Stopped Information Complexity for Stochastic Bandit Convex Optimization

Yunbei Xu
National University of Singapore
yunbei@nus.edu.sg

Abstract

The leading $\sqrt{T}$ dimension factor in stationary stochastic bandit convex optimization is known only to lie between d and $\widetilde{O}(d^{3/2})$. Standard information ratios, the algorithmic information ratio, and decision–estimation coefficients optimize one query–observation pair at a time. We introduce *stopped information complexity*, a local cost-to-progress ratio optimized over variable-length, observation-adaptive experiments. The cost charges scale-weighted query count and exploratory regret; progress credits information about a retained near-optimal action and the first constant-factor reduction in posterior Bayes risk.

The statewise and measurable-policy profiles are

$$\Lambda^{\mathrm{state}} = \sup_{s} \inf_{\mathcal{E}} \rho(s, \mathcal{E}), \qquad \Lambda^{\mathrm{policy}} = \inf_{\sigma \text{ measurable}} \sup_{s} \rho(s, \sigma(s)).$$

We prove that they coincide for Gaussian BCO whenever the loss class is Borel in the uniform topology, and for every finite stopped model, including experiments with no deterministic query cap. We also prove that the finite-horizon BCO Bayes–minimax identity continues to hold under Gaussian feedback, so worst-prior Bayes regret equals minimax regret. Under a mild constant-scale nondegeneracy condition, the common polynomial dimension exponent α_{stop} therefore exactly determines the minimax dimension exponent:

$$\beta_{\mathfrak{F}} = \frac{1 + \alpha_{\mathrm{stop}}}{2}.$$

For the dimple family of Bakhtiari, Lattimore, and Szepesvári, both one-query and fresh-uniform-direction experiments have ratio $\Omega(d^2)$, whereas an unrestricted adaptive stopped experiment attains $\widetilde{O}(d)$. Thus the known one-query obstruction is not robust to adaptive continuation. Determining whether the common stopped profile scales linearly, quadratically, or at a strictly intermediate rate in d is equivalent, at the exponent level, to resolving the remaining minimax dimension dependence.

1 Introduction

Let $C \subset \mathbb{R}^d$ be a convex body. A continuous convex loss $F : C \to [0, 1]$ is drawn once and remains fixed. At round t, a learner chooses $X_t \in C$ and observes

$$O_t = F(X_t) + \xi_t, \qquad \xi_t \overset{\text{iid}}{\sim} N(0, 1).$$

Its cumulative regret is

$$R_T = \sum_{t=1}^{T} \{F(X_t) - \min_{x \in C} F(x)\}.$$

This is stationary stochastic bandit convex optimization (BCO), or noisy zeroth-order convex optimization with cumulative rather than terminal loss. In the nontrivial long-horizon regime, the best general dimension dependence lies between $\Omega(d\sqrt{T})$ and $\widetilde{O}(d^{3/2}\sqrt{T})$ (Dani et al., 2008; Shamir, 2013; Lattimore and György, 2023; Fokkema et al., 2024). Throughout, the "full bounded convex class" means all continuous $[0,1]$-valued convex losses on C.

The Bayesian formulation has the same value as the minimax problem at every finite horizon. Minimax duality has long been used in BCO to pass from uniform prior-wise bounds to worst-case regret (Bubeck et al., 2015; Bubeck and Eldan, 2016; Lattimore and Szepesvári, 2019). Proposition 2.1 gives a direct proof for the continuous-action Gaussian model considered here:

$$\inf_{\pi} \sup_{f} \mathbb{E}_f R_T = \sup_{Q} \inf_{\pi} \mathbb{E}_Q R_T. \tag{1}$$

Here Q ranges over Borel priors on the loss class, and the policy on the right may depend on Q. Thus the main characterization can be stated directly for minimax regret.

Why one query can be too small. The information ratio compares squared conditional regret with the information in one query (Russo and Van Roy, 2016, 2018; Lattimore and György, 2021b), while the decision–estimation coefficient (DEC) balances immediate loss against statistical separation in one interaction (Foster et al., 2021, 2022, 2023). The algorithmic information ratio (AIR) gives an exact posterior-coordinate regret identity whose logarithmic potential telescopes across rounds (Xu and Zeevi, 2025). These methods are sequential, but their local optimization is over one query–observation pair and is recomputed after each observation. This distinction matters in BCO: one observation may reveal the geometry in which the next probe should be taken. Indeed, Bakhtiari et al. (2025) construct a smooth convex dimple family whose one-query coefficient is of order d^2. Such a calculation need not remain hard when later directions can depend on earlier observations.

Interactive Fano inequalities already compare the laws of complete adaptive transcripts and are therefore multistep from the outset (Chen et al., 2024), but they are lower-bound tools. Building on the AIR identity, Xu (2026a) propagates its information potential through a Bellman-sufficient continuation recursion. For a matching upper-and-lower characterization in BCO, we take the local control to be a stopped experiment that chooses its query directions, duration, and terminal action adaptively from the observations it collects.

Let $r_{\mathrm{B}}(h)$ be posterior Bayes risk after history h, and fix a scale b with $r_{\mathrm{B}}(h) \asymp b$. For the analysis, draw a near-optimal action A_b once and retain it across all later experiments at scale b. For an adaptive experiment stopped no later than the first history at which $r_{\mathrm{B}} \le \gamma b$, let Z be its complete transcript, N_Z its number of queries, R_Z its exploratory regret, and M the event that the smaller risk scale is reached. The local quantity studied in this paper is

$$\Pi_b(h; A_b) = \inf_{Z} \frac{b^2 \mathbb{E}[N_Z \mid h] + b\mathbb{E}[R_Z \mid h]}{I(A_b; Z \mid h) + d\mathbb{P}(\mathsf{M} \mid h)}. \tag{2}$$

The numerator charges samples and exploratory loss. The denominator records two forms of progress: information about a target retained across later experiments, or a constant-factor reduction in posterior risk. The mutual-information chain rule controls the first term across experiments, while a martingale argument controls repeated returns to the same risk scale.

Exact characterization. For an active posterior state s and a first-hit experiment $\mathcal{E}$, write $\rho(s, \mathcal{E})$ for the ratio in (2). The upper and lower transfers involve two quantifier orders:

$$\underbrace{\sup_{s} \inf_{\mathcal{E}} \rho(s, \mathcal{E})}_{\text{statewise profile } \Lambda^{\mathrm{state}}} \quad \le \quad \underbrace{\inf_{\sigma \text{ measurable}} \sup_{s} \rho(s, \sigma(s))}_{\text{policy profile } \Lambda^{\mathrm{policy}}}. \tag{3}$$

This is not a classical minimax gap: σ may choose a different experiment at every state. The only issue is whether these statewise choices can be made measurably. We prove that they can for the continuous Gaussian BCO model used here. We also prove equality for every finite stopped model, without imposing a common query cap; a measurable policy may choose its finite cap from the posterior state.

For every nondegenerate class sequence, meaning $\liminf_{d\to\infty} \Lambda(d,2) > 0$, the common stopped exponent satisfies

$$\boxed{\beta_{\mathfrak{F}} = \frac{1+\alpha_{\mathrm{stop}}}{2}.} \tag{4}$$

Here $\beta_{\mathfrak{F}}$ is the polynomial dimension exponent of minimax regret over polynomial horizons. Thus the growth of the stopped profile characterizes the unresolved minimax dimension dependence at the exponent level. The characterization concerns the exponent, not each finite-(d,T) regret value.

Adaptive continuation. Our second result shows that the stopped value can be polynomially smaller than its one-step restriction. On the dimple family, at $\varepsilon = d^{-1}$ and $b \asymp d^{-1}$,

$$\Pi_b^{[1]} = \Omega(d^2), \qquad \Pi_b^{\mathrm{fresh}} = \Omega(d^2), \qquad \Pi_b = \widetilde{O}(d). \tag{5}$$

Here $\Pi_b^{[1]}$ allows one query, while Π_b^{fresh} allows adaptive radii but uses a fresh uniform direction at each query. The unrestricted experiment first learns a weak direction, amplifies it by tangent finite differences, and then contracts its angular error geometrically. The one-step obstruction therefore loses a factor $\widetilde{\Omega}(d)$ once the experiment may choose later directions from earlier observations. This is a local separation, not an unrestricted minimax lower or upper bound.

Together, the two theorems reduce the minimax dimension gap to a concrete question. For the full bounded convex class, the common profile lies between linear and quadratic growth, up to logarithmic factors. Linear losses have stopped-complexity exponent one. The same holds for quadratic losses with uniformly bounded condition number, even when the center and Hessian orientation are unknown, and the apparent exponent-two obstruction on the dimple family collapses to nearly linear complexity. No example analyzed here has unrestricted stopped exponent greater than one. The central quantitative question is the growth of the common $\Lambda(d,B)$ on polynomial risk scales. A bound of order $d\,\mathrm{polylog}(dB)$ would give $\widetilde{O}(d\sqrt{T})$ minimax regret. Conversely, an order-$d^{1+\delta}$ lower bound would require a prior whose hardness survives every posterior update and would yield a stronger regret lower bound at the corresponding horizon.

Relation to prior work. Bubeck et al. (2015) used finite-horizon minimax duality to reduce adversarial BCO to a Bayesian problem and combined it with a local-to-global property of one-dimensional convex functions; Bubeck and Eldan (2016) subsequently coupled the same Bayesian reduction with multiscale exploration and the information ratio. Lattimore and Szepesvári (2019) gave a measure-theoretic minimax theorem for finite-action partial monitoring with general signal spaces. The direct Gaussian proof of Proposition 2.1 handles the continuous action space used here. For the stationary stochastic model studied here, Agarwal et al. (2013) first obtained $\widetilde{O}(\mathrm{poly}(d)\sqrt{T})$ regret; later localization and second-order methods progressively improved the dimension dependence (Lattimore and György, 2021a, 2023; Fokkema et al., 2024). Information ratios and DEC formalize the one-round comparisons summarized above. Re-solving them after each observation, together with Cauchy–Schwarz and the information chain rule, produces sequential regret bounds (Russo and Van Roy, 2016; Lattimore and György, 2021b; Foster et al., 2021, 2023). Xu and Zeevi (2025) introduced AIR and used its posterior-coordinate regret identity to turn optimized algorithmic

beliefs into frequentist algorithms. Neu et al. (2024) connect Bayesian information-directed sampling with frequentist DEC analyses in contextual bandits.

Information-ratio arguments can also be multiscale. Gouverneur et al. (2024) organize chain-link ratios across successive action quantizations and recover the tight $O(d\sqrt{T})$ Bayesian rate for linear bandits. Their construction uses fixed two-query updates; it does not optimize observation-dependent continuation and stopping within one local experiment.

The interactive Fano method has a different scope. It is a lower-bound method formulated for the laws of complete transcripts generated by arbitrary adaptive algorithms and is therefore multistep from the outset (Chen et al., 2024). A capped stopped transcript is a special case. Our distinction is consequently not between sequential and nonsequential learning theories, but between a local quantity that optimizes one BCO query and one that optimizes continuation and stopping within a variable-length scalar-feedback experiment. In the PAC/query setting, Hanneke and Wang (2025) give a complementary PAC characterization for stochastic noisy bandits and show that adaptivity can be necessary for optimal query complexity; our object instead concerns the polynomial dimension dependence of cumulative regret in BCO.

Near-optimal learning targets arise in rate–distortion and satisficing analyses (Dong and Van Roy, 2018; Russo and Van Roy, 2022; Arumugam and Van Roy, 2021), and scale-sensitive information ratios provide a related one-query localization (Bubeck and Sellke, 2023). Replacing a global class description by a target- or hypothesis-dependent quantity also appears in recent pointwise theories (Li and Xu, 2026; Xu, 2026b). In linear best-arm identification, Maiti et al. (2026) show that adaptive sampling can have a polynomial advantage when early measurements orient later ones. Our construction specializes the exact-posterior indexed recursion of Xu (2026a) to a sublevel-set target retained across experiments, with stopping at the first lower posterior-risk scale. Observation-adaptive experiment choice and stopping have a classical history in sequential design and active hypothesis testing (Chernoff, 1959; Naghshvar and Javidi, 2013). Those theories minimize sampling and declaration costs for identifying a hypothesis. Here every observation is also a regret-bearing BCO action, and progress may instead be the first reduction of posterior optimization risk.

Independent recent work of Zhang et al. (2026) proves nearly sharp lower bounds for fully adaptive algorithms in noiseless exact-value convex optimization, using a posterior-mean energy argument. Its objective is terminal accuracy and its queries carry no cumulative loss. That lower bound does not directly give a stopped-complexity or cumulative-regret lower bound, while our Gaussian cumulative-regret results do not address its noiseless exact-value model. Its posterior-stability mechanism is nevertheless relevant to the search for a superlinear stopped-profile lower bound.

Section 2 states the characterization and closes the measurable-policy issue. Appendix A proves the two regret transfers, and Appendix B gives the adaptive separation. Technical calibrations and the three-stage dimple proof are included for completeness.

2 Stopped profiles and the minimax regret exponent

Formally, a loss-class sequence $\mathfrak{F} = (\mathfrak{F}_d(C))$ assigns to each dimension d and each known compact convex body $C \subset \mathbb{R}^d$ with nonempty interior a Borel class of continuous $[0,1]$-valued convex losses. Suprema over C below range over these bodies, and $\Pi_T(C)$ denotes the randomized nonanticipating Borel policies on C for horizon T.

We first state the main result. Fix $0 < c_A < \gamma < \vartheta < 1$. At risk scale b, use the persistent target

$$A_b^\circ \mid (F = f) \sim \mathrm{Unif}\{x \in C : f(x) - \min_C f \le c_A b\}. \tag{6}$$

Its construction and information bound are proved in Lemma A.1. A state $s = (Q, b)$ is active when $\vartheta b < r_{\mathrm{B}}(Q) \le b$. An admissible experiment uses a positive, integrable number of queries and stops no later than the first posterior with Bayes risk at most γb. If Z is its transcript, N_Z its length, R_Z its exploratory regret, and M the lower-scale hit, put

$$\rho(s, \mathcal{E}) = \frac{b^2 \mathbb{E}_s N_Z + b \mathbb{E}_s R_Z}{I_s(A_b^\circ; Z) + d \mathbb{P}_s(\mathsf{M})}. \tag{7}$$

For a known body C, let $\mathsf{S}_{\mathfrak{F},C}(d, B)$ be the active states with $B^{-1} \le b \le 1$; it is standard Borel because posterior risk is Borel. Define the *statewise profile* and the *measurable-policy profile* by

$$\Lambda^{\mathrm{state}}_{\mathfrak{F},C}(d, B) = \sup_{s \in \mathsf{S}_{\mathfrak{F},C}(d,B)} \inf_{\mathcal{E}} \rho(s, \mathcal{E}), \tag{8}$$

$$\Lambda^{\mathrm{policy}}_{\mathfrak{F},C}(d, B) = \inf_{\sigma_C \text{ universally measurable}} \sup_{s \in \mathsf{S}_{\mathfrak{F},C}(d,B)} \rho(s, \sigma_C(s)), \tag{9}$$

Here σ_C is one universally measurable block policy on the augmented belief state: from the initial state s and each subsequent local history, it chooses the next action or stops; $\sigma_C(s)$ denotes the experiment that it induces from s. Since C is known, write

$$\Lambda^{\diamond}_{\mathfrak{F}}(d, B) := \sup_C \Lambda^{\diamond}_{\mathfrak{F},C}(d, B), \qquad \diamond \in \{\mathrm{state}, \mathrm{policy}\}. \tag{10}$$

Always $\Lambda^{\mathrm{state}} \le \Lambda^{\mathrm{policy}}$.

For $\diamond \in \{\mathrm{state}, \mathrm{policy}\}$, define

$$\alpha^{\diamond}_{\mathrm{stop}}(\mathfrak{F}) = \sup_{m \in \mathbb{N}} \limsup_{d \to \infty} \frac{\log\{1 + \Lambda^{\diamond}_{\mathfrak{F}}(d, d^m)\}}{\log d}. \tag{11}$$

Let

$$\mathfrak{R}^{\mathrm{M}}_{\mathfrak{F}}(d, T) := \sup_C \inf_{\pi \in \Pi_T(C)} \sup_{f \in \mathfrak{F}_d(C)} \mathbb{E}_f R_T, \tag{12}$$

$$\mathfrak{R}^{\mathrm{B}}_{\mathfrak{F}}(d, T) := \sup_C \sup_{Q \text{ on } \mathfrak{F}_d(C)} \inf_{\pi \in \Pi_T(C)} \mathbb{E}_Q R_T, \tag{13}$$

where the expectations include the Gaussian noise and the learner's randomization.

Proposition 2.1 (Bayes–minimax identity for Gaussian BCO)**.** *For every finite d, T, every compact convex body C, and every Borel class of convex losses $\mathfrak{F}_d(C) \subset \mathcal{C}(C, [0, 1])$,*

$$\inf_{\pi \in \Pi_T(C)} \sup_{f \in \mathfrak{F}_d(C)} \mathbb{E}_f R_T = \sup_{Q \text{ on } \mathfrak{F}_d(C)} \inf_{\pi \in \Pi_T(C)} \mathbb{E}_Q R_T. \tag{14}$$

Consequently, $\mathfrak{R}^{\mathrm{M}}_{\mathfrak{F}}(d, T) = \mathfrak{R}^{\mathrm{B}}_{\mathfrak{F}}(d, T)$. Moreover, the supremum over Q is unchanged if it is restricted to finitely supported priors.

Write $\mathfrak{R}_{\mathfrak{F}}$ for this common value and define

$$\beta_{\mathfrak{F}} := \sup_{M \in \mathbb{N}} \limsup_{d \to \infty} \sup_{2 \le T \le d^M} \frac{\log\{1 + \mathfrak{R}_{\mathfrak{F}}(d, T)/\sqrt{T}\}}{\log d}. \tag{15}$$

Theorem 2.2 (Two-sided stopped-complexity transfer)**.** *For every loss-class sequence* $\mathfrak{F}$*,*

$$\mathfrak{R}_{\mathfrak{F}}(d,T) \le \min\left\{T,\, \widetilde{O}\left(\sqrt{d\{1+\Lambda_{\mathfrak{F}}^{\mathrm{policy}}(d,T)\}T}\right)\right\}. \tag{16}$$

Conversely, for every $0 < \lambda < \Lambda_{\mathfrak{F}}^{\mathrm{state}}(d,B)$*, some active initial prior and scale* $B^{-1} \le b \le 1$ *satisfy, whenever* $d\lambda/b^2 \ge 8$*,*

$$n = \left\lfloor \frac{d\lambda}{4b^2} \right\rfloor, \tag{17}$$

$$\mathfrak{R}_{\mathfrak{F}}(d,n) \ge c_\gamma \sqrt{d\lambda n}. \tag{18}$$

Consequently, if $\liminf_{d\to\infty} \Lambda_{\mathfrak{F}}^{\mathrm{state}}(d,2) > 0$*, then*

$$\frac{1+\alpha_{\mathrm{stop}}^{\mathrm{state}}}{2} \le \beta_{\mathfrak{F}} \le \frac{1+\alpha_{\mathrm{stop}}^{\mathrm{policy}}}{2}. \tag{19}$$

The next result shows that the two profiles in (8)– (9) have the same value in the models studied here.

Theorem 2.3 (Measurable implementation of statewise experiments)**.** *Suppose* C *is compact and* $\mathfrak{F}_d(C)$ *is a Borel subset of* $\mathcal{C}(C,[0,1])$ *with the uniform topology. Under Gaussian feedback, for every Borel target kernel into a standard Borel space and every* $B < \infty$*,*

$$\Lambda_{\mathfrak{F},C}^{\mathrm{state}}(d,B) = \Lambda_{\mathfrak{F},C}^{\mathrm{policy}}(d,B). \tag{20}$$

The same equality holds, without a deterministic query cap, for every finite stopped model with finite hypothesis, action, and observation alphabets and a Borel target kernel into a standard Borel space. In both cases, for every $\epsilon > 0$ *there is a universally measurable policy whose experiment has a finite cap depending on the state and whose ratio is within* ϵ *of the statewise infimum at every active state.*

The full class of $[0,1]$-valued continuous convex losses is closed in $\mathcal{C}(C,[0,1])$, so Theorem 2.3 applies. Write Λ and α_{stop} for the common profile and exponent.

Theorem 2.4 (Exact minimax exponent characterization)**.** *Under the hypotheses of Theorem 2.3, if* $\liminf_{d\to\infty} \Lambda_{\mathfrak{F}}(d,2) > 0$*, then*

$$\boxed{\beta_{\mathfrak{F}} = \frac{1+\alpha_{\mathrm{stop}}}{2}.} \tag{21}$$

For the full bounded convex class and every $B \ge 2$*,*

$$cd \le \Lambda(d,B) \le Cd^2\,\mathrm{polylog}(dB), \tag{22}$$

and hence $1 \le \alpha_{\mathrm{stop}} \le 2$*.*

Proof ideas. For the upper transfer, run a stopped experiment whenever posterior risk enters a geometric scale. The near-optimal target is retained at that scale, so its information gains telescope, while optional stopping controls returns that first reach the next scale; Cauchy–Schwarz then gives the regret bound. Conversely, a low-regret policy at the calibrated horizon can itself be stopped at the first risk reduction, producing an experiment that contradicts a large statewise ratio. For measurable implementation, the capped Gaussian problem is a belief-state dynamic program whose

near-minimizers form one universally measurable block policy. Finally, prefix truncation preserves cost, hit probability, and information in the limit, removing the deterministic query cap.

For $C = B_2^d(1)$, $\theta \in S_2^{d-1}(1)$, and $s = \sqrt{1+2\varepsilon}$, define

$$F_\theta(x) = \varepsilon + \frac{1}{2}\|x\|^2 - \frac{1}{2}\big(s - \|\theta - x\|\big)_+^2. \tag{23}$$

These convex losses are uniquely minimized at θ and have minimum zero.

Theorem 2.5 (Adaptive continuation on the dimple family)**.** *At $\varepsilon = d^{-1}$ and $b = b_0 \asymp d^{-1}$ in the dimple family (23), take the initial prior $\Theta \sim \mathrm{Unif}(S_2^{d-1}(1))$ and use the persistent sublevel-set target (6). At this initial state, the one-query restriction and the class that may adapt radii but draws a fresh uniform direction at every query satisfy*

$$\Pi_b^{[1]} = \Omega_\gamma(d^2), \qquad \Pi_b^{\mathrm{fresh}} = \Omega_\gamma(d^2), \qquad \Pi_b = \widetilde{O}_\gamma(d). \tag{24}$$

Thus observation-dependent direction selection reduces this local complexity by a factor $\widetilde{\Omega}(d)$. Matching upper bounds for the two restricted classes are not claimed.

Scope. Theorem 2.4, using the transfer and implementation results in Theorems 2.2 and 2.3, characterizes the polynomial dimension exponent of minimax regret rather than each finite-(d, T) value. Measurable implementation is exact for the stated continuous Gaussian and finite models, and does not require a uniform deterministic cap across states. Proposition 2.1 supplies the value-level Bayes–minimax passage, so no separate duality condition remains in the characterization. What remains distinct is constructive: the minimax identity allows prior-dependent Bayes policies and does not by itself produce an efficient prior-free algorithm. The comparison in Theorem 2.5 concerns restricted and unrestricted local experiments; it does not by itself improve either endpoint of the full-class minimax gap.

3 Consequences and the remaining question

For the full bounded convex class, the common profile satisfies

$$cd \le \Lambda(d, B) \le Cd^2 \,\mathrm{polylog}(dB), \qquad B \ge 2.$$

The lower endpoint is attained by linear losses and by uniformly well-conditioned quadratics, even when their minimizers and Hessian orientations are unknown. The dimple family appears at first to favor the quadratic endpoint: its one-query and fresh-direction values are at least order d^2, but its unrestricted stopped value is $\widetilde{O}(d)$. No unrestricted prior with stopped exponent strictly larger than one appears among our examples.

The dimple experiment makes the role of continuation concrete. Here a is the target terminal loss and q is the current angular error:

stage	progress	(queries, regret)
orthogonal scan	find v_0 with $\langle \theta, v_0 \rangle = \widetilde{\Omega}(\max\{\sqrt{\varepsilon}, d^{-1/2}\})$	$\widetilde{O}(d^2/\varepsilon^2,\ d^2/\varepsilon)$
tangent amplification	use observation-adapted tangent differences to enter a fixed cone	$\widetilde{O}(d^2/\varepsilon^2,\ d^2/\varepsilon)$
angular contraction	invert boundary observations and repeat $q \mapsto q/2$ until $q = O(a)$	$\widetilde{O}(d^2/a^2,\ d^2/a)$

Only the last two stages choose later directions from earlier observations; this is precisely the adaptivity excluded by the fresh-direction lower bound.

The central question is therefore

$$\Lambda(d, B) \overset{?}{\leq} Cd[\log(edB)]^q \quad \text{for the full bounded convex class.} \tag{25}$$

Such a bound would give $\widetilde{O}(d\sqrt{T})$ minimax regret. Conversely, an order-$d^{1+\delta}$ lower bound must survive every adaptive first-hit experiment and every posterior it can reach; it would yield a larger regret exponent through Theorem 2.4. Whether the common stopped exponent is always one, can equal two, or can be strictly fractional is open.

The profile is invariant under invertible affine changes of variables: such a map transports the target and every stopped experiment while preserving observations, regret, information, and the first-hit event. This supports a geometric interpretation, but we avoid calling the exponent an intrinsic dimension because invariance under the fixed threshold constants and alternative admissible target kernels has not been established.

A Proof of the transfer theorem and calibrations

A.1 Bayes–minimax duality for the Gaussian model

Proof of Proposition 2.1. Let $\gamma = N(0,1)$ and $\mathsf{H}_T = (C \times \mathbb{R})^T$. Under the reference observation law γ, let $\mathcal{S}_T$ be the set of strategic measures

$$P(dh_T) = \prod_{t=1}^{T} \pi_t(dx_t \mid h_{t-1})\, \gamma(dy_t), \tag{26}$$

where the π_t are Borel kernels. This set is convex and weakly compact. Indeed, C is compact and every $P \in \mathcal{S}_T$ has $Y_{1:T}$-marginal $\gamma^{\otimes T}$, so the family is tight. It is closed because, for every bounded continuous a and b,

$$\int a(h_{t-1}, x_t) b(y_t)\, dP = \left(\int b\, d\gamma \right) \int a(h_{t-1}, x_t)\, dP, \tag{27}$$

and these identities pass to weak limits. Conversely, standard-Borel disintegration identifies every $P \in \mathcal{S}_T$ with a randomized nonanticipating Borel policy. Gaussian likelihoods are strictly positive, so policies identified outside a reference-null set have the same risk under every f.

For $f \in \mathfrak{F}_d(C)$ define

$$L_f(h_T) = \exp\left\{ \sum_{t=1}^{T} \left(y_t f(x_t) - \tfrac{1}{2} f(x_t)^2 \right) \right\}, \tag{28}$$

$$u(P, f) = \int L_f(h_T) \sum_{t=1}^{T} \{ f(x_t) - \min_C f \}\, dP(h_T). \tag{29}$$

The Gaussian change-of-measure formula gives $u(P, f) = \mathbb{E}_f^{\pi} R_T$. Moreover, $u(\cdot, f)$ is affine and continuous on $\mathcal{S}_T$, while $u(P, \cdot)$ is continuous in the uniform topology. To see this, note that the integrand is continuous and bounded by

$$T \exp\left(\sum_{t=1}^{T} |y_t| \right),$$

which is uniformly integrable over $\mathcal{S}_T$ because all its members have the same Gaussian $Y_{1:T}$-marginal.

Let $\mathcal{Q}_0$ be the convex set of finitely supported priors on $\mathfrak{F}_d(C)$ and put $u(P, Q) = \int u(P, f) Q(df)$. Sion's theorem (Sion, 1958), applied to the compact convex set $\mathcal{S}_T$ and the convex set $\mathcal{Q}_0$, yields

$$\min_{P \in \mathcal{S}_T} \sup_f u(P, f) = \sup_{Q \in \mathcal{Q}_0} \min_{P \in \mathcal{S}_T} u(P, Q). \tag{30}$$

The left side is minimax regret, and the right side is worst Bayes regret over finitely supported priors. Allowing all Borel priors cannot decrease the latter value and, by weak duality, cannot increase it beyond the left side. This proves (14) and the finite-support assertion. □

A.2 An affine-covariant near-optimal target

Assume that every loss in the prior has range at most $0 < \Delta \le 1$ on C. The analysis needs a near-optimal random index whose information can telescope without depending on an arbitrary choice among multiple minimizers. For $0 < \eta < \Delta$, put

$$S_\eta(f) = \{x \in C : f(x) - \min_C f \le \eta\}, \qquad A_\eta^\circ \mid (F = f) \sim \mathrm{Unif}(S_\eta(f)). \tag{31}$$

The conditional draw uses private randomness unavailable to the policy. Near-optimal random learning targets are standard in rate–distortion and satisficing analyses (Dong and Van Roy, 2018; Arumugam and Van Roy, 2021; Russo and Van Roy, 2022). The role here is more specific: the target is affine covariant, retained across repeated local experiments, and paired with a posterior-risk hitting event.

Lemma A.1 (Sublevel-set near-optimal target)**.** *Equip the continuous losses on C with the uniform topology. The map $(f, \eta) \mapsto \mathrm{Unif}(S_\eta(f))$ is a Borel probability kernel on the loss space times $(0, \Delta)$. In particular, for every fixed η, the coupling (31) satisfies*

$$F(A_\eta^\circ) - F_C^\star \le \eta \quad \text{almost surely}, \qquad I(F; A_\eta^\circ) \le d \log \frac{\Delta}{\eta}. \tag{32}$$

If T is an invertible affine map, then the sublevel-set target for $f \circ T^{-1}$ on $T(C)$ has the law of $T(A_\eta^\circ)$. Thus the channel is affine covariant and independent of a minimizer representation or of the prior on F.

Proof. The graph of the correspondence $(f, \eta) \mapsto S_\eta(f)$ is Borel because $f \mapsto \min_C f$ is continuous in the uniform norm and $(f, x) \mapsto f(x)$ is continuous. More explicitly, $(f, \eta, x) \mapsto \mathbf{1}\{x \in S_\eta(f)\}$ is a bounded jointly Borel function. Its integral against Lebesgue measure is therefore Borel in (f, η), and the same parameter-integration argument, now treating (f, η) jointly as the parameter and multiplying by the indicator of an arbitrary Borel set in C, proves that $(f, \eta) \mapsto \mathrm{Unif}(S_\eta(f))$ is a Borel probability kernel once positivity of the denominator is shown.

Fix any $y \in \arg\min_C f$ and put $\alpha = \eta/\Delta$. For every $x \in C$, convexity gives

$$f((1-\alpha)y + \alpha x) \le (1-\alpha) f(y) + \alpha f(x) \le f_C^\star + \eta.$$

Consequently, $(1-\alpha)y + \alpha C \subseteq S_\eta(f)$ and $\mathrm{vol}(S_\eta(f)) \ge \alpha^d \,\mathrm{vol}(C) > 0$. The support definition gives the near-optimality statement. Using the uniform law on C as a reference,

$$I(F; A_\eta^\circ) \le \mathbb{E} D_{\mathrm{KL}}(P_{A_\eta^\circ \mid F} \,\|\, \mathrm{Unif}(C)) = \mathbb{E} \log \frac{\mathrm{vol}(C)}{\mathrm{vol}(S_\eta(F))} \le d \log(1/\alpha).$$

Finally, invertible affine maps carry sublevel sets to sublevel sets and uniform measures to uniform measures, proving covariance. □

For the reduction, fix a tolerance η and write $A = A_\eta^\circ$. On an auxiliary probability space, conditionally on F, draw the countable family of bucket targets $A_j \sim A_{c_A b_j}^\circ$, for a fixed $c_A \in (0, 1)$, before the interaction begins, using independent private coins. The learner never observes these coins. Target A_j is activated for the analysis when bucket j is first entered and is retained across every later return to that bucket. This ex-ante coupling lets its information drift telescope through the random bucket-entry history.

Let h be a reached history and write

$$r_{\mathrm{B}}(h) = \inf_{x \in C} \mathbb{E}[F(x) - F_C^\star \mid h], \qquad \Phi(h) = I(F; A \mid h). \tag{33}$$

Throughout, conditional information at a random history is interpreted statewise: $I(A; Z \mid H)$ denotes the H-measurable function which, at $H = h$, equals $D_{\mathrm{KL}}(P_{A,Z|h} \| P_{A|h} \otimes P_{Z|h})$. An admissible experiment has a complete, almost surely finite stopped transcript Z, including its length, actions, observations, and private coins. Its actions cannot use the hidden auxiliary randomization in A. This operational restriction ensures the Markov relation

$$A \perp Z \mid (F, h). \tag{34}$$

In particular, after every history generated without access to the target's private coin, the conditional law of A given F remains its original kernel $K_b(\cdot \mid F)$.

Lemma A.2 (Exact information drift)**.** *Every admissible experiment satisfies*

$$\Phi(h) - \mathbb{E}[\Phi(h, Z) \mid h] = I(A; Z \mid h). \tag{35}$$

Consequently, along any adaptive sequence of admissible blocks,

$$\mathbb{E} \sum_k I(A; Z_k \mid H_k) \le I(F; A). \tag{36}$$

Proof. Expand $I(A; F, Z \mid h)$ in two orders and use $I(A; Z \mid F, h) = 0$:

$$I(A; F \mid h) = I(A; Z \mid h) + \mathbb{E}[I(A; F \mid h, Z) \mid h].$$

Iterating this identity proves (36); the remaining conditional mutual information is nonnegative. □

The information term $I(A; Z \mid h)$ is the posterior average of the coordinate log gain underlying AIR (Xu and Zeevi, 2025). Under the Markov relation (34), the potential $I(F; A \mid h)$ has exactly this fixed-index drift, as used by Xu (2026a). The BCO-specific step is to couple the drift to a near-optimal target and the first hitting time of a lower posterior-risk level.

A frequentist high-probability recommendation is not automatically a posterior guarantee: the bound must hold under the terminal posterior on every accepted transcript. The following elementary acceptance rule enforces this condition.

Lemma A.3 (Posterior-risk acceptance)**.** *Suppose a stopped routine returns $\widehat{x} = \widehat{x}(Z)$ and*

$$\mathbb{E}[F(\widehat{x}) - F_C^\star \mid h] \le ab.$$

For $\gamma > a$, accept the recommendation only on

$$\mathsf{M} = \{\mathbb{E}[F(\widehat{x}) - F_C^\star \mid h, Z] \le \gamma b\}.$$

Then every accepted transcript has terminal posterior regret at most γb, and

$$\mathbb{P}(\mathsf{M} \mid h) \ge 1 - a/\gamma.$$

Proof. The terminal posterior regret is nonnegative and has expectation at most ab by the tower property. Apply Markov's inequality. □

The lemma validates accepted transcripts, but does not pay for rejected ones. A complete experiment must still charge its entire expected query and regret cost.

Whenever a capped routine below is analyzed through this rule, we couple it to its full scheduled run but execute its first-hit truncation: after each scalar observation, stop as soon as $r_{\mathrm{B}} \le \gamma b$ and return a posterior Bayes action. The gated terminal event implies that a hit occurred by the cap, while truncation only decreases nonnegative query and regret costs. Thus all explicit linear and dimple bounds below concern the first-hit complexity in Definition A.4.

A.3 Stopped information complexity

Fix constants

$$0 < \gamma < \vartheta < 1 \tag{37}$$

and call (h, b) *active* when

$$\vartheta b < r_{\mathrm{B}}(h) \le b. \tag{38}$$

For an experiment Z, let N_Z be its length and

$$R_Z = \sum_{t=1}^{N_Z} \{F(X_t) - F_C^\star\}.$$

The experiment stops when its goal has already been reached. Starting from $H_0 = h$, let

$$\tau_b = \inf\{t \ge 1 : r_{\mathrm{B}}(H_t) \le \gamma b\}. \tag{39}$$

A *first-hit experiment* chooses a positive adapted stopping time σ, stops at $\tau = \sigma \wedge \tau_b$, and has a complete transcript $Z = H_\tau \setminus H_0$. It declares $\mathsf{M} = \{\tau_b \le \sigma\}$ and, on M, returns a measurable posterior Bayes action x_Z. Thus

$$\mathbb{E}[F(x_Z) - F_C^\star \mid h, Z] = r_{\mathrm{B}}(h, Z) \le \gamma b \quad \text{on } \mathsf{M}. \tag{40}$$

On M^c the experiment stops before the posterior reaches the smaller scale. No pathwise charge is assigned to either event; the inequalities below control the entire experiment in expectation. The first-hit convention prevents an experiment from continuing after solution merely to harvest target information that the regret problem no longer needs.

The information and risk-reduction terms combine into one statewise quantity.

Definition A.4 (Stopped information complexity)**.** *Fix an active state (h, b, A) satisfying (38), where A is the persistent near-optimal target sampled once for this geometric scale. Let $\mathcal{B}(h, b, A)$ be the collection of admissible positive-length, almost surely finite first-hit experiments (Z, M, x_Z) described around (39), with $\mathbb{E}[N_Z \mid h] < \infty$. Define*

$$\Pi_b(h; A) = \inf_{(Z,\mathsf{M},x_Z)\in\mathcal{B}(h,b,A)} \frac{b^2\mathbb{E}[N_Z \mid h] + b\mathbb{E}[R_Z \mid h]}{I(A; Z \mid h) + d\mathbb{P}(\mathsf{M} \mid h)}. \tag{41}$$

The ratio is $+\infty$ when its denominator is zero. The target's private randomization remains inaccessible to the policy, and the complete transcript includes the stopping time, actions, observations, and policy coins. All complexities below use this first-hit convention.

Definition A.4 is a stopped analogue of the one-round tradeoff measured by localized and constrained variants of the decision–estimation coefficient (Foster et al., 2021, 2023). It replaces one randomized query by a variable-length BCO experiment whose continuation rule and stopping time are optimized jointly. It also retains a near-optimal target across repeated experiments and counts the first hit of a lower Bayes-risk level as progress. The coordinate log identity underlying AIR supplies the telescoping algebra (Xu and Zeevi, 2025); its fixed-index Bellman extension (Xu, 2026a) supplies the continuation recursion.

Proposition A.5 (One query and adaptive continuation)**.** *Let $\mathcal{B}_1(h, b, A)$ be the subclass in Definition A.4 with $N_Z = 1$ almost surely, and let $\Pi_b^{[1]}(h; A)$ be the corresponding restricted infimum. Then*

$$\Pi_b(h; A) \le \Pi_b^{[1]}(h; A). \tag{42}$$

For a one-query law ν whose posterior remains active almost surely (so $\mathsf{M} = \varnothing$), write

$$i_\nu = I(A; X, O \mid h), \qquad \ell_\nu = \mathbb{E}[F(X) - F_C^\star \mid h], \qquad \Gamma_\nu = \frac{\ell_\nu^2}{i_\nu}. \tag{43}$$

When $i_\nu > 0$,

$$\frac{b^2 + b\ell_\nu}{i_\nu} = \Gamma_\nu \left\{ \left(\frac{b}{\ell_\nu} \right)^2 + \frac{b}{\ell_\nu} \right\}. \tag{44}$$

Hence, at a state satisfying (38), the two ratios are comparable up to constants on query laws with $\ell_\nu \le Kb$, for fixed $K < \infty$. They need not be comparable for high-loss probes.

Assume $I(F; A \mid h) < \infty$. For $\lambda > 0$, define the one-query offset

$$\mathfrak{D}_{b,\lambda}^{[1]}(h; A) = \inf_{(Z, \mathsf{M}, x_Z) \in \mathcal{B}_1(h, b, A)} \left[b^2 + b\mathbb{E}[R_Z \mid h] - \lambda \{ I(A; Z \mid h) + d\mathbb{P}(\mathsf{M} \mid h) \} \right]. \tag{45}$$

Then

$$\Pi_b^{[1]}(h; A) \le \lambda \quad \Longleftrightarrow \quad \mathfrak{D}_{b,\lambda}^{[1]}(h; A) \le 0. \tag{46}$$

For finite problems and experiments capped at m queries, iterating the same offset with its posterior continuation gives the standard information-penalized optimal-stopping recursion. Thus Π_b is the adaptive optimal-stopping extension of the one-query cost–information tradeoff.

Proof. The class inclusion gives (42), and (44) is algebra. Under (38), every query law has $\ell_\nu \ge r_{\mathrm{B}}(h) > \vartheta b$; together with $\ell_\nu \le Kb$, this bounds the factor in braces above and below by positive constants depending only on K and ϑ. Finally, write $c(Z) = b^2 + b\mathbb{E}R_Z$ and $g(Z) = I(A; Z) + dP(\mathsf{M})$. Data processing gives the uniform bound $g(Z) \le I(F; A \mid h) + d < \infty$. Hence the infimum of $c(Z)/g(Z)$ is at most λ exactly when the infimum of $c(Z) - \lambda g(Z)$ is nonpositive, including the boundary case in which the ratio approaches λ without attaining it. The convention for $g(Z) = 0$ is harmless because $c(Z) \ge b^2 > 0$. This proves (46). □

One-round coefficients and stopped experiments. For one randomized BCO query, the information ratio is ℓ_ν^2/i_ν. Offset DEC and constrained regret–DEC likewise optimize one randomized decision and its induced observation, balancing immediate loss against statistical separation from a reference model. The associated algorithms re-solve these one-round problems after each update. Here one round means one native BCO query–observation pair; treating a stopped policy as one abstract decision merely moves the continuation problem into the action space. In the present scalar-feedback model, Π_b instead optimizes every query, adaptation, and stopping decision within a variable-length experiment. We claim no general numerical equivalence between Π_b, DEC, and the information ratio. Equation (44) is the exact algebraic relation for each active information-only one-query law; the stopped value is governed by the optimal-stopping recursion above.

A.4 Upper and lower transfers

Theorem A.6 (From local experiments to global regret)**.** *Fix $0 < c_A < \gamma$. Conditionally on F, draw independent private targets $A_j = A^\circ_{c_A b_j}$ for the geometric risk scales $b_j = \Delta\vartheta^j$ before interaction, and retain A_j across every return to scale j. Suppose that, at every active state, a universally measurable rule selects a first-hit experiment whose ratio in (41) is at most Λ_d, uniformly over scales*

and posterior states. Then, for every $0 < \eta < \Delta/2$,

$$\mathrm{BReg}_T \le C_{\gamma,\vartheta,c_A}\Lambda_d \frac{d\{1+\log(\Delta/\eta)\}}{\eta} + 2T\eta, \tag{47}$$

$$\mathbb{E}N_{\mathrm{explore}} \le C_{\gamma,\vartheta,c_A}\Lambda_d \frac{d\{1+\log(\Delta/\eta)\}}{\eta^2}. \tag{48}$$

Consequently,

$$\mathrm{BReg}_T = \widetilde{O}(\sqrt{d\Lambda_d T}). \tag{49}$$

Thus $\Lambda_d = \widetilde{O}(d^{2-\kappa})$ *gives* $\widetilde{O}(d^{(3-\kappa)/2}\sqrt{T})$.

Proof. Invoke the selected experiment whenever $r_{\mathrm{B}} > 2\eta$ and play a posterior Bayes action otherwise. The ratio bound implies, at scale b_j,

$$\mathbb{E}[R_Z \mid h] \le \frac{\Lambda_d}{b_j}\{I(A_j; Z \mid h) + d\mathbb{P}(\mathsf{M} \mid h)\},$$

and the analogous query bound with b_j^{-2}. For each fixed j, apply Lemma A.2 to the complete block sequence with target A_j and discard the nonnegative drift terms outside bucket j. This gives

$$\mathbb{E}\sum_{k:j(k)=j} I(A_j; Z_k \mid H_k) \le I(F; A_j).$$

Lemma A.1 bounds the right side by $d\log\{\Delta/(c_A b_j)\}$. Geometric summation down to scale 2η therefore bounds the information terms by the displayed logarithmic multiples of d/η and d/η^2.

It remains to count risk-reduction outcomes. Condition on one such outcome at scale b_j, and let x be the returned action. At every later history,

$$q_t = \mathbb{E}[F(x) - F_C^\star \mid H_t]$$

is a bounded nonnegative martingale and initially satisfies $q_0 \le \gamma b_j$. Returning to the same bucket requires $r_{\mathrm{B}}(H_t) > \vartheta b_j$, hence $q_t > \vartheta b_j$. Optional stopping gives conditional return probability at most γ/ϑ. Thus the expected number of risk-reduction outcomes in one bucket is at most $(1-\gamma/\vartheta)^{-1}$. Summing b_j^{-1} and b_j^{-2} over the geometric buckets gives $O(\eta^{-1})$ and $O(\eta^{-2})$, respectively. Below scale 2η, a posterior Bayes action incurs conditional expected regret at most 2η per round. If the horizon cuts an experiment short, extend that experiment only for the analysis; its actual prefix has no larger nonnegative query or regret cost. Combining these bounds proves (47)–(48); optimizing in η proves (49). □

The same ratio turns a local bound that holds for every stopped experiment into a Bayes-regret lower bound. This converse is why the lower-bound direction of the core question is stronger than disproving one sufficient condition. For an initial law Q, write $\Pi_b(Q; A)$ for the stopped complexity at the empty history.

Proposition A.7 (From stopped complexity to a regret lower bound)**.** *Suppose an* initial *prior, a target coupling fixed independently of the policy under test, an active scale* b, *and* $\lambda > 0$ *satisfy*

$$\Pi_b(Q; A) \ge \lambda. \tag{50}$$

Then, for every integer $n \ge 1$ *and every* $0 < a < \gamma$, *every* n*-round adaptive policy satisfies*

$$\mathrm{BReg}_n \ge \min\left\{abn, \left[\frac{\lambda d(1-a/\gamma)}{b} - bn\right]_+\right\}. \tag{51}$$

In particular, if $d\lambda/b^2 \geq 8$, then at

$$n = \left\lfloor \frac{d\lambda}{4b^2} \right\rfloor \tag{52}$$

every n-round adaptive policy has

$$\mathrm{BReg}_n \geq c_\gamma \frac{d\lambda}{b} = \Omega_\gamma(\sqrt{d\lambda n}). \tag{53}$$

Because this is a Bayes lower bound for an initial prior, it immediately lower-bounds the minimax regret of every loss class containing $\mathrm{supp}(Q)$*. At a reached state the same argument gives only a conditional continuation bound unless the state is made initial or is reached with controlled probability.*

Proof. Fix a policy and let $R_n = \mathbb{E}\sum_{t=1}^n \{F(X_t) - F_C^\star\}$. If $R_n > abn$, the first term in (51) applies. Otherwise put $\bar{X} = n^{-1}\sum_{t=1}^n X_t$. Convexity gives

$$\mathbb{E}[F(\bar{X}) - F_C^\star] \leq R_n/n \leq ab.$$

Let

$$q_n = \mathbb{E}[F(\bar{X}) - F_C^\star \mid H_n].$$

Markov's inequality gives $\mathbb{P}(q_n \leq \gamma b) \geq 1 - a/\gamma$. On this event $r_{\mathrm{B}}(H_n) \leq q_n$, so the first-hit time $\tau_b = \inf\{t \geq 1 : r_{\mathrm{B}}(H_t) \leq \gamma b\}$ is at most n. Couple the policy to the first-hit block with $\tau = n \wedge \tau_b$ and $\mathsf{M} = \{\tau_b \leq n\}$. Nonnegativity of regret gives $\mathbb{E}\tau \leq n$ and $\mathbb{E}R_\tau \leq R_n$, while $\mathbb{P}(\mathsf{M}) \geq 1 - a/\gamma$. Hence (50) gives

$$b^2 n + bR_n \geq b^2\mathbb{E}\tau + b\mathbb{E}R_\tau \geq \lambda\{I(A; Z_\tau) + d\mathbb{P}(\mathsf{M})\} \geq \lambda d(1 - a/\gamma),$$

which proves (51). Take $a = \gamma/2$ and then (52); the assumption $d\lambda/b^2 \geq 8$ absorbs the floor and gives (53). □

A.5 Linear and quadratic calibrations

The next proposition calibrates the smallest possible universal stopped complexity. It controls arbitrary adaptive measurements, including information about joint parities or other nonlinear functions of the hidden signs.

Proposition A.8 (A linear lower bound)**.** *Let $C = B_2^d(1)$, let $0 < b \leq 1/2$, draw $V \sim \mathrm{Unif}\{-1, 1\}^d$, put $\theta_V = V/\sqrt{d}$, and define*

$$F_V(x) = b\{1 - \langle \theta_V, x\rangle\}. \tag{54}$$

Write Q_{lin} for this prior. Then $F_V \in [0, 2b]$, $F_V^\star = 0$, and the initial Bayes risk is b. For every randomized adaptive stopped transcript Z with random length N_Z satisfying $\mathbb{E}N_Z < \infty$, and every target A obtained from (V, U) with private randomization U conditionally independent of Z given V,

$$I(A; Z) \leq I(V; Z) \leq \frac{b^2}{d}\mathbb{E}N_Z. \tag{55}$$

If (Z, M, x_Z) satisfies the posterior risk bound at level γb, then

$$\mathbb{E}N_Z \geq c_\gamma \left\{ \frac{d}{b^2} I(A; Z) + \frac{d^2}{b^2}\mathbb{P}(\mathsf{M}) \right\}. \tag{56}$$

Consequently

$$\Pi_b(Q_{\mathrm{lin}}; A) \geq c_\gamma d. \tag{57}$$

Consequently every universal stopped-complexity upper bound must be at least of order d.

Proof. Write P_v for the law of the complete stopped transcript under $V = v$, let μ be uniform on the cube, and let $\overline{P} = 2^{-d}\sum_v P_v$. The tensorized two-point entropy inequality states that, for every positive f on the cube,

$$\mathrm{Ent}_\mu(f) \le \frac{1}{4}\sum_{i=1}^{d} \mathbb{E}_\mu[(f(V) - f(V^{(i)}))(\log f(V) - \log f(V^{(i)}))],$$

where $V^{(i)}$ flips coordinate i. Apply it pointwise to $f_z(v) = dP_v/d\overline{P}(z)$ after replacing f_z by $f_z + \varepsilon$, and let $\varepsilon \downarrow 0$; this handles densities that vanish. Integrating in z gives

$$I(V;Z) \le \frac{1}{4}\,2^{-d}\sum_{v,i}\{D_{\mathrm{KL}}(P_v\|P_{v^{(i)}}) + D_{\mathrm{KL}}(P_{v^{(i)}}\|P_v)\}. \tag{58}$$

For adaptive Gaussian observations, apply the finite-horizon chain rule to $N_Z \wedge m$ and let $m \to \infty$ by monotone convergence. This yields

$$D_{\mathrm{KL}}(P_v\|P_{v^{(i)}}) = \frac{2b^2}{d}\mathbb{E}_v \sum_{t\le N_Z} X_{t,i}^2.$$

Average this identity in (58) and use $\|X_t\|_2 \le 1$ to obtain the second inequality in (55); the first is data processing.

On a terminal transcript $z \in \mathsf{M}$, let $m_z = \mathbb{E}[\theta_V \mid Z = z]$. The posterior risk bound gives

$$1 - \langle m_z, x_Z\rangle \le \gamma, \qquad \|m_z\|_2 \ge 1 - \gamma.$$

The law of θ_V is $1/d$-sub-Gaussian in every unit direction. The variational formula for relative entropy therefore gives

$$D_{\mathrm{KL}}(P_{V|Z=z}\|\mu) \ge \frac{d}{2}\|m_z\|_2^2.$$

Averaging over $z \in \mathsf{M}$ shows

$$I(V;Z) \ge \frac{d}{2}(1-\gamma)^2\mathbb{P}(\mathsf{M}). \tag{59}$$

Combine (55) and (59). They respectively bound $dI(A;Z)$ and $d^2\mathbb{P}(\mathsf{M})$ by constant multiples of $b^2\mathbb{E}N_Z$, proving (56) and then (57). □

The lower endpoint is attained throughout the unit-ball linear class, not only on the preceding prior.

Proposition A.9 (Stopped complexity of linear losses)**.** *Fix $0 < \beta \le 1/2$ and consider*

$$F_\theta(x) = \beta\{1 - \langle\theta, x\rangle\}, \qquad x \in B_2^d(1), \quad \theta \in S_2^{d-1}(1). \tag{60}$$

For every posterior Q on Θ, every active risk scale a with $\vartheta a < r_{\mathrm{B}}(Q) \le a$, and every admissible target coupling A, one has

$$\Pi_a(Q;A) \le C_{\gamma,\vartheta}d. \tag{61}$$

Together with Proposition A.8, this shows that the worst stopped complexity within the class (60) is $\Theta_{\gamma,\vartheta}(d)$ up to universal constants. Thus a superlinear obstruction, if one exists, must use structure absent from this fixed-amplitude Euclidean class.

Proof. Put $u = \min\{a, \beta\}$ and

$$m = \left\lceil C_0 \frac{d}{\beta u} \right\rceil,$$

where $C_0 = C_0(\gamma)$ is chosen below. Query each coordinate action e_i exactly m times. If $\bar{O}_i$ is its averaged observation, set

$$\widetilde{\theta}_i = 1 - \bar{O}_i/\beta, \qquad \widehat{x} = \frac{\widetilde{\theta}}{\|\widetilde{\theta}\|_2},$$

with an arbitrary definition on the null event $\widetilde{\theta} = 0$. Conditionally on every θ,

$$\mathbb{E}_\theta \|\widetilde{\theta} - \theta\|_2^2 = \frac{d}{m\beta^2}.$$

For every nonzero vector y and unit vector θ,

$$\left\| \frac{y}{\|y\|_2} - \theta \right\|_2 \le 2\|y - \theta\|_2,$$

because $|\|y\|_2 - 1| \le \|y - \theta\|_2$. Therefore, uniformly in θ,

$$\mathbb{E}_\theta F_\theta(\widehat{x}) = \frac{\beta}{2} \mathbb{E}_\theta \|\widehat{x} - \theta\|_2^2 \le \frac{2d}{m\beta} \le \frac{2u}{C_0}.$$

Choose $C_0 \ge 8/\gamma$. Under every posterior the returned action has expected regret at most $\gamma a/4$. Posterior-risk acceptance at scale a therefore gives an event G of probability at least $3/4$ on which $\widehat{x}$ has posterior risk at most γa. Under the first-hit convention, $\mathsf{G} \subseteq \mathsf{M}$, so $\mathbb{P}(\mathsf{M}) \ge 3/4$ and actual prefix costs are no larger than the caps below.

The block has deterministic length $N = dm$, and every query has regret at most 2β, so $R_Z \le 2\beta N$. If $a \le \beta$, then

$$a^2 N + a\mathbb{E}R_Z \le C_1 d^2.$$

If $a > \beta$, the origin has regret exactly β, so activity implies $\beta \ge r_{\mathrm{B}}(Q) > \vartheta a$. The same display then holds with C_1 enlarged by a factor depending only on ϑ. Since the denominator of (41) is at least $d\mathbb{P}(\mathsf{M}) \ge 3d/4$, the block proves (61). The lower bound follows by taking Q to be the Rademacher prior of Proposition A.8 and $a = \beta$. □

The linear calculation is not an artifact of a flat loss surface. The same exponent persists when both the minimizer and the orientation of a curved loss are unknown. This is a calculation of the stopped profile, not the first favorable-regime $d\sqrt{T}$ result for curved losses: Ito (2020) obtained optimal dimension dependence for smooth, strongly convex online BCO under an interior-minimizer condition.

Proposition A.10 (Uniformly well-conditioned quadratics have exponent one)**.** *Fix $R \ge 1$ and, on $C = B_2^d(1)$, consider*

$$\mathfrak{Q}_R = \left\{ F_{H,y}(x) = \frac{1}{4R}(x - y)^\top H (x - y) : I \preceq H \preceq RI,\ \|y\|_2 \le \frac{1}{4} \right\}. \tag{62}$$

Every member of $\mathfrak{Q}_R$ takes values in $[0, 25/64]$. For every posterior on $\mathfrak{Q}_R$, every active scale a, and every admissible private target coupling A,

$$\Pi_a(Q; A) \le C_{R,\gamma,\vartheta} d. \tag{63}$$

At a constant active scale, a shifted-hypercube subfamily satisfies, for every such target,

$$\Pi_a(Q; A) \geq c_{R,\gamma} d.$$

Consequently, for every $B \geq 64R$, both profiles of $\mathfrak{Q}_R$ are $\Theta_R(d)$. In particular, both stopped-complexity exponents equal one.

Proof. Write $F(x) = \frac{1}{2}(x-y)^\top B(x-y)$, where $\mu I \preceq B \preceq LI$ with $\mu = (2R)^{-1}$ and $L = 1/2$. Project all iterates onto $B_2^d(1/2)$, start from zero, and put $h = 1/4$. At iteration k, query $x_k \pm he_i$ once for every coordinate and use the central differences as a gradient estimate $\widehat{g}_k$. All query points lie in C, and quadraticity gives

$$\mathbb{E}[\widehat{g}_k \mid x_k, F] = \nabla F(x_k), \qquad \mathbb{E}[\|\widehat{g}_k - \nabla F(x_k)\|_2^2 \mid x_k, F] = 8d.$$

Projected stochastic gradient descent with $\eta_k = 2/\{\mu(k+k_0)\}$ and sufficiently large $k_0 = k_0(R)$ obeys the standard strongly-convex recursion

$$\mathbb{E}\|x_k - y\|_2^2 \leq C_R \min\left\{1, \frac{d}{k+k_0}\right\}.$$

Indeed, nonexpansiveness of projection, strong convexity, and $\|\nabla F(x)\| \leq L\|x-y\|$ give this bound by induction from

$$\mathbb{E}[\|x_{k+1} - y\|^2 \mid x_k, F] \leq (1 - 2\mu\eta_k + L^2\eta_k^2)\|x_k - y\|^2 + 8d\eta_k^2.$$

Choose k_0 so that $L^2\eta_k^2 \leq \mu\eta_k/2$. Writing $s_k = \mathbb{E}\|x_k - y\|^2$, the preceding display becomes

$$s_{k+1} \leq \left(1 - \frac{3}{k+k_0}\right) s_k + \frac{C_R d}{(k+k_0)^2}.$$

The claim follows by induction with $s_k \leq C_R d/(k+k_0)$ after enlarging C_R; the trivial bound $s_k \leq (3/2)^2$ supplies the initial range. Taking $K = \lceil C_{R,\gamma} d/a \rceil$ makes $\mathbb{E}_F F(x_K) \leq \gamma a/16$, uniformly in F. Moreover the exact pair identity

$$\sum_{i=1}^{d} \{F(x_k + he_i) + F(x_k - he_i)\} = 2dF(x_k) + h^2 \operatorname{tr}(B)$$

and the preceding decay imply

$$N_Z = 2dK \leq C_{R,\gamma} \frac{d^2}{a}, \qquad \mathbb{E}_F R_Z \leq C_{R,\gamma} \frac{d^2}{a}.$$

Indeed $\mathbb{E}F(x_k) \leq C_R \min\{1, d/(k+k_0)\}$ and

$$\sum_{k=1}^{K} \min\left\{1, \frac{d}{k+k_0}\right\} \leq C_R d\{1 + \log(1/a)\}.$$

Thus the $2dF(x_k)$ terms contribute at most $O_R(d^2\{1+\log(1/a)\}) = O_R(d^2/a)$. The common curvature term $h^2 \operatorname{tr}(B) = O_R(d)$ contributes $O_R(dK) = O_R(d^2/a)$. The bounds hold under every posterior. Posterior-risk acceptance and first-hit truncation therefore give $\mathbb{P}(\mathsf{M}) \geq 15/16$. Using only the hit term in the denominator of (41) proves (63).

For the converse, fix $H = I$, draw $V \sim \mathrm{Unif}\{-1,1\}^d$, and set $y_V = rV/\sqrt{d}$ with $r = 1/4$. The initial Bayes action is zero and its risk is the constant $b = r^2/(4R)$. Flipping bit i changes a query mean by $-rv_ix_i/(R\sqrt{d})$. The cube-entropy comparison and stopped Gaussian chain rule used in Proposition A.8 therefore give

$$I(V;Z) \le \frac{C_R}{d}\mathbb{E}N_Z.$$

If $m_Z = \mathbb{E}[V \mid Z]$, the terminal posterior Bayes risk is $b\{1 - \|m_Z\|^2/d\}$. Hence on M one has $\|m_Z\|^2 \ge (1-\gamma)d$, and the sub-Gaussian entropy variational bound yields $I(V;Z) \ge c_\gamma d\mathbb{P}(\mathsf{M})$. Finally $I(A;Z) \le I(V;Z)$ by data processing. Thus

$$I(A;Z) + d\mathbb{P}(\mathsf{M}) \le \frac{C_{R,\gamma}}{d}\mathbb{E}N_Z,$$

whereas the numerator in (41) is at least $b^2\mathbb{E}N_Z$. This proves the order-d lower bound. The explicit upper routine and this constant-scale prior give the two matching statements. □

A.6 From regret bounds to stopped experiments

Proposition A.11 (A stopped experiment from a regret bound)**.** *Suppose an algorithm has, uniformly over fixed losses and all n, δ, a high-probability regret bound*

$$R_n \le \{D\sqrt{n} + A\}\, L(n,d,\delta)$$

outside an event of probability δ, where $A \ge 0$ and, for fixed constants $C_L, q < \infty$,

$$L(n,d,\delta) \le C_L[\log\{end/\delta\}]^q. \tag{64}$$

Then, from any posterior state and scale b, it yields a first-hit experiment with

$$\mathbb{E}N_Z \le \widetilde{O}(1 + D^2/b^2 + A/b)\mathbb{P}(\mathsf{M}), \qquad \mathbb{E}R_Z \le \widetilde{O}(b + D^2/b + A)\mathbb{P}(\mathsf{M}), \qquad \mathbb{P}(\mathsf{M}) \ge \tfrac{1}{2}.$$

In particular, the computation-unrestricted bound of Fokkema et al. (2024) for arbitrary bodies, and their efficient bound for symmetric bodies, have $D = \widetilde{O}(d^{3/2})$ and supply stopped complexity $\widetilde{O}(d^2)$; their additive polynomial terms fit the displayed A/b allowance. Their efficient arbitrary-body result instead has $D = \widetilde{O}(d^{7/4})$.

Proof. Take $\delta = \gamma b/4$ and let n be the least integer for which $DL(n,d,\delta)/\sqrt{n} + AL(n,d,\delta)/n \le \gamma b/4$. The explicit bound (64) and a standard doubling argument give

$$n \le C_{\gamma,C_L,q}\left(\frac{D^2}{b^2} + \frac{A}{b} + 1\right)\left[\log\left\{ed\left(\frac{D^2}{b^3} + \frac{A}{b^2} + \frac{1}{b}\right)\right\}\right]^{2q+2}.$$

Thus $n = \widetilde{O}(1 + D^2/b^2 + A/b)$. Since regret is at most n on the failure event, $\mathbb{E}R_n \le \gamma bn/2$. For $\bar{X} = n^{-1}\sum_{t=1}^n X_t$, convexity gives

$$\mathbb{E}[F(\bar{X}) - F_C^\star] \le \mathbb{E}R_n/n \le \gamma b/2.$$

Apply Lemma A.3 with $a = \gamma/2$ to the terminal posterior regret of $\bar{X}$. Then $\mathbb{P}(\mathsf{M}) \ge 1/2$, so the deterministic schedule can be truncated at the first posterior state with Bayes risk at most γb. The gated terminal event implies that such a hit has occurred, while prefix length and regret can only decrease. Since $\mathbb{P}(\mathsf{M}) \ge 1/2$, the schedule cap and expected regret are absorbed by the displayed hit-probability terms. □

B A multistep separation on the dimple family

The dimple construction of Bakhtiari et al. (2025, Lemma 10) makes the distinction between adaptive radii and observation-dependent direction selection exact. The following consequence is self-contained except for the known convexity of the displayed losses. It rules out a natural class of local searches; it is not a lower bound for unrestricted stochastic BCO.

Proposition B.1 (Lower bound for nonadaptive designs and fresh uniform directions)**.** *Let d be sufficiently large, $C = B_2^d(1)$, $d^{-1} \le \varepsilon \le \varepsilon_0$ for a small universal constant ε_0, $s = \sqrt{1+2\varepsilon}$, and $\Theta \sim \mathrm{Unif}(S_2^{d-1}(1))$. Recall the family F_θ from (23); its losses take values in $[0,1]$. Their prior Bayes action is $z = 0$ and*

$$b_0 = F_\theta(0) = s - 1 = \frac{2\varepsilon}{\sqrt{1+2\varepsilon}+1} \asymp \varepsilon, \qquad \nabla F_\theta(0) = -b_0\theta. \tag{65}$$

Every nonadaptive batch $X_1, \dots, X_N$, possibly randomized independently of Θ but fixed before its observations, satisfies

$$I(\Theta; O_{1:N} \mid X_{1:N}) \le CN(\varepsilon^4 + d^{-4}). \tag{66}$$

The same bound holds sequentially for any adaptive experiment which, at time t, first draws $B_t \sim \mathrm{Unif}(S_2^{d-1}(1))$ independently of the past and of Θ, then chooses an arbitrary history- and B_t-measurable radius $R_t \in [0,1]$, queries $X_t = R_t B_t$, and observes $O_t = F_\Theta(X_t) + \xi_t$. For every current posterior,

$$I(\Theta; B_t, O_t \mid H_t) \le C(\varepsilon^4 + d^{-4}). \tag{67}$$

Consequently, for this fresh-uniform-direction class, a stopped transcript Z of expected length N satisfies $I(\Theta; Z) \le CN(\varepsilon^4 + d^{-4})$.

If an event $\mathsf{M} \in \sigma(Z)$ reports an action x_Z with posterior regret at most γb_0, where $\gamma < 1$, then

$$I(\Theta; Z) \ge \frac{d-1}{2}(1-\gamma)^2 \mathbb{P}(\mathsf{M}). \tag{68}$$

Thus either $\Omega(d)$ information or a constant-probability target-risk hit requires

$$N = \Omega_\gamma\left(\frac{d}{\varepsilon^4 + d^{-4}}\right) = \Omega_\gamma(d/\varepsilon^4), \tag{69}$$

whereas the sample scale corresponding to complexity $O(d)$ at $b_0 \asymp \varepsilon$ is d^2/ε^2. The restricted class loses the polynomial factor $1/(d\varepsilon^2)$ whenever $d^{-1} \lesssim \varepsilon \ll d^{-1/2}$. Its direction law, not merely its radii, must respond to the observations.

Proof. Put $q(x) = \varepsilon + \|x\|^2/2$. For $u = \langle \theta, B \rangle$,

$$|F_\theta(rB) - q(rB)| = \frac{1}{2}\big(s - \sqrt{1 + r^2 - 2ru}\big)_+^2.$$

Minimizing the square root over $r \in [0,1]$ shows that the supremum of the right side is $(s-1)^2/2$ when $u \le 0$, and $(s - \sqrt{1-u^2})^2/2$ when $u > 0$. On $|u| \le 1/2$, its square is at most $C(\varepsilon^4 + u^8)$; the complementary spherical cap has exponentially small mass. Since

$$\mathbb{E}u^8 = \frac{105}{d(d+2)(d+4)(d+6)},$$

rotation invariance gives, uniformly in θ,

$$\mathbb{E}_B \sup_{0\le r\le 1} |F_\theta(rB) - q(rB)|^2 \le C(\varepsilon^4 + d^{-4}). \tag{70}$$

For a nonadaptive batch, condition on its public random seed, apply the same bound after interchanging the Haar variables Θ and B by rotational invariance for each fixed design direction, and then apply the Gaussian KL bound relative to the baseline observations $q(X_t) + \xi_t$, and sum (70) over t. For the sequential class, the same estimate remains true after replacing the prior of Θ by an arbitrary posterior because the fresh B_t is uniform. Conditional Gaussian-channel capacity, with the known $q(R_tB_t)$ subtracted, proves (67); the stopped chain rule proves the transcript bound.

Convexity at the origin and (65) give

$$F_\theta(x) \ge b_0 - b_0\langle\theta, x\rangle.$$

On an accepted transcript, with $m_Z = \mathbb{E}[\Theta \mid Z]$, posterior regret at most γb_0 therefore forces $\|m_Z\| \ge 1 - \gamma$. The uniform law σ on the sphere is $1/(d-1)$-sub-Gaussian, so the variational formula gives

$$D_{\mathrm{KL}}(Q\|\sigma) \ge \frac{d-1}{2}\|\mathbb{E}_Q\Theta\|^2.$$

Apply this to each posterior and average over M to obtain (68). □

The next result shows that this restricted obstruction is removed when the probe directions may depend on the observations.

Theorem B.2 (Observation-adaptive experiment for the dimple family)**.** *For the family* (23)*, every $d^{-1} \le \varepsilon \le \varepsilon_0$, and all sufficiently large d, there are universal constants $0 < \gamma_0 < \gamma < 1$ such that, for every target scale $0 < a \le b_0$, a capped observation-adaptive experiment satisfies, uniformly over $\theta \in S_2^{d-1}(1)$,*

$$\mathbb{P}_\theta\{F_\theta(\widehat{x}) \le \gamma_0 a\} \ge 1 - ca, \qquad N_Z \le \widetilde{O}(d^2/a^2), \qquad \mathbb{E}_\theta R_Z \le \widetilde{O}(d^2/a). \tag{71}$$

For every prior on Θ, applying posterior-risk acceptance to $\widehat{x}$ therefore produces an event G of constant probability on which $\widehat{x}$ has terminal posterior regret at most γa. The first-hit event M contains G. For $b \le b_0$ run the routine at $a = b$; for $b_0 < b < b_0/\gamma$ run it at $a = b_0$; and for $b \ge b_0/\gamma$, no posterior state is active because its Bayes risk is at most $b_0 \le \gamma b$. These experiments have stopped complexity $\widetilde{O}(d)$ throughout the active range. At the critical scale $a = b_0 \asymp \varepsilon$, the adaptive cost is

$$N_Z = \widetilde{O}(d^2/\varepsilon^2), \qquad \mathbb{E}R_Z = \widetilde{O}(d^2/\varepsilon). \tag{72}$$

The restricted lower bound in Proposition B.1 exceeds the $O(d)$ benchmark by the factor $1/(d\varepsilon^2)$ whenever that factor diverges, and the adaptive routine matches the target up to polylogarithmic factors. At the Bakhtiari–Lattimore–Szepesvári scale $\varepsilon = 8\log(d)/d$, the respective query bounds are $\widetilde{O}(d^4/\log^2 d)$ and $\Omega(d^5/\log^4 d)$, a separation by a factor $\widetilde{\Omega}(d)$.

At $\varepsilon = d^{-1}$, these bounds yield the following separation. Let Π_b^{fresh} denote Definition A.4 with the infimum restricted to the sequential fresh-uniform-direction policies in Proposition B.1. Every one-query experiment is covered by the $N = 1$ nonadaptive information bound.

Proof of Theorem 2.5. For a one-query experiment, the $N = 1$ case of (66) gives the first bound below. For a fresh-uniform-direction first-hit experiment, (67) and the stopped chain rule give the second:

$$I(\Theta; Z) \le Cd^{-4} \quad \text{or} \quad I(\Theta; Z) \le Cd^{-4}\mathbb{E}N_Z, \qquad I(\Theta; Z) \ge c_\gamma d\mathbb{P}(\mathsf{M}).$$

The last inequality gives $d\mathbb{P}(\mathsf{M}) \le C_\gamma I(\Theta; Z)$. The sublevel-set target is a randomized function of Θ whose private coin is inaccessible, so $I(A; Z) \le I(\Theta; Z)$. Hence

$$I(A; Z) + d\mathbb{P}(\mathsf{M}) \le C_\gamma d^{-4} \mathbb{E} N_Z,$$

where $\mathbb{E}N_Z = 1$ in the one-query case. Since $b^2 \asymp d^{-2}$, the numerator of (41) is at least $cd^{-2}\mathbb{E}N_Z$, proving the two restricted lower bounds. The risk-reduction experiment in Theorem B.2 has $b^2\mathbb{E}N_Z + b\mathbb{E}R_Z = \widetilde{O}(d^2)$ and $\mathbb{P}(\mathsf{M}) \ge c_\gamma$, which proves the unrestricted upper bound. □

The proof, given in Appendix D, has three stages. A small centered orthogonal design creates correlation $\widetilde{\Omega}(\max\{\sqrt{\varepsilon}, d^{-1/2}\})$. Tangent finite differences amplify this to a fixed angle. On the unit sphere the active loss has the exact form

$$F_\theta(x) = \psi(\|\theta - x\|), \qquad \psi(r) = sr - r^2/2,$$

so observation-adapted boundary probes invert the tangent coordinates and halve the angular error geometrically down to $O(a)$. This last loss-sensitive stage is essential: constant correlation alone does not give loss at the target scale.

The gain comes from choosing later query directions using earlier observations; adapting only the radii does not suffice. This is a family-specific result, not a minimax lower or upper bound for unrestricted BCO.

C Posterior-stable lower bounds and measurable selection

This section records two logically separate tools. Section C.1 isolates a posterior-stability condition that a future superlinear stopped-profile lower bound could use; it is not needed for the characterization proved above. Section C.2 proves the measurable-implementation theorem. Readers concerned only with profile equality may proceed directly to Section C.2.

C.1 A lower-bound criterion

The following criterion adapts the Bellman-supersolution lower-bound argument of Xu (2026a) to the present first-hit problem.

Proposition C.1 (Bellman lower-bound criterion)**.** *Fix an initial active state h_0, a scale b, a persistent target A, and $\lambda > 0$. Suppose there is a measurable, finite-valued nonnegative function G on descendant histories, with all conditional expectations below well defined, such that*

$$G(h_0) = 0, \qquad G(h) \ge d \quad \text{whenever } r_{\mathrm{B}}(h) \le \gamma b, \tag{73}$$

and every admissible randomized query $Y = (X, O)$ at every descendant history h reachable from h_0 before the first hit—equivalently, every relevant descendant with $r_{\mathrm{B}}(h) > \gamma b$, whether or not it remains in the original active bucket—satisfies

$$I(A; Y \mid h) + \mathbb{E}[G(h, Y) - G(h) \mid h] \le \frac{b^2 + b\ell(h)}{\lambda}, \qquad \ell(h) = \mathbb{E}[F(X) - F_C^\star \mid h]. \tag{74}$$

Then

$$\Pi_b(h_0; A) \ge \lambda. \tag{75}$$

Proof. Apply (74) along an arbitrary first-hit experiment with stopping time τ, first replacing τ by $\tau \wedge m$. The chain rule gives the sum of the information terms as the mutual information in the truncated transcript, while the G terms and the Bayes-regret terms telescope in expectation. As $m \to \infty$, accumulated cost and transcript information increase to their stopped values. Since $\tau < \infty$ almost surely, the truncated terminal histories are eventually equal to the terminal history; Fatou's lemma applies to the nonnegative G term. Hence

$$I(A; Z \mid h_0) + \mathbb{E}G(h_0, Z) \le \frac{b^2 \mathbb{E}N_Z + b\mathbb{E}R_Z}{\lambda}.$$

On M, the posterior-risk condition implies $r_{\mathrm{B}}(h_0, Z) \le \gamma b$; off M, nonnegativity of G applies. Thus $\mathbb{E}G(h_0, Z) \ge d\mathbb{P}(\mathsf{M})$, and taking the infimum over experiments proves (75). □

The following specialization states the uniformity required of a lower-bound construction. An initial weak-channel estimate is not enough: the channel bound must survive every posterior reached by an arbitrary policy.

Lemma C.2 (A KL criterion for stopped lower bounds)**.** *Let a parameter Θ generate F, let $Q_h = \mathbb{P}_{\Theta|h}$, and write $Q_0 = \mathbb{P}_\Theta$. Suppose all relative entropies below are finite and, conditionally on every descendant history reachable before the first hit, $A - \Theta - Y$ is a Markov chain for the next randomized query and observation $Y = (X, O)$. If, for some $c_0 > 0$,*

$$r_{\mathrm{B}}(h) \le \gamma b \quad \Longrightarrow \quad D_{\mathrm{KL}}(Q_h \| Q_0) \ge c_0 d, \tag{76}$$

and every randomized query at every such pre-hit descendant history satisfies

$$I(\Theta; Y \mid h) \le \frac{c_0}{1 + c_0} \frac{b^2 + b\ell(h)}{\lambda}, \tag{77}$$

then $\Pi_b(Q_0; A) \ge \lambda$.

Proof. Set

$$G(h) = c_0^{-1} D_{\mathrm{KL}}(Q_h \| Q_0).$$

The conditional relative-entropy chain rule gives

$$\mathbb{E}[G(h, Y) - G(h) \mid h] = c_0^{-1} I(\Theta; Y \mid h),$$

while data processing gives $I(A; Y \mid h) \le I(\Theta; Y \mid h)$. Therefore (77) implies (74). Moreover $G(h_0) = 0$, $G \ge 0$, and (76) gives the terminal boundary $G \ge d$. Proposition C.1 applies. □

For finite problems and a fixed query cap, backward induction makes this criterion exact: the Bellman value is

$$\big[\text{information} + \text{terminal } d \text{ credit} - \text{cost}/\lambda\big]_+.$$

The function G in Proposition C.1 is a single nonnegative supersolution that proves the lower inequality simultaneously for every finite cap. We omit the standard recursion because the posterior-stable inequality (74), rather than its finite unrolling, is what a future superlinear lower-bound construction would need to verify.

C.2 Measurable implementation

We use two standard facts about relative entropy. They are recorded here because the selection argument depends on measurability of the information term and on approximation by finite prefixes.

Lemma C.3 (Relative entropy under measurable limits)**.** *On a Polish space, $(P, Q) \mapsto D_{\mathrm{KL}}(P\|Q)$ is lower semicontinuous, hence Borel, for the weak topology on probability laws. If $\mathcal{G}_L \uparrow \mathcal{G}$ generates the full sigma-field, then*

$$D_{\mathrm{KL}}(P|_{\mathcal{G}_L}\|Q|_{\mathcal{G}_L}) \uparrow D_{\mathrm{KL}}(P\|Q). \tag{78}$$

Consequently, if Z_L are increasing prefixes generating a complete transcript Z, then $I(A; Z_L) \uparrow I(A; Z)$.

Proof. The variational formula

$$D_{\mathrm{KL}}(P\|Q) = \sup_{\phi} \left\{ \int \phi \, dP - \log \int e^{\phi} \, dQ \right\}$$

over bounded continuous ϕ proves lower semicontinuity, because each displayed functional is continuous in (P, Q); Borel measurability follows. Applying the same formula to bounded functions measurable with respect to $\mathcal{G}_L$, and then using martingale approximation of bounded $\mathcal{G}$-measurable functions, proves (78). For mutual information take the two laws to be $P_{A,Z}$ and $P_A \otimes P_Z$ and restrict both to $\sigma(A, Z_L)$. □

Lemma C.4 (Finite-horizon selection for Gaussian BCO)**.** *Let C be compact and let $\mathfrak{F}_d(C)$ be a Borel subset of $\mathcal{C}(C, [0, 1])$ under the uniform topology. Fix a Borel target kernel into a standard Borel space A and a query cap $L < \infty$. The capped statewise value*

$$v_L(s) = \inf_{\mathcal{E}: N_Z \le L} \rho(s, \mathcal{E}) \tag{79}$$

is universally measurable. For every $\epsilon > 0$, there is a universally measurable state-to-experiment rule $\sigma_{L,\epsilon}$ satisfying

$$\rho(s, \sigma_{L,\epsilon}(s)) \le v_L(s) + \epsilon \quad \text{at every active state } s. \tag{80}$$

Proof. The hidden variable for one scale is (F, A), which takes values in a standard Borel space. Its joint posterior μ is therefore a point of the standard Borel belief space $\mathcal{P}(\mathfrak{F}_d(C) \times \mathsf{A})$. Define

$$\ell(\mu, x) = \int \{f(x) - \min_C f\} \, \mu(df, da), \qquad r(\mu) = \min_{x \in C} \ell(\mu, x).$$

If the initial state is $s = (Q, b)$ and the target channel is $K_b(da \mid f)$, then

$$\mu_s(df, da) = Q(df) K_b(da \mid f).$$

The map $s \mapsto \mu_s$ is Borel by measurable parameter integration. The integrand is bounded and jointly continuous in (f, x), so ℓ is Borel and continuous in x; compactness of C makes r Borel and gives a Borel Bayes-action selector. The Gaussian observation kernel is Borel, and its strictly positive density gives the explicit Borel update

$$\Psi(\mu, x, o)(df, da) = \frac{\varphi(o - f(x)) \, \mu(df, da)}{\int \varphi(o - f'(x)) \, \mu(df', da')},$$

where φ is the standard Gaussian density. By Lemma C.3,

$$i(\mu, x) = I_\mu(A; O \mid X = x)$$

is Borel. It is finite uniformly in (μ, x): data processing and the Gaussian capacity bound give $i(\mu, x) \le I(F(x); F(x) + \xi) \le \frac{1}{2}\log(5/4)$.

For fixed $\lambda > 0$, the depth-k offset problem has the belief-state recursion

$$\begin{aligned} J_k^\lambda(\mu, b) = \inf_{x \in C} \Big\{ & b^2 + b\ell(\mu, x) - \lambda i(\mu, x) \\ & + \int \big[-\lambda d \mathbf{1}\{r(\mu') \le \gamma b\} + \mathbf{1}\{r(\mu') > \gamma b\} W_{k-1}^\lambda(\mu', b)\big] P_\mu(do \mid x) \Big\}, \qquad (81) \\ W_0^\lambda = 0, \qquad & W_k^\lambda = \min\{0, J_k^\lambda\}, \qquad \mu' = \Psi(\mu, x, o). \end{aligned}$$

The root uses J_L^λ, rather than W_L^λ, because an experiment has positive length; W_k^λ permits voluntary stopping after the first query. The posterior belief is a sufficient controlled state, so any capped history-dependent policy can be Markovized without changing its expected offset value (equivalently, the relevant belief–action occupancy measures). If τ is its stopping time, the stopped information chain rule reads

$$I(A; Z) = \mathbb{E} \sum_{t=1}^{\tau} i(\mu_{t-1}, X_t). \qquad (82)$$

Together these facts show by induction that J_L^λ equals

$$\inf_{\mathcal{E}: N_Z \le L} \{c_s(\mathcal{E}) - \lambda g_s(\mathcal{E})\}, \quad c_s = b^2 \mathbb{E}_s N_Z + b\mathbb{E}_s R_Z, \quad g_s = I_s(A; Z) + d\mathbb{P}_s(\mathsf{M}). \qquad (83)$$

Randomization causes no difficulty: its complete seed is independent of (F, A) given the state and is included in the transcript, so the offset is the average of the offsets of its deterministic realizations.

All primitive kernels and costs above are Borel, and the only negative one-step term is finite because $i(\mu, x) \le \frac{1}{2}\log(5/4)$. For fixed L and λ the stage cost can therefore be shifted to be nonnegative before applying the standard integration theorem. The recursion preserves lower semianalyticity jointly in (μ, b, λ), and the measurable selection theorem gives universally measurable near-minimizers at every stage, with stagewise errors allocated to sum to any prescribed total (Bertsekas and Shreve, 1978, Chapter 7). Finally, for rational $q > 0$,

$$v_L(s) < q \quad \Longleftrightarrow \quad J_L^q(\mu, b) < 0. \qquad (84)$$

The right-hand side is universally measurable, so the rational strict sublevel sets in (84) first show that v_L is universally measurable. To construct the policy, enumerate $\mathbb{Q}_{>0} = \{q_1, q_2, \ldots\}$. On $\{v_L < \infty\}$, choose measurably the first index $j(s)$ such that

$$v_L(s) < q_{j(s)} \le v_L(s) + \epsilon/2,$$

and then the first $m(s)$ satisfying

$$\frac{1}{m(s)} < -\frac{1}{2} J_L^{q_{j(s)}}(\mu_s, b).$$

For each pair (j, m), the dynamic program has a universally measurable $1/m$-optimal offset policy. Patch these countably many policies on the sets $\{j(s) = j, m(s) = m\}$. Its induced cost and progress satisfy

$$c_s - q_{j(s)} g_s \le J_L^{q_{j(s)}}(\mu_s, b) + \frac{1}{m(s)} < 0.$$

Thus $g_s > 0$ and $\rho(s, \sigma_{L,\epsilon}(s)) < q_{j(s)} \le v_L(s) + \epsilon/2$. On $\{v_L = \infty\}$, any measurable one-query rule suffices and the desired inequality is vacuous. The patched stagewise rules form one universally measurable block policy on the augmented state, proving (80). □

Proposition C.5 (Selection and truncation)**.** *Let S be a standard Borel state space, let $\mathcal{A}_L(s)$ be the admissible first-hit experiments using at most L queries, and let $v_L(s)$ be their statewise infimum. Suppose that every v_L has the measurable selectors in (80). Then the unrestricted value*

$$v(s) = \inf_{\mathcal{E} \in \mathcal{A}(s)} \rho(s, \mathcal{E}) \tag{85}$$

is universally measurable and has a universally measurable ϵ-optimal selector for every $\epsilon > 0$. The selected experiment may be taken to have a finite deterministic cap $L(s)$ depending on the state. Consequently the statewise and measurable-policy profiles coincide.

Proof. For an admissible experiment $\mathcal{E}$, stop it after at most L queries and denote the resulting first-hit experiment by $\mathcal{E}^{[L]}$. Encode each truncation as a nested transcript prefix with a stopped marker and dummy padding, so that $\sigma(Z^{[L]}) \uparrow \sigma(Z)$. Since regret is nonnegative and $\mathbb{E}N_Z < \infty$,

$$c_s(\mathcal{E}^{[L]}) \uparrow c_s(\mathcal{E}), \qquad \mathbb{P}_s(\mathsf{M}^{[L]}) \uparrow \mathbb{P}_s(\mathsf{M}).$$

The prefix sigma-fields increase to that of the complete almost surely finite transcript, so Lemma C.3 gives $I_s(A; Z^{[L]}) \uparrow I_s(A; Z)$. Thus $g_s(\mathcal{E}^{[L]}) \uparrow g_s(\mathcal{E})$ and

$$\rho(s, \mathcal{E}^{[L]}) \longrightarrow \rho(s, \mathcal{E}), \tag{86}$$

including the zero-denominator case under the value $+\infty$ convention. It follows that $v = \inf_{L \ge 1} v_L$, hence v is universally measurable.

Choose measurably

$$L_\epsilon(s) = \min\{L : v_L(s) \le v(s) + \epsilon/2\}$$

and, on the universally measurable set $\{s : L_\epsilon(s) = L\}$, use $\sigma_{L,\epsilon/2}$. The patched rule stores $L_\epsilon(s)$ at block entry and is one universally measurable policy on the augmented state. It has ratio at most $v(s) + \epsilon$ at every state. Taking the supremum over states and then $\epsilon \downarrow 0$ proves equality of the profiles. □

For the next result it is useful to isolate the selection issue from the convex geometry. A *finite stopped model* $\mathcal{D} = (\mathcal{F}, \mathcal{X}, \mathcal{O}, P, \ell)$ has finite hypothesis, action, and observation sets, observation law $P_f(\cdot \mid x)$, and bounded loss $\ell(f, x)$. Policies, posterior Bayes actions, and the regret benchmark all use the same action set $\mathcal{X}$. Fix an external normalization parameter $d > 0$ for the terminal credit in (7), and a Borel kernel $K(\cdot \mid f, b)$ into a standard Borel target space. Define active posterior states, first-hit experiments, ρ, and the two profiles exactly as above, with $\min_{x \in \mathcal{X}} \mathbb{E}[\ell(F, x) - \min_y \ell(F, y)]$ as posterior risk; write the resulting state space and profiles (suppressing K in the state notation) as $\mathsf{S}_{\mathcal{D}}(d, B)$ and $\Lambda^{\mathrm{state}}_{\mathcal{D},K}, \Lambda^{\mathrm{policy}}_{\mathcal{D},K}$. This finite model includes fixed discretizations of BCO, but the following claim is only about measurable selection within the stated alphabets.

Lemma C.6 (Finite experiment trees are measurable)**.** *Fix a finite stopped model $\mathcal{D}$, a Borel target kernel K into a standard Borel target space, and fixed tie-breaking for terminal Bayes actions. For every integer query cap $1 \le L < \infty$, the randomized depth-L policy trees form a compact metrizable space $\mathcal{P}_L$. The graph of admissible first-hit trees is Borel in $\mathsf{S}_{\mathcal{D}}(d, B) \times \mathcal{P}_L$, and the stopped cost, target information, and hit probability are Borel on this graph. Consequently, the extended ratio ρ, with value $+\infty$ when its denominator vanishes, is lower semianalytic.*

Proof. There are finitely many deterministic contingent trees of depth at most L. At state s, a randomized policy draws a pure-tree index J from a mixing law p_s in the corresponding finite simplex, independently of (F, A) conditional on s, and then executes tree J. The mixing law may depend measurably on s. We include the entire ex-ante seed J among the policy coins in the complete transcript. Thus randomized trees form a compact metrizable probability simplex.

The prior state lies in a finite-dimensional probability simplex. At every positive-probability history, Bayes' rule is a rational function of the prior and the tree kernels; define the posterior arbitrarily on a zero-probability history. With this convention posterior updating is Borel. Posterior risk is the minimum of finitely many Borel expected losses, fixed tie-breaking makes a terminal Bayes action Borel, and the first-hit and admissibility conditions are finite combinations of Borel conditions. Thus the feasible graph is Borel.

The capped transcript takes values in a finite set. The Borel target kernel and the finite likelihood recursion make the joint law of (F, A, Z) a Borel probability kernel in the state and policy tree. Expected query cost, regret, and hit probability are finite sums of its marginals. Lemma C.3 shows that relative entropy, and hence mutual information, is Borel after choosing compatible Polish realizations of the target and transcript spaces. The stated extended ratio is therefore Borel on the feasible graph and, in particular, lower semianalytic. □

Proposition C.7 (Finite stopped models)**.** *For every finite stopped model $\mathcal{D}$, Borel target kernel K into a standard Borel target space, and normalization $d > 0$, the unrestricted profiles coincide:*

$$\Lambda^{\mathrm{state}}_{\mathcal{D},K}(d, B) = \Lambda^{\mathrm{policy}}_{\mathcal{D},K}(d, B). \tag{87}$$

The equality also holds separately under every fixed finite cap. For every $\epsilon > 0$, one may choose a universally measurable selector that is additively ϵ-optimal at every active state and whose deterministic cap is finite but may depend on the state.

Proof. For each fixed L, Lemma C.6 and the universally measurable selection theorem for lower-semianalytic minimization (Bertsekas and Shreve, 1978, Proposition 7.50) give a selector $\sigma_{L,\epsilon}$ satisfying

$$\rho(s, \sigma_{L,\epsilon}(s)) \le \inf_{\mathcal{E} \in \mathcal{A}_L(s)} \rho(s, \mathcal{E}) + \epsilon \qquad \text{for every active state } s.$$

Here $\mathcal{A}_L(s)$ is the nonempty set of admissible trees at s. Taking the supremum over s, then the infimum over selectors, and finally $\epsilon \downarrow 0$ proves the reverse inequality to the always-valid relation $\Lambda^{\mathrm{state,[L]}}_{\mathcal{D},K} \le \Lambda^{\mathrm{policy,[L]}}_{\mathcal{D},K}$. The uncapped statement and the state-dependent finite cap now follow from Proposition C.5. □

Proof of Theorem 2.2. For a fixed prior, the upper bound is Theorem A.6 applied to a measurable additive-one selector and the sublevel-set targets from Lemma A.1. Taking the supremum over priors gives the same bound for $\mathfrak{R}^{\mathrm{B}}_{\mathfrak{F}}$, and Proposition 2.1 transfers it to $\mathfrak{R}_{\mathfrak{F}}$. Conversely, the statewise supremum and Proposition A.7 produce a prior with the stated Bayes lower bound, which Proposition 2.1 identifies with a minimax lower bound. To make the exponent step explicit, take $B = d^q$. Then $b \ge d^{-q}$, while the general profile bound in Proposition A.11, applied to the computation-unrestricted stochastic-BCO regret bound, gives $\Lambda^{\mathrm{state}}_{\mathfrak{F}}(d, d^q) \le d^{2+o(1)}$ for every subclass of bounded convex losses. Hence $\lambda \le d^{2+o(1)}$, and the corresponding horizon in (17) satisfies $n \le d^{2q+3+o(1)}$. It therefore lies in a polynomial horizon window, and taking dimension growth rates gives the exponent sandwich. □

Proof of Theorem 2.3. In Gaussian BCO, Lemma C.4 supplies a universally measurable near-minimizer under every finite cap. Proposition C.5 then supplies one measurable state-dependent

experiment for the uncapped value, proving (20). For finite stopped models, the same conclusion is Proposition C.7. □

Proof of Theorem 2.4. Theorem 2.3 identifies the two profiles and hence their polynomial growth exponents. The exponent sandwich in Theorem 2.2 becomes (21). The dense linear prior in Proposition A.8 gives the lower bound in (22); the general BCO regret bound and Proposition A.11 give the upper bound. □

D Proof of the observation-adaptive dimple experiment

Throughout this appendix fix a target $0 < a \le b_0$ and total failure probability $\delta = c_\delta a$, where $c_\delta > 0$ will be chosen after $\gamma_0 < \gamma$, and put $\Lambda = 1 + \log\{16d(1 + \log(d/a))/\delta\}$. Constants below are universal and may change from line to line. We use repeatedly the standard consequence of sphere concentration that, for a Haar orthonormal basis (u_i) in a fixed k-dimensional subspace and fixed unit w,

$$\max_i |\langle w, u_i\rangle| \le C\sqrt{\Lambda/k} \tag{88}$$

outside probability $e^{-c\Lambda}$ when $\Lambda \le k$; when $\Lambda > k$, we use the deterministic bound 1. All repetitions below use fresh Gaussian noise. Each stage has a deterministic query cap, the number of stages is at most $C(1 + \log(d/a))$, and confidence levels are divided across these caps so that the union of all concentration failures is at most δ.

Every use of Haar concentration is conditional on the transcript before the corresponding basis is drawn. At that point the current iterate, the hidden direction, and the activity margin are fixed, while the new basis is Haar in the stated tangent space. Thus the displayed tail bound applies conditionally at each stage. Iterated conditioning and the allocated union bound give one simultaneous event on which all activity, estimation, and contraction invariants below hold. The deterministic caps define the routine on its complement, so no assertion relies on an indefinitely repeated success event.

Whenever $r = \|\theta - x\| < s$, expanding (23) gives the exact identity

$$F_\theta(x) = s\|\theta - x\| + \langle \theta, x\rangle - 1. \tag{89}$$

Lemma D.1 (Orthogonal initialization)**.** *With probability at least $1 - \delta/3$, the following construction returns a unit vector v_0 satisfying*

$$\langle \theta, v_0\rangle \ge c \min\left\{1, \max\left(\sqrt{\varepsilon}, \sqrt{\Lambda/d}\right)\right\}, \tag{90}$$

using at most $Cd^2\Lambda^2/\varepsilon^2$ queries and $Cd^2\Lambda^2/\varepsilon$ regret on the joint success event. If D is the noiseless centered-difference vector and ζ is its Gaussian estimation error, then

$$D = -b_0\theta + e_0, \quad \|e_0\| \le Ch_0^2, \quad |\langle \theta, \zeta\rangle| \le \frac{C\varepsilon}{h_0\sqrt{d\Lambda}}, \quad \|\zeta\| \le \frac{C\varepsilon}{h_0\Lambda}\left(1 + \sqrt{\frac{\Lambda}{d}}\right). \tag{91}$$

Proof. Draw a Haar basis $(u_i)_{i=1}^d$, set

$$h_0 = c_0 \min\{\sqrt{\varepsilon}, \varepsilon\sqrt{d/\Lambda}\}, \qquad n_0 = \left\lceil Cd\Lambda^2/\varepsilon^2 \right\rceil, \tag{92}$$

and estimate every centered difference

$$D_i^0 = \frac{F_\theta(h_0u_i) - F_\theta(-h_0u_i)}{2h_0}$$

using n_0 repetitions at each of the two points. On (88),

$$h_0^2 + 2h_0 \max_i |\langle \theta, u_i \rangle| < 2\varepsilon$$

for small enough c_0, so both points lie in the active dimple. Directly from (89),

$$D_i^0 = -\lambda_i \langle \theta, u_i \rangle, \qquad \lambda_i = \frac{2s}{r_i^+ + r_i^-} - 1, \qquad |\lambda_i - b_0| \le C h_0^2, \tag{93}$$

where $r_i^{\pm} = \|\theta \mp h_0 u_i\|$. Hence the mean difference vector is $-b_0\theta + e_0$ with $\|e_0\| \le Ch_0^2$.

The Gaussian error vector ζ is isotropic with coordinate variance $(2n_0h_0^2)^{-1}$. At the allocated joint failure level,

$$|\langle \theta, \zeta \rangle| \le \frac{C\varepsilon}{h_0\sqrt{d\Lambda}},$$
$$\|\zeta\| \le \begin{cases} C\varepsilon/(h_0\Lambda), & \Lambda \le d, \\ C\varepsilon/(h_0\sqrt{d\Lambda}), & \Lambda > d. \end{cases}$$

Let $t = \min\{1, \max(\sqrt{\varepsilon}, \sqrt{\Lambda/d})\}$. When $\Lambda \le d$, the definition of h_0 gives $h_0 \asymp c_0\varepsilon/t$. Hence, after first choosing c_0 small and then the repetition constant large,

$$\langle \theta, -\widehat{D}^0 \rangle \ge b_0 - Ch_0^2 - |\langle \theta, \zeta \rangle| \ge c\varepsilon, \qquad \|\widehat{D}^0\| \le C\varepsilon/t.$$

When $\Lambda > d$, one has $t = 1$; the second line of the noise bound, $\varepsilon \ge d^{-1}$, and the definition of h_0 instead give the same two inequalities with $t = 1$. Therefore

$$v_0 = -\widehat{D}^0/\|\widehat{D}^0\|, \qquad \langle \theta, v_0 \rangle \ge \min\Big\{c_\star, c_1 \max\Big(\sqrt{\varepsilon}, \sqrt{\Lambda/d}\Big)\Big\} =: c_{\text{init}}. \tag{94}$$

Here $c_\star > 0$ is a small universal constant. This stage uses $\widetilde{O}(d^2/\varepsilon^2)$ queries. Its query points have loss $O(\varepsilon)$, so its regret is $\widetilde{O}(d^2/\varepsilon)$; these are bounded by $\widetilde{O}(d^2/a^2)$ and $\widetilde{O}(d^2/a)$ because $a \le b_0 \asymp \varepsilon$. □

Lemma D.2 (Tangent amplification)**.** *Conditionally on the initialization event, the capped routine in this stage reaches $\|\theta - v\| \le q_\star$ with failure probability at most $\delta/3$, using*

$$N_{\text{amp}} \le \frac{Cd^2\Lambda}{c_{\text{init}}^4}, \qquad R_{\text{amp}} \le \frac{Cd^2\Lambda}{c_{\text{init}}^2}, \tag{95}$$

up to lower-order $C\Lambda(1+\log(1/c_{\text{init}}))$ acceptance-test costs. At every continuing stage with correlation at least c, its probes are active, its noiseless tangent estimate g and Gaussian error ζ satisfy

$$\langle \theta, g \rangle \ge c_1 c^2, \quad \|g\| \le Cc, \quad |\langle \theta, \zeta \rangle| \le \tfrac{1}{2} c_1 c^2, \quad \|\zeta\| \le Cc, \tag{96}$$

and the new correlation lower bound grows by a fixed factor.

Proof. Fix $q_\star$ first. Suppose a unit vector v satisfies $\langle \theta, v \rangle \ge c$, where $c_{\text{init}} \le c \le 1 - q_\star^2/8$. All activity and Taylor estimates below are uniform on this interval. In particular, $c^2 \ge c_2\varepsilon$ after changing universal constants. On the unit sphere the value is the nondecreasing function

$$\phi_\varepsilon(q) = \varepsilon + \frac{1}{2} - \frac{1}{2}(s - q)_+^2, \qquad q = \|\theta - v\|.$$

Choose $q_\star > 0$ so small that $q_\star < s/4$. Uniformly over $0 < \varepsilon \le \varepsilon_0$, the gap

$$\Delta_\star = \phi_\varepsilon(q_\star) - \phi_\varepsilon(q_\star/2)$$

is bounded below by a positive universal constant. Compare the empirical value at v with the midpoint of these two values. Gaussian concentration with $O(\Lambda/\Delta_\star^2) = O(\Lambda)$ repetitions separates $q \le q_\star/2$ from $q \ge q_\star$; on the overlap it may take either branch. Thus stopping implies $q \le q_\star$, while continuing implies $q \ge q_\star/2$. On the active part the same value is the strictly increasing function

$$\psi(q) = sq - q^2/2, \tag{97}$$

for a fixed sufficiently small $q_\star > 0$.

On the latter branch, put $x = (c/2)v$ and draw a Haar tangent basis $(u_i)_{i=1}^{d-1}$ of $v^\perp$. When $\Lambda \le d$, set $h = c_2 c$; by (88) and $c \ge c_{\mathrm{init}}$, all $x \pm hu_i$ remain in the active dimple. On the capped branch $\Lambda > d$, set $h = c_2 c^2$ instead. Since $\|\theta - x\|^2 \le 1 - 3c^2/4$, this choice is deterministically active for small c_2; here $c \ge \min\{c_\star, c_1\} =: c_{\mathrm{low}} > 0$, and changing h only changes universal constants in the repetition budget. If $q_i = \langle \theta, u_i \rangle$, their exact central differences satisfy

$$D_i = -\lambda_i q_i, \qquad \lambda_i = \frac{2s}{r_i^+ + r_i^-} - 1. \tag{98}$$

Writing $r_0 = \|\theta - x\|$ and $\lambda_0 = s/r_0 - 1$, elementary Taylor bounds on the square root, uniformly for $r_i^\pm \ge 1/2$, give

$$\lambda_0 \ge c_3 c^2, \qquad \lambda_0 \le Cc, \qquad |\lambda_i - \lambda_0| \le Ch^2. \tag{99}$$

Put

$$g = -\sum_i D_i u_i = \sum_i \lambda_i q_i u_i, \qquad w = P_{v^\perp}\theta = \sum_i q_i u_i.$$

On the continuing branch, $\|\theta - v\| \ge q_\star/2$, while the correlation lower bound and the choice of k below keep $\langle \theta, v \rangle$ positive. Thus $\|w\|$ is bounded below by a positive constant depending only on $q_\star$. Equations (99) and $h \le c_2 c$ give, for sufficiently small c_2,

$$\langle \theta, g \rangle = \sum_i \lambda_i q_i^2 \ge c_4 c^2, \qquad \|g\|^2 = \sum_i \lambda_i^2 q_i^2 \le Cc^2.$$

Estimate each D_i with $n_c = \lceil Cd\Lambda/c^4 \rceil$ repetitions at each side. Choosing the repetition constant uniformly over the capped stages, Gaussian concentration in $v^\perp$ and the lower bound $c \ge c_{\mathrm{init}} \gtrsim \sqrt{\Lambda/d}$ (in the nontrivial $\Lambda \le d$ case) make the error ζ satisfy

$$\|\zeta\| \le c_5 c, \qquad |\langle \theta, \zeta \rangle| \le (c_4/2)c^2$$

on the joint event. Thus the normalized negative estimate $u \in v^\perp$ obeys

$$\langle \theta, u \rangle \ge k_0 c \tag{100}$$

for a universal $k_0 > 0$. Choose $0 < k \le \min\{k_0, q_\star/4\}$ once and for all. Replacing v by $(v + ku)/\sqrt{1+k^2}$ raises its correlation lower bound from c to at least $\sqrt{1+k^2}\, c$; on the continuing branch the choice of k keeps this lower bound at most one. Iterate until it exceeds $1 - q_\star^2/8$. At that point $\|\theta - v\| \le q_\star/2$, so the radial gate must stop. There are at most $C \log(2/c_{\mathrm{init}}) = O(\log d)$ stages. The stage at correlation c uses $O(d^2\Lambda/c^4)$ queries. Moreover $F_\theta(x \pm hu_i) \le C(\varepsilon + c^2) \le Cc^2$, so geometric summation from c_{init} gives

$$N_{\mathrm{amp}} = \widetilde{O}(d^2/\varepsilon^2), \qquad R_{\mathrm{amp}} = \widetilde{O}(d^2/\varepsilon). \tag{101}$$

□

Lemma D.3 (Angular contraction)**.** *Conditionally on the first two stage events, this capped stage returns $\widehat{x}$ with $F_\theta(\widehat{x}) \le \gamma_0 a$ outside probability $\delta/3$, and has deterministic query cap $Cd^2\Lambda/a^2$ and success-event regret at most $Cd^2\Lambda/a$. At any intermediate scale $\|\theta - v\| \le Q \le q_\star$, one step uses at most $Cd^2\Lambda/Q^2$ queries, incurs at most $Cd^2\Lambda/Q$ regret on its success event, and returns v_{new} with $\|\theta - v_{\mathrm{new}}\| \le Q/2$.*

Proof. Now $q = \|\theta - v\| \le q_\star$. Maintain the deterministic invariant $\|\theta - v_j\| \le Q_j$, where $Q_j = q_\star 2^{-j}$, and at one step write $v = v_j$ and $Q = Q_j$. Take $h = c_6 Q$ and draw a Haar tangent basis (u_i) of $v^\perp$. Define

$$\alpha_h = (1+h^2)^{-1/2}, \qquad t = h(1+h^2)^{-1/2}, \qquad x_i^\pm = \alpha_h v \pm t u_i.$$

All these points lie on the unit sphere and within distance CQ of θ, hence in the active dimple. Write

$$\theta = c_v v + \sum_i q_i u_i, \qquad c_v = 1 - q^2/2, \qquad G(p) = \psi(\sqrt{2-2p}).$$

Then exactly

$$F_\theta(x_i^\pm) = G(\alpha_h c_v \pm t q_i), \qquad q_i = \frac{G^{-1}(F_\theta(x_i^+)) - G^{-1}(F_\theta(x_i^-))}{2t}. \tag{102}$$

Let

$$I_Q = \{p \in [-1,1] : \sqrt{2-2p} \le C_0 Q\}.$$

All inner products in (102) lie in I_Q. On this interval G is strictly monotone and, writing $r = \sqrt{2-2p}$,

$$|(G^{-1})'(G(p))| = \frac{r}{s-r} \le CQ.$$

Project each empirical function value onto $G(I_Q)$ before applying G^{-1}. Estimating each value to accuracy $c_7 Q/\sqrt{d}$, using $\lceil Cd\Lambda/Q^2 \rceil$ repetitions per site, changes each inverse inner product by at most $Cc_7 Q^2/\sqrt{d}$. Since $t \asymp Q$, every tangent coefficient changes by at most $Cc_7 Q/\sqrt{d}$. Consequently $\|\widehat{w} - w\| \le Cc_7 Q$, where $w = \sum_i q_i u_i$ and $\widehat{w} = \sum_i \widehat{q}_i u_i$. Clip $\widehat{w}$ to norm at most $1/2$ and set

$$v_{\mathrm{new}} = \sqrt{1 - \|\widehat{w}\|^2}\, v + \widehat{w}.$$

On the ball $\|w\| \le 1/2$, the map $w \mapsto \sqrt{1-\|w\|^2} v + w$ is 2-Lipschitz. Choose $q_\star$ and then c_7 sufficiently small. Projection onto the radius-$1/2$ ball is nonexpansive, so $\|v_{\mathrm{new}} - \theta\| \le 2Cc_7 Q \le Q/2$.

At angular scale Q, the query count and regret are

$$N_Q = \widetilde{O}(d^2/Q^2), \qquad R_Q = \widetilde{O}(d^2/Q), \tag{103}$$

because every probe has loss $\psi(O(Q)) = O(Q)$. Summing geometrically until $Q \le c_5 a$ proves the query and regret bounds in (71). At termination, $F_\theta(\widehat{x}) = \psi(\|\theta - \widehat{x}\|) \le \gamma_0 a$. □

Proof of Theorem B.2. Run Lemmas D.1– D.3 in order, assigning failure probability $\delta/3$ to each. Their deterministic caps and geometric sums give $N_Z \le \widetilde{O}(d^2/a^2)$ and success-event regret $\widetilde{O}(d^2/a)$.

On the complement of the joint concentration event the capped routine uses no more than its displayed query budget and incurs at most one unit of regret per query. Since its probability is at most $\delta = c_\delta a$, this adds only $\widetilde{O}(d^2/a)$ expected regret. The frequentist guarantee is uniform in θ, so under every prior

$$\mathbb{E}[F_\Theta(\widehat{x})] \le (\gamma_0 + c_\delta) a =: a_0 a.$$

Choose c_δ so that $a_0 < \gamma$ and apply Lemma A.3. It accepts with probability at least $1 - a_0/\gamma > 0$ and gives the required terminal posterior-risk bound. □

Acknowledgements

The author used ChatGPT and Codex as research and editorial aids. The author assumes full responsibility for the statements and proofs in the paper.

References

Alekh Agarwal, Dean P. Foster, Daniel Hsu, Sham M. Kakade, and Alexander Rakhlin. Stochastic convex optimization with bandit feedback. *SIAM Journal on Optimization*, 23(1):213–240, 2013. doi: 10.1137/110850827. URL https://doi.org/10.1137/110850827.

Dilip Arumugam and Benjamin Van Roy. Deciding what to learn: A rate-distortion approach. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 373–382. PMLR, 2021. URL https://proceedings.mlr.press/v139/arumugam21a.html.

Alireza Bakhtiari, Tor Lattimore, and Csaba Szepesvári. Thompson sampling for bandit convex optimisation. In *Proceedings of the 38th Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 231–263. PMLR, 2025. URL https://proceedings.mlr.press/v291/bakhtiari25a.html.

Dimitri P. Bertsekas and Steven E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Academic Press, New York, 1978.

Sébastien Bubeck and Ronen Eldan. Multi-scale exploration of convex functions and bandit convex optimization. In *Proceedings of the 29th Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 583–589. PMLR, 2016. URL https://proceedings.mlr.press/v49/bubeck16.html.

Sébastien Bubeck and Mark Sellke. First-order bayesian regret analysis of thompson sampling. *IEEE Transactions on Information Theory*, 69(3):1795–1823, 2023. URL https://ieeexplore.ieee.org/document/9915622/.

Sébastien Bubeck, Ofer Dekel, Tomer Koren, and Yuval Peres. Bandit convex optimization: $\sqrt{T}$ regret in one dimension. In *Proceedings of the 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 266–278. PMLR, 2015. URL https://proceedings.mlr.press/v40/Bubeck15a.html.

Fan Chen, Dylan J. Foster, Yanjun Han, Jian Qian, Alexander Rakhlin, and Yunbei Xu. Assouad, fano, and le cam with interaction: A unifying lower bound framework and characterization for bandit learnability. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-2407. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/8a23a95e26d016711c0d70f79ade3c95-Abstract-Conference.html.

Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30:755–770, 1959. doi: 10.1214/aoms/1177706205. URL https://doi.org/10.1214/aoms/1177706205.

Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. In *Proceedings of the 21st Annual Conference on Learning Theory*, pages 355–366, 2008. URL https://homes.cs.washington.edu/~sham/papers/ml/bandit_linear_long.pdf.

Shi Dong and Benjamin Van Roy. An information-theoretic analysis for thompson sampling with many actions. In *Advances in Neural Information Processing Systems 31*, pages 4161–4169, 2018. URL https://arxiv.org/abs/1805.11845.

Hidde Fokkema, Dirk van der Hoeven, Tor Lattimore, and Jack J. Mayo. Online newton method for bandit convex optimisation. In *Proceedings of Thirty Seventh Conference on Learning Theory*,

volume 247 of *Proceedings of Machine Learning Research*, pages 1713–1714. PMLR, 2024. URL https://proceedings.mlr.press/v247/fokkema24a.html. Full version: arXiv:2406.06506.

Dylan J. Foster, Sham M. Kakade, Jian Qian, and Alexander Rakhlin. The statistical complexity of interactive decision making, 2021. URL https://arxiv.org/abs/2112.13487.

Dylan J. Foster, Alexander Rakhlin, Ayush Sekhari, and Karthik Sridharan. On the complexity of adversarial decision making. In *Advances in Neural Information Processing Systems*, volume 35, pages 35404–35417, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/e5daf532498ebba4fe6544d49c811678-Abstract-Conference.html.

Dylan J. Foster, Noah Golowich, and Yanjun Han. Tight guarantees for interactive decision making with the decision-estimation coefficient. In *Proceedings of the 36th Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 3969–4043. PMLR, 2023. URL https://proceedings.mlr.press/v195/foster23b.html.

Amaury Gouverneur, Borja Rodríguez-Gálvez, Tobias J. Oechtering, and Mikael Skoglund. Chained information-theoretic bounds and tight regret rate for linear bandit problems, 2024. URL https://arxiv.org/abs/2403.03361.

Steve Hanneke and Kun Wang. A complete characterization of learnability for stochastic noisy bandits. In *Proceedings of the 36th International Conference on Algorithmic Learning Theory*, volume 272 of *Proceedings of Machine Learning Research*, pages 560–577. PMLR, 2025. URL https://proceedings.mlr.press/v272/hanneke25c.html.

Shinji Ito. An optimal algorithm for bandit convex optimization with strongly-convex and smooth loss. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 2229–2239. PMLR, 2020. URL https://proceedings.mlr.press/v108/ito20a.html.

Tor Lattimore and András György. Improved regret for zeroth-order stochastic convex bandits. In *Proceedings of the 34th Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2938–2964. PMLR, 2021a. URL https://proceedings.mlr.press/v134/lattimore21a.html.

Tor Lattimore and András György. Mirror descent and the information ratio. In *Proceedings of the 34th Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2965–2992. PMLR, 2021b. URL https://proceedings.mlr.press/v134/lattimore21b.html.

Tor Lattimore and András György. A second-order method for stochastic bandit convex optimisation. In *Proceedings of the 36th Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 2067–2094. PMLR, 2023. URL https://proceedings.mlr.press/v195/lattimore23a.html.

Tor Lattimore and Csaba Szepesvári. An information-theoretic approach to minimax regret in partial monitoring. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 2111–2139. PMLR, 2019. URL https://proceedings.mlr.press/v99/lattimore19a.html.

Shaojie Li and Yunbei Xu. Pointwise generalization in deep neural networks, 2026. URL https://arxiv.org/abs/2605.18598.

Arnab Maiti, Yunbei Xu, and Kevin Jamieson. On the power of adaptivity for ε-best arm identification in linear bandits. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pages 4911–4968. PMLR, 2026. URL https://proceedings.mlr.press/v336/maiti26a.html.

Mohammad Naghshvar and Tara Javidi. Active sequential hypothesis testing. *The Annals of Statistics*, 41(6):2703–2738, 2013. doi: 10.1214/13-AOS1144. URL https://doi.org/10.1214/13-AOS1144.

Gergely Neu, Matteo Papini, and Ludovic Schwartz. Optimistic information directed sampling. In *Proceedings of the 37th Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 3970–4006. PMLR, 2024. URL https://proceedings.mlr.press/v247/neu24a.html.

Daniel Russo and Benjamin Van Roy. An information-theoretic analysis of thompson sampling. *Journal of Machine Learning Research*, 17(68):1–30, 2016. URL https://jmlr.org/papers/v17/14-087.html.

Daniel Russo and Benjamin Van Roy. Learning to optimize via information-directed sampling. *Operations Research*, 66(1):230–252, 2018. doi: 10.1287/opre.2017.1663. URL https://pubsonline.informs.org/doi/10.1287/opre.2017.1663.

Daniel Russo and Benjamin Van Roy. Satisficing in time-sensitive bandit learning. *Mathematics of Operations Research*, 47(4):2815–2839, 2022. doi: 10.1287/moor.2021.1229. URL https://doi.org/10.1287/moor.2021.1229.

Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pages 3–24. PMLR, 2013. URL https://proceedings.mlr.press/v30/Shamir13.html.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958. doi: 10.2140/pjm.1958.8.171. URL https://doi.org/10.2140/pjm.1958.8.171.

Yunbei Xu. Bellman-sufficient information complexity, 2026a. URL https://arxiv.org/abs/2606.11171.

Yunbei Xu. Pointwise complexity for gaussian fields: Upper envelopes, algorithmic lower bounds, and separation, 2026b. URL https://arxiv.org/abs/2606.07931.

Yunbei Xu and Assaf Zeevi. Bayesian design principles for frequentist sequential learning. *Journal of the ACM*, 72(5):37:1–37:65, October 2025. doi: 10.1145/3766898. URL https://dl.acm.org/doi/10.1145/3766898.

Haihan Zhang, Chenheng Zhang, Zhiquan Qi, and Zhouchen Lin. Near-optimal lower bounds for exact zeroth-order convex optimization, 2026. URL https://arxiv.org/abs/2607.16558.