ACCESSIBILITY LIMITS FOR BEST-ARM IDENTIFICATION UNDER COUPLED OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study fixed-confidence identification of the best of K candidate objectives that are optimized jointly. The quality of arm i is its standalone optimum $f_i^\star = \inf_z f_i(z)$, but the parameters of all arms are tied by an affine constraint $Ax = b$: each query returns a noisy value and gradient of one arm at its current parameters and triggers a prescribed projected gradient step that can move every arm, with no resets. Let $S_i = E_i^\top \operatorname{Proj}_{\ker A} E_i$, where E_i embeds the parameters of arm i. For smooth strongly convex arms, the complete noiseless value–gradient oracle on the constraint set determines every standalone optimum if and only if every S_i is positive definite; otherwise two instances agree on all feasible values and gradients but have different best arms. When all S_i are positive definite, repeated queries at a frozen state are allowed, and the prescribed step has a positive gain floor, we give Geometric Block Certificates, a δ-correct algorithm whose cost is $\widetilde{O}(\Delta_{\min}^{-2})$ on uniformly controlled families, and lower bounds showing that its access-dependent factor is tight on explicit constant-gain classes. For quadratic arms with known Hessians, a transformed observation reduces the problem to standard best-arm identification even when access is singular, given either the gradient-noise covariance or two independent queries at the same state. In experiments under the stochastic oracle, the algorithm never returned a wrong arm and its cost followed the predicted dependence on the gap and on access; on tied linear heads for digits, certified stopping was correct on all 450 runs.

1 INTRODUCTION

Choosing the best of several candidate models is often done before the chosen model is trained in earnest. When the candidates are independent, this is a best-arm identification (BAI) problem: each candidate is an arm, evaluating it yields a noisy sample, and the goal is to find the best arm with probability at least $1 - \delta$ from as few samples as possible (Kaufmann et al., 2016; Garivier & Kaufmann, 2016). In many pipelines, however, the candidates are trained *together* and share parameters. One-shot architecture search trains all subnetworks inside one supernet (Pham et al., 2018; Bender et al., 2018), multi-head models tie parameters across heads, and resource-allocation problems keep the candidates on a common budget. A gradient step taken to evaluate one candidate then changes the parameters of the others. After selection, the chosen candidate is trained on its own, so the quantity to compare is each candidate's *standalone* optimal value, not its value during the shared run. Shared weights are known to distort supernet rankings (Xu et al., 2022).

Parameter sharing creates two difficulties absent from standard BAI. The point at which a candidate can be evaluated depends on all past updates, including those made for other candidates, and cannot be reset. And the sharing constraint may keep a candidate away from its standalone minimizer forever, so that no amount of sampling reveals its standalone optimum. We ask when the best candidate can be identified under such coupling, and at what cost.

Problem. Arm $i \in [K]$ is an unknown differentiable objective f_i with standalone optimum $f_i^\star = \inf_z f_i(z)$; the target is the unique $i^\star = \arg\min_i f_i^\star$. The joint state $x = (x_1, \ldots, x_K)$ must stay in an affine set $\mathcal{F} = \{x : Ax = b\}$. In each round the learner picks an arm and a batch size, observes noisy values and gradients of that arm at its current block, and the application then applies a prescribed projected gradient update, which may move every block. Resets and off-trajectory queries

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.

are unavailable. We call this *coupled best-function identification*; its cost is the number of oracle calls.

Contributions. Let $P = \mathrm{Proj}_{\ker A}$, let E_i embed arm i's block into the joint state, and let $S_i = E_i^\top P E_i$. All our results are stated in terms of the matrices S_i.

- **Identifiability.** For smooth strongly convex arms with public curvature bounds, the complete noiseless value–gradient oracle on $\mathcal{F}$ determines every standalone optimum if and only if $S_i \succ 0$ for all i. If some S_i is singular, two instances agree in value and full gradient at every feasible state yet have different best arms (Theorem 1). The missing information is where an arm can be evaluated, not a component of its gradient.
- **Algorithm.** Suppose every $S_i \succ 0$, the application allows repeated queries at a frozen state, and its prescribed update is a radial response whose gain is at least a public $\theta_i^- > 0$. Then Geometric Block Certificates (Algorithm 1) are δ-correct, stop almost surely, and have finite expected cost; on uniformly controlled families with a fixed number of arms, the cost is $\widetilde{O}(\Delta_{\min}^{-2})$ with high probability and in expectation, where $\Delta_{\min}$ is the smallest gap (Theorem 2, Corollary 1). The bound contains an explicit access factor $\Gamma_q = \sum_i (1 + 1/(\mu_i \rho_i \theta_i^-))$, where μ_i is the strong-convexity constant of arm i and $\rho_i \le \lambda_{\min}(S_i)$.
- **Lower bounds.** The terms $1/(\mu_j \rho_j \theta_j)$ of Γ_q are tight on explicit classes with constant gains: there every δ-correct rule pays $\Omega(\sum_j (\mu_j \rho_j \theta_j)^{-1})$ oracle calls as $\mu_j \rho_j \theta_j \downarrow 0$, and a matching procedure exists (Corollary 2). On a single base instance, the cost of moving to informative states and the statistical Δ^{-2} cost add up (Theorem 3).
- **Known quadratics.** For quadratic arms with known Hessians and a known gradient-noise covariance, a transformed observation has conditional mean $f_i^\star$ at every state, so the problem reduces to standard BAI even when access is singular (Theorem 4); a cross-fitted variant that uses two independent queries at the same state needs no knowledge of the covariance (Theorem 6).
- **Experiments.** Under the stochastic oracle of the theory, with $\lambda_{\min}(S_i)$ as small as $1/64$, Algorithm 1 made no wrong stop in 7,000 runs; its cost scaled as $\Delta_{\min}^{-2}$, and most of its growth as $\lambda_{\min}(S_i)$ decreased is explained by Γ_q (Figure 1). With a full-batch oracle on tied linear heads for all 45 digit pairs, certified stopping was correct on 450/450 runs, versus 448/450 for a one-sweep heuristic, with 1.0% more row evaluations (Table 4).

2 Related Work

Best-arm and best-function identification. Fixed-confidence BAI finds the stationary distribution with the extremal mean (Kaufmann et al., 2016; Garivier & Kaufmann, 2016). Best function identification (Dorn et al., 2026) is the closest model: arms are functions and quality is their minimum. Its oracle uses arm-specific query points and its F-LCB algorithm runs a separate optimizer for each arm, so querying one arm never moves another; this is the uncoupled case $S_i = I$ of our protocol.

Learning curves and one-trajectory methods. Rested best-model identification, rising and improving bandits, and hyperparameter racing (Cella et al., 2021; Mussi et al., 2024; Takemori et al., 2024; Blum et al., 2026; Jamieson & Talwalkar, 2016; Li et al., 2018) allocate computation among learning curves, but each curve advances only when its own arm is pulled. Optimizer selection, hyper-heuristics, and warm-start switching (Meidani et al., 2022; Lagos & Pereira, 2024; Jankovic et al., 2022; Vermetten et al., 2023; Antipov & Doerr, 2026) follow one trajectory but for a single objective; in our notation all f_i coincide. In bandits with action-induced dynamics, a pull can advance neighboring reward curves or change future losses through an interaction matrix or a hidden state (Genalti et al., 2024; Sato & Ito, 2025; Mussi et al., 2023; Bacchiocchi et al., 2024), but the goal is cumulative reward; pure exploration in online MDPs identifies a policy rather than standalone minima (Al Marjani et al., 2021).

Structured pure exploration and weight sharing. Feedback-graph, restless Markov, and non-stationary linear pure exploration (Russo et al., 2025; Karthik et al., 2024; Xiong et al., 2024; Maynard-Zhang et al., 2026), linear partial monitoring (Kirschner et al., 2023), and linear, combinatorial, multi-fidelity, and constrained pure exploration (Degenne et al., 2020; Du et al., 2021; Poiani et al.,

Table 1: Summary of results. $\widetilde{O}_{\mathrm{ctl}}(\Delta^{-2})$ denotes $O_{\mathrm{ctl}}(\Delta^{-2}\{\log(K/\delta) + \log\log(e/\Delta)\})$ with a constant that depends only on the listed controls. Lower bounds fix the remaining construction and curvature parameters and take $\delta \le \delta_0 < 1/2$. Symbols are defined in Section 4.

Setting	Assumptions	Result	Type
Identifiability	Smooth μ-strongly convex, L-smooth arms, public $0 < \mu < L$; complete noiseless oracle on $\mathcal{F}$; any S_i	All optima identified iff all $S_i \succ 0$; otherwise two instances are indistinguishable, and every rule that stops almost surely on both errs with probability $\ge 1/2$ on one of them	Exact characterization
Positive access	Fixed K; frozen batches; prescribed batch-invariant radial response; uniform curvature, gain, noise, geometry, and anchor controls	$\widetilde{O}_{\mathrm{ctl}}(\Delta_{\min}^{-2})$ cost with high probability and in expectation	Upper bound
Constant gain	Prescribed time-constant θ_j; fixed witness ratios; $\mu_j\rho_j\theta_j \downarrow 0$; explicit K	$\bar{N}_j \asymp (\mu_j\rho_j\theta_j)^{-1}$ per challenger, with a matching upper bound	Tight on the class
Plateau	Constant-gain finite class; fixed K; $\varepsilon = \mu\rho\theta \downarrow 0$, $\Delta = \Theta(\rho\varepsilon^2)$	$\Theta((K-1)\{\varepsilon^{-1} + \sigma^2\Delta^{-2}\log(1/\delta)\})$ on one base instance	Tight on the class
Known quadratics	Common dimension; known Hessians; public tails; known gradient covariance or a frozen independent pair; singular access allowed	Reduction to standard BAI with stationary means	Upper bound

2022; 2024; Carlsson et al., 2024; Bian & Tan, 2025) couple information across arms or restrict the admissible measurements and answers, but not the optimization states. Weight-sharing architecture search (Pham et al., 2018; Bender et al., 2018; Xu et al., 2022) is the closest application; it develops search procedures and reduces interference, not fixed-confidence guarantees for standalone optima.

3 Problem Setup

Notation. We write $[K] = \{1, \dots, K\}$, $\|\cdot\|$ for the Euclidean norm, † for the Moore–Penrose pseudo-inverse, $\mathrm{Proj}_{\mathcal{C}}$ for the Euclidean projection onto a closed convex set $\mathcal{C}$, and $(u)_+ = \max\{u, 0\}$; $\bullet \in \{f, g\}$ marks value and gradient quantities. In lower bounds, $\Pr_0, \mathbb{E}_0$ refer to the base instance and kl is the binary relative entropy. Appendix A.1 lists all symbols.

Arms and coupling. Fix $K \ge 2$. Arm i has an unknown differentiable objective $f_i : \mathbb{R}^{d_i} \to \mathbb{R}$, a block of parameters $x_i \in \mathbb{R}^{d_i}$, and standalone optimum

$$f_i^\star = \inf_x f_i(x), \qquad i^\star \in \arg\min_{i \in [K]} f_i^\star.$$

We assume that $i^\star$ is unique. The target is defined for each arm trained on its own; it is not the minimum of a joint constrained objective. The joint state $x = (x_1, \dots, x_K)$ lies in a nonempty affine set

$$\mathcal{F} = \{x \in \mathbb{R}^{d_{\mathrm{tot}}} : Ax = b\}, \qquad d_{\mathrm{tot}} = \sum_i d_i.$$

Two examples with $d_i = d$ for all arms are *consensus*, $x_1 = \cdots = x_K$, and *fixed-sum balance*, $\sum_i x_i = b$ for some $b \in \mathbb{R}^d$ (zero-sum balance when $b = 0$). Write $E_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d_{\mathrm{tot}}}$ for the canonical embedding of block i.

Oracle and transitions. Rounds $t = 1, 2, \dots$ index state transitions. Before transition t, the policy chooses, based on the past, an arm i_t and a batch size $B_t \in \mathbb{N}$ allowed by the application. The state is held fixed while B_t independent value–gradient pairs are drawn; pair $r \in [B_t]$ is

$$F_{t,r} = f_{i_t}(x_{i_t,t}) + \zeta^f_{i_t,t,r}, \qquad G_{t,r} = \nabla f_{i_t}(x_{i_t,t}) + \zeta^g_{i_t,t,r}.$$

The learner observes the full gradient $G_{t,r}$, not only its projection onto the feasible directions. We call a single pair a *primitive query* and measure cost by the number of primitive queries, $C_t = \sum_{s \le t} B_s$. The case $B_t = 1$ is an ordinary query, and we then write F_t, G_t. Let $\mathcal{H}_{t-1}$ be the σ-field generated by the initial state, past choices, raw observations, states, and internal randomness; the choice (i_t, B_t) is $\mathcal{H}_{t-1}$-measurable, i.e., *predictable*. The noise conditions are stated with the results that use them.

Assumption 1 (Prescribed projected transition)**.** *The initial state lies in $\mathcal{F}$. For every history and selected arm, the application fixes in advance the set of allowed batch sizes and a transition (η_t, Φ_t), with $\eta_t \ge 0$, $\Phi_t : \mathbb{R}^{d_{i_t}} \to \mathbb{R}^{d_{i_t}}$, and $\Phi_t(0) = 0$. The transition does not depend on the batch size, and the policy cannot choose or tune it: it chooses only (i_t, B_t). After the batch, exactly one transition is applied, using the averaged gradient $\bar{G}_t = B_t^{-1} \sum_{r=1}^{B_t} G_{t,r}$: $\widetilde{x}_{t+1} = x_t - \eta_t E_{i_t} \Phi_t(\bar{G}_t)$ and $x_{t+1} = \mathrm{Proj}_{\mathcal{F}}(\widetilde{x}_{t+1})$. All raw pairs remain available to the learner.*

The trivial transition $x_{t+1} = x_t$ is allowed. Batches with $B_t > 1$ require the application to allow several queries before an update; if every primitive query triggers an update, only the results with $B_t = 1$ apply. All algorithms below assume that the batch sizes they request are allowed. Section 6 also uses a stronger *full-batch* oracle, which returns the exact value and gradient of a finite-sum objective; it is defined separately in Corollary 4.

Correctness and cost. A fixed-confidence rule stops at a time τ and recommends $\widehat{J}_\tau$. It is δ-*safe* if $\Pr(\tau < \infty, \widehat{J}_\tau \ne i^\star) \le \delta$, and δ-*correct* if in addition $\Pr(\tau < \infty) = 1$. Its cost is C_τ. With a fixed budget T of primitive queries, a rule recommends $\widehat{J}_T$ and incurs simple regret $R(T) = f^\star_{\widehat{J}_T} - f^\star_{i^\star}$. We write $N_i(t) = \sum_{s \le t} \mathbf{1}\{i_s = i\}$, $\tau_{i,n} = \inf\{t : N_i(t) = n\}$, and, when $N_i(t) \ge 1$, $\tau_i(t) := \tau_{i,N_i(t)} = \max\{s \le t : i_s = i\}$ for the last query of arm i up to round t.

Access geometry. Let $P = \mathrm{Proj}_{\ker A}$ and

$$S_i = E_i^\top P E_i \succeq 0.$$

For any $x^\circ \in \mathcal{F}$, the feasible values of block i form the affine set $x_i^\circ + \mathrm{range}(S_i)$ (Theorem 1). When arm i is queried with the identity response and step η, its block moves by $x_i^+ = x_i - \eta S_i \bar{G}_t$. Thus S_i describes both where arm i can be evaluated and how fast a projected update moves it. Consensus gives $S_i = I/K$, and fixed-sum balance gives $S_i = (1 - 1/K)I$ (Appendix A.5).

4 Main Results

4.1 When is the best arm identifiable?

Theorem 1 (Identifiability)**.** *Fix any $x^\circ \in \mathcal{F}$. The feasible locations of block i are*

$$E_i^\top \mathcal{F} = x_i^\circ + \mathrm{range}(S_i). \tag{1}$$

Consider the class of all smooth μ-strongly convex, L-smooth arms with fixed public bounds $0 < \mu < L$. The complete noiseless value–gradient oracle on $\mathcal{F}$ determines every standalone optimum $(f_i^\star)_{i=1}^K$ for all instances in this class if and only if $S_i \succ 0$ for every arm. If some S_i is singular, there are two instances with the same constraint, initial state, transition, and noise law, and with identical values and full gradients at every feasible state, but with different unique best arms. Consequently, every rule that terminates almost surely on both instances has worst-case error at least $1/2$.

Proof. See Appendix A.3.1.

The theorem concerns what the complete feasible oracle reveals, not whether a given transition reaches enough feasible points with finitely many samples (Section 4.2). Restricted classes can remain identifiable under singular access (Section 5).

4.2 Geometric Block Certificates

Certificates for the optimum value.

Assumption 2 (Smooth gradient-dominated arms)**.** *Arm i is L_i-smooth, attains its minimum at a point $x_i^\star$, and satisfies the μ_i-PL inequality*

$$f_i(x) - f_i^\star \le \frac{\|\nabla f_i(x)\|^2}{2\mu_i}.$$

Assumption 3 (Conditional first-order noise)**.** *Value noise is conditionally centered scalar sub-Gaussian and gradient noise is conditionally centered directional sub-Gaussian, with known proxies. Whenever a result uses a radius for the gradient norm, its norm tail (including the dependence on dimension) is assumed explicitly.*

If a frozen batch returns F, G with $|F - f_i(x)| \le r^f$ and $\|G - \nabla f_i(x)\| \le r^g$, then the PL inequality and the descent lemma give

$$\ell = F - r^f - \frac{(\|G\| + r^g)^2}{2\mu_i} \;\le\; f_i^\star \;\le\; u = F + r^f - \frac{(\|G\| - r^g)_+^2}{2L_i}. \tag{2}$$

The interval brackets the fixed number $f_i^\star$, so later updates cannot invalidate it, and intervals from different rounds can be intersected. Write $[\underline{V}_i, \overline{V}_i]$ for the current intersection of arm i's intervals within the family being tested (all intervals so far, or those of the current phase).

Proposition 1 (Certificate safety)**.** *If all stopping intervals are simultaneously valid with probability at least $1 - \delta$, recommending $\hat{\imath}$ only when $\overline{V}_{\hat{\imath}} < \min_{i \ne \hat{\imath}} \underline{V}_i$ is δ-safe for every predictable policy.*

Proof. This is Theorem 5 in Appendix A.2, applied to intersected certificates.

Safety does not imply termination: the query policy must make every relevant interval shrink. Under coupling this is the main difficulty, since a certificate is narrow only when the gradient of the queried arm is small, i.e., when its block is near its own standalone minimizer.

Assumption 4 (Access, response, and batch controls)**.** *Every f_i is μ_i-strongly convex and L_i-smooth; dimensions and minimizers may differ across arms. The access bounds*

$$0 < \rho_i \le \lambda_{\min}(S_i) \le \lambda_{\max}(S_i) \le R_i$$

are known. For every history at which arm i may be queried, the application fixes a measurable radial response $\Phi_t(v) = a_t(v)v$ and a step η_t, both independent of the batch size. For a public $\theta_i^- > 0$ we require, on every path and for every argument v,

$$\theta_i^- \le \eta_t a_t(v) \le \theta_i^+ := \frac{2}{\mu_i\rho_i + L_i R_i}, \qquad q_i := 1 - \mu_i\rho_i\theta_i^- \in [0, 1). \tag{3}$$

The response map is fixed before the batch, but its realized gain may depend on the averaged gradient of the batch. The identity response with the geometry-scaled step is the case $\theta_i^- = \theta_i^+$, for which $q_i = (L_iR_i - \mu_i\rho_i)/(L_iR_i + \mu_i\rho_i)$. For every requested frozen batch of B fresh pairs of arm i at round t, the averaged noises $\zeta_i^\bullet = B^{-1}\sum_{r=1}^B \zeta_{i,t,r}^\bullet$ satisfy

$$\begin{aligned} \Pr\left(|\zeta_i^f| > \frac{u}{\sqrt{B}} \mid \mathcal{H}_{t-1}\right) &\le 2e^{-u^2/(2(\sigma_i^f)^2)}, \\ \Pr\left(\|\zeta_i^g\| > \frac{u}{\sqrt{B}} \mid \mathcal{H}_{t-1}\right) &\le 2e^{-u^2/(2(\sigma_i^g)^2)}, \qquad u \ge 0. \end{aligned} \tag{4}$$

The norm proxy σ_i^g includes any dependence on dimension; a zero proxy forces zero noise.

Algorithm. Algorithm 1 runs in phases $h = 0, 1, \ldots$ with $B_h = 2^h$ and visits the arms in a fixed order. For arm i it first runs

$$m_{i,h} = \max\left\{1, \left\lceil \frac{\log B_h}{-\log q_i} \right\rceil\right\}$$

optimization batches of sizes $C_{i,h,j} = \lceil B_h q_i^{m_{i,h}-j} \rceil$, $j = 1, \ldots, m_{i,h}$ (one batch of size B_h if $q_i = 0$), each followed by its prescribed update. These batches move the joint state so that block i approaches its own minimizer; the remaining distance shrinks as $B_h^{-1/2}$ up to logarithmic factors

Input: confidence δ, public controls, fixed arm order.
for phases $h = 0, 1, \ldots$ **do** reset the phase-local intersections.
  for $i = 1, \ldots, K$ **do**
    for $j = 1, \ldots, m_{i,h}$ **do** average $C_{i,h,j}$ fresh pairs; apply the prescribed update with their mean gradient.
    Average a fresh batch of B_h pairs; add its interval to both families; apply the prescribed update.
  Test the families whose intersections are all nonempty, lifetime first. Return the smallest-index $\hat{\imath}$ with $\overline{V}_{\hat{\imath}} < \min_{i \neq \hat{\imath}} \underline{V}_i$, if one exists.
end for

Algorithm 1: Geometric Block Certificates

(Appendix Eq. (24)). It then draws a certificate batch of B_h fresh pairs, forms the interval $[\ell_{i,h}, u_{i,h}]$ of Eq. (2) with radii

$$r^{\bullet}_{i,h} = \sigma^{\bullet}_i \sqrt{\frac{2}{B_h} \log \frac{2R^{\mathrm{evt}}_h}{\delta_h}}, \qquad \delta_h = \frac{6\delta}{\pi^2 (h+1)^2}, \quad R^{\mathrm{evt}}_h = \sum_i m_{i,h} + 2K,$$

and applies that batch's prescribed update. Two families of intersections are kept: lifetime intersections, which persist, and phase-local intersections, which are reset at the start of each phase.

For the unique best arm $i^\star$, let $\Delta_i = f^\star_i - f^\star_{i^\star}$ for $i \neq i^\star$ and $\Delta_{\min} = \min_{i \neq i^\star} \Delta_i > 0$. Let $w_i(h)$ be the deterministic bound on the phase-h certificate width in Appendix Eq. (25), so that $u_{i,h} - \ell_{i,h} \leq w_i(h)$, and let

$$h_\delta = \min\{h : w_{i^\star}(h) + w_i(h) < \Delta_i \ \forall i \neq i^\star\}, \quad T_h = \sum_{r=0}^{h} \sum_i (m_{i,r} + 1).$$

Theorem 2 (Geometric Block Certificates)**.** *Under the protocol and Assumptions 2, 3, and 4, $h_\delta < \infty$, and Algorithm 1 is δ-correct, stops almost surely, has $\mathbb{E}C_\tau < \infty$, and satisfies*

$$\Pr(\tau \leq T_{h_\delta}, \widehat{J}_\tau = i^\star) \geq 1 - \delta.$$

On the same event,

$$C_\tau \leq (2^{h_\delta + 1} - 1)\Gamma_q + \sum_{r=0}^{h_\delta} \sum_i m_{i,r}, \qquad \Gamma_q = \sum_i \left(1 + \frac{1}{1 - q_i}\right) = \sum_i \left(1 + \frac{1}{\mu_i \rho_i \theta^-_i}\right). \quad (5)$$

Corollary 1 (Uniform gap rate)**.** *Fix K and a family with common positive lower bounds on $\mu_i, \rho_i, \mu_i\rho_i\theta^-_i$, and finite common upper bounds on L_i, R_i, the noise proxies, and the initial and transformed anchor radii. Uniformly over this family, for $\delta \in (0,1)$ and $0 < \Delta_{\min} \leq 1$,*

$$\max\{2^{h_\delta}, \mathbb{E}C_\tau\} = O\left(\Delta^{-2}_{\min} \left\{\log \frac{K}{\delta} + \log\log \frac{e}{\Delta_{\min}}\right\}\right),$$

where the hidden constant depends only on the displayed controls.

Proof. See Appendix A.3.2 for both results.

Discussion. Corollary 1 is uniform over families in which $\Delta_{\min} \downarrow 0$ with the other controls fixed; the transformed anchor radius of arm i is $\|S_i^{-1/2}(x^\star_i - x^\circ_i)\|$ for a feasible reference point x° (Appendix Eq. (20)). Access enters Theorem 2 through Γ_q; for the identity response with the exact step, arm i contributes $3/2 + L_i R_i/(2\mu_i\rho_i)$. The other factors (the factor $\sum_{q_i>0}(-\log q_i)^{-1}$ in the number of updates, and the factor $\rho_i^{-1/2}$ and the anchor radii in the bound on the state) are sufficient but not shown to be necessary; Appendix Eqs. (21)–(30) display every term. For the full-batch oracle, which evaluates all n_i rows of a finite empirical objective before each update, Corollary 4 gives a deterministic stopping phase $h_{\mathrm{pool}} = O_{\mathrm{ctl}}(\log(e + 1/\Delta_{\min}))$ and a total of $C_{\mathrm{pool}} = O_{\mathrm{ctl}}(\sum_i n_i\{1 + \mathbf{1}\{q_i > 0\}/(-\log q_i)\} \log^2(e + 1/\Delta_{\min}))$ row evaluations.

4.3 Lower bounds

The first bound shows that the terms $1/(\mu_j\rho_j\theta_j)$ of Γ_q are tight on a class with constant gains. Appendix Proposition 5 shows that the $\Delta^{-2}_{\min}$ dependence in Corollary 1 is also necessary.

Corollary 2 (Constant-gain reciprocal progress)**.** *For $j < K$, fix $0 < \mu_j < L_j$, $0 < \rho_j \le 1/2$, $\alpha_j, g_j > 0$, and an application-prescribed, time-constant gain $0 < \theta_j \le 2/(\mu_j\rho_j + L_j)$. For $\delta \in (0, 1/2)$, Appendix Proposition 8 constructs one noiseless base instance and $K-1$ alternatives such that, with*

$$q_j = 1 - \mu_j\rho_j\theta_j, \qquad \bar{N}_j = \left\lfloor \frac{\log(1 + \mu_j\alpha_j/g_j)}{-\log q_j} \right\rfloor,$$

every rule that is δ-correct on this class has $\mathbb{E}_0 C_\tau \ge (1-2\delta)\sum_{j<K} \bar{N}_j$, while a deterministic procedure for this class costs at most $\sum_{j<K}(\bar{N}_j + 2)$. For fixed positive $\alpha_j, g_j, \mu_j, L_j$,

$$\bar{N}_j \sim \frac{\log(1 + \mu_j\alpha_j/g_j)}{\mu_j\rho_j\theta_j} \quad \text{as } \mu_j\rho_j\theta_j \downarrow 0.$$

Hence the terms $1/(\mu_j\rho_j\theta_j)$ of Γ_q are tight on this class as $\mu_j\rho_j\theta_j \downarrow 0$ with the witness fixed, and the cost is linear in K when the challengers are homogeneous or of a common order.

Proof. See Appendix A.4.3.

Appendix Proposition 7 gives a bound $\Omega((K-1)L/(\mu\rho))$ for challengers with common constants μ, L, ρ that holds for every common adaptive radial response with nonnegative gains up to the safe ceiling. If θ_j is fixed and $\rho_j \downarrow 0$, a pair of instances with a common value and gradient in the low-access direction needs anchor radius $\Omega(\rho_j^{-1/2})$ at a fixed gap (Lemma 4); rescaling to bounded radius gives gap $\Theta(\rho_j)$.

Next, on one instance, the costs of reaching informative states and of sampling there add up.

Theorem 3 (Additive access and statistical cost)**.** *Fix $K \ge 2$, $0 < \mu < L$, $0 < \rho \le 1/2$, $\sigma > 0$, an integer traversal length $n \ge 1$, $\delta \in (0, 1/2)$, and an application-prescribed, time-constant gain $0 < \theta \le 2/(\mu\rho + L)$. There is an explicit $\bar{\Delta} > 0$, obtained by instantiating Appendix Proposition 9 with plateau scale $\lambda = 1$, such that, for every $0 < \Delta \le \bar{\Delta}$, there are one base and $K-1$ alternative smooth μ-strongly convex, L-smooth instances with the same affine geometry, initial state, identity-response transitions with challenger gain θ, allowed frozen batches of every positive size, and independent Gaussian value noise $\mathcal{N}(0, \sigma^2)$ per primitive query, with exact gradients. Every instance has gap Δ and least block-access eigenvalue ρ, and every rule that is δ-correct on this class satisfies on the base instance*

$$\mathbb{E}_0 C_\tau \ge (K-1)\left\{(1-2\delta)(n+1) + \frac{\sigma^2}{2\Delta^2}\,\mathrm{kl}(1-\delta, \delta)\right\}. \tag{6}$$

The primitive queries counted by the two terms are disjoint. On this class, a δ-correct procedure has deterministic cost at most

$$(K-1)\left\{n + 1 + \left\lceil \frac{2\sigma^2}{\Delta^2}\log\frac{K-1}{\delta}\right\rceil\right\}.$$

Proof. See Appendix A.4.4.

Corollary 3 (Bounded-radius constant-gain specialization)**.** *In the plateau construction of Proposition 9, let $\varepsilon = \mu\rho\theta \downarrow 0$ and, for fixed $\xi, \Lambda > 0$, set $n = \lceil \xi/\varepsilon \rceil$ and $\lambda = \Lambda\sqrt{\rho}$, and choose the admissible gap*

$$\Delta_{\rho,\theta} = \Theta(\rho\varepsilon^2) = \Theta(\rho^3\theta^2).$$

The transformed anchor radii remain uniformly bounded. For fixed K, any fixed $\delta_0 < 1/2$, and $0 < \delta \le \delta_0$, the minimax cost on this class is of order

$$(K-1)\left\{(\mu\rho\theta)^{-1} + \sigma^2\Delta_{\rho,\theta}^{-2}\log(1/\delta)\right\}$$

up to constants and the confidence allocation over finitely many arms.

Proof. See Appendix A.4.4.

On this class the traversal and statistical costs add; the result does not characterize the general upper bound. By Lemma 4, a pair of instances with a common value and gradient g cannot combine bounded anchor radius, a fixed gap, and vanishing access $g^\top Sg/\|g\|^2$ in the direction of g; that regime would need a different construction.

Secondary results. Appendix A.7 gives four secondary results: constrained standalone problems, learned curvature, a fixed-budget guarantee, and a Gaussian lower bound for Section 5.

5 KNOWN-HESSIAN QUADRATICS

Theorem 1 rules out identification over all smooth strongly convex arms when access is singular, so additional structure is needed. If arm i is a quadratic with a known Hessian, its value and gradient at a single point determine $f_i^\star$, wherever that point is.

Assumption 5 (Known-Hessian quadratics). *For a common dimension d_q, $f_i(x) = \frac{1}{2}(x-b_i)^\top H_i(x-b_i) + c_i$, where $H_i \succeq 0$ is known and b_i, c_i are unknown. The trace-corrected statistic assumes centered value and gradient noise with a known, time-constant conditional gradient covariance Σ_i. Both statistics below require public tail and state-tube bounds that hold uniformly over predictable feasible trajectories; the cross-fitted statistic also assumes a sure bound on the distance between the state and the center and a frozen pair of conditionally independent draws at the same state.*

Let $A_i = H_i^\dagger$. For a query of arm i, define

$$Y_t = F_t - \tfrac{1}{2} G_t^\top A_i G_t + \tfrac{1}{2}\operatorname{tr}(A_i \Sigma_i). \tag{7}$$

Since $H_i A_i H_i = H_i$, every feasible state satisfies $\mathbb{E}[Y_t \mid \mathcal{H}_{t-1}, i_t = i] = c_i = f_i^\star$: the coupling affects the noise of Y_t but not its mean. Appendix Proposition 16 gives the exact moment generating function under Gaussian noise and observable tail parameters.

If Σ_i is unknown, hold the state fixed for two conditionally independent pairs and set

$$Y_t^{\mathrm{cf}} = F_t^{(1)} - \tfrac{1}{2}(G_t^{(1)})^\top A_i G_t^{(2)}. \tag{8}$$

The prescribed update then uses the average of the two gradients, and $\mathbb{E}[Y_t^{\mathrm{cf}} \mid \mathcal{H}_{t-1}, i_t = i] = f_i^\star$ without knowledge of Σ_i. The two pairs must come from the same state; an update between them breaks this identity (Remark 1). Each such query costs two primitive pairs, and Theorem 6 gives its tail parameters in Eq. (68).

Given history-uniform sub-gamma parameters (v_i, κ_i) for the centered transformed samples, Appendix Corollary 6 gives an anytime confidence radius $\beta_{i,n}(\delta)$. Sweep-LUCB (Algorithm 2) queries whichever of the leader and the challenger has the wider interval, adds sparse sweeps over all arms, and stops when the intervals separate.

Theorem 4 (Quadratic identification). *Assume $K \geq 2$, $\delta \in (0,1)$, the quadratic model, a unique best arm, and the conditional sub-gamma condition of Appendix Corollary 6. Define*

$$\Delta_i = f_i^\star - f_{i^\star}^\star, \quad \gamma_{i^\star} = \min_{i \neq i^\star} \Delta_i, \quad \gamma_i = \Delta_i \ (i \neq i^\star),$$

$$M_i = \#\{n \geq 1 : 4\beta_{i,n} \geq \gamma_i\}, \qquad M_\delta = \sum_i M_i.$$

Algorithm 2 is δ-correct and its stopping time τ_Y, counted in transformed queries, satisfies

$$\Pr\Big(\tau_Y \leq K + M_\delta + K\lfloor \log_2(M_\delta + 1)\rfloor, \widehat{J}_{\tau_Y} = i^\star\Big) \geq 1 - \delta,$$

where, for $\omega_i = v_i/\gamma_i^2 + \kappa_i/\gamma_i$,

$$M_\delta = O\left(\sum_i \omega_i \left[\log\frac{K}{\delta} + \log\big(1 + \log(e + \omega_i)\big)\right]\right).$$

Proof. See Appendix A.6.2; the cross-fitted variant is Theorem 6 in Appendix A.6.1.

The algorithm needs neither the gaps nor M_δ; looser public tail bounds only delay stopping.

6 EXPERIMENTS

We test Algorithm 1 under the stochastic oracle with poorly conditioned access. Appendix B gives details, the model-selection application summarized at the end of this section, and further experiments, including those for Section 5.

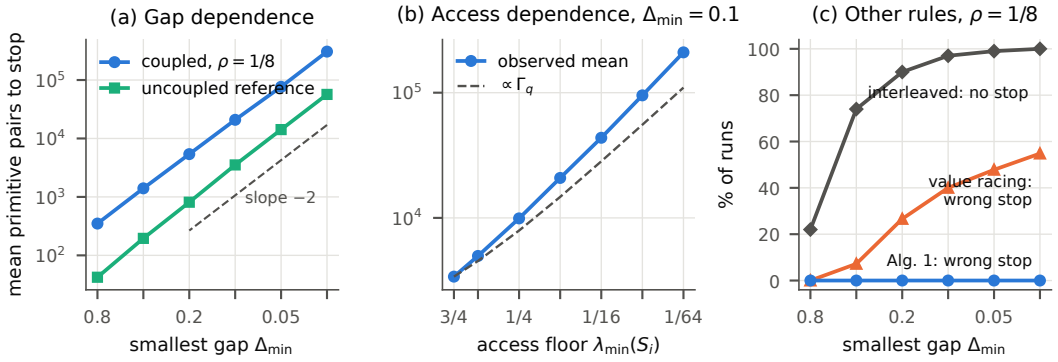


Figure 1: Algorithm 1 under the stochastic oracle (500 runs per point; bands are 95% bootstrap intervals over instances). (a) Mean cost against the gap for $\rho = 1/8$ and without coupling. (b) Mean cost against the access floor at $\Delta_{\min} = 0.1$, with Γ_q rescaled to the first point. (c) Share of runs in which value racing stopped with a wrong answer and in which the interleaved rule had not stopped by phase 24.

6.1 Stochastic oracle with poorly conditioned access

Setup. We use $K = 4$ arms with $d_i = 4$ and a weighted balance constraint $\sum_i W_i x_i = 0$ whose diagonal weights give each arm one poorly accessible coordinate, so that the access floor is $\lambda_{\min}(S_i) = \rho$; $\rho = 3/4$ is zero-sum balance. The arms are $f_i(z) = f_i^\star + \frac{1}{4}\|z - x_i^\star\|^2 + \sum_k \log\cosh(z_k - x_{ik}^\star)$ ($\mu_i = 1/2$, $L_i = 3/2$), with minimizers $x_i^\star$ drawn from the unit sphere and optima $(f_i^\star)_i$ a random permutation of $\Delta_{\min} \cdot (0, 1, 1.5, 2)$. Gaussian noise has proxies $\sigma_i^f = \sigma_i^g = 0.1$; batch means are drawn from their exact law. Updates use the geometry-scaled gain, and $\delta = 0.05$ unless stated otherwise. Each setting has 100 instances and 5 noise paths per instance. As a reference, the *uncoupled* run applies Algorithm 1 to the same arms without the constraint ($S_i = I$). Two further rules use the same oracle and transitions as Algorithm 1: *interleaved* uses exactly its batches but alternates between arms, and *value racing* is a standard racing rule that treats the observed values as samples of $f_i^\star$.

Results. Over 14 settings (two of them with $\delta = 0.01$ and 0.2) and 7,000 runs, Algorithm 1 always stopped and never returned a wrong arm (95% Wilson upper bound 0.76% per setting). For $\Delta_{\min} \le 0.2$ the log–log slope of the mean cost against the gap is -1.94 with coupling and -2.04 without (Figure 1a), consistent with Corollary 1. At the two smallest gaps the coupled cost is 5.40 times the uncoupled cost, of which the ratio of the access factors Γ_q accounts for a factor 4.33. As ρ decreases from $3/4$ to $1/64$, the mean cost grows from 3,396 to 209,926 and Γ_q from 12 to 388; the ratio of the two changes by a factor of only 1.91, and the median stopping phase by at most one (Figure 1b). Most of the dependence on access is thus explained by Γ_q, whose terms Corollary 2 shows to be tight on constant-gain classes. With the same batches, the interleaved rule did not stop within the cap on 95–100% of the runs with $\Delta_{\min} \le 0.1$, because no block gets close to its own minimizer, and it never stopped wrongly (Proposition 1). Value racing stopped quickly but returned a wrong arm on up to 55% of the runs (Figure 1c).

Application: tied linear heads on digits. Appendix B.2 applies the method to choosing among four linear heads that share parameters on 8×8 digits ($S_i = \frac{1}{2} I_{32}$). With the exact full-batch oracle of Corollary 4, certified stopping was correct on $450/450$ runs over all 45 digit pairs, versus $448/450$ for a one-sweep heuristic, with 1.00% more row evaluations (Table 4). Under the stochastic oracle it also stopped correctly, but the small gaps (about $5 \cdot 10^{-4}$) made it expensive (median 1.64×10^9 primitive queries).

Limitations. The upper bound needs frozen batches, a gain floor, and uniform controls; the lower bounds leave open vanishing access at a fixed gap and the anchor and state-envelope factors of Theorem 2. The digits comparison uses the full-batch oracle; under the stochastic oracle the cost is large when the gaps are small relative to worst-case noise proxies, as on that task.

AI USE STATEMENT

We used OpenAI Codex throughout research and writing. It assisted in finding counterexamples and proof ingredients (mostly for the lower bounds); deriving complete proofs from our proof sketches; and designing, implementing, and running experiments. It also assisted with literature search, code, and, in part, verification of the mathematics in Lean. Claude (Anthropic) assisted with revising the prose, reorganizing the appendix, and implementing and running the fresh-sampling experiment of Section 6.1. Pilot and reporting seeds were kept disjoint, and executable invariants checked the reported theorem implications. The authors reviewed the proofs, code, outputs, and rendered submission, and take responsibility for every claim and artifact. No dataset or empirical value was fabricated or manually altered.

REPRODUCIBILITY STATEMENT

Appendix A contains complete proofs and begins with a guide to where each main-text result is proved. The anonymous supplement contains the experiment runners, frozen pilot configurations, disjoint reporting seeds, raw and aggregated CSV files, manifests with script hashes and hardware metadata, and figure sources. The fresh-sampling experiment is regenerated end to end by `supplement/scripts/fresh_sampling_experiment.py` and `supplement/scripts/make_fresh_sampling_report.py` (about 20 seconds on two CPU cores). The public digits dataset is loaded through scikit-learn (Pedregosa et al., 2011); diagnostics not reported in the paper, and their datasets, are not distributed. Table 7 maps each reported number to its artifact.

REFERENCES

Aymen Al Marjani, Aurélien Garivier, and Alexandre Proutiere. Navigating to the best policy in markov decision processes. In *Advances in Neural Information Processing Systems*, volume 34, pp. 25852–25864, 2021.

Ethem Alpaydin and Cevdet Kaynak. Optical recognition of handwritten digits. UCI Machine Learning Repository, 1998. Dataset, DOI: 10.24432/C50P49.

Denis Antipov and Carola Doerr. When switching algorithms helps: A theoretical study of online algorithm selection. *arXiv preprint arXiv:2604.07473*, 2026. doi: 10.48550/arXiv.2604.07473.

Francesco Bacchiocchi, Gianmarco Genalti, Davide Maran, Marco Mussi, Marcello Restelli, Nicola Gatti, and Alberto Maria Metelli. Autoregressive bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 937–945, 2024.

Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, and Quoc Le. Understanding and simplifying one-shot architecture search. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 550–559, 2018.

Jie Bian and Vincent Y. F. Tan. Asymptotically optimal linear best feasible arm identification with fixed budget. In *Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 296–331, 2025.

Avrim Blum, Marten Garicano, Kavya Ravichandran, and Dravyansh Sharma. Algorithm design and stronger guarantees for the improving multi-armed bandits problem. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 585–619. PMLR, 2026.

Emil Carlsson, Debabrota Basu, Fredrik Johansson, and Devdatt Dubhashi. Pure exploration in bandits with linear constraints. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 334–342, 2024.

Leonardo Cella, Massimiliano Pontil, and Claudio Gentile. Best model identification: A rested bandit formulation. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1362–1372, 2021.

Rémy Degenne, Pierre Menard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 2432–2442, 2020.

Yuriy Dorn, Aleksandr Katrutsa, Ilgam Latypov, and Anastasia Soboleva. Functional multi-armed bandit and the best function identification problems. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*, pp. 2913–2921, 2026. doi: 10.65109/SFMN9947.

Yihan Du, Yuko Kuroki, and Wei Chen. Combinatorial pure exploration with full-bandit or partial linear feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 7262–7270, 2021. doi: 10.1609/aaai.v35i8.16892.

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

Gianmarco Genalti, Marco Mussi, Nicola Gatti, Marcello Restelli, Matteo Castiglioni, and Alberto Maria Metelli. Graph-triggered rising bandits. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 15351–15380, 2024.

Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pp. 240–248, 2016.

Anja Jankovic, Diederick Vermetten, Ana Kostovska, Jacob de Nobel, Tome Eftimov, and Carola Doerr. Trajectory-based algorithm selection with warm-starting. In *Proceedings of the 2022 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8. IEEE, 2022. doi: 10.1109/CEC55065.2022.9870222.

P. N. Karthik, Vincent Y. F. Tan, Arpan Mukherjee, and Ali Tajer. Optimal best arm identification with fixed confidence in restless bandits. *IEEE Transactions on Information Theory*, 70(10):7349–7384, 2024. doi: 10.1109/TIT.2024.3419924.

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

Johannes Kirschner, Tor Lattimore, and Andreas Krause. Linear partial monitoring for sequential decision making: Algorithms, regret bounds and applications. *Journal of Machine Learning Research*, 24(346):1–45, 2023.

Felipe Lagos and Jordi Pereira. Multi-armed bandit-based hyper-heuristics for combinatorial optimization problems. *European Journal of Operational Research*, 312(1):70–91, 2024. doi: 10.1016/j.ejor.2023.06.016.

Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185):1–52, 2018.

Leo Maynard-Zhang, Zhihan Xiong, Kevin Jamieson, and Maryam Fazel. On the complexity of best-arm identification in non-stationary linear bandits. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 5021–5052. PMLR, 2026.

Kazem Meidani, Seyedali Mirjalili, and Amir Barati Farimani. MAB-OS: Multi-armed bandits metaheuristic optimizer selection. *Applied Soft Computing*, 128:109452, 2022. doi: 10.1016/j.asoc.2022.109452.

Marco Mussi, Alberto Maria Metelli, and Marcello Restelli. Dynamical linear bandits. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25563–25587, 2023.

Marco Mussi, Alessandro Montenegro, Francesco Trovò, Marcello Restelli, and Alberto Maria Metelli. Best arm identification for stochastic rising bandits. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 36953–36989, 2024.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake VanderPlas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, and Jeff Dean. Efficient neural architecture search via parameter sharing. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4095–4104, 2018.

Riccardo Poiani, Alberto Maria Metelli, and Marcello Restelli. Multi-fidelity best-arm identification. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Riccardo Poiani, Rémy Degenne, Emilie Kaufmann, Alberto Maria Metelli, and Marcello Restelli. Optimal multi-fidelity best-arm identification. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Alessio Russo, Yichen Song, and Aldo Pacchiano. Pure exploration with feedback graphs. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1810–1818, 2025.

Ryoma Sato and Shinji Ito. Influential bandits: Pulling an arm may change the environment. *Transactions on Machine Learning Research*, 2025.

Sho Takemori, Yuhei Umeda, and Aditya Gopalan. Model-based best arm identification for decreasing bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1567–1575, 2024.

Diederick Vermetten, Hao Wang, Kevin Sim, and Emma Hart. To switch or not to switch: Predicting the benefit of switching between algorithms based on trajectory features. In *Applications of Evolutionary Computation (EvoApplications 2023)*, volume 13989 of *Lecture Notes in Computer Science*, pp. 335–350. Springer, 2023. doi: 10.1007/978-3-031-30229-9_22.

Zhihan Xiong, Romain Camilleri, Maryam Fazel, Lalit Jain, and Kevin Jamieson. A/B testing and best-arm identification for linear bandits with robustness to non-stationarity. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1585–1593, 2024.

Jin Xu, Xu Tan, Kaitao Song, Renqian Luo, Yichong Leng, Tao Qin, Tie-Yan Liu, and Jian Li. Analyzing and mitigating interference in neural architecture search. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 24646–24662, 2022.

APPENDIX CONTENTS

A Proofs and Technical Details **14**

A.1 Notation . . . 14

A.2 Certificate Preliminaries . . . 16

A.3 Accessibility Boundary and Geometric Block Certificates . . . 18

A.3.1 Proof of Theorem 1 . . . 18

A.3.2 Proof of Theorem 2 and Corollary 1 . . . 19

A.3.3 Full-batch oracle . . . 26

A.4 Lower Bounds . . . 27

A.4.1 Necessity of the $\Delta_{\min}^{-2}$ dependence . . . 27

A.4.2 Hidden-tail constructions for access-dependent bounds . . . 28

A.4.3 Proof of Corollary 2 . . . 31

A.4.4 Proof of Theorem 3 and Corollary 3 . . . 33

A.4.5 Singular access . . . 36

A.5 Canonical Couplings . . . 37

A.6 Known-Hessian Quadratics . . . 38

A.6.1 Transformed observations, cross-fitting, and confidence sequences . . . 38

A.6.2 Proof of Theorem 4 . . . 45

A.7 Secondary Results . . . 46

A.7.1 Constrained standalone problems . . . 46

A.7.2 Learned curvature . . . 48

A.7.3 Fixed-budget recommendation . . . 51

A.7.4 Optimality of the quadratic reduction . . . 52

B Experimental Details **53**

B.1 Fresh-sampling experiment . . . 53

B.2 Tied linear heads on digits . . . 54

B.3 Additional synthetic experiments . . . 56

B.4 Artifact map . . . 61

A Proofs and Technical Details

Where the proofs are.

- Theorem 1: Appendix A.3.1, using Proposition 10 (Appendix A.4.5).
- Proposition 1: Theorem 5 in Appendix A.2.
- Theorem 2 and Corollary 1: Appendix A.3.2; the full-batch variant is Corollary 4 (Appendix A.3.3).
- Corollary 2: Appendix A.4.3.
- Theorem 3 and Corollary 3: Appendix A.4.4.
- Theorem 4: Appendix A.6.2; Theorem 6: Appendix A.6.1.

Main and secondary results. Appendices A.2–A.6 contain the results stated or cited in the main text and the lemmas used in their proofs. Appendix A.7 contains four *secondary* results. They are not needed for the main text, but each extends it in a direction the main text does not cover: certificates for arms whose standalone problem has a convex constraint (Proposition 17), identification of quadratic arms with learned rather than known curvature (Theorem 7), a fixed-budget guarantee for the quadratic reduction (Corollary 9), and a Gaussian lower bound that matches the $\sum_{i\neq i^\star} v_i \Delta_i^{-2} \log(1/\delta)$ term of Theorem 4 (Theorem 8).

A.1 Notation

The table below lists the symbols used throughout the paper. When a statement concerns a single arm, μ, L may stand for μ_i, L_i. Other letters (for example the parameters $a, g, c, d, \lambda, \chi$ of a lower-bound construction, or events written $\mathcal{A}, \mathcal{E}$) are local to the construction or proof in which they are defined.

Symbol	Meaning
Problem and oracle (Section 3)	
K, $[K]$, $i^\star$	number of arms, $\{1,\dots,K\}$, unique best arm
f_i, d_i, $x_i^\star$	objective of arm i on $\mathbb{R}^{d_i}$, a minimizer of f_i
$f_i^\star$, Δ_i, $\Delta_{\min}$	standalone optimum $\inf f_i$, gap $f_i^\star - f_{i^\star}^\star$, $\min_{i\neq i^\star}\Delta_i$
$x=(x_1,\dots,x_K)$, d_{tot}	joint state and its dimension $\sum_i d_i$
A, b, $\mathcal{F}$	constraint $Ax=b$ and feasible set $\mathcal{F}=\{x: Ax=b\}$
E_i, P, S_i	embedding of block i, $\mathrm{Proj}_{\ker A}$, access matrix $E_i^\top P E_i$
t, i_t, B_t, r	round, queried arm, batch size, pair index in $[B_t]$
$F_{t,r}$, $G_{t,r}$, $\zeta^\bullet_{i,t,r}$	noisy value and gradient of pair r, and their noises
F_t, G_t, $\bar G_t$, $\zeta_i^\bullet$	pair when $B_t=1$, averaged gradient of a batch, averaged noises of a batch
$\mathcal{H}_{t-1}$	σ-field of the past; *predictable* means $\mathcal{H}_{t-1}$-measurable
η_t, Φ_t, $a_t(v)$, θ_t	step, response map, radial gain, realized gain $\eta_t a_t(\bar G_t)$
C_t, τ, $\widehat{J}$	cost $\sum_{s\le t} B_s$, stopping time, recommended arm
$N_i(t)$, $\tau_{i,n}$, $\tau_i(t)$	number of queries of arm i, round of its n-th query, its last query up to t
T, $R(T)$, δ	fixed budget, simple regret, confidence level
Certificates and Algorithm 1 (Section 4.2, Appendices A.2–A.3)	
μ_i, L_i	strong-convexity (or PL) and smoothness constants of arm i
ρ_i, R_i	public bounds $\rho_i \le \lambda_{\min}(S_i)$, $\lambda_{\max}(S_i)\le R_i$
θ_i^-, θ_i^+, q_i	gain floor, safe ceiling $2/(\mu_i\rho_i + L_iR_i)$, contraction $1-\mu_i\rho_i\theta_i^-$
μ_i^{eff}, L_i^{eff}	$\mu_i\rho_i$ and L_iR_i
σ_i^f, σ_i^g	tail proxies of the averaged value noise and gradient-noise norm
ℓ, u; $\underline{V}_i$, $\overline{V}_i$	certificate endpoints; current interval of arm i used for stopping
$\hat\imath$, $\check\imath$	recommended (leader) arm and challenger
h, B_h, $m_{i,h}$, $C_{i,h,j}$	phase, 2^h, number and sizes of optimization batches
$r^\bullet_{i,h}$, $\alpha^\bullet_{i,h}$, δ_h, R_h^{evt}	certificate radii, thresholds $r^\bullet_{i,h}\sqrt{B_h}$, per-phase confidence, event count
$w_i(h)$, h_δ, T_h	width bound, deciding phase, number of rounds through phase h
Γ_q	access factor $\sum_i(1+1/(\mu_i\rho_i\theta_i^-))$
x°, $\mathcal{R}_i$	feasible anchor, transformed anchor radius $\|S_i^{-1/2}(x_i^\star - x_i^\circ)\|$
n_i, h_{pool}, C_{pool}	rows of arm i; stopping phase and row evaluations of the full-batch oracle
O_{ctl}	constant depending only on the controls of Eq. (22)
Lower bounds (Section 4, Appendix A.4)	
$\Pr_0$, $\mathbb{E}_0$; $\Pr_j$, $\mathbb{E}_j$	base instance; alternative j of a construction
$\bar N_j$, $\bar N_\rho$	number of challenger updates during which an alternative is indistinguishable from the base
$\mathcal{N}(m,s^2)$, KL, kl	Gaussian law, relative entropy, binary relative entropy
Known-Hessian quadratics (Section 5, Appendix A.6)	
d_q, H_i, b_i, c_i	common dimension, Hessian, center, offset ($c_i = f_i^\star$)
Σ_i, A_i	gradient-noise covariance, $H_i^\dagger$
Y_t, Y_t^{cf}	trace-corrected and cross-fitted transformed observations
$\sigma_{f,i}$, $\sigma_{g,i}$	sub-Gaussian (moment) proxies of value noise and of gradient noise in every direction
v_i, κ_i, $\beta_{i,n}(\delta)$	sub-gamma parameters, anytime confidence radius
$\overline{Y}_{i,n}$, $\iota_n(\delta)$	mean of the first n transformed samples of arm i, log-confidence level
γ_i, M_i, M_δ, ω_i	target gaps, bad counts, $\sum_i M_i$, complexity $v_i/\gamma_i^2 + \kappa_i/\gamma_i$
τ_Y	stopping time counted in transformed queries
Conventions	
$\bullet \in \{f,g\}$	value or gradient version of a quantity
$\|\cdot\|$, $\|\cdot\|_{\mathrm{op}}$, $\|\cdot\|_F$	Euclidean norm, spectral norm, Frobenius norm
†, $\mathrm{Proj}_{\mathcal{C}}$, $(u)_+$	Moore–Penrose pseudo-inverse, Euclidean projection onto $\mathcal{C}$, $\max\{u,0\}$

A.2 Certificate Preliminaries

Proposition 2 (Deterministic certificate under smoothness and gradient domination)**.** *Let $f_i : \mathbb{R}^{d_i} \to \mathbb{R}$ have finite infimum $f_i^\star$, be L-smooth, and satisfy the μ-Polyak–Łojasiewicz (PL) inequality*

$$f_i(x) - f_i^\star \le \frac{\|\nabla f_i(x)\|^2}{2\mu} \qquad \text{for every } x \in \mathbb{R}^{d_i},$$

where $0 < \mu \le L$. Then every $x_i \in \mathbb{R}^{d_i}$ satisfies

$$f_i(x_i) - \frac{1}{2\mu}\|\nabla f_i(x_i)\|^2 \le f_i^\star \le f_i(x_i) - \frac{1}{2L}\|\nabla f_i(x_i)\|^2.$$

Proof. The lower bound is exactly the PL inequality rearranged. For the upper bound, the descent lemma gives

$$f_i\left(x_i - \frac{1}{L}\nabla f_i(x_i)\right) \le f_i(x_i) - \frac{1}{2L}\|\nabla f_i(x_i)\|^2.$$

Since $f_i^\star$ is the infimum of f_i,

$$f_i^\star \le f_i\left(x_i - \frac{1}{L}\nabla f_i(x_i)\right) \le f_i(x_i) - \frac{1}{2L}\|\nabla f_i(x_i)\|^2.$$

□

Assumption 2 implies the PL inequality with $\mu = \mu_i$, so all subsequent strongly convex results remain special cases of this proposition. Convexity is not needed for the smoothness-based upper bound.

Proposition 3 (Single-query noisy certificate under smoothness and PL)**.** *Suppose arm i is L-smooth and μ-PL as in Proposition 2, and assume the oracle model of Assumption 3. Consider a query to arm i at round s, and write*

$$F_s = f_i(x_{i,s}) + \zeta_{i,s}^f, \qquad G_s = \nabla f_i(x_{i,s}) + \zeta_{i,s}^g.$$

Fix deterministic radii $r_s^f, r_s^g \ge 0$. On the event

$$|\zeta_{i,s}^f| \le r_s^f, \qquad \|\zeta_{i,s}^g\| \le r_s^g,$$

the arm optimum satisfies

$$\ell_{i,s}^{\mathrm{sc}} \le f_i^\star \le u_{i,s}^{\mathrm{sc}},$$

where, as in Eq. (2)*,*

$$\ell_{i,s}^{\mathrm{sc}} = F_s - r_s^f - \frac{(\|G_s\| + r_s^g)^2}{2\mu},$$

and

$$u_{i,s}^{\mathrm{sc}} = F_s + r_s^f - \frac{(\|G_s\| - r_s^g)_+^2}{2L}.$$

Proof. Let $g_s = \nabla f_i(x_{i,s})$. By Proposition 2,

$$f_i(x_{i,s}) - \frac{\|g_s\|^2}{2\mu} \le f_i^\star \le f_i(x_{i,s}) - \frac{\|g_s\|^2}{2L}.$$

Since $F_s = f_i(x_{i,s}) + \zeta_{i,s}^f$, we have

$$f_i(x_{i,s}) \ge F_s - r_s^f, \qquad f_i(x_{i,s}) \le F_s + r_s^f.$$

Also,

$$\|g_s\| \le \|G_s\| + \|\zeta_{i,s}^g\| \le \|G_s\| + r_s^g,$$

and by the reverse triangle inequality,

$$\|g_s\| \ge \big(\|G_s\| - \|\zeta_{i,s}^g\|\big)_+ \ge (\|G_s\| - r_s^g)_+.$$

Substituting these bounds into the deterministic certificate gives

$$f_i^\star \geq F_s - r_s^f - \frac{(\|G_s\| + r_s^g)^2}{2\mu} = \ell_{i,s}^{\mathrm{sc}},$$

and

$$f_i^\star \leq F_s + r_s^f - \frac{(\|G_s\| - r_s^g)_+^2}{2L} = u_{i,s}^{\mathrm{sc}}.$$

□

Proposition 4 (Recent certificate width under queried-gradient and noise envelopes)**.** *Suppose every arm is L-smooth and μ-PL, and assume Assumption 3. For each arm i and query index $n \geq 1$, let $\tau_{i,n}$ be the round of the n-th query to arm i. Suppose that on some event $\mathcal{E}$, there exist nonincreasing envelopes $\bar{r}_i^f(n), \bar{r}_i^g(n), \bar{g}_i(n) \geq 0$ such that*

$$|\zeta_{i,\tau_{i,n}}^f| \leq \bar{r}_i^f(n), \qquad \|\zeta_{i,\tau_{i,n}}^g\| \leq \bar{r}_i^g(n), \qquad \|\nabla f_i(x_{i,\tau_{i,n}})\| \leq \bar{g}_i(n),$$

for all arms $i \in [K]$ and all $n \geq 1$. For a round t with $N_i(t) \geq 1$, let $W_i^{\mathrm{rec}}(t) = u_{i,s}^{\mathrm{sc}} - \ell_{i,s}^{\mathrm{sc}}$, $s = \tau_i(t)$, be the width of arm i's most recent certificate, formed with the radii $r_s^\bullet = \bar{r}_i^\bullet(N_i(t))$. Then on $\mathcal{E}$, for every arm i and every round t with $N_i(t) \geq 1$,

$$W_i^{\mathrm{rec}}(t) \leq 2\bar{r}_i^f\big(N_i(t)\big) + \frac{\big(\bar{g}_i\big(N_i(t)\big) + 2\bar{r}_i^g\big(N_i(t)\big)\big)^2}{2\mu} - \frac{\bar{g}_i\big(N_i(t)\big)^2}{2L}.$$

In particular, the recent width admits the nonincreasing envelope

$$\bar{w}_i(n) = 2\bar{r}_i^f(n) + \frac{\big(\bar{g}_i(n) + 2\bar{r}_i^g(n)\big)^2}{2\mu} - \frac{\bar{g}_i(n)^2}{2L}.$$

Proof. Fix an outcome in $\mathcal{E}$, an arm i, and a round t with $N_i(t) \geq 1$. Let $n = N_i(t)$ and $s = \tau_{i,n}$, so that the recent certificate of arm i at time t is the primitive certificate generated at round s. By Proposition 3,

$$\ell_{i,s}^{\mathrm{sc}} = F_s - \bar{r}_i^f(n) - \frac{(\|G_s\| + \bar{r}_i^g(n))^2}{2\mu},$$

and

$$u_{i,s}^{\mathrm{sc}} = F_s + \bar{r}_i^f(n) - \frac{(\|G_s\| - \bar{r}_i^g(n))_+^2}{2L}.$$

Therefore, with $y = \|G_s\|$ and $r = \bar{r}_i^g(n)$,

$$W_i^{\mathrm{rec}}(t) = u_{i,s}^{\mathrm{sc}} - \ell_{i,s}^{\mathrm{sc}} = 2\bar{r}_i^f(n) + \frac{(y+r)^2}{2\mu} - \frac{(y-r)_+^2}{2L}.$$

On $\mathcal{E}$,

$$\|G_s\| \leq \|\nabla f_i(x_{i,s})\| + \|\zeta_{i,s}^g\| \leq \bar{g}_i(n) + \bar{r}_i^g(n),$$

so

$$0 \leq y \leq \bar{g}_i(n) + r.$$

The function

$$\phi_r(y) = \frac{(y+r)^2}{2\mu} - \frac{(y-r)_+^2}{2L}$$

is nondecreasing on $[0, \infty)$: its derivative is $(y+r)/\mu$ for $y < r$, and $y(1/\mu - 1/L) + r(1/\mu + 1/L)$ for $y > r$. Hence

$$W_i^{\mathrm{rec}}(t) \leq 2\bar{r}_i^f(n) + \frac{\big(\bar{g}_i(n) + 2\bar{r}_i^g(n)\big)^2}{2\mu} - \frac{\bar{g}_i(n)^2}{2L},$$

which is exactly the claimed bound. The monotonicity of $\bar{w}_i(\cdot)$ follows because the displayed expression is nondecreasing in each of $\bar{r}_i^f, \bar{r}_i^g, \bar{g}_i$. □

Theorem 5 (Safe certificate stopping under arbitrary predictable queries)**.** *Assume the optimal arm $i^\star$ is unique and there is an event $\mathcal{E}$ on which*

$$\underline{V}_i(t) \le f_i^\star \le \overline{V}_i(t) \qquad \textit{for all } i \in [K],\ t \ge 1.$$

Let the queried arm i_t be chosen by any $\mathcal{H}_{t-1}$-measurable rule. Define the certificate stopping time by

$$\tau_{\mathrm{stop}} = \inf\left\{t \ge K : \overline{V}_{\hat{\imath}_t}(t) < \underline{V}_{\check{\imath}_t}(t)\right\},$$

where

$$\hat{\imath}_t \in \operatorname*{arg\,min}_{1\le i\le K} \overline{V}_i(t), \qquad \check{\imath}_t \in \operatorname*{arg\,min}_{i\ne\hat{\imath}_t} \underline{V}_i(t).$$

If the algorithm stops at τ_{stop} and recommends

$$\widehat{J}_{\tau_{\mathrm{stop}}} = \hat{\imath}_{\tau_{\mathrm{stop}}},$$

then on $\mathcal{E}$ the recommendation is correct. Consequently,

$$\Pr\Big(\tau_{\mathrm{stop}} < \infty \implies \widehat{J}_{\tau_{\mathrm{stop}}} = i^\star\Big) \ge \Pr(\mathcal{E}).$$

Proof. Work on $\mathcal{E}$. The query rule affects which observations appear in the history and therefore which certificate intervals are constructed, but by assumption every stopping interval remains valid on $\mathcal{E}$. Suppose the algorithm stops at some finite time τ_{stop} but recommends a suboptimal arm $\hat{\imath}_{\tau_{\mathrm{stop}}} \ne i^\star$. By definition of the challenger,

$$\underline{V}_{\check{\imath}_{\tau_{\mathrm{stop}}}}(\tau_{\mathrm{stop}}) \le \underline{V}_{i^\star}(\tau_{\mathrm{stop}}) \le f_{i^\star}^\star.$$

By validity of the stopping interval for the recommended arm,

$$f_{i^\star}^\star < f_{\hat{\imath}_{\tau_{\mathrm{stop}}}}^\star \le \overline{V}_{\hat{\imath}_{\tau_{\mathrm{stop}}}}(\tau_{\mathrm{stop}}).$$

Combining this with the stopping condition yields

$$\overline{V}_{\hat{\imath}_{\tau_{\mathrm{stop}}}}(\tau_{\mathrm{stop}}) < \underline{V}_{\check{\imath}_{\tau_{\mathrm{stop}}}}(\tau_{\mathrm{stop}}) \le f_{i^\star}^\star < f_{\hat{\imath}_{\tau_{\mathrm{stop}}}}^\star \le \overline{V}_{\hat{\imath}_{\tau_{\mathrm{stop}}}}(\tau_{\mathrm{stop}}),$$

a contradiction. Hence the recommendation must be $i^\star$. □

Proposition 1 is the special case in which the stopping intervals are the intersections of primitive certificates: an intersection of intervals that each contain $f_i^\star$ also contains $f_i^\star$, so the intersected intervals are valid on the same event.

A.3 Accessibility Boundary and Geometric Block Certificates

This subsection proves Theorems 1 and 2 and Corollary 1. The analysis allows different dimensions, strong-convexity and smoothness constants, and minimizers across arms, and any affine coupling in which every block is accessible through the feasible tangent space. The application-fixed response may be radial and may depend on the current averaged gradient, provided its effective gain stays in the public safe sector below.

A.3.1 Proof of Theorem 1

Proof of Theorem 1. Write $\mathcal{T} = \ker A = \mathrm{range}(P)$. Since $\mathcal{F} = x^\circ + \mathcal{T}$, its image in block i is

$$E_i^\top \mathcal{F} = x_i^\circ + \mathrm{range}(E_i^\top P).$$

Since $P = P^\top = P^2$, we have $S_i = (PE_i)^\top (PE_i)$. In finite dimensions,

$$\ker(S_i) = \ker(PE_i), \qquad \mathrm{range}(S_i) = \mathrm{range}(E_i^\top P),$$

where the second equality follows by taking orthogonal complements of the first. This proves Eq. (1).

If $S_i \succ 0$, its range is $\mathbb{R}^{d_i}$. Hence for each $z \in \mathbb{R}^{d_i}$, the explicit feasible lift

$$x(z) = x^\circ + PE_i S_i^{-1}(z - x_i^\circ)$$

has $x_i(z) = z$, so the complete feasible value–gradient oracle reveals $f_i(z)$ (and its gradient) at every point. If this holds for all arms, it determines every function and therefore every $f_i^\star$.

Conversely, if some S_i is singular, Proposition 10 constructs, for any fixed public bounds $0 < \mu < L$, two smooth strongly convex instances with identical value and full-gradient laws at every feasible state but different standalone optima and opposite unique best arms. Thus the vector of standalone optima is not a function of the complete feasible oracle. Together these implications prove the claimed if-and-only-if structural boundary for the unrestricted class with distinct public curvature bounds. They do not imply that a particular transition visits enough feasible points: finite-sample achievability in Theorem 2 additionally uses a positive-gain radial response in a geometry-scaled safe sector and the frozen-batch interface. □

A.3.2 Proof of Theorem 2 and Corollary 1

Projected block geometry. Let P be the Euclidean orthogonal projector onto $\ker A$, and define

$$S_i = E_i^\top P E_i, \qquad 0 < \rho_i \le \lambda_{\min}(S_i) \le \lambda_{\max}(S_i) \le R_i. \tag{9}$$

Thus S_i is positive definite. This condition holds with $S_i = I/K$ under consensus and $S_i = (1-1/K)I$ under fixed-sum balance (including zero-sum balance). At every allowed arm-i pre-batch history, retain Assumption 1 and suppose the application fixes the same step and measurable radial map $\Phi_t(v) = a_t(v)v$ across every allowed batch size. For every response argument v, assume the public pathwise bounds

$$0 < \theta_i^- \le \eta_t a_t(v) \le \theta_i^+ := \frac{2}{\mu_i\rho_i + L_i R_i}. \tag{10}$$

The response map is predictable before the batch, whereas its realized gain $\theta_t = \eta_t a_t(\bar G_t)$ may depend on the fresh averaged gradient. For a feasible x, Euclidean projection gives the exact update

$$\mathrm{Proj}_{\mathcal{F}}(x - \eta_t E_i \Phi_t(\bar G_t)) = x - \theta_t P E_i \bar G_t, \qquad x_i^+ = x_i - \theta_t S_i \bar G_t. \tag{11}$$

Arm i is assumed μ_i-strongly convex and L_i-smooth on $\mathbb{R}^{d_i}$, with its own minimizer $x_i^\star$. Put

$$\mu_i^{\mathrm{eff}} = \mu_i\rho_i, \qquad L_i^{\mathrm{eff}} = L_i R_i, \qquad q_i = 1 - \mu_i^{\mathrm{eff}}\theta_i^- \in [0,1). \tag{12}$$

The inclusion follows from $\theta_i^- \le 2/(\mu_i^{\mathrm{eff}} + L_i^{\mathrm{eff}}) \le 1/\mu_i^{\mathrm{eff}}$. The identity response with the exact geometry-scaled step is recovered by $\theta_i^- = \theta_i^+$, when $q_i = (L_i^{\mathrm{eff}} - \mu_i^{\mathrm{eff}})/(L_i^{\mathrm{eff}} + \mu_i^{\mathrm{eff}})$.

For every requested frozen-state batch of B fresh joint pairs of arm i at round t, assume that the averaged noises $\zeta_i^\bullet = B^{-1}\sum_{r=1}^B \zeta_{i,t,r}^\bullet$ satisfy, conditionally on the pre-batch history $\mathcal{H} = \mathcal{H}_{t-1}$,

$$\begin{aligned} \Pr\left(|\zeta_i^f| > \frac{u}{\sqrt{B}} \,\middle|\, \mathcal{H}\right) &\le 2\exp\left\{-\frac{u^2}{2(\sigma_i^f)^2}\right\}, \\ \Pr\left(\|\zeta_i^g\| > \frac{u}{\sqrt{B}} \,\middle|\, \mathcal{H}\right) &\le 2\exp\left\{-\frac{u^2}{2(\sigma_i^g)^2}\right\}, \end{aligned} \qquad u \ge 0. \tag{13}$$

The norm-tail proxy σ_i^g includes any dimension dependence. When a proxy is zero, the corresponding tail assumption forces that noise to vanish almost surely.

Geometric phase schedule. For $h = 0, 1, \ldots$, let

$$B_h = 2^h, \qquad \delta_h = \frac{6\delta}{\pi^2(h+1)^2}, \qquad m_{i,h} = \begin{cases} \max\left\{1, \left\lceil \dfrac{\log B_h}{-\log q_i} \right\rceil\right\}, & 0 < q_i < 1, \\ 1, & q_i = 0. \end{cases} \tag{14}$$

The number of noise events retained in the phase proof and their thresholds are

$$R_h^{\mathrm{evt}} = \sum_{i=1}^K m_{i,h} + 2K, \qquad \alpha_{i,h}^\bullet = \sigma_i^\bullet \sqrt{2\log\frac{2R_h^{\mathrm{evt}}}{\delta_h}}, \qquad \bullet \in \{f, g\}. \tag{15}$$

In phase h, visit arms in the fixed order $1, \dots, K$. For arm i, perform $m_{i,h}$ consecutive optimization macro-queries. If $0 < q_i < 1$, optimization query j averages

$$C_{i,h,j} = \left\lceil B_h q_i^{m_{i,h}-j} \right\rceil, \qquad j = 1, \dots, m_{i,h}, \tag{16}$$

fresh pairs; if $q_i = 0$, set $C_{i,h,1} = B_h$. Discard the averaged values and apply Eq. (11) using the averaged gradient. At the resulting state, freeze arm i, average B_h further fresh pairs into $F_{i,h}, G_{i,h}$, and form

$$\begin{aligned} \ell_{i,h} &= F_{i,h} - r^f_{i,h} - \frac{(\|G_{i,h}\| + r^g_{i,h})^2}{2\mu_i}, \\ u_{i,h} &= F_{i,h} + r^f_{i,h} - \frac{(\|G_{i,h}\| - r^g_{i,h})^2_+}{2L_i}, \qquad r^\bullet_{i,h} = \frac{\alpha^\bullet_{i,h}}{\sqrt{B_h}}. \end{aligned} \tag{17}$$

Insert this interval into both a lifetime intersection and the current phase-local intersection, and then apply the mandatory transition once using the certificate-batch gradient.

At the start of each phase, reset only the phase-local intersections to $(-\infty, +\infty)$. The physical state and lifetime intersections persist. After the K arm segments, test the lifetime family before the phase-local family. A family containing an empty arm intersection is ineligible. Write $[\underline{V}_i, \overline{V}_i]$ for arm i's intersection in the family currently being tested. In an eligible family, return the smallest-index $\hat{\imath}$ satisfying $\overline{V}_{\hat{\imath}} < \min_{j \neq \hat{\imath}} \underline{V}_j$, and continue if neither family contains such an arm. These priorities and empty-intersection conventions define the procedure on every sample path.

The macro-query and primitive-pair counts through phase h are

$$\begin{aligned} T_h &= \sum_{r=0}^{h} \sum_{i=1}^{K} (m_{i,r} + 1), \\ C_{T_h} &= \sum_{r=0}^{h} \sum_{i=1}^{K} \left(B_r + \sum_{j=1}^{m_{i,r}} C_{i,r,j} \right). \end{aligned} \tag{18}$$

Every macro-query is one round, so T_h counts rounds and C_{T_h} is the cost of Section 3 at the end of phase h.

Lemma 1 (Projected block contraction)**.** *For*

$$y = S_i^{-1/2}(x_i - x_i^\star), \qquad \phi_i(y) = f_i(x_i^\star + S_i^{1/2} y),$$

the function ϕ_i is μ_i^{eff}-strongly convex and L_i^{eff}-smooth. One arm-i update with realized gain θ_t and averaged gradient error ζ^g satisfies pathwise

$$\|y^+\| \le q_i \|y\| + \theta_i^+ \sqrt{R_i} \|\zeta^g\|. \tag{19}$$

When $q_i = 0$, the assumptions force equal effective curvature and the unique exact gain, so the noiseless map sends y to zero in one update.

Proof. Equation (11) gives

$$y^+ = y - \theta_t \nabla \phi_i(y) - \theta_t S_i^{1/2} \zeta^g.$$

Under a linear change of variables, strong convexity and smoothness give the constants $\mu_i^{\text{eff}} = \mu_i \rho_i$ and $L_i^{\text{eff}} = L_i R_i$. For any $0 \le \theta \le 2/(\mu_i^{\text{eff}} + L_i^{\text{eff}})$, the gradient map of a μ_i^{eff}-strongly convex, L_i^{eff}-smooth function obeys

$$\|y - z - \theta\{\nabla \phi_i(y) - \nabla \phi_i(z)\}\| \le (1 - \mu_i^{\text{eff}} \theta) \|y - z\|.$$

For C^2 objectives this follows by integrating the Hessian along the segment and bounding every eigenvalue in $[\mu_i^{\text{eff}}, L_i^{\text{eff}}]$; the general $C^{1,1}$ case follows by mollification, which preserves both constants and gives locally uniform gradient convergence. Set $z = 0$, freeze the realized scalar θ_t, and use $\theta_t \ge \theta_i^-$ and $\theta_t \le \theta_i^+$. This yields the displayed bound because $\|S_i^{1/2}\|_{\text{op}} \le \sqrt{R_i}$. No centeredness or independence of $\theta_t \zeta^g$ is used. If $q_i = 0$, then $\mu_i^{\text{eff}} \theta_i^- = 1$; the sector endpoint forces $\mu_i^{\text{eff}} = L_i^{\text{eff}}$ and $\theta_i^- = \theta_i^+ = 1/\mu_i^{\text{eff}}$. □

Lemma 2 (Full-state nonexpansiveness)**.** *For feasible states and any $0 \le \theta \le \theta_i^+$, define*

$$\mathsf{T}_{i,\theta}(x) = x - \theta P E_i \nabla f_i(x_i).$$

Then $\mathsf{T}_{i,\theta}$ is nonexpansive in the ambient Euclidean norm.

Proof. On the tangent space $\ker A$, the convex function $x \mapsto f_i(E_i^\top x)$ has tangent gradient $PE_i\nabla f_i(x_i)$. This gradient is L_iR_i-Lipschitz because

$$\|PE_i\|_{\mathrm{op}}^2 = \|E_i^\top P E_i\|_{\mathrm{op}} \le R_i.$$

Cocoercivity therefore gives, for feasible x, x',

$$\begin{aligned}\|\mathsf{T}_{i,\theta}(x) - \mathsf{T}_{i,\theta}(x')\|^2 \le \|x - x'\|^2 \\ - \left(\frac{2\theta}{L_iR_i} - \theta^2\right) \|PE_i\{\nabla f_i(x_i) - \nabla f_i(x_i')\}\|^2.\end{aligned}$$

Since $\theta \le \theta_i^+ \le 2/(L_iR_i)$, the subtracted coefficient is nonnegative. □

Feasible anchors and a block-start envelope. Fix any analysis-only $x^\circ \in \mathcal{F}$, write $x_i^\circ = E_i^\top x^\circ$, and set

$$\begin{aligned} r_i^\circ &= x^\circ + PE_iS_i^{-1}(x_i^\star - x_i^\circ), \\ \mathcal{R}_i &= \|r_i^\circ - x^\circ\| = \|S_i^{-1/2}(x_i^\star - x_i^\circ)\|. \end{aligned} \tag{20}$$

The anchor is feasible, its i-th block is $x_i^\star$, and hence $\mathsf{T}_{i,\theta}(r_i^\circ) = r_i^\circ$ for every allowed θ. Let x_{init} be the initial joint state and

$$\begin{aligned} D_x &= \|x_{\mathrm{init}} - x^\circ\|, \qquad \bar\alpha_h = \max_{\substack{0\le r\le h \\ i\in[K]}} \alpha_{i,r}^g, \\ d_i^{\mathrm{seg}} &= \theta_i^+\sqrt{R_i}\left\{1 + \frac{1}{1-\sqrt{q_i}}\right\}, \\ D_h &= D_x + (h+1)\left\{2\sum_{i=1}^K \mathcal{R}_i + \bar\alpha_h \sum_{i=1}^K d_i^{\mathrm{seg}}\right\}, \\ D_{i,h} &= \frac{D_h + \mathcal{R}_i}{\sqrt{\rho_i}}. \end{aligned} \tag{21}$$

The convention $1/(1-\sqrt{q_i}) = 1$ applies at $q_i = 0$.

For rate statements that allow the gap and confidence level to vary, call a family of instances *uniformly controlled* if K is fixed and there are common constants $\underline{\mu}, \underline{\rho}, \underline{\varepsilon} > 0$ and $\overline{L}, \overline{R}, \overline{\sigma}, \overline{D} < \infty$ such that every arm and instance satisfies

$$\begin{aligned} &\mu_i \ge \underline{\mu}, \quad L_i \le \overline{L}, \quad \rho_i \ge \underline{\rho}, \quad R_i \le \overline{R}, \\ &\mu_i\rho_i\theta_i^- \ge \underline{\varepsilon}, \quad \sigma_i^f \vee \sigma_i^g \le \overline{\sigma}, \\ &D_x + \sum_{i=1}^K \mathcal{R}_i \le \overline{D}, \end{aligned} \tag{22}$$

for some feasible analysis anchor x°, which may vary with the instance. We write $O_{\mathrm{ctl}}(\cdot)$ when the hidden constant depends only on K and these common controls. In particular, it is independent of the particular instance, $\Delta_{\min}$, and δ. These conditions imply $q_i \le \overline{q} := 1 - \underline{\varepsilon} < 1$, uniformly bounded θ_i^+, d_i^{seg}, and Γ_q, and uniformly linear growth of $m_{i,h}$ in $h+1$.

Lemma 3 (Polynomial block-start envelope)**.** *On the event that all configured gradient bounds through phase h hold, every reached state through that phase satisfies $\|x - x^\circ\| \le D_h$. Immediately before any arm-i segment in phase h,*

$$\|S_i^{-1/2}(x_i - x_i^\star)\| \le D_{i,h}.$$

Moreover, uniformly over any uniformly controlled family,

$$D_h,\ D_{i,h} = O_{\mathrm{ctl}}\left((h+1)\sqrt{\log\frac{K(h+2)}{\delta}}\right).$$

Proof. The equality for $\mathcal{R}_i$ follows from $P^2 = P$ and $E_i^\top P E_i = S_i$. All noiseless maps in an arm-i segment share the fixed point r_i°. Lemma 2 and a one-noise-at-a-time comparison give

$$\|x_{\text{end}} - r_i^\circ\| \le \|x_{\text{start}} - r_i^\circ\| + \theta_i^+ \sqrt{R_i} \sum_j \|\zeta_j^g\|.$$

For $0 < q_i < 1$,

$$\sum_{j=1}^{m_{i,h}} C_{i,h,j}^{-1/2} \le B_h^{-1/2} \sum_{k=0}^{m_{i,h}-1} q_i^{-k/2} \le \frac{1}{1-\sqrt{q_i}}. \tag{23}$$

For $h \ge 1$, the last inequality follows from the minimality of $m_{i,h}$, which gives $B_h^{-1/2} q_i^{-(m_{i,h}-1)/2} < 1$, followed by a geometric sum; for $h = 0$ it follows directly. At $q_i = 0$, the single optimization batch contributes $B_h^{-1/2} \le 1$. The certificate transition contributes at most $\theta_i^+ \sqrt{R_i} \bar{\alpha}_h / \sqrt{B_h}$.

Converting twice between r_i° and x°, an entire segment, and likewise any prefix of it, increases the x°-distance by at most $2\mathcal{R}_i + d_i^{\text{seg}} \bar{\alpha}_h$. Summing these increments over the $K(h+1)$ reached segments proves the first claim. Since r_i° has i-th block $x_i^\star$,

$$\begin{aligned} \|S_i^{-1/2}(x_i - x_i^\star)\| &\le \rho_i^{-1/2} \|x_i - x_i^\star\| \\ &\le \rho_i^{-1/2} \|x - r_i^\circ\| \le D_{i,h}. \end{aligned}$$

Finally, Eq. (22) implies uniformly that $R_h^{\text{evt}} = O_{\text{ctl}}(h+1)$ and $\bar{\alpha}_h = O_{\text{ctl}}(\sqrt{\log(K(h+2)/\delta)})$. It also bounds the anchor sums and segment-response factors in D_h, proving the stated uniform order. □

Certification width and analysis phase. Unrolling Lemma 1 over the optimization queries uses

$$\frac{q_i^{m_{i,h}-j}}{\sqrt{C_{i,h,j}}} \le \frac{q_i^{(m_{i,h}-j)/2}}{\sqrt{B_h}}, \qquad q_i^{m_{i,h}} \le B_h^{-1}.$$

Thus, on the simultaneous confidence event, immediately before the certificate batch,

$$\|S_i^{-1/2}(x_i - x_i^\star)\| \le \frac{D_{i,h}}{B_h} + \frac{\theta_i^+ \sqrt{R_i}\, \alpha_{i,h}^g}{(1-\sqrt{q_i})\sqrt{B_h}}. \tag{24}$$

At $q_i = 0$, the actual inherited contribution $q_i^{m_{i,h}} D_{i,h}$ vanishes; Eq. (24) remains valid with the displayed $D_{i,h}/B_h$ as a conservative upper bound and $1 - \sqrt{q_i} = 1$. Define

$$\begin{aligned} Q_i(h) &= L_i \sqrt{R_i} \left\{ \frac{D_{i,h}}{\sqrt{B_h}} + \frac{\theta_i^+ \sqrt{R_i}\, \alpha_{i,h}^g}{1-\sqrt{q_i}} \right\}, \\ w_i(h) &= \frac{2\alpha_{i,h}^f}{\sqrt{B_h}} + \frac{1}{B_h} \left\{ \frac{(Q_i(h) + 2\alpha_{i,h}^g)^2}{2\mu_i} - \frac{Q_i(h)^2}{2L_i} \right\}. \end{aligned} \tag{25}$$

Smoothness and Eq. (24) give

$$\|\nabla f_i(x_i)\| \le Q_i(h)/\sqrt{B_h}$$

at certification, and the monotone width calculation in Proposition 4 gives $u_{i,h} - \ell_{i,h} \le w_i(h)$. For the unique best arm $i^\star$, write

$$\Delta_i = f_i^\star - f_{i^\star}^\star \quad (i \ne i^\star), \qquad \Delta_{\min} = \min_{i \ne i^\star} \Delta_i > 0,$$

and let

$$h_\delta = \min\{h \ge 0 : w_{i^\star}(h) + w_i(h) < \Delta_i \text{ for every } i \ne i^\star\}. \tag{26}$$

This is an analysis quantity, not an algorithmic input.

Proof of Theorem 2. For each phase h, retain the gradient event for every reached optimization batch and the value and gradient events for every reached certificate batch. The reach event is measurable

before its batch is drawn. When $\sigma_i^\bullet > 0$, Eqs. (13) and (15) bound the conditional failure probability of each retained event by

$$2\exp\left\{-\frac{(\alpha_{i,h}^\bullet)^2}{2(\sigma_i^\bullet)^2}\right\} = \frac{\delta_h}{R_h^{\mathrm{evt}}}.$$

For a zero proxy the event holds almost surely. Sequential conditioning and a union bound over the R_h^{evt} possible events show that the reached-phase failure probability is at most δ_h. Hence the all-phase event $\mathcal{E}$ satisfies

$$\Pr(\mathcal{E}) \geq 1 - \sum_{h\geq 0} \delta_h = 1 - \delta.$$

On $\mathcal{E}$, Eq. (17) contains $f_i^\star$ by Proposition 3. Intersections of these intervals remain valid, so an eligible strict separator can only recommend $i^\star$. The exact primitive width, with $y = \|G_{i,h}\|$, is

$$2r_{i,h}^f + \frac{(y + r_{i,h}^g)^2}{2\mu_i} - \frac{(y - r_{i,h}^g)_+^2}{2L_i}.$$

It is nondecreasing in $y \geq 0$. At certification,

$$\|G_{i,h}\| \leq \frac{Q_i(h) + \alpha_{i,h}^g}{\sqrt{B_h}},$$

so substitution gives exactly the width $w_i(h)$ in Eq. (25).

At the end of phase h_δ, the current phase contains a valid interval for every arm; write $[\underline{V}_i^{\mathrm{ph}}, \overline{V}_i^{\mathrm{ph}}]$ for the phase-local intersections. For every $i \neq i^\star$,

$$\begin{aligned}\overline{V}_{i^\star}^{\mathrm{ph}} &\leq f_{i^\star}^\star + w_{i^\star}(h_\delta) \\ &< f_i^\star - w_i(h_\delta) \leq \underline{V}_i^{\mathrm{ph}}.\end{aligned}$$

Thus the algorithm stops correctly by T_{h_δ} on $\mathcal{E}$. It follows that

$$\Pr(\tau \leq T_{h_\delta}, \widehat{J}_\tau = i^\star) \geq 1 - \delta, \qquad \Pr(\tau < \infty, \widehat{J}_\tau \neq i^\star) \leq \delta.$$

We next verify finiteness and cost. Any fixed admissible instance is a controlled singleton with finite constants, which already proves the asserted instancewise finiteness; the same argument is uniform over the controlled families below. Put

$$\mathcal{L}_h = \log(K/\delta) + \log(h+2).$$

Equations (21) and (25) give

$$w_i(h) = O_{\mathrm{ctl}}\left(\sqrt{\frac{\mathcal{L}_h}{2^h}} + \frac{\mathcal{L}_h}{2^h} + \frac{h^2\mathcal{L}_h}{4^h}\right). \tag{27}$$

Hence $h_\delta < \infty$, and the elementary exponential-versus-log inversion yields

$$2^{h_\delta} = O_{\mathrm{ctl}}\left(\Delta_{\min}^{-2}\left\{\log\frac{K}{\delta} + \log\log\frac{e}{\Delta_{\min}}\right\}\right). \tag{28}$$

Indeed, it suffices to choose 2^h larger than a fixed multiple of $\Delta_{\min}^{-2}\{\log(K/\delta) + \log(h+2)\}$; the proposed right-hand side makes this inequality self-consistent. The remaining two terms in Eq. (27) are then smaller after increasing the fixed constant.

The batch definition and $\lceil u \rceil \leq u + 1$ give

$$\sum_{j=1}^{m_{i,h}} C_{i,h,j} \leq B_h \sum_{k=0}^{m_{i,h}-1} q_i^k + m_{i,h} \leq \frac{B_h}{1-q_i} + m_{i,h}.$$

Consequently,

$$C_{T_h} \leq (2^{h+1}-1)\sum_{i=1}^{K}\left(1 + \frac{1}{1-q_i}\right) + \sum_{r=0}^{h}\sum_{i=1}^{K} m_{i,r}. \tag{29}$$

Also, for $\mathcal{J} = \{i : 0 < q_i < 1\}$,

$$T_h \le 2K(h+1) + \frac{(\log 2)h(h+1)}{2} \sum_{i \in \mathcal{J}} \frac{1}{-\log q_i}. \tag{30}$$

Combining these bounds with Eq. (28), uniformly over the controlled family, all $\delta \in (0,1)$, and $0 < \Delta_{\min} \le 1$,

$$C_{T_{h_\delta}} = O_{\mathrm{ctl}}\left(\Delta_{\min}^{-2}\left\{\log\frac{K}{\delta} + \log\log\frac{e}{\Delta_{\min}}\right\}\right).$$

For fixed δ, uniformly over the same family, $T_{h_\delta} = O_{\mathrm{ctl},\delta}(\log^2(e/\Delta_{\min}))$, where the constant may also depend on δ. Without fixing δ, Eqs. (30) and (28) give

$$T_{h_\delta} = O_{\mathrm{ctl}}\left(\left\{\log\frac{e}{\Delta_{\min}} + \log\log\frac{eK}{\delta}\right\}^2\right).$$

These are the deterministic bounds on the high-probability deadline event.

It remains to prove termination outside $\mathcal{E}$. Let $\mathcal{E}_h$ denote the reached-phase event used above. Since $\sum_h \Pr(\mathcal{E}_h^c) \le \sum_h \delta_h < \infty$, the first Borel–Cantelli lemma, without independence, shows that only finitely many reached phases fail almost surely. On $\{\tau = \infty\}$, every phase is reached. After the last bad phase, the state produced by the finitely many earlier observations is a finite random starting point, and the segment argument in Lemma 3 gives

$$\|x - x^\circ\| = O(h\sqrt{\log h})$$

almost surely along late phases. Unrolling Lemma 1 through a late good arm block then gives

$$\|S_i^{-1/2}(x_i - x_i^\star)\| \le O\left(\frac{h\sqrt{\log h}}{B_h}\right) + O\left(\sqrt{\frac{\log h}{B_h}}\right)\sum_{k\ge 0} q_i^{k/2} \longrightarrow 0.$$

This includes $q_i = 0$, for which the sum equals one and the noiseless initial term is zero. The final certificate-batch noises and radii also converge to zero in every sufficiently late good phase. Therefore

$$F_{i,h} \to f_i^\star, \qquad G_{i,h} \to 0,$$

and both endpoints of every phase-local certificate converge to $f_i^\star$. A sufficiently late phase places every endpoint within $\Delta_{\min}/4$ of its target and forces strict separation, contradicting $\tau = \infty$. Hence $\Pr(\tau < \infty) = 1$, proving δ-correctness.

It remains to control the expectation without defining a trajectory after stopping. Enumerate the potential batches in the deterministic phase schedule. For a potential batch ν, let $\mathcal{G}_\nu$ be its pre-batch reach event, B_ν its size, $\zeta_\nu^\bullet$ its averaged noises, and $\sigma_\nu^\bullet = \sigma_{i_\nu}^\bullet$, where i_ν is its arm. Define the killed errors $\widetilde{Z}_\nu^\bullet = \mathbf{1}_{\mathcal{G}_\nu} Z_\nu^\bullet$, where $Z_\nu^f = |\zeta_\nu^f|$ and $Z_\nu^g = \|\zeta_\nu^g\|$. Conditional tail integration gives, for every $p > 0$,

$$\|\widetilde{Z}_\nu^\bullet\|_{L^p} \le \frac{\mathsf{m}_p \sigma_\nu^\bullet}{\sqrt{B_\nu}}, \qquad \mathsf{m}_p = \{p 2^{p/2}\Gamma(p/2)\}^{1/p}. \tag{31}$$

Here $\|Z\|_{L^p} = (\mathbb{E}|Z|^p)^{1/p}$ and Γ is Euler's Gamma function. The zero-proxy case is immediate because the corresponding noise is almost surely zero. For $\sigma_\nu^\bullet > 0$, $\mathcal{G}_\nu$ is pre-batch measurable and

$$\mathbb{E}[(\widetilde{Z}_\nu^\bullet)^p] \le 2p\int_0^\infty t^{p-1} e^{-B_\nu t^2/(2(\sigma_\nu^\bullet)^2)} dt = \mathsf{m}_p^p (\sigma_\nu^\bullet)^p B_\nu^{-p/2}.$$

For a potential state x with reach event $\mathcal{G}$, write $\widetilde{X}_{\mathcal{G}} = \mathbf{1}_{\mathcal{G}}\|x - x^\circ\|$, and, when x is arm i's phase-h certificate state, write $\widetilde{D}_{i,h}^{\mathrm{cert}} = \mathbf{1}_{\mathcal{G}}\|S_i^{-1/2}(x_i - x_i^\star)\|$. Set

$$\begin{aligned}
\mathsf{A}_h &= D_x + (h+1)\sum_i \left[2\mathcal{R}_i + \theta_i^+\sqrt{R_i}\,\mathsf{m}_8\sigma_i^g\left\{1 + \frac{1}{1-\sqrt{q_i}}\right\}\right], \\
\mathsf{Z}_{i,h} &= \frac{\mathsf{A}_h + \mathcal{R}_i}{\sqrt{\rho_i}}, \\
\mathsf{d}_{i,h} &= \frac{\mathsf{Z}_{i,h}}{B_h} + \frac{\theta_i^+\sqrt{R_i}\,\mathsf{m}_8\sigma_i^g}{(1-\sqrt{q_i})\sqrt{B_h}}.
\end{aligned} \tag{32}$$

The reciprocal is one at $q_i = 0$.

Multiply the segment-prefix inequality from Lemma 3 by the later state's reach indicator. Later reach implies every earlier batch reach, so Minkowski, Eq. (31), and $\sum_j C_{i,h,j}^{-1/2} \le (1 - \sqrt{q_i})^{-1}$ give

$$\|\widetilde{X}_{\mathcal{G}}\|_{L^8} \le \mathsf{A}_h.$$

Unrolling Lemma 1 on the certificate reach event then gives

$$\|\widetilde{D}_{i,h}^{\mathrm{cert}}\|_{L^8} \le \mathsf{d}_{i,h},$$

using $q_i^{m_{i,h}} \le B_h^{-1}$ and $q_i^{m_{i,h}-j}/\sqrt{C_{i,h,j}} \le q_i^{(m_{i,h}-j)/2}/\sqrt{B_h}$.

Define

$$\begin{aligned}\mathsf{e}_{i,h} := \frac{L_i R_i}{2}\mathsf{d}_{i,h}^2 + \frac{\mathsf{m}_4\sigma_i^f}{\sqrt{B_h}} + r_{i,h}^f \\ + \frac{1}{2\mu_i}\left(L_i\sqrt{R_i}\,\mathsf{d}_{i,h} + \frac{\mathsf{m}_8\sigma_i^g}{\sqrt{B_h}} + r_{i,h}^g\right)^2.\end{aligned} \tag{33}$$

Let h_τ be the stopping phase and let $\widetilde{\mathsf{E}}_{i,h}$ be the maximum absolute deviation of the two phase-h certificate endpoints from $f_i^\star$, multiplied by $\mathbf{1}\{h_\tau \ge h\}$. With $\widetilde{Z}_{i,h}^{\bullet}$ the killed errors of arm i's phase-h certificate batch, smoothness and the certificate formula give pointwise

$$\begin{aligned}\widetilde{\mathsf{E}}_{i,h} \le \frac{L_i R_i}{2}(\widetilde{D}_{i,h}^{\mathrm{cert}})^2 + \widetilde{Z}_{i,h}^f + r_{i,h}^f \\ + \frac{1}{2\mu_i}\left(L_i\sqrt{R_i}\widetilde{D}_{i,h}^{\mathrm{cert}} + \widetilde{Z}_{i,h}^g + r_{i,h}^g\right)^2.\end{aligned}$$

Thus Minkowski and Eq. (31) imply $\|\widetilde{\mathsf{E}}_{i,h}\|_{L^4} \le \mathsf{e}_{i,h}$. If phase h is reached and every endpoint error is below $\Delta_{\min}/3$, its phase-local family strictly separates $i^\star$. Therefore

$$\Pr(h_\tau > h) \le \sum_i \left(\frac{3\mathsf{e}_{i,h}}{\Delta_{\min}}\right)^4. \tag{34}$$

Write $c_0 = C_{T_0}$ and $c_h = C_{T_h} - C_{T_{h-1}}$ for $h \ge 1$. These phase costs are deterministic, so Eqs. (29) and (34) yield

$$\mathbb{E}C_\tau \le c_0 + \sum_{h\ge 1} c_h \min\left\{1, \sum_i \left(\frac{3\mathsf{e}_{i,h-1}}{\Delta_{\min}}\right)^4\right\}. \tag{35}$$

For a fixed instance,

$$\mathsf{e}_{i,h} = O\left(\sqrt{\frac{\log(h+2)}{2^h}} + \frac{(h+1)^2}{4^h}\right).$$

Since $c_h = O(2^h + h)$, the tail of Eq. (35) is dominated by

$$\sum_h (2^h + h)\left\{\frac{\log^2(h+2)}{4^h} + \frac{(h+1)^8}{16^h}\right\} < \infty.$$

This proves $\mathbb{E}C_\tau < \infty$.

Uniformly over Eq. (22),

$$\mathsf{e}_{i,h} \le C_{\mathrm{ctl}}\left\{\sqrt{\frac{\mathcal{L}_h}{2^h}} + \frac{\mathcal{L}_h}{2^h} + \frac{(h+1)^2}{4^h}\right\}, \qquad \mathcal{L}_h = \log(K/\delta) + \log(h+2). \tag{36}$$

Choose h_0 with 2^{h_0} of the order in Eq. (28). Increasing its controlled constant makes $\mathsf{e}_{i,h_0} \le c\Delta_{\min}$, while $\mathcal{L}_{h_0+s}/\mathcal{L}_{h_0} \le C(s+1)$. Hence

$$\Delta_{\min}^{-4}\mathsf{e}_{i,h_0+s}^4 \le C_{\mathrm{ctl}}(s+1)^8\,4^{-s}.$$

The expected tail after h_0 is at most $C_{\mathrm{ctl}}2^{h_0}\sum_{s\ge 0}(s+1)^8 2^{-s}$, and the cost through h_0 has the same order. This proves the expected-cost rate stated in Corollary 1. □

A.3.3 FULL-BATCH ORACLE

Corollary 4 (Exact finite-pool identification)**.** *Replace the stochastic oracle by the following stronger full-batch oracle. Arm i is the exact average over a fixed finite pool of n_i rows, including its deterministic regularizer and cost; an allowed macro evaluates all rows once at the frozen state, returns the exact finite empirical value and gradient, costs n_i primitive pairs, and then triggers one mandatory update. The target is the standalone optimum of this finite empirical objective.*

In phase h, before an exact full-pool certificate macro for arm i, take

$$m_{i,h}^{\mathrm{pool}} = \begin{cases} \max\{1, \lceil h \log 2/(-\log q_i)\rceil\}, & 0 < q_i < 1, \\ 1, & q_i = 0, \end{cases}$$

exact full-pool optimization macros. Define

$$\begin{aligned} D_h^{\mathrm{pool}} &= D_x + 2(h+1)\sum_j \mathcal{R}_j, \qquad D_{i,h}^{\mathrm{pool}} = \frac{D_h^{\mathrm{pool}} + \mathcal{R}_i}{\sqrt{\rho_i}}, \\ w_i^{\mathrm{pool}}(h) &= \frac{L_i^2 R_i}{2}\left(\frac{1}{\mu_i} - \frac{1}{L_i}\right)(D_{i,h}^{\mathrm{pool}})^2\, 4^{-h}, \\ h_{\mathrm{pool}} &= \min\{h : w_{i^\star}^{\mathrm{pool}}(h) + w_i^{\mathrm{pool}}(h) < \Delta_i \ \forall i \neq i^\star\}. \end{aligned} \tag{37}$$

The resulting phase-local strict-separation rule is deterministically correct and stops by phase $h_{\mathrm{pool}} < \infty$; its total number C_{pool} of row evaluations satisfies

$$\begin{aligned} C_{\mathrm{pool}} \le\ & 2(h_{\mathrm{pool}}+1)\sum_i n_i \\ & + \frac{(\log 2) h_{\mathrm{pool}}(h_{\mathrm{pool}}+1)}{2}\sum_{i:q_i>0}\frac{n_i}{-\log q_i}. \end{aligned} \tag{38}$$

Uniformly over the controlled families of Eq. (22)*,*

$$\begin{aligned} h_{\mathrm{pool}} &= O_{\mathrm{ctl}}(\log(e + 1/\Delta_{\min})), \\ C_{\mathrm{pool}} &= O_{\mathrm{ctl}}\left(\sum_i n_i \left\{1 + \frac{\mathbf{1}\{q_i > 0\}}{-\log q_i}\right\} \log^2\left(e + \frac{1}{\Delta_{\min}}\right)\right). \end{aligned}$$

Proof. Every exact arm-i update is nonexpansive about r_i°. Consequently, any prefix of that arm's segment satisfies

$$\|x_{\mathrm{end}} - x^\circ\| \le \|x_{\mathrm{start}} - x^\circ\| + 2\mathcal{R}_i.$$

Summing over earlier segments bounds every state in phase h by D_h^{pool}, and hence its arm-i transformed start distance by $D_{i,h}^{\mathrm{pool}}$. The exact contraction and the block length give

$$\|S_i^{-1/2}(x_i^{\mathrm{cert}} - x_i^\star)\| \le q_i^{m_{i,h}^{\mathrm{pool}}} D_{i,h}^{\mathrm{pool}} \le 2^{-h} D_{i,h}^{\mathrm{pool}};$$

at $q_i = 0$, the left side is zero after one optimization macro. Thus $\|\nabla f_i(x_i^{\mathrm{cert}})\| \le L_i\sqrt{R_i}\, 2^{-h} D_{i,h}^{\mathrm{pool}}$. The exact zero-radius certificate contains $f_i^\star$ and has width

$$\frac{\|\nabla f_i(x_i^{\mathrm{cert}})\|^2}{2}\left(\frac{1}{\mu_i} - \frac{1}{L_i}\right) \le w_i^{\mathrm{pool}}(h).$$

This proves separation by h_{pool}.

For $q_i > 0$, $m_{i,h}^{\mathrm{pool}} \le 1 + h\log 2/(-\log q_i)$; at zero it equals one. Summing the $m_{i,h}^{\mathrm{pool}} + 1$ full-pool macros through h_{pool} proves Eq. (38). Finally, $w_i^{\mathrm{pool}}(h) = O_{\mathrm{ctl}}((h+1)^2 4^{-h})$. For every $\Delta_{\min} > 0$, exponential–polynomial inversion gives $h_{\mathrm{pool}} = O_{\mathrm{ctl}}(1 + \max\{0, \log(1/\Delta_{\min})\})$, which is equivalent up to universal constants to $O_{\mathrm{ctl}}(\log(e + 1/\Delta_{\min}))$; substituting this bound in Eq. (38) gives the displayed cost order. □

A.4 Lower Bounds

This subsection proves the lower bounds stated in Section 4. Proposition 5 shows that the $\Delta_{\min}^{-2}$ dependence of Corollary 1 cannot be removed. Propositions 6 and 7, Lemma 4, and Corollary 5 construct the hidden-tail instances behind the access-dependent bounds; Propositions 8 and 9 prove Corollary 2 and Theorem 3; and Proposition 10 gives the singular-access construction used in Theorem 1.

Conventions. Each construction has a base instance, indexed by 0, and alternatives, indexed by 1 or by a challenger j. We write $\Pr_0, \mathbb{E}_0$ and $\Pr_j, \mathbb{E}_j$ for probability and expectation of the stopped transcript under these instances, and $i_0^\star, i_1^\star$ for the best arms of a pair of instances. A rule is *uniformly* δ-correct on a class if it is δ-correct on every instance of the class.

A.4.1 Necessity of the $\Delta_{\min}^{-2}$ dependence

Proposition 5 (A nonquadratic Δ^{-2} lower-bound subclass)**.** *Fix $\delta \in (0, 1/2)$ and positive value-noise proxies $\sigma_i^f > 0$. There is a nonquadratic subclass satisfying the assumptions above on which every δ-correct rule has*

$$\mathbb{E}[C_\tau] \geq \sum_{i \neq i^\star} \frac{2(\sigma_i^f)^2}{\Delta_i^2} \operatorname{kl}(1-\delta, \delta). \tag{39}$$

Moreover, for any $p \in (0, 1-\delta)$, define the joint correct-deadline quantile

$$Q_{1-p}^{\mathrm{corr}}(C_\tau) = \inf\{n \in \mathbb{N} : \Pr(C_\tau \leq n, \widehat{J}_\tau = i^\star) \geq 1-p\}.$$

Here $\inf \varnothing = +\infty$. *Then*

$$Q_{1-p}^{\mathrm{corr}}(C_\tau) \geq \max_{i \neq i^\star} \frac{2(\sigma_i^f)^2}{\Delta_i^2} \operatorname{kl}(1-p, \delta). \tag{40}$$

Proof. In one dimension take, for fixed $\alpha > a > 0$,

$$f_i(x) = c_i + \frac{\alpha}{2}x^2 + a(1 - \cos x).$$

Its Hessian lies in $[\alpha - a, \alpha + a]$, it is not quadratic, and zero is its unique minimizer. Initialize at zero, use exact gradients, and give arm i independent $\mathcal{N}(0, (\sigma_i^f)^2)$ primitive value noise. Every projected update remains at zero, so the transcript is ordinary Gaussian best-arm identification with mean $c_i = f_i^\star$.

Call this instance 0. For a challenger i, lower only c_i to $c_{i^\star} - \varepsilon$, and call the result instance i. The standard stopped change-of-measure inequality gives, with $C_{i,\tau}$ denoting the number of primitive pairs drawn from arm i,

$$\mathbb{E}[C_{i,\tau}] \frac{(\Delta_i + \varepsilon)^2}{2(\sigma_i^f)^2} \geq \operatorname{kl}(1-\delta, \delta).$$

Letting $\varepsilon \downarrow 0$ and summing over challengers proves the expected-cost claim.

For the quantile claim, fix n with $\Pr_0(C_\tau \leq n, \widehat{J}_\tau = i^\star) \geq 1-p$, and let $\mathcal{A}_n = \{\widehat{J}_\tau = i^\star, C_\tau \leq n\}$. Then

$$\Pr_0(\mathcal{A}_n) \geq 1-p, \qquad \Pr_i(\mathcal{A}_n) \leq \delta.$$

Expose raw pairs sequentially while keeping every action within a macro-batch fixed before its draws, pad the transcript after stopping, and truncate it after n primitive draws; a predictable batch request that would cross n makes $\mathcal{A}_n$ false. The truncated chain rule and data processing give

$$n \frac{(\Delta_i + \varepsilon)^2}{2(\sigma_i^f)^2} \geq \operatorname{kl}(1-p, \delta),$$

because $1 - p > \delta$. Letting $\varepsilon \downarrow 0$, taking the maximum over challengers, and then the smallest eligible n proves Eq. (40). Thus the Δ^{-2} primitive exponent is necessary both in expectation and at any fixed confidence quantile. □

A.4.2 HIDDEN-TAIL CONSTRUCTIONS FOR ACCESS-DEPENDENT BOUNDS

Proposition 6 (Positive access still has a geometry cost)**.** *Fix $0 < \mu < L$, $0 < \rho \le 1/2$, $a, g, \lambda > 0$, and $\delta \in (0, 1/2)$. Put*

$$\chi_\rho = \frac{L - \mu\rho}{L + \mu\rho}, \qquad \bar{N}_\rho = \left\lfloor \frac{\log(1 + \mu a/g)}{-\log \chi_\rho} \right\rfloor, \qquad \Delta = \frac{g^2}{4}\left(\frac{1}{\mu} - \frac{1}{L}\right). \tag{41}$$

There are two noiseless $K = 2$ instances satisfying the hypotheses of Theorem 2, with the same affine set, initial state, public constants, allowed batches, and prescribed transitions, such that every block is accessible, the smallest block-access eigenvalue is ρ, the best arm is different in the two instances, and both gaps are the fixed value Δ. Their transcript laws agree on every event that stops by primitive cost $\bar{N}_\rho$. Consequently every rule that is uniformly δ-correct on this class satisfies, on each constructed instance,

$$\Pr(C_\tau \le \bar{N}_\rho) \le 2\delta, \qquad \Pr(C_\tau \le \bar{N}_\rho, \widehat{J}_\tau = i^\star) \le \delta, \qquad \mathbb{E}[C_\tau] \ge (1 - 2\delta)\bar{N}_\rho. \tag{42}$$

Moreover,

$$\bar{N}_\rho \ge \frac{L - \mu\rho}{2\mu\rho} \log\left(1 + \frac{\mu a}{g}\right) - 1, \tag{43}$$

so this family has a $\Omega(L/(\mu\rho))$ access-geometry obstruction as $\rho \downarrow 0$. More sharply, for fixed a, g, μ, L,

$$\lim_{\rho \downarrow 0} \frac{\bar{N}_\rho}{1 + (1 - \chi_\rho)^{-1}} = \log\left(1 + \frac{\mu a}{g}\right). \tag{44}$$

For the joint correct-deadline quantile defined in Proposition 5, $Q^{\mathrm{corr}}_{1-p}(C_\tau) > \bar{N}_\rho$ whenever $p < 1 - \delta$.

Proof. Write $R = 1 - \rho$. Use ambient coordinates $x_1 = (u, w) \in \mathbb{R}^2$ and $x_2 = y \in \mathbb{R}$, and let the feasible tangent space be spanned by the orthonormal vectors

$$t_1 = (\sqrt{\rho}, 0, \sqrt{R}), \qquad t_2 = (0, 1, 0).$$

Equivalently, $\mathcal{F} = \{(u, w, y) : \sqrt{R}\, u - \sqrt{\rho}\, y = 0\}$, with initial state zero. Its tangent projector and block-access matrices are

$$P = \begin{pmatrix} \rho & 0 & \sqrt{\rho R} \\ 0 & 1 & 0 \\ \sqrt{\rho R} & 0 & R \end{pmatrix}, \qquad S_1 = \mathrm{diag}(\rho, 1), \qquad S_2 = R. \tag{45}$$

Thus both arms are strictly accessible and the least block eigenvalue is ρ.

For $s \in \{\mu, L\}$, define

$$\psi_s(u) = \begin{cases} c + g(u + a) + \frac{\mu}{2}(u + a)^2, & u \ge -a, \\ c + g(u + a) + \frac{s}{2}(u + a)^2, & u < -a, \end{cases} \qquad f_{1,s}(u, w) = \psi_s(u) + \frac{\mu}{2} w^2. \tag{46}$$

The branches have the same value and derivative at $-a$. The derivative of ψ_s is therefore globally L-Lipschitz and has secant slopes in $[\mu, L]$; hence $f_{1,s}$ is μ-strongly convex and L-smooth. Its unique minimizer and optimum are

$$(u_s^\star, w_s^\star) = (-a - g/s, 0), \qquad f_{1,s}^\star = c - \frac{g^2}{2s}.$$

Let

$$f_2(y) = d + \frac{\lambda}{2} y^2, \qquad d = c - \frac{g^2}{4}\left(\frac{1}{\mu} + \frac{1}{L}\right). \tag{47}$$

Instance 0 uses $f_{1,\mu}$, and instance 1 uses $f_{1,L}$; arm 2 is common. Since $c - g^2/(2\mu) < d < c - g^2/(2L)$, arm 1 is uniquely best in instance 0, arm 2 is uniquely best in instance 1, and both gaps equal Δ. Exact value–gradient pairs satisfy all batch-tail assumptions with zero noise proxies. For the common helper arm, the theorem parameters can be taken as $\mu_2 = L_2 = \lambda$, $\rho_2 = R_2 = R = 1 - \rho$, and $\theta_2^- = \theta_2^+ = 1/(\lambda R)$, so $q_2 = 0$.

We next show that neither adaptive arm selection nor the helper arm can reveal the hidden branch early. The prescribed arm-1 step is $\eta_1 = 2/(\mu\rho + L)$. While $u \geq -a$, both instances have $\psi_s'(u) = g + \mu(u + a)$. For $z = u + a + g/\mu$, the projected arm-1 update therefore obeys

$$z^+ = z - \eta_1 \rho\mu z = \chi_\rho z, \qquad z_0 = a + g/\mu. \tag{48}$$

For arm 2, the prescribed step is $\eta_2 = 1/(\lambda R)$. Feasibility gives $y = \sqrt{R/\rho}\, u$, so Eq. (45) implies

$$u^+ = u - \eta_2 \sqrt{\rho R}\, \lambda y = 0, \qquad y^+ = y - \eta_2 R\lambda y = 0.$$

Thus an arm-2 update resets z to z_0; it cannot accelerate motion toward the hidden half-line. The coordinate w stays zero.

After any m completed updates, an elementary induction over contractions and resets gives $z_m \geq z_0\chi_\rho^m$. The boundary $u = -a$ corresponds to $z = g/\mu$, and the definition of $\bar{N}_\rho$ gives $z_m \geq g/\mu$ for every $m \leq \bar{N}_\rho$. One completed macro-update costs at least one primitive pair, while extra frozen replicates create no additional transition and all exact pairs at a common-region state coincide. Hence every history that can stop by primitive cost $\bar{N}_\rho$ stays in $u \geq -a$, where the two objectives have identical values and full gradients. Coupling the policy's internal randomness proves equality of the two stopped transcript laws on such events.

Let $\mathcal{A} = \{C_\tau \leq \bar{N}_\rho\}$, and denote this common restricted law by $\mathbb{Q}$. Uniform δ-correctness and the opposite best arms give

$$\mathbb{Q}(\mathcal{A}, \widehat{J}_\tau = 1) \leq \delta, \qquad \mathbb{Q}(\mathcal{A}, \widehat{J}_\tau = 2) \leq \delta.$$

Because $K = 2$, $\mathbb{Q}(\mathcal{A}) \leq 2\delta$, proving the first inequality in Eq. (42). On either instance, its correct recommendation on $\mathcal{A}$ is the other instance's error event, so the second inequality follows as well. Finally $\mathbb{E}[C_\tau] \geq \bar{N}_\rho \Pr(C_\tau > \bar{N}_\rho)$, which gives the expected bound. Since the joint correct-stop probability through $\bar{N}_\rho$ is at most δ, the quantile assertion follows whenever $1 - p > \delta$.

It remains to expose the geometry. Since

$$-\log\chi_\rho = \log\left(1 + \frac{2\mu\rho}{L - \mu\rho}\right) \leq \frac{2\mu\rho}{L - \mu\rho},$$

substitution into Eq. (41) and $\lfloor x \rfloor \geq x - 1$ prove Eq. (43).

For comparison with the achievability constants, arm 1 in this construction has $R_1 = 1$ and hence

$$1 + \frac{1}{1 - \chi_\rho} = \frac{3}{2} + \frac{L}{2\mu\rho}.$$

Also $-\log\chi_\rho \sim 2\mu\rho/L$, so the product of this quantity and the preceding display tends to one. Replacing the unrounded definition of $\bar{N}_\rho$ by its floor changes the ratio in Eq. (44) by $o(1)$, proving that equation. Thus the lower bound matches this arm's exact-step contribution to Γ_q at the reciprocal-access scale, including its limiting multiplier. On the other hand, with the natural anchor $x^\circ = 0$,

$$\mathcal{R}_{1,s} = \|S_1^{-1/2}(u_s^\star, w_s^\star)\| = \frac{a + g/s}{\sqrt{\rho}}, \qquad s \in \{\mu, L\}.$$

The family therefore leaves the uniformly controlled class in Eq. (22) as $\rho \downarrow 0$. The construction does not show that the other geometry-dependent factors in the state envelope or in h_δ are necessary, and it is not an instance-wise lower bound at fixed transformed anchor radius. □

Lemma 4 (A low-access common jet forces radius–gap tradeoff)**.** *Let $S \succ 0$, and let $f^{(0)}, f^{(1)}$ be μ-strongly convex and L-smooth, where $0 < \mu < L$. Suppose that at some x, for $q \in \{0, 1\}$,*

$$f^{(q)}(x) = c, \qquad \nabla f^{(q)}(x) = g \neq 0, \qquad \|S^{-1/2}(x - x_q^\star)\| \leq D.$$

Define $\rho_{\mathrm{eff}}(g) = g^\top S g/\|g\|^2$. Then

$$\left|(f^{(1)})^\star - (f^{(0)})^\star\right| \leq \frac{\|g\|^2}{2}\left(\frac{1}{\mu} - \frac{1}{L}\right) \leq \frac{L(L - \mu)}{2\mu}\,\rho_{\mathrm{eff}}(g) D^2. \tag{49}$$

Consequently, if a common competitor reverses the best arm with gap at least $\Delta > 0$ in each instance, then

$$D^2 \geq \frac{4\mu\Delta}{L(L - \mu)\rho_{\mathrm{eff}}(g)}. \tag{50}$$

Proof. Strong convexity and one smoothness descent step put both optimum values in $[c - \|g\|^2/(2\mu), c - \|g\|^2/(2L)]$, proving the first inequality. Cocoercivity and $\nabla f^{(q)}(x_q^\star) = 0$ give

$$\frac{\|g\|^2}{L} \le \langle g, x - x_q^\star \rangle \le \|S^{1/2} g\| \, \|S^{-1/2}(x - x_q^\star)\| \le \|g\| \sqrt{\rho_{\text{eff}}(g)} D.$$

Thus $\|g\| \le LD\sqrt{\rho_{\text{eff}}(g)}$, which proves the second inequality. Opposite gaps of size Δ force the two optimum values to differ by at least 2Δ, proving Eq. (50). The witness of Proposition 6 has g in the ρ-eigenspace, so $\rho_{\text{eff}}(g) = \rho$. □

Corollary 5 (Fixed-radius traversal toward singular access)**.** *Fix $0 < \mu < L$, $\bar{a}, \bar{g}, \lambda > 0$, and $\delta \in (0, 1/2)$. In the construction of Proposition 6, for every $0 < \rho \le 1/2$ set*

$$a_\rho = \bar{a}\sqrt{\rho}, \qquad g_\rho = \bar{g}\sqrt{\rho}, \qquad \Delta_\rho = \frac{\bar{g}^2 \rho}{4}\left(\frac{1}{\mu} - \frac{1}{L}\right).$$

Then curvature, noise, initial-state, and transformed-anchor bounds are uniform in ρ, while both instances have gap $\Delta_\rho = \Theta(\rho)$. With

$$\chi_\rho = \frac{L - \mu\rho}{L + \mu\rho}, \qquad \bar{N}_\rho = \left\lfloor \frac{\log(1 + \mu\bar{a}/\bar{g})}{-\log \chi_\rho} \right\rfloor,$$

their transcript laws agree through primitive cost $\bar{N}_\rho$, and every uniformly δ-correct rule obeys

$$\mathbb{E}[C_\tau] \ge (1 - 2\delta)\bar{N}_\rho, \qquad \lim_{\rho \downarrow 0} \rho \bar{N}_\rho = \frac{L}{2\mu} \log\left(1 + \frac{\mu\bar{a}}{\bar{g}}\right).$$

Thus $\bar{N}_\rho = \Theta(\rho^{-1}) = \Theta(\Delta_\rho^{-1})$ at fixed transformed radius. Transcript agreement in fact holds for every common application-fixed, pre-batch, batch-size-invariant radial specification in the paired instances whose realized gains $\theta_{i,t}$ (the gain θ_t of an arm-i round) satisfy

$$0 \le \theta_{1,t} \le \frac{2}{\mu\rho + L}, \qquad 0 \le \theta_{2,t} \le \frac{1}{\lambda(1 - \rho)}.$$

This bound concerns a family with shrinking gap; it is neither an additional cost at a fixed gap nor a product with the statistical Δ^{-2} term.

Proof. For arm 1's tail curvature $s \in \{\mu, L\}$, the physical minimizer and transformed anchor radius at $x^\circ = 0$ are

$$u_{s,\rho}^\star = -\sqrt{\rho}\left(\bar{a} + \frac{\bar{g}}{s}\right), \qquad \|S_1^{-1/2}(u_{s,\rho}^\star, 0)\| = \bar{a} + \frac{\bar{g}}{s}.$$

Arm 2 has radius zero. Substitution in Eq. (41) gives the displayed gap, $\bar{N}_\rho$, and the same transcript argument. Moreover

$$\frac{L - \mu\rho}{2\mu\rho} \log\left(1 + \frac{\mu\bar{a}}{\bar{g}}\right) - 1 \le \bar{N}_\rho \le \frac{L + \mu\rho}{2\mu\rho} \log\left(1 + \frac{\mu\bar{a}}{\bar{g}}\right),$$

which proves the limit and order.

For the stated radial robustness, until the hidden boundary the shifted residual $z = u + a_\rho + g_\rho/\mu$ obeys $z^+ = (1 - \mu\rho\theta_{1,t})z \ge \chi_\rho z$. Feasibility makes an arm-2 update $u^+ = (1 - \lambda(1 - \rho)\theta_{2,t})u$; its coefficient lies in $[0, 1]$, so it moves $u \in [-a_\rho, 0]$ toward zero and cannot accelerate boundary entry. Before that entry the two instances expose identical values and gradients, hence the common response realizes identical gains. Frozen batches cannot increase the number of completed transitions per primitive call. The original coupling and two-error argument therefore proves the same deadline and expectation bounds. Here ρ and Δ_ρ vanish; under the exact step, $1 - \chi_\rho = \Theta(\rho)$. □

Proposition 7 (K-dependent access-geometry cost)**.** *Fix the parameters and $\bar{N}_\rho, \Delta$ of Eq.* (41)*, let $K \ge 2$, and fix $\delta \in (0, 1/2)$. There are one noiseless K-arm base instance and $K - 1$ alternatives with the same affine set, initial state, public oracle interface and noise law, and response/update maps. The oracle kernels are identical until the corresponding challenger's hidden boundary. Arm*

K is uniquely best in the base, challenger j is uniquely best in alternative j, and the best–runner-up gap is Δ in every instance. Every rule uniformly δ-correct on these instances satisfies on the base

$$\mathbb{E}_0 C_\tau \geq (K-1)(1-2\delta)\bar{N}_\rho = \Omega\left(\frac{(K-1)L}{\mu\rho}\right) \tag{51}$$

as $\rho \downarrow 0$, for fixed a, g, μ, L. The bound holds for every common application-fixed, pre-batch, batch-size-invariant radial response whose gains are nonnegative and obey the safe upper endpoints. Under $a_\rho = \bar{a}\sqrt{\rho}$, $g_\rho = \bar{g}\sqrt{\rho}$, every per-arm transformed radius is bounded in ρ, $\Delta_\rho = \Theta(\rho)$, and

$$\mathbb{E}_0 C_\tau = \Omega((K-1)/\rho) = \Omega((K-1)/\Delta_\rho).$$

When K varies, the aggregate transformed-anchor control grows as $O(K)$.

Proof. Use challengers $j = 1, \ldots, K-1$, each with $(u_j, w_j) \in \mathbb{R}^2$, and a common arm K with $y \in \mathbb{R}^{K-1}$. Impose

$$\sqrt{1-\rho}\, u_j - \sqrt{\rho}\, y_j = 0, \qquad j = 1, \ldots, K-1,$$

and initialize at zero. The tangent projector is the orthogonal direct sum of the (u_j, y_j) projectors in Eq. (45), together with the unconstrained w_j directions. Hence

$$S_j = \mathrm{diag}(\rho, 1), \qquad S_K = (1-\rho) I_{K-1}.$$

For $s \in \{\mu, L\}$, let

$$f_{j,s}(u_j, w_j) = \psi_s(u_j) + \frac{\mu}{2} w_j^2,$$

where ψ_s is Eq. (46), and set

$$f_K(y) = d + \frac{\lambda}{2}\|y\|^2, \qquad d = c - \frac{g^2}{4}\left(\frac{1}{\mu} + \frac{1}{L}\right).$$

The base uses $s = L$ for every challenger; alternative j changes only its j-th challenger to $s = \mu$. Thus the base best arm is K; alternative j's best arm is j; the best–runner-up gap is Δ, while its remaining challenger gaps are 2Δ. The public parameters are

$$(\mu_j, L_j, \rho_j, R_j) = (\mu, L, \rho, 1), \qquad (\mu_K, L_K, \rho_K, R_K) = (\lambda, \lambda, 1-\rho, 1-\rho).$$

Before challenger j's hidden boundary, its shifted residual satisfies

$$z_j^+ = (1 - \mu\rho\theta_{j,t}) z_j \geq \chi_\rho z_j$$

for every $0 \leq \theta_{j,t} \leq 2/(\mu\rho + L)$. Other challengers leave u_j fixed. A common-arm update with $0 \leq \theta_{K,t} \leq 1/\{\lambda(1-\rho)\}$ moves $u_j \in [-a, 0]$ toward zero, and at the upper endpoint resets it to zero. Therefore no policy reveals challenger j's hidden tail in its first $\bar{N}_\rho$ macro-queries.

Let $N_j(\tau)$ be the number of challenger-j macro-queries and $\mathcal{A}_j = \{N_j(\tau) \leq \bar{N}_\rho\}$. The stopped transcript restricted to $\mathcal{A}_j$ has the same law under the base and alternative j. Hence

$$\Pr_0(\mathcal{A}_j, \widehat{J}_\tau = K) = \Pr_j(\mathcal{A}_j, \widehat{J}_\tau = K) \leq \delta.$$

Since $\Pr_0(\widehat{J}_\tau \neq K) \leq \delta$, $\Pr_0(\mathcal{A}_j) \leq 2\delta$, and thus $\mathbb{E}_0 N_j(\tau) \geq (1-2\delta)\bar{N}_\rho$. Every macro-query costs at least one primitive pair, so summing over the disjoint challenger counts proves Eq. (51).

Finally, under $a_\rho = \bar{a}\sqrt{\rho}$, $g_\rho = \bar{g}\sqrt{\rho}$, challenger tail s has transformed radius $\bar{a} + \bar{g}/s$, whereas the common arm has radius zero. Corollary 5 gives $\Delta_\rho = \Theta(\rho)$ and $\bar{N}_\rho = \Theta(\rho^{-1})$. There are $K-1$ bounded challenger radii, so their aggregate is $O(K)$. □

A.4.3 Proof of Corollary 2

Proposition 8 (Constant-gain reciprocal-progress direct sum)**.** *Fix $K \geq 2$ and $\delta \in (0, 1/2)$. For every $j < K$, fix*

$$0 < \mu_j < L_j, \qquad 0 < \rho_j \leq \tfrac{1}{2}, \qquad \alpha_j, g_j > 0, \qquad 0 < \theta_j \leq \frac{2}{\mu_j\rho_j + L_j},$$

where θ_j is an application-prescribed, time-constant effective identity-response gain. Put

$$q_j = 1 - \mu_j \rho_j \theta_j, \qquad \bar{N}_j = \left\lfloor \frac{\log(1 + \mu_j \alpha_j / g_j)}{-\log q_j} \right\rfloor, \qquad \Delta_j = \frac{g_j^2}{4} \left(\frac{1}{\mu_j} - \frac{1}{L_j} \right). \tag{52}$$

There are one noiseless K-arm base and $K-1$ alternatives with the same affine set, initial state, public oracle interface, allowed frozen batches, and transition specification. Every block is positively accessible; the helper is uniquely best on the base, challenger j is uniquely best on alternative j, the base gap is $\min_{j<K} \Delta_j$, and alternative j's gap is Δ_j. Every rule uniformly δ-correct on this finite class satisfies on the base

$$\mathbb{E}_0 C_\tau \geq (1 - 2\delta) \sum_{j=1}^{K-1} \bar{N}_j. \tag{53}$$

Conversely, a deterministic procedure identifies the instance using at most

$$\sum_{j=1}^{K-1} (\bar{N}_j + 2) \tag{54}$$

primitive calls. Along any family on which $\varepsilon_j := \mu_j \rho_j \theta_j \downarrow 0$ while $\alpha_j, g_j, \mu_j, L_j$ remain fixed and positive,

$$\frac{\bar{N}_j}{1 + (1 - q_j)^{-1}} \longrightarrow \log \left(1 + \frac{\mu_j \alpha_j}{g_j} \right). \tag{55}$$

Thus the reciprocal progress summands $1 + (\mu_j \rho_j \theta_j)^{-1}$ are finite-class minimax-tight along this fixed-witness degeneration. The dependence is linear in K for homogeneous challengers, or uniformly controlled challengers whose effective progress parameters vanish at a common order.

Proof. Give challenger j coordinates $x_j = (u_j, w_j) \in \mathbb{R}^2$ and the helper coordinates $y = (y_1, \ldots, y_{K-1})$. Impose

$$\sqrt{1 - \rho_j}\, u_j - \sqrt{\rho_j}\, y_j = 0, \qquad j < K,$$

and initialize every coordinate at zero. The tangent projector is the orthogonal direct sum of the two-coordinate projectors on (u_j, y_j), with the w_j directions free. Consequently

$$S_j = \operatorname{diag}(\rho_j, 1), \qquad S_K = \operatorname{diag}(1 - \rho_1, \ldots, 1 - \rho_{K-1}).$$

For $s \in \{\mu_j, L_j\}$, define

$$\psi_{j,s}(u) = \begin{cases} c_j + g_j(u + \alpha_j) + \dfrac{\mu_j}{2}(u + \alpha_j)^2, & u \geq -\alpha_j, \\ c_j + g_j(u + \alpha_j) + \dfrac{s}{2}(u + \alpha_j)^2, & u < -\alpha_j, \end{cases}$$

$$f_{j,s}(u, w) = \psi_{j,s}(u) + \frac{\mu_j}{2} w^2, \qquad c_j = \frac{g_j^2}{4} \left(\frac{1}{\mu_j} + \frac{1}{L_j} \right).$$

The matched value and derivative at $-\alpha_j$, together with derivative secants in $[\mu_j, L_j]$, show that $f_{j,s}$ is μ_j-strongly convex and L_j-smooth. Its minimizer and optimum are

$$(u^\star_{j,s}, w^\star_{j,s}) = \left(-\alpha_j - \frac{g_j}{s}, 0 \right), \qquad f^\star_{j,s} = c_j - \frac{g_j^2}{2s}.$$

Hence $f^\star_{j,L_j} = \Delta_j$ and $f^\star_{j,\mu_j} = -\Delta_j$. Let the common helper be $f_K(y) = \|y\|^2/2$. The base uses $s = L_j$ for every challenger; alternative j changes only its j-th challenger to $s = \mu_j$. This proves the best-arm and gap claims.

Challenger j uses identity response with the prescribed constant step θ_j. The helper uses identity response with step one. For the helper, $\mu_K = L_K = 1$, $\rho_K = \min_j(1-\rho_j)$, and $R_K = \max_j(1-\rho_j)$; since $1 \leq 2/(\rho_K + R_K)$, this step obeys the same safe sector as the upper theorem. All value–gradient pairs are exact and every positive frozen batch size is allowed.

Until challenger j enters $u_j < -\alpha_j$, write $z_j = u_j + \alpha_j + g_j/\mu_j$. Its projected update obeys

$$z_j^+ = (1 - \mu_j \rho_j \theta_j) z_j = q_j z_j.$$

The safe-step condition implies $q_j \in (0,1)$. Other challengers leave u_j fixed. Feasibility gives $y_j = \sqrt{(1-\rho_j)/\rho_j}\, u_j$, so a helper update satisfies

$$u_j^+ = u_j - \sqrt{\rho_j(1-\rho_j)}\, y_j = \rho_j u_j.$$

While $u_j \in [-\alpha_j, 0]$, this moves u_j toward zero and cannot accelerate entry into the hidden half-line. Thus after at most m challenger-j updates, interspersed with arbitrary actions,

$$z_j \geq \left(\alpha_j + \frac{g_j}{\mu_j}\right) q_j^m.$$

The boundary has $z_j = g_j/\mu_j$, and the definition of $\bar{N}_j$ shows that every one of the first $\bar{N}_j + 1$ challenger query locations is on the common branch. The base and alternative j therefore expose identical exact values and full gradients at all those locations; extra frozen replicates add cost but no transition.

Let $N_j(\tau)$ be the number of challenger-j macro-queries and $\mathcal{A}_j = \{N_j(\tau) \leq \bar{N}_j\}$. The stopped transcript restricted to $\mathcal{A}_j$ has the same law under the base and alternative j. Formally, pad a stopped history by a common cemetery symbol, establish equality on finite raw pair prefixes by induction, and pass to their increasing union. Uniform δ-correctness gives

$$\Pr_0(\mathcal{A}_j, \widehat{J}_\tau = K) = \Pr_j(\mathcal{A}_j, \widehat{J}_\tau = K) \leq \delta, \qquad \Pr_0(\mathcal{A}_j, \widehat{J}_\tau \neq K) \leq \delta.$$

Hence $\Pr_0(\mathcal{A}_j) \leq 2\delta$ and $\mathbb{E}_0 N_j(\tau) \geq (1-2\delta)\bar{N}_j$. The challenger counts are disjoint and every macro costs at least one primitive pair, proving Eq. (53).

For the upper bound, make $\bar{N}_j + 2$ consecutive singleton queries to each challenger. After $\bar{N}_j + 1$ completed updates, the final query lies strictly in $u_j < -\alpha_j$, including when the unrounded ratio in Eq. (52) is an integer. The two hidden curvatures then expose different exact full gradients. Detect the unique changed challenger if present, and otherwise recommend the helper. This error-free rule proves Eq. (54).

Finally, $-\log(1-\varepsilon_j) \sim \varepsilon_j$; the floor changes the ratio by $o(1)$. Since $1 + (1-q_j)^{-1} = 1 + \varepsilon_j^{-1}$, this proves Eq. (55) and the stated scoped optimality consequence. □

A.4.4 Proof of Theorem 3 and Corollary 3

Proposition 9 (Same-instance additive access–statistics lower bound)**.** *Fix*

$$K \geq 2, \quad 0 < \mu < L, \quad 0 < \rho \leq \tfrac{1}{2}, \quad \sigma > 0, \quad n \geq 1, \quad \lambda > 0,$$
$$\delta \in (0, \tfrac{1}{2}), \qquad 0 < \theta \leq \frac{2}{\mu\rho + L}.$$

Let

$$\kappa = \frac{\mu + L}{2}, \qquad \chi = 1 - \kappa\rho\theta \in (0,1),$$

and put $a = \lambda(\chi^{-n} - 1)$, $w = (1-\chi)\lambda$, and $M_2 = 10\sqrt{3}/3$. For every $0 < \Delta \leq (L-\mu)w^2/(2M_2)$, there are one base and $K-1$ alternative K-arm instances satisfying the main oracle protocol, with the same affine set, initial state, identity-response transitions with application-prescribed, time-constant challenger gain θ, allowed frozen batches, and Gaussian value-noise law $\mathcal{N}(0,\sigma^2)$. Gradients are exact, the least block-access eigenvalue is ρ, and every instance has best–runner-up gap Δ. Every rule uniformly δ-correct on this finite class satisfies on the base

$$\mathbb{E}_0 C_\tau \geq (K-1)\left\{(1-2\delta)(n+1) + \frac{\sigma^2}{2\Delta^2}\,\mathrm{kl}(1-\delta,\delta)\right\}. \tag{56}$$

Conversely, on this declared family there is a δ-correct procedure with deterministic cost at most

$$(K-1)\left\{n + 1 + \left\lceil \frac{2\sigma^2}{\Delta^2}\log\frac{K-1}{\delta}\right\rceil\right\}. \tag{57}$$

For fixed $\xi, \Lambda > 0$, along any parameter sequence with $\varepsilon := \mu\rho\theta \downarrow 0$, take

$$n_{\rho,\theta} = \left\lceil \frac{\xi}{\varepsilon} \right\rceil, \qquad \lambda_\rho = \Lambda\sqrt{\rho}, \qquad \Delta_{\rho,\theta} = \alpha(1-\chi)^2\lambda_\rho^2, \qquad 0 < \alpha \le \frac{L-\mu}{2M_2}.$$

Then every per-arm transformed anchor radius is bounded uniformly along the sequence,

$$\Delta_{\rho,\theta} = \alpha\Lambda^2\kappa^2\rho^3\theta^2 = \Theta(\rho\varepsilon^2),$$

and Eqs. (56)– (57) *match, for fixed K, any fixed $\delta_0 < 1/2$, and $0 < \delta \le \delta_0$, at order*

$$(K-1)\left\{(\mu\rho\theta)^{-1} + \sigma^2\Delta_{\rho,\theta}^{-2}\log(1/\delta)\right\}$$

up to constants and finite-K confidence allocation.

Proof. The gain restriction and $\rho < 1$ give

$$0 < \kappa\rho\theta \le \frac{\rho(\mu+L)}{\mu\rho+L} < 1,$$

so the stated χ lies in $(0,1)$. Define the C^2 plateau

$$\Psi(s) = \begin{cases} 0, & s \ge 0, \\ 10(-s)^3 - 15(-s)^4 + 6(-s)^5, & -1 < s < 0, \\ 1, & s \le -1. \end{cases}$$

Its endpoint values and first two derivatives match, and direct differentiation gives $\sup_s |\Psi'(s)| = 15/8$ and $\sup_s |\Psi''(s)| = M_2$. Put $u_0 = -a - \lambda = -\lambda\chi^{-n}$, and for $s \in \{+1, -1\}$ define

$$\phi_s(u,v) = c + \frac{\kappa}{2}\{(u-u_0)^2 + v^2\} + s\Delta\Psi\left(\frac{u+a}{w}\right). \tag{58}$$

The perturbation changes only the u-curvature, by at most $\Delta M_2/w^2 \le (L-\mu)/2$. Hence every ϕ_s is μ-strongly convex and L-smooth. Since $u_0 < -a - w$, the plateau is constant at the common minimizer and

$$\phi_s^\star = c + s\Delta.$$

For challengers $j = 1, \dots, K-1$, let $x_j = (u_j, v_j)$; let arm K have $y = (y_1, \dots, y_{K-1})$. Impose

$$\sqrt{1-\rho}\, u_j - \sqrt{\rho}\, y_j = 0, \qquad j = 1, \dots, K-1,$$

and initialize all coordinates at zero. The tangent projector is the orthogonal direct sum of the two-coordinate projectors on (u_j, y_j), with each v_j free. Therefore

$$S_j = \mathrm{diag}(\rho, 1), \qquad S_K = (1-\rho)I_{K-1}.$$

Challengers use Eq. (58), the identity response, and the prescribed constant step θ. The helper objective and step are

$$f_K(y) = c + \frac{\kappa}{2}\|y\|^2, \qquad \eta_K = \frac{1}{\kappa(1-\rho)}.$$

Thus the challenger step is inside the safe sector for public controls $(\mu, L, \rho, R = 1)$; the helper has $(\mu_K, L_K, \rho_K, R_K) = (\kappa, \kappa, 1-\rho, 1-\rho)$, and a helper update resets every (u_j, y_j) pair to zero. Each challenger primitive returns its exact gradient and an independent value perturbed by $\mathcal{N}(0, \sigma^2)$; give the helper the same independent value-noise law and an exact gradient. All positive frozen batch sizes are allowed.

We next verify the gate alignment. Starting from a reset, write $z = u_j - u_0 = u_j + a + \lambda$. At every common-side point $u_j \ge -a$, both signs have u-gradient κz, so the projected update gives

$$z^+ = \chi z, \qquad z_m = \lambda\chi^{m-n}$$

after m consecutive challenger updates. For $m = 0, \dots, n$, $z_m \ge \lambda$, equivalently $u_j \ge -a$, and the query sees $\Psi = 0$. The update following the $(n+1)$-st such query lands at

$$z_{n+1} = \chi\lambda = \lambda - w, \qquad u_{j,n+1} = -a - w,$$

where $\Psi = 1$ and $\Psi' = 0$. All later challenger queries before a reset remain on that saturated side. Other challengers leave u_j fixed, the helper resets it, and frozen replication adds no intervening transition. Thus no reachable query lies in the transition layer. The two signs expose identical values and full gradients for the first $n+1$ challenger macro-queries after a reset; at every saturated query their gradients remain identical and their values differ by exactly 2Δ.

Let the base instance use ϕ_+ for every challenger, and let alternative j change only challenger j to ϕ_-. The helper is uniquely best on the base; challenger j is uniquely best on alternative j, with gap Δ in every instance. Let C_j^{com} count primitive pairs drawn from challenger j at common-side states, C_j^{sat} those drawn at saturated states, and $\mathcal{A}_j = \{C_j^{\mathrm{sat}} = 0\}$.

The stopped transcript restricted to $\mathcal{A}_j$ has the same law under the base and alternative j. Formally, pad a stopped history with a common cemetery symbol, restrict first to histories stopping within m raw pairs, and induct over these finite prefixes. On $\mathcal{A}_j$, the changed objective has not produced an instance-dependent response, so every state and conditional kernel agrees. Summing the prefixes and sending $m \to \infty$ proves the restricted stopped-law equality. Uniform δ-correctness then gives

$$\Pr_0(\mathcal{A}_j, \widehat{J}_\tau = K) = \Pr_j(\mathcal{A}_j, \widehat{J}_\tau = K) \le \delta, \qquad \Pr_0(\mathcal{A}_j, \widehat{J}_\tau \neq K) \le \delta.$$

Hence $\Pr_0(\mathcal{A}_j) \le 2\delta$. Reaching a saturated query requires the $n+1$ common macro-queries above, each costing at least one primitive pair; therefore

$$\mathbb{E}_0 C_j^{\mathrm{com}} \ge (1 - 2\delta)(n+1). \tag{59}$$

At a saturated challenger-j query, where the unperturbed part of the objective has a common value $\bar{\phi}$, the conditional KL per raw pair is

$$\mathrm{KL}\big(\mathcal{N}(\bar{\phi} + \Delta, \sigma^2) \,\|\, \mathcal{N}(\bar{\phi} - \Delta, \sigma^2)\big) = \frac{2\Delta^2}{\sigma^2};$$

all other kernels agree. If $\mathbb{E}_0 C_j^{\mathrm{sat}} = \infty$, the desired result is immediate. Otherwise truncate the raw-pair transcript, apply the adaptive chain rule (batch actions are fixed before their draws), and pass to the stopped transcript by monotone convergence and lower semicontinuity. Data processing through $\{\widehat{J}_\tau = K\}$ yields

$$\frac{2\Delta^2}{\sigma^2} \mathbb{E}_0 C_j^{\mathrm{sat}} \ge \mathrm{kl}(1 - \delta, \delta). \tag{60}$$

The C_j^{com} and C_j^{sat} pairs are disjoint, as are counts belonging to different challengers. Summing Eqs. (59) and (60) proves Eq. (56).

For the upper bound, query each challenger consecutively with $n+1$ singleton macros. Its next state is saturated. Freeze there and draw

$$B_\Delta = \left\lceil \frac{2\sigma^2}{\Delta^2} \log \frac{K-1}{\delta} \right\rceil$$

fresh pairs. The common quadratic part is known, so thresholding the batch mean at its midpoint classifies the sign with error at most $\exp\{-B_\Delta \Delta^2/(2\sigma^2)\}$. A union bound proves δ-correctness and Eq. (57).

Finally, under the displayed scaling, with $\varepsilon = \mu\rho\theta$,

$$\|S_j^{-1/2}(u_0, 0)\| = \Lambda \chi^{-n_{\rho,\theta}} \longrightarrow \Lambda \exp\left\{\frac{\kappa\xi}{\mu}\right\},$$

whereas the helper radius is zero. Moreover

$$1 - \chi = \kappa\rho\theta = \frac{\kappa}{\mu}\varepsilon, \qquad w = \kappa\Lambda\rho^{3/2}\theta, \qquad \Delta_{\rho,\theta} = \alpha\Lambda^2\kappa^2\rho^3\theta^2 = \Theta(\rho\varepsilon^2).$$

Substitution into the two cost bounds proves the final assertion. □

A.4.5 SINGULAR ACCESS

Proposition 10 (Singular block access prevents uniform identification)**.** *Fix $K \geq 2$, a nonempty affine feasible set $\mathcal{F} = \{x : Ax = b\}$, and $P = \mathrm{Proj}_{\ker A}$. Suppose that $S_i = E_i^\top P E_i$ is singular for some arm i. For every pair of public constants $0 < \mu < L < \infty$ and every $\delta \in (0, 1/2)$, there are two instances with the same A, b, initial feasible state, public (μ, L) bounds, transition specification, and oracle-noise law such that: every objective is μ-strongly convex and L-smooth with a unique minimizer; the value and full-gradient oracle laws agree at every feasible state; but the two instances have different unique best arms. Consequently, no rule is uniformly δ-correct over this class.*

The equality of transcript laws holds under arbitrary adaptive arm choices and allowed frozen-state batch sizes and under every common predictable projected transition of Assumption 1. More quantitatively, for any rule that terminates almost surely on both instances,

$$\max\left\{\Pr_0(\widehat{J}_\tau \neq i_0^\star), \Pr_1(\widehat{J}_\tau \neq i_1^\star)\right\} \geq \frac{1}{2}. \tag{61}$$

Proof. Fix $x^\circ \in \mathcal{F}$. Since S_i is positive semidefinite and singular, choose a unit vector $v \in \ker(S_i)$. Then

$$0 = v^\top S_i v = \|PE_i v\|^2, \qquad \text{so} \qquad PE_i v = 0.$$

For every $x \in \mathcal{F}$, the difference $x - x^\circ$ lies in $\ker A = \mathrm{range}(P)$, and hence

$$v^\top (x_i - x_i^\circ) = \langle E_i v, x - x^\circ \rangle = \langle PE_i v, x - x^\circ \rangle = 0. \tag{62}$$

Thus the v-coordinate of block i is invariant over the entire feasible set, not only along a particular query path.

Let $\alpha = (\mu + L)/2$, choose $\lambda > 0$ and $c \in \mathbb{R}$, and, for $s \in \{\mu, L\}$, define

$$f_{i,s}(z) = c + \frac{\alpha}{2}\left\|(I - vv^\top)(z - x_i^\circ)\right\|^2 + \frac{s}{2}\left(v^\top (z - x_i^\circ)\right)^2 + \lambda v^\top (z - x_i^\circ). \tag{63}$$

Its Hessian is $\alpha(I - vv^\top) + svv^\top$, so both choices are μ-strongly convex and L-smooth. They have the unique minimizers $x_i^\circ - (\lambda/s)v$ and standalone optimum values

$$f_{i,s}^\star = c - \frac{\lambda^2}{2s}. \tag{64}$$

On the other hand, Eq. (62) implies that, at every feasible block value $z = x_i$,

$$f_{i,\mu}(z) = f_{i,L}(z), \qquad \nabla f_{i,\mu}(z) = \nabla f_{i,L}(z) = \alpha(z - x_i^\circ) + \lambda v. \tag{65}$$

In particular, observing the full gradient reveals the common inaccessible slope λv, but not the inaccessible curvature s that determines the decrease to the standalone optimum.

Choose $j \neq i$, and set

$$m = c - \frac{\lambda^2}{4}\left(\frac{1}{\mu} + \frac{1}{L}\right), \qquad f_j(z) = m + \frac{\alpha}{2}\|z - x_j^\circ\|^2.$$

If there are further arms, give each of them the objective $c + \lambda^2 + (\alpha/2)\|z - x_k^\circ\|^2$. All these objectives satisfy the same public (μ, L) bounds and are unchanged between the two instances. Because $\mu < L$,

$$c - \frac{\lambda^2}{2\mu} < m < c - \frac{\lambda^2}{2L}.$$

Let instance 0 use $f_{i,\mu}$, and let instance 1 use $f_{i,L}$. Arm i is the unique best arm in instance 0, whereas arm j is the unique best arm in instance 1.

It remains to verify that adaptivity, batching, and state motion cannot distinguish the instances. First take noiseless value–gradient oracles. Couple the policy's internal randomization in the two instances. Initially the histories and feasible states agree. Conditional on equal histories, the policy makes the same arm and batch request, every primitive pair in that batch agrees by Eq. (65), and the common predictable transition applies the same response to the same averaged gradient. Projection therefore produces the same next feasible state. Induction gives pathwise equality of the complete transcripts, including the stopping decision and recommendation. The same argument works after adding any

common, possibly joint and history-dependent, conditional noise kernel: couple equal oracle noise at each primitive draw. Hence the transcript laws are identical for arbitrary adaptive batching and for every common predictable transition allowed by Assumption 1.

Let $\mathbb{Q}$ be their common law of the terminal recommendation for a rule that terminates almost surely on both instances. Since the two best arms i and j are distinct,

$$\mathbb{Q}(\widehat{J}_\tau = i) + \mathbb{Q}(\widehat{J}_\tau = j) \le 1.$$

At least one of these two probabilities is at most $1/2$, which proves Eq. (61). A uniformly δ-correct rule would terminate almost surely and make each of these two probabilities at least $1 - \delta > 1/2$, a contradiction. □

Scope of block accessibility. Positive definiteness of S_i is sufficient for arbitrary standalone minimizers, but is stronger than necessary for a fixed instance. If S_i is singular, the immovable component of every feasible block lies in $\ker(S_i)$; an instance-specific extension can be formulated when $x_i^\star - x_i^\circ \in \mathrm{range}(S_i)$, using the smallest positive eigenvalue on that range. Without such reachability, an arbitrary minimizer displacement cannot be attained by any feasible transition. Proposition 10 shows that this is not only an optimization obstruction: without additional off-feasible structure, singular access makes the best standalone arm statistically nonidentifiable over the full smooth strongly convex class.

A.5 Canonical Couplings

The two propositions below give the projected update for fixed-sum balance and for consensus. Reading off the queried block gives $S_i = (1 - 1/K)I$ and $S_i = I/K$, respectively, the values quoted in Section 3.

Proposition 11 (Projection formula for the balance constraint)**.** *Let*

$$\mathcal{F}_{\mathrm{bal}} = \left\{ x \in (\mathbb{R}^d)^K : \sum_{i=1}^K x_i = b \right\}.$$

Suppose $x_t \in \mathcal{F}_{\mathrm{bal}}$, arm i_t is queried, and the tentative update is

$$\widetilde{x}_{t+1} = x_t - \eta_t E_{i_t} G_t,$$

where E_{i_t} injects the queried gradient block $G_t \in \mathbb{R}^d$ into the i_t-th arm position. Then the Euclidean projection onto $\mathcal{F}_{\mathrm{bal}}$ is blockwise

$$x_{j,t+1} = \widetilde{x}_{j,t+1} - \frac{\sum_{\ell=1}^K \widetilde{x}_{\ell,t+1} - b}{K}, \qquad j \in [K],$$

and therefore

$$x_{i_t,t+1} = x_{i_t,t} - \eta_t \left(1 - \frac{1}{K}\right) G_t,$$

while for every $j \neq i_t$,

$$x_{j,t+1} = x_{j,t} + \frac{\eta_t}{K} G_t.$$

Proof. Since $x_t \in \mathcal{F}_{\mathrm{bal}}$, we have $\sum_{j=1}^K x_{j,t} = b$. Hence

$$\sum_{j=1}^K \widetilde{x}_{j,t+1} = \sum_{j=1}^K x_{j,t} - \eta_t G_t = b - \eta_t G_t.$$

Projection onto the affine hyperplane $\mathcal{F}_{\mathrm{bal}}$ subtracts the average vector-valued constraint violation from every arm block, so

$$x_{j,t+1} = \widetilde{x}_{j,t+1} + \frac{\eta_t}{K} G_t, \qquad j \in [K].$$

Reading off the queried and nonqueried blocks gives the claimed formulas. □

Proposition 12 (Projection formula for consensus coupling)**.** *Let $d \geq 1$ and*

$$\mathcal{F}_{\mathrm{con}} = \{x \in (\mathbb{R}^d)^K : x_1 = \cdots = x_K\}.$$

Suppose

$$x_t = (z_t, \ldots, z_t) \in \mathcal{F}_{\mathrm{con}},$$

arm i_t is queried, and the tentative update is

$$\widetilde{x}_{t+1} = x_t - \eta_t E_{i_t} G_t,$$

where E_{i_t} injects the queried gradient block $G_t \in \mathbb{R}^d$ into the i_t-th arm position. Then the Euclidean projection onto $\mathcal{F}_{\mathrm{con}}$ is

$$x_{j,t+1} = z_t - \frac{\eta_t}{K} G_t, \qquad j \in [K].$$

Equivalently, the feasible state remains consensus-valued and the shared iterate evolves as

$$z_{t+1} = z_t - \frac{\eta_t}{K} G_t.$$

Proof. Projection onto $\mathcal{F}_{\mathrm{con}}$ replaces every arm block by the mean of the tentative blocks. Since

$$\widetilde{x}_{i_t,t+1} = z_t - \eta_t G_t \qquad \text{and} \qquad \widetilde{x}_{j,t+1} = z_t \quad (j \neq i_t),$$

their mean is

$$\frac{1}{K} \sum_{j=1}^{K} \widetilde{x}_{j,t+1} = z_t - \frac{\eta_t}{K} G_t.$$

Assigning this mean to every block gives the claimed formula. □

A.6 Known-Hessian Quadratics

This subsection proves Theorem 4 and Theorem 6. The key identity is the exact certificate for quadratics.

A.6.1 Transformed observations, cross-fitting, and confidence sequences

Input: δ, feasible initial state, known H_i, a selected one-pair or cross-fitted transform, application-supplied projector/transition, and observable $(v_i, \kappa_i)_{i=1}^K$ validating the conditional sub-gamma condition of Corollary 6. A *query* of arm i draws one transformed sample and applies the prescribed transition. With N_i the current number of samples of arm i, write $\underline{V}_i = \overline{Y}_{i,N_i} - \beta_{i,N_i}(\delta)$ and $\overline{V}_i = \overline{Y}_{i,N_i} + \beta_{i,N_i}(\delta)$.

1. Query each arm once; set $h = 0$.
2. Set $\hat{\imath} \in \arg\min_i \overline{V}_i$ and $\check{\imath} \in \arg\min_{i \neq \hat{\imath}} \underline{V}_i$. If $\overline{V}_{\hat{\imath}} < \underline{V}_{\check{\imath}}$, return $\hat{\imath}$.
3. Repeat 2^h times: recompute $\hat{\imath}, \check{\imath}$; query $i \in \arg\max_{j \in \{\hat{\imath}, \check{\imath}\}} \beta_{j,N_j}(\delta)$; recompute $\hat{\imath}, \check{\imath}$; if $\overline{V}_{\hat{\imath}} < \underline{V}_{\check{\imath}}$, return $\hat{\imath}$.
4. If no stop occurred, query every arm once in increasing index order, recomputing $\hat{\imath}, \check{\imath}$ and returning $\hat{\imath}$ after any query for which $\overline{V}_{\hat{\imath}} < \underline{V}_{\check{\imath}}$. Then set $h \leftarrow h + 1$ and return to Step 2.

Algorithm 2: Sweep-LUCB

Proposition 13 (Exact certificate for convex vector quadratics)**.** *For the vector-quadratic arm*

$$f_i(x_i) = \frac{1}{2}(x_i - b_i)^\top H_i (x_i - b_i) + c_i, \qquad H_i \succeq 0,$$

the optimum value satisfies

$$f_i^\star = c_i = f_i(x_i) - \frac{1}{2} \nabla f_i(x_i)^\top H_i^\dagger \nabla f_i(x_i)$$

for every $x_i \in \mathbb{R}^{d_q}$. The identity is pointwise, so it remains true at arbitrary predictable queried states, irrespective of the coupling, projections, or preceding query rule.

Proof. Put $y_i = x_i - b_i$. Since H_i is symmetric positive semidefinite, $H_i H_i^\dagger H_i = H_i$. Therefore

$$f_i(x_i) - \frac{1}{2}\nabla f_i(x_i)^\top H_i^\dagger \nabla f_i(x_i) = \frac{1}{2} y_i^\top H_i y_i + c_i - \frac{1}{2} y_i^\top H_i H_i^\dagger H_i y_i$$
$$= c_i.$$

The quadratic term is nonnegative and vanishes at $x_i = b_i$, so $f_i^\star = c_i$. Positive definiteness is needed only for uniqueness of the state minimizer, not for the value identity. □

Proposition 14 (Variance-corrected transformed observation)**.** *Consider the noisy vector-quadratic model for arm i,*

$$f_i(x_i) = \frac{1}{2}(x_i - b_i)^\top H_i (x_i - b_i) + c_i, \quad H_i \succeq 0,$$
$$F_t = f_i(x_{i,t}) + \zeta_t^f, \qquad G_t = \nabla f_i(x_{i,t}) + \zeta_t^g,$$

where $A_i = H_i^\dagger$, and where the queried arm and state are chosen predictably. Assume only conditional centering and a finite conditional gradient second moment:

$$\mathbb{E}[\zeta_t^f \mid \mathcal{H}_{t-1}, i_t = i] = 0, \qquad \mathbb{E}[\zeta_t^g \mid \mathcal{H}_{t-1}, i_t = i] = 0,$$

$$\mathbb{E}\big[\zeta_t^g (\zeta_t^g)^\top \mid \mathcal{H}_{t-1}, i_t = i\big] = \Sigma_{i,t},$$

where $\Sigma_{i,t}$ may vary predictably with time and history. It is enough that the scalar $\operatorname{tr}(A_i \Sigma_{i,t})$ be known before the response. Define the transformed observation

$$Y_t = F_t - \frac{1}{2} G_t^\top A_i G_t + \frac{1}{2}\operatorname{tr}(A_i \Sigma_{i,t}), \qquad i = i_t.$$

Then

$$\mathbb{E}[Y_t \mid \mathcal{H}_{t-1}, i_t = i] = c_i = f_i^\star.$$

No independence between value and gradient noise, across arms, or across rounds is required. The result holds for arbitrary predictable state trajectories generated by any coupling or projections.

Proof. Condition on $\mathcal{H}_{t-1}$ and $i_t = i$, and write $g_t = \nabla f_i(x_{i,t})$. The predictable quantities are then fixed, and conditional centering gives

$$\mathbb{E}[(g_t + \zeta_t^g)^\top A_i (g_t + \zeta_t^g)] = g_t^\top A_i g_t + \operatorname{tr}(A_i \Sigma_{i,t}).$$

The trace correction cancels the second term, and Proposition 13 gives $f_i(x_{i,t}) - g_t^\top A_i g_t / 2 = c_i = f_i^\star$. □

Covariance-free cross-fitting. When the gradient covariance is unknown, two conditionally independent value–gradient draws at the same frozen state remove the trace correction. The construction needs neither Gaussian noise nor a known covariance.

Lemma 5 (Bilinear sub-gamma bound)**.** *Let $\zeta, \zeta' \in \mathbb{R}^d$ be independent, centered, and σ-sub-Gaussian in every direction:*

$$\mathbb{E}e^{\langle u, \zeta\rangle},\ \mathbb{E}e^{\langle u, \zeta'\rangle} \le e^{\sigma^2 \|u\|^2/2}, \qquad u \in \mathbb{R}^d.$$

For a fixed matrix $A \in \mathbb{R}^{d\times d}$ and $|\lambda| < (\sigma^2 \|A\|_{\mathrm{op}})^{-1}$,

$$\log \mathbb{E}e^{\lambda \zeta^\top A \zeta'} \le \frac{\lambda^2 \sigma^4 \|A\|_F^2}{2\{1 - \sigma^2 \|A\|_{\mathrm{op}} |\lambda|\}}. \tag{66}$$

When $\sigma\|A\|_{\mathrm{op}} = 0$, the domain is all of $\mathbb{R}$. The same statement holds conditionally when independence, centering, and the directional sub-Gaussian bounds hold given a sigma-field and A is measurable with respect to that sigma-field.

Proof. Conditioning first on ζ gives

$$\mathbb{E}[e^{\lambda\zeta^\top A\zeta'} \mid \zeta] \le \exp\left\{\frac{\lambda^2\sigma^2}{2}\|A^\top\zeta\|^2\right\}.$$

For $\eta \ge 0$, introduce an independent $z \sim \mathcal{N}(0, I_d)$. The Gaussian linearization identity and the sub-Gaussian MGF of ζ give

$$\begin{aligned}\mathbb{E}e^{\eta\|A^\top\zeta\|^2} &= \mathbb{E}_z\mathbb{E}_\zeta e^{\sqrt{2\eta}\,z^\top A^\top\zeta} \\ &\le \mathbb{E}_z e^{\eta\sigma^2\|Az\|^2} = \det(I - 2\eta\sigma^2 A^\top A)^{-1/2},\end{aligned}$$

whenever the determinant is finite. With $\eta = \lambda^2\sigma^2/2$, let s_j be the singular values of A and put $r = \sigma^2\|A\|_{\mathrm{op}}|\lambda| < 1$. Then

$$-\frac{1}{2}\sum_j \log(1 - \lambda^2\sigma^4 s_j^2) \le \frac{\lambda^2\sigma^4\|A\|_F^2}{2(1-r^2)} \le \frac{\lambda^2\sigma^4\|A\|_F^2}{2(1-r)},$$

which proves Eq. (66) for either sign of λ. The conditional version follows after conditioning on the given sigma-field. □

For the next result, a *cross-fitted macro-query* of arm i freezes the current state, draws two conditionally independent primitive joint value–gradient pairs at that state, computes the statistic below, and only then performs exactly one application-fixed projected transition using the average of the two gradients. There is no transition between the primitive draws. Arm selection and stopping are performed only between macro-queries.

Theorem 6 (Covariance-free cross-fitted quadratic identification)**.** *Assume $K \ge 2$, $\delta \in (0,1)$, the quadratic functional form of Assumption 5, the exogenous-transition clauses of Assumption 1, and a unique best arm. Replace the known-covariance clause of Assumption 5 by the noise conditions below. Put $A_i = H_i^\dagger$. At every cross-fitted macro-query of arm i, write*

$$F_t^{(1)} = f_i(x_{i,t}) + \zeta_{i,t}^f, \qquad G_t^{(a)} = g_{i,t} + \zeta_{i,t}^{g,(a)}, \quad a \in \{1,2\}, \qquad g_{i,t} = \nabla f_i(x_{i,t}),$$

where, conditionally on the pre-macro-query history, the two gradient noises are independent and centered, and satisfy

$$\mathbb{E}e^{\langle u, \zeta_{i,t}^{g,(a)}\rangle} \le e^{\sigma_{g,i}^2\|u\|^2/2}, \qquad u \in \mathbb{R}^{d_q}.$$

The centered value noise is conditionally $\sigma_{f,i}$-sub-Gaussian. It may depend on $\zeta_{i,t}^{g,(1)}$, and no value–gradient independence is assumed. These moment proxies differ from the tail proxies σ_i^f, σ_i^g of Assumption 4; $\sigma_{g,i}$ bounds every direction and carries no dimension factor. Suppose the public state–center envelope gives the sure predictable bound

$$\|A_i g_{i,t}\| \le D_i.$$

Define

$$Y_t^{\mathrm{cf}} = F_t^{(1)} - \frac{1}{2}(G_t^{(1)})^\top A_i G_t^{(2)}, \tag{67}$$

and the observable constants

$$v_i^{\mathrm{cf}} = 4\sigma_{f,i}^2 + 2\sigma_{g,i}^2 D_i^2 + \sigma_{g,i}^4\|A_i\|_F^2, \qquad \kappa_i^{\mathrm{cf}} = 2\sigma_{g,i}^2\|A_i\|_{\mathrm{op}}. \tag{68}$$

Then, at every predictable coupled state,

$$\mathbb{E}[Y_t^{\mathrm{cf}} \mid \mathcal{H}_{t-1}, i_t = i] = c_i = f_i^\star$$

and, for $|\lambda| < 1/\kappa_i^{\mathrm{cf}}$,

$$\log\mathbb{E}\Big[e^{\lambda(Y_t^{\mathrm{cf}} - c_i)}\Big|\mathcal{H}_{t-1}, i_t = i\Big] \le \frac{\lambda^2 v_i^{\mathrm{cf}}}{2(1 - \kappa_i^{\mathrm{cf}}|\lambda|)}. \tag{69}$$

Run Algorithm 2 at the macro-query level, using Eq. (67) *and substituting $(v_i^{\mathrm{cf}}, \kappa_i^{\mathrm{cf}})$ in the radius $\beta_{i,n}(\delta)$ of Corollary 6. Define γ_i as in Theorem 4,*

$$M_i^{\mathrm{cf}} = \#\{n \ge 1 : 4\beta_{i,n}^{\mathrm{cf}} \ge \gamma_i\}, \qquad M_\delta^{\mathrm{cf}} = \sum_i M_i^{\mathrm{cf}}.$$

The rule is δ-correct and, with probability at least $1-\delta$, returns $i^\star$ by

$$\tau_Y \le K + M_\delta^{\mathrm{cf}} + K\left\lfloor \log_2(M_\delta^{\mathrm{cf}}+1)\right\rfloor. \tag{70}$$

Every macro-query uses exactly two primitive joint pairs, so on the same event

$$C_\tau \le 2\left\{K + M_\delta^{\mathrm{cf}} + K\left\lfloor \log_2(M_\delta^{\mathrm{cf}}+1)\right\rfloor\right\}.$$

Thus neither Gaussian noise nor a known gradient covariance is needed.

Proof. Condition on the history before a macro-query and abbreviate $g = g_{i,t}$, $A = A_i$. Conditional independence and centering give

$$\mathbb{E}[(G_t^{(1)})^\top A G_t^{(2)} \mid \mathcal{H}_{t-1}, i_t = i] = g^\top A g.$$

The Moore–Penrose identity $H_i A_i H_i = H_i$ gives $f_i(x_{i,t}) - g^\top A_i g/2 = c_i$, proving the stationary conditional mean. The centered noise is

$$Z_t = \zeta_{i,t}^f - \frac{1}{2}\langle A_i g, \zeta_{i,t}^{g,(1)}\rangle - \frac{1}{2}\langle A_i g, \zeta_{i,t}^{g,(2)}\rangle - \frac{1}{2}(\zeta_{i,t}^{g,(1)})^\top A_i \zeta_{i,t}^{g,(2)}.$$

Generalized Hölder with four equal exponents is valid even though the value noise may depend on the first gradient noise. The four contributions to the conditional log-MGF are at most

$$2\lambda^2\sigma_{f,i}^2, \qquad \frac{1}{2}\lambda^2\sigma_{g,i}^2 D_i^2, \qquad \frac{1}{2}\lambda^2\sigma_{g,i}^2 D_i^2,$$

and, by Lemma 5,

$$\frac{\lambda^2\sigma_{g,i}^4\|A_i\|_F^2}{2\{1-2\sigma_{g,i}^2\|A_i\|_{\mathrm{op}}|\lambda|\}}.$$

Putting the first three terms over the last denominator proves Eqs. (68)–(69). The query-clock confidence sequence, charging argument, geometric-sweep count, correctness, and almost-sure termination proof of Theorem 4 now apply verbatim in macro-query units. Multiplication by two gives the exact primitive-pair accounting. □

Remark 1 (An intervening update destroys cross-fitted centering)**.** *The same-state clause in Theorem 6 is essential. In one dimension, take $f(x) = x^2/2 + c$, exact gradients, and initial state $x = 1$. If the mandatory update after the first ordinary query moves the state to zero, then $F^{(1)} = c + 1/2$, $G^{(1)} = 1$, and $G^{(2)} = 0$, so Eq.* (67) *equals $c + 1/2$, not c. Two ordinary queries separated by transitions therefore cannot replace one frozen two-pair macro-query.*

Proposition 15 (A simple sufficient condition for transformed sub-Gaussianity)**.** *Consider the setup of Proposition 14. Fix an arm i, write $A_i = H_i^\dagger$, and suppose that whenever arm i is queried,*

$$|\zeta_t^f| \le B_i^f, \qquad \|\zeta_t^g\| \le B_i^g, \qquad \|A_i \nabla f_i(x_{i,t})\| \le D_i$$

almost surely given $\mathcal{H}_{t-1}$ and $i_t = i$. Then the transformed noise $Y_t - f_i^\star$ is conditionally σ_i^Y-sub-Gaussian with

$$\sigma_i^Y = B_i^f + D_i B_i^g + \frac{1}{4}\|A_i\|_{\mathrm{op}}(B_i^g)^2.$$

For $H_i \succ 0$, the last state condition is exactly $\|x_{i,t} - b_i\| \le D_i$.

Proof. Let $g_t = \nabla f_i(x_{i,t})$. The transformed noise is

$$Y_t - f_i^\star = \zeta_t^f - (A_i g_t)^\top \zeta_t^g - \frac{1}{2}(\zeta_t^g)^\top A_i \zeta_t^g + \frac{1}{2}\operatorname{tr}(A_i \Sigma_{i,t}).$$

Because $A_i \succeq 0$, conditionally on the past this centered random variable lies in an interval of length at most

$$2B_i^f + 2D_i B_i^g + \frac{1}{2}\|A_i\|_{\mathrm{op}}(B_i^g)^2.$$

Indeed, the deterministic trace correction only translates the support; the quadratic term ranges from $-\|A_i\|_{\mathrm{op}}(B_i^g)^2/2$ to zero. Hoeffding's lemma and Proposition 14 therefore give

$$\mathbb{E}[\exp(\lambda(Y_t - f_i^\star))|\mathcal{H}_{t-1}, i_t = i] \le \exp\left(\frac{\lambda^2(\sigma_i^Y)^2}{2}\right),$$

because σ_i^Y is one half of the displayed support length. □

Lemma 6 (Query-ordered mgf transfer)**.** *Fix an arm i, let i_t be $\mathcal{H}_{t-1}$-measurable, and write $Z_t = Y_t - f_i^\star$ when $i_t = i$. Suppose that for every λ in an interval $\Lambda \subseteq \mathbb{R}$,*

$$\mathbb{E}[\exp(\lambda Z_t)|\mathcal{H}_{t-1}, i_t = i] \le \exp\big(\psi_{i,t}(\lambda)\big),$$

where $\psi_{i,t}(\lambda)$ is $\mathcal{H}_{t-1}$-measurable. Then

$$\mathcal{M}_{i,t}(\lambda) = \exp\left\{\lambda \sum_{s=1}^{t} \mathbf{1}\{i_s = i\} Z_s - \sum_{s=1}^{t} \mathbf{1}\{i_s = i\}\psi_{i,s}(\lambda)\right\}$$

is a nonnegative $(\mathcal{H}_t)$-supermartingale with initial value one. Consequently, if $\tau_{i,n}$ is attained and $Y_{i,m} = Y_{\tau_{i,m}}$, the same process in query clock is

$$\mathcal{M}_{i,\tau_{i,n}}(\lambda) = \exp\left\{\lambda \sum_{m=1}^{n} (Y_{i,m} - f_i^\star) - \sum_{m=1}^{n} \psi_{i,\tau_{i,m}}(\lambda)\right\}.$$

All maximal inequalities below therefore hold over attained query counts; they do not assume that an arm is queried again.

Proof. If $i_t \neq i$, the multiplicative increment of $\mathcal{M}_{i,t}$ is one. If $i_t = i$, predictability of the arm choice and the assumed conditional MGF bound show that its conditional mean is at most one. Thus $\mathbb{E}[\mathcal{M}_{i,t}(\lambda) \mid \mathcal{H}_{t-1}] \le \mathcal{M}_{i,t-1}(\lambda)$. Evaluating the displayed global-time process at an attained $\tau_{i,n}$ gives the query-clock identity. □

Corollary 6 (Anytime confidence sequence for transformed samples)**.** *Assume that, whenever arm i is queried, $Z_t = Y_t - f_i^\star$ is conditionally centered and has deterministic, history-uniform two-sided sub-gamma parameters $v_i, \kappa_i \ge 0$:*

$$\log \mathbb{E}[e^{\lambda Z_t} \mid \mathcal{H}_{t-1}, i_t = i] \le \frac{\lambda^2 v_i}{2(1 - \kappa_i|\lambda|)}, \qquad |\lambda| < \frac{1}{\kappa_i}.$$

When $\kappa_i = 0$, the domain is understood to be all of $\mathbb{R}$. Let $Y_{i,1}, Y_{i,2}, \ldots$ be the attained transformed samples in query order, and define

$$\overline{Y}_{i,n} = \frac{1}{n}\sum_{m=1}^{n} Y_{i,m}, \qquad \iota_n(\delta) = \log\left(\frac{\pi^2 K(1 + \lfloor \log_2 n \rfloor)^2}{3\delta}\right),$$

$$\beta_{i,n}(\delta) = 2\sqrt{\frac{v_i \iota_n(\delta)}{n}} + \frac{\kappa_i \iota_n(\delta)}{n}, \qquad n \ge 1.$$

Then

$$\Pr\big(\forall i \in [K],\ \forall \text{ attained } n \ge 1,\ \big|\overline{Y}_{i,n} - f_i^\star\big| \le \beta_{i,n}(\delta)\big) \ge 1 - \delta.$$

Proof. Put $\mathcal{S}_{i,n} = \sum_{m=1}^{n}(Y_{i,m} - f_i^\star)$. First fix a query horizon N and $\iota > 0$. Lemma 6, stopped when the query count reaches N, and Ville's inequality give the maximal Bernstein bound

$$\Pr\Big(\exists \text{ attained } m \le N :\ \mathcal{S}_{i,m} \ge \sqrt{2Nv_i\iota} + \kappa_i\iota\Big) \le e^{-\iota}.$$

Indeed, for $v_i > 0$, use the fixed parameter

$$\lambda = \frac{\sqrt{2\iota/(Nv_i)}}{1 + \kappa_i\sqrt{2\iota/(Nv_i)}} < \frac{1}{\kappa_i};$$

then $\lambda(\sqrt{2Nv_i\iota} + \kappa_i\iota) - N\lambda^2 v_i/[2(1 - \kappa_i\lambda)] = \iota$. The cases $v_i = 0$ or $\kappa_i = 0$ follow by the corresponding limit. Applying the same argument to $-\mathcal{S}_{i,m}$ gives the lower tail.

For epoch $2^r \le n < 2^{r+1}$, let $\iota_r = \log\{\pi^2 K(r+1)^2/(3\delta)\}$. Throughout that epoch,

$$n\beta_{i,n} = 2\sqrt{v_i n \iota_r} + \kappa_i\iota_r \ge \sqrt{2v_i 2^{r+1}\iota_r} + \kappa_i\iota_r.$$

The two-sided failure probability for arm i in epoch r is therefore at most $2e^{-\iota_r}$. Finally,

$$\sum_{i=1}^{K}\sum_{r=0}^{\infty} 2e^{-\iota_r} = \frac{6\delta}{\pi^2}\sum_{r=0}^{\infty}\frac{1}{(r+1)^2} = \delta.$$

□

Proposition 16 (Exact Gaussian transform and sub-gamma constants)**.** *Consider the setup of Proposition 14. Fix an arm i, write*

$$A_i = H_i^\dagger, \qquad \Pi_i = A_i H_i,$$

and condition on $\mathcal{H}_{t-1}$ and the predictable event $i_t = i$. Suppose

$$\zeta_t^f \sim \mathcal{N}(0, \sigma_{f,i,t}^2), \qquad \zeta_t^g = \Sigma_{i,t}^{1/2} z_t, \qquad z_t \sim \mathcal{N}(0, I_{d_q}),$$

where $\sigma_{f,i,t}^2$ and $\Sigma_{i,t} \succeq 0$ are predictable and ζ_t^f is conditionally independent of z_t. The scalar trace correction $\operatorname{tr}(A_i \Sigma_{i,t})$ is assumed known before the response. Define

$$\Omega_{i,t} = \Sigma_{i,t}^{1/2} A_i \Sigma_{i,t}^{1/2}, \qquad \mathsf{h}_{i,t} = \Sigma_{i,t}^{1/2} A_i \nabla f_i(x_{i,t}),$$

$$\kappa_{i,t} = \|\Omega_{i,t}\|_{\mathrm{op}}, \qquad v_{i,t} = \sigma_{f,i,t}^2 + \|\mathsf{h}_{i,t}\|^2 + \frac{1}{2}\|\Omega_{i,t}\|_F^2.$$

Then the known-covariance transform is conditionally centered at $f_i^\star = c_i$, and, whenever $I + \lambda \Omega_{i,t} \succ 0$, its exact conditional log-MGF is

$$\log \mathbb{E}\Big[e^{\lambda(Y_t - f_i^\star)} \Big| \mathcal{H}_{t-1}, i_t = i\Big] = \frac{\lambda^2 \sigma_{f,i,t}^2}{2} + \frac{1}{2}\{\lambda \operatorname{tr} \Omega_{i,t} - \log\det(I + \lambda \Omega_{i,t})\} + \frac{\lambda^2}{2} \mathsf{h}_{i,t}^\top (I + \lambda \Omega_{i,t})^{-1} \mathsf{h}_{i,t}.$$

When $\kappa_{i,t} > 0$, the exact domain is $\lambda > -1/\kappa_{i,t}$; when $\kappa_{i,t} = 0$, all real λ are allowed. On the symmetric domain $|\lambda| < 1/\kappa_{i,t}$,

$$\log \mathbb{E}\Big[e^{\lambda(Y_t - f_i^\star)} \Big| \mathcal{H}_{t-1}, i_t = i\Big] \le \frac{\lambda^2 v_{i,t}}{2(1 - \kappa_{i,t}|\lambda|)}.$$

Moreover,

$$\operatorname{Var}(Y_t - f_i^\star \mid \mathcal{H}_{t-1}, i_t = i) = v_{i,t}.$$

Finally,

$$A_i \nabla f_i(x_{i,t}) = \Pi_i (x_{i,t} - b_i).$$

Thus this vector equals $x_{i,t} - b_i$ when $H_i \succ 0$, but only the projection onto $\operatorname{range}(H_i)$ when H_i is singular.

Proof. Put $y_t = x_{i,t} - b_i$ and $g_t = H_i y_t$. The Moore–Penrose identity $H_i A_i H_i = H_i$ gives

$$g_t^\top A_i g_t = y_t^\top H_i A_i H_i y_t = y_t^\top H_i y_t,$$

so $f_i(x_{i,t}) - g_t^\top A_i g_t / 2 = c_i$. Also $A_i g_t = \Pi_i y_t$. Expanding the noisy transform therefore yields

$$Y_t - c_i = \zeta_t^f - (A_i g_t)^\top \zeta_t^g - \frac{1}{2}\Big\{\zeta_t^{g\top} A_i \zeta_t^g - \operatorname{tr}(A_i \Sigma_{i,t})\Big\}.$$

With $\Omega = \Omega_{i,t}$, $\mathsf{h} = \mathsf{h}_{i,t}$, and $z = z_t$, this becomes

$$Y_t - c_i = \zeta_t^f - \mathsf{h}^\top z - \frac{1}{2}(z^\top \Omega z - \operatorname{tr} \Omega).$$

Conditional centering follows immediately. When $I + \lambda\Omega \succ 0$, completion of the Gaussian square gives

$$\mathbb{E} \exp\left\{-\lambda \mathsf{h}^\top z - \frac{\lambda}{2} z^\top \Omega z\right\} = \det(I + \lambda\Omega)^{-1/2} \exp\left\{\frac{\lambda^2}{2} \mathsf{h}^\top (I + \lambda\Omega)^{-1} \mathsf{h}\right\}.$$

Multiplying by the centering factor $\exp\{\lambda \operatorname{tr}(\Omega)/2\}$ and by the Gaussian value-noise MGF proves the exact expression. Because $\Omega \succeq 0$, the positive definiteness condition is exactly $\lambda > -1/\|\Omega\|_{\mathrm{op}}$, with the stated zero-scale convention.

For $|u| < 1$,

$$u - \log(1 + u) \le \frac{u^2}{2(1 - |u|)}.$$

Applying this inequality eigenvalue by eigenvalue gives

$$\frac{1}{2}\{\lambda \operatorname{tr}\Omega - \log\det(I+\lambda\Omega)\} \le \frac{\lambda^2\|\Omega\|_F^2}{4(1-\|\Omega\|_{\mathrm{op}}|\lambda|)}.$$

The resolvent bound

$$\|(I+\lambda\Omega)^{-1}\|_{\mathrm{op}} \le \frac{1}{1-\|\Omega\|_{\mathrm{op}}|\lambda|}$$

and enlargement of the value-noise term by the same denominator prove the sub-gamma inequality. The Gaussian linear term and the centered quadratic term are uncorrelated because their product is an odd polynomial in z, whereas

$$\operatorname{Var}\left(\frac{1}{2}z^\top\Omega z\right) = \frac{1}{2}\|\Omega\|_F^2.$$

Independence of the value noise proves the variance identity. □

Corollary 7 (Observable covariance domination)**.** *In the setting of Proposition 16, suppose that throughout the time range of interest the known deterministic bounds*

$$\sigma_{f,i,t}^2 \le \bar{\sigma}_{f,i}^2, \qquad 0 \preceq \Sigma_{i,t} \preceq \bar{\Sigma}_i, \qquad \|A_i\nabla f_i(x_{i,t})\| \le D_i$$

hold. The actual trace correction must remain predictable. Define

$$\overline{\Omega}_i = A_i^{1/2}\bar{\Sigma}_i A_i^{1/2}, \qquad \Lambda_i = \|\Pi_i\bar{\Sigma}_i\Pi_i\|_{\mathrm{op}},$$

$$\bar{v}_i = \bar{\sigma}_{f,i}^2 + \Lambda_i D_i^2 + \frac{1}{2}\|\overline{\Omega}_i\|_F^2, \qquad \bar{\kappa}_i = \|\overline{\Omega}_i\|_{\mathrm{op}}.$$

Then $v_{i,t} \le \bar{v}_i$ and $\kappa_{i,t} \le \bar{\kappa}_i$. Hence the conditional sub-gamma bound holds uniformly with the deterministic, observable pair $(\bar{v}_i, \bar{\kappa}_i)$.

Proof. The vector $A_i\nabla f_i(x_{i,t})$ lies in $\operatorname{range}(H_i)$, so

$$\begin{aligned}&(A_i\nabla f_i)^\top\Sigma_{i,t}(A_i\nabla f_i)\\ &\quad\le (A_i\nabla f_i)^\top\bar{\Sigma}_i(A_i\nabla f_i) \le \Lambda_i D_i^2.\end{aligned}$$

The matrices $\Omega_{i,t} = \Sigma_{i,t}^{1/2}A_i\Sigma_{i,t}^{1/2}$ and $\widetilde{\Omega}_{i,t} = A_i^{1/2}\Sigma_{i,t}A_i^{1/2}$ have the same nonzero eigenvalues, while

$$0 \preceq \widetilde{\Omega}_{i,t} \preceq A_i^{1/2}\bar{\Sigma}_i A_i^{1/2} = \overline{\Omega}_i.$$

Eigenvalue monotonicity therefore yields

$$\|\Omega_{i,t}\|_{\mathrm{op}} \le \|\overline{\Omega}_i\|_{\mathrm{op}}, \qquad \|\Omega_{i,t}\|_F^2 \le \|\overline{\Omega}_i\|_F^2.$$

Substitution in Proposition 16 proves the claim. □

Corollary 8 (Observable isotropic calibration)**.** *Suppose $\Sigma_{i,t} = \sigma_{g,i}^2 I$, $\sigma_{f,i,t}^2 \le \sigma_{f,i}^2$, and a deterministic known bound $\|H_i^\dagger\nabla f_i(x_{i,t})\| \le D_i$ holds. Then usable uniform sub-gamma parameters are*

$$\boxed{\bar{v}_i = \sigma_{f,i}^2 + \sigma_{g,i}^2 D_i^2 + \frac{1}{2}\sigma_{g,i}^4\|H_i^\dagger\|_F^2, \qquad \bar{\kappa}_i = \sigma_{g,i}^2\|H_i^\dagger\|_{\mathrm{op}}.}$$

For $H_i \succ 0$, the bounded vector is $x_{i,t} - b_i$. For singular H_i, it is only $\Pi_i(x_{i,t} - b_i)$.

Proof. Here $\Omega_{i,t} = \sigma_{g,i}^2 A_i$ and $\mathsf{h}_{i,t} = \sigma_{g,i}A_i\nabla f_i(x_{i,t})$. Insert these expressions in Proposition 16 and use $A_i\nabla f_i = \Pi_i(x_{i,t} - b_i)$. □

A.6.2 PROOF OF THEOREM 4

Proof of Theorem 4. Step 1 (reduction and simultaneous confidence). Assumption 1 makes $x_{i,t}$ $\mathcal{H}_{t-1}$-measurable, including in the frozen case, as the conditional calculations require. Proposition 14 gives $\mathbb{E}[Y_t \mid \mathcal{H}_{t-1}, i_t = i] = f_i^\star$, so each transformed observation estimates the fixed arm mean, regardless of the state path. Under the theorem's deterministic sub-gamma parameters, Corollary 6 supplies an event $\mathcal{E}_\delta$ of probability at least $1-\delta$ on which, for every arm and attained count,

$$|\overline{Y}_{i,n} - f_i^\star| \le \beta_{i,n}, \qquad \beta_{i,n} = 2\sqrt{\frac{v_i \iota_n}{n}} + \frac{\kappa_i \iota_n}{n},$$

where $\iota_n = \log\{\pi^2 K(1+\lfloor \log_2 n \rfloor)^2/(3\delta)\}$. Write $\underline{V}_i = \overline{Y}_{i,N_i} - \beta_{i,N_i}$ and $\overline{V}_i = \overline{Y}_{i,N_i} + \beta_{i,N_i}$, where N_i is the current count.

Step 2 (correctness upon stopping). At every decision point let $\hat{\imath} \in \arg\min_i \overline{V}_i$ and $\check{\imath} \in \arg\min_{j \ne \hat{\imath}} \underline{V}_j$. If $\overline{V}_{\hat{\imath}} < \underline{V}_{\check{\imath}}$, the rule stops and recommends $\hat{\imath}$. On $\mathcal{E}_\delta$, for every $j \ne \hat{\imath}$,

$$f_{\hat{\imath}}^\star \le \overline{V}_{\hat{\imath}} < \underline{V}_{\check{\imath}} \le \underline{V}_j \le f_j^\star.$$

Thus any finite recommendation is correct on $\mathcal{E}_\delta$.

Step 3 (geometric sweeps and almost-sure termination). After one warm-start sample from every arm, phase $h = 0, 1, \ldots$ contains at most 2^h ordinary LUCB pulls. Before each such pull the rule checks the stopping condition; if it continues, it pulls the wider of $\hat{\imath}, \check{\imath}$. If the phase finishes without a stop, it performs one all-arm sweep, checking for a stop as observations arrive.

The fixed conditional sub-gamma assumption implies uniformly bounded conditional second moments. Hence the martingale strong law in each arm's query clock gives $\overline{Y}_{i,n} \to f_i^\star$ almost surely whenever that arm is queried infinitely often. If the algorithm never stopped, every finite phase would finish and infinitely many all-arm sweeps would occur, so every arm would be queried infinitely often. Since $\beta_{i,n} \to 0$ and the optimum is unique, eventually $\overline{V}_{i^\star} < \underline{V}_j$ for all $j \ne i^\star$, forcing a stop. This contradiction proves $\tau_Y < \infty$ almost surely, including outside $\mathcal{E}_\delta$. Together with Step 2, the rule is δ-correct.

Step 4 (the charging lemma). Put $\Delta_i = f_i^\star - f_{i^\star}^\star$ and

$$\gamma_i = \Delta_i \quad (i \ne i^\star), \qquad \gamma_{i^\star} = \Delta_{\min} := \min_{j \ne i^\star} \Delta_j.$$

On $\mathcal{E}_\delta$, intervals of radius $r_i = \beta_{i,N_i}$ obey

$$f_i^\star - 2r_i \le \underline{V}_i \le f_i^\star \le \overline{V}_i \le f_i^\star + 2r_i.$$

At a nonstopping ordinary decision let $k \in \arg\max_{j \in \{\hat{\imath}, \check{\imath}\}} r_j$. We show $4r_k \ge \gamma_k$ in all three possible configurations.

If $\hat{\imath} = i^\star$, nonstopping gives $\overline{V}_{\hat{\imath}} \ge \underline{V}_{\check{\imath}}$, whence

$$\Delta_{\check{\imath}} \le 2(r_{\hat{\imath}} + r_{\check{\imath}}) \le 4r_k.$$

If $k = \check{\imath}$, this is $4r_k \ge \gamma_k$; if $k = \hat{\imath}$, use $\gamma_{\hat{\imath}} = \Delta_{\min} \le \Delta_{\check{\imath}}$.

If $\hat{\imath} \ne i^\star$ and $\check{\imath} = i^\star$, the leader property gives $\overline{V}_{\hat{\imath}} \le \overline{V}_{i^\star}$, so $\Delta_{\hat{\imath}} \le 2r_{i^\star}$. If $k = \check{\imath}$, then $2r_k \ge \Delta_{\min} = \gamma_k$; if $k = \hat{\imath}$, then $r_{\hat{\imath}} \ge r_{i^\star}$ and $2r_k \ge \Delta_{\hat{\imath}} = \gamma_k$.

Finally, suppose $\hat{\imath} \ne i^\star$ and $\check{\imath} \ne i^\star$. Since the optimal arm is eligible in the challenger minimization, $\underline{V}_{\check{\imath}} \le \underline{V}_{i^\star} \le f_{i^\star}^\star$, and hence $\Delta_{\check{\imath}} \le 2r_{\check{\imath}}$. This charges $k = \check{\imath}$. If $k = \hat{\imath}$, use also $\overline{V}_{\hat{\imath}} \le \overline{V}_{\check{\imath}} = \underline{V}_{\check{\imath}} + 2r_{\check{\imath}} \le f_{i^\star}^\star + 2r_{\check{\imath}}$, which gives $\Delta_{\hat{\imath}} \le 2r_{\check{\imath}} \le 2r_{\hat{\imath}}$. Thus $4r_k \ge \gamma_k$ always. The factor four is attained when $\hat{\imath} = i^\star$, both radii equal $\Delta_{\check{\imath}}/4$, and $\overline{V}_{\hat{\imath}} = \underline{V}_{\check{\imath}}$.

Step 5 (ordinary pulls and sweep overhead). Define the bad-count cardinalities, rather than a first crossing,

$$M_i = \#\{n \ge 1 : 4\beta_{i,n} \ge \gamma_i\}, \qquad M_\delta = \sum_{i=1}^K M_i.$$

The distinction matters because the dyadic radius can jump upward at powers of two. Every ordinary pull of arm i occurs at a distinct pre-pull count in this set, so the total number N_{ord} of ordinary pulls

is at most M_δ on $\mathcal{E}_\delta$. If N_{sw} all-arm sweeps are begun, completing the preceding phases requires $N_{\mathrm{ord}} \geq 1 + 2 + \cdots + 2^{N_{\mathrm{sw}}-1} = 2^{N_{\mathrm{sw}}} - 1$. Therefore

$$N_{\mathrm{sw}} \leq \lfloor \log_2(N_{\mathrm{ord}} + 1) \rfloor \leq \lfloor \log_2(M_\delta + 1) \rfloor.$$

Including the warm start and allowing a partial final sweep proves

$$\tau_Y \leq K + M_\delta + K\lfloor \log_2(M_\delta + 1) \rfloor \quad \text{on } \mathcal{E}_\delta.$$

Step 6 (explicit order of M_δ). Let

$$\omega_i = \frac{v_i}{\gamma_i^2} + \frac{\kappa_i}{\gamma_i}.$$

If $\omega_i = 0$, then $M_i = 0$. Otherwise, if $4\beta_{i,n} \geq \gamma_i$, necessarily

$$n \leq 64\omega_i \iota_n.$$

To verify the constant, write $p = (v_i/\gamma_i^2)/\omega_i \in [0, 1]$. If $n > 64\omega_i\iota_n$, then

$$\frac{\beta_{i,n}}{\gamma_i} < \frac{\sqrt{p}}{4} + \frac{1-p}{64} \leq \frac{1}{4};$$

the final function is increasing on $[0, 1]$ and equals $1/4$ at one, a contradiction.

For completeness, this implicit inequality has the claimed log–log inverse. Let $y = n/\omega_i$. Since $n \geq 1$, for an absolute constant C_0,

$$\log(1 + \log_2 n) \leq C_0 + \log\big(1 + \log(e + \omega_i)\big) + \log\big(1 + \log(e + y)\big).$$

The continuous function $\log(1 + \log(e + y)) - y/256$ is bounded above on $[0, \infty)$. Substituting this bound into $y \leq 64\iota_n$ absorbs at most $y/2$ into the left side and gives

$$n = O\left(\omega_i \left[\log \frac{K}{\delta} + \log\big(1 + \log(e + \omega_i)\big)\right]\right).$$

The cardinality of the bad-count set is no larger than its largest element. Summing the last display over arms proves the stated order for M_δ. □

Remark 2 (Instantiating the tail parameters)**.** *Theorem 4 requires deterministic uniform sub-gamma parameters for the transformed samples. Proposition 15 supplies them for bounded noise, and Proposition 16 with Corollaries 7 and 8 supplies observable parameters for Gaussian noise. The sparse sweeps in Algorithm 2 guarantee that every arm is sampled infinitely often, which yields almost-sure termination under a unique best arm even outside the confidence event.*

A.7 Secondary Results

The results in this subsection are not used in the main text. Each extends the main results in one direction; their proofs are self-contained given the lemmas above.

A.7.1 Constrained standalone problems

If arm i is deployed on a closed convex domain $\mathcal{X}_i$, its standalone optimum is $f_i^\star = \min_{x \in \mathcal{X}_i} f_i(x)$. The gradient-norm certificate of Eq. (2) is then no longer valid at the boundary; the certificate below replaces the gradient by the gradient mapping and remains exact at a constrained minimizer.

Proposition 17 (Boundary-exact certificate on a convex domain)**.** *Suppose f_i is μ-strongly convex and L-smooth on a nonempty closed convex arm-local domain $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$, with constrained minimizer $x_i^\star \in \mathcal{X}_i$ and value $f_i^\star = \min_{x \in \mathcal{X}_i} f_i(x)$. Define the gradient-mapping operator*

$$G_i^L(x_i) = L\big(x_i - \mathrm{Proj}_{\mathcal{X}_i}(x_i - \tfrac{1}{L}\nabla f_i(x_i))\big), \qquad x_i \in \mathcal{X}_i.$$

For any $a \in (0, \mu]$, also define

$$Q_{a,i}(x_i; v) = \min_{y \in \mathcal{X}_i} \left\{ \langle v, y - x_i \rangle + \frac{a}{2}\|y - x_i\|^2 \right\}.$$

Then, for every $x_i \in \mathcal{X}_i$,

$$f_i(x_i) + Q_{a,i}(x_i; \nabla f_i(x_i)) \;\le\; f_i^\star \;\le\; f_i(x_i) - \frac{\|G_i^L(x_i)\|^2}{2L}.$$

Moreover,

$$Q_{a,i}(x_i; v) = -\frac{\|v\|^2}{2a} + \frac{a}{2}\operatorname{dist}^2(x_i - v/a, \mathcal{X}_i),$$

where $\operatorname{dist}(z, \mathcal{X}_i) = \min_{y\in\mathcal{X}_i} \|z - y\|$*, so the lower endpoint is computable by one Euclidean projection.*

Proof. For every $y \in \mathcal{X}_i$, strong convexity and $a \le \mu$ give

$$f_i(y) \ge f_i(x_i) + \langle \nabla f_i(x_i), y - x_i\rangle + \frac{a}{2}\|y - x_i\|^2.$$

Minimizing the right-hand side over $y \in \mathcal{X}_i$ proves the lower bound. Completing the square inside this minimum gives the displayed distance formula.

Upper bound. Set $x_i^+ = \operatorname{Proj}_{\mathcal{X}_i}(x_i - \nabla f_i(x_i)/L)$, so $G_i^L(x_i) = L(x_i - x_i^+)$ and $x_i^+ \in \mathcal{X}_i$. By L-smoothness,

$$f_i(x_i^+) \le f_i(x_i) + \nabla f_i(x_i)^\top (x_i^+ - x_i) + \tfrac{L}{2}\|x_i^+ - x_i\|^2.$$

Projection optimality (variational characterization) applied at the feasible point $y = x_i \in \mathcal{X}_i$ gives $\langle x_i - \nabla f_i(x_i)/L - x_i^+,\, x_i - x_i^+\rangle \le 0$, which rearranges to $\nabla f_i(x_i)^\top (x_i^+ - x_i) \le -L\|x_i - x_i^+\|^2$. Substituting,

$$f_i(x_i^+) \le f_i(x_i) - L\|x_i - x_i^+\|^2 + \tfrac{L}{2}\|x_i - x_i^+\|^2 = f_i(x_i) - \tfrac{1}{2L}\|G_i^L(x_i)\|^2.$$

Since $x_i^+ \in \mathcal{X}_i$, $f_i^\star \le f_i(x_i^+)$, proving the upper bound. □

Remark 3 (Boundary exactness)**.** *At a constrained minimizer, first-order optimality gives $\langle \nabla f_i(x_i^\star), y - x_i^\star\rangle \ge 0$ for every $y \in \mathcal{X}_i$. Hence $Q_{a,i}(x_i^\star; \nabla f_i(x_i^\star)) = 0$. Projection optimality also gives $G_i^L(x_i^\star) = 0$. Both deterministic endpoints therefore equal $f_i^\star$, including when $x_i^\star$ lies on an active boundary.*

Proposition 18 (Robust noisy constrained certificate)**.** *Under the assumptions of Proposition 17, fix $a \in (0, \mu]$. At a query point $x_{i,t} \in \mathcal{X}_i$, suppose $|F_t - f_i(x_{i,t})| \le r_t^f$ and $\|G_t - \nabla f_i(x_{i,t})\| \le r_t^g$. Define*

$$Q_{a,i}^-(x_{i,t}; G_t, r_t^g) = \min_{y\in\mathcal{X}_i} \Big\{ \langle G_t, y - x_{i,t}\rangle - r_t^g\|y - x_{i,t}\| + \frac{a}{2}\|y - x_{i,t}\|^2 \Big\}$$

and the noisy gradient mapping $\widehat{G}_i^L(x_{i,t}) = L(x_{i,t} - \operatorname{Proj}_{\mathcal{X}_i}(x_{i,t} - G_t/L))$. Then

$$F_t - r_t^f + Q_{a,i}^-(x_{i,t}; G_t, r_t^g) \le f_i^\star \le F_t + r_t^f - \frac{(\|\widehat{G}_i^L(x_{i,t})\| - r_t^g)_+^2}{2L}.$$

The robust lower term is finite because its quadratic term is coercive; it can equivalently be written $\min_{\|e\|\le r_t^g} Q_{a,i}(x_{i,t}; G_t - e)$. A fully projection-computable relaxation is available for every $\lambda \in (0, a)$:

$$F_t - r_t^f + Q_{a-\lambda,i}(x_{i,t}; G_t) - \frac{(r_t^g)^2}{2\lambda} \le f_i^\star.$$

Each such endpoint uses one Euclidean projection through the distance formula in Proposition 17; for example, $\lambda = a/2$ is an explicit parameter-free choice once a is fixed.

Proof. For $d = y - x_{i,t}$, Cauchy–Schwarz gives $\langle \nabla f_i(x_{i,t}), d\rangle \ge \langle G_t, d\rangle - r_t^g\|d\|$. The deterministic lower proof and $F_t - r_t^f \le f_i(x_{i,t})$ then give the first inequality. Nonexpansiveness of projection gives $\|\widehat{G}_i^L - G_i^L\| \le r_t^g$, hence $\|G_i^L\| \ge (\|\widehat{G}_i^L\| - r_t^g)_+$. Combining this with the deterministic upper bound and $f_i(x_{i,t}) \le F_t + r_t^f$ gives the second. Finally, minimizing first over e in $\min_{\|e\|\le r_t^g, y\in\mathcal{X}_i}\{\langle G_t - e, d\rangle + a\|d\|^2/2\}$ yields the stated equivalent form. Young's inequality $-r_t^g\|d\| \ge -\lambda\|d\|^2/2 - (r_t^g)^2/(2\lambda)$ then lower-bounds $Q_{a,i}^-$ by the displayed projection-computable term. □

A.7.2 LEARNED CURVATURE

We next allow the curvature in the cross-fitted statistic to be learned from the past. The approximation must be predictable: using either current gradient replicate to construct the matrix used in the same statistic would invalidate both its conditional mean calculation and its tail bound.

Theorem 7 (Bias-robust identification with predictable curvature)**.** *Assume $K \geq 2$, $\delta, \delta_H > 0$ with $\delta + \delta_H < 1$, a unique best arm, and*

$$f_i(x) = \frac{1}{2}(x - b_i)^\top H_i (x - b_i) + c_i, \qquad \mu_i I \preceq H_i \preceq L_i I,$$

where the valid spectral constants $\mu_i, L_i > 0$ are known. Assume the centered conditional noise and same-state replicate conditions of Theorem 6, and the sure public tube

$$\|x_{i,t} - b_i\| \leq D_i^{\mathrm{pub}} < \infty \quad \textit{for every reachable state.} \tag{71}$$

At a post-calibration macro-query, let $\widetilde{H}_{i,t}$ be a symmetric positive-definite matrix measurable before either primitive draw, put $\widetilde{A}_{i,t} = \widetilde{H}_{i,t}^{-1}$, and define

$$Y_t^{\mathrm{rob}} = F_t^{(1)} - \frac{1}{2}(G_t^{(1)})^\top \widetilde{A}_{i,t} G_t^{(2)}. \tag{72}$$

Every matrix actually used by the rule must satisfy $\lambda_{\min}(\widetilde{H}_{i,t}) \geq \mu_i/2$ on every path, enforced by an observed eigenvalue gate or deterministic spectral flooring. Consequently, the centered part of Eq. (72) has the sure uniform sub-gamma parameters

$$v_i^{\mathrm{rob}} = 4\sigma_{f,i}^2 + \frac{8\sigma_{g,i}^2 L_i^2 (D_i^{\mathrm{pub}})^2}{\mu_i^2} + \frac{4\sigma_{g,i}^4 d_i}{\mu_i^2}, \qquad \kappa_i^{\mathrm{rob}} = \frac{4\sigma_{g,i}^2}{\mu_i}. \tag{73}$$

Let $\beta_{i,n}^{\mathrm{rob}}$ be the radius of Corollary 6 with these parameters.

Suppose there is an event $\mathcal{E}_H$, with $\Pr(\mathcal{E}_H) \geq 1 - \delta_H$, and deterministic known sequences $\bar{q}_{i,s} < \mu_i/2$, defined for every counterfactual post-calibration arm count $s \geq 1$, such that whenever the s-th post-calibration query of arm i is attained,

$$\|\widetilde{H}_{i,t} - H_i\|_{\mathrm{op}} \leq \bar{q}_{i,s} \qquad \textit{on } \mathcal{E}_H. \tag{74}$$

Define the deterministic envelopes

$$\bar{e}_{i,s} = \frac{L_i (D_i^{\mathrm{pub}})^2 \bar{q}_{i,s}}{2(\mu_i - \bar{q}_{i,s})}, \qquad \bar{r}_{i,n} = \beta_{i,n}^{\mathrm{rob}} + \frac{1}{n}\sum_{s=1}^{n} \bar{e}_{i,s}. \tag{75}$$

The bias-robust rule forms all confidence intervals with $\bar{r}_{i,n}$, and uses this total radius, rather than $\beta_{i,n}^{\mathrm{rob}}$ alone, in every wider-arm comparison. Let $T_{\mathrm{cal}}, C_{\mathrm{cal}}$ be deterministic upper bounds, on $\mathcal{E}_H$, for its calibration macro-queries and primitive pairs. With γ_i as in Theorem 4, put

$$\overline{M}_i = \#\{n \geq 1 : 4\bar{r}_{i,n} \geq \gamma_i\}, \qquad \overline{M}_\delta = \sum_i \overline{M}_i. \tag{76}$$

If $\bar{q}_{i,s} \to 0$ for every arm, then $\overline{M}_\delta < \infty$ and

$$\begin{aligned} \Pr\Big(&\tau \leq T_{\mathrm{cal}} + K + \overline{M}_\delta + K\lfloor \log_2(\overline{M}_\delta + 1)\rfloor, \\ &C_\tau \leq C_{\mathrm{cal}} + 2\{K + \overline{M}_\delta + K\lfloor \log_2(\overline{M}_\delta + 1)\rfloor\},\ \widehat{J}_\tau = i^\star\Big) \geq 1 - \delta - \delta_H. \end{aligned} \tag{77}$$

The deterministic continuation in Eq. (75) is defined whether or not a sample is reached; hence $\overline{M}_\delta$ is an a-priori instance-dependent bound, not a random post-stop quantity.

If, additionally, calibration is almost surely finite and, along every arm's attained query clock, $\widetilde{H}_{i,t} \to H_i$ almost surely, then the rule stops almost surely and is $(\delta + \delta_H)$-correct. Finally, if $\bar{q}_{i,s} = O(\sqrt{\log s / s})$, then

$$\frac{1}{n}\sum_{s=1}^{n} \bar{e}_{i,s} = O(\sqrt{\log n / n}),$$

so learned curvature preserves the $\widetilde{O}(\Delta_{\min}^{-2})$ gap exponent.

Proof. Condition on a post-calibration macro-query and abbreviate $x = x_{i,t}$, $\widetilde{H} = \widetilde{H}_{i,t}$, and $\widetilde{A} = \widetilde{H}^{-1}$. The predictable conditional mean is $c_i + e_{i,t}$, where

$$e_{i,t} = \frac{1}{2}(x - b_i)^\top [H_i - H_i \widetilde{A} H_i](x - b_i).$$

On $\mathcal{E}_H$, the identity

$$H_i - H_i \widetilde{H}^{-1} H_i = H_i \widetilde{H}^{-1}(\widetilde{H} - H_i)$$

and Weyl's inequality give, at the attained count s,

$$|e_{i,t}| \le \frac{L_i (D_i^{\mathrm{pub}})^2 \bar{q}_{i,s}}{2(\mu_i - \bar{q}_{i,s})} = \bar{e}_{i,s}. \tag{78}$$

The pathwise eigenvalue gate and Eq. (71) give

$$\|\widetilde{A}\|_{\mathrm{op}} \le 2/\mu_i, \quad \|\widetilde{A}\|_F \le 2\sqrt{d_i}/\mu_i, \quad \|\widetilde{A}\nabla f_i(x)\| \le 2L_i D_i^{\mathrm{pub}}/\mu_i.$$

Applying Theorem 6's MGF calculation conditionally on the predictable $\widetilde{A}$ proves Eq. (73) on every path. Let $Z_{i,s}$ be the centered martingale difference at the s-th attained post-calibration query. Corollary 6 supplies an event $\mathcal{E}_Z$, with probability at least $1 - \delta$, on which

$$\left| \frac{1}{n} \sum_{s=1}^{n} Z_{i,s} \right| \le \beta_{i,n}^{\mathrm{rob}} \quad \text{for every arm and attained } n.$$

On $\mathcal{E}_H \cap \mathcal{E}_Z$, whose probability is at least $1 - \delta - \delta_H$, Eq. (78) therefore makes every interval of radius $\bar{r}_{i,n}$ valid.

The proof of Theorem 4 uses only simultaneous coverage and the fact that the queried member of the leader–challenger pair has the larger interval radius. Its charging lemma therefore gives $4\bar{r}_{i,N_i} \ge \gamma_i$ for every ordinary nonstopping pull. Because $\bar{r}_{i,n} \to 0$, each $\overline{M}_i$ is finite. Charging distinct arm counts yields at most $\overline{M}_\delta$ ordinary pulls, and the geometric phases add at most $K\lfloor \log_2(\overline{M}_\delta + 1) \rfloor$ sweep queries. Adding calibration and multiplying every post-calibration macro-query by two proves Eq. (77); coverage proves the recommendation.

For almost-sure termination, suppose the rule did not stop. Sparse sweeps would query every arm infinitely often. The uniform sub-gamma parameters imply uniformly bounded conditional second moments, so the martingale strong law gives $n^{-1} \sum_{s \le n} Z_{i,s} \to 0$ along each arm's query clock. Almost-sure curvature consistency, continuity of matrix inversion at $H_i \succ 0$, and the sure tube imply $e_{i,t} \to 0$, hence its Cesàro mean also vanishes. The configured radii vanish by $\bar{q}_{i,s} \to 0$. Unique optimality then forces strict interval separation after finitely many samples, a contradiction. Thus stopping is almost sure, including outside $\mathcal{E}_H \cap \mathcal{E}_Z$, and the joint-event correctness statement gives $(\delta + \delta_H)$-correctness. The final rate follows by summing the displayed square-root sequence. □

Proposition 19 (First-order curvature regression under persistent excitation)**.** *Assume the setting of Theorem 7. Enumerate arm i's gradient observations in its own query clock and set*

$$\phi_{i,s} = (x_{i,s}^\top, 1)^\top \in \mathbb{R}^{p_i}, \qquad p_i = d_i + 1, \qquad \Theta_i = [H_i, -H_i b_i].$$

Then $G_{i,s} = \Theta_i \phi_{i,s} + \zeta_{i,s}^g$. Suppose $\phi_{i,s}$ is predictable, the valid bounds $\overline{\phi}_i, \overline{\Theta}_i$ are known, and $\|\phi_{i,s}\| \le \overline{\phi}_i$, $\|\Theta_i\|_{\mathrm{op}} \le \overline{\Theta}_i$. Assume the gradient noise is conditionally $\sigma_{g,i}$-sub-Gaussian in every direction. For a fixed $\lambda_0 > 0$, define

$$\mathcal{V}_{i,n} = \lambda_0 I + \sum_{s=1}^{n} \phi_{i,s} \phi_{i,s}^\top, \qquad \widehat{\Theta}_{i,n} = \left(\sum_{s=1}^{n} G_{i,s} \phi_{i,s}^\top \right) \mathcal{V}_{i,n}^{-1},$$

let $\widehat{H}_{i,n}^{\mathrm{sym}}$ be the symmetrization of the first d_i columns of $\widehat{\Theta}_{i,n}$, and put

$$\delta_{i,n} = \frac{6\delta_H}{\pi^2 K n^2},$$

$$\Upsilon_{i,n} = 2\sigma_{g,i}\sqrt{2\left\{\frac{p_i}{2}\log\left(1+\frac{n\overline{\phi}_i^2}{\lambda_0 p_i}\right)+d_i\log 5+\log\frac{1}{\delta_{i,n}}\right\}}+\sqrt{\lambda_0}\overline{\Theta}_i, \tag{79}$$

$$q_{i,n} = \frac{\Upsilon_{i,n}}{\sqrt{\lambda_{\min}(\mathcal{V}_{i,n})}}. \tag{80}$$

Then, with probability at least $1-\delta_H$, simultaneously for every arm and $n \geq 1$,

$$\|\widehat{H}_{i,n}^{\mathrm{sym}} - H_i\|_{\mathrm{op}} \leq q_{i,n}. \tag{81}$$

Suppose moreover that the application-fixed transition and calibration schedule guarantee, for a known $\lambda_i^{\mathrm{PE}} > 0$,

$$\lambda_{\min}(\mathcal{V}_{i,n}) \geq \lambda_0 + \lambda_i^{\mathrm{PE}}(n-p_i+1)_+ \qquad \textit{for every arm and } n. \tag{82}$$

The burn-in is necessary: before p_i rank-one observations, the update to $\lambda_0 I$ cannot have positive minimum eigenvalue in every direction. Define the deterministic radii

$$q_{i,n}^{\mathrm{PE}} = \frac{\Upsilon_{i,n}}{\sqrt{\lambda_0+\lambda_i^{\mathrm{PE}}(n-p_i+1)_+}}, \qquad \check{q}_{i,n}^{\mathrm{PE}} = \min_{1\leq u\leq n} q_{i,u}^{\mathrm{PE}},$$

and

$$n_{\mathrm{cal}} = \max_i \min\{n \geq 1 : \check{q}_{i,n}^{\mathrm{PE}} < \mu_i/2\}. \tag{83}$$

After n_{cal} one-pair cyclic calibration observations per arm, each followed by its application-fixed mandatory transition, retain for each arm the best-so-far symmetrized estimator attaining $\check{q}_{i,n_{\mathrm{cal}}}^{\mathrm{PE}}$. On the event Eq. (81)*, each retained estimate is positive definite, and the deterministic calibration bounds in Theorem 7 are*

$$T_{\mathrm{cal}} = C_{\mathrm{cal}} = Kn_{\mathrm{cal}}. \tag{84}$$

An estimate formed from a calibration or post-calibration observation is used only from the following macro-query. Off the curvature event, calibration continues until the observed minimum-eigenvalue gate passes. After every later regression observation, the candidate is the best-so-far symmetrized estimate attaining the smallest deterministic $q_{i,u}^{\mathrm{PE}}$; a candidate that fails the gate never replaces the last safe estimator. The first gradient from every later cross-fitted macro-query may be reused in the regression. Before arm i's s-th post-calibration transformed sample, there are at least $n_{\mathrm{cal}}+s-1$ regression observations, so the deterministic sequence required by Theorem 7 is

$$\bar{q}_{i,s} = \check{q}_{i,n_{\mathrm{cal}}+s-1}^{\mathrm{PE}}. \tag{85}$$

It tends to zero at rate $O(\sqrt{(d_i\log s+\log(K/\delta_H))/s})$.

With $\overline{M}_\delta$ formed from this sequence through Eqs. (75)–(76)*, the explicit bounds of Theorem 7 specialize to*

$$\tau \leq Kn_{\mathrm{cal}} + K + \overline{M}_\delta + K\lfloor\log_2(\overline{M}_\delta+1)\rfloor,$$
$$C_\tau \leq Kn_{\mathrm{cal}} + 2\{K + \overline{M}_\delta + K\lfloor\log_2(\overline{M}_\delta+1)\rfloor\}$$

on the same $1-\delta-\delta_H$ event.

The retained estimates converge to H_i almost surely, the eigenvalue gate passes after finite calibration almost surely, and therefore the almost-sure termination conclusion of Theorem 7 applies. The calibration cost is gap-independent up to logarithms. No such conclusion is available from first-order data without an independent curvature stream or a persistently exciting design: frozen or rank-deficient state trajectories need not identify every Hessian direction.

Proof. Write $\Xi_{i,n} = \sum_{s=1}^n \zeta_{i,s}^g \phi_{i,s}^\top$. For a fixed unit vector $u \in \mathbb{R}^{d_i}$, the scalar noises $u^\top\zeta_{i,s}^g$ are conditionally $\sigma_{g,i}$-sub-Gaussian. The exponential-mixture proof for a self-normalized linear martingale, with failure probability $\delta_{i,n}/5^{d_i}$, gives

$$\|u^\top\Xi_{i,n}\mathcal{V}_{i,n}^{-1/2}\| \leq \sigma_{g,i}\sqrt{2\log\left(\frac{\det(\mathcal{V}_{i,n})^{1/2}}{\det(\lambda_0 I)^{1/2}(\delta_{i,n}/5^{d_i})}\right)}.$$

A $1/2$-net of the unit sphere has at most 5^{d_i} points and satisfies $\|M\|_{\mathrm{op}} \le 2\max_u \|u^\top M\|$. Also,

$$\frac{\det(\mathcal{V}_{i,n})}{\det(\lambda_0 I)} \le \left(1 + \frac{n\overline{\phi}_i^2}{\lambda_0 p_i}\right)^{p_i}.$$

The ridge identity

$$(\widehat{\Theta}_{i,n} - \Theta_i)\mathcal{V}_{i,n}^{1/2} = \Xi_{i,n}\mathcal{V}_{i,n}^{-1/2} - \lambda_0\Theta_i\mathcal{V}_{i,n}^{-1/2}$$

and $\|\lambda_0\Theta_i\mathcal{V}_{i,n}^{-1/2}\|_{\mathrm{op}} \le \sqrt{\lambda_0}\overline{\Theta}_i$ give $\Upsilon_{i,n}$ after the net bound. Division by $\sqrt{\lambda_{\min}(\mathcal{V}_{i,n})}$, restriction to the first d_i columns, and symmetrization cannot increase the operator error. Finally,

$$\sum_{i=1}^{K}\sum_{n=1}^{\infty}\delta_{i,n} = \delta_H,$$

which proves Eq. (81).

Under Eq. (82), $q_{i,n} \le q_{i,n}^{\mathrm{PE}}$ on the same event, and $q_{i,n}^{\mathrm{PE}} = O(\sqrt{(d_i\log n + \log(K/\delta_H))/n})$. Hence $n_{\mathrm{cal}} < \infty$, the retained best-so-far estimate is within $\mu_i/2$ of H_i, and it is positive definite. The cyclic calibration accounting gives Eq. (84); the one-step delay gives predictability; and the number of regression observations before a post-calibration query gives Eq. (85).

For almost-sure consistency, the individual failure probabilities $\delta_{i,n}$ are summable. The first Borel–Cantelli lemma and persistent excitation imply that $\|\widehat{H}_{i,n}^{\mathrm{sym}} - H_i\|_{\mathrm{op}} \to 0$ almost surely. Thus the observed eigenvalue gate eventually passes, calibration is almost surely finite, and the retained best-so-far estimates converge. The remaining claims follow from Theorem 7. □

A.7.3 Fixed-budget recommendation

The confidence sequence of Corollary 6 also gives a fixed-budget guarantee for the simple regret $R(T)$ of Section 3.

Corollary 9 (Uniform transformed fixed-budget rate)**.** *Assume the setting of Corollary 6. Fix an integer $n \ge 1$, collect exactly n transformed observations from every arm along any predictable coupled-state trajectory, and suppose each transformed observation uses exactly B primitive pairs. Set the primitive budget $T = BKn$. At the end of the budget, let*

$$\overline{V}_{i,n} = \overline{Y}_{i,n} + \beta_{i,n}(\delta), \qquad \widehat{J}_T \in \arg\min_{i\in[K]} \overline{V}_{i,n}.$$

Then, with probability at least $1-\delta$,

$$\begin{aligned} R(T) := f^\star_{\widehat{J}_T} - f^\star_{i^\star} &\le 2\beta_{i^\star,n}(\delta) \\ &= 4\sqrt{\frac{v_{i^\star}\iota_n(\delta)}{n}} + \frac{2\kappa_{i^\star}\iota_n(\delta)}{n} \\ &\le 4\sqrt{\frac{v_{\max}\iota_n(\delta)}{n}} + \frac{2\kappa_{\max}\iota_n(\delta)}{n}, \end{aligned}$$

where $v_{\max} = \max_i v_i$ and $\kappa_{\max} = \max_i \kappa_i$.

For an arbitrary primitive budget $T \ge BK$, take $n = \lfloor T/(BK)\rfloor$, allocate the first Kn transformed observations uniformly, and ignore the remaining primitive budget in the recommendation. The same bound with this integer n holds. In particular, because $n \ge T/(2BK)$,

$$R(T) \le 4\sqrt{\frac{2BKv_{\max}\iota_n(\delta)}{T}} + \frac{4BK\kappa_{\max}\iota_n(\delta)}{T} \qquad \textit{with probability at least } 1-\delta.$$

Proof. Let $\mathcal{E}_\delta$ be the simultaneous event from Corollary 6. On this event, $f_i^\star \le \overline{V}_{i,n}$ for every arm. Since $\widehat{J}_T$ minimizes the upper endpoint,

$$f^\star_{\widehat{J}_T} \le \overline{V}_{\widehat{J}_T,n} \le \overline{V}_{i^\star,n} = \overline{Y}_{i^\star,n} + \beta_{i^\star,n} \le f^\star_{i^\star} + 2\beta_{i^\star,n}.$$

Subtracting $f^\star_{i^\star}$ proves the first inequality. Substitution of the radius and domination by $v_{\max}, \kappa_{\max}$ prove the next two. For a general budget, the recommendation is measurable with respect to the first Kn observations and the same confidence event applies. Finally, $\lfloor y\rfloor \ge y/2$ for $y \ge 1$, which gives the displayed budget-only denominator bound. □

A.7.4 Optimality of the quadratic reduction

The next lower bound matches the $\sum_{i\neq i^\star} v_i \Delta_i^{-2} \log(1/\delta)$ term of Theorem 4 on a Gaussian quadratic subclass, on which $v_i = \sigma_i^2$ is the value-noise variance. It is the classical Gaussian best-arm lower bound embedded in the coupled protocol and does not identify a coupling-specific cost.

Theorem 8 (Gaussian subclass lower bound)**.** *Let $\delta \in (0, 1/2)$. Any algorithm that is δ-correct on every instance of the coupled class satisfies the following bound on the Gaussian quadratic subclass. Let $\alpha > 0$ be known and, for $i \in [K]$, set*

$$f_i(x) = \frac{\alpha}{2}x^2 + c_i, \qquad G_{t,r} = 0,$$
$$F_{t,r} \mid \mathcal{H}_{t-1}, i_t = i \sim \mathcal{N}(c_i, \sigma_i^2).$$

The primitive value noises are conditionally independent within and across attained batches, and $\sigma_i > 0$. Use either consensus or zero-sum coupling initialized at the zero state. If $i^\star$ is the unique minimizing offset and $\Delta_i = c_i - c_{i^\star}$, then

$$\mathbb{E}[C_\tau] \geq \sum_{i\neq i^\star} \frac{2\sigma_i^2}{\Delta_i^2} \operatorname{kl}(1-\delta, \delta),$$

where

$$\operatorname{kl}(p, q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$

Since $\operatorname{kl}(1-\delta, \delta) = (1-2\delta)\log\{(1-\delta)/\delta\} \sim \log(1/\delta)$ as $\delta \downarrow 0$, the suboptimal-arm $\sum_{i\neq i^\star} \sigma_i^2 \Delta_i^{-2} \log(1/\delta)$ dependence cannot be removed in general. This is the standard Gaussian BAI lower bound embedded in the coupled protocol; it does not show a coupling-specific cost, nor that the sweep and time-uniform overheads are optimal.

Proof. At the zero state every exact gradient is zero. A projected update based on $G_t = 0$ therefore leaves the state at zero under either canonical coupling. The transformed statistic is simply $Y_t = F_t$, and the gradient and state observations contain no information about the offsets. Thus the stopped transcript is that of adaptive sampling from independent Gaussian arms. Write $C_{i,\tau}$ for the total number of primitive pairs drawn from arm i, including all pairs inside its attained macro-queries.

Call the given instance 0. For a suboptimal arm i and $\varepsilon > 0$, define alternative i by lowering the offset of arm i to $c_{i^\star} - \varepsilon$ and leaving every other offset unchanged. Arm i is uniquely optimal under this alternative. Let $\mathcal{A} = \{\widehat{J}_\tau = i^\star\}$. Delta-correctness gives

$$\Pr_0(\mathcal{A}) \geq 1-\delta, \qquad \Pr_i(\mathcal{A}) \leq \delta.$$

If $\mathbb{E}_0[C_\tau] = \infty$, the conclusion is immediate, so assume it is finite. Let $\Pr_0^\tau$ and $\Pr_i^\tau$ denote the stopped-history laws, and let $\mathcal{L}_j^0$ and $\mathcal{L}_j^i$ be the laws of a primitive value observation of arm j under the two instances. The chain rule for relative entropy, first at $\tau \wedge m$ and then as $m \to \infty$, gives

$$\operatorname{KL}(\Pr_0^\tau \,\|\, \Pr_i^\tau) = \sum_{j=1}^K \mathbb{E}_0[C_{j,\tau}] \operatorname{KL}(\mathcal{L}_j^0 \| \mathcal{L}_j^i).$$

Finite expected primitive cost supplies the required integrability. Data processing to the binary event $\mathcal{A}$, together with the monotonicity of binary relative entropy over $\Pr_0(\mathcal{A}) \geq 1-\delta > \delta \geq \Pr_i(\mathcal{A})$, yields

$$\sum_{j=1}^K \mathbb{E}_0[C_{j,\tau}] \operatorname{KL}(\mathcal{L}_j^0 \| \mathcal{L}_j^i) \geq \operatorname{kl}(1-\delta, \delta).$$

Only arm i changes, and its equal-variance Gaussian divergence is

$$\operatorname{KL}\big(\mathcal{N}(c_i, \sigma_i^2) \,\|\, \mathcal{N}(c_{i^\star} - \varepsilon, \sigma_i^2)\big) = \frac{(\Delta_i + \varepsilon)^2}{2\sigma_i^2}.$$

Consequently,

$$\mathbb{E}_0[C_{i,\tau}] \geq \frac{2\sigma_i^2}{(\Delta_i + \varepsilon)^2} \operatorname{kl}(1-\delta, \delta).$$

Let $\varepsilon \downarrow 0$, sum the resulting inequalities over all $i \neq i^\star$, and use $C_\tau = \sum_i C_{i,\tau}$. □

B EXPERIMENTAL DETAILS

Every numerical table in this section is produced by a released generator from raw CSV files; the generators record input hashes and check the accounting (objective instances, independent units, primitive pairs) before writing the table. Table 7 lists the generator of each table and figure.

B.1 FRESH-SAMPLING EXPERIMENT

Coupling. We use $K = 4$ arms with $d_i = 4$ and the weighted balance constraint $\sum_{i=1}^{K} W_i x_i = 0$, where W_i is diagonal with $(W_i)_{kk} = \omega_\rho$ if $k = i$ and $(W_i)_{kk} = 1$ otherwise, and $\omega_\rho^2 = (1-\rho)(K-1)/\rho$. A direct computation of $P = I - A^\top (AA^\top)^{-1} A$ with $A = [W_1, \dots, W_K]$ gives

$$S_i = \operatorname{diag}\Big(1 - \tfrac{\rho}{K-1}, \dots, \underbrace{\rho}_{k=i}, \dots, 1 - \tfrac{\rho}{K-1}\Big),$$

so $\lambda_{\min}(S_i) = \rho$ for every arm. Each arm has one poorly accessible direction. At $\rho = 3/4$ all weights equal one, which is ordinary zero-sum balance with $S_i = \frac{3}{4} I$.

Objectives and oracle. Arm i is $f_i(z) = f_i^\star + \frac{1}{4}\|z - x_i^\star\|^2 + \sum_{k=1}^{4} \log\cosh(z_k - x_{ik}^\star)$, which is $\frac{1}{2}$-strongly convex and $\frac{3}{2}$-smooth with minimizer $x_i^\star$. Each instance draws the four minimizers uniformly from the unit sphere and sets $(f_i^\star)_i$ to a random permutation of $\Delta_{\min} \cdot (0, 1, 1.5, 2)$. A primitive query returns $f_i(x_i) + \zeta^f$ and $\nabla f_i(x_i) + \zeta^g$ with independent $\zeta^f \sim \mathcal{N}(0, 0.1^2)$ and $\zeta^g \sim \mathcal{N}(0, 0.05^2 I_4)$. The value proxy is $\sigma_i^f = 0.1$ for every arm. For the gradient-norm proxy, the Chernoff bound $\Pr(\|Z\|^2 > s^2 v) \le 2^{d/2} e^{-v/4}$ for $Z \sim \mathcal{N}(0, s^2 I_d)$ implies $\Pr(\|Z\| > u) \le 2e^{-u^2/(2\sigma^2)}$ with $\sigma^2 = s^2 \max\{d, 4\}$, so $\sigma_i^g = 0.1$ satisfies Eq. (4). Because the noise is Gaussian, the mean of a frozen batch of B fresh pairs has the exact law $\mathcal{N}(0, 0.1^2/B) \times \mathcal{N}(0, 0.05^2 I_4/B)$; we sample it directly, so every run is an exact simulation of the fresh-sampling oracle and the reported cost counts all B primitive pairs. The initial state is $x = 0$, and every update is the identity response with the exact geometry-scaled gain $\theta_i^- = \theta_i^+ = 2/(\mu_i \rho_i + L_i R_i)$, with $\rho_i = \lambda_{\min}(S_i)$ and $R_i = \lambda_{\max}(S_i)$, followed by the Euclidean projection onto the constraint set. Unless stated otherwise $\delta = 0.05$; the phase cap is $h \le 24$.

Units. Each cell uses 100 instances and 5 independent noise paths per instance (500 paths). The instances are the same in every cell, so gap, access, and method comparisons are matched; noise seeds differ across cells. Intervals are percentile intervals from 2,000 bootstrap resamples of instances.

Comparators. The *uncoupled reference* runs Algorithm 1 on the same objectives and noise without the constraint ($P = I$, $S_i = I$, gain $2/(\mu_i + L_i)$), i.e., rested best function identification. The other three rules use the same oracle, gains, projection, and transitions as Algorithm 1. *Interleaved* uses exactly the batches of Algorithm 1 in each phase but interleaves the optimization batches of different arms; its certificates and stopping rule are those of Algorithm 1, so it is δ-safe by Proposition 1. *Round-robin certificates* queries each arm once per phase with a B_h-pair certificate batch. *Value racing* uses the round-robin schedule but treats $F \pm r^f$ as a confidence interval for $f_i^\star$; this is a standard stationary racing rule that ignores the difference between $f_i(x_i)$ and $f_i^\star$.

Results. Table 2 lists every cell. Across 14 cells and 7,000 paths, Algorithm 1 made 0 wrong stops and stopped on every path; every certificate interval formed before stopping, by any of the certificate-based rules, contained its $f_i^\star$. With zero failures in 500 paths, the two-sided 95% Wilson upper bound on the per-path error rate is 0.76% in each cell (3.7% at the instance level). Increasing δ from 0.01 to 0.2 at $\rho = 1/8$, $\Delta_{\min} = 0.05$ lowered the mean cost from 87,498 to 68,606 pairs without producing an error, so the confidence allocation is conservative. Along the access sweep the median stopping phase varied only between 7 and 8. The mean cost grew from 3,396 to 209,926 and Γ_q from 12 to 388, so Γ_q accounts for most of the growth; the remaining factor 1.91 comes from runs that stop one phase later.

Table 2: All cells of the fresh-sampling experiment (500 paths per cell). "Wrong" counts stops with an incorrect answer; "No stop" counts paths that had not stopped by phase 24. For rules with unstopped paths, the parenthetical number is the median cost of the stopped paths only, which is biased downward. The access sweep shares the cell $\rho = 1/8$, $\Delta_{\min} = 0.1$ with the gap sweep.

Sweep	Rule	ρ	$\Delta_{\min}$	Wrong	No stop	Mean primitive pairs [95% CI]
gap	Algorithm 1	1/8	0.8	0/500	0/500	350 [335, 365]
gap	Algorithm 1	1/8	0.4	0/500	0/500	1,402 [1,360, 1,444]
gap	Algorithm 1	1/8	0.2	0/500	0/500	5,396 [5,241, 5,548]
gap	Algorithm 1	1/8	0.1	0/500	0/500	20,851 [20,274, 21,400]
gap	Algorithm 1	1/8	0.05	0/500	0/500	76,558 [74,050, 79,013]
gap	Algorithm 1	1/8	0.025	0/500	0/500	306,378 [298,065, 314,477]
gap	Algorithm 1, uncoupled	1	0.8	0/500	0/500	42 [40, 44]
gap	Algorithm 1, uncoupled	1	0.4	0/500	0/500	195 [189, 202]
gap	Algorithm 1, uncoupled	1	0.2	0/500	0/500	812 [784, 838]
gap	Algorithm 1, uncoupled	1	0.1	0/500	0/500	3,527 [3,423, 3,641]
gap	Algorithm 1, uncoupled	1	0.05	0/500	0/500	14,190 [13,772, 14,632]
gap	Algorithm 1, uncoupled	1	0.025	0/500	0/500	56,715 [55,142, 58,385]
gap	Interleaved (same batches)	1/8	0.8	0/500	110/500	(6,600)
gap	Interleaved (same batches)	1/8	0.4	0/500	370/500	(26,720)
gap	Interleaved (same batches)	1/8	0.2	0/500	450/500	(80,134)
gap	Interleaved (same batches)	1/8	0.1	0/500	485/500	(426,592)
gap	Interleaved (same batches)	1/8	0.05	0/500	495/500	(13,633,096)
gap	Interleaved (same batches)	1/8	0.025	0/500	500/500	–
gap	Round-robin certificates	1/8	0.8	0/500	100/500	(508)
gap	Round-robin certificates	1/8	0.4	0/500	366/500	(4,092)
gap	Round-robin certificates	1/8	0.2	0/500	449/500	(4,092)
gap	Round-robin certificates	1/8	0.1	0/500	482/500	(65,532)
gap	Round-robin certificates	1/8	0.05	0/500	496/500	(100,663,292)
gap	Round-robin certificates	1/8	0.025	0/500	500/500	–
gap	Value racing	1/8	0.8	0/500	0/500	56 [9, 147]
gap	Value racing	1/8	0.4	36/500	0/500	1,394 [185, 3,666]
gap	Value racing	1/8	0.2	133/500	0/500	35,329 [387, 103,795]
gap	Value racing	1/8	0.1	200/500	0/500	3,979 [1,817, 7,124]
gap	Value racing	1/8	0.05	239/500	0/500	10,576 [3,758, 19,923]
gap	Value racing	1/8	0.025	274/500	0/500	10,517 [4,855, 18,944]
access	Algorithm 1	3/4	0.1	0/500	0/500	3,396 [3,285, 3,521]
access	Algorithm 1	1/2	0.1	0/500	0/500	4,967 [4,805, 5,121]
access	Algorithm 1	1/4	0.1	0/500	0/500	9,908 [9,591, 10,225]
access	Algorithm 1	1/16	0.1	0/500	0/500	43,644 [42,664, 44,649]
access	Algorithm 1	1/32	0.1	0/500	0/500	95,222 [93,505, 96,838]
access	Algorithm 1	1/64	0.1	0/500	0/500	209,926 [204,539, 215,514]
access	Interleaved (same batches)	3/4	0.1	0/500	500/500	–
access	Interleaved (same batches)	1/2	0.1	0/500	495/500	(65,592)
access	Interleaved (same batches)	1/4	0.1	0/500	495/500	(459,080)
access	Interleaved (same batches)	1/16	0.1	0/500	480/500	(205,572)
access	Interleaved (same batches)	1/32	0.1	0/500	475/500	(402,840)
access	Interleaved (same batches)	1/64	0.1	0/500	475/500	(797,516)
access	Value racing	3/4	0.1	7/500	0/500	831 [779, 885]
access	Value racing	1/2	0.1	221/500	0/500	16,208 [6,007, 29,698]
access	Value racing	1/4	0.1	201/500	0/500	4,409 [1,323, 9,038]
access	Value racing	1/16	0.1	196/500	0/500	1,836 [1,085, 2,925]
access	Value racing	1/32	0.1	176/500	0/500	1,438 [962, 2,053]
access	Value racing	1/64	0.1	175/500	0/500	2,585 [1,586, 3,797]
$\delta = 0.01$	Algorithm 1	1/8	0.05	0/500	0/500	87,498 [84,883, 89,954]
$\delta = 0.2$	Algorithm 1	1/8	0.05	0/500	0/500	68,606 [66,258, 71,009]

B.2 TIED LINEAR HEADS ON DIGITS

Task. Each 8×8 digit (Alpaydin & Kaynak, 1998) is classified by one of four linear heads on 32 pixels (left, right, top, or bottom half). During selection, overlapping coefficients are tied; at deployment the tie is released and only the selected head is refit. Each pixel belongs to two heads, so $S_i = \frac{1}{2} I_{32}$. Head i minimizes $f_i(w) = n_i^{-1} \sum_r (p(a_r^\top w) - y_r)^2 + \frac{1}{8}\|w\|^2 + c_i^{\mathrm{acq}}$ over its masked features a_r, labels y_r, sigmoid p, and known acquisition cost c_i^{acq}; this objective is strongly convex and smooth with the constants below.

Curvature constants. For a normalized masked feature a with $\|a\| \le 1$, binary label y, and $p(s) = (1 + e^{-s})^{-1}$, write $\psi_y(s) = (p(s) - y)^2$. Direct differentiation gives

$$\psi_y''(s) = 2p(1-p)\{p(1-p) + (p-y)(1-2p)\}.$$

Optimizing this expression over $p \in [0,1]$ and $y \in \{0,1\}$ yields

$$-\beta_- \le \psi_y''(s) \le \beta_+, \quad \beta_- = \frac{55\sqrt{33}}{2304} - \frac{13}{768}, \quad \beta_+ = \frac{13}{768} + \frac{55\sqrt{33}}{2304}.$$

Since $\|a\| \le 1$, the data term has Hessian eigenvalues in $[-\beta_-, \beta_+]$. With the regularizer $\frac{\lambda}{2}\|w\|^2$, $\lambda = 1/4$, the objective above is therefore μ-strongly convex and L-smooth with

$$\mu = \frac{1}{4} - \beta_-, \qquad L = \frac{1}{4} + \beta_+.$$

Every pixel has exactly two coefficient copies. The Euclidean projector acts on each pair by $\frac{1}{2}\left(\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right)$, so $S_i = \frac{1}{2} I_{32}$ and the geometry-scaled gain is $\theta_i^- = \theta_i^+ = 4/(\mu + L)$.

Two oracles. The *fresh-sampling* oracle draws B training rows independently with replacement at a frozen coefficient; each primitive pair is one row's loss and gradient, and the regularizer and acquisition cost are added exactly. The loss lies in $[0,1]$, so $\sigma_i^f = 1/2$; the data-gradient norm is at most $8/27$, so the Hilbert-space Hoeffding inequality gives $\sigma_i^g = 16/27$. This is the oracle of Algorithm 1. The *full-batch* oracle evaluates each of the arm's n_i training rows once at the frozen coefficient and returns the exact empirical value and gradient before the mandatory update; it costs n_i row evaluations and has no noise. It is the exact finite-pool oracle of Corollary 4, whose phase schedule and zero-radius certificates we use unchanged.

Fresh-sampling results on digits. We ran Algorithm 1 under the fresh-sampling oracle on 16 splits of 4 purposively chosen digit pairs, with 8 oracle seeds per split (128 paths, which are not independent units). All 128/128 paths stopped correctly, at phases 23–27, with a median of 1.64×10^9 primitive pairs. The cost is large because the certified gaps between the standalone optima of the heads are small (about $5 \cdot 10^{-4}$ to $2 \cdot 10^{-3}$) relative to the worst-case proxies $\sigma_i^f = 1/2$ and $\sigma_i^g = 16/27$. Costs are computed exactly without materializing the batches: a with-replacement batch of B rows is determined by its multinomial row counts, which we sample directly. A checkpointed active-elimination variant, which uses the same sampled information, projection, step, and transitions, also stopped correctly on all paths and reduced the cost by 31.4% on average (95% CI [27.1%, 35.8%]; Table 3). These results confirm that the theory's oracle is executable on this task and that it is expensive there; this is why the comparison below uses the full-batch oracle.

Table 3: Algorithm 1 and a checkpointed active-elimination variant under the fresh-sampling oracle on 16 digit splits, with 8 oracle paths per split. Both rules use the same oracle, tied projection, step, and transitions. Split-level failures: 0/16 (two-sided 95% Wilson upper bound 19.4%). Costs count sampled primitive pairs, not runtime.

Certified stopping method	Split units	Paths/unit	Correct/stopped	Median conceptual pairs
Primary geometric certificates	16	8	128/128	1.64×10^9
Checkpointed active elimination	16	8	128/128	9.74×10^8
Paired split-level mean saving: 31.4% [27.1%, 35.8%] (95% t interval).				

Full-batch comparison. The full-batch comparison uses a fresh set of splits: all 45 unordered digit pairs with ten splits each. The one-sweep heuristic runs phase zero and returns the arm with the smallest upper endpoint without requiring separation. Both rules share the initial state, full-batch oracle, head order, step, projection, and every mandatory transition. The target is the winner among independently bracketed offline empirical standalone optima; no rule sees these brackets, and neither rule was tuned on these splits. Paired effects first average the ten splits within a digit pair and then form Student-t intervals over the 45 pair means; since all tasks come from one fixed dataset, the intervals are descriptive. The release contains all 9,900 raw timing repetitions; these are benchmark repeats, not independent units. An independent validator reconstructs the 450 splits and optimum brackets, the 900 policy paths, the 3,612 trace rows, the timing medians, the paired effects, and the generated TEX fragments.

Table 4: Tied heads on all digit pairs with the full-batch oracle. Both rules use the same oracle and transitions; "rows" counts training-row evaluations. Intervals in the text are paired 95% intervals over the 45 pair means; timing repetitions are not independent units.

Rule	Correct	Med. macros	Med. rows	Med. online ms	Mean regret
Certified sequential	450/450	8	1872	0.829	0.00000000
One-sweep incumbent	448/450	8	1872	0.778	0.00000001

Full-batch results. Certified stopping was correct on $450/450$ runs and the heuristic on $448/450$ (Table 4). Certification continued past phase zero on three runs, including the two on which the heuristic erred. Its mean overhead per digit pair was 1.00% $[-0.49\%, 2.49\%]$ in row evaluations and 9.3% $[6.9\%, 11.7\%]$ in wall-clock time. The heuristic's two errors had negligible regret, and paired intervals for the held-out loss and accuracy differences contain zero. Compared with always deploying the left head, certification reduced the mean squared test loss by 0.00147 $[0.00097, 0.00197]$ and increased accuracy by 8.3 points $[3.5, 13.2]$.

Stopping-rule ablation. An earlier complete-universe run compared sequential stopping with a comparator forced to run the same certificate rule through phase one (the depth was fixed on 90 calibration paths disjoint from the 450 report paths). Both rules were correct on all 450 paths; the sequential rule stopped at phase zero on 449 of them and used a median of 8 macros and 1872 row evaluations, versus 20 and 4680. This measures the benefit of early stopping over a fixed depth, not a selection advantage.

Full-batch results on the purposive splits. On the 16 splits of the fresh-sampling study, the full-batch rule stopped correctly in phase zero on every split with 8 macros and a median of 1,880 row evaluations (Table 5). The two oracles have different strength and cost units, so we do not report a speedup ratio.

Table 5: Full-batch oracle on the 16 digit splits of Table 3. Each macro-step evaluates the full training pool at a frozen state and makes one tied projected update. Rows are training-row evaluations; wall time is the median online loop time on the recorded platform. The two oracles differ in strength and cost units, so no speedup ratio is reported.

Method	Split units	Correct/stopped	Median rows	Macros	Median wall
Deterministic full pool (stronger oracle)	16	16/16	1,880	8	0.430 ms

In all digits experiments the standalone brackets are used only for evaluation, the tie is released before the selected head is refit, and test data are used only after selection.

B.3 Additional synthetic experiments

Table 6 summarizes the remaining experiments and states, for each, the number of objective instances and the unit used for uncertainty. Rows that share an objective instance (different policies, budgets, or cells) are correlated and are not additional independent units. The pooled Wilson bounds in the sensitivity and distinct-minimizer rows pool heterogeneous cells and are descriptive.

Table 6: Core empirical results with objective and replication accounting. "Objective" counts deterministic arm-function collections; "units" gives the independent or clustered unit used for uncertainty; parenthetical counts are correlated policy or configuration rows. Intervals are two-sided 95%. Fixed-confidence audits use nominal $\delta = 0.05$ except that the quadratic-review row combines $\delta = 0.05$ control suites with the stated multi-δ sweep. All caps are predeclared.

Objective design; uncertainty units (path rows)	Observed outcome	Protocol and interval
Observable quadratic identification 1 fixed objective; 300 CRN noise units (1,500)	4 LUCB medians 480–488 queries versus 1324 for round robin; every path stopped correctly	Sweep-LUCB paired mean query savings 95% CI [826.6, 871.0]; joint failure-rate Wilson upper 1.26%; cap 5,286
Nested fixed budget same 1 fixed objective; 1,000 nested-noise units (9,000)	Mean regret 0.1006 at $T = 8$ and 0.0000 at $T = 128$; correct rate 0.651 to 1.000	Paired mean regret decrease 95% CI [0.0916, 0.1096]
Quadratic sensitivity 5 objective specifications in 11 fixed cells; 60 noise paths/cell, 660 cell–noise units (1,980)	1,320/1,320 swept-rule paths stopped; zero wrong stop or coverage failure	Descriptive pooled audit Wilson upper 0.58% across heterogeneous fixed cells, not a common-population bound; cap 3,000
Distinct-minimizer geometric block 1,040 independently seeded objective–noise units (1,040)	Median primitive cost 128 at gap 0.10 under both couplings (200 paths each); zero theorem-audit failures	Descriptive pooled audit Wilson upper 0.37% across gaps and couplings, not a common-population bound; phase cap 18
Access-geometry conditioning 64 matched objective-instance indices across 12 cells (768)	At access floor $1/32$, descriptive median cost ratio 25.38 weak / 11.34 strong	Matched-instance bootstrap intervals for the ratio of marginal medians: [25.38, 58.00] / [11.34, 18.36]; phase cap 18; not a lower bound
Quadratic review controls suite-specific objectives; 800 digest-pair clusters (4,300)	Sweep-LUCB paired mean advantage has a positive 95% CI in 22/22 cells; zero failures	Main control suites use nominal $\delta = 0.05$; the nominal-confidence suite sweeps $\{0.02, 0.05, 0.10, 0.20\}$; scaling caps 3,000–5,000; descriptive pooled failure-rate Wilson upper 0.48% across four heterogeneous suites, not a common-population estimate
Covariance-free cross-fit 1 fixed objective; 100 noise-digest units (600)	Cross-fit medians 185/190 primitives versus 27.5 for the known-covariance trace; zero failures	Paired premium 95% CIs [152.7, 159.2] and [156.8, 165.1]; joint failure-rate upper 3.70%; cap 800 macro-queries

The distinct-minimizer experiment runs Algorithm 1 under the fresh-sampling oracle on shifted-cosine objectives with exact consensus and zero-sum projectors. The conditioning experiment regenerates objectives and places the minimizers either along weak or along strong eigendirections of the projectors, for access floors down to $1/32$ (Figure 2). Both describe the behavior of the algorithm; neither is a lower bound.

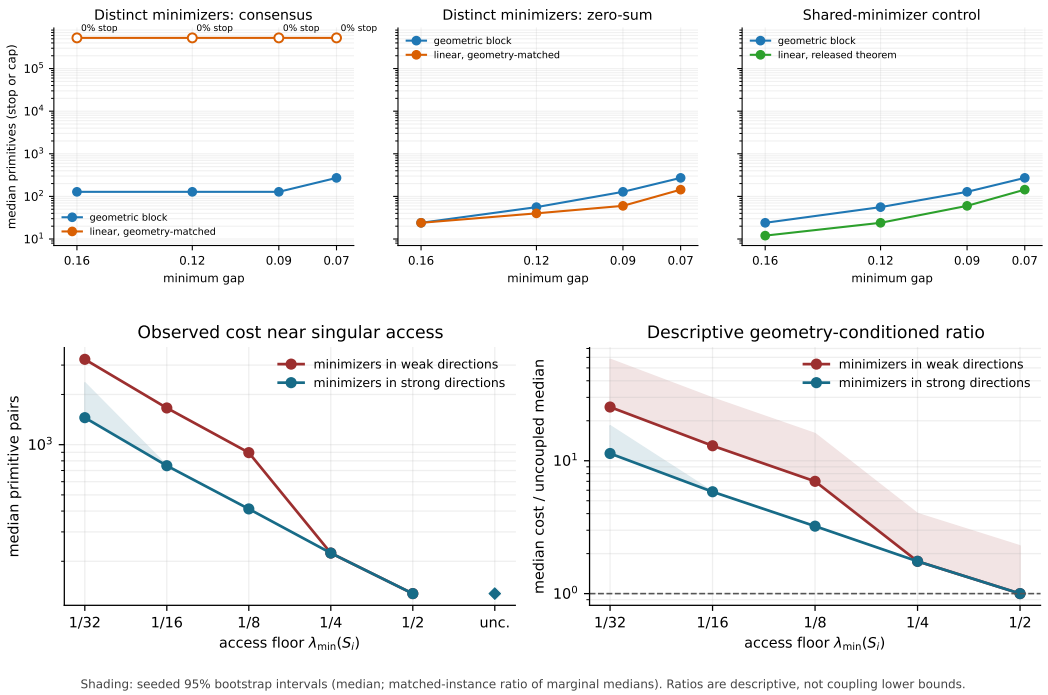


Figure 2: Additional fresh-sampling experiments for Algorithm 1. Top: cost over the predeclared gaps; comparator paths that reach the cap are shown as censored. Bottom: dependence on the access floor and on the orientation of the minimizers. Uncertainty is computed over the independent units in Table 6. Ratio bands use a matched-instance bootstrap: an objective instance is resampled once for the target and the uncoupled cell, while the noise draws of the two cells remain independent. Because costs lie on the geometric phase lattice, the percentile intervals can be asymmetric or share an endpoint with the point estimate.

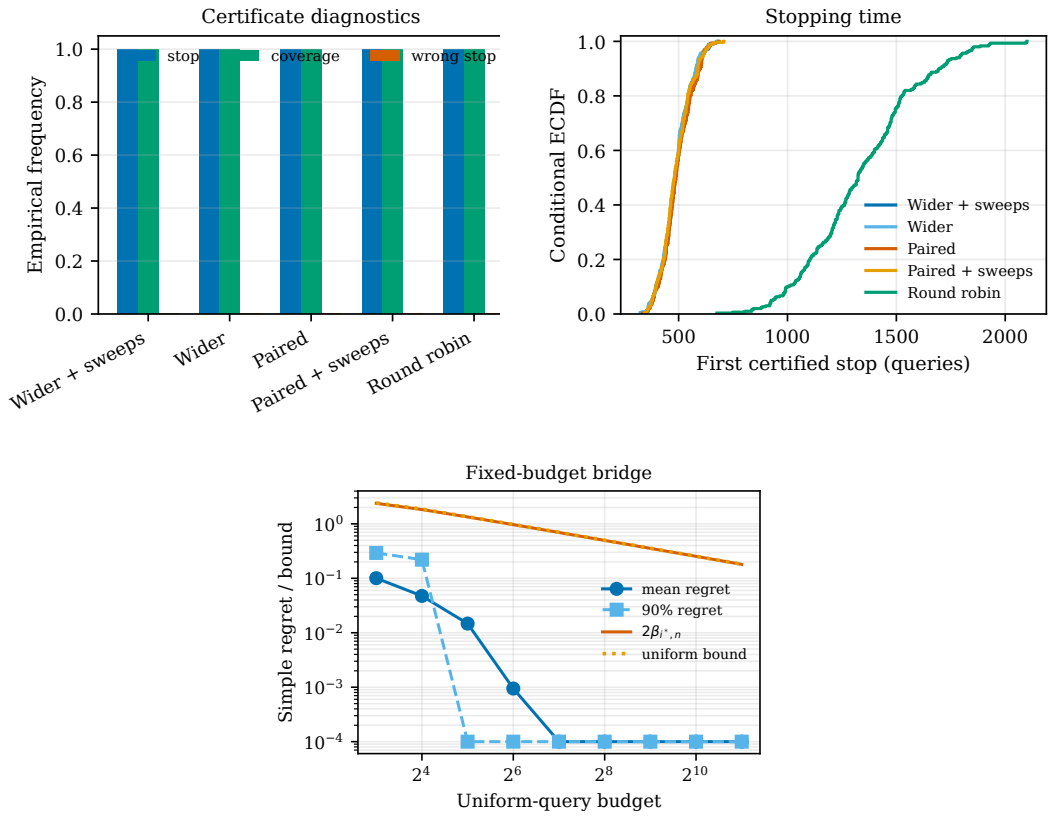


Figure 3: Known-covariance trace-corrected identification: certificate diagnostics, stopping time, and the fixed-budget guarantee of Corollary 9. Counts and intervals are in Table 6.

The quadratic experiments test Theorem 4 rather than Algorithm 1. The plotted policies in Figure 3 use the known-covariance transform; the LUCB variants differ only in their query schedule, and Theorem 4 gives the stopping bound for Sweep-LUCB. The cross-fit experiment (Figure 4) draws two complete value–gradient pairs at the same frozen state, applies one mandatory transition, and does not give the noise covariance to the policy. The known-covariance method uses strictly more information and is shown for scale only.

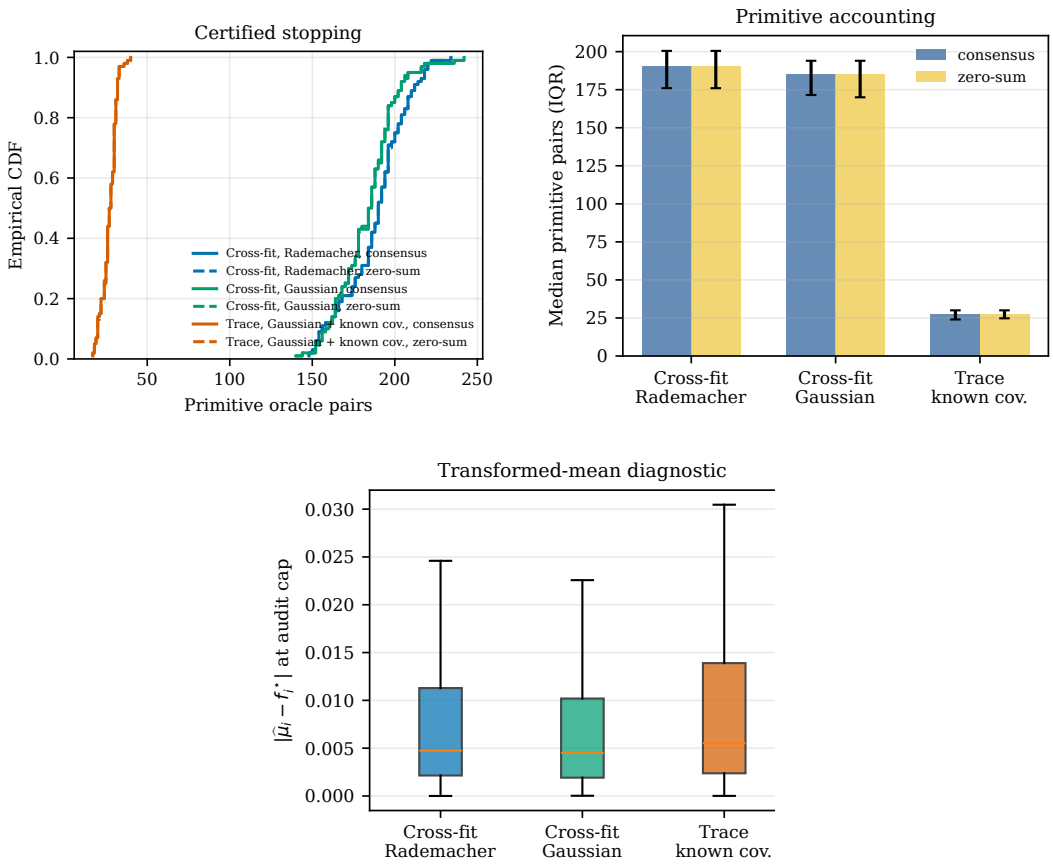


Figure 4: Same-state covariance-free cross-fitting under Gaussian and bounded non-Gaussian noise: stopping cost, primitive accounting, and transformed-mean diagnostics. The known-covariance method uses more information and is shown for scale only.

B.4 ARTIFACT MAP

Table 7: Artifacts behind each part of the paper, in the supplementary archive. Every numerical table is written by the listed generator; the archive's README gives one reproduction command per table and figure.

Paper object	Artifact	Generator / check
Fresh-sampling experiment (Figure 1, Table 2)	`supplement/results/fresh_sampling_v1`	`supplement/scripts/fresh_sampling_experiment.py`; `supplement/scripts/make_fresh_sampling_report.py`; manifest with script and data hashes
Tied-head application	`supplement/results/digits_tied_heads_competitive_full_pool_v1`; `supplement/results/digits_tied_heads_all_pairs_full_pool_v1`; `supplement/results/digits_tied_heads_full_pool_report_v1`; `supplement/results/digits_tied_heads_report_v2`; `supplement/results/digits_tied_heads_elimination_report_v1`	competitive, stopping-ablation, full-batch, and fresh-sampling generators; independent replay and accounting validators
Additional synthetic table	`supplement/generated/release_empirics_v1`	`supplement/scripts/generate_release_empirics.py`; objective, unit, and path accounting; source-hash manifest
Geometric achievability	`supplement/results/geometric_block_review_v1/report`	raw-run replay; feasibility, certificate, count, and deadline checks
Geometry conditioning	`supplement/results/geometry_conditioning_review_v1/report`	independent projector reconstruction and seed replay
Quadratic trace / cross-fit	`supplement/results/iclr_quadratic_upgrade/report`; `supplement/results/iclr_quadratic_sensitivity/report_v1`; `supplement/results/quadratic_review_upgrade/report_v3`; `supplement/results/crossfit_review_v2/report`	raw and summary reconstruction; protocol validators