primitive with

$$\mathbb{E}\,\mathrm{Gap}_A(\bar{x}_T, \bar{y}_T) \le c_0 B \sqrt{\frac{k}{T}}$$

for every k, T and every bounded-feedback game. The dense fine-accuracy lower bound rules out such a universal primitive.

Corollary N.5 (No universal $\sqrt{k/T}$ expected-gap primitive)**.** *There is no universal constant c_0 and no one-cell bounded-feedback self-play primitive satisfying*

$$\sup_{A\in[-1,1]^{k\times k}} \mathbb{E}_A \operatorname{Gap}_A(\bar{x}_T, \bar{y}_T) \le c_0 B\sqrt{\frac{k}{T}}$$

for every $k \ge 2$ and every $T \ge 1$.

Proof. If such a primitive existed, choose $T = \Theta(k\varepsilon^{-2})$. Markov's inequality would give a constant-success $O(k\varepsilon^{-2})$ proposal. The exact Gaussian-to-bounded mean factory and fresh verification-restart wrapper would then produce a globally correct Gaussian policy with expected complexity

$$O\left(k\varepsilon^{-2}\left[1+\log\frac{1}{\delta}\right]\right)$$

for every accuracy. At constant confidence, this contradicts the established dense lower bound

$$\Omega(k\varepsilon^{-2}\log k)$$

when $\varepsilon \lesssim \log k/k$. Therefore the premise is false, not unresolved. □

O Formal Fixed-Candidate Characteristic Definitions

This section defines information quantities for a prescribed candidate. They are used in Appendix S to compare two acceptance criteria, not to claim an instance-optimal discovery algorithm. Besides the ε-Nash answer set $\mathcal{C}_\varepsilon^{\mathrm{ind}}(A)$, define $\mathcal{C}_\varepsilon^{\mathrm{DG}}(A) = \{z : \operatorname{Gap}_A(z) \le \varepsilon\}$. Nonnegativity of both regrets gives $\mathcal{C}_\varepsilon^{\mathrm{DG}}(A) \subseteq \mathcal{C}_\varepsilon^{\mathrm{ind}}(A) \subseteq \mathcal{C}_{2\varepsilon}^{\mathrm{DG}}(A)$. These criteria need not agree at the same accuracy.

Define

$$\Phi_A(y) = \max_i e_i^\top Ay, \qquad \Psi_A(x) = \min_j x^\top Ae_j.$$

For a criterion $\mathsf{C} \in \{\mathrm{ind}, \mathrm{DG}\}$ and a fixed candidate z, let

$$\mathrm{Alt}_{\mathsf{C}}(z) = \{B \in [-1,1]^{n\times m} : z \notin \mathcal{C}_\varepsilon^{\mathsf{C}}(B)\}.$$

For an allocation $w \in \Delta_{nm}$, define

$$D_{\mathsf{C}}(A; z, w) = \inf_{B\in\mathrm{Alt}_{\mathsf{C}}(z)} \frac{1}{2}\sum_{i,j} w_{ij}(A_{ij}-B_{ij})^2,$$

$$D_{\mathsf{C}}^\star(A,\varepsilon) = \sup_{z\in\mathcal{C}_\varepsilon^{\mathsf{C}}(A)} \sup_{w\in\Delta_{nm}} D_{\mathsf{C}}(A; z, w),$$

$$T_{\mathsf{C}}^\star(A,\varepsilon) = \frac{1}{D_{\mathsf{C}}^\star(A,\varepsilon)}.$$

This is an oracle fixed-answer information characteristic. It optimizes over the terminal answer and the allocation. It is not, by itself, an end-to-end unknown-output sample complexity.

For a fixed individual-regret candidate $z = (x, y)$, the violating alternatives form the union of the intersections of $n + m$ open linear halfspaces with the payoff box. Define direction matrices

$$G_i^R = (e_i - x)y^\top, \qquad G_j^C = x(y - e_j)^\top,$$

and write $\Delta_G = \varepsilon - \langle G, A\rangle$ for each direction G. Let $\mathcal{G}_\varepsilon$ contain those row and column directions for which $\sum_{a,b}|G_{ab}| > \varepsilon$. Every other direction has no strictly violating point in the payoff box and contributes $+\infty$. Suppose $\Delta_G > 0$ for every $G \in \mathcal{G}_\varepsilon$ and, for each such direction, the minimization of the weighted quadratic distance over the closed halfspace $\langle G, B\rangle \ge \varepsilon$, without box constraints, has a minimizer in the payoff box. Then

$$D_{\mathrm{ind}}(A; z, w) = \min_{G\in\mathcal{G}_\varepsilon} \frac{\Delta_G^2}{2\sum_{a,b} G_{ab}^2/w_{ab}}, \tag{19}$$

where $\min\varnothing = +\infty$. In the denominator, $0/0$ contributes zero and a nonzero direction on an unsampled cell contributes $+\infty$.

Appendix table of contents

Derivation of Equation (19). The maximum of $\langle G, B\rangle$ over the payoff box is $\sum_{a,b}|G_{ab}|$, so strict violation is possible exactly for $G \in \mathcal{G}_\varepsilon$. Fix such a direction and write $H = B - A$ and $S_G = \sum_{a,b} G_{ab}^2/w_{ab}$. For $\langle G, B\rangle \geq \varepsilon$ and $S_G < \infty$, weighted Cauchy–Schwarz gives

$$\Delta_G^2 \leq S_G \left(\sum_{a,b} w_{ab} H_{ab}^2 \right).$$

The closed-halfspace minimum is $\Delta_G^2/(2S_G)$, attained by $H_{ab} = \Delta_G G_{ab}/(w_{ab} S_G)$ on sampled cells. If $S_G = +\infty$, changing only unsampled cells attains a zero unconstrained minimum. In either case, the hypothesis supplies a closed-halfspace minimizer B_0 in the payoff box. Choose a strictly violating point B_+ in the box. For $0 < t \leq 1$, $(1-t)B_0 + tB_+$ is strictly violating and remains in the box, while its cost tends to that of B_0 as $t \downarrow 0$. Thus the strict infimum equals the closed minimum. Taking the minimum over the eligible directions proves the formula. □

A GLR statistic and plug-in allocation tracking are possible research components, not an established end-to-end algorithm for this characteristic. A proof must handle discontinuous answer selection, nonunique optimal allocations, payoff-box boundary alternatives, and global stopping. In particular, finite-answer tracking cannot be transferred automatically to this infinite answer set (Poiani et al., 2026). We make no asymptotic optimality or instance-optimality claim for such a construction.

P Rank-two response compression

This section proves the two bounds in Theorem 4.5. The headline skew bound is Theorem P.11. The headline rectangular bound is Theorem P.12. The direct and sequential solvers provide reusable ingredients for these bounds. The following map separates their roles.

Proof module	Role	Main dependency
Appendix P.1	Stable skew cross, direct solver, and lower bound	Fresh PAC scans
Appendix P.2	Deterministic planar primitive and sequential implementation	Stable cross
Appendix P.3	Final skew query bound by sharing coordinate samples	Planar primitive
Appendix P.4	Final rectangular query bound by bounded-coordinate feasibility	PAC scans and volume purification
Appendices P.5 to P.7	Two sufficient near-skew neighborhoods with compressed and safe interfaces	Planar primitive and shared sampling
Appendix P.8	Change of feedback law for the stated algorithms	Bounded-feedback concentration

The exact-matrix results do not assume supplied representations. The rectangular proof can be read independently of the skew proof path: Lemmas P.13 and P.14 discover and purify a cross, Lemmas P.15 and P.16 transfer physical-coordinate errors, and the final optimization accounts for the three logarithmic schedules. Its logarithms are not matched by a lower bound. The robustness modules concern near-skew structure. They do not establish a fast guarantee for arbitrary rectangular rank-two perturbations. Appendix Q, after these modules, combines the established proposal contracts into one globally valid policy.

P.1 Stable skew crosses and a direct solver

Write $f_A(q) = \max_i e_i^\top Aq$. For a real skew-symmetric matrix, $f_A(q) = R_A(q,q) = C_A(q,q)$, since $q^\top Aq = 0$. Every real skew-symmetric matrix of rank at most two admits a representation $A_{ij} = a_i b_j - b_i a_j$. The vectors a, b are not supplied to the algorithm.

Theorem P.1 (Computable skew rank-two certification)**.** *There are universal constants $c, C > 0$ such that, for $n \geq 2$, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$, the Gaussian cell-query model on*

$$\mathcal{S}_2(n) = \{A \in [-1,1]^{n\times n} : A = -A^\top,\ \mathrm{rank}(A) \leq 2\}$$

has minimax expected complexity satisfying

$$c(n-1)\varepsilon^{-2}\log(1/\delta) \;\le\; \mathcal{M}^{\mathrm{G}}_{S_2}(n,\varepsilon,\delta) \;\le\; Cn\varepsilon^{-2}\log(cn/\delta).$$

The upper bound is a deterministic query cap, and the output is one policy q, supported on at most three labels, with $f_A(q) \le \varepsilon$ with probability at least $1-\delta$. The algorithm knows no factorization, anchor, support, conditioning parameter, or spikiness parameter. Its postprocessing uses polynomially many arithmetic operations and can be formulated as an LP with $O(n)$ variables, constraints, and nonzero coefficients.

The upper bound has the optimal accuracy exponent 2. The remaining $\log n$ is not claimed necessary. For comparison with the unrestricted algorithm, one may take the smaller of the two public upper bounds.

Lemma P.2 (Two absolute scans stabilize a skew rank-two cross)**.** *Suppose $A = -A^\top$ and $\mathrm{rank}(A) \le 2$. Let a, b, c satisfy*

$$|A_{ba}| \ge \max_i |A_{ia}| - t, \qquad |A_{cb}| \ge \max_i |A_{ib}| - t, \qquad |A_{ba}| > 62t.$$

Then, writing $w = A_{bc}$ and $p = |w|$, we have $p > 61t$ and

$$\max_i |A_{ib}| \le \frac{62}{61}p, \qquad \max_i |A_{ic}| \le \frac{63}{62}\left(2 + \frac{1}{61}\right)p < 3p.$$

In particular, both columns of the cross are bounded by $3|w|$. Moreover, with $u_i = A_{ib}$ and $v_i = A_{ic}$,

$$A_{ij} = \frac{u_i v_j - v_i u_j}{w} \quad \textit{for all } i, j. \tag{20}$$

Proof. Put $d = |A_{ba}|$, $M_a = \max_i |A_{ia}|$, and $M_b = \max_i |A_{ib}|$. Since $M_b \ge d$ and $p \ge M_b - t$, we have $p > d - t > 61t$ and $M_b \le p + t \le (62/61)p$. A nonzero entry of a skew matrix of rank at most two gives a nonsingular 2×2 principal cross. Vanishing 4×4 Pfaffians, or the factorization $A_{ij} = a_i b_j - b_i a_j$, gives

$$A_{ic} = \frac{A_{ib}A_{ac} - A_{ia}A_{bc}}{A_{ab}}.$$

Consequently

$$|A_{ic}| \le \frac{M_a}{d}(M_b + p) \le \left(1 + \frac{t}{d}\right)(2p + t) \le \frac{63}{62}\left(2 + \frac{1}{61}\right)p.$$

The same rank-two identity on the nonzero pivot A_{bc} proves equation 20. □

Lemma P.3 (Stable noisy cross reconstruction)**.** *Suppose equation 20 holds and $|u_i|, |v_i| \le 3p$, where $p = |w| > 0$. Let $|\widehat{u}_i - u_i|, |\widehat{v}_i - v_i| \le h$ for all i, $|\widehat{w} - w| \le h$, and $h \le p/2$. Define*

$$\widehat{A}_{ij} = \frac{\widehat{u}_i\widehat{v}_j - \widehat{v}_i\widehat{u}_j}{\widehat{w}}.$$

Then $\widehat{A}$ is skew-symmetric, has rank at most two, and $\|\widehat{A} - A\|_\infty \le 62h$.

Proof. The numerator perturbation is at most $12ph + 2h^2$. Also the exact numerator is at most $18p^2$ in absolute value. Thus

$$|\widehat{A}_{ij} - A_{ij}| \le \frac{12ph + 2h^2}{p - h} + \frac{18p^2 h}{p(p-h)} = \frac{30ph + 2h^2}{p - h} \le 62h.$$

Skew symmetry and the rank bound follow directly from the formula. □

Algorithm and risk budget. Set $t = \varepsilon/128$ and $h = \varepsilon/256$. Each of the first four grouped calls below is given failure probability $\delta/8$. Every call uses fresh observations conditional on all preceding labels and estimates.

1. Fix $a = 1$. Run a t-PAC maximum on column a, returning j. Estimate A_{ja} to accuracy t on a fresh stream. If this estimate is at most $\varepsilon/2$, return $q = e_a$.
2. Run a t-PAC absolute maximum on column a, returning b. Then run a t-PAC absolute maximum on column b, returning c.
3. After b, c are frozen, estimate all $u_i = A_{ib}$ and $v_i = A_{ic}$, and separately $w = A_{bc}$, to accuracy h, with total risk $\delta/2$. For Gaussian variance-one observations, it suffices to use
$$N_h = \left\lceil 2h^{-2} \log \frac{4(2n+1)}{\delta} \right\rceil$$
fresh samples for each of the $2n + 1$ means. Exact diagonal zeroes may be inserted without sampling, but are not needed for the cap.
4. If $b = c$ or $|\widehat{w}| \le h$, return e_1. Otherwise form $\widehat{A}$ implicitly using the noisy-cross formula and find a symmetric equilibrium q of $\widehat{A}$. Return (q, q). Optional guards on invalid numerical solver output return e_1. In exact real arithmetic they are never activated on the good event.

The four early risk groups are the PAC maximum, its scalar estimate, and the two absolute scans. Thus the total failure probability is at most $4\delta/8 + \delta/2 = \delta$. If a model-independent public implementation requires all internal routines to return on arbitrary transcripts, every PAC routine uses its deterministic public cap and the final numerical routine solves the always feasible skew-game LP described below.

Correctness. On the simultaneous good event, if the initial branch returns, then
$$f_A(e_a) = \max_i A_{ia} \le A_{ja} + t \le \widehat{A}_{ja} + 2t \le \varepsilon/2 + 2t < \varepsilon.$$
Otherwise $A_{ja} > \varepsilon/2 - t = 63t$, so the first absolute scan gives $|A_{ba}| > 62t$. Lemma P.2 implies $|w| > 61t$, and all its other conclusions. Since $h = t/2 < |w|/2$, the guard does not activate. The reconstruction lemma yields
$$\|\widehat{A} - A\|_\infty \le 62h = 31\varepsilon/128 < \varepsilon.$$
An exact symmetric equilibrium of a skew-symmetric zero-sum game satisfies $\max_i e_i^\top \widehat{A} q = 0$. Therefore
$$f_A(q) \le f_{\widehat{A}}(q) + \|\widehat{A} - A\|_\infty \le 31\varepsilon/128 < \varepsilon.$$
The initial constant number of PAC extrema costs $O(n\varepsilon^{-2} \log(c/\delta))$. The last cross estimates cost $O(n\varepsilon^{-2} \log(cn/\delta))$. No inverse singular value occurs.

Sparse, polynomial-time postprocessing. Let $s = \operatorname{sign}(\widehat{w})$, and introduce two variables U, V. Find any feasible solution to
$$q \ge 0, \quad \sum_i q_i = 1, \quad U = \sum_i q_i \widehat{u}_i, \quad V = \sum_i q_i \widehat{v}_i, \quad s(\widehat{u}_i V - \widehat{v}_i U) \le 0 \quad (i \in [n]).$$
This system is feasible by the minimax theorem for $\widehat{A}$ and has $O(n)$ nonzero coefficients. It need not explicitly construct the $n \times n$ matrix. A planar Carathéodory reduction preserves U, V and reduces the support of q to at most three, preserving every constraint. It can be implemented by repeated affine-dependence elimination, in polynomial time. Every output therefore has support at most three.

Lower bound. Use the transitive subfamily $A_{ij} = z_i - z_j$, which is skew-symmetric and has rank at most two. At the base put $z_1 = 0$ and $z_i = -4\varepsilon$ for $i > 1$. Alternative $h > 1$ changes only z_h to 2ε. These matrices belong to $[-1, 1]^{n \times n}$ for $\varepsilon \le 1/8$. For any pair (x, y), the row regret is $\max_i z_i - \sum_i x_i z_i$. Base correctness forces $x_1 \ge 3/4$, whereas row correctness under alternative h forces
$$x_h \ge 1/2 + 2 \sum_{i \notin \{1,h\}} x_i.$$

Appendix table of contents

Hence the base-correct region and alternative-h-correct region are disjoint. If E is the base-correct event, global correctness gives $P_0(E) \geq 1-\delta$ and $P_h(E) \leq \delta$. A query whose two labels exclude h has identical distribution, and any other query changes its mean by at most 6ε. Thus stopped Gaussian change of measure implies

$$18\varepsilon^2 \mathbb{E}_0 N_h \geq \mathrm{kl}(1-\delta, \delta) \geq \tfrac{1}{4}\log(1/\delta).$$

Here N_h counts queries incident to h. Summing over $h > 1$, each physical query is counted at most twice, so $36\varepsilon^2 \mathbb{E}_0 \tau \geq (n-1)\log(1/\delta)/4$. This argument applies even to algorithms allowed to output two unrelated mixed policies and therefore proves the stated lower bound.

A genuinely nontransitive member. Take $a = (1, -1/2, -1/2)$ and $b = (0, \sqrt{3}/2, -\sqrt{3}/2)$. Then $A_{ij} = a_i b_j - b_i a_j$ has rank two, has all off-diagonal magnitudes $\sqrt{3}/2$, and contains the directed cycle $A_{12}, A_{23}, A_{31} > 0$. The uniform policy is an exact equilibrium. The theorem covers this entire cyclic class, rather than only $A_{ij} = z_i - z_j$.

Compatibility with globally safe structure selection. Off the rank-two skew class, all sample caps and guard branches remain well defined. The resulting output is a candidate, not a certified answer. A fresh full-box verifier may accept this candidate or reject it. Racing this proposal–verification branch with a globally correct unrestricted branch therefore preserves correctness without trusting the structure. The theorem supplies the structured branch's constant-success proposal cost. The independent verifier supplies validity under model mismatch.

Corollary P.4 (Unknown-radius skew rank-two tube)**.** *Suppose the observed matrix $A \in [-1,1]^{n\times n}$ is skew-symmetric and there exists a skew-symmetric matrix B of rank at most two with $\|A-B\|_\infty \leq \rho \leq \varepsilon/128$. The algorithm above, without knowing ρ or B, has the same cap and returns a genuine ε-equilibrium of A with probability at least $1-\delta$. More precisely, its structural branch satisfies*

$$f_A(q) \leq 31\varepsilon/128 + 63\rho.$$

Proof. The initial pure-output branch is valid directly on A by skew symmetry. Otherwise, its selected label satisfies $A_{ja} > 63t$, where $t = \varepsilon/128$. In B this gives $|B_{ja}| > 63t - \rho$. Every t-near absolute scan in A is a $t' = t + 2\rho$-near absolute scan in B. Thus $d = |B_{ba}| > 62t - 3\rho \geq 59t$, and $p = |B_{bc}| > 61t - 5\rho \geq 56t$. The calculation in the two-scan lemma, now with $t' \leq 3t$, gives

$$\max_i |B_{ib}|/p \leq 1 + 3/56 < 3, \quad \max_i |B_{ic}|/p \leq (1+3/59)(2+3/56) < 3.$$

Every empirical cross coordinate and pivot is within $h+\rho$ of its counterpart in B, where $h = t/2$. Since $h + \rho \leq 3t/2 < p/2$, the noisy-cross lemma gives $\|\widehat{A} - B\|_\infty \leq 62(h+\rho)$. Also $|\widehat{w}| \geq p - h - \rho > h$, so the guard passes. A symmetric equilibrium of $\widehat{A}$ therefore satisfies

$$f_A(q) \leq \|A-B\|_\infty + \|B-\widehat{A}\|_\infty \leq 62h + 63\rho \leq 94\varepsilon/128 < \varepsilon.$$

All risk events refer to estimation of A, so their probabilities and query caps are unchanged. □

P.2 Planar support geometry and three-action output

This section supplies the planar geometry used by the sharper skew bound. It first pays separately for each support direction. The next section reuses coordinate samples across directions. The final compression step limits the output to three actions without additional observations.

Theorem P.5 (Planar skew rank-two certification)**.** *For every $n \geq 2$, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$, there is a deterministically capped Gaussian cell-query algorithm such that, for every $A \in [-1,1]^{n\times n}$ with $A = -A^\top$ and $\mathrm{rank}(A) \leq 2$, its output is a distribution q on at most three labels satisfying $\max_i e_i^\top A q \leq \varepsilon$ with probability at least $1-\delta$. Its query cap is*

$$Cn\varepsilon^{-2}\log(2/\varepsilon)\log\left(\frac{\log(2/\varepsilon)}{\delta}\right).$$

The policy requires no anchor, factorization, conditioning or spikiness information and uses only PAC one-dimensional maxima, selected-label scalar estimates, and arithmetic on at most four output weights. Combining it with Theorem P.1 gives

$$\mathcal{M}^{\mathrm{G}}_{S_2} \leq Cn\varepsilon^{-2}\min\left\{\log(cn/\delta),\ \log(2/\varepsilon)\log\left(\frac{\log(2/\varepsilon)}{\delta}\right)\right\}.$$

Together with the rank-two lower bound, this leaves only logarithmic accuracy factors in the joint-parameter rate. It is not asserted to be the exact rank-two minimax law.

The geometric primitive is stated explicitly so that the oracle and normalization costs can be checked independently.

Lemma P.6 (Planar support primitive)**.** *Let $z_i = (u_i, v_i) \in [-K, K]^2$ with $K \geq 1$, including known labels a, b with $z_a = (0, 1)$ and $z_b = (-1, 0)$. Put*

$$f_i(t) = (1-t)u_i - tv_i, \quad F(t) = \max_i f_i(t), \quad s_i = u_i + v_i.$$

Suppose each requested $t \in [0, 1]$ returns a label i and estimates $\widehat{z}_i$ satisfying

$$f_i(t) \geq F(t) - \tau, \qquad \|\widehat{z}_i - z_i\|_\infty \leq \tau. \tag{P}$$

For $0 < \tau \leq 1/(64K)$, at most

$$M = 8 + 2\lceil \log_2(2K/\tau) \rceil$$

such calls suffice to produce q on at most three labels with

$$\max_i \det\left(z_i, \sum_j q_j z_j\right) \leq 64K\tau.$$

The number of calls is capped on every transcript, whether or not (P) holds. On cap exhaustion, a fixed pure distribution is returned.

Proof. Throughout, every f_i and F is $2K$-Lipschitz. Define $\widehat{f}_i(t) = (1-t)\widehat{u}_i - t\widehat{v}_i$ and $\widehat{s}_i = \widehat{u}_i + \widehat{v}_i$. Assumption (P) implies

$$|\widehat{f}_i(t) - F(t)| \leq 2\tau, \qquad |\widehat{s}_i - s_i| \leq 2\tau.$$

First construct a scalar $t_D \in [0, 1]$ and a distribution p on at most two returned labels such that, writing $(U, V) = \sum_i p_i z_i$,

$$\min(U, -V) \geq F(t_D) - 6\tau. \tag{21}$$

Query at 0. If its response has $\widehat{s} \leq 2\tau$, set p to that pure label and $t_D = 0$. Otherwise query at 1. If its response has $\widehat{s} \geq -2\tau$, return that pure label and $t_D = 1$. In the remaining case, initialize $L = 0, R = 1$, retaining the two responses, so $\widehat{s}_L > 2\tau$ and $\widehat{s}_R < -2\tau$. While $R - L > \tau/(2K)$, query its midpoint t. If $|\widehat{s}| \leq 2\tau$, return its pure response and $t_D = t$. If $\widehat{s} > 2\tau$, replace L and its response by this call. Otherwise replace R and its response. At termination set $t_D = (L + R)/2$ and

$$\lambda = \frac{-\widehat{s}_R}{\widehat{s}_L - \widehat{s}_R}, \qquad p = \lambda e_{i_L} + (1-\lambda)e_{i_R}.$$

The strict opposite signs imply $0 < \lambda < 1$. The estimated coordinate sum of this mixture is zero, whence $|U + V| \leq 2\tau$.

For a stop at 0, $s_i \leq 4\tau$, so $\min(u_i, -v_i) = u_i - \max(s_i, 0) \geq F(0) - 5\tau$. The stop at 1 is symmetric. For an interior singleton stop, $|s_i| \leq 4\tau$, and $\min(u_i, -v_i) \geq f_i(t) - |s_i| \geq F(t) - 5\tau$. These cases imply equation 21.

For the bracket case, let $v_* = \min_{t\in[0,1]} F(t)$ and $h = R - L \leq \tau/(2K)$. Both stored responses satisfy $f_{i_L}(t_D), f_{i_R}(t_D) \geq v_* - 2\tau$. Therefore $\min(U, -V) \geq v_* - 4\tau$. For any minimizer t_* in the final bracket, Lipschitzness gives $F(t_D) \leq v_* + \tau$. If $t_* < L$, the stored response has $s_L > 0$, hence

$$v_* = F(t_*) \geq f_{i_L}(t_*) = f_{i_L}(L) + s_L(L - t_*) \geq F(L) - \tau.$$

It follows that $F(t_D) \leq v_* + 2\tau$. The case $t_* > R$ is symmetric. This proves equation 21 in all cases.

Query once more at t_D and write its estimate as $\widehat{F}_D$. If $\widehat{F}_D \geq -16\tau$, then $F(t_D) \geq -18\tau$ and equation 21 gives $U \geq -24\tau$ and $V \leq 24\tau$. Form $(\widehat{U}, \widehat{V})$ by mixing the stored estimates with the already selected weights p, and set

$$x = \max(\widehat{U}, 0), \qquad y = \max(-\widehat{V}, 0), \qquad q = \frac{p + xe_b + ye_a}{1 + x + y}.$$

Appendix table of contents

Its unnormalized feature vector is $(U - x, V + y)$. Each coordinate has absolute value at most 25τ, and normalization cannot enlarge it. Consequently

$$\max_i \det(z_i, z_q) \le K(|(z_q)_1| + |(z_q)_2|) \le 50K\tau.$$

This intermediate output has at most four labeled occurrences. Apply Lemma P.7 to their stored estimates, using the exact anchor coordinates. The resulting distribution has at most three actions and determinant bound $54K\tau < 64K\tau$. No additional oracle call is used.

It remains to treat $\widehat{F}_D < -16\tau$, for which the fixed point t_D satisfies $F(t_D) < -14\tau$. If the original oracle estimate at 0 is at most 4τ, return e_a: its exploitability in determinant units is $F(0) \le 6\tau$. Otherwise maintain a root bracket $[l, r] = [0, t_D]$. At each midpoint t, query the oracle. If $|\widehat{f}_i(t)| \le 4\tau$, return e_i. If $\widehat{f}_i(t) > 4\tau$, replace $l = t$. If it is below -4τ, replace $r = t$. Throughout, $F(l) > 2\tau$ and $F(r) < -2\tau$. Thus the procedure must return before the bracket width reaches $\tau/(2K)$. Otherwise its endpoint values contradict $2K$-Lipschitzness. Impose an explicit hard cap of $\lceil \log_2(2K/\tau) \rceil + 1$ midpoint queries, returning e_a if exhausted. This fallback is unreachable whenever (P) holds.

For a returned label, $F(t) \le 6\tau$ and $-7\tau \le f_i(t) \le 6\tau$. Since $t < t_D$ and $f_i(t_D) \le F(t_D) < -14\tau < f_i(t)$, we have $s_i > 0$. For every j, the determinant identity gives

$$\det(z_j, z_i) = s_i f_j(t) - s_j f_i(t) \le (2K)(6\tau) + (2K)(7\tau) = 26K\tau.$$

This is the required approximate pure equilibrium. The first saddle routine uses two endpoint calls and at most $\lceil \log_2(2K/\tau) \rceil$ midpoint calls, the new call at t_D uses one, and the root search obeys its explicit cap. Their sum is below M. □

Proof of Theorem P.5. Run the seed/early-certification step and the two absolute scans of Theorem P.1, with $t = \varepsilon/128$. Give each of their four grouped calls risk $\delta/16$. On their good event, the procedure either accepts correctly in the early branch or selects a cross (b, c) with $p = |A_{bc}| > 61\varepsilon/128$ and

$$|A_{ib}|, |A_{ic}| \le 3p \quad (i \in [n]).$$

Estimate $w = A_{bc}$ on a fresh stream to error $h_0 = \varepsilon/2^{20}$ with risk $\delta/4$. If the estimate is negative, swap b, c and negate the estimate. Denote its resulting positive value by $\widetilde{p}$. If $\widetilde{p} \notin [\varepsilon/4, 2]$, return e_1. On the simultaneous good event, this guard passes, the new true pivot is $p > 0$, $|\widetilde{p} - p| \le h_0$, and $p/\widetilde{p} \in [1/2, 2]$.

Define the true but unknown normalized features

$$z_i = (A_{ib}/p, A_{ic}/p).$$

They lie in $[-3, 3]^2$, their anchor labels satisfy $z_b = (0, 1), z_c = (-1, 0)$, and $A_{ij} = p \det(z_i, z_j)$. Set $K = 3$ and freeze

$$\tau = \frac{\varepsilon}{1024\widetilde{p}}, \qquad \eta = \frac{\varepsilon}{4096}, \qquad M_* = 8 + 2\left\lceil \log_2 \frac{12288}{\varepsilon} \right\rceil.$$

The guard implies $0 < \tau \le 1/256 < 1/(64K)$, and the geometric primitive uses no more than M_* oracle calls.

We now implement each oracle call of Lemma P.6 with risk $\beta = \delta/(2M_*)$ and physical precision η. At direction $t \in [0, 1]$, a one-cell observation of the physical score

$$(1 - t)A_{ib} - tA_{ic} = pf_i(t)$$

is obtained by sampling column b with probability $1 - t$ and using the observed Gaussian sample, or sampling column c with probability t and negating its sample. The centered subgaussian proxy is at most two. Use an η-PAC maximum of these n arms with risk $\beta/2$. After its returned label is frozen, estimate its two physical coordinates on fresh streams to accuracy η, with combined risk $\beta/2$, and divide their estimates by $\widetilde{p}$. For the known anchor labels, exact normalized coordinates may instead be inserted.

Condition on the preceding transcript and the good pivot estimate. With conditional probability at least $1 - \beta$, the returned true normalized score is within $\eta/p \le \tau/2$ of its maximum. Moreover, for either coordinate a_i with $|a_i| \le 3p$,

$$\left| \frac{\widehat{a}_i}{\widetilde{p}} - \frac{a_i}{p} \right| \le \frac{\eta + 3h_0}{\widetilde{p}} < \tau.$$

Thus (P) holds with the prescribed τ. The cost of the PAC scan and the two selected-label estimates is $O(n\varepsilon^{-2}\log(c/\beta))$, with a deterministic cap. Only selected labels are estimated, so there is no union bound over all n coordinates.

The simultaneous good event has probability at least $1-4\delta/16-\delta/4-M_*\beta = 1-\delta$. On it, the planar primitive gives

$$\max_i e_i^\top Aq = p\max_i \det(z_i, z_q) \le 192p\tau = \frac{192}{1024}\frac{p}{\widetilde{p}}\varepsilon \le 3\varepsilon/8 < \varepsilon.$$

The seed calls and pivot estimate cost $O(n\varepsilon^{-2}\log(c/\delta))$. The at most $M_* = O(\log(2/\varepsilon))$ oracle calls cost the displayed bound. Every subroutine and root search has a public deterministic cap, and all guard failures return a fixed policy, so the total cap holds on all transcripts. The work beyond PAC scans consists of interval bisection, scalar arithmetic, and mixing at most four labels, hence is polynomial. □

P.2.1 THREE-ACTION OUTPUT WITHOUT EXTRA OBSERVATIONS

The support bound in the planar construction can be sharpened from four to three without another observation. This uses the coordinates already estimated after each selected label is frozen. It does not require the true factorization.

Lemma P.7 (Compression with estimated features)**.** *Let $z_i \in [-K, K]^2$ and let q be supported on at most four labeled occurrences. Suppose each occurrence has a stored estimate satisfying $\|\widehat{z}_i - z_i\|_\infty \le \tau$. There is a distribution q' on at most three of these occurrences, computable from q and the estimates alone, such that*

$$\max_j \det(z_j, z_{q'}) \le \max_j \det(z_j, z_q) + 4K\tau, \qquad z_q = \sum_i q_i z_i.$$

Repeated labels may be merged after compression, so the same bound holds for the number of distinct actions.

Proof. If four positive weights remain, the 3×4 matrix with columns $(1, \widehat{z}_i)$ has a nonzero null vector a. Since $\sum_i a_i = 0$, it has both positive and negative entries. With $t = \min_{i:a_i>0} q_i/a_i$, replace q by $q' = q - ta$. All weights remain nonnegative, their sum and estimated feature mean are preserved, and at least one weight vanishes. Fewer than four positive weights need no change. Preservation of the estimated mean gives

$$\|z_{q'} - z_q\|_\infty \le \|z_{q'} - \widehat{z}_{q'}\|_\infty + \|\widehat{z}_q - z_q\|_\infty \le 2\tau.$$

Thus $|\det(z_j, z_{q'} - z_q)| \le 4K\tau$ for every j. The argument is pathwise on the coordinate-accuracy event, even when all weights and labels were selected from the same estimates. If a label occurs twice with different stored estimates, keep its occurrences distinct during this calculation and merge its final weights afterwards. □

Only the near-origin branch of Lemma P.6 can return more than three labels. Its uncompressed bound is $50K\tau$, so compression gives $54K\tau < 64K\tau$. Consequently the original determinant bound, physical accuracy budget, query cap, confidence budget, and all-transcript guards are unchanged. Only a constant-dimensional linear system is solved. Arithmetic is charged under the same real-arithmetic convention as the other structural results.

Proposition P.8 (Three actions are necessary in general)**.** *For every $0 < \varepsilon < 1/3$, a skew rank-two game in $[-1,1]^{3\times 3}$ exists whose symmetric ε-equilibria all use three actions.*

Proof. Take the rock–paper–scissors matrix $A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$. It is skew-symmetric of rank two. By cyclic relabeling, any distribution on at most two actions has the form $q = (t, 1-t, 0)$, $0 \le t \le 1$. Its exploitability is $\max\{1-t, -t, 2t-1\} \ge 1/3$, with equality at $t = 2/3$. The uniform distribution is an exact equilibrium with three actions. □

P.3 Shared physical sampling for the skew bound

The sequential planar algorithm pays separately for each support direction. This section shares coordinate observations across a finite direction grid, preserves its support values by elimination, and then runs the planar geometry offline. It removes the factor $\log\log(1/\varepsilon)$ at constant confidence. The direction grid increases arithmetic work, but all computations remain polynomial in the population size and inverse accuracy.

Lemma P.9 (Finite-direction PAC compression)**.** *Let $a_1, \dots, a_n \in [-1,1]^d$ be unknown vectors, with a fresh $\mathcal{N}(a_{ij}, 1)$ observation available for every requested coordinate. Let $w_1, \dots, w_D$ be known vectors satisfying $\|w_h\|_1 \le 1$. For $0 < \eta \le 1$ and $0 < \beta \le 1/4$, a deterministically capped algorithm returns a nonempty set S, with $|S| \le \min\{n, 4D\}$, such that, with probability at least $1-\beta$,*

$$\max_{i \in S} w_h^\top a_i \ge \max_{i \in [n]} w_h^\top a_i - \eta \qquad (h \in [D]).$$

Its query cap is $Cdn\eta^{-2}\log(cdD/\beta)$. If $n \le 4D$, it may return all labels without sampling.

Proof. Start with $S_0 = [n]$. At stage ℓ, if $s_\ell = |S_\ell| > 4D$, put

$$e_\ell = \frac{\eta}{4}(3/4)^\ell, \qquad \alpha_\ell = \frac{\beta}{2^{\ell+1}}, \qquad b_\ell = \frac{\alpha_\ell}{8D}.$$

Estimate every coordinate of every active label from

$$N_\ell = \left\lceil 8 e_\ell^{-2} \log \frac{2d}{b_\ell} \right\rceil$$

fresh observations. Conditional on the complete stage-entry history, every vector estimate $\widehat{a}_i$ has ℓ_∞ error at most $e_\ell/2$, except with probability b_ℓ. For each direction retain the $r_\ell = \lceil s_\ell/(4D) \rceil$ largest empirical scores, with deterministic tie breaking. Let $S_{\ell+1}$ be the union of these sets. On every transcript it is nonempty and

$$|S_{\ell+1}| \le D \left\lceil \frac{s_\ell}{4D} \right\rceil \le s_\ell/4 + D < s_\ell/2. \tag{22}$$

Call an active label bad if its vector estimate exceeds the stated error. For each direction fix one true maximizing active label, using a deterministic choice in case of a tie. The probability that any of these at most D labels is bad is at most Db_ℓ. The conditional expected number of bad active labels is at most $s_\ell b_\ell$. Markov's inequality therefore gives

$$\Pr\{\text{at least } r_\ell \text{ bad labels} \mid \mathcal{H}_\ell\} \le \frac{s_\ell b_\ell}{r_\ell} \le 4Db_\ell.$$

Outside these events, every direction preserves its maximum to error e_ℓ. Indeed, if its true maximizing label is retained, this is immediate. Otherwise all r_ℓ retained empirical scores are at least its empirical score, and at least one retained label is accurate. Since $\|w_h\|_1 \le 1$, the latter label's true score is within e_ℓ of the true maximum. This holds for all directions simultaneously, with failure probability at most $5Db_\ell \le \alpha_\ell$.

Stop when at most $4D$ labels remain. Conditional union bounds over reached stages are valid because each stage uses fresh samples. As $\sum_\ell \alpha_\ell = \beta$ and $\sum_\ell e_\ell = \eta$, the support losses telescope to the claimed guarantee. No simultaneous confidence event over all original vectors is used.

For the deterministic sample count, equation 22 gives $s_\ell \le n2^{-\ell}$, so

$$\begin{aligned}\sum_\ell d s_\ell N_\ell &\le Cdn\eta^{-2} \sum_{\ell \ge 0} (8/9)^\ell \left[\log(cdD/\beta) + (\ell+1)\log 2\right] + dn \sum_{\ell \ge 0} 2^{-\ell} \\ &\le C' dn\eta^{-2} \log(c' dD/\beta).\end{aligned}$$

Both geometric sums converge. In particular, the number of elimination stages does not introduce an additional population-size logarithm. □

Lemma P.10 (A continuous planar support interface). *Suppose $a_i = (u_i, v_i) \in [-1,1]^2$ are accessible through Gaussian coordinate queries. Define $g_i(t) = (1-t)u_i - tv_i$ and $G(t) = \max_i g_i(t)$ for $t \in [0,1]$. For $0 < \eta \le 1$ and $0 < \beta \le 1/4$, a capped sampling procedure constructs a nonempty set S and stored estimates $\widehat{a}_i$, $i \in S$, such that, with probability at least $1 - 2\beta$, simultaneously*

$$\|\widehat{a}_i - a_i\|_\infty \le \eta \quad (i \in S), \qquad G(t) - \max_{i \in S} g_i(t) \le 2\eta \quad (t \in [0,1]).$$

Consequently the label maximizing the stored empirical score over S is 4η-optimal for every $t \in [0,1]$, including adaptive offline choices. The query cap is

$$Cn\eta^{-2}\left[\log\big(c\min\{n, \eta^{-1}\}\big) + \log(1/\beta)\right]. \tag{23}$$

Proof. Let $J = \lceil 4/\eta \rceil$, $D = J + 1$, and use the grid $t_h = h/J$, $h = 0, \dots, J$, with directions $w_h = (1 - t_h, -t_h)$. Apply Lemma P.9 at error η and risk β. Every individual score and both support envelopes are 2-Lipschitz. For a grid point within $\eta/4$ of any t,

$$G(t) \le G(t_h) + 2|t - t_h| \le \max_{i \in S} g_i(t_h) + \eta + 2|t - t_h| \le \max_{i \in S} g_i(t) + 2\eta.$$

After S is frozen, estimate both coordinates of every retained label from

$$\left\lceil 2\eta^{-2} \log \frac{4|S|}{\beta} \right\rceil$$

fresh observations per coordinate. Their simultaneous error is at most η with conditional probability at least $1 - \beta$. An empirical maximizer over S then loses at most 2η within S, yielding total loss at most 4η for all real directions on the same event.

If $n \le 4D$, omit elimination and estimate all labels, at cost $Cn\eta^{-2}\log(cn/\beta)$. Otherwise elimination costs $Cn\eta^{-2}\log(cD/\beta)$ and final estimation concerns at most $4D$ labels. Since $D = \Theta(\eta^{-1})$, these two branches give equation 23. The factor four in the branch threshold is absorbed in the universal constant. Every later support call is deterministically controlled by the same event and requires no new samples. □

Theorem P.11 (Compressed planar skew rank-two certification). *For $n \ge 2$, $0 < \varepsilon \le 1/8$, and $0 < \delta \le 1/4$, there is a deterministically capped Gaussian cell-query algorithm such that, for every $A \in [-1,1]^{n \times n}$ with $A = -A^\top$ and $\operatorname{rank}(A) \le 2$, its output q has support size at most three and satisfies $\max_i e_i^\top A q \le \varepsilon$ with probability at least $1 - \delta$. Its query cap is*

$$Cn\varepsilon^{-2}\left[\log\big(2 + \min\{n, \varepsilon^{-1}\}\big) + \log(1/\delta)\right]. \tag{24}$$

It requires no anchor, factorization, conditioning, or spikiness information. The arithmetic cost is polynomial in n and $1/\varepsilon$.

Proof. Use the seed and two absolute scans of Theorem P.5, assigning risk $\delta/16$ to each of its four grouped calls, and estimate the pivot with error $h_0 = \varepsilon/2^{20}$ and risk $\delta/4$. On their joint good event, the procedure either accepts correctly in the early branch or obtains an oriented true pivot $p = A_{bc} > 0$ and a positive estimate $\widetilde{p}$ satisfying

$$|A_{ib}|, |A_{ic}| \le 3p, \qquad |\widetilde{p} - p| \le h_0, \qquad p/\widetilde{p},\ \widetilde{p}/p \le 2.$$

The existing public guard $\varepsilon/4 \le \widetilde{p} \le 2$ passes. Define the unknown normalized features $z_i = (A_{ib}/p, A_{ic}/p)$, which have the known anchors $(0,1), (-1,0)$ and satisfy $A_{ij} = p\det(z_i, z_j)$.

Set $\eta_c = \varepsilon/32768$ and $\tau = \varepsilon/(1024\widetilde{p})$. Apply Lemma P.10 to the raw physical vectors $(A_{ib}, A_{ic}) \in [-1,1]^2$, with $\eta = \eta_c$ and $\beta = \delta/4$. No noisy division is used in elimination or in empirical score ranking. For every later direction t, choose the empirical maximizer of the physical score over S, and return its stored coordinates divided by $\widetilde{p}$. Exactly known normalized anchor coordinates may be substituted after the label is selected.

On the simultaneous sampling event, the returned label satisfies

$$F(t) - f_i(t) \le 4\eta_c/p = \frac{\varepsilon}{8192p} \le \tau/4 < \tau.$$

Its coordinate error is at most

$$\frac{\eta_c + 3h_0}{\widetilde{p}} = \frac{35\varepsilon}{2^{20}\widetilde{p}} < \tau.$$

Thus hypothesis (P) of Lemma P.6 holds for all directions on one event, including arbitrary adaptive reuse of the stored coordinates. Run that primitive offline, with its public query-count cap and fixed-output cap fallback. Its geometric conclusion gives

$$\max_i e_i^\top Aq \le 192p\tau \le 3\varepsilon/8 < \varepsilon.$$

Its at-most-three support labels can include the original anchors whether or not they survived elimination.

The seed, pivot, elimination, and final coordinate estimates each consume at most $\delta/4$ risk. Their total is δ. No risk is spent on the offline geometric calls. Equation (23), with $\eta_c = \Theta(\varepsilon)$, gives equation 24 after including seed and pivot costs. Public sample schedules, strict active-set shrinkage, and the original geometric cap ensure a deterministic total cap on every transcript. Evaluating the $O(1/\varepsilon)$ grid directions and the compressed envelope, and running the scalar geometry, require only polynomial arithmetic. □

P.4 Rectangular rank two at physical precision

The rectangular proof has two tasks: discover a cross with bounded interpolation on both sides, then find mixtures whose errors are small in the two selected payoff coordinates. The first task controls error transfer. The second uses convex-hull membership and feasibility searches. Their query costs are combined in the final subsection.

Theorem P.12 (Rectangular rank-two certification)**.** *There are universal positive constants C, c and a matrix-independent, deterministically capped policy with the following property. For $n, m \ge 1$, $0 < \varepsilon \le 1/8$, $0 < \delta \le 1/4$, and every $A \in [-1,1]^{n\times m}$ of rank at most two, it outputs a pair with both individual regrets at most ε, with probability at least $1-\delta$, using at most*

$$C(n+m)\varepsilon^{-2}L^3\log(cL/\delta), \qquad L = \log(2/\varepsilon),$$

physical Gaussian cell observations. No factorization, anchor, feature vector, conditioning parameter, or spikiness bound is supplied. Each output distribution is supported on at most three selected labels. The 1×1 case is returned without sampling.

All arithmetic below is in the same real-arithmetic query model as the main text. The finite-dimensional projections and ellipsoid updates are explicit and have polynomial arithmetic cost. Statistical estimates are clipped to their known coordinate boxes. This cannot increase their error on their good events. Every loop has a public finite cap, every inverse has a public singular-value screen, and any failed guard returns the fixed pure pair (e_1, e_1). Such a fallback is a proposal if this algorithm is used inside the globally safe wrapper.

P.4.1 Creating a useful skeleton, or returning an equilibrium

Lemma P.13 (A strategically discovered well-conditioned cross)**.** *For $0 < e \le 1/8$, a capped procedure uses*

$$C(n+m)e^{-2}\log(2/e)\log\big(c\log(2/e)/\beta\big)$$

queries and, with probability at least $1-\beta$, either returns an e-equilibrium of A, or returns two distinct rows I_0 and two distinct columns J_0 such that

$$C_0 = A_{I_0,J_0}, \qquad \sigma_{\min}(C_0) \ge e/160.$$

The second alternative can occur on the good event only if A has rank exactly two. The guarantee assumes only $\operatorname{rank}(A) \le 2$.

Proof. Set $t = e/2^{20}$. All PAC calls below have accuracy t, and all scalar estimates have error at most t, unless another accuracy is specified. Their risks are divided over the public $O(\log(2/e))$ call schedule. Every selected label or selected mixture is frozen before its subsequent scalar observations are drawn. We reason on the simultaneous good event.

Start at a fixed pure pair. A PAC row maximum, a PAC column minimum, and fresh scalar estimates give estimated individual regrets. If both are at most $e/2$, return the pair: its true regrets are at most $e/2 + 3t < e$. Otherwise the two estimated endpoint payoffs in a rejected comparison differ by more than $e/2$. The endpoint of larger estimated magnitude has true magnitude at least $e/4 - t > e/8$. Use it as a pivot (a, b).

At each current pivot, estimate its payoff, find a PAC absolute maximum in its row and its column, and estimate the two selected payoffs. If both estimated magnitudes are at most $3/2$ times the pivot's estimated magnitude, stop the ascent. Otherwise replace the pivot by a violating selected cell of largest estimated magnitude. If its old true magnitude is $p \geq e/8$, a replacement has true magnitude at least $3p/2 - 5t/2 \geq 4p/3$. Thus there are at most

$$K_{\mathrm{asc}} = 2 + \lceil \log(8/e)/\log(4/3) \rceil$$

passes before successful termination, since every payoff has magnitude at most one. On termination, writing $c = A_{ab}$ and $p = |c|$, the PAC and scalar errors give

$$p \geq e/8, \qquad \max_i |A_{ib}| \leq 2p, \qquad \max_j |A_{aj}| \leq 2p. \tag{25}$$

Every request beyond the pass cap returns the fixed-pair fallback.

Define the rank-one surrogate $B_{ij} = A_{ib}A_{aj}/c$. Find near maximum and minimum endpoints of column b and row a, obtaining sets S, T of size at most two. On fresh streams estimate the pivot and the selected row/column coordinates to accuracy t, form the at-most-2×2 estimated surrogate, and solve it exactly. The guard $|\widehat{c}| \geq e/16$ is passed on the good event. The ratio bounds in equation 25 imply that each estimated surrogate entry differs from its true value by at most $10t$. The restricted saddle pair $z = (x, y)$ consequently has restricted B-gap at most $20t$. Rank-one endpoint transfer adds at most $2t$ per side. Thus

$$\mathrm{Gap}_B(x, y) \leq 24t. \tag{26}$$

Run fresh PAC row and column best-response searches for z in the original matrix, and estimate the two selected response values to error t. Let $\widehat{g}$ be their estimated difference. If $\widehat{g} \leq e/2$, return z, since its true duality gap is at most $\widehat{g} + 4t < e$. Otherwise let i_R, j_C be the selected labels and put $E = A - B$. The selected true A-response difference exceeds $e/2 - 2t$, whereas its B-response difference is at most $24t$. Hence

$$E_{i_R,:}y - x^\top E_{:,j_C} > e/2 - 26t > e/3.$$

At least one of the two residual mixture values has magnitude greater than $e/6$.

Here these residuals can be sampled at constant variance without accurate inverse estimation. Estimate $A_{i_R b}$, $A_{a j_C}$ and c freshly to accuracy t, writing $\bar{A}_{i_R b}$ and $\bar{A}_{a j_C}$ for the fresh empirical cell means and $\widehat{c}$ for the pivot estimate. Freeze

$$\widehat{\alpha} = \bar{A}_{i_R b}/\widehat{c}, \qquad \widehat{\beta} = \bar{A}_{a j_C}/\widehat{c}.$$

The public guards $|\widehat{\alpha}|, |\widehat{\beta}| \leq 3$ are passed. Each differs from its true ratio by at most $4t/p$. Therefore the score

$$A_{i_R j} - \widehat{\alpha} A_{aj}$$

differs from $E_{i_R j}$ by at most $8t$, uniformly over j. The analogous column score $A_{i j_C} - \widehat{\beta} A_{ib}$ has the same guarantee. Two independent cell samples implement each score with a universal subgaussian proxy. Sampling the corresponding mixtures estimates each residual mixture value with total error at most $9t$. Select the larger estimated absolute value. Its true residual mixture magnitude is greater than $e/6 - 18t > e/8$. A PAC absolute maximum over that entire residual row or column, with score accuracy t, returns a cell (i, j) with

$$|E_{ij}| > e/8 - 17t > e/16.$$

Every primitive here has a deterministic $O((n+m)e^{-2})$ cap at constant risk, and retains the logarithm of the allocated risk at general confidence.

Append this row and column to the pivot. Let $d = E_{ij}$, so the resulting block is

$$C_0 = \begin{pmatrix} c & u \\ v & d + uv/c \end{pmatrix}, \qquad |u|, |v| \leq 2p, \quad |d| > e/16.$$

Appendix table of contents

Its determinant is cd, and $\|C_0\|_2 \le \|C_0\|_F \le 9p + |d|$. Consequently

$$\sigma_{\min}(C_0) = \frac{|cd|}{\sigma_{\max}(C_0)} \ge \frac{p|d|}{9p + |d|} \ge \frac{1}{10} \min\{p, |d|\} \ge e/160.$$

This bound uses the relative control of the whole first pivot row and column, not merely a determinant lower bound. Since residuals vanish on the pivot row and column, the appended labels are distinct. The Schur rank identity gives $\operatorname{rank}(A) = 1 + \operatorname{rank}(E)$. Thus a nonzero residual is impossible in rank at most one and the new block is nonsingular in rank two. To implement the final screen, estimate its four entries freshly to error t and require the estimated smallest singular value to be at least $e/320$. The proved bound and Weyl's inequality make this screen pass. Duplicate-label and screen failures return the fallback off the good event. □

P.4.2 Empirical volume purification at physical precision

Lemma P.14 (Capped empirical row purification)**.** *Fix two column labels J, and suppose that two supplied rows I_0 satisfy $\sigma_{\min}(A_{I_0,J}) \ge \kappa e$, for a known constant $\kappa \in (0, 1]$. There is a capped procedure which, with probability at least $1 - \beta$, returns two rows I such that*

$$\max_i \|A_{i,J} A_{I,J}^{-1}\|_\infty \le 2, \qquad \sigma_{\min}(A_{I,J}) \ge \kappa e/8.$$

Its cost is $C_\kappa n e^{-2} \log(2/e) \log(c_\kappa \log(2/e)/\beta)$. The transposed statement purifies columns. No accurate estimate of the true inverse is required.

Proof. Put $h = \kappa e/2^{16}$. Estimate each coordinate of the initial two row points $p_i = A_{i,J}$ to error h, using fresh samples and clipping to $[-1, 1]$. Store these empirical points. Maintain a set R of selected labels and their stored estimates. At every pass choose the two empirical points in R of largest absolute determinant, with deterministic tie breaking, and let $\widehat{C}$ be this empirical 2×2 matrix. Screen $\sigma_{\min}(\widehat{C}) \ge \kappa e/4$ before inverting it.

For each coordinate $a = 1, 2$, use a $1/8$-PAC absolute maximum of $p_i \widehat{C}^{-1} e_a$ over all n rows, then estimate its selected score to error $1/8$. These are directly observed linear combinations of the two physical cells with weights $\widehat{C}^{-1} e_a$. Their subgaussian proxy is at most $C/(\kappa e)^2$ on every screened transcript. If both selected estimated magnitudes are at most $3/2$, return the two current labels. Otherwise select a violating label and coordinate, estimate its two physical coordinates freshly to error h, append this point to R, and repeat. A duplicate label or an exhausted append cap causes the fixed-pair fallback in the encompassing algorithm.

We justify the screen and the cap on the simultaneous good event. By maximality of the empirical determinant, every stored point has both interpolation coordinates in $[-1, 1]$ in the current empirical basis. In particular, the initial empirical block can be written $\widehat{C}_0 = M\widehat{C}$ with $\|M\|_2 \le 2$. Since $\sigma_{\min}(\widehat{C}_0) \ge \kappa e - 2h \ge \kappa e/2$,

$$\sigma_{\min}(\widehat{C}) \ge \kappa e/4.$$

Thus the inverse screen always passes and $\|\widehat{C}^{-1}\|_2 \le 4/(\kappa e)$. A fresh coordinate error of at most h changes either interpolation score by at most $4\sqrt{2}h/(\kappa e) < 1/16$.

For an appended point, its true selected score exceeds $3/2 - 1/8 = 11/8$. Its stored empirical score therefore exceeds $21/16 > 5/4$. Replacing that coordinate's basis row increases the empirical determinant by more than $5/4$. Re-maximizing the determinant can only improve it. All empirical coordinates lie in $[-1, 1]$, so every such determinant is at most two in magnitude. The initial empirical determinant is at least $(\kappa e)^2/4$. Thus it suffices to allow

$$K_{\mathrm{vol}} = 2 + \left\lceil \frac{\log(32/(\kappa e)^2)}{\log(5/4)} \right\rceil$$

appends. A previously stored label has empirical score at most one and true score at most $1 + 1/16$, so it cannot trigger an append on this event.

On termination, $\gamma_i = p_i \widehat{C}^{-1}$ satisfies $\|\gamma_i\|_\infty \le 3/2 + 1/8 + 1/8 = 7/4$ for every row. Writing the true selected block as $C = \widehat{C} + F$, we have $\|F\|_2 \le 2h$ and $\|F\widehat{C}^{-1}\|_2 \le 8h/(\kappa e) < 2^{-12}$.

Consequently

$$p_i C^{-1} = \gamma_i (I + F\widehat{C}^{-1})^{-1}, \qquad \|p_i C^{-1}\|_\infty \le \frac{7/4}{1 - \sqrt{2}\, 8h/(\kappa e)} < 2.$$

Weyl's inequality gives $\sigma_{\min}(C) \ge \kappa e/4 - 2h \ge \kappa e/8$. The at most $K_{\mathrm{vol}} + 1$ passes have two PAC scans with proxy $C/(\kappa e)^2$ at constant accuracy, and a constant number of scalar or selected-coordinate estimates at precision proportional to e. Dividing the risk over this public schedule proves the claimed cost. The matrix inverse used to define a query is the inverse of the stored empirical matrix and is computed exactly after its screen. No assertion of $O(1)$ absolute error in the true inverse has been used. □

Apply the lemma first to rows, and then to columns with the resulting rows held fixed. Exact rank two makes row interpolation independent of the choice of a nonsingular column pair. Therefore the second purification preserves the first one's row guarantee. Starting from Lemma P.13, the final true block $C = A_{I,J}$ satisfies

$$\sigma_{\min}(C) \ge \varepsilon/10240, \qquad \|\alpha_i\|_\infty \le 2, \quad \|\beta_j\|_\infty \le 2, \tag{27}$$

where $A_{i,:} = \alpha_i^\top A_{I,:}$ and $A_{:,j} = A_{:,J}\beta_j$. No side information was assumed to obtain these bounds. If a routine has already returned an equilibrium, this stage is not run.

P.4.3 Weak membership from noisy selected support points

Lemma P.15 (Two-dimensional weak hull membership)**.** *Let unknown points $q_j \in [-1,1]^2$ be accessible as follows. For each unit vector u, an oracle returns a label j with $u^\top q_j \ge \max_k u^\top q_k - \eta$ and a fresh estimate $\widehat{q}_j \in [-1,1]^2$ with $\|\widehat{q}_j - q_j\|_\infty \le \eta$. For $z \in [-12,12]^2$ and $0 < \eta \le 1/64$, there is a procedure using at most $C\log(2/\eta)$ oracle calls that returns either*

1. *a distribution y on at most three returned labels with $\|\sum_j y_j q_j - z\|_2 \le 42\eta$, or*
2. *a unit direction u satisfying $u^\top z > \max_j u^\top q_j + 5\eta$.*

The conclusions hold whenever every invoked oracle meets its stated guarantees. The call cap and a valid output distribution or direction are defined on every transcript. For directions of norm at most one, divide by their norm before invoking the support oracle. Its additive guarantee only becomes stronger after rescaling back.

Proof. The following elementary central-cut ellipsoid fact will also be used below. For an ellipse $\{x : (x-c)^\top Q^{-1}(x-c) \le 1\}$ and a nonzero vector g, its intersection with $g^\top(x-c) \le 0$ is contained in the ellipse with

$$c' = c - \frac{Qg}{3\sqrt{g^\top Q g}}, \qquad Q' = \frac{4}{3}\left(Q - \frac{2Qgg^\top Q}{3g^\top Q g}\right).$$

The area is multiplied by $\chi = 4/(3\sqrt{3}) < 1$. To verify containment, transform to the unit disk and the cut $x_1 \le 0$. The proposed ellipse has center $(-1/3, 0)$ and squared radii $4/9, 4/3$. On the half disk its quadratic form is at most $1 + (3/2)x_1(x_1+1) \le 1$.

Run this central-cut method in the dual variable ℓ, starting from the unit disk. If a center satisfies $\|\ell\|_2 > 1$, make the central cut pointing toward the unit disk. Otherwise call the support oracle in direction ℓ (any direction if $\ell = 0$), store its empirical point, and put $g = z - \widehat{q}_j$. If $\|g\|_2 \le \eta$, return that pure label. If $\ell^\top g > 8\eta$, return $u = \ell/\|\ell\|_2$. This is a genuine separator because

$$\max_k \ell^\top q_k \le \ell^\top \widehat{q}_j + (1+\sqrt{2})\eta < \ell^\top z - 5\eta,$$

and normalization only increases the gap. Otherwise retain the central halfspace $g^\top(\ell' - \ell) \ge 0$ and continue.

Let $B = 20$, so $\|z - \widehat{q}_j\|_2 < B$ for every stored point, and put $r = \eta/B$. Use a public number K of iterations with $\chi^K < r^2$. If no early output is produced, project z onto the convex hull of all stored

empirical points and return a convex representation of that projection using at most three points. Such a representation is computed by examining the finite planar hull and its edges or triangles. The stored hull is nonempty, since the initial center is zero and hence the first iteration invokes the support oracle.

To prove its accuracy, suppose the distance from z to this final empirical hull were greater than 40η. Projection onto a compact convex set provides a unit vector ℓ_* such that $\ell_*^\top(z - \widehat{q}_j) > 40\eta$ for every stored point. The disk of radius r centered at $\ell_*/2$ lies in the dual unit disk and satisfies

$$(\ell')^\top(z - \widehat{q}_j) > 20\eta - Br = 19\eta$$

for every point and every ℓ' in this disk. It is retained by every support cut, because the queried center had inner product at most 8η, and by every unit-disk cut. The final ellipse would therefore contain area at least πr^2, contrary to its area $\pi\chi^K$. The empirical projection distance is at most 40η. Its true mixture differs from its empirical counterpart by at most $\sqrt{2}\eta$, proving the claimed 42η. The pure-label exit has error at most $(1 + \sqrt{2})\eta$. If malformed numerical output is encountered, return a fixed distribution. This branch is not reached under the assumptions. □

For physical column points $q_j = A_{I,j}$, the required support oracle is a PAC maximum of $u_1 A_{i_1 j} + u_2 A_{i_2 j}$, followed by fresh estimates of only the selected two coordinates. Its subgaussian proxy is universal and its cost is $Cm\eta^{-2}\log(c/\alpha)$ at conditional risk α. There is no union bound over all m columns.

P.4.4 A BOUNDED-COORDINATE FEASIBILITY ORACLE

Choose

$$\eta = \varepsilon/1024, \qquad h = \varepsilon/2^{40}.$$

Estimate the four entries of the purified C on fresh streams to error h, clipping each entry to $[-1, 1]$ and obtaining $\widehat{C}$. Screen $\sigma_{\min}(\widehat{C}) \geq \varepsilon/20480$ and $\|\widehat{C}\|_2 \leq 3$ before forming its inverse. These screens pass under equation 27. The small h ensures that both $\|C\widehat{C}^{-1} - I\|_2$ and $\|\widehat{C}^{-1}C - I\|_2$ are at most $1/100$. Let

$$p_i = A_{i,J}, \quad q_j = A_{I,j}, \quad Q = \mathrm{conv}\{q_j\}, \quad V_A^\star = \min_{y \in \Delta_m} \max_i e_i^\top A y.$$

These actual physical coordinates, rather than estimated normalized features of every action, are the only vectors queried by the procedure.

For a target $v \in [-2, 2]$, run a central-cut ellipsoid in $b \in [-4, 4]^2$, starting with a disk of radius eight. If its center is outside that box, make the corresponding coordinate cut. Otherwise form $z = \widehat{C}b \in [-12, 12]^2$ and call Lemma P.15 at tolerance η.

If a separating unit vector u is returned, make the central cut

$$(\widehat{C}^\top u)^\top(b' - b) \leq 0.$$

If a mixture y is returned, run an η-PAC row maximum of $p_i b$. The coefficient vector b has ℓ_1 norm at most eight, so one-query linear scores have a universal subgaussian proxy: for $b \neq 0$, select coordinate k with probability $|b_k|/\|b\|_1$, query cell (i, J_k), and return $\|b\|_1 \operatorname{sign}(b_k) X$. Its mean is $p_i b$ and its centered proxy is at most $2\|b\|_1^2 \leq 128$. For $b = 0$, all scores are exactly zero and any row can be selected without a PAC query. After selecting row i, estimate its two physical coordinates to accuracy $\eta/16$, obtaining $\widehat{p}_i$. If

$$\widehat{p}_i b \leq v + 2\eta,$$

return y. Otherwise make the central cut $\widehat{p}_i(b' - b) \leq 0$. Allow $K_{\mathrm{out}} = C\log(2/\eta)$ iterations, with the universal constant chosen so that the final ellipse would have area less than $\pi(\eta/4)^2$, and return the symbol FAIL if no mixture has been returned. If a violated row inequality has $\widehat{p}_i = 0$, return FAIL immediately. This can occur at an infeasible target and must not abort the encompassing optimization. The proof below shows that it cannot occur at a target $v \geq V_A^\star + 32h$ on the good event. Other failed numerical screens cause the encompassing fallback. A membership-cut normal cannot vanish because $\widehat{C}$ is screened nonsingular and the returned direction is a unit vector.

Lemma P.16 (Feasibility guarantee)**.** *On the simultaneous good event, a returned mixture satisfies*

$$\max_i e_i^\top Ay \le v + 172\eta + 32h.$$

If $v \ge V_A^\star + 32h$, the procedure returns a mixture before its cap. Each call makes $O(\log^2(2/\eta))$ physical support-oracle calls and $O(\log(2/\eta))$ additional row PAC calls.

Proof. For a returned mixture, the PAC and selected-coordinate errors give

$$\max_i p_i b \le \widehat{p}_i b + \eta + 8(\eta/16) \le v + 4\eta.$$

Its physical column mixture satisfies $\|q_y - \widehat{C}b\|_\infty \le 42\eta$. Because $A_{i,:} = \alpha_i^\top A_{I,:}$, $p_i = \alpha_i^\top C$, and $\|\alpha_i\|_1 \le 4$, we obtain

$$|e_i^\top Ay - p_i b| \le 4(42\eta + 8h) = 168\eta + 32h.$$

This proves the output guarantee without estimating any α_i.

For termination, take an optimal column mixture and write $q_* = C\beta_*$, where $\|\beta_*\|_\infty \le 2$ by column dominance. Put $b_* = \widehat{C}^{-1}q_*$. The relative block perturbation bounds give $\|b_*\|_\infty < 3$. Moreover, using $C\widehat{C}^{-1}C - C = -C\widehat{C}^{-1}(\widehat{C} - C)$,

$$|p_i b_* - e_i^\top Ay_*| \le 32h, \qquad \max_i p_i b_* \le V_A^\star + 32h \le v.$$

Consider the disk $\|b' - b_*\|_2 \le r_0 = \eta/4$. It lies in the box $[-4, 4]^2$. A membership separator at the current center obeys $u^\top \widehat{C}b > \max_j u^\top q_j + 5\eta$. At every point in the disk,

$$u^\top \widehat{C}b' \le u^\top q_* + \|\widehat{C}\|_2 r_0 \le \max_j u^\top q_j + 3\eta/4,$$

so the disk is retained by that central cut. For a row cut, the stored estimate has $\|\widehat{p}_i\|_2 \le 2$ and

$$\widehat{p}_i b' \le p_i b_* + \|\widehat{p}_i - p_i\|_1 \|b_*\|_\infty + 2r_0 \le v + 3\eta/8 + \eta/2 < v + 2\eta.$$

The current center violates the last threshold, so the whole disk is again retained. All cuts preserve this disk, contradicting the final area cap if no mixture has been returned. The oracle-call count follows by multiplying the two public logarithmic ellipsoid schedules. □

P.4.5 Optimization, query accounting, and output

Run the feasibility procedure at $v = 2$ and store its successful mixture. If it fails, take the guarded fallback. Initialize $\ell = -2, u = 2$. For a public $O(\log(2/\eta))$ number of binary-search iterations, call the procedure at $v = (\ell + u)/2$. On success replace u by v and store the new mixture. On FAIL replace ℓ by v. Stop when $u - \ell \le \eta$. On the good event every FAIL satisfies $v < V_A^\star + 32h$, while the initial lower endpoint is below that threshold. Therefore the last stored mixture has

$$\max_i e_i^\top Ay \le V_A^\star + 173\eta + 64h < V_A^\star + \varepsilon/4.$$

Apply the same optimization procedure to $-A^\top$, using the transposed purified skeleton and fresh oracle streams. Its output x satisfies

$$\min_j x^\top Ae_j \ge V_A^\star - \varepsilon/4.$$

The output pair thus has duality gap at most $\varepsilon/2$, hence each individual regret at most ε.

For a completely public risk schedule, write $\chi = 4/(3\sqrt{3})$ and take

$$K_{\mathrm{mem}} = 1 + \left\lceil \frac{2\log(20/\eta)}{\log(1/\chi)} \right\rceil, \quad K_{\mathrm{out}} = 1 + \left\lceil \frac{2\log(32/\eta)}{\log(1/\chi)} \right\rceil, \quad K_{\mathrm{bin}} = 2 + \lceil \log_2(4/\eta) \rceil.$$

Let K_1, K_2 be the displayed volume append caps with $\kappa = 1/160, 1/1280$, respectively, and let K_{asc} be the displayed ascent cap. All required subroutines have at most

$$M = 64\,[1 + K_{\mathrm{asc}} + K_1 + K_2 + 2K_{\mathrm{bin}}K_{\mathrm{out}}(K_{\mathrm{mem}} + 1)] \le CL^3$$

grouped PAC/selected-estimate calls. Allocate conditional risk δ/M to each possible call, dividing it further among its constant number of PAC and scalar events. Selected labels and estimated blocks are frozen before their dependent observations. A union bound is therefore valid under adaptive selection. Every physical precision is a universal constant times ε, except the purification scans: their accuracy is constant and their screened score proxy is $C\varepsilon^{-2}$, giving the same physical rate. Each support call costs at most $C(n+m)\varepsilon^{-2}\log(cM/\delta)$. The total is

$$C(n+m)\varepsilon^{-2}L^3\log(cL/\delta).$$

At most three selected column labels suffice for each membership mixture, and the final transpose call gives the same support bound for the row strategy. Every returned pair from an earlier branch already has the claimed guarantee on this same good event. The simultaneous event has probability at least $1-\delta$. On its complement, the public guards and caps preserve totality and the deterministic query budget, completing the proof of Theorem P.12.

P.5 Perturbing the planar support primitive

We now allow the observed game to deviate from exact skew rank two. The first lemma quantifies how coordinate and support errors affect the planar construction. The results that follow apply this bound to structured perturbations, keeping the error radius separate from the sampling tolerance.

Lemma P.17 (Extended range of the planar support primitive)**.** *In the notation of the planar support lemma, let $K \geq 1$ and $0 < \tau \leq 1/(16K)$. Assume each support call returns $i, \widehat{z}_i$ with*

$$f_i(t) \geq F(t) - \tau, \qquad \|\widehat{z}_i - z_i\|_\infty \leq \tau.$$

The same saddle-and-root routine, followed by the estimated-coordinate compression of Lemma P.7, returns a distribution q on at most three labels with

$$\max_i \det(z_i, z_q) \leq 54K\tau, \qquad z_q = \sum_j q_j z_j,$$

using at most $8 + 2\lceil\log_2(2K/\tau)\rceil$ support calls. As before, the root search has its explicit cap and returns a fixed policy on exhaustion. The cap holds on every transcript.

Proof. We record why the larger tolerance range is legitimate, rather than substituting it into a lemma with a narrower hypothesis. In the saddle subroutine, all arguments use only that f_i, F are $2K$-Lipschitz, $|\widehat{f}_i(t) - F(t)| \leq 2\tau$, and $|\widehat{s}_i - s_i| \leq 2\tau$. They give, for the returned dual point t_D and one- or two-label mixture with feature (U, V),

$$\min(U, -V) \geq F(t_D) - 6\tau.$$

None of these inequalities assumes $64K\tau \leq 1$. In the branch $\widehat{F}_D \geq -16\tau$, the same anchor correction has both feature coordinates bounded by 25τ in absolute value, hence exploitability at most $50K\tau$. In the negative branch, the endpoint output has exploitability at most 6τ. A root output satisfies $s_i > 0$, $F(t) \leq 6\tau$, and $|f_i(t)| \leq 7\tau$, giving exploitability at most $26K\tau$. Since $K \geq 1$, both bounds are at most $50K\tau$ before compression. Only the anchor-correction branch can contain four labeled occurrences. Applying Lemma P.7 to the stored coordinates changes each determinant by at most $4K\tau$ and leaves at most three occurrences. The other branches are already pure. Hence the final bound is $54K\tau$ without any additional observation. The two bisections halve their interval each time. Their terminal width is $\tau/(2K)$. Our enlarged range still gives $\tau/(2K) < 1$, and therefore $J = \lceil\log_2(2K/\tau)\rceil$ is positive. The original counting argument, including the root fallback after $J+1$ iterations, applies unchanged. Thus the earlier condition $\tau \leq 1/(64K)$, which originated from parameterizing a target $\alpha = 64K\tau \leq 1$, is unnecessary for the geometric proof. The explicitly stated larger range suffices. □

Theorem P.18 (Unknown-radius skew rank-two tube, without a dimension logarithm)**.** *Let $n \geq 2$, $0 < \varepsilon \leq 1/8$, and $0 < \delta \leq 1/4$. Suppose the observed matrix $A \in [-1,1]^{n\times n}$ is skew-symmetric and there is a skew-symmetric B of rank at most two satisfying*

$$\|A - B\|_\infty \leq \rho \leq \varepsilon/512.$$

The algorithm specified below is not given B, ρ, a factorization, an anchor, or any conditioning information. It returns a distribution q supported on at most three labels with

$$f_A(q) := \max_i e_i^\top Aq \le \varepsilon$$

with probability at least $1-\delta$, using a deterministic query cap

$$Cn\varepsilon^{-2}\log(2/\varepsilon)\log\left(\frac{\log(2/\varepsilon)}{\delta}\right).$$

In its structural branch, the stronger bound is

$$f_A(q) \le \frac{8019}{8192}\varepsilon + \rho \le \frac{8035}{8192}\varepsilon < \varepsilon. \tag{28}$$

The initial pure-output branch instead satisfies $f_A(q) \le 33\varepsilon/64$.

Exact changes to the previous algorithm. Retain the initial PAC scan, scalar screen, two absolute scans, and pivot estimate, with their physical precisions

$$t = \varepsilon/128, \qquad h_0 = \varepsilon/2^{20}, \qquad \eta = \varepsilon/4096.$$

After orienting the estimated pivot positively, change its public guard to $\varepsilon/3 \le \widetilde{p} \le 2$. Use

$$K = 9/4, \qquad \tau = \frac{\varepsilon}{128\widetilde{p}}, \qquad M_* = 8 + 2\left\lceil \log_2 \frac{1152}{\varepsilon} \right\rceil.$$

Run exactly the same planar saddle-and-root procedure, now under Lemma P.17. Every physical PAC support scan still uses accuracy η, and each selected physical coordinate is still estimated to accuracy η before division by $\widetilde{p}$. Give the four initial grouped calls risk $\delta/16$ each, the pivot estimate risk $\delta/4$, and each support call total risk $\beta = \delta/(2M_*)$, split between its PAC scan and its selected-coordinate estimates. All estimates following an adaptive label use fresh observations. Every guard failure and public-cap exhaustion returns a fixed policy. These are the only modifications. In particular, there is no estimate of ρ and no simultaneous estimation of all n feature vectors.

Proof. Condition first on the four initial estimation events. The pure branch is valid directly in the observed skew matrix A: its PAC response and fresh scalar estimate give

$$f_A(e_a) \le \varepsilon/2 + 2t = 33\varepsilon/64.$$

If it rejects, the first absolute scan returns b with $|A_{ba}| > 62t$. The second absolute scan returns c with $|A_{bc}| > 61t$. These two lower bounds use skew symmetry of A, but do not use its rank.

Write $d = |B_{ba}|$, $p = |B_{bc}|$, and $t' = t + 2\rho$. Every t-near absolute scan in A is a t'-near absolute scan in B. In addition, the observed-pivot lower bounds give directly

$$d > 62t - \rho \ge 247t/4, \qquad p > 61t - \rho \ge 243t/4, \qquad t' \le 3t/2.$$

Let $M_a^B = \max_i |B_{ia}|$ and $M_b^B = \max_i |B_{ib}|$. Then $M_a^B \le d + t'$ and $M_b^B \le p + t'$. The rank-two skew identity in B gives

$$B_{ic} = \frac{B_{ib}B_{ac} - B_{ia}B_{bc}}{B_{ab}}.$$

Consequently the two columns of the new cross satisfy

$$\frac{\max_i |B_{ib}|}{p} \le 1 + \frac{6}{243} = \frac{83}{81} < \frac{9}{4},$$
$$\frac{\max_i |B_{ic}|}{p} \le \left(1 + \frac{6}{247}\right)\left(2 + \frac{6}{243}\right) = \frac{41492}{20007} < \frac{9}{4}.$$

Thus B has a stable cross with the smaller common constant $K = 9/4$, uniformly over all allowed model perturbations.

The fresh pivot estimate differs from A_{bc} by at most h_0. Orient it positively by swapping b, c if necessary, and call it $\widetilde{p}$. Since $|A_{bc}| > 61t$ and $\rho + h_0 < 61t$, its sign agrees with the corresponding

nonzero B pivot. After orientation that true pivot is $p > 0$. The guard passes on the good event because it is a scalar estimate of an observed entry of magnitude in $(61\varepsilon/128, 1]$. Moreover,

$$|\widetilde{p} - p| \le \rho + h_0 \le \varepsilon/256, \qquad p > \varepsilon/3.$$

It follows that $|\widetilde{p} - p|/p < 3/256$ and hence, with slack,

$$\frac{p}{\widetilde{p}} \le \frac{33}{32}, \qquad \frac{\widetilde{p}}{p} \le \frac{33}{32}. \tag{29}$$

Use the true but unknown normalized B features

$$z_i = (B_{ib}/p, B_{ic}/p).$$

They lie in $[-K, K]^2$, have known anchor coordinates $z_b = (0, 1), z_c = (-1, 0)$, and satisfy $B_{ij} = p \det(z_i, z_j)$. The strengthened guard ensures

$$0 < \tau \le 3/128 < 1/36 = 1/(16K),$$

so the extended planar lemma applies. It also ensures $2K/\tau \le 1152/\varepsilon$, establishing the public count M_* on every transcript that passes the guard.

It remains to verify the oracle relative to B, although every physical sample is from A. At direction $s \in [0, 1]$, the physical score is

$$g_i^A(s) = (1 - s)A_{ib} - sA_{ic}.$$

The analogous B score is $g_i^B(s) = pf_i(s)$, and $|g_i^A(s) - g_i^B(s)| \le \rho$ for every label. Therefore an η-PAC maximum for g^A returns a label whose B score is within $\eta + 2\rho$ of $\max_i g_i^B$. By equation 29,

$$\frac{\eta + 2\rho}{p\tau} \le \frac{17\varepsilon/4096}{\varepsilon/128} \frac{\widetilde{p}}{p} \le \frac{561}{1024} < 1.$$

Thus its normalized true score meets the required τ-optimality. This calculation does not use an unjustified condition-number-free inverse estimate. It uses the explicit relative pivot bound.

For a selected physical coordinate, a fresh estimate $\widehat{a}_i$ satisfies $|\widehat{a}_i - B_{ib}| \le \eta + \rho$ (and likewise for column c). Since $|B_{ib}| \le Kp$,

$$\begin{aligned}\left|\frac{\widehat{a}_i}{\widetilde{p}} - \frac{B_{ib}}{p}\right| &\le \frac{\eta + \rho + K(h_0 + \rho)}{\widetilde{p}} \\ &\le \frac{27657}{32768}\tau < \tau.\end{aligned}$$

The exact fraction uses $K = 9/4$, $\rho \le \varepsilon/512$, $\eta = \varepsilon/4096$, $h_0 = \varepsilon/2^{20}$, and $\tau\widetilde{p} = \varepsilon/128$. For an anchor label, the known exact normalized coordinates may instead be inserted. This proves every part of the support-oracle contract relative to the latent matrix B.

All PAC and scalar events concern samples from A. The comparisons with B are deterministic perturbation bounds. Conditional on each adaptive history, the current support direction, cross labels, and pivot estimate are frozen before fresh samples are drawn. The support call meets its physical estimation events with conditional probability at least $1 - \beta$. A union bound gives simultaneous probability at least

$$1 - 4\delta/16 - \delta/4 - M_*\beta = 1 - \delta.$$

On this event, Lemma P.17 yields

$$f_B(q) \le 54Kp\tau \le \frac{243}{256}\frac{33}{32}\varepsilon = \frac{8019}{8192}\varepsilon.$$

Finally $f_A(q) \le f_B(q) + \rho$, proving equation 28.

Each physical score is implemented using one randomized cell query: choose column b with probability $1 - s$ and use the observation, or choose column c with probability s and negate it. Its centered subgaussian proxy is at most two, because all observed means belong to $[-1, 1]$. Thus each support call costs $O(n\eta^{-2}\log(c/\beta)) = O(n\varepsilon^{-2}\log(cM_*/\delta))$. The initial calls and pivot estimate have the unchanged $O(n\varepsilon^{-2}\log(c/\delta))$ cap. Since $M_* = O(\log(2/\varepsilon))$, the stated bound follows. All caps and guards are public. Outside the good event, the algorithm is still a total, capped candidate generator. Its accuracy guarantee uses the good event only. No assumption on perturbation events or radius estimates has been introduced. □

Relation to the unmodified policy. For clarity, the theorem above concerns the explicitly retuned policy. The original planar policy, with $K = 3$ and $\tau = \varepsilon/(1024\widetilde{p})$, also has unknown-radius robustness for $\rho \le \varepsilon/8192$. Indeed its score error is at most $(\eta + 2\rho)/p \le \tau$, and its normalized coordinate error is at most

$$\frac{\eta + 4\rho + 3h_0}{\widetilde{p}} \le \frac{771}{1024}\tau < \tau.$$

The original stable-cross, guard, and risk arguments then give $f_A(q) \le 3\varepsilon/8 + \rho$ in the structural branch. The broader $\varepsilon/512$ tube is obtained by the stated changes to K, the planar tolerance, the pivot guard, and the primitive's tolerance range. It should not be attributed to the unmodified implementation.

P.6 A wider neighborhood of shifted skew games

This section proves robustness of the fast planar construction. The observed matrix need not be skew-symmetric or low rank. Its distance to the structural class is not an algorithm input. The improvement comes from solving an exact surrogate made from the observed cross, followed by a deterministic perturbation argument. Observations of the perturbed matrix are never claimed to be unbiased observations of a hidden structured game.

Theorem P.19 (Unknown-radius fast skew-rank-two tube)**.** *Let $n \ge 2$, $0 < \varepsilon \le 1/8$, and $0 < \delta \le 1/4$. There is a Gaussian physical-cell-query algorithm taking only (n, ε, δ) as input. On every $A \in [-1,1]^{n\times n}$ it has the deterministic cap*

$$Cn\varepsilon^{-2}\log(2/\varepsilon)\log\left(\frac{\log(2/\varepsilon)}{\delta}\right)$$

and returns a distribution q on at most three labels. Suppose there exist a real skew-symmetric matrix B with $\operatorname{rank}(B) \le 2$ and a scalar c such that $\|A - (B + c\mathbf{1}\mathbf{1}^\top)\|_\infty \le \rho \le \varepsilon/16$. Then, with probability at least $1 - \delta$, its symmetric pair (q, q) satisfies

$$\max\{R_A(q,q), C_A(q,q)\} \le \frac{7}{8}\varepsilon.$$

In particular it is an ε-equilibrium of the actually queried matrix A. No factorization, anchor, conditioning parameter, skewness test, rank test, or radius bound is supplied to the algorithm. On the non-pure branch the sharper bound is

$$\max\{R_A(q,q), C_A(q,q)\} \le 13\rho + \frac{3}{64}\varepsilon. \tag{RPT}$$

The support-three claim uses the estimated-coordinate compression of Lemma P.7.

Symmetrization costs one physical query per observation. Set $S = (A - A^\top)/2$. To sample a scalar with mean S_{ij}, choose with probability one half between querying (i, j) and using its Gaussian observation, or querying (j, i) and negating its observation. These are fresh independent choices for each requested observation. Its mean is S_{ij}, and its centered subgaussian proxy is at most two, since it is a unit Gaussian plus an independent random mean in $[-1, 1]$. On the diagonal one may use the known zero without sampling. Moreover,

$$\|S - B\|_\infty \le \rho, \qquad \|A - S - c\mathbf{1}\mathbf{1}^\top\|_\infty \le \rho. \tag{SYM}$$

The second inequality follows by writing $E = A - (B + c\mathbf{1}\mathbf{1}^\top)$ and $A - S - c\mathbf{1}\mathbf{1}^\top = (E + E^\top)/2$. Thus it does not require A to be skew. The constant shift cancels in both individual regrets.

Algorithm. Every invocation uses fresh observations conditional on all preceding labels and estimates. Set

$$t = \varepsilon/1024, \quad H = 3\varepsilon/4 - 2t = 766t, \quad h_0 = \varepsilon/2^{20},$$
$$\eta = \varepsilon/32768, \qquad M_* = 8 + 2\left\lceil \log_2(98304/\varepsilon)\right\rceil.$$

1. Fix $a = 1$. Run a t-PAC maximum on column a of S, obtaining j, then estimate S_{ja} to accuracy t on a fresh stream. If that scalar estimate is at most H, return $q = e_a$.

2. Run a t-PAC absolute maximum on column a of S, returning b. Then run one on column b, returning c. Give each of the four grouped calls so far (the first PAC call, its scalar estimate, and the two absolute scans) risk $\delta/16$.
3. Estimate S_{bc} on a fresh stream to error h_0, with risk $\delta/4$. If this estimate is negative, swap b, c and negate the estimate. Denote the resulting nonnegative estimate by $\widetilde{p}$. If $\widetilde{p} \notin [\varepsilon/2, 2]$, return e_1.
4. Set $\tau = \varepsilon/(8192\widetilde{p})$ and run the planar support primitive with $K = 3$, precision τ, and its public hard caps. At each requested direction $u \in [0, 1]$, use an η-PAC maximum of the physical means $(1-u)S_{ib} - uS_{ic}$. Give it risk $\beta/2$, where $\beta = \delta/(2M_*)$. Then, after its label is frozen, estimate its two cross coordinates on fresh streams to accuracy η, with combined risk $\beta/2$, and divide by $\widetilde{p}$. Exact anchor coordinates can be inserted for b, c. If the primitive returns more than three labels, compress its final mixture to three using the estimated coordinates of those labels. Return the resulting q.

Each physical-score observation in the last step first selects column b or c and a sign, and then a symmetrization orientation. Its random conditional Gaussian mean is still in $[-1, 1]$, so the centered proxy is at most two. One observation is still exactly one physical Gaussian query. An absolute maximum is implemented by PAC maximization over the $2n$ signed arms. All routine costs and risks therefore have the same orders as for the original Gaussian PAC primitive.

Lemma P.20 (Observed-cross residual under unknown perturbation)**.** *On the simultaneous good event of the first four calls, if the pure branch was not taken, then the selected observed cross satisfies*

$$d := |S_{ba}| > 764t, \qquad p := |S_{bc}| > 763t.$$

If $\|S - B\|_\infty \le \rho \le 64t$ for a skew matrix B of rank at most two, the two observed columns obey

$$\max_i |S_{ib}| \le 3p, \qquad \max_i |S_{ic}| \le 3p.$$

The exact observed-cross matrix

$$D_{ij} := \frac{S_{ib}S_{jc} - S_{ic}S_{jb}}{S_{bc}} \tag{CROSS}$$

is skew-symmetric of rank at most two, preserves both observed cross columns, and satisfies

$$\|D - B\|_\infty \le 11\rho. \tag{RES}$$

The notation D here is a surrogate matrix, not an ambient dimension.

Proof. No structural assumption is needed for the pivot lower bounds. Rejection of the pure branch gives $S_{ja} > H - t = 765t$. The first absolute scan gives $d > 764t$. Skew symmetry of the actually scanned matrix S implies that column b contains the entry S_{ab} of magnitude d. Therefore the second absolute scan gives $p > d - t > 763t$.

Let $M_a = \max_i |S_{ia}|$ and $M_b = \max_i |S_{ib}|$. The two PAC guarantees give $M_a \le d + t$ and $M_b \le p + t$. For hidden-game quantities write $d_B = |B_{ba}|$, $p_B = |B_{bc}|$ and $M_x^B = \max_i |B_{ix}|$. Since $\rho \le 64t$,

$$d_B \ge d - \rho > 700t, \qquad p_B \ge p - \rho > 699t.$$

In particular, both hidden pivots are nonzero. Furthermore,

$$r_a := \frac{M_a^B}{d_B} \le \frac{d + t + \rho}{d - \rho} = 1 + \frac{t + 2\rho}{d - \rho} < 1 + \frac{129}{700} < \frac{6}{5},$$
$$M_b^B \le p + t + \rho < \left(1 + \frac{65}{763}\right) p < \frac{11}{10} p,$$
$$p_B \le p + \rho < \left(1 + \frac{64}{763}\right) p < \frac{11}{10} p.$$

The exact rank-two identity at the nonzero hidden pivot B_{ab} gives

$$B_{ic} = \frac{B_{ib}B_{ac} - B_{ia}B_{bc}}{B_{ab}}, \qquad B_{ij} = \frac{B_{ia}B_{jb} - B_{ib}B_{ja}}{B_{ab}}.$$

Hence

$$M_c^B \le r_a(M_b^B + p_B) \le \frac{66}{25}p, \qquad \max_{ij}|B_{ij}| \le 2r_a M_b^B \le \frac{66}{25}p. \tag{BC}$$

It follows that $M_b \le p + t < 3p$ and

$$\max_i |S_{ic}| \le M_c^B + \rho < \left(\frac{66}{25} + \frac{64}{763}\right)p < 3p.$$

Write $w = S_{bc}$ and $w_B = B_{bc}$. The exact hidden-cross identity says $B_{ij}w_B = B_{ib}B_{jc} - B_{ic}B_{jb}$. The numerator perturbation between this expression and the numerator in (CROSS) is bounded by $2(M_b^B + M_c^B)\rho + 2\rho^2$. Thus

$$\begin{aligned}|D_{ij} - B_{ij}| &\le \frac{2(M_b^B + M_c^B)\rho + 2\rho^2}{p} + |B_{ij}|\frac{|w_B - w|}{p} \\ &\le \left[2\left(\frac{11}{10} + \frac{66}{25}\right) + \frac{66}{25} + \frac{128}{763}\right]\rho < 11\rho.\end{aligned}$$

The remaining assertions about D follow from (CROSS), using exact skewness and diagonal zeroes of S. Swapping b, c leaves D invariant. The column bounds and residual estimate are consequently unaffected by the later orientation convention. □

Proof of the theorem. Condition first on the good events of the seed and scan calls. If the pure output is returned, its exploitability in S obeys

$$f_S(e_a) = \max_i S_{ia} \le S_{ja} + t \le \widehat{S}_{ja} + 2t \le H + 2t = 3\varepsilon/4.$$

For each individual regret, replacing a matrix by one within entrywise distance ρ changes that regret by at most 2ρ. Since S is skew, the constant shift cancels, and the second part of (SYM) holds, this gives

$$\max\{R_A(e_a, e_a), C_A(e_a, e_a)\} \le 3\varepsilon/4 + 2\rho \le 7\varepsilon/8. \tag{PURE}$$

Otherwise, the residual lemma applies. On the additional good event of the fresh pivot estimate, the guard passes, the oriented true pivot is $p = S_{bc} > 0$, $|\widetilde{p} - p| \le h_0$, and $p/\widetilde{p} \in [1/2, 2]$. Define the true, unknown observed-cross features

$$z_i = (S_{ib}/p, S_{ic}/p).$$

They lie in $[-3, 3]^2$, contain the anchors $z_b = (0, 1)$ and $z_c = (-1, 0)$, and represent the exact surrogate as $D_{ij} = p\det(z_i, z_j)$. Importantly, every support score actually sampled has mean $p[(1-u)(z_i)_1 - u(z_i)_2]$. Thus no ρ-bias is hidden in its oracle.

Given all preceding transcript and the good pivot event, a support call is good with conditional probability at least $1 - \beta$. Its true normalized score is within $\eta/p \le \tau/2$ of its maximum. Each estimated normalized coordinate obeys

$$\left|\frac{\widehat{a}_i}{\widetilde{p}} - \frac{a_i}{p}\right| \le \frac{\eta + 3h_0}{\widetilde{p}} = \frac{35\varepsilon}{2^{20}\widetilde{p}} < \tau.$$

Here $|a_i| \le 3p$ was used. Therefore the same planar primitive applies to D, with its original geometric $64K\tau$ error constant. The guard implies $\tau \le 1/4096 < 1/(64K)$ and $2K/\tau \le 98304/\varepsilon$, so it makes at most M_* calls.

For completeness, the support-three compression does not increase the primitive's stated error constant. Its only branch with more than three labels is the near-origin branch, whose proved error before compression is $50K\tau$. Insert exact estimates of the two anchors and use the coordinate estimates stored for its at most two other labels. A planar Carathéodory elimination preserves the convex combination of these estimated features and retains at most three labeled occurrences. If a label appears twice with different estimates, keep its occurrences separate during elimination and merge its final weights afterwards. Since each estimated feature is within τ in ℓ_∞ of its true feature, the old and new true means differ by at most 2τ in each coordinate. Every determinant with an opponent feature changes by at most $4K\tau$. Its final error is therefore at most $54K\tau < 64K\tau$. All other branches already output a pure label.

Consequently the returned q satisfies

$$f_D(q) \le 192p\tau = \frac{192}{8192}\frac{p}{\widetilde{p}}\varepsilon \le \frac{3}{64}\varepsilon.$$

Combining the residual lemma with $\|A-(B+c\mathbf{1}\mathbf{1}^\top)\|_\infty \le \rho$, and using that both individual regrets of (q,q) in skew B equal $f_B(q)$, gives

$$\max\{R_A(q,q), C_A(q,q)\} \le f_B(q) + 2\rho \le f_D(q) + 11\rho + 2\rho \le \frac{3}{64}\varepsilon + 13\rho \le \frac{55}{64}\varepsilon < \frac{7}{8}\varepsilon.$$

This proves (RPT) and the required margin in both branches.

The simultaneous good event has probability at least $1 - 4\delta/16 - \delta/4 - M_*\beta = 1-\delta$. Conditional freshness makes this union valid although all support directions and selected labels are adaptive. Each seed call costs $O(n\varepsilon^{-2}\log(c/\delta))$. The scalar pivot cost is $O(\varepsilon^{-2}\log(c/\delta))$ with an absolute constant. Each support call costs $O(n\varepsilon^{-2}\log(c/\beta))$. Summing at most M_* calls gives the displayed deterministic cap. All public guard failures and hard-cap exhaustions return a fixed simplex point on arbitrary matrices and arbitrary transcripts. The final compression is finite-dimensional arithmetic on at most four labels. Thus the algorithm is total and polynomial in the same arithmetic model as the baseline primitive. □

Corollary P.21 (Full-box-safe acceleration near a skew rank-two game)**.** *Replace the skew-rank-two branch of the existing safe-racing policy by the algorithm above at target $\xi = \varepsilon/2$ and failure $1/8$. The policy remains globally correct on all square matrices. On every A satisfying*

$$\operatorname{dist}_\infty\big(A, \{B + c\mathbf{1}\mathbf{1}^\top : B = -B^\top, \operatorname{rank}(B) \le 2,\ c \in \mathbb{R}\}\big) \le \varepsilon/32,$$

its expected physical query cost is

$$Cn\varepsilon^{-2}\left[\log(2/\varepsilon)\log\big(c\log(2/\varepsilon)\big) + \log(c/\delta)\right].$$

The existing unrestricted guarantee holds simultaneously. The policy does not need to know that the distance condition holds.

Proof. The proposal's individual-regret margin is $7\xi/8 = 7\varepsilon/16$, exactly the completeness margin of the original fresh verifier. Apply the established safe-racing lemma. The closed distance condition has an attained minimizer: the diagonal of A bounds c along a minimizing sequence, after which the rank-at-most-two skew parts can be restricted to a compact entrywise ball. □

P.7 Compressed robustness and its safe interface

Shared sampling preserves the deterministic error bounds of Appendix P.5. The first two corollaries give the standalone guarantees, first for skew observations and then for arbitrary observations near a shifted skew game. The final corollary records their interface with the safe scheduler in Appendix Q.

Corollary P.22 (Compressed certification under unknown skew model error)**.** *Suppose the observed $A \in [-1,1]^{n\times n}$ is skew-symmetric and $\|A-B\|_\infty \le \rho \le \varepsilon/512$ for some unknown skew-symmetric B of rank at most two. There is an algorithm with the cap equation 24, not given B or ρ, returning a distribution on at most three labels with $\max_i e_i^\top Aq \le \varepsilon$ with probability at least $1-\delta$. The stronger structural and early-output bounds in Theorem P.18 remain valid.*

Proof. Use the retuned seed, pivot guard, and normalized parameters of Theorem P.18:

$$K = 9/4, \quad \varepsilon/3 \le \widetilde{p} \le 2, \quad \tau = \frac{\varepsilon}{128\widetilde{p}}.$$

Its robustness proof requires only that every physical support label be $\varepsilon/4096$-optimal and that its two physical coordinate estimates have error at most $\varepsilon/4096$. Replace its sequential calls by Lemma P.10 with $\eta_c = \varepsilon/32768$. On its single good event the physical selection loss is $4\eta_c = \varepsilon/8192 \le \varepsilon/4096$ and all returned coordinate errors are at most $\eta_c \le \varepsilon/4096$.

The deterministic comparisons with the latent matrix B, including its stable-cross bound, relative pivot inequalities, score perturbation $\eta+2\rho$, and coordinate perturbation $\eta+\rho+K(h_0+\rho)$, therefore continue to hold with at least their previous slack. No fresh-sampling assumption is needed inside the

deterministic planar primitive once its oracle conditions hold simultaneously. Its extended range is supplied by Lemma P.17. The same four $\delta/4$ sampling allocations prove correctness, and the new compressed sample count proves the cap. The structural output still obeys

$$f_A(q) \le (8019/8192)\varepsilon + \rho \le (8035/8192)\varepsilon,$$

while the early pure output obeys $f_A(q) \le 33\varepsilon/64$. □

Corollary P.23 (Compressed certification with nonskew error and an unknown shift)**.** *Let $A \in [-1,1]^{n\times n}$ and suppose that, for some unknown scalar c and skew-symmetric rank-at-most-two matrix B,*

$$\|A - (B + c\mathbf{1}\mathbf{1}^\top)\|_\infty \le \rho \le \varepsilon/512.$$

There is a deterministically capped algorithm, not given B, c, or ρ, which uses the query order in equation 24 and returns q on at most three actions such that both individual regrets of (q,q) in the actually queried matrix A are at most ε, with probability at least $1-\delta$. Thus neither exact skew symmetry nor exact rank two is required of the observed matrix.

Proof. Put $S = (A - A^\top)/2$ and $H = (A + A^\top)/2$. The structural assumption gives $\|S - B\|_\infty \le \rho$ and $\|H - c\mathbf{1}\mathbf{1}^\top\|_\infty \le \rho$. An exact $\mathcal{N}(S_{ij}, 1)$ observation is obtained from two independent physical samples $Y \sim \mathcal{N}(A_{ij}, 1)$, $Y' \sim \mathcal{N}(A_{ji}, 1)$ and fresh private $Z \sim \mathcal{N}(0, 1/2)$ by returning $(Y - Y')/2 + Z$. This costs two physical queries and preserves the adaptive Gaussian oracle law.

Run Corollary P.22 on S. A constant shift cancels in both regrets, and entrywise perturbation by at most ρ changes each individual regret by at most 2ρ. In the structural branch, therefore,

$$\max\{R_A(q,q), C_A(q,q)\} \le \frac{8019}{8192}\varepsilon + 3\rho \le \frac{8067}{8192}\varepsilon < \varepsilon.$$

The early pure branch is bounded by $33\varepsilon/64 + 2\rho < \varepsilon$. The simulator changes only the universal constant in the deterministic cap. All support and failure-probability claims are inherited. □

Corollary P.24 (Safe acceleration near a shifted skew rank-two game)**.** *For square games, the generic safe race can use a capped branch which is globally safe after fresh verification and, without knowing a structural witness or approximation radius, satisfies*

$$\mathbb{E}_A \tau \le Cn\varepsilon^{-2}\left[\log\big(2 + \min\{n, \varepsilon^{-1}\}\big) + \log(1/\delta)\right] \tag{30}$$

whenever

$$\inf_{B=-B^\top,\ \mathrm{rank}(B)\le 2,\ c\in\mathbb{R}} \|A - (B + c\mathbf{1}\mathbf{1}^\top)\|_\infty \le \varepsilon/2048.$$

The unrestricted generic guarantee is retained simultaneously. The final pair on this branch is (q,q), with q supported on at most three labels.

Proof. The displayed infimum is attained: in a bounded minimizing sequence the diagonal determines a bounded constant shift, the skew parts are bounded, and rank at most two and skew symmetry are closed conditions. Let $S = (A - A^\top)/2$ and $H = (A + A^\top)/2$. If $\|A - (B + c\mathbf{1}\mathbf{1}^\top)\|_\infty \le \rho$, then

$$\|S - B\|_\infty \le \rho, \qquad \|H - c\mathbf{1}\mathbf{1}^\top\|_\infty \le \rho.$$

The constant shift cancels in both regrets. Consequently, for every q,

$$R_A(q,q),\ C_A(q,q) \le f_S(q) + 2\rho.$$

An exact unit-Gaussian query to S_{ij} is simulated by two independent physical observations $Y \sim \mathcal{N}(A_{ij}, 1)$ and $Y' \sim \mathcal{N}(A_{ji}, 1)$ and fresh private $Z \sim \mathcal{N}(0, 1/2)$, returning $(Y - Y')/2 + Z$. Even when $i = j$, use independent physical observations. This preserves the adaptive Gaussian oracle law and costs two physical queries per virtual observation.

Run Corollary P.22 on this oracle at target $\xi = \varepsilon/4$ and failure probability $1/8$. The radius condition gives $\rho \le \xi/512$, so with probability at least $7/8$,

$$R_A(q,q),\ C_A(q,q) \le \varepsilon/4 + 2\varepsilon/2048 = 257\varepsilon/1024 < 7\varepsilon/16.$$

This is the good-proposal margin of the existing safe race. The branch is deterministically capped even when the structural condition fails. Then it still returns only a proposal, and fresh full-game

verification controls every accepted output. Apply Lemma Q.2 with a constant scheduling share for this branch. Its constant-confidence proposal cost is the first logarithm in equation 30. The outer verifier contributes the additive confidence term. The generic branch gives almost-sure stopping and the unrestricted bound for every matrix. Neither the simulator nor the proposal uses B, c, or ρ. □

The remaining accuracy logarithm in equation 24 is not proved necessary. The current compression proof pays $\log D$ to preserve all D grid support values. Removing that charge requires a further argument. This section does not claim the exact skew-rank-two minimax law at constant confidence.

P.8 Binary feedback implementations

Corollary P.25 (Rank-two certification from binary rollouts)**.** *Under the Rademacher observation model of Section 2, Theorems P.11, P.12 and P.19 and Corollary P.23 retain their deterministic query orders, structural assumptions, support bounds, and individual-regret guarantees, up to universal sampling constants. Their globally valid, safely adaptive expected bounds also hold, using bounded-feedback self-play directly for the unrestricted branch.*

Proof. Every physical observation $W \in \{-1, 1\}$ has its specified mean and centered subgaussian proxy at most one. Frozen strategy mixtures and the random orientation/sign symmetrization preserve this property and cost one physical query. A planar support score is therefore an unbiased bounded observation with the required mean.

The rectangular proof uses only three further interfaces. A frozen linear combination of two cells is sampled with two independent observations and has proxy proportional to the squared coefficient norm. Every such norm is bounded by the same public inverse or ratio screen used in the Gaussian proof. A support score in a unit direction is sampled by choosing one of its two coordinates proportionally to absolute weight and rescaling the signed observation. Finally, all PAC labels are frozen before their coordinates are estimated afresh. Hoeffding concentration supplies exactly the scalar and PAC contracts used by the proofs, with no larger asymptotic cost.

No routine samples a possibly unbounded reconstructed entry as a binary mean: planar surrogates are accessed through observed cross columns, and rectangular residual or inverse scores are real-valued combinations of actual observations. Conditional freshness, deterministic guards, and query caps are unchanged. Applying the same bounded-feedback verifier and racing lemma gives the safely adaptive conclusions. No Gaussian factory is used by these structured branches. □

For Bernoulli win observations X, the convention is $W = 2X - 1$ and $A = 2P - \mathbf{1}\mathbf{1}^\top$, where $P_{ij} = \mathbb{E}X_{ij}$. The structural hypothesis concerns this centered payoff A. Low rank of P alone does not imply the same rank for A. Individual regrets in A are twice those measured in win-probability units. A nonlinear link applied to a low-rank score matrix does not automatically satisfy the payoff-rank assumption either.

Q Globally Valid Structure-Adaptive Certification

Structural assumptions can determine the speed without determining the validity of the answer. We prove this using independently verified proposals and a schedule measured in physical cell observations. Computation is not charged, as in the rest of the paper. No rank test is needed.

Q.1 The verifier's completeness margin

Lemma Q.1 (Individual-regret completeness)**.** *The verifier of Lemma D.1, with its original thresholds, accepts with conditional probability at least* $1 - \alpha$ *whenever*

$$\max\{R_A(x, y), C_A(x, y)\} \le 7\varepsilon/16.$$

Its soundness, simultaneous reported bounds, and deterministic physical-query cap are unchanged. The condition need only hold for the candidate fixed before the verifier's fresh observations are drawn.

Proof. On the simultaneous event in the proof of Lemma D.1, the two scalar estimation errors contribute at most $2a = \varepsilon/16$ to each estimated regret. The selected row response cannot exceed the true maximum, and the selected column response cannot fall below the true minimum. Each estimated regret is consequently at most $7\varepsilon/16 + \varepsilon/16 = \varepsilon/2$, so both acceptance tests pass. The rest is exactly the same event and the same algorithm as in that lemma. □

Q.2 A physical-query scheduling lemma

Fix a finite collection of proposal routines indexed by $b = 0, \dots, d$. Each routine is almost surely total: it returns a simplex pair after almost surely finite local computation and finitely many queries. Between consecutive physical queries, every local computation block terminates almost surely. The structured routines have deterministic query caps. No universal CPU-time bound for arbitrary proposal programs is assumed. A deterministic cap or guard failure returns a fixed simplex pair. A malformed output is also replaced by that fixed pair. Returning a pair from a proposal, including from its internal verification step, never returns an answer from the outer algorithm.

Branch b runs repeated fresh proposals, each followed by the verifier of Lemma D.1. In its kth cycle use

$$w_b = 2^{-(b+1)}, \qquad \alpha_{b,k} = \frac{6\delta w_b}{\pi^2 k^2}. \tag{31}$$

Branches use independent randomization and disjoint fresh observation streams. A branch only reads its own transcript. In particular, a mixture draw, a PAC observation, and every input requested by the Gaussian factory belong to the branch that requests that physical cell observation.

At physical query slot $t = 1, 2, \dots$, let $v_2(t)$ be the largest integer $a \geq 0$ for which 2^a divides t. Serve branch $v_2(t)$ if $v_2(t) \leq d$, and branch 0 otherwise. A service performs one physical observation, resuming any paused subroutine, followed by its finite local computation up to the next requested observation or acceptance. In particular, a Gaussian-factory invocation is paused between its physical inputs. Its entire possibly random-length invocation is never one scheduling slot. Initialize a branch lazily on its first service. Stop immediately when any outer verifier accepts, and return that candidate and its reported bounds. The nontrivial cases considered below require a positive finite number of verifier observations in every cycle, so there is no infinite sequence of zero-query restarts between two slots.

Lemma Q.2 (Safe racing with expected physical cost)**.** *The racing policy returns an incorrect answer or invalid reported bound with probability at most δ, for every matrix in the payoff box, regardless of the proposal routines' structural assumptions.*

For a fixed A, suppose branch b has the following guarantees conditional on every history at the beginning of one of its cycles: its expected proposal cost is at most $P_b(A) < \infty$, and with probability at least $p_b > 0$ its proposal satisfies $\max\{R_A, C_A\} \leq 7\varepsilon/16$. Then the racing policy stops almost surely and its total number τ of physical queries satisfies

$$\mathbb{E}_A \tau \leq C_{p_b} 2^{b+1} \left[P_b(A) + \frac{n+m}{\varepsilon^2} \left(1 + \log \frac{1}{\delta w_b} \right) \right]. \tag{32}$$

If several branches have these guarantees, the minimum of their displayed bounds is valid. No deterministic cap on the total τ is asserted.

Proof. It is useful to pre-generate independent infinite observation arrays for each branch and each cell, together with independent private randomness. Define each branch's isolated infinite execution, stopping it at its first acceptance. The interleaved execution is a prefix of the corresponding isolated execution in every branch. All unused observations retain the specified cell law, even when the queried labels depend on past data.

For the kth reached verifier of branch b, condition on its complete proposal transcript before any verifier observations are revealed. Its simultaneous good event fails with probability at most $\alpha_{b,k}$. It follows, without conditioning on which branch ultimately wins, that

$$\mathbb{P}_A\{\text{some reached verifier has a failed good event}\} \leq \sum_{b=0}^{d} \sum_{k \geq 1} \alpha_{b,k} = \delta \sum_{b=0}^{d} w_b \leq \delta.$$

On the complement, every possible accepted output and reported confidence bound are valid. Structural proposal failures need no additional share of the global error budget: they only change which candidate is verified.

Now consider an isolated branch satisfying the stated proposal assumptions. Write $H_{b,k}$ for the event that cycle k is reached. Conditional on its entry history, a good candidate is generated with probability at least p_b. Conditional on any such candidate transcript, fresh verification accepts with probability at least $1 - \alpha_{b,k} \geq 3/4$ by Lemma Q.1. Therefore

$$\mathbb{P}_A(H_{b,k}) \leq q_b^{k-1}, \qquad q_b = 1 - 3p_b/4 < 1.$$

This is a conditional argument at cycle entry, not an assumption that restart events or cycle lengths are independent.

Let $S_{b,k}$ be the proposal's physical cost in cycle k when reached. Its conditional expected cost is at most $P_b(A)$, even though the cost may be correlated with proposal quality or acceptance. The deterministic verifier cap is at most

$$V_{b,k} = \frac{C(n+m)}{\varepsilon^2}\left(1 + \log\frac{1}{\delta w_b} + 2\log k\right).$$

For the isolated branch's local stopping count σ_b, Tonelli's theorem and conditioning at cycle entry give

$$\mathbb{E}_A\sigma_b \leq \sum_{k\geq 1} q_b^{k-1}\{P_b(A) + V_{b,k}\} \leq C_{p_b}\left[P_b(A) + \frac{n+m}{\varepsilon^2}\left(1 + \log\frac{1}{\delta w_b}\right)\right].$$

Both series $\sum q_b^{k-1}$ and $\sum q_b^{k-1}\log k$ are finite. Every invocation is almost surely finite, so $\sigma_b < \infty$ almost surely.

For $b \geq 1$, its jth scheduled observation occurs at $2^b(2j-1) \leq 2^{b+1}j$. Branch 0 receives all odd slots, so its jth observation occurs no later than $2j - 1 \leq 2j$. Consequently, pathwise,

$$\tau \leq 2^{b+1}\sigma_b.$$

If another branch accepts earlier, the inequality only improves. Taking expectations proves equation 32, including almost-sure stopping. The comparison is on coupled full executions, so it does not erroneously condition a verifier on the outcome of the race. □

Q.3 A larger unknown-radius rank-one neighborhood

The following strengthened perturbation calculation also improves the standalone robustness theorem. It uses only the existing algorithm.

Lemma Q.3 (Stable rank-one proposals in an entrywise tube)**.** *Let $0 < \xi \leq 1/8$ and $0 < \beta \leq 1/4$. Run algorithm L.1 at target accuracy ξ and failure probability β on $A \in [-1,1]^{n\times m}$. If some rank-at-most-one matrix B satisfies $\|A - B\|_\infty \leq \rho \leq \xi/16$, then, with probability at least $1-\beta$, the returned pair satisfies*

$$\max\{R_A, C_A\} \leq \frac{51}{64}\xi. \tag{33}$$

The algorithm does not use B or ρ, and has its original deterministic cap $C(n+m)\xi^{-2}\log(c/\beta)$ on every matrix in the payoff box.

Proof. Set $t = \xi/64$, so $\rho \leq 4t$, and condition on all PAC and scalar good events of the original algorithm. Their simultaneous probability is at least $1-\beta$, independently of the rank of A. If the initial pure pair is returned, its individual regrets are at most $3\xi/4 + 3t = 51\xi/64$.

Otherwise the seed and the two actual scans in A give, successively,

$$|A_{a_0b_0}| > 21t, \qquad |A_{ab_0}| > 20t, \qquad |A_{ab}| > 19t.$$

Thus $|B_{ab_0}| > 16t$ and $|B_{ab}| > 15t$. In particular B has rank exactly one and these two entries are nonzero. The first absolute scan implies

$$\max_i |B_{ib_0}| \leq |B_{ab_0}| + t + 2\rho, \qquad \frac{\max_i |B_{ib_0}|}{|B_{ab_0}|} \leq 1 + \frac{9}{16} = \frac{25}{16}.$$

Appendix table of contents

For a rank-one matrix this last ratio is independent of the chosen nonzero column. The second scan likewise gives

$$\frac{\max_i |B_{ib}|}{|B_{ab}|} \le \frac{25}{16}, \qquad \frac{\max_j |B_{aj}|}{|B_{ab}|} \le 1 + \frac{9}{15} = \frac{8}{5}.$$

Each t-near extremal endpoint selected in A is a $(t + 2\rho)$-near extremal endpoint in B. The rank-one endpoint transfer argument therefore gives, for the returned strategies supported on the selected S, T,

$$\mathrm{Gap}_B(x, y) \le \mathrm{Gap}_{B|S\times T}(x, y) + \left(\frac{8}{5} + \frac{25}{16}\right)(t + 2\rho).$$

Here and on the restricted game, the exact identity $\mathrm{Gap}_A(x, y) = \max_i e_i^\top Ay - \min_j x^\top Ae_j$ gives the sharper perturbation bound

$$|\,\mathrm{Gap}_A(x, y) - \mathrm{Gap}_B(x, y)| \le 2\rho.$$

The estimated restricted equilibrium has restricted A-gap at most $2t$. Applying the perturbation bound once to the restricted game and once to the full game yields

$$\begin{aligned}\mathrm{Gap}_A(x, y) &\le 2t + 4\rho + \frac{253}{80}(t + 2\rho)\\ &\le \left(18 + 9\frac{253}{80}\right) t = \frac{3717}{80} t < 51t = \frac{51}{64}\xi.\end{aligned}$$

The individual regrets are nonnegative and sum to this gap. This proves equation 33 on both algorithmic branches. Every query cap and selection-freshness condition is inherited from the same finite algorithm. No inversion of noisy estimates is introduced. □

Q.4 A single policy with no rank or radius input

Write $D = n + m$ and

$$G(n, m, \varepsilon, \delta) = \varepsilon^{-2}\left[n \log(en) + m \log(em) + D \log \frac{1}{\delta}\right].$$

For $r \ge 3$, let $P_r(n, m, \varepsilon)$ be a public deterministic cap for algorithm M.1 run with rank bound r, target accuracy $\varepsilon/4$, and internal failure probability $1/8$. In particular,

$$P_r(n, m, \varepsilon) \le C_r D \varepsilon^{-(4r+6)}. \tag{34}$$

The more informative bound equation 18, evaluated at $(\varepsilon/4, 1/8)$, can be used in place of this simplification.

Theorem Q.4 (Global validity with automatic structural acceleration)**.** *There is a single policy, given only $n, m, \varepsilon, \delta$, such that for every $A \in [-1, 1]^{n\times m}$, $0 < \varepsilon \le 1/8$, and $0 < \delta \le 1/4$, it stops almost surely and is (ε, δ)-PAC-Nash on the entire payoff box. With probability at least $1 - \delta$, it also returns the simultaneous deviation bounds of Lemma D.1. Its expected physical cost satisfies*

$$\mathbb{E}_A\tau \le CG(n, m, \varepsilon, \delta) \qquad \textit{for every } A. \tag{35}$$

If $\mathrm{dist}_\infty(A, \{B : \mathrm{rank}(B) \le 1\}) \le \varepsilon/32$, the same policy satisfies

$$\mathbb{E}_A\tau \le CD\varepsilon^{-2} \log \frac{c}{\delta}. \tag{36}$$

For $3 \le r \le \min\{n, m\}$ with $\mathrm{rank}(A) \le r$, it additionally satisfies

$$\mathbb{E}_A\tau \le C2^{r+3}\left[P_r(n, m, \varepsilon) + D\varepsilon^{-2}\left(\log \frac{c}{\delta} + r\right)\right] \le C_r D\left[\varepsilon^{-(4r+6)} + \varepsilon^{-2} \log \frac{c}{\delta}\right]. \tag{37}$$

For every matrix of rank at most two, it additionally satisfies

$$\mathbb{E}_A\tau \le C(n + m)\varepsilon^{-2}[L_\varepsilon^3 \log(cL_\varepsilon) + \log(c/\delta)], \qquad L_\varepsilon = \log(2/\varepsilon). \tag{38}$$

When $n = m$ and A lies within $\varepsilon/32$ of a matrix $B + c\mathbf{1}\mathbf{1}^\top$, with $B = -B^\top$ and rank at most two, the same policy additionally satisfies

$$\mathbb{E}_A\tau \le Cn\varepsilon^{-2}\left[L_\varepsilon \log(cL_\varepsilon) + \log(c/\delta)\right]. \tag{39}$$

On the narrower $\varepsilon/2048$ neighborhood of this class it further satisfies

$$\mathbb{E}_A\tau \le Cn\varepsilon^{-2}\left[\log(2 + \min\{n, \varepsilon^{-1}\}) + \log(c/\delta)\right]. \tag{40}$$

The final simplified C_r bound also holds for every fixed $r > \min\{n, m\}$. All applicable bounds hold simultaneously, so their minimum is valid. The 1×1 game is returned with zero queries. The policy is not given the true rank or an approximation-radius bound. It need not estimate either.

Proof. For nontrivial dimensions put $d = \min\{n, m\}$ and use the schedule above with indices $0, \dots, d+2$. Branch 0 uses $\mathsf{GProp}(\varepsilon)$ from Lemma E.4 as its proposal. Branch 1 uses algorithm L.1 at target $\varepsilon/2$ and failure $1/8$. Branch 2 uses the wide-tube planar algorithm of Theorem P.19 at target $\varepsilon/2$ and failure $1/8$ when $n = m$, returning (q, q), and returns a fixed pair otherwise. Branch 3 uses the nonskew compressed algorithm of Corollary P.23 at target $\varepsilon/4$ and failure $1/8$ when $n = m$, converts its output to (q, q), and returns a fixed pair otherwise. Each virtual Gaussian draw is charged as two physical observations and may be paused between those observations. If $d \geq 2$, branch 4 uses Theorem P.12 at target $\varepsilon/4$ and failure $1/8$. For $3 \leq r \leq d$, branch $r+2$ uses the capped fixed-rank proposal defining P_r. Every cycle starts a new proposal, and every outer verifier uses fresh physical samples with the risk in equation 31.

Branch 0 has good-proposal probability at least $15/16$ on the full box, expected cost at most $C\varepsilon^{-2}[n\log(en) + m\log(em)]$, and almost-sure termination by Lemma E.4. Thus Lemma Q.2 gives global stopping, correctness, and equation 35.

On the stated rank-one tube, Lemma Q.3 with $\xi = \varepsilon/2$ gives good-proposal probability at least $7/8$, because $51\varepsilon/128 < 7\varepsilon/16$. Its cap is $CD\varepsilon^{-2}$. Applying the same scheduling lemma to branch 1 yields equation 36.

When $\operatorname{rank}(A) \leq r \leq d$ and $r \geq 3$, the fixed-rank theorem at target $\varepsilon/4$ makes both proposal regrets at most $\varepsilon/4$ with probability at least $7/8$, at cost at most P_r. The branch remains capped even when $\operatorname{rank}(A) > r$: the public call schedule, inversion and feature screens, duplicate-label and pivot guards, row-swap caps, and final $s = r$ rejection branch return a fixed pair. That pair is only a proposal and is never exempt from outer verification. Therefore Lemma Q.2 applies to branch $r+2$ and gives the first inequality in equation 37. The second follows from equation 34, absorbing fixed-r constants. If $r > d$, use branch $d+2$ (or branch 1 when $d = 1$) and enlarge the displayed r-dependent constant. For rank at most two and $d \geq 2$, branch 4 returns regrets at most $\varepsilon/4 < 7\varepsilon/16$ with probability at least $7/8$ and has cap $C(n+m)\varepsilon^{-2}L_\varepsilon^3\log(cL_\varepsilon)$. The scheduling share 2^{-5} is constant, so Lemma Q.2 gives equation 38. If $d = 1$, branch 1 already gives that bound. Its proposal term is also at most $C(n+m)\varepsilon^{-14}$ on the stated accuracy range, preserving the simplified fixed-r bound for $r = 2$. The bound involving P_r is used only for $r \geq 3$. On the wider stated near-skew tube, branch 2 has success probability at least $7/8$: Theorem P.19, at target $\varepsilon/2$, allows radius $(\varepsilon/2)/16 = \varepsilon/32$ and gives both proposal regrets at most $7\varepsilon/16$. Its constant-confidence cap is $Cn\varepsilon^{-2}L_\varepsilon\log(cL_\varepsilon)$. The safe-racing lemma gives equation 39.

On the narrower stated near-skew tube, branch 3 has success probability at least $7/8$ and a constant-confidence cap $Cn\varepsilon^{-2}\log(2 + \min\{n, \varepsilon^{-1}\})$. Indeed the radius is at most $(\varepsilon/4)/512$, and Corollary P.23 bounds both proposal regrets by $\varepsilon/4 < 7\varepsilon/16$ on the good event. The safe-racing lemma gives equation 40. Off their respective tubes, both planar branches' public guards and caps still return only proposals, and outer verification supplies safety. All proposal guarantees remain valid conditional on each cycle-entry history, because each invocation is restarted with fresh data and private randomization. □

The confidence cost in equation 37 is additive. The potentially expensive structured proposal is run only at constant confidence. Only the inexpensive verifier pays $\log(1/\delta)$. This does not improve the unresolved rank or accuracy dependence of the proposal, nor imply a deterministic total-cost bound or efficient net enumeration.

R Response compression with a fixed smaller action set

When one player has few actions, a finite grid of that player's mixtures can preserve all relevant responses of the other player. At fixed accuracy, the grid size depends only on the smaller action set, giving linear query dependence on the larger one.

Let $r = \min\{n, m\}$ and define

$$N_r(\varepsilon) = \binom{\lceil 16r/\varepsilon \rceil + r - 1}{r - 1}.$$

Theorem R.1 (Fixed-width response compression)**.** *For every $n, m \geq 1$, $0 < \varepsilon \leq 1/4$, and $0 < \delta \leq 1/4$, there is a matrix-independent, globally (ε, δ)-PAC-Nash policy with deterministic*

query cap

$$C\varepsilon^{-2}N_r(\varepsilon)(n+m)\log\left(\frac{cN_r(\varepsilon)r}{\delta}\right).$$

The argument uses only the PAC extremum interface in Lemma C.5, simplex Lipschitzness, Gaussian concentration, and finite-game minimax duality.

Proof of Theorem R.1. We first assume that $m = r \le n$. The opposite orientation will be reduced to this case at the end. Set

$$h = a = b = \frac{\varepsilon}{16}, \qquad L = \left\lceil \frac{r}{h} \right\rceil = \left\lceil \frac{16r}{\varepsilon} \right\rceil,$$

and define the denominator-L grid

$$\mathcal{G}_L = \left\{ \frac{z}{L} : z \in \mathbb{Z}^r_{\ge 0},\ \sum_{j=1}^{r} z_j = L \right\} \subseteq \Delta_r.$$

Stars and bars gives

$$N := |\mathcal{G}_L| = \binom{L+r-1}{r-1} = N_r(\varepsilon).$$

The grid covers the small simplex. Fix $y \in \Delta_r$. Write $Ly_j = q_j + f_j$, where $q_j = \lfloor Ly_j \rfloor$ and $0 \le f_j < 1$. The number

$$s := \sum_{j=1}^{r} f_j = L - \sum_{j=1}^{r} q_j$$

is an integer in $\{0, \dots, r-1\}$. If $s > 0$, select the s coordinates with largest fractional parts and add one to the corresponding q_j. If $s = 0$, make no change. Let z be the resulting integer vector. Then $\sum_j z_j = L$, so $g = z/L$ belongs to $\mathcal{G}_L$.

Let H be the selected coordinate set. Since the selected fractional parts are the s largest,

$$\sum_{j \in H} f_j \ge s\left(\frac{1}{r}\sum_{j=1}^{r} f_j\right) = \frac{s^2}{r}.$$

Consequently,

$$\begin{aligned} L\|g - y\|_1 &= \sum_{j \in H}(1 - f_j) + \sum_{j \notin H} f_j \\ &= 2s - 2\sum_{j \in H} f_j \le 2s\left(1 - \frac{s}{r}\right) \le \frac{r}{2} \le r. \end{aligned}$$

Thus $\|g - y\|_1 \le r/L \le h$. Every point of Δ_r is therefore within ℓ_1-distance h of a grid point.

PAC row calls on the grid. For $i \in [n]$ and $y \in \Delta_r$, define

$$\mu_i(y) = e_i^\top A y, \qquad F(y) = \max_{i \in [n]} \mu_i(y), \qquad V_A^\star = \min_{y \in \Delta_r} F(y).$$

At a fixed grid point g, one pull of row arm i draws a fresh $J \sim g$ using private randomization and queries (i, J). The observation $Y = A_{iJ} + Z$, where $Z \sim \mathcal{N}(0,1)$ is also fresh, has mean $\mu_i(g)$. Different arm pulls and different grid calls use disjoint observation streams. The centered random variable $A_{iJ} - \mu_i(g)$ lies in an interval of length at most two and hence has subgaussian proxy at most one by Hoeffding's lemma. Adding the independent Gaussian term gives a universal subgaussian proxy.

At every $g \in \mathcal{G}_L$, apply Lemma C.5 with accuracy a and failure probability $\delta/(2N)$. Let i_g be the returned row and put

$$S = \{i_g : g \in \mathcal{G}_L\}.$$

Then $|S| \le N$. Each call has a deterministic cap

$$C_0 n a^{-2} \log\left(\frac{c_0 N}{\delta}\right).$$

A union bound shows that, with probability at least $1 - \delta/2$, the event

$$\mathcal{E}_1 = \{\mu_{i_g}(g) \ge F(g) - a \text{ for every } g \in \mathcal{G}_L\}$$

holds.

Let $A_S = A_{S,[r]}$ be the game restricted to the selected rows. For $y \in \Delta_r$, define

$$F_S(y) = \max_{i \in S} e_i^\top A y, \qquad V^\star_{A_S} = \min_{y \in \Delta_r} F_S(y).$$

Because all entries of A lie in $[-1, 1]$, both envelopes are one-Lipschitz in ℓ_1:

$$|F(y) - F(y')| \le \|y - y'\|_1, \qquad |F_S(y) - F_S(y')| \le \|y - y'\|_1.$$

On $\mathcal{E}_1$, $F_S(g) \ge F(g) - a$ at every grid point. For arbitrary y, choose $g \in \mathcal{G}_L$ with $\|g - y\|_1 \le h$. Then

$$F(y) \le F(g) + h \le F_S(g) + a + h \le F_S(y) + a + 2h. \tag{41}$$

Thus the selected rows approximate the full best-response envelope uniformly, not only on the grid.

Since $S \subseteq [n]$, we have $F_S \le F$ and therefore $V^\star_{A_S} \le V^\star_A$. Let y_S minimize F_S, and choose $g_S \in \mathcal{G}_L$ with $\|g_S - y_S\|_1 \le h$. On $\mathcal{E}_1$,

$$V^\star_{A_S} = F_S(y_S) \ge F_S(g_S) - h \ge F(g_S) - a - h \ge V^\star_A - a - h.$$

Hence

$$V^\star_A - a - h \le V^\star_{A_S} \le V^\star_A. \tag{42}$$

Fresh estimation of the restricted game. Conditional on the complete first-stage transcript and the resulting set S, take fresh Gaussian observations from every cell in $S \times [r]$. Use

$$M_0 = \left\lceil 2b^{-2} \log\left(\frac{4Nr}{\delta}\right)\right\rceil$$

observations per cell and let $\bar{A}_{ij}$ be the corresponding sample mean. Write $\bar{A}_S = (\bar{A}_{ij})_{i \in S,\, j \in [r]}$ for the estimated restricted matrix. For each estimated cell,

$$\mathbb{P}\big(|\bar{A}_{ij} - A_{ij}| > b \mid S\big) \le 2 \exp\left(-\frac{M_0 b^2}{2}\right) \le \frac{\delta}{2Nr}.$$

Since $|S| r \le Nr$, a conditional union bound gives an event

$$\mathcal{E}_2 = \left\{\max_{i \in S,\, j \in [r]} |\bar{A}_{ij} - A_{ij}| \le b\right\}$$

with conditional probability at least $1 - \delta/2$. This stage uses at most

$$C_1 N r b^{-2} \log\left(\frac{c_1 N r}{\delta}\right)$$

observations.

Solve the estimated restricted zero-sum game $\bar{A}_S$ exactly. Let $(\widehat{x}_S, \widehat{y}) \in \Delta_S \times \Delta_r$ be a saddle point and let $V^\star_{\bar{A}_S}$ be its value. Embed $\widehat{x}_S$ into Δ_n by putting zero mass outside S, and call the embedded strategy $\widehat{x}$.

On $\mathcal{E}_2$, every mixed payoff in the true and estimated restricted games differs by at most b, so $|V^\star_{\bar{A}_S} - V^\star_{A_S}| \le b$. The saddle-point identities for the estimated game give

$$F_S(\widehat{y}) \le \max_{i \in S} e_i^\top \bar{A}_S \widehat{y} + b = V^\star_{\bar{A}_S} + b \le V^\star_{A_S} + 2b, \tag{43}$$

$$\min_{j \in [r]} \widehat{x}^\top A e_j \ge \min_{j \in [r]} \widehat{x}_S^\top \bar{A}_S e_j - b = V^\star_{\bar{A}_S} - b \ge V^\star_{A_S} - 2b. \tag{44}$$

Appendix table of contents

Define the row strategy's security level $G(x) = \min_{j\in[r]} x^\top A e_j$. Combining Equations (41), (43) and (44) directly gives

$$F(\widehat{y}) - G(\widehat{x}) \le a + 2h + 4b = \frac{7\varepsilon}{16}.$$

Since

$$R_A(\widehat{x},\widehat{y}) + C_A(\widehat{x},\widehat{y}) = F(\widehat{y}) - G(\widehat{x})$$

and both regrets are nonnegative, each is at most $7\varepsilon/16 < \varepsilon$. Let $\mathcal{H}_1$ be the full first-stage transcript. Freshness of the second stage gives

$$\begin{aligned}\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) &\ge 1 - \mathbb{P}(\mathcal{E}_1^c) - \mathbb{E}[\mathbb{P}(\mathcal{E}_2^c \mid \mathcal{H}_1)] \\ &\ge 1 - \delta.\end{aligned}$$

Query accounting. The total number of physical Gaussian observations is at most

$$\begin{aligned}T &\le C_0 N n a^{-2} \log\left(\frac{c_0 N}{\delta}\right) + C_1 N r b^{-2} \log\left(\frac{c_1 N r}{\delta}\right) \\ &\le C\varepsilon^{-2} N(n+r) \log\left(\frac{cNr}{\delta}\right).\end{aligned}$$

Since $m = r$, this is the stated cap. Every PAC call and every cell-estimation stage has a deterministic cap, so the total stopping time is bounded by the displayed deterministic quantity. Deduplicating S can only reduce the cost.

The opposite orientation. If $n < m$, define $B = -A^\top \in [-1,1]^{m\times n}$. The column side of B has size $n = r$, so the preceding construction applies. A query to B_{ji} is implemented by querying A_{ij} and negating the observation. If (x_B, y_B) is the output for B, return $x_A = y_B$ and $y_A = x_B$. Direct calculation gives

$$R_A(x_A, y_A) = C_B(x_B, y_B), \qquad C_A(x_A, y_A) = R_B(x_B, y_B).$$

Thus the regrets are preserved up to exchange and the physical-query count is unchanged. When $n = m = 1$, return the unique pair without sampling.

The grid and LP computations are not charged in the cell-query model. The grid can be exponentially large when r grows, so this construction is not a balanced-regime algorithmic closure. □

Corollary R.2 (Fixed-small-side and one-sided consequences)**.** *For every fixed $r \ge 2$ and fixed $\varepsilon \in (0, 1/8]$, the dependence of $\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta)$ on $n+m$ and $\log(1/\delta)$ is minimax-optimal uniformly over dimensions with $\min\{n,m\} = r$ and $0 < \delta \le 1/4$. If exactly one of n, m equals one and $\max\{n,m\} \ge 2$, then, uniformly over $0 < \varepsilon \le 1/8$ and $0 < \delta \le 1/4$,*

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta) = \Theta\left((n+m-1)\varepsilon^{-2}\log\frac{1}{\delta}\right).$$

The 1×1 game has zero query complexity.

Proof. Let $q = \max\{n,m\}$ and $r = \min\{n,m\}$. The disjoint-star theorem gives, for every nontrivial dimension pair,

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta) \ge \frac{D}{18\kappa\varepsilon^2}\,\mathrm{kl}\left(\frac{\delta}{D}, 1-\delta\right),$$

where $D = n + m - 2$ and $\kappa = \mathbf{1}\{n > 1\} + \mathbf{1}\{m > 1\}$. For $0 < \delta \le 1/4$, put $p = \delta/D$. Since $p \le 1/4$,

$$\begin{aligned}\mathrm{kl}(p, 1-\delta) &= -H(p) - p\log(1-\delta) - (1-p)\log\delta \\ &\ge (1-p)\log(1/\delta) - \log 2 \ge \frac{1}{4}\log(1/\delta).\end{aligned}$$

The last inequality uses $\log(1/\delta) \ge \log 4$.

If $r \ge 2$, then $\kappa = 2$ and $D = q + r - 2 \ge (q+r)/2$. For each fixed (r,ε), $N_r(\varepsilon)$ is constant and the logarithm in Theorem R.1 is at most a constant multiple, depending on (r,ε), of $\log(1/\delta)$. This proves the first claim.

If $r = 1$ and $q \ge 2$, then $N_1(\varepsilon) = 1$, $D = q - 1 \ge q/2$, and $\kappa = 1$. The upper and lower bounds therefore match uniformly in $\varepsilon \le 1/8$, giving the displayed one-sided rate. The 1×1 case is immediate. □

S Individual-Regret and Duality-Gap Characteristics Can Differ Linearly

At the same accuracy, controlling each player's regret and controlling their sum define different answer sets. The inclusions between these sets do not ensure comparable fixed-candidate information costs. Using the definitions in Appendix O, the following family gives a linear separation between the two characteristics.

S.1 Construction

Fix $m \geq 6$ and $0 < \varepsilon \leq 1/4$. Define

$$A^{(m)} = -\mathbf{1}_{3\times m} + \varepsilon C^{(m)},$$

where

$$C_{\bullet,1} = \begin{pmatrix} 1.8 \\ 1.2 \\ 0.9 \end{pmatrix}, \qquad C_{\bullet,2} = \begin{pmatrix} 1.2 \\ 1.8 \\ 0.1 \end{pmatrix}, \qquad C_{\bullet,j} = \begin{pmatrix} 1.6 \\ 1.6 \\ 1.6 \end{pmatrix} \quad (j \geq 3).$$

The constant -1 shift does not affect regrets.

Lemma S.1 (Unique exact equilibrium)**.** *The unique exact equilibrium is*

$$x^\star = \left(\frac{1}{2}, \frac{1}{2}, 0\right), \qquad y^\star = \left(\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0\right),$$

with value $-1 + 1.5\varepsilon$.

Proof. The displayed pair is an equilibrium: against $x^\star$, columns 1 and 2 both give $-1 + 1.5\varepsilon$, whereas every distractor gives $-1 + 1.6\varepsilon$. Against $y^\star$, rows 1 and 2 both give $-1 + 1.5\varepsilon$, whereas row 3 gives $-1 + 0.5\varepsilon$.

For uniqueness of x, remove the common -1 and divide by ε. Any maximin strategy must make both columns 1 and 2 at least 1.5:

$$1.8x_1 + 1.2x_2 + 0.9x_3 \geq 1.5,$$
$$1.2x_1 + 1.8x_2 + 0.1x_3 \geq 1.5.$$

Adding and using $x_1 + x_2 + x_3 = 1$ gives $3(1 - x_3) + x_3 \geq 3$, so $x_3 = 0$. The two inequalities then force $x_1 \geq 1/2$ and $x_1 \leq 1/2$.

For uniqueness of y, let $d := \sum_{j\geq 3} y_j$. Requiring rows 1 and 2 to have payoff at most 1.5 and adding the inequalities gives $3(y_1 + y_2) + 3.2d \leq 3$. Since $y_1 + y_2 = 1 - d$, this becomes $3 + 0.2d \leq 3$, so $d = 0$. The two row inequalities then force $y_1 = y_2 = 1/2$. □

S.2 Individual-regret certificate

Lemma S.2 (Dimension-free individual-regret upper characteristic)**.** *For the individual-regret answer set,*

$$T^\star_{\mathrm{ind}}(A^{(m)}, \varepsilon) \leq \frac{1200}{\varepsilon^2}.$$

Proof. Take the pure pair $z_0 = (e_3, e_1)$. Its row regret is $(1.8 - 0.9)\varepsilon = 0.9\varepsilon$ and its column regret is $(0.9 - 0.1)\varepsilon = 0.8\varepsilon$, so it is an individual-regret ε-equilibrium.

Allocate weight $1/3$ to each of entries $(1, 1), (2, 1), (3, 1)$ and zero elsewhere.

The tight row alternative is row 1. Its current deviation is 0.9ε, so the slack is 0.1ε. To increase $B_{11} - B_{31}$ by 0.1ε, the minimum weighted Gaussian cost is attained by moving the two entries by $\pm 0.05\varepsilon$:

$$\frac{1}{2} \cdot \frac{1}{3}\left[(0.05\varepsilon)^2 + (0.05\varepsilon)^2\right] = \frac{\varepsilon^2}{1200}.$$

The row-2 alternative is farther away.

For a column violation, an unobserved entry B_{3j} may be moved for free but cannot go below -1. Since the sampled entry is $A_{31} = -1 + 0.9\varepsilon$, even setting an unobserved comparison entry to -1 leaves only 0.9ε column regret. Therefore every column-violation alternative must increase B_{31} by at least 0.1ε. Its cost is at least

$$\frac{1}{2} \cdot \frac{1}{3}(0.1\varepsilon)^2 = \frac{\varepsilon^2}{600}.$$

Thus the minimum information rate under this design is at least $\varepsilon^2/1200$, proving the claim. □

S.3 Duality-gap lower characteristic

Lemma S.3 (Every duality-gap answer has many dangerous distractors)**.** *For every $z = (x, y) \in \mathcal{C}_\varepsilon^{\mathrm{DG}}(A^{(m)})$, at least $m - 5$ distractor columns $j \geq 3$ satisfy the following: if column j is replaced by the all--1 column, then z is no longer a duality-gap ε-answer.*

Proof. For every y, $\Phi_A(y) \geq V_A^\star = -1 + 1.5\varepsilon$. Fix a distractor j and let $B^{(j)}$ be obtained by replacing column j by $-\mathbf{1}$. Since this new column is available to the minimizer, $\Psi_{B^{(j)}}(x) \leq -1$. All three entries of column j are lowered by 1.6ε, so every row payoff, and hence their maximum, changes by exactly $-1.6\varepsilon y_j$:

$$\Phi_{B^{(j)}}(y) = \Phi_A(y) - 1.6\varepsilon y_j.$$

Thus the new duality gap is at least

$$\Phi_{B^{(j)}}(y) - \Psi_{B^{(j)}}(x) \geq (1.5 - 1.6y_j)\varepsilon.$$

This is strictly larger than ε whenever $y_j < 5/16$. A probability vector has at most three coordinates of size at least $5/16$. Among the $m - 2$ distractors, at least $m - 5$ therefore satisfy the strict inequality. □

Theorem S.4 (Linear characteristic separation)**.** *For the family above,*

$$T_{\mathrm{DG}}^\star(A^{(m)}, \varepsilon) \geq \frac{25(m-5)}{32\varepsilon^2},$$

and hence

$$\frac{T_{\mathrm{DG}}^\star(A^{(m)}, \varepsilon)}{T_{\mathrm{ind}}^\star(A^{(m)}, \varepsilon)} \geq \frac{m-5}{1536}.$$

In particular, the ratio is unbounded with dimension.

Proof. Fix any duality-gap answer z and any allocation w. Let $W_j := \sum_{i=1}^3 w_{ij}$ be the total weight on column j. By Lemma S.3, at least $m - 5$ distractors define valid alternatives. Their total weight is at most one, so one such j has

$$W_j \leq \frac{1}{m-5}.$$

Replacing that column by $-\mathbf{1}$ moves each of its three entries by 1.6ε. Therefore

$$D_{\mathrm{DG}}(A; z, w) \leq \frac{1}{2}(1.6\varepsilon)^2 W_j \leq \frac{32}{25}\frac{\varepsilon^2}{m-5}.$$

This holds for every z, w, so

$$D_{\mathrm{DG}}^\star(A^{(m)}, \varepsilon) \leq \frac{32\varepsilon^2}{25(m-5)}.$$

Taking reciprocals proves the first bound. Combining it with Lemma S.2 yields

$$\frac{T_{\mathrm{DG}}^\star}{T_{\mathrm{ind}}^\star} \geq \frac{25(m-5)}{32\varepsilon^2} \cdot \frac{\varepsilon^2}{1200} = \frac{m-5}{1536}.$$

□

Remark S.5 (Why the separation matters). The individual-regret extension is statistically substantive. The larger individual-regret answer set contains certificates that can ignore linearly many columns, whereas every duality-gap answer must protect against almost all distractors.

T Relation to the COLT 2025 Open Problem and the Remaining Balanced Regime

This appendix places the full-box results in the broader program initiated by the COLT 2025 open problem of Maiti (2025). It also records several precise statements that clarify the unresolved balanced regime. The first separates support discovery from certification. The second shows why pseudo-regret is insufficient. The final discussion explains why a tempting universal expected-gap shortcut is ruled out by the dense lower bound.

T.1 Full-box minimax and instance-dependent quantifiers

The COLT question asks for the optimal *instance-dependent* sample complexity of finding an approximate Nash equilibrium from noisy payoff queries. The full-box quantity studied in this paper has a different order of quantifiers:

$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta) = \inf_{\substack{\pi:\ \forall B\in[-1,1]^{n\times m},\\ \mathbb{P}^{\pi}_{B}\{\widehat{z}\in\mathcal{C}^{\mathrm{ind}}_{\varepsilon}(B)\}\ge 1-\delta}} \sup_{A\in[-1,1]^{n\times m}} \mathbb{E}^{\pi}_{A}\tau. \tag{45}$$

Thus a single policy, selected from $(n, m, \varepsilon, \delta)$ but not from the unknown matrix, must be correct on every matrix in the box before its worst-case cost is evaluated.

A convenient pointwise benchmark instead fixes a reference matrix in the objective:

$$H_{\mathrm{pt}}(A,\varepsilon,\delta) = \inf_{\substack{\pi:\ \forall B\in[-1,1]^{n\times m},\\ \mathbb{P}^{\pi}_{B}\{\widehat{z}\in\mathcal{C}^{\mathrm{ind}}_{\varepsilon}(B)\}\ge 1-\delta}} \mathbb{E}^{\pi}_{A}\tau. \tag{46}$$

Even here the selected policy remains globally correct, but the infimum may favor the reference matrix A. A complete instance-dependent theory must go further by identifying the local alternative geometry and, ideally, by giving one adaptive unknown-instance policy that attains the resulting benchmark across matrices. Consequently, the present results establish sharp worst-case benchmarks that such a theory must recover, but they do not constitute a complete solution of the instance-dependent COLT problem.

The regimes closed in the present full-box analysis are the following.

1. For arbitrary n, m with $(n, m) \neq (1, 1)$ and $0 < \varepsilon \le 1/8$, if $\log(1/\delta) \gtrsim \max\{\log(en), \log(em)\}$, then
$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta) = \Theta\left((n+m)\varepsilon^{-2}\log\frac{1}{\delta}\right).$$
This is the confidence-dominated full-box minimax regime.
2. For every fixed integer $r \ge 2$ and fixed $\varepsilon \in (0, 1/8]$, uniformly over dimensions satisfying $\min\{n, m\} = r$ and over $0 < \delta \le 1/4$,
$$\mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta) = \Theta_{r,\varepsilon}\left((n+m)\varepsilon^{-2}\log\frac{1}{\delta}\right).$$
The one-sided case $r = 1$ has the corresponding exact best-arm rate, apart from the zero-query 1×1 game.
3. If $n = m = k$ and $\varepsilon \lesssim \log(ek)/k$, the full-box bounds match at
$$\Theta\left(k\varepsilon^{-2}\left[\log(ek) + \log\frac{1}{\delta}\right]\right).$$
This is a fine-accuracy balanced regime rather than the fixed-accuracy balanced regime discussed below.

T.2 What fixed-accuracy discovery now requires

Fix $\varepsilon_0 \in (0, 1/16]$ and $\delta_0 \in (0, 1/4]$. Corollary J.2 gives

$$ck\varepsilon_0^{-2}\log\log(ek) \;\le\; \mathcal{M}^{\mathrm{G}}_{k,k}(\varepsilon_0,\delta_0) \;\le\; Ck\varepsilon_0^{-2}\log(ek). \tag{47}$$

Appendix table of contents

Discovery is therefore superlinear even at fixed accuracy and confidence. Together with the fixed-rank upper and lower bounds, this proves a strict separation between unrestricted and every fixed-rank class. It does not identify the exact unrestricted rate: a factor of order $\log k/\log\log k$ remains between the two displayed bounds.

The two dense witnesses serve different regimes. The additive witness in Theorem I.1 uses $O(s\log k)$ columns but requires $s = O(1/\varepsilon)$. The intersection witness in Theorem J.1 uses $O(2^s \log k)$ columns and permits $s = \Theta(\log k)$ independently of fixed accuracy. Its probability mass is 16ε, rather than a sum of s separate masses. The same adjacent-layer information argument then supplies $\log s = \Theta(\log\log k)$. Neither witness supplies $s = k^{\Omega(1)}$ at fixed accuracy inside a balanced table. The fixed-small-side algorithm also does not settle the remaining gap, since its simplex grid is exponential when that side grows with k.

The exact-payoff model is a separate question.

Proposition T.1 (One-way exact-payoff-query implication)**.** *Let $Q^{\mathrm{ex}}_{n,m}(\varepsilon,\delta)$ denote the minimax expected number of* distinct *exact payoff entries queried under the same full-box correctness and almost-sure stopping requirements. Then*

$$Q^{\mathrm{ex}}_{n,m}(\varepsilon,\delta) \le \mathcal{M}^{\mathrm{G}}_{n,m}(\varepsilon,\delta).$$

More specifically, a Gaussian policy with worst-case expected cost B yields an exact-query policy with at most $\lceil B/h\rceil$ distinct queries and success probability at least $1-\delta-h$, for every $0<h<1-\delta$.

Proof. Given exact-query access to A, run the Gaussian policy. On the first request to cell (i,j), query its mean exactly and cache it. Return $A_{ij}+Z$ with fresh private $Z\sim\mathcal{N}(0,1)$ on each simulated request. The complete transcript, output, and stopping law are identical to those of the Gaussian policy. The number of distinct exact queries is pathwise at most the Gaussian observation count. Infimizing proves the expected-cost statement. Truncation after $\lceil B/h\rceil$ simulated observations loses success probability at most h by Markov's inequality, proving the second statement. □

An exact-query lower bound can thus rule out a corresponding Gaussian upper bound, once accuracy, success probability, and budget form are matched. The converse implication for lower bounds is unavailable. In particular, the new Gaussian $\Omega(k\log\log k)$ lower bound does not prove an exact-query lower bound of that order. Its hard dictionary may be public, in which case one exact observation from each row pair identifies all membership bits. The statistical difficulty is acquiring those bits reliably from noisy observations. Nor does an arbitrary exact-query upper bound provide a Gaussian algorithm: infinitesimal exact comparisons may require arbitrarily many noisy observations to resolve.

T.3 Boundary results and barriers

T.3.1 Known equilibrium-support labels remove the logarithm

For an integer $s\ge 1$, define

$$N^{\sup}_s(\varepsilon) = \binom{\lceil 48s/\varepsilon\rceil + s - 1}{s-1}.$$

Theorem T.2 (Known-support Gaussian closure)**.** *Let $A\in[-1,1]^{n\times m}$. Suppose that sets $R\subseteq[n]$ and $C\subseteq[m]$ are supplied before interaction, with $|R|=p$ and $|C|=q$, and that A has a minimax equilibrium $(x^\star,y^\star)$ satisfying*

$$\mathrm{Supp}(x^\star)\subseteq R, \qquad \mathrm{Supp}(y^\star)\subseteq C.$$

For $0<\varepsilon\le 1/8$ and $0<\delta\le 1/4$, there is a policy that receives R and C as side information, but does not depend on A, and returns $(\widehat{x},\widehat{y})$ with

$$R_A(\widehat{x},\widehat{y})\le\varepsilon, \qquad C_A(\widehat{x},\widehat{y})\le\varepsilon$$

with probability at least $1-\delta$. Its number of Gaussian cell observations is at most

$$C_0 N^{\sup}_p(\varepsilon)N^{\sup}_q(\varepsilon)(n+m)\varepsilon^{-2}\log\left(\frac{c_0 N^{\sup}_p(\varepsilon)N^{\sup}_q(\varepsilon)}{\delta}\right).$$

In particular, for fixed (p,q,ε) this is $O_{p,q,\varepsilon}(k\log(1/\delta))$ on the corresponding $k\times k$ class. No uniqueness, strict complementarity, or lower bound on equilibrium masses is required.

Proof. Put $h = \varepsilon/48$. For the simplex supported on any supplied s-element set, use the denominator-L_s grid with $L_s = \lceil s/h \rceil$. Rounding simplex coordinates while preserving their sum gives an ℓ_1 covering radius at most $s/L_s \le h$, and the grid has at most $N_s^{\sup}(\varepsilon)$ points. Enumerate the $K = N_p^{\sup}(\varepsilon) N_q^{\sup}(\varepsilon)$ pairs $z = (x, y)$ in the product of the two grids.

For a fixed candidate, a row-deviation sample draws $J \sim y$ and queries (i, J), while a column-deviation sample draws $I \sim x$ and queries (I, j). A candidate-payoff sample draws $(I, J) \sim x \otimes y$ and queries that cell. These observations have the required means and a universal subgaussian proxy. Capped PAC maximum and minimum calls, followed by scalar mean estimation on fresh streams, give estimates $\widehat{R}(z)$ and $\widehat{C}(z)$ such that

$$|\widehat{R}(z) - R_A(z)| \le \frac{3\varepsilon}{16}, \qquad |\widehat{C}(z) - C_A(z)| \le \frac{3\varepsilon}{16}.$$

For one candidate, the cost is at most $C(n+m)\varepsilon^{-2}\log(cK/\delta)$. PAC extremum selection has no $\log n$ or $\log m$ union penalty. Assigning failure probability $O(\delta/K)$ to each candidate makes these two bounds simultaneous over the product grid with probability at least $1 - \delta$.

Let (x', y') be grid points within h in ℓ_1 of $(x^\star, y^\star)$. Since $A \in [-1, 1]^{n \times m}$ and $(x^\star, y^\star)$ is an exact equilibrium,

$$R_A(x', y') \le \|x' - x^\star\|_1 + 2\|y' - y^\star\|_1 \le 3h,$$

and symmetrically $C_A(x', y') \le 2\|x' - x^\star\|_1 + \|y' - y^\star\|_1 \le 3h$. Thus this candidate has both true regrets at most $\varepsilon/16$, and on the simultaneous event its two estimates are at most $\varepsilon/4$.

Return any grid pair whose two estimated regrets are at most $\varepsilon/2$. Such a pair exists on the simultaneous event. Every returned pair then has both true regrets at most $\varepsilon/2 + 3\varepsilon/16 < \varepsilon$. If no pair passes, return an arbitrary fixed pair. This last branch can occur only outside the simultaneous event. Summing the per-candidate costs proves the stated bound. □

The theorem locates the conditional difficulty sharply. Once two fixed-size support-label sets are supplied, even a balanced game admits a linear fixed-accuracy certificate. The theorem makes no correctness claim for a matrix whose equilibria all use labels outside the supplied sets. Searching over unknown support labels would reintroduce dimension-dependent discovery, so this is not an unrestricted full-box closure.

T.3.2 Pseudo-regret alone does not certify the output

Proposition T.3 (Pseudo-regret and final-regret separation)**.** *For every $k \ge 2$ and $T \ge 1$, there is a legal fixed-matrix strategy process on $A = I_k$ for which both comparator-wise pseudo-regrets are zero, while the averaged output has duality gap $1 - 1/k$.*

Proof. Let $u = (1/k, \dots, 1/k)$ and draw $J \sim \mathrm{Unif}([k])$ before the first round. Set $x_t = u$ and $y_t = e_J$ for every t. Define

$$\mathrm{PR}_x = \max_{i \in [k]} \mathbb{E}\left[\sum_{t=1}^{T} e_i^\top A y_t\right] - \mathbb{E}\left[\sum_{t=1}^{T} x_t^\top A y_t\right],$$

$$\mathrm{PR}_y = \mathbb{E}\left[\sum_{t=1}^{T} x_t^\top A y_t\right] - \min_{j \in [k]} \mathbb{E}\left[\sum_{t=1}^{T} x_t^\top A e_j\right].$$

For every fixed i, $\mathbb{E}\sum_t e_i^\top A y_t = T/k$, and $x_t^\top A y_t = 1/k$ deterministically. Hence $\mathrm{PR}_x = 0$. Every fixed column against u also has payoff $1/k$, which gives $\mathrm{PR}_y = 0$.

The averages are $\bar{x}_T = u$ and $\bar{y}_T = e_J$. The best row against e_J earns 1, whereas $u^\top I_k e_J = 1/k$. Therefore

$$R_A(\bar{x}_T, \bar{y}_T) = 1 - \frac{1}{k}, \qquad C_A(\bar{x}_T, \bar{y}_T) = 0, \qquad \mathrm{Gap}_A(\bar{x}_T, \bar{y}_T) = 1 - \frac{1}{k}.$$

The matrix is fixed before play. The only randomization is the private label drawn before the first round. □

The proposition concerns the location of the extrema relative to the expectation. A guarantee with $\max_i \mathbb{E}[\cdot]$ controls a different object from $\mathbb{E}[\max_i \cdot]$. It therefore cannot by itself justify a proposal for the fresh verification wrapper. This is an obstruction to that proof interface, not a lower bound on unrestricted Gaussian minimax complexity.

T.3.3 WHY THE STRONG-GAP SHORTCUT CANNOT BE UNIVERSAL

One might hope for a bounded-feedback primitive with expected duality gap $O(\sqrt{k/T})$ uniformly over all k, T and all bounded games. Combined with the fresh verification wrapper, such a primitive would contradict even the fixed-accuracy lower bound in Corollary J.2. Corollary N.5 gives the earlier contradiction using the fine-accuracy bound. More generally, at fixed accuracy no uniformly constant-success Gaussian proposal can have worst-case expected cost $O(k)$: the fresh verifier would turn it into a globally correct $O(k)$ policy. The remaining question is which superlinear discovery rate is optimal.