

Parallel First-Order Optimization in One Dimension: A Tight Round–Width Tradeoff via Moment Shielding

Xiaoyu Li

University of New South Wales
xiaoyu.li2@unsw.edu.au

Abstract

How much can simultaneous queries replace adaptive rounds in optimization? On an interval, gradient signs give a simple benchmark: q queries divide a bracket into $q + 1$ pieces. Exact numerical replies need not respect this picture; two simultaneous gradients already determine any strongly convex quadratic. We show that the search benchmark is nevertheless minimax optimal over C^∞ (infinitely differentiable) objectives with a fixed curvature range. For every fixed $\kappa = L/\mu > 1$, the optimal number of rounds for δ -accurate minimizer localization is

$$\Theta_\kappa \left(\max \left\{ 1, \frac{\log(1/\delta)}{\log(q+1)} \right\} \right),$$

uniformly in the integer batch width $q \geq 1$ at sufficiently small δ , even for randomized algorithms receiving exact function values and gradients. The construction uses a compactly supported smooth perturbation that is linear on the interval reserved for the next level. This preserves curvature bounds at every finite depth and makes all descendants agree exactly outside their active interval. A fixed uniform prior then limits the conditional information of one batch to $\log_2 q + 3$ bits. The same tradeoff holds on the larger C^2 class and for objective-gap accuracy. A known-coordinate restriction transfers the lower bound to every dimension without changing the condition number, but supplies no dimension dependence. The contribution is an exact smooth-oracle realization of the classical parallel-search law, with a self-contained randomized lower-bound proof.

1 Introduction

Parallel queries buy more observations; another round buys the opportunity to decide where to look next. These resources need not be interchangeable. Consider their simplest optimization setting: an unknown strongly convex function on $[0, 1]$, with q locations queried simultaneously in each round. After the replies arrive, the next batch may depend on everything learned so far. How many rounds are needed to locate the minimizer within distance δ ?

Gradient signs suggest an answer. Because the derivative is increasing, q equally spaced queries divide a bracket into $q + 1$ pieces, and their signs identify a piece containing the minimizer. Repetition gives accuracy of order $(q + 1)^{-R}$ after R rounds. But this is only an upper bound. An optimization oracle returns numbers, not just signs, and numbers can defeat the search picture entirely. For a quadratic,

$$f'(x) = ax + b,$$

two simultaneous gradients determine a and b , hence the exact constrained minimizer. No accuracy-dependent sequence of rounds is needed.

This easy example isolates the question we study:

Can a one-dimensional, infinitely differentiable, strongly convex objective hide its minimizer at successive scales, even when every query reveals both the exact function value and the exact gradient?

The answer is yes, with the same order of round complexity as parallel bracketing. The issue is not whether a real number can contain many bits. It is whether a batch chosen before its replies arrive can reach the parts of the function where those bits are stored.

Why one dimension? In one dimension there are no hidden directions and no competing local minima: the derivative is ordered and has at most one zero. A lower bound here therefore isolates the cost of adaptive localization from those geometric obstacles. The comparison with quadratics is equally important. Our class controls the range of the curvature, but imposes no common bound on its higher derivatives. It permits nonquadratic changes on successively smaller scales, even though every individual function is C^∞ . This is a minimax statement over a function class, not a claim that every fixed smooth instance requires logarithmically many local iterations. Nor do we assert a complexity phase transition between dimension one and a fixed larger dimension. The scalar setting makes the information obstruction explicit; dimension-sensitive extensions require an additional argument.

Result. Fix $\kappa > 1$ and consider all $f \in C^\infty([0, 1])$ satisfying $1 \leq f'' \leq \kappa$. One query returns $(f(x), f'(x))$. Algorithms may use private randomness, choose their batch sizes from the history subject to a hard cap q , and output a point they never queried. They must succeed with probability at least $2/3$ on every function and stop within a hard cap of R rounds. For

$$s_\kappa = \frac{\kappa - 1}{12(\kappa + 1)}, \quad 0 < \delta \leq s_\kappa/4,$$

we prove, uniformly over integers $q \geq 1$,

$$R_A^*(q, \delta) = \Theta_\kappa \left(\max \left\{ 1, \frac{\log(1/\delta)}{\log(q+1)} \right\} \right), \quad A \in \{\text{rand}, \text{det}\}. \quad (1.1)$$

The same bounds hold on the larger C^2 class. The maximum accounts for the regime in which a single sufficiently wide batch suffices; the constants depend on κ , and are made explicit in [Theorem 3.1](#).

Mechanism. The proof joins two simple interfaces. First, a localized change in the objective leaves both values and gradients unchanged outside its support. We arrange such changes recursively, while reserving a linear part of each perturbation for the next level. There the perturbation changes the gradient but not the curvature, so the child's curvature replaces, rather than adds to, the parent's. The condition number is consequently independent of depth. In curvature coordinates, exterior equality is precisely the cancellation of two moments: mass and first moment. We retain the name *moment shielding* for this exact compatibility condition.

Second, index the resulting functions by binary strings. Their minimizers occupy separated leaf intervals, and all descendants of a prefix agree outside its interval. A query can reveal several bits by entering several nested intervals, but under a uniform hidden string its penetration depth has a geometric tail. The longest

prefix revealed by q simultaneous probes has entropy at most $\log_2 q + 3$. A posterior-cylinder induction preserves this bound under adaptive batching. Decoding the separated minimizer then gives the lower bound. This argument controls the full numerical replies; it does not replace them by comparisons.

Position relative to prior work. The search law itself is classical. [Gal and Miranker \(1977\)](#) studied optimal sequential and parallel exact-value root search under two-sided divided-difference bounds. With the root function $g = f'$, this includes the deterministic gradient-search problem for an interior minimizer and gives sharper condition-dependent contractions than equal-cell bracketing. Sequential zero finding was further developed by [Sikorski \(1982\)](#), [Novak \(1991\)](#), and [Novak and Ritter \(1993\)](#); parallel information-based complexity was studied by [Heinrich and Kern \(1991\)](#). Infinite differentiability alone is therefore not a new source of root-search hardness. The combination addressed here is fixed positive curvature bounds, exact objective values *and* gradients, and randomized parallel rounds.

The discrete analogue is ordered search. In the comparison model, with the output required to be a queried location, [Brânzei et al. \(2020, Theorem 3, version 5\)](#) give the success-sensitive lower bound for N ordered locations, success probability $p > 0$, total queries Q , and R rounds:

$$\mathbb{E}Q \geq pR(N^{1/R} - 2).$$

A hard width cap implies $\mathbb{E}Q \leq qR$, and hence $N \leq (q/p + 2)^R$. Thus converting an expected-total-query lower bound to a width lower bound is immediate *within the same oracle model*. Our model allows unqueried outputs. What needs proof here is the continuous realization: an exact smooth reply may disclose many address bits at once. Our compact construction and conditional entropy argument supply that missing control. We do not claim the stronger success dependence of the discrete result.

In convex optimization, [Nemirovski \(1994\)](#) established parallel lower bounds for nonsmooth objectives in dimension d . [Balkanski and Singer \(2018\)](#) and [Diakonikolas and Guzmán \(2019\)](#) developed randomized parallel lower bounds using higher-dimensional constructions. Low-dimensional questions also appear explicitly in the discussion of [Diakonikolas and Guzmán \(2019\)](#); the distinction is not simply that all prior work needs a growing dimension. In stochastic batching, [Duchi et al. \(2018, Condition 5 and Lemmas 16–18\)](#) already use nested localization with exact exterior agreement of function values and gradients. Their information analysis uses noisy replies; here we combine depth-independent curvature bounds, a fixed geometric scale, and exact-reply entropy control.

Our contribution is the smooth exact-pair construction and its randomized information analysis, with a matching-order rate. We are not aware of a theorem with this full combination of features in the primary sources inspected, but make no absolute priority claim. The objective-gap conversion and known-coordinate lift in [Section 6](#) record the scope of the result; the lift does not supply a lower bound growing with dimension.

Reading guide. [Section 2](#) fixes the model and gives the upper bound. [Section 3](#) states the quantitative theorem and the geometric interface. [Section 4](#) constructs the smooth family; [Section 5](#) proves the information bound and completes the theorem. [Sections 6](#) and [7](#) give consequences and limitations. The explicit bump and elementary constant calculations are in [Appendix A](#).

2 Model and elementary upper bound

2.1 Function class and oracle

For $0 < \mu < L$ and $p \in \{2, \infty\}$, let

$$\mathcal{F}_{\mu,L}^{(p)} := \{f \in C^p([0, 1]) : \mu \leq f''(x) \leq L \text{ for every } x \in [0, 1]\}. \quad (2.1)$$

Here $C^p([0, 1])$ has its usual one-sided endpoint interpretation; our hard instances in fact extend smoothly to all of $\mathbb{R}$. Unless specified otherwise, the complexity quantities below refer to $\mathcal{F}_{\mu,L}^{(\infty)}$. Every stated upper bound also holds on $\mathcal{F}_{\mu,L}^{(2)}$, and every lower bound uses C^∞ instances, so the same conclusions hold for both classes. Every $f \in \mathcal{F}_{\mu,L}^{(\infty)}$ is μ -strongly convex and has a unique constrained minimizer, denoted by x_f^* . Our unit-cost oracle is the exact standard first-order oracle

$$\mathcal{O}_f(x) := (f(x), f'(x)). \quad (2.2)$$

Thus the lower bound below continues to hold for the weaker gradient-only oracle. The value channel is kept because neutralizing it is one of the main points of the construction.

We make the interaction model explicit.

Assumption 2.1 (Regularity). The unknown objective belongs to $\mathcal{F}_{\mu,L}^{(\infty)}$, and $\kappa := L/\mu > 1$ is fixed.

Assumption 2.2 (Batched adaptivity). Before round r , the algorithm chooses a batch of at most q points as a function of its private random seed and the replies from rounds $1, \dots, r-1$. All points in the batch are fixed before any reply from round r is returned.

Assumption 2.3 (Exact information). The reply at every queried point is the pair in (2.2), with no noise or finite-precision restriction. One pair counts as one query.

Assumption 2.4 (Worst-case guarantee). The algorithm has a deterministic cap of R rounds and, for every $f \in \mathcal{F}_{\mu,L}^{(\infty)}$, returns $\hat{x} \in [0, 1]$, not necessarily previously queried, such that

$$\mathbb{P}_f(|\hat{x} - x_f^*| \leq \delta) \geq \frac{2}{3}.$$

The probability is over the algorithm's private coins.

Let $R_{\text{rand}}^*(q, \delta; \mu, L)$ be the least admissible round cap under [Assumptions 2.1 to 2.4](#). The deterministic quantity $R_{\text{det}}^*(q, \delta; \mu, L)$ is defined by suppressing the private seed. All information-theoretic logarithms are in base two. The base of logarithms in order notation is immaterial.

Remark 2.5 (Normalization). It is enough to prove the lower bound for $\mu = 1$ and $L = \kappa$. Indeed, if $f'' \in [1, \kappa]$, then $(\mu f)'' \in [\mu, L]$, the minimizer is unchanged, and the exact oracle reply is simply multiplied by μ . No information is lost or gained by this deterministic rescaling.

2.2 What parallel bracketing already gives

The upper bound is elementary, but recording it first clarifies the benchmark that the lower bound must match.

Proposition 2.6 (q -ary bracketing). *For every $q \geq 1$, $0 < \delta < 1/2$, and $0 < \mu \leq L$,*

$$R_{\text{det}}^*(q, \delta; \mu, L) \leq \left\lceil \log_{q+1} \frac{1}{2\delta} \right\rceil.$$

The algorithm uses gradients only.

Proof. Maintain a closed interval $I = [\ell, r]$ known to contain x_f^* ; initially $I = [0, 1]$. Write $D = r - \ell$, and query

$$x_j = \ell + \frac{jD}{q+1}, \quad j = 1, \dots, q,$$

in one batch. Since f' is strictly increasing, the signs select one of the $q+1$ consecutive subintervals.

More explicitly, if $f'(x_1) \geq 0$, then the minimizer lies in $[\ell, x_1]$. If $f'(x_q) \leq 0$, it lies in $[x_q, r]$. Otherwise there is a unique j for which $f'(x_j) < 0 < f'(x_{j+1})$, and the minimizer lies in $[x_j, x_{j+1}]$. If a queried derivative is zero, its query point is the minimizer and the algorithm may stop. In every nonterminal case the new interval has length $D/(q+1)$.

After R rounds the maintained interval has length at most $(q+1)^{-R}$. Its midpoint is within $(q+1)^{-R}/2$ of the minimizer. The displayed choice of R makes this at most δ . $\square$

Remark 2.7 (Why this is not yet a lower bound). The proof of [Proposition 2.6](#) uses only gradient signs. The oracle, however, returns exact real numbers. For the quadratic $f_r(x) = \frac{\mu}{2}(x-r)^2$, a single gradient query reveals r exactly. Consequently, counting the $q+1$ geometric cells cut out by a batch does not limit the information carried by that batch. The lower bound must control exact response values, not merely signs.

3 Main result and proof architecture

We now state the round lower bound with explicit constants. The dependence on κ is deliberately visible: the theorem is uniform in the batch width q , but its constants are for a fixed condition number.

Theorem 3.1 (Tight round-width tradeoff). *Fix $\kappa > 1$, and define*

$$s := \frac{\kappa - 1}{12(\kappa + 1)}, \quad n_\delta := \left\lceil \log_{1/s} \frac{1}{4\delta} \right\rceil. \quad (3.1)$$

Let $q \geq 1$. Every private-coin algorithm satisfying [Assumptions 2.2 to 2.4](#) on $\mathcal{F}_{1,\kappa}^{(\infty)}$ obeys

$$R \geq \frac{\frac{2}{3}n_\delta - 1}{\log_2 q + 3} \quad \text{whenever } n_\delta \geq 1. \quad (3.2)$$

Moreover, if $0 < \delta \leq s/4$, then, with

$$A = \log_2(1/\delta), \quad B = \log_2(1/s), \quad H = \log_2(q+1),$$

one has the uniform comparison

$$\frac{1}{12B} \max \left\{ 1, \frac{A}{H} \right\} \leq R_{\text{rand}}^*(q, \delta; 1, \kappa) \leq R_{\text{det}}^*(q, \delta; 1, \kappa) \leq 2 \max \left\{ 1, \frac{A}{H} \right\}. \quad (3.3)$$

Consequently, for every fixed $\kappa > 1$ and under the same condition $0 < \delta \leq s/4$,

$$R_A^*(q, \delta; \mu, L) = \Theta_\kappa \left(\max \left\{ 1, \frac{\log(1/\delta)}{\log(q+1)} \right\} \right), \quad A \in \{\text{rand}, \text{det}\},$$

in order, uniformly over $q \geq 1$, where $L/\mu = \kappa$. Every statement remains true when the function class is replaced by $\mathcal{F}_{\mu,L}^{(2)}$.

Where the lower bound comes from. The construction below hides n bits in minimizers separated by s^n , while a batch reveals at most $\log_2 q + 3$ bits in conditional information. Accuracy δ permits decoding when $4\delta \leq s^n$. This gives the explicit lower bound; [Proposition 2.6](#) supplies the upper bound. A separate two-instance argument gives the one-round floor, so the comparison stays valid even for extremely large q .

3.1 The hard family behind the theorem

The analytic and information-theoretic parts of the proof meet in the following construction.

Theorem 3.2 (Moment-shielded hard family). *Fix $\kappa > 1$, let s be as in (3.1), and let $n \geq 1$. There are functions*

$$\{f_\theta : \theta \in \{0, 1\}^n\} \subset \mathcal{F}_{1,\kappa}^{(\infty)}$$

and closed intervals $J_\sigma \subseteq [0, 1]$, indexed by binary prefixes σ of length at most n , with the following properties.

(H1) $J_\emptyset = [0, 1]$ and $|J_\sigma| = s^{|\sigma|}$. Each child $J_{\sigma i}$ lies in J_σ , and the interiors of the two siblings $J_{\sigma 0}$ and $J_{\sigma 1}$ are disjoint.

(H2) If θ and θ' extend σ , then

$$(f_\theta(x), f'_\theta(x)) = (f_{\theta'}(x), f'_{\theta'}(x)) \quad \text{for every } x \notin \text{int } J_\sigma.$$

(H3) The minimizers x_θ^* satisfy

$$\min_{\theta \neq \theta'} |x_\theta^* - x_{\theta'}^*| = s^n.$$

(H4) $f_\theta(0) = f_\theta(1) = 0$ for every θ .

The functions actually extend to $C^\infty(\mathbb{R})$. The nesting and disjointness in Item (H1) are essential: together with exact exterior agreement, they are the entire interface used by the information argument. The root separation is exact; the constants in the resulting minimax comparison are not claimed optimal.

3.2 A two-level picture

[Figure 1](#) shows the geometry without the inessential scale distortion. The first bit selects one of two adjacent child shells. Inside that shell, a rescaled copy of the same construction stores the second bit. The objective outside the selected shell has already been fixed. A query there may reveal the selected prefix, but it cannot distinguish any of its extensions.

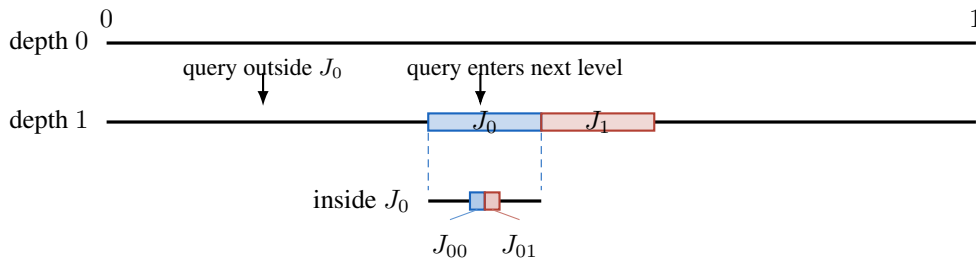


Figure 1: The recursive shell geometry, drawn out of scale. Descendants of a prefix agree in value and gradient outside that prefix's shell. A spatial query can enter deeply only along one address of the binary tree.

This picture suggests the correct replacement for interval counting. A point does not merely choose a geometric cell; it guesses an address in a spatial tree. Under a uniform hidden address, the probability that one fixed point guesses $t - 1$ successive bits is at most $2^{-(t-1)}$. The deepest of q such probes therefore has a geometric tail and expected depth $O(\log(q + 1))$.

4 A smooth family with an invisible interior

We construct the family in [Theorem 3.2](#). It is useful to separate two requirements. A change made inside an interval must leave the exact oracle unchanged outside it. At the same time, inserting finer changes must not accumulate curvature. Compact support handles the first requirement; a linear plateau handles the second.

4.1 Exterior equality and two moments

Suppose two objectives agree to the left of $[\alpha, \beta]$, and their curvatures differ there by an integrable function h , zero outside that interval. Their derivative and value differences are

$$Vh(x) = \int_{\alpha}^x h(t) dt, \quad V^2h(x) = \int_{\alpha}^x (x - t)h(t) dt$$

for $x \geq \alpha$, and zero to the left. For $x \geq \beta$, these become

$$Vh(x) = \int_{\alpha}^{\beta} h(t) dt, \quad V^2h(x) = x \int_{\alpha}^{\beta} h(t) dt - \int_{\alpha}^{\beta} th(t) dt.$$

Thus exact equality of both oracle channels outside the interval is equivalent to two linear constraints.

Lemma 4.1 (Moment shielding). *Let $h \in L^1([0, 1])$ vanish almost everywhere outside $[\alpha, \beta]$. Normalize all successive antiderivatives to vanish to the left of α , so*

$$V^k h(x) = \frac{1}{(k-1)!} \int_{\alpha}^x (x-t)^{k-1} h(t) dt, \quad x \geq \alpha. \quad (4.1)$$

If $\int_{\alpha}^{\beta} t^j h(t) dt = 0$ for $j = 0, \dots, r-1$, then $V^k h$ vanishes outside $[\alpha, \beta]$ for every $1 \leq k \leq r$. For $r = 2$, these are exactly the mass and first-moment constraints needed to shield gradients and values.

Proof. There is nothing to prove to the left of α . To the right of β , replace the upper integration limit in (4.1) by β and expand $(x-t)^{k-1}$ as a polynomial in t . Every term has degree at most $r-1$, so every integral vanishes. For $r = 2$, necessity follows from the two explicit exterior formulas above. A general necessity statement is proved in [Appendix A](#). $\square$

There is a convenient way to satisfy the two constraints automatically. Choose a compactly supported objective perturbation B , and let $h = B''$. If B and B' vanish at the support endpoints, integration by parts gives

$$\int h = 0, \quad \int th = 0.$$

We use this function-level viewpoint in the recursion. It removes the need to tune the two curvature moments separately.

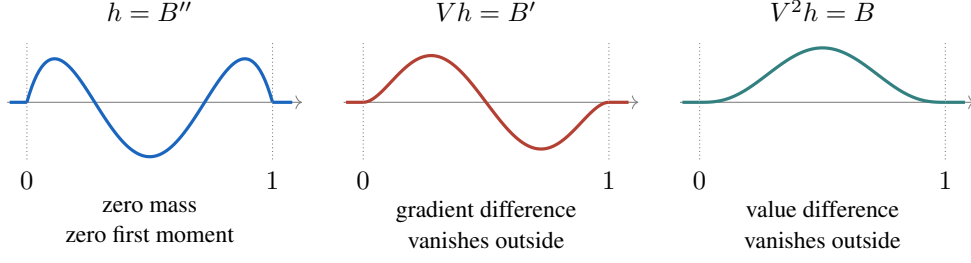


Figure 2: Two moments shield two oracle channels. These are derivatives of the same function: $B(u) = 64u^3(1 - u)^3$ on $[0, 1]$, extended by zero. The plotted vertical coordinates are $B''/24$, $B'/4$, and B , respectively; horizontal coordinates are shared. This C^2 example illustrates the moment identity, not the C^∞ hard-family bump used below.

4.2 A perturbation that leaves room for its child

Write

$$K = \kappa - 1, \quad M = \frac{\kappa + 1}{2}, \quad s = \frac{K}{12(\kappa + 1)}, \quad \lambda = \frac{Ms}{2} = \frac{K}{48}. \quad (4.2)$$

The baseline objective is

$$Q(x) = \frac{M}{2}(x^2 - x).$$

Its minimizer is $1/2$, and its constant curvature M lies halfway between 1 and κ . The two candidate child intervals are

$$J_0 = [1/2 - s, 1/2], \quad J_1 = [1/2, 1/2 + s].$$

We require a compact perturbation that is exactly linear on both children.

Lemma 4.2 (Smooth linear plateau). *There exists $b \in C_c^\infty(\mathbb{R})$ such that*

$$\text{supp } b \subseteq [1/48, 47/48], \quad b(x) = x - \frac{1}{2} \quad \text{if } |x - \frac{1}{2}| \leq 5/48, \quad \|b''\|_\infty \leq 75/4. \quad (4.3)$$

The constants are inessential, but the plateau is not. Since $s < 1/12 < 5/48$, we have $b' = 1$ and $b'' = 0$ on $J_0 \cup J_1$. An explicit compact cubic followed by a compact mollification proves [Lemma 4.2](#); all joins and bounds are checked in [Appendix A.1](#). The mollification preserves the linear part exactly.

4.3 The recursion

Let

$$\ell_0 = \frac{1}{2} - s, \quad \ell_1 = \frac{1}{2}, \quad \chi_i(u) = \ell_i + su, \quad e_0 = 1, \quad e_1 = -1.$$

Define $f_\emptyset = Q$ on all of $\mathbb{R}$, and for every finite binary word $i\omega$ set

$$f_{i\omega}(x) = Q(x) + \lambda e_i b(x) + s^2(f_\omega - Q)\left(\frac{x - \ell_i}{s}\right). \quad (4.4)$$

The first perturbation moves the minimizer toward the selected child. The last term inserts the unresolved part of the objective into that child.

Smoothness and support. Inductively, $f_\omega - Q$ is smooth and compactly supported in $(0, 1)$. Its rescaled copy in (4.4) is therefore smooth and compactly supported in $\text{int } J_i$. Since b also has compact support in $(0, 1)$, the same assertion holds for $f_{i\omega} - Q$. No piecewise gluing is needed. In particular, every finite-depth objective is C^∞ on $\mathbb{R}$, and

$$f_\theta(0) = f_\theta(1) = 0, \quad f'_\theta(0) = -M/2, \quad f'_\theta(1) = M/2. \quad (4.5)$$

Curvature does not accumulate. Inside the selected child, differentiating twice and using $b'' = 0$ gives

$$f''_{i\omega}(\chi_i(u)) = f''_{\omega}(u), \quad 0 \leq u \leq 1. \quad (4.6)$$

Outside the child, the inserted term vanishes, leaving $f''_{i\omega} = M + \lambda e_i b''$. Starting from $Q'' = M$, induction consequently gives

$$|f''_{\theta} - M| \leq \frac{75\lambda}{4} = \frac{25K}{64}, \quad 1 + \frac{7K}{64} \leq f''_{\theta} \leq 1 + \frac{57K}{64}. \quad (4.7)$$

This interval lies inside $[1, \kappa]$, uniformly in depth. Equation (4.6) is the reason: on the region where another level is inserted, the preceding perturbation contributes no curvature.

The root follows the selected child. For $u \in [0, 1]$, the same plateau gives $b'(\chi_i(u)) = 1$, and hence

$$\begin{aligned} f'_{i\omega}(\chi_i(u)) &= M(\ell_i + su - \tfrac{1}{2}) + \lambda e_i + s(f'_{\omega} - Q')(u) \\ &= sM(u - \tfrac{1}{2}) + s(f'_{\omega} - Q')(u) = sf'_{\omega}(u). \end{aligned} \quad (4.8)$$

For $i = 0$, the constant in the first line is $-Ms + \lambda = -Ms/2$; for $i = 1$, it is $-\lambda = -Ms/2$. This verifies the middle equality for both branches. Strict convexity and (4.5) now imply

$$x_{\theta}^* = \chi_{\theta}(1/2), \quad \chi_{\theta} = \chi_{\theta_1} \circ \cdots \circ \chi_{\theta_n}. \quad (4.9)$$

Exact exterior agreement. For every prefix σ , set $J_{\sigma} = \chi_{\sigma}([0, 1])$, with the identity map for the empty prefix. If $J_{\theta_{1:k}} = [A_k, A_k + D_k]$, then $D_k = s^k$. Unrolling (4.4) yields the finite sum

$$f_{\theta}(x) = Q(x) + \lambda \sum_{k=0}^{n-1} e_{\theta_{k+1}} D_k^2 b\left(\frac{x - A_k}{D_k}\right). \quad (4.10)$$

Two leaves extending σ have identical terms for $k < |\sigma|$. Every later difference is supported strictly inside J_{σ} . Their objective values and all their derivatives therefore agree on the exterior, including its boundary. In particular, the exact pair (f, f') carries no information about the suffix at an exterior point. There is no suffix-dependent additive constant.

Completion of Theorem 3.2. The affine maps give $J_{\sigma i} \subset J_{\sigma}$ and $|J_{\sigma}| = s^{|\sigma|}$; sibling interiors are disjoint. The preceding paragraphs prove the regularity, curvature range, shielding, and endpoint normalization. Finally, (4.9) identifies each minimizer with the midpoint of its leaf interval. Distinct depth- n intervals have disjoint interiors and common length s^n , so their midpoints are at least s^n apart. Two leaves differing only in their last bit occupy adjacent sibling intervals, whose midpoints are exactly s^n apart. This proves every item of the theorem. $\square$

The construction uses only finite sums. Higher derivatives need not have bounds uniform in n : differentiating a term in (4.10) r times introduces D_k^{2-r} . Thus the result imposes infinite differentiability on each instance, but does not impose a common Hessian-Lipschitz constant or an analytic extension.

5 Information revealed by a parallel batch

We now prove the lower bound in Theorem 3.1. The hard objective is fixed before the algorithm draws its private coins: throughout the section, Θ is uniform on $\{0, 1\}^n$, and the unknown function is f_{Θ} from Theorem 3.2. This fixed prior avoids the quantifier problem of a resisting oracle that chooses its path after observing the algorithm's random seed.

5.1 The exit-prefix oracle

For a binary prefix σ , write

$$\text{Cyl}(\sigma) := \{\theta \in \{0, 1\}^n : \theta \text{ extends } \sigma\}.$$

Under the uniform leaf prior, conditioning on $\Theta \in \text{Cyl}(\sigma)$ leaves a uniform random suffix.

Fix such a prefix σ and a query location x . For $\theta \in \text{Cyl}(\sigma)$, define the *exit prefix* $W_x^\sigma(\theta)$ as follows. If $x \notin \text{int } J_\sigma$, set $W_x^\sigma(\theta) = \sigma$. Otherwise descend along the shells selected by θ , and stop at the shortest extension τ of σ for which $x \notin \text{int } J_\tau$. If no exit occurs before depth n , set $W_x^\sigma(\theta) = \theta$.

The label is an analytical device; the original oracle does not return it. It is nevertheless useful because it is more informative than the exact numerical pair.

Lemma 5.1 (A batch reveals one prefix). *Condition on a cylinder $\text{Cyl}(\sigma)$, and fix a batch $x_1, \dots, x_b$, where $0 \leq b \leq q$. Let W be the longest among the exit prefixes $W_{x_i}^\sigma(\Theta)$, with $W = \sigma$ when $b = 0$. Then:*

- (i) *the complete exact batch reply $(O_{f_\Theta}(x_i))_{i=1}^b$ is determined by W , the query locations, and σ ;*
- (ii) *every nonempty fiber $\{\theta : W(\theta) = w\}$ is the cylinder $\text{Cyl}(w)$;*
- (iii) *for $D = |W| - |\sigma|$ and every $t \geq 1$,*

$$\mathbb{P}(D \geq t \mid \Theta \in \text{Cyl}(\sigma)) \leq \min\{1, b 2^{-(t-1)}\}. \quad (5.1)$$

Proof. Consider first one query x , and suppose its exit prefix is τ . If $\tau \neq \theta$, then $x \notin \text{int } J_\tau$. By [Theorem 3.2](#) and [Item \(H2\)](#), every descendant of τ returns the same value–gradient pair at x . If $\tau = \theta$, the fiber is a singleton. Thus the exact reply is determined by the exit prefix.

For a label τ shorter than n , the event $W_x^\sigma(\theta) = \tau$ depends only on the shells along τ : the query lies in the interiors of the shells associated with the proper extensions between σ and τ , and lies outside the interior of J_τ . These conditions are independent of the suffix after τ . Hence every nonempty single-query fiber is exactly $\text{Cyl}(\tau)$. A depth- n label has a singleton fiber, whether the query stops by exiting that interval or simply by reaching the depth cap.

All exit labels produced by a batch are prefixes of the same leaf θ . They are therefore nested, and their intersection is the cylinder indexed by the longest label W . The preceding argument also shows stability under every extension of W : a query that exits at W remains outside the same fixed shell, and a query that exited earlier remains outside its earlier shell. Consequently the batch fiber is exactly $\text{Cyl}(W)$, and W determines all individual labels and hence all exact replies. This proves (i) and (ii).

For (iii), fix one query x_i . In order for its exit label to extend σ by at least t bits, the point must lie in the interiors of the selected shells for the first $t - 1$ new levels. The depth- $(t - 1)$ descendant shells of J_σ have disjoint interiors. A fixed spatial point can therefore be contained in at most one of the 2^{t-1} possible descendant shells. Since the suffix of Θ is uniform,

$$\mathbb{P}\left(|W_{x_i}^\sigma(\Theta)| - |\sigma| \geq t \mid \Theta \in \text{Cyl}(\sigma)\right) \leq 2^{-(t-1)}.$$

If the maximum depth D is at least t , this event occurs for at least one of the b queries. A union bound, followed by the trivial upper bound one, yields (5.1). $\square$

A depth-three, two-query example. Condition on the first hidden bit being 0, so the prior is uniform on $\text{Cyl}(0)$. Suppose a realized two-point batch has exit labels 01 and 010. The labels are nested, their longest member is $W = 010$, and the complete numerical reply is determined by that prefix. The surviving leaves

are exactly $\text{Cyl}(010)$; their conditional probability is $2^{-(3-1)} = 1/4$, so this augmented label carries two bits of self-information, exactly the two new levels that it reveals. If $n = 3$ the posterior is a point mass; if $n > 3$ the suffix after 010 remains uniform.

This example isolates the four invariants used below: every nonempty single-query fiber is a cylinder; all labels in one batch are nested; the longest label leaves a uniform suffix; and the self-information of that label equals its added depth. The only remaining question is how large that depth can be, which is answered by the geometric tail in (5.1).

Lemma 5.2 (Conditional information of one batch). *Under the conditions of Lemma 5.1,*

$$H(W \mid \Theta \in \text{Cyl}(\sigma)) = \mathbb{E}[D \mid \Theta \in \text{Cyl}(\sigma)] \leq \log_2 b + 3$$

for $b \geq 1$, while the entropy is zero for $b = 0$.

Proof. By Lemma 5.1, every nonempty fiber indexed by a label w is a cylinder. Conditional on $\text{Cyl}(\sigma)$, its probability is

$$2^{-(|w| - |\sigma|)}.$$

Therefore the self-information of the realized label is exactly $|W| - |\sigma| = D$, and taking expectations gives $H(W) = \mathbb{E}D$.

For $b \geq 1$, let $\ell = \lceil \log_2 b \rceil$. The tail-sum formula and (5.1) give

$$\begin{aligned} \mathbb{E}D &= \sum_{t \geq 1} \mathbb{P}(D \geq t) \\ &\leq (\ell + 1) + b \sum_{t=\ell+2}^{\infty} 2^{-(t-1)} \\ &= \ell + 1 + b 2^{-\ell} \leq \ell + 2 \leq \log_2 b + 3. \end{aligned}$$

The finite depth $n - |\sigma|$ can only decrease the sum. □

5.2 Adaptive transcripts

We next pass from a fixed batch to R adaptive rounds. Let U denote the algorithm's complete private seed, sampled independently of Θ . Conceptually augment every round's exact reply by the longest exit prefix from Lemma 5.1. The algorithm is run unchanged: it chooses future queries from U and the exact numerical component only, and ignores the additional label.

Proposition 5.3 (Information per round). *Let $Y_{1:R}$ be the exact transcript of any private-coin algorithm that uses at most q queries in each of at most R rounds. Under the uniform depth- n hard prior,*

$$I(\Theta; Y_{1:R}, U) \leq R(\log_2 q + 3). \tag{5.2}$$

Proof. We first make the posterior invariant explicit. Condition on the private seed U . Before the first round, the compatible leaves form the cylinder $\text{Cyl}(\emptyset)$, and the posterior is uniform on it. Inductively, suppose that a realized augmented transcript through round $r - 1$ leaves the uniform posterior on a cylinder $\text{Cyl}(\sigma)$. The round- r batch is then fixed, because its locations are determined by U and the past exact replies. Item (ii) of Lemma 5.1 says that every nonempty fiber of its augmented reply is a subcylinder $\text{Cyl}(w)$. Conditioning on the realized reply therefore leaves a uniform posterior on that subcylinder. This proves the invariant for every round, including batches containing boundary or repeated query points.

Under this conditioning, [Lemma 5.2](#) bounds the entropy of the next augmented reply by $\log_2 q + 3$. Its conditional mutual information with Θ is no larger than this entropy.

Applying the chain rule over rounds gives

$$I(\Theta; \tilde{Y}_{1:R} \mid U) \leq R(\log_2 q + 3),$$

where $\tilde{Y}$ denotes the augmented transcript. Since U is independent of Θ , $I(\Theta; U) = 0$. Finally, the exact transcript Y is a deterministic projection of $\tilde{Y}$, so data processing gives

$$I(\Theta; Y_{1:R}, U) \leq I(\Theta; \tilde{Y}_{1:R}, U) \leq R(\log_2 q + 3).$$

If the algorithm stops before round R , pad the transcript with empty batches, which carry zero information. $\square$

5.3 From localization to decoding

The following elementary entropy calculation converts localization into recovery of the hidden address.

Lemma 5.4 (Finite Fano bound). *Let Θ be uniform on a set of size 2^n , and let T be any transcript from which an estimator $\hat{\Theta}$ has error probability at most $1/3$. Then*

$$I(\Theta; T) \geq \frac{2n}{3} - 1.$$

Proof. Let $E = \mathbf{1}\{\hat{\Theta} \neq \Theta\}$, and write $p = \mathbb{P}(E = 1) \leq 1/3$. Let

$$h_2(p) := -p \log_2 p - (1 - p) \log_2 (1 - p)$$

denote the binary entropy, with the usual convention $0 \log_2 0 = 0$. Since E is determined by (Θ, T) ,

$$\begin{aligned} H(\Theta \mid T) &\leq H(E \mid T) + H(\Theta \mid E, T) \\ &\leq h_2(p) + p \log_2 (2^n - 1) \leq 1 + \frac{n}{3}. \end{aligned}$$

Because $H(\Theta) = n$, subtracting the conditional entropy proves the claim. $\square$

Proposition 5.5 (Explicit round lower bound). *Let $n \geq 1$, and suppose that*

$$4\delta \leq s^n.$$

Every algorithm that is δ -accurate with probability at least $2/3$ on every function in the depth- n hard family satisfies

$$R \geq \frac{\frac{2}{3}n - 1}{\log_2 q + 3}.$$

Proof. By [Theorem 3.2](#) and [Item \(H3\)](#), distinct leaf minimizers are at distance at least $s^n \geq 4\delta$. Given the algorithm's output $\hat{x}$, decode the leaf whose minimizer is nearest to $\hat{x}$. Whenever $|\hat{x} - x_\Theta^*| \leq \delta$, every other minimizer is at distance at least $s^n - \delta \geq 3\delta$, so the decoder is correct.

Worst-case success at least $2/3$ implies average success at least $2/3$ under the uniform hard prior. Applying [Lemma 5.4](#) to the full transcript, including the private seed, gives

$$I(\Theta; Y_{1:R}, U) \geq \frac{2n}{3} - 1.$$

Combining this with [\(5.2\)](#) proves the proposition. $\square$

5.4 Completion of the main theorem

Proof of Theorem 3.1. Choose $n = n_\delta$ as in (3.1). Since $0 < s < 1$ and

$$n_\delta \leq \log_{1/s} \frac{1}{4\delta},$$

raising s to both sides reverses the inequality and gives

$$s^{n_\delta} \geq 4\delta.$$

When $n_\delta \geq 1$, Proposition 5.5 yields (3.2).

We separately record the one-round floor needed for the uniform statement. If $\delta \leq s/4$, use the two depth-one functions. Their minimizers are separated by s , so their closed δ -success regions are disjoint. A zero-round randomized output has the same distribution under both functions; the probabilities it assigns to two disjoint regions sum to at most one. Hence it cannot have success probability at least $2/3$ on both functions, and

$$R_{\text{rand}}^*(q, \delta; 1, \kappa) \geq 1. \quad (5.3)$$

The explicit lower bound, the one-round floor, and Proposition 2.6 imply (3.3); the elementary estimates keeping the constants uniform in q are given in Appendix A.2. The lower bound uses C^∞ instances and the upper bound holds on C^2 , so class inclusion gives both versions. Normalization by μ completes the general statement. $\square$

6 Consequences and changes of viewpoint

The main theorem is stated in minimizer distance because this is the cleanest loss for the decoding argument. We first translate it to objective suboptimality, taking care of constrained boundary minimizers. We then record the exact coordinate restriction that carries the result to every dimension.

6.1 Objective-gap accuracy

Let $R_{\text{gap,rand}}^*(q, \varepsilon; \mu, L)$ denote the least hard round cap of a private-coin algorithm that, for every $f \in \mathcal{F}_{\mu,L}^{(\infty)}$, returns $\hat{x} \in [0, 1]$ satisfying

$$f(\hat{x}) - f(x_f^*) \leq \varepsilon$$

with probability at least $2/3$. Define the deterministic quantity $R_{\text{gap,det}}^*$ analogously.

Corollary 6.1 (Objective-gap tradeoff). *Fix $\kappa = L/\mu > 1$, let*

$$s = \frac{\kappa - 1}{12(\kappa + 1)}, \quad \bar{\varepsilon} = \frac{\varepsilon}{\mu},$$

and suppose $0 < \bar{\varepsilon} \leq s^2/32$. Then, uniformly over $q \geq 1$,

$$R_{\text{gap,A}}^*(q, \varepsilon; \mu, L) = \Theta_\kappa \left(\max \left\{ 1, \frac{\log(1/\bar{\varepsilon})}{\log(q+1)} \right\} \right), \quad \text{A} \in \{\text{rand}, \text{det}\}.$$

More explicitly, with

$$n_\varepsilon = \left\lceil \log_{1/s} \frac{1}{4\sqrt{2\bar{\varepsilon}}} \right\rceil,$$

every randomized algorithm satisfies

$$R \geq 1, \quad R \geq \frac{\frac{2}{3}n_\varepsilon - 1}{\log_2 q + 3},$$

whereas a deterministic algorithm satisfies

$$R_{\text{gap,det}}^*(q, \varepsilon; \mu, L) \leq \left\lceil \frac{2}{q} \right\rceil + \max \left\{ 0, \left\lceil \log_{q+1} \sqrt{\frac{L}{8\varepsilon}} \right\rceil \right\}. \quad (6.1)$$

Proof. Strong convexity at a constrained minimizer gives, for every $x \in [0, 1]$,

$$f(x) - f(x_f^*) \geq f'(x_f^*)(x - x_f^*) + \frac{\mu}{2}|x - x_f^*|^2. \quad (6.2)$$

The first-order optimality condition on the interval states that

$$f'(x_f^*)(x - x_f^*) \geq 0.$$

For an interior minimizer this term is zero; at $x_f^* = 0$, both factors are nonnegative; at $x_f^* = 1$, both factors are nonpositive. Therefore (6.2) implies

$$f(x) - f(x_f^*) \geq \frac{\mu}{2}|x - x_f^*|^2. \quad (6.3)$$

An ε -accurate output is consequently

$$\delta := \sqrt{\frac{2\varepsilon}{\mu}} = \sqrt{2\bar{\varepsilon}}$$

close to the minimizer.

Multiply the normalized hard family by μ . Its curvature belongs to $[\mu, L]$, its minimizers are stationary and interior, and its exact value–gradient pairs are scaled consistently. Since

$$\bar{\varepsilon} \leq \frac{s^2}{32} \iff \delta \leq \frac{s}{4},$$

Theorem 3.1 applies with the displayed value of n_ε and proves the two lower bounds.

For the upper bound on the full constrained class, first query both endpoints. This costs $\lceil 2/q \rceil$ rounds. Since f' is strictly increasing, $f'(0) \geq 0$ identifies $x_f^* = 0$, while $f'(1) \leq 0$ identifies $x_f^* = 1$. In the remaining case,

$$f'(0) < 0 < f'(1),$$

and the minimizer is the unique stationary point. Apply the q -ary bracketing algorithm for T additional rounds. Its final midpoint is at distance at most $(q+1)^{-T}/2$ from x_f^* . Smoothness and $f'(x_f^*) = 0$ then give

$$f(\hat{x}) - f(x_f^*) \leq \frac{L}{2}|\hat{x} - x_f^*|^2 \leq \frac{L}{8}(q+1)^{-2T}.$$

Taking

$$T = \max \left\{ 0, \left\lceil \log_{q+1} \sqrt{\frac{L}{8\varepsilon}} \right\rceil \right\}$$

proves (6.1).

Finally,

$$\log \frac{1}{\delta} = \frac{1}{2} \log \frac{1}{\bar{\varepsilon}} - \frac{1}{2} \log 2.$$

On the stated accuracy range, $\log(1/\delta) = \Theta_\kappa(\log(1/\bar{\varepsilon}))$. The endpoint cost, ceilings, and the one-round lower bound are absorbed by the outer maximum uniformly in q . $\square$

Remark 6.2 (Why the endpoint phase is necessary). At a nonstationary boundary minimizer, smoothness does not give a purely quadratic upper bound on the gap. For

$$f_A(x) = Ax + \frac{\mu}{2}x^2, \quad A > 0,$$

the constrained minimizer is zero, but the gap contains the linear term Ax . Querying the endpoints first separates this case from the interior stationary regime. No such preprocessing is needed for distance loss, because gradient signs still maintain a bracket containing a boundary minimizer.

6.2 A lossless coordinate restriction

Proposition 6.3 (Known-coordinate lift). *Let $d \geq 1$, let*

$$X_d = [0, 1] \times [-1, 1]^{d-1},$$

and for $f \in \mathcal{F}_{\mu, L}^{(\infty)}$ define

$$F_f(x) = f(x_1) + \frac{\mu}{2} \sum_{j=2}^d x_j^2. \quad (6.4)$$

Then F_f is μ -strongly convex and L -smooth on X_d , its minimizer is $(x_f^*, 0, \dots, 0)$, and one exact high-dimensional first-order query $(F_f(x), \nabla F_f(x))$ can be simulated by one one-dimensional query $(f(x_1), f'(x_1))$. Consequently, every q -query, R -round lower bound proved above holds in every dimension with the same (q, R, μ, L, δ) .

Proof. The Hessian of F_f is diagonal:

$$\nabla^2 F_f(x) = \text{diag}(f''(x_1), \mu, \dots, \mu).$$

Its eigenvalues lie in $[\mu, L]$, which proves strong convexity and smoothness. Separability gives the stated minimizer, including when x_f^* is an endpoint.

For a query $x = (t, z)$, the one-dimensional oracle pair determines

$$F_f(t, z) = f(t) + \frac{\mu}{2} \|z\|_2^2, \quad \nabla F_f(t, z) = (f'(t), \mu z).$$

All transverse terms are known from the query point. In a batch, repeated first coordinates need be queried only once, so at most q one-dimensional pairs are used. Running a high-dimensional algorithm with the same private seed therefore reproduces its exact transcript and preserves the round boundaries.

If the high-dimensional output is $\hat{x}$, then

$$|\hat{x}_1 - x_f^*| \leq \|\hat{x} - (x_f^*, 0, \dots, 0)\|_2.$$

Thus a successful high-dimensional algorithm would yield a successful one-dimensional one with unchanged resources. For objective gap, the same conclusion follows from

$$F_f(t, z) - \min_{X_d} F_f = f(t) - f(x_f^*) + \frac{\mu}{2} \|z\|_2^2 \geq f(t) - f(x_f^*).$$

□

Remark 6.4 (What the lift does not provide). The difficult coordinate in (6.4) is public, and the lower bound does not grow with d . This is a lossless restriction argument, not a dimension-dependent parallel lower bound. Hiding a ridge direction naively does not solve the problem: for objectives of the form

$$f(\langle u, x \rangle) + \frac{\mu}{2} \|P_{u^\perp} x\|_2^2,$$

the gradient at the origin is often $f'(0)u$, which reveals u in one query. A genuine high-dimensional extension requires a composition principle that limits what one vector-valued reply can read from several hidden codes.

6.3 The boundary $\kappa = 1$

The assumption $\kappa > 1$ is necessary for the phenomenon as stated. If $\kappa = 1$, continuity and $\mu \leq f'' \leq L = \mu$ force $f'' \equiv \mu$. Hence

$$f'(x) = \mu x + c,$$

and one exact gradient query determines c and therefore the constrained minimizer $\Pi_{[0,1]}(-c/\mu)$. Consistently, the hard-family scale

$$s = \frac{\kappa - 1}{12(\kappa + 1)}$$

collapses to zero as $\kappa \downarrow 1$.

7 What the theorem explains, and what it leaves open

The theorem identifies a cost of adaptivity that remains after the geometry has been reduced to an interval. A query reaches one spatial address; parallelism permits more addresses to be tried before feedback arrives. The hidden suffix remains uniform after each augmented batch, and the deepest of q probes carries only $O(\log(q + 1))$ new bits in expectation. This is why the rate has the shape of ordered search even though a single exact reply may reveal many bits.

For fixed $\kappa > 1$, a single query per round requires $\Theta_\kappa(\log(1/\delta))$ rounds in the minimax sense. If $q = (1/\delta)^\alpha$, with fixed $\alpha > 0$, the rate becomes $\Theta_\kappa(\max\{1, 1/\alpha\})$. Large batches reduce synchronization cost, but this theorem does not characterize the least total work for each prescribed number of rounds.

Smoothness and scale. Every hard instance is C^∞ , and its condition number is bounded independently of the required accuracy. What the class does not control is a common scale on which the curvature must vary slowly. The factors D_k^{2-r} in higher derivatives allow progressively finer structure. Thus infinite differentiability alone does not remove this minimax obstruction. Adding a common Hessian-Lipschitz or higher-derivative bound would be a different question. Analytic functions require another construction altogether: two real-analytic functions on a connected neighborhood that agree on an open exterior interval agree everywhere, so the exact compact shielding used here is unavailable.

One dimension and fixed dimension. The known-coordinate lift proves that the scalar obstruction persists in any dimension. It does not prove that the obstruction becomes stronger. A vector gradient may reveal information about several coordinates simultaneously, so placing several independent scalar copies side by side does not by itself multiply the number of rounds. The value of the one-dimensional theorem is to make

one source of hardness fully explicit: adaptive access to unresolved local structure under an exact numerical oracle. A genuine dimension-sensitive theorem must additionally control how many such structures one vector reply can expose.

Remaining quantitative questions. The present bound is uniform in q , but its constants are not sharp in κ ; indeed the construction scale vanishes as $\kappa \downarrow 1$, when the class collapses to quadratics of known curvature. Exact randomized constants and success-sensitive converses remain open within the present analysis. The Fano step uses only a fixed success probability and does not establish a strong converse for very small success probabilities.

The interaction model also matters. History-dependent widths $b_r \leq q$, repeated points, empty batches, and early stopping under a hard round cap are already covered. A width allowed to exceed every uniform hard cap, subject only to a total or expected-query budget, is not characterized by the theorem. Nor does it determine the minimax tradeoff under noisy or finite-precision replies.

The construction and its information interface are separate. The former realizes exact exterior agreement with fixed curvature bounds; the latter turns that agreement and disjoint interval geometry into a round lower bound. This separation is the reusable part of the argument. Any broader extension must preserve the appropriate oracle interface, not only the appearance of a nested search tree.

References

- Eric Balkanski and Yaron Singer. Parallelization does not accelerate convex optimization: Adaptivity lower bounds for non-smooth convex minimization, 2018.
- Simina Brânzei, Dimitris Paparas, and Nicholas Recker. Searching, sorting, and cake cutting in rounds, 2020. URL <https://arxiv.org/abs/2012.00738v5>. Version 5, revised November 2023.
- Jelena Diakonikolas and Cristóbal Guzmán. Lower bounds for parallel and randomized convex optimization. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1132–1157, 2019. URL <https://proceedings.mlr.press/v99/diakonikolas19c.html>.
- John Duchi, Feng Ruan, and Chulhee Yun. Minimax bounds on stochastic batched convex optimization. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 3065–3162, 2018. URL <https://proceedings.mlr.press/v75/duchi18a.html>.
- Shmuel Gal and Willard Miranker. Optimal sequential and parallel search for finding a root. *Journal of Combinatorial Theory, Series A*, 23(1):1–14, 1977. doi: 10.1016/0097-3165(77)90074-7.
- Stefan Heinrich and Jörg-Detlef Kern. Parallel information-based complexity. *Journal of Complexity*, 7(4): 339–370, 1991. doi: 10.1016/0885-064X(91)90024-R.
- Arkadi Nemirovski. On parallel complexity of nonsmooth convex optimization. *Journal of Complexity*, 10(4):451–463, 1994. doi: 10.1006/jcom.1994.1025.
- Erich Novak. Determining zeroes of increasing lipschitz functions. *Aequationes Mathematicae*, 41(1):161–167, 1991. doi: 10.1007/BF02227451.

Erich Novak and Klaus Ritter. Some complexity results for zero finding for univariate functions. *Journal of Complexity*, 9(1):15–40, 1993. doi: 10.1006/jcom.1993.1003.

K. Sikorski. Bisection is optimal. *Numerische Mathematik*, 40(1):111–117, 1982. doi: 10.1007/BF01459080.

A Additional details

A.1 An explicit smooth linear plateau

Proof of Lemma 4.2. Define an odd function $B_0 : \mathbb{R} \rightarrow \mathbb{R}$ by specifying its nonnegative half:

$$B_0(y) = \begin{cases} y, & 0 \leq y \leq 1/8, \\ P(y - 1/8), & 1/8 \leq y \leq 11/24, \\ 0, & y \geq 11/24, \end{cases} \quad P(t) = \frac{1}{8} + t - \frac{75}{8}t^2 + \frac{63}{4}t^3.$$

The transition interval has length $1/3$. Direct differentiation gives

$$P'(t) = 1 - \frac{75}{4}t + \frac{189}{4}t^2, \quad P''(t) = -\frac{75}{4} + \frac{189}{2}t.$$

In particular,

$$P(0) = \frac{1}{8}, \quad P'(0) = 1, \quad P(1/3) = 0, \quad P'(1/3) = 0.$$

Thus B_0 and B'_0 match at both positive joins; odd reflection gives the negative joins, and $B_0(y) = y$ near zero. Consequently $B_0 \in C_c^1(\mathbb{R})$. Its weak second derivative is the piecewise ordinary second derivative, with no point masses because B'_0 is continuous. On the positive transition, P'' runs from $-75/4$ to $51/4$; hence

$$\|B''_0\|_{L^\infty} \leq 75/4.$$

Let $a = 1/48$, and choose the normalized even mollifier

$$\eta(t) = \begin{cases} c \exp(-1/(1 - (t/a)^2)), & |t| < a, \\ 0, & |t| \geq a, \end{cases} \quad \int_{\mathbb{R}} \eta(t) dt = 1.$$

It is nonnegative and C^∞ ; at $|t| = a$, each one-sided derivative tends to zero: writing $z = 1 - (t/a)^2$, one has $e^{-1/z} = o(z^m)$ as $z \downarrow 0$ for every $m > 0$, which dominates the inverse-power factors produced by differentiation. Set

$$b(x) = (B_0 * \eta)(x - \frac{1}{2}).$$

Convolution with a smooth compact kernel is smooth, by differentiation under the integral. Its support is contained in

$$\frac{1}{2} + [-11/24, 11/24] + [-1/48, 1/48] = [1/48, 47/48].$$

If $y = x - 1/2$ satisfies $|y| \leq 5/48$, then $|y - t| \leq 1/8$ for every $t \in \text{supp } \eta$. Therefore

$$b(x) = \int_{\mathbb{R}} (y - t)\eta(t) dt = y = x - \frac{1}{2},$$

using the zero first moment of the even kernel. Finally, weak differentiation of the convolution agrees with classical differentiation:

$$b''(x) = (B_0'' * \eta)(x - \tfrac{1}{2}), \quad |b''(x)| \leq \int_{\mathbb{R}} |B_0''(x - \tfrac{1}{2} - t)| \eta(t) dt \leq 75/4.$$

This proves all three properties. $\square$

A.2 Uniform constants in the round bound

Constant calculation for Theorem 3.1. Set $A = \log_2(1/\delta)$, $B = \log_2(1/s) > 1$, and $H = \log_2(q+1) \geq 1$. Then $n_\delta = \lfloor (A-2)/B \rfloor$. If $A \geq 5B+4$, the floor inequality gives

$$\frac{2}{3}n_\delta - 1 \geq \frac{2A-4-5B}{3B} \geq \frac{A}{3B}, \quad \log_2 q + 3 \leq H + 3H = 4H.$$

The explicit lower bound therefore yields $R \geq A/(12BH)$. Together with $R \geq 1$ from (5.3), this implies the left side of (3.3), whether or not $A/H \geq 1$. If $A < 5B+4 < 12B$, the same conclusion follows directly from

$$\frac{1}{12B} \max\{1, A/H\} \leq 1 \leq R.$$

For the upper bound, $\delta \leq s/4$ ensures $A > 1$. Hence

$$\left\lceil \log_{q+1} \frac{1}{2\delta} \right\rceil = \left\lceil \frac{A-1}{H} \right\rceil \leq 1 + \frac{A-1}{H} \leq 2 \max\{1, A/H\}.$$

Now apply Proposition 2.6 and the inclusion of deterministic algorithms among randomized ones. Scaling by μ , as in Remark 2.5, gives the general curvature parameters. $\square$

A.3 The number of shielding moments is exact

Lemma 4.1 gives a sufficient condition for hiding several successive antiderivatives. At the right endpoint of the active interval, the same condition is also necessary.

Proposition A.1 (Necessity of the moment list). *Let $h \in L^1([0, 1])$ be supported almost everywhere on $[\alpha, \beta]$, and fix $r \geq 1$. The following are equivalent:*

- (i) $V^k h(\beta) = 0$ for $k = 1, \dots, r$;
- (ii) $\int_{\alpha}^{\beta} t^j h(t) dt = 0$ for $j = 0, \dots, r-1$.

Proof. By (4.1),

$$(k-1)! V^k h(\beta) = \int_{\alpha}^{\beta} (\beta-t)^{k-1} h(t) dt.$$

The polynomials

$$1, (\beta-t), \dots, (\beta-t)^{r-1}$$

form a basis for the vector space of polynomials in t of degree at most $r-1$. Therefore the r integrals in (i) vanish if and only if the integral of h against every polynomial of degree at most $r-1$ vanishes. Taking the monomial basis $1, t, \dots, t^{r-1}$ gives (ii). $\square$

For the standard first-order oracle, $r = 2$. The proposition formalizes the claim that mass and first moment are not merely one convenient way to obtain shielding: they are exactly the two linear constraints required to annihilate the gradient and value differences at the far boundary.

A.4 Changing the length of the domain

The unit interval is a normalization. Suppose that the original domain is $[\alpha, \alpha + D]$ with $D > 0$, and define

$$\tilde{f}(u) = \frac{f(\alpha + Du)}{\mu D^2}, \quad 0 \leq u \leq 1.$$

Then

$$\tilde{f}'(u) = \frac{f'(\alpha + Du)}{\mu D}, \quad \tilde{f}''(u) = \frac{f''(\alpha + Du)}{\mu}.$$

Thus $\mu \leq f'' \leq L$ is equivalent to $1 \leq \tilde{f}'' \leq \kappa$. Exact pair queries in either coordinate system deterministically simulate exact pair queries in the other. A physical distance tolerance δ_x becomes δ_x/D , and a physical objective-gap tolerance ε becomes

$$\frac{\varepsilon}{\mu D^2}.$$

These are the dimensionless quantities that should appear inside the logarithms of [Theorem 3.1](#) and [Corollary 6.1](#).