

Scale-Adaptive Cut-and-Flow for Stationary-Point Search on Periodic Domains

Xiaoyu Li

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

Abstract

On a periodic domain, recursive localization becomes cheaper when each separator is sampled at the scale of the region that remains. For a function $f : \mathbb{T}^d \rightarrow \mathbb{R}$ with a 1-periodic C^1 lift and 1-Lipschitz gradient, we turn this observation into a deterministic exact value–gradient algorithm that finds a queried point x with $\|\nabla f(x)\| \leq \varepsilon$ in

$$O_d\left(\varepsilon^{-2(d-1)/(d+1)}\right)$$

queries; in dimension three the bound is $O(1/\varepsilon)$. At spatial scale R , sampling a separator and descending below its true minimum cost $(R/\delta)^{d-1} + \delta^2/\varepsilon^2$. Balancing at the current scale makes these costs geometrically summable. A face-complete net retains quadratic accuracy at lower-dimensional face minima. As a finite stress test, the same principle gives an $O(\sqrt{K}n)$ coefficient-query algorithm for periodic integer potentials whose second differences are bounded by K . As a boundary test, we show why the argument does not directly compactify $\mathbb{R}^d$: a bounded smooth function may admit no bounded exact trap, and fixed full-range additive walls face a radius–cancellation dichotomy. Thus the unrestricted three-dimensional problem remains open.

1 Introduction

Why does gradient descent pay $1/\varepsilon^2$ when space has only three dimensions? The usual proof gives a precise but incomplete answer. If the gradient is larger than ε , one unit gradient step decreases a 1-smooth function by more than $\varepsilon^2/2$. A bounded range can finance only $O(\varepsilon^{-2})$ such steps. Dimension never enters: the argument treats a curve in $\mathbb{R}^3$ exactly as it treats one in a space with a million coordinates.

The natural way to use dimension is to stop following a single trajectory and inspect the space around it. In three dimensions, a box can be divided by a two-dimensional separator. If that separator is known to lie above a current pivot, one half can be discarded without losing the point trapped inside. Repeating the operation localizes a stationary point. The quantitative question is simple: how accurately should one sample each separator, and how much descent is needed to turn an approximate sample into an exact certificate?

Everything in this paper follows from one answer:

the separator mesh should track the scale of the region being cut.

Suppose the current box has longest side R , and a mesh of width δ is placed on its middle face. In three dimensions the mesh costs $O((R/\delta)^2)$ queries. Smoothness makes its sampled minimum accurate up to

$O(\delta^2)$, and gradient descent needs $O(\delta^2/\varepsilon^2)$ further queries either to find an ε -stationary point or to descend below the true separator minimum. Thus one cut costs

$$O\left(\left(\frac{R}{\delta}\right)^2 + \frac{\delta^2}{\varepsilon^2}\right). \quad (1.1)$$

The choice $\delta = \sqrt{R\varepsilon}$ makes this $O(R/\varepsilon)$. Longest-side bisection produces a geometric sequence of scales, so $\sum R = O(1)$. A small box is cheap because it is not sampled with the mesh chosen for the original box.

We develop this idea first on the flat torus, where periodicity supplies an initial boundary. For $d \geq 2$, let $f : \mathbb{T}^d \rightarrow \mathbb{R}$ have a 1-periodic C^1 lift whose gradient is 1-Lipschitz. A query returns the exact pair $(f(x), \nabla f(x))$, and the output must be a queried point. We prove that a deterministic algorithm finds x with $\|\nabla f(x)\| \leq \varepsilon$ using

$$O_d\left(\varepsilon^{-\frac{2(d-1)}{d+1}}\right) \quad (1.2)$$

queries. In particular, the complexity on $\mathbb{T}^3$ is $O(1/\varepsilon)$.

The closest precedents use different boundaries, oracles, and notions of stationarity, so there is no model-free ranking. The direct method-level comparison is the Cut-and-Flow theorem of [Bubeck and Mikulincer \(2024, Theorem 3\)](#). With combined value–gradient queries on a box, its fixed-resolution analysis costs

$$O_d\left(\log(d/\varepsilon) \varepsilon^{-2(d-1)/(d+1)}\right)$$

for a projected boundary criterion. Choosing the resolution from the current scale removes that logarithm. The normalized compact-box method of [Vavasis \(1993\)](#), also for a projected criterion, gives $O_d(\varepsilon^{-2d/(d+2)})$, or $O(\varepsilon^{-6/5})$ in dimension three.

A stronger zero-order comparator is due to [Hollender and Zampetakis \(2025, Theorem 4.2\)](#). For an L -smooth function on a box, their algorithm finds an ε -KKT point using

$$O(d)^{(5d-1)/4} \left(\sqrt{L/\varepsilon}\right)^{d-1} \quad (1.3)$$

value queries. With L fixed, their ε -exponent is better in dimension two, equal to ours in dimension three, and larger in dimensions $d \geq 4$. Neither result dominates the other: value-only box-KKT search and combined-oracle ordinary stationarity on a torus are different problems. Moreover, their proof already uses scale-varying, face-complete nets. The precise delta of our continuous theorem is therefore narrower: an exact-barrier Cut-and-Flow proof on a boundaryless periodic domain, with a log-free amortization relative to the original Cut-and-Flow bound.

The proof revolves around one invariant. For a box H and a queried pivot $p \in H$, maintain

$$f(p) < \min_{x \in \partial H} f(x). \quad (1.4)$$

We call (H, p) an *exact trap*. A descent trajectory starting no higher than p cannot leave H : every point on a unit gradient segment has value no larger than its starting point, whereas every boundary point has larger value. The whole segment matters here; checking only the discrete iterates would leave open the possibility of crossing the boundary between two queries.

Now bisect a longest side of H and call the middle face E . Query a mesh on E , let y be its sampled minimum, and descend from the lower-valued point among p and y . Unless stationarity is found, a short descent ends at a point q below $\min_E f$. The trajectory remains in H , and q lies strictly in one half. Keeping

that half restores the trap: its old boundary is above p , and its new boundary is the separator E , now above q .

There is one boundary case that cannot be hidden in notation. A minimizer of a two-dimensional face may lie in the relative interior of an edge. An arbitrary Euclidean net that merely contains the vertices need not place a nearby query on that edge; a normal linear term can then make the sampling error $\Theta(\delta)$ rather than $O(\delta^2)$. The needed convention is already present in the “nice δ -nets” of [Hollender and Zampetakis \(2025, Definition 4.3 and Lemmas 8–9\)](#) and appears again in [Zhou et al. \(2025, Definition 28\)](#). We call such a net *boundary-complete*: every coordinate subface receives its own mesh. At a minimizer, take the smallest stratum containing it. The tangential gradient vanishes there, and the smoothness inequality gives the desired quadratic error. This changes the query count only by a dimension-dependent constant. [Appendix A](#) gives the exact-minimum specialization used here and a two-dimensional diagnostic showing why an arbitrary net is insufficient.

The same resource balance has a finite counterpart. There is a useful complexity contrast behind this model. Generic direction-preserving discrete Brouwer search on a three-dimensional grid has deterministic query complexity $\Theta(n^2)$ ([Chen and Deng, 2008](#)). Our oracle reveals much more structure: exact values of one scalar potential, hence an integrable edge field with zero periods, together with bounded second differences. The theorem below shows that these promises change the query landscape; it is not an algorithm for the generic Brouwer oracle.

We introduce an integer potential on $(\mathbb{Z}/n\mathbb{Z})^3$ whose pure and mixed second differences have magnitude at most K . The associated *balanced-cube* task asks for a constant-size cell on which, in each coordinate direction, the relevant first differences take both weak signs. We prove an

$$O(\sqrt{K} n) \tag{1.5}$$

coefficient-query bound, linear in n for fixed K and independent of the amplitude of the potential. Here an h -spaced separator mesh costs $O((R/h)^2)$, bounded second differences make its value error $O(Kh^2)$, and integrality turns that error into a bound on the number of strict descent moves. The choice $h \asymp R^{1/2}K^{-1/4}$ again makes the scale costs geometric. A periodic cubic-spline representation shows how approximate spline stationarity decodes to the balanced sign condition. Thus the linear bound is a structural separation between potential access and a generic direction-field oracle, not a claim that three-dimensional fixed-point search is generically linear.

Compactness is not cosmetic. The bounded 1-smooth function

$$f_\infty(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x_1) \tag{1.6}$$

has no critical point and, more strongly, no bounded region contains a pivot strictly below its entire boundary. A correct algorithm may have to move toward infinity, where the gradient becomes small, rather than trap a critical point. A fixed additive wall does not repair this for free: a wall that raises the boundary by the full possible range of the objective either has radius $\Omega(1/\varepsilon)$ or can be cancelled at an interior point by a legal objective whose original gradient still has norm 2ε .

These statements delimit the paper. We do not prove an $O(1/\varepsilon)$ upper bound for arbitrary bounded smooth functions on $\mathbb{R}^3$. We do not prove an oracle lower bound against adaptive walls or continuation. We do not prove that the torus or balanced-cube rate is tight. Let $Q_3^{(1)}(\varepsilon)$ be the minimum worst-case number of exact value–gradient queries needed to return a queried ε -stationary point, uniformly over 1-smooth functions $f : \mathbb{R}^3 \rightarrow [0, 1]$. The unrestricted three-dimensional frontier remains

$$\Omega(\varepsilon^{-1}) \leq Q_3^{(1)}(\varepsilon) \leq O(\varepsilon^{-2}); \tag{1.7}$$

the lower bound follows from the two-dimensional construction of [Hollender and Zampetakis \(2025\)](#), embedded as a function independent of the third coordinate, while the upper bound is gradient descent. For the exact-trap program studied here, these results isolate the missing operation. Once a trap exists, recursive cutting is inexpensive; extending this architecture to the noncompact domain requires either initializing a trap or producing an escape certificate when no bounded trap exists.

Organization. [Section 2](#) formalizes the oracle model, boundary-complete nets, and the trap invariant. [Section 3.1](#) proves the scale-adaptive torus theorem; [Section 3.2](#) gives the discrete periodic analogue. The obstruction at infinity appears in [Section 3.3](#). Complete auxiliary proofs, including the face-net counterexample, the spline bridge, and the sharp quadratic-wall witness, are collected in the appendices.

2 Geometry, descent, and exact traps

All norms are Euclidean. We identify the flat torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ with the unit cube whose opposite faces are identified. A function on $\mathbb{T}^d$ is called 1-smooth when its 1-periodic lift, denoted by the same letter, is C^1 and satisfies

$$\|\nabla f(x) - \nabla f(y)\| \leq \|x - y\| \quad (x, y \in \mathbb{R}^d). \quad (2.1)$$

An exact combined-oracle query at x returns $(f(x), \nabla f(x))$. Algorithms are deterministic, arithmetic and comparisons are exact, and the output must be a queried point. We write $Q_{\mathbb{T}^d}^{(1)}(\varepsilon)$ for the minimum worst-case number of such queries needed to return an ε -stationary point in this periodic class.

Throughout, sampled minima and choices among lower neighbors are resolved by a fixed lexicographic order. This convention makes every displayed procedure deterministic; none of the proofs depends on which fixed order is chosen.

The smoothness inequality will be used in two different ways. The first is the usual descent estimate, strengthened from an endpoint statement to the whole gradient segment.

Lemma 2.1 (Full-segment descent). *If f is 1-smooth, then for every $x \in \mathbb{R}^d$ and $0 \leq t \leq 1$,*

$$f(x - t\nabla f(x)) \leq f(x) - t(1 - t/2)\|\nabla f(x)\|^2. \quad (2.2)$$

In particular, a unit step from a point with gradient norm larger than ε decreases the value by more than $\varepsilon^2/2$.

Proof. The descent lemma gives

$$f(x + s) \leq f(x) + \langle \nabla f(x), s \rangle + \frac{1}{2}\|s\|^2.$$

Substitute $s = -t\nabla f(x)$. The endpoint claim is the case $t = 1$. □

The second use of smoothness is to estimate the minimum of a coordinate face. The qualification “coordinate” is important: it gives a canonical collection of lower-dimensional strata. Our mesh convention is the “nice δ -net” condition of [Hollender and Zampetakis \(2025, Definition 4.3\)](#): the mesh remains a net after restriction to every face. We use different terminology only to foreground the boundary strata needed by the exact-minimum argument.

Definition 2.2 (Boundary-complete mesh). Let E be an m -dimensional coordinate rectangle. A coordinate stratum of E is obtained by declaring each coordinate free, fixed at its lower endpoint, or fixed at its upper endpoint. A mesh $N_\delta(E)$ is *boundary-complete with spacing δ* if, on every stratum, it is the Cartesian product of one-dimensional endpoint-complete grids whose consecutive gaps are at most δ .

Thus every point of a k -dimensional stratum has a mesh point in the same stratum at distance at most $\sqrt{k}\delta/2$. If all side lengths of E are at most R , then, for fixed m ,

$$|N_\delta(E)| = O_m(1 + (R/\delta)^m). \quad (2.3)$$

One may realize the definition by a single endpoint-complete Cartesian grid; the explicit stratum language records the property the proof needs.

Lemma 2.3 (Quadratic face approximation). *Let E be an m -dimensional coordinate rectangle, let f be 1-smooth on a neighborhood of E , and let $N_\delta(E)$ be boundary-complete. Then*

$$\min_{q \in N_\delta(E)} f(q) \leq \min_{x \in E} f(x) + \frac{m\delta^2}{8}. \quad (2.4)$$

Proof. Let u minimize f on E , and let F be the unique smallest coordinate stratum whose relative interior contains u . First-order optimality on the affine hull of F gives

$$\langle \nabla f(u), z - u \rangle = 0 \quad (z \in F). \quad (2.5)$$

If $k = \dim F$, boundary completeness supplies $q \in N_\delta(E) \cap F$ with $\|q - u\| \leq \sqrt{k}\delta/2$. Smoothness and (2.5) yield

$$f(q) \leq f(u) + \langle \nabla f(u), q - u \rangle + \frac{1}{2}\|q - u\|^2 \leq f(u) + \frac{k\delta^2}{8} \leq f(u) + \frac{m\delta^2}{8}.$$

The sampled minimum is no larger than $f(q)$. □

The smallest-stratum step is load-bearing. An arbitrary Euclidean net that only promises to contain the vertices can have linear, rather than quadratic, error when the minimizer lies inside an edge. A concrete example appears in [Appendix A](#).

We now state the invariant that turns a sampled separator into a localization algorithm. For a coordinate box $H = \prod_{i=1}^d [\ell_i, r_i]$, write

$$R(H) = \max_i (r_i - \ell_i).$$

Definition 2.4 (Exact trap). A coordinate box H and a queried point $p \in \text{int } H$ form an *exact trap* if

$$f(p) < \min_{z \in \partial H} f(z). \quad (2.6)$$

The point p is the *pivot*.

Lemma 2.5 (A trapped descent cannot escape). *Suppose (H, p) is an exact trap, and let $x_0 \in H$ satisfy $f(x_0) \leq f(p)$. Every unit-gradient segment generated from x_0 remains in H for as long as the iteration is run.*

Proof. By [Lemma 2.1](#), the value at every point of every unit-gradient segment is at most $f(x_0) \leq f(p)$. If a segment first left H , it would meet some $z \in \partial H$ with $f(z) \leq f(p)$, contradicting (2.6). □

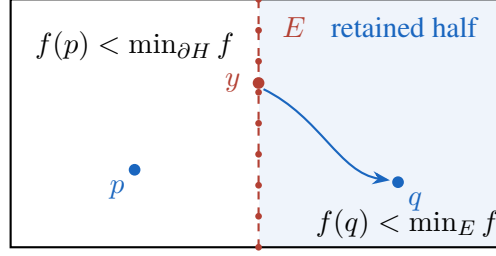


Figure 1: One sealed cut, shown in a two-dimensional cross-section. The middle face is sampled, descent starts from the lower-valued point in $\{p, y\}$, and the half containing the resulting point q inherits the exact trap.

2.1 A finite shadow of stationarity

The discrete analogue lives on $\Lambda_n = (\mathbb{Z}/n\mathbb{Z})^3$, with $n \geq 4$. For an array $a : \Lambda_n \rightarrow \mathbb{Z}$, define backward differences

$$\Delta_r a_v = a_v - a_{v-e_r}, \quad \Delta_r \Delta_s a_v = \Delta_r (\Delta_s a)_v. \quad (2.7)$$

The oracle returns one exact coefficient a_v per query. This is a unit-cost query model: the bit length of an integer answer is not charged, and no runtime or bit-complexity bound is claimed. Given $A > 0$ and $K \geq 0$, the promise is

$$a_v \in \mathbb{Z} \cap [-An^2, An^2], \quad |\Delta_r \Delta_s a_v| \leq K \quad (v \in \Lambda_n, 1 \leq r, s \leq 3). \quad (2.8)$$

For a cell index $c \in \Lambda_n$, put

$$I_j(c) = \{c_j - 3, c_j - 2, c_j - 1, c_j\} \pmod{n}, \quad (2.9)$$

$$S_r(c) = \{v : v_r \in \{c_r - 2, c_r - 1, c_r\}, \quad v_s \in I_s(c) \text{ for } s \neq r\}. \quad (2.10)$$

Definition 2.6 (Balanced cube). A cell c is *balanced* if, for each $r = 1, 2, 3$,

$$\min_{v \in S_r(c)} \Delta_r a_v \leq 0 \leq \max_{v \in S_r(c)} \Delta_r a_v. \quad (2.11)$$

Let $T_{\text{BC}}(n; A, K)$ denote the deterministic worst-case coefficient-query complexity of finding such a cell.

The stencil is not arbitrary. A periodized tensor cubic spline maps every approximate stationary point to a cell satisfying (2.11).

Proposition 2.7 (Spline interpretation). *For every array satisfying (2.8), there is a C^2 , $n\mathbb{Z}^3$ -periodic tensor cubic spline G_a such that*

$$|G_a(x)| \leq An^2, \quad \text{Lip}(\nabla G_a) \leq 3K.$$

Each exact query $(G_a(x), \nabla G_a(x))$ can be simulated with at most 64 coefficient queries. If $\|\nabla G_a(x)\| \leq 1/2$, then $c = \lfloor x \rfloor \pmod{n}$ is balanced. In particular, the balanced-cube problem is total.

The construction and all derivative identities are proved in [Appendix B](#). The algorithm in [Section 3.2](#) works directly with coefficients; the spline proposition explains the stationarity content of its output but is not used in its query bound.

3 Scale-adaptive trapping

3.1 The continuous periodic theorem

Set

$$C_d = \frac{d-1}{8}, \quad \alpha_d = \frac{2(d-1)}{d+1}. \quad (3.1)$$

The constant C_d is the face-approximation loss from [Lemma 2.3](#); the exponent α_d will emerge by balancing two costs, not by an induction.

Theorem 3.1 (Scale-adaptive Cut-and-Flow). *Let $d \geq 2$ and $0 < \varepsilon \leq 1$. For every 1-smooth $f : \mathbb{T}^d \rightarrow \mathbb{R}$, a deterministic exact combined-oracle algorithm returns a queried point x with*

$$\|\nabla f(x)\| \leq \varepsilon$$

using

$$O_d(\varepsilon^{-\alpha_d}) = O_d(\varepsilon^{-2(d-1)/(d+1)}) \quad (3.2)$$

queries. In particular,

$$Q_{\mathbb{T}^3}^{(1)}(\varepsilon) = O(1/\varepsilon). \quad (3.3)$$

Proof idea. Maintain a queried pivot strictly below the entire boundary of the current box. A boundary-complete δ -mesh estimates the minimum of a middle face to error $O_d(\delta^2)$, and $O_d(\delta^2/\varepsilon^2)$ unsuccessful gradient steps descend below the true face minimum and seal one half. Choosing δ from the current longest side R makes one cut cost $O_d((R/\varepsilon)^{\alpha_d})$. Longest-side bisection makes these scale costs summable, while the periodic seam creates the initial trap. We implement this outline in three steps: one sealed cut, seam initialization, and scale summation.

Lemma 3.2 (One sealed cut). *Let (H, p) be an exact trap, let $R = R(H)$, and let E be the common face obtained by bisecting a longest side of H . Fix $\delta > 0$. After*

$$O_d\left(1 + (R/\delta)^{d-1} + \delta^2/\varepsilon^2\right) \quad (3.4)$$

queries, an algorithm either returns a queried ε -stationary point or produces an exact trap on one of the two half-boxes.

Proof. Query a boundary-complete δ -mesh of E , let y minimize the queried values, and let $\bar{p}$ be the lower-valued point in $\{p, y\}$. Both points have been queried.

Let u minimize f on E . If $u \in \partial E$, then $\partial E \subseteq \partial H$, so the old trap gives $f(p) < f(u)$. In this case $f(\bar{p}) < \min_E f$. If u is in the relative interior of E , [Lemma 2.3](#) gives

$$f(y) \leq f(u) + C_d \delta^2.$$

In both cases we have the single inequality

$$f(\bar{p}) \leq \min_E f + C_d \delta^2. \quad (3.5)$$

Starting from $\bar{p}$, perform unit gradient steps. At every queried point, stop if the gradient norm is at most ε . Otherwise execute at most

$$M_\delta = \left\lceil \frac{2C_d \delta^2}{\varepsilon^2} \right\rceil + 1 \quad (3.6)$$

steps. If all M_δ steps are executed, [Lemma 2.1](#) and (3.5) show that the queried endpoint q satisfies

$$f(q) < \min_E f. \quad (3.7)$$

The whole trajectory remains in H . Indeed, $f(\bar{p}) \leq f(p)$ by construction, so [Lemma 2.5](#) applies. By (3.7), $q \notin E$. Retain the half-box containing q . Its boundary is contained in $\partial H \cup E$; the new pivot q lies below both pieces, so it forms a new exact trap. Finally, (2.3) and (3.6) give (3.4). $\square$

Lemma 3.3 (Periodic seam initialization). *For every $\delta > 0$, after*

$$O_d \left(1 + \delta^{-(d-1)} + \delta^2 / \varepsilon^2 \right) \quad (3.8)$$

queries, the algorithm either returns a queried ε -stationary point or produces an exact trap in the cut cube $[0, 1]^d$.

Proof. Cut the torus along the coordinate seams and write $\Sigma = \partial[0, 1]^d$ for their union. Query a boundary-complete δ -mesh on each of the $2d$ faces and let y be the minimum of all queried values. If u minimizes f on Σ , then u belongs to at least one face. Applying [Lemma 2.3](#) on that face gives

$$f(y) \leq \min_{\Sigma} f + C_d \delta^2. \quad (3.9)$$

Run unit-step gradient descent intrinsically on the torus from y , with the budget M_δ from (3.6). If no queried iterate is ε -stationary, the endpoint p has value strictly below $\min_{\Sigma} f$. It therefore has a representative in the interior of the cut cube and satisfies

$$f(p) < \min_{z \in \partial[0, 1]^d} f(z).$$

This is the desired exact trap. The mesh and descent counts give (3.8). Notice that wrapping is harmless before the trap is created; afterward, [Lemma 2.5](#) prevents the lifted trajectory from crossing the seam. $\square$

Proof of Theorem 3.1. At a box of longest side R , use

$$\delta_R = \min \left\{ R, R^{(d-1)/(d+1)} \varepsilon^{2/(d+1)} \right\}. \quad (3.10)$$

Ignoring the truncation for one line, balancing the two nonconstant terms in (3.4) means solving

$$(R/\delta)^{d-1} = \delta^2 / \varepsilon^2.$$

This is exactly the second term in (3.10), and both sides then equal $(R/\varepsilon)^{\alpha_d}$. If the untruncated choice exceeds R , then $R < \varepsilon$; choosing $\delta_R = R$ makes both the grid and descent costs $O_d(1)$. Hence every cut costs

$$O_d(1 + (R/\varepsilon)^{\alpha_d}). \quad (3.11)$$

Always bisect a longest side, and stop once $\text{diam}(H) \leq \varepsilon$. At every nonterminal cut,

$$R \geq \text{diam}(H) / \sqrt{d} > \varepsilon / \sqrt{d},$$

so the additive constant in (3.11) is absorbed into the main term, up to a factor depending only on d . Longest-side bisection keeps the aspect ratio at most two and, if R_t is the longest side before cut t ,

$$R_{t+d} \leq R_t / 2. \quad (3.12)$$

Consequently,

$$\sum_t R_t^{\alpha_d} \leq \frac{d}{1 - 2^{-\alpha_d}} = O_d(1). \quad (3.13)$$

Indeed, grouping cuts into consecutive blocks of d , each block reduces the longest side by at least a factor two, and the displayed bound is the resulting geometric series. The total cost of all cuts is therefore $O_d(\varepsilon^{-\alpha_d})$.

Use (3.10) with $R = 1$ in Lemma 3.3. The initialization has the same order. It remains to justify the stopping rule. Let x_* minimize f over the final box H . The trap is strict, so x_* lies in the interior and $\nabla f(x_*) = 0$. The pivot p has already been queried, and (2.1) gives

$$\|\nabla f(p)\| = \|\nabla f(p) - \nabla f(x_*)\| \leq \|p - x_*\| \leq \text{diam}(H) \leq \varepsilon.$$

Thus the algorithm outputs a queried ε -stationary point within the claimed budget. $\square$

In dimension three, (3.4) reads

$$O\left((R/\delta)^2 + \delta^2/\varepsilon^2 + 1\right).$$

The choice $\delta = \sqrt{R\varepsilon}$ makes the leading cost $O(R/\varepsilon)$, and (3.13) reduces to the elementary fact $\sum_t R_t = O(1)$. This is the entire logarithm removal.

Remark 3.4 (Rescaling). If the torus has side length S and ∇f is L -Lipschitz, rescale $g(u) = f(Su)/(LS^2)$. Then g is 1-smooth and an $\varepsilon/(LS)$ -stationary point of g is an ε -stationary point of f . Thus Theorem 3.1 becomes $O_d((LS/\varepsilon)^{\alpha_d})$ when $L > 0$; the case $L = 0$ is immediate.

3.2 A discrete periodic mirror

The next theorem uses the same invariant without calculus. Smoothness becomes bounded second differences, and the guaranteed decrease $\varepsilon^2/2$ becomes the unit decrease of a strict move in an integer potential.

Theorem 3.5 (Linear fixed-curvature balanced-cube search). *There is a universal constant C such that, for every $n \geq 4$, $A > 0$, and $K \geq 1$,*

$$T_{\text{BC}}(n; A, K) \leq C\sqrt{K}n + C. \quad (3.14)$$

The bound is independent of A . If $0 \leq K < 1$, every promised array is constant, so any cell is balanced without a query.

Proof idea. The invariant is the same exact trap, now for an integer potential. Bounded second differences replace smoothness and give $O(Kh^2)$ face error, while integrality turns that error into at most $O(Kh^2)$ strict descent moves. Balancing this against the $O((R/h)^2)$ separator cost gives $h \asymp R^{1/2}K^{-1/4}$ and $O(\sqrt{K}R)$ queries at scale R . Periodic boundary layers initialize the trap, and a terminal weak local minimum directly supplies all three balanced signs.

We begin with the discrete counterpart of Lemma 2.3. A cyclic h -mesh has wraparound gaps at most h ; on a nonperiodic rectangle, the one-dimensional meshes include both endpoints.

Lemma 3.6 (Discrete quadratic face approximation). *Let E be a two-dimensional coordinate rectangle or periodic coordinate layer, and let u be a relative-interior minimizer of a on E . Query a Cartesian mesh with gaps at most the integer $h \geq 1$, and let y minimize its queried values. Define*

$$L_h = 2\lfloor h/2 \rfloor, \quad B_h = \frac{K}{2}L_h(L_h + 1). \quad (3.15)$$

Then

$$a_y \leq a_u + B_h. \quad (3.16)$$

Proof. Choose a nearest mesh point w in each free coordinate. There is a coordinate path $u = z_0, z_1, \dots, z_L = w$ within E of length $L \leq 2\lfloor h/2 \rfloor = L_h$.

Fix an oriented free-coordinate edge. At u , the increments in its two orientations are nonnegative because u is a relative-interior minimum. Their sum is a pure second difference in magnitude at most K , so either oriented increment lies in $[0, K]$. Moving the base of a fixed oriented increment by one face edge changes it by a pure or mixed second difference, again in absolute value at most K . The base of the t th edge of the path is at graph distance at most $t - 1$ from u . Its oriented increment is therefore at most $K + (t - 1)K = Kt$. Summing the path increments gives

$$a_w - a_u \leq K \sum_{t=1}^L t = \frac{K}{2} L(L + 1) \leq B_h.$$

Since y is the sampled minimum, $a_y \leq a_w$. □

Proof of Theorem 3.5. We write a consecutive-coordinate lattice box as H and call its vertices with at least one endpoint coordinate ∂H . The discrete trap is

$$p \in H \setminus \partial H, \quad a_p < \min_{v \in \partial H} a_v. \quad (3.17)$$

Initialization. Represent Λ_n by $\{0, \dots, n - 1\}^3$, and let Σ be the union of the six boundary layers $v_r \in \{0, n - 1\}$. Query a cyclic h -mesh in the two free coordinates of each layer, and let y minimize all queried coefficients. If u minimizes a on Σ , it minimizes every periodic layer containing it. Applying Lemma 3.6 on one such layer gives

$$a_y \leq \min_{\Sigma} a + B_h. \quad (3.18)$$

Starting at y , repeatedly query the six torus neighbors and move to a strictly lower one. If none exists, y is a *weak torus local minimum*, meaning $a_y \leq a_{y \pm e_r}$ for every $r \in \{1, 2, 3\}$. Otherwise each move decreases the integer value by at least one. After $\lfloor B_h \rfloor + 1$ moves, the current queried point p satisfies $a_p < \min_{\Sigma} a$, which is (3.17) for the cut cube.

One cut. Suppose (3.17) holds. Bisect a longest side into two inclusive lattice boxes that overlap on their median layer E . Query an endpoint-complete h -mesh of E , let y be its sampled minimum, and begin strict neighbor descent at the lower-valued point $\bar{p} \in \{p, y\}$.

Let u minimize a on E . If $u \in \partial E$, then $\partial E \subseteq \partial H$ and the old trap gives $a_p < a_u$. Otherwise Lemma 3.6 gives $a_y \leq a_u + B_h$. In both cases,

$$a_{\bar{p}} \leq \min_E a + B_h. \quad (3.19)$$

Unless a local minimum is encountered first, after $\lfloor B_h \rfloor + 1$ strict moves the current point q satisfies

$$a_q < \min_E a. \quad (3.20)$$

The descent path cannot leave H . Every visited value is at most $a_{\bar{p}} \leq a_p$, while an exit path must visit a boundary vertex whose value is strictly larger than a_p . Equation (3.20) also implies $q \notin E$. Retaining the half containing q restores (3.17) because its boundary is contained in $\partial H \cup E$.

Termination and decoding. Stop before a degenerate cut, once every side has at most four lattice edges, and query the remaining constant-size box. Its global minimum v is not on the boundary by (3.17); all six torus neighbors therefore lie in the box, so v is a weak torus local minimum. The same conclusion holds if neighbor descent stopped earlier. Set $c = v + (1, 1, 1)$. For every coordinate r , local minimality gives

$$\Delta_r a_v = a_v - a_{v-e_r} \leq 0, \quad \Delta_r a_{v+e_r} = a_{v+e_r} - a_v \geq 0.$$

Both indices belong to $S_r(c)$, so (2.11) holds.

Queries. Let R be the longest side length in lattice edges. A separator mesh costs $O(1 + (R/h)^2)$ coefficient queries. With cached values, the descent budget costs $O(1 + Kh^2)$ queries when $h > 1$. Choose

$$h_R = \max \left\{ 1, \left\lceil R^{1/2} K^{-1/4} \right\rceil \right\}. \quad (3.21)$$

If the unrounded optimizer is at least one, both costs are $O(\sqrt{K} R)$. If it is below one, then $h_R = 1$, the entire separator is queried, and $L_{h_R} = B_{h_R} = 0$. Only one attempted strict move is needed, and $R \leq \sqrt{K}$ implies $R^2 \leq \sqrt{K} R$. The same bound therefore holds in this regime.

The six-layer initialization has the same cost with $R = O(n)$. Longest-side bisection satisfies $R_{t+3} \leq \lceil R_t/2 \rceil$ until the terminal constant scale, and hence

$$\sum_t R_t = O(n). \quad (3.22)$$

As in the continuous proof, blocks of three cuts halve the longest side up to the terminal additive ceiling; the full ceiling bookkeeping appears in Appendix C. The additive constant per cut contributes $O(\log n)$, absorbed by $O(n)$. This proves (3.14) for $K \geq 1$.

Finally, if $K < 1$, every integer pure second difference vanishes. Along each cyclic coordinate fiber, the corresponding first difference is constant; its sum around the fiber is zero, so it vanishes. This holds in all three directions, and the array is constant. $\square$

Three assumptions have three separate jobs. Bounded second differences give the quadratic face estimate, integrality counts strict descent moves, and periodicity seals the initial cut. The amplitude parameter A is never used. No matching $\Omega(n)$ lower bound is known.

3.3 Why the unrestricted domain is different

The preceding algorithms begin only after a low pivot has been enclosed by a higher boundary. Periodicity creates such a boundary by cutting a seam. On $\mathbb{R}^d$, the missing initialization is a real obstruction rather than a notational inconvenience.

Theorem 3.7 (Trap and wall obstructions). *For every finite $d \geq 1$, the following statements hold.*

(i) *The function*

$$f_\infty(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x_1) \quad (3.23)$$

maps $\mathbb{R}^d$ into $(0, 1)$, has 1-Lipschitz gradient, and has no critical point. For every bounded nonempty open set $U \subset \mathbb{R}^d$ and every $p \in U$,

$$f_\infty(p) \geq \min_{z \in \partial U} f_\infty(z). \quad (3.24)$$

Thus f_∞ admits no bounded exact trap.

(ii) Let $0 < \varepsilon \leq 1/4$. Let $H \subset \mathbb{R}^d$ be a compact convex body, $p \in \text{int } H$, and let P be C^1 on a neighborhood of H . Suppose the fixed wall raises the boundary by the full possible objective range:

$$P(z) - P(p) > 1 \quad (z \in \partial H). \quad (3.25)$$

Then at least one of the following holds:

$$\text{dist}(p, \partial H) > \frac{1}{2\varepsilon}; \quad (3.26)$$

or there exist a globally bounded 1-smooth $f : \mathbb{R}^d \rightarrow [0, 1]$ and $x \in \text{int } H$ such that

$$\nabla(f + P)(x) = 0, \quad \|\nabla f(x)\| = 2\varepsilon. \quad (3.27)$$

Proof idea. The monotone ridge defeats every bounded trap because moving in the $-e_1$ direction always lowers its value. For a fixed wall, split according to whether $\|\nabla P\|$ ever exceeds 2ε . If it does not, the fundamental theorem of calculus forces the boundary to be farther than $1/(2\varepsilon)$. If it does, continuity from an interior minimizer of P produces a point with $\|\nabla P\| = 2\varepsilon$, where a bounded tanh ridge cancels the wall exactly.

Proof. For part (i), direct differentiation gives

$$\nabla f_\infty(x) = \frac{1}{\pi(1 + x_1^2)} e_1 \neq 0. \quad (3.28)$$

Moreover,

$$\text{Lip}(\nabla f_\infty) = \sup_t \frac{2|t|}{\pi(1 + t^2)^2} = \frac{9}{8\pi\sqrt{3}} < 1.$$

Fix bounded open U and $p \in U$. The first coordinate attains its minimum on $\overline{U}$ at a point z . Such a point lies on ∂U , since an interior point can be moved slightly in the $-e_1$ direction. The same local move at p shows $z_1 < p_1$. Strict monotonicity of the arctangent gives $f_\infty(z) < f_\infty(p)$, proving (3.24).

For part (ii), first suppose

$$\sup_{x \in H} \|\nabla P(x)\| \leq 2\varepsilon. \quad (3.29)$$

Let z be a nearest boundary point to p . Convexity puts the segment $[p, z]$ inside H , so (3.25) and the fundamental theorem of calculus give

$$1 < P(z) - P(p) \leq \int_0^1 \|\nabla P(p + t(z - p))\| \|z - p\| dt \leq 2\varepsilon \|z - p\|.$$

This is (3.26).

Otherwise, some $b \in H$ satisfies $\|\nabla P(b)\| > 2\varepsilon$. A minimizer a of P on H lies in the interior: every boundary value is larger than $P(p)$ by (3.25). Hence $\nabla P(a) = 0$. Along the segment from a to b , continuity supplies an interior point x with $\|\nabla P(x)\| = 2\varepsilon$. If b lies on the boundary, take the first level crossing; it occurs before the endpoint, and every open segment from an interior point of a convex body stays in the interior.

Put $u = \nabla P(x)/(2\varepsilon)$ and define

$$f(y) = \frac{1}{2} - 2\varepsilon \tanh\langle u, y - x \rangle. \quad (3.30)$$

Since $\varepsilon \leq 1/4$, this function maps into $[0, 1]$. At x , $\nabla f(x) = -2\varepsilon u = -\nabla P(x)$. Writing $s = \langle u, y - x \rangle$, its Hessian has operator norm

$$4\varepsilon |\operatorname{sech}^2(s) \tanh(s)| \leq 4\varepsilon \leq 1.$$

Thus f is legal and (3.27) holds. $\square$

The wall condition (3.25) is sufficient to dominate every $f : \mathbb{R}^d \rightarrow [0, 1]$: for $z \in \partial H$,

$$(f + P)(z) - (f + P)(p) > -1 + 1 = 0.$$

The theorem says that this robust compactification has a price. In the first branch, a box has scale $R = \Omega(1/\varepsilon)$. Even under the favorable additional assumption that $f + P$ retains the $O(1)$ smoothness needed by the trapped-box routine, its rescaled $O(R/\varepsilon)$ guarantee is then quadratic. In the second branch, an arbitrary stationary point of $f + P$ need not be stationary for f .

The quantifiers matter. [Theorem 3.7](#) rules out the direct strategy of choosing one f -independent full-range wall and immediately accepting an arbitrary stationary point of the regularized objective. It is not an oracle lower bound. It does not exclude adaptive or data-dependent walls, continuation, a specified stationary-point selection rule, or using a cancellation point as an escape certificate. A quadratic regularizer exhibits the radius–bias tradeoff with matching scales; see [Appendix D](#).

4 What the scale calculation reveals

The continuous and discrete proofs are the same argument paid for in different currencies. Their correspondence is summarized in [Table 1](#).

Table 1: The two scale-matched trapping arguments.

	Smooth periodic function	Integral periodic potential
separator cost	$(R/\delta)^{d-1}$	$(R/h)^2$
face-minimum error	$O(\delta^2)$	$O(Kh^2)$
descent currency	$> \varepsilon^2/2$ per failed query	≥ 1 per strict move
scale-matched mesh	$R^{(d-1)/(d+1)} \varepsilon^{2/(d+1)}$	$R^{1/2} K^{-1/4}$
three-dimensional scale cost	$O(R/\varepsilon)$	$O(\sqrt{K} R)$
initial trap	periodic seam	periodic boundary layers

This dictionary separates the reusable mechanism from its assumptions. The trap needs exact value comparisons. Quadratic face approximation needs either smoothness or bounded second differences. Progress below the sampled face needs a countable decrease: $\varepsilon^2/2$ in the smooth problem and one unit in the integral problem. In the two recursive algorithms, the decisive geometric use of periodicity is to create the first boundary. Cyclicity is also used by the spline interpretation and by the $K < 1$ degeneracy argument.

The exponent in [Theorem 3.1](#) should be read accordingly. It is not an optimality claim. Scale-varying face-complete nets already appear in the GFPT algorithm of [Hollender and Zampetakis \(2025\)](#); their value-only box-KKT bound has ε -exponent $(d-1)/2$. Hence their exponent is smaller for $d = 2$, equal for $d = 3$, and larger for $d \geq 4$, although the different oracle and output criterion prevent a dominance claim. Relative to the original combined-oracle Cut-and-Flow analysis of [Bubeck and Mikulincer \(2024\)](#), the exact-barrier

amortization here removes the fixed-resolution logarithm. As d grows, our exponent $2(d - 1)/(d + 1)$ approaches the dimension-free gradient-descent exponent two. This matches the geometric intuition: sampling a codimension-one separator becomes expensive in high dimension.

The noncompact theorem changes the question. The monotone ridge (3.23) is not hard in the ordinary sense—its gradient becomes arbitrarily small at infinity. It is hard only for the logic of exact trapping: there may be no critical point to enclose. Any extension of this exact-trap architecture to the unrestricted problem must therefore support two outcomes. It should either create a sealed region or certify that moving farther out is productive. We call this missing primitive *trap or escape*.

Three concrete questions remain.

1. **Can trap or escape be implemented subquadratically in $\mathbb{R}^3$?** An adaptive wall must handle points where $\|\nabla P\|$ is comparable to ε and cancellation can occur. The next step is either to turn such a point into a valid cut or escape certificate, or to prove that locating and interpreting these points costs quadratic query complexity.
2. **Is balanced-cube search linear for a reason that is also necessary?** For fixed K , Theorem 3.5 gives $O(n)$ queries, but no matching $\Omega(n)$ lower bound is known. Exact potential values, integrability, and bounded curvature obstruct a direct import of the usual hidden-path lower bounds from generic local search.
3. **Which compact geometries admit cheap seams?** The recursive proof only needs an efficiently sampled initial boundary and coordinate-like cuts whose new boundary can be sealed. Product manifolds and domains with a known cut system are natural candidates, but the cost must include the geometry of that system rather than hide it in a compactness assumption.

The first question is the one tied to the original three-dimensional frontier. The present paper does not close that frontier. Its narrower conclusion is that, after initialization, the recursion is no longer mysterious: buy each separator at the scale of the region that remains.

References

- Sébastien Bubeck and Dan Mikulincer. How to trap a gradient flow. *SIAM Journal on Computing*, 53(4): 803–824, 2024. doi: 10.1137/21M1397854.
- Xi Chen and Xiaotie Deng. Matching algorithmic bounds for finding a Brouwer fixed point. *Journal of the ACM*, 55(3):13:1–13:26, 2008. doi: 10.1145/1379759.1379761.
- Alexandros Hollender and Manolis Zampetakis. The computational complexity of finding stationary points in non-convex optimization. *Mathematical Programming*, 213:281–341, 2025. doi: 10.1007/s10107-024-02139-3.
- Stephen A. Vavasis. Black-box complexity of local minimization. *SIAM Journal on Optimization*, 3(1): 60–80, 1993. doi: 10.1137/0803004.
- Huanjian Zhou, Andi Han, Akiko Takeda, and Masashi Sugiyama. The adaptive complexity of finding a stationary point. In *Proceedings of the Thirty-Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 6091–6123, 2025.

A Why boundary completeness is necessary

[Lemma 2.3](#) used the mesh point lying in the same smallest stratum as the face minimizer. This face-complete convention is the “nice δ -net” of [Hollender and Zampetakis \(2025, Definition 4.3\)](#); the smallest-face proof also appears in their Lemma 9. The following example is a diagnostic for our specialization: it shows that an arbitrary Euclidean net cannot replace that property, even if every vertex is queried.

Proposition A.1 (A vertex-complete net with linear error). *For every $0 < \delta < 1/4$, there are a 1-smooth function on $[0, 1]^2$ and a Euclidean δ -net N containing all four vertices such that*

$$\min_{q \in N} f(q) - \min_{x \in [0, 1]^2} f(x) \geq \delta/2. \quad (\text{A.1})$$

Proof. Let

$$f(x, y) = x + \frac{1}{2} \left(y - \frac{1}{2} \right)^2. \quad (\text{A.2})$$

Its Hessian is $\text{diag}(0, 1)$, so it is 1-smooth, and its minimum is $f(0, 1/2) = 0$.

Choose a finite set $X \subset [\delta/2, 1]$ containing $\delta/2$ and 1, with consecutive gaps at most $\delta/2$. Choose $Y \subset [0, 1]$ containing both endpoints, also with consecutive gaps at most $\delta/2$, and set

$$N = (X \times Y) \cup \{(0, 0), (0, 1)\}. \quad (\text{A.3})$$

Every point of the square is within horizontal distance $\delta/2$ and vertical distance $\delta/4$ of $X \times Y$, hence within $\sqrt{5} \delta/4 < \delta$. The four vertices belong to N .

Every nonvertex query has $x \geq \delta/2$ and therefore value at least $\delta/2$. The two added vertices on $x = 0$ have value $1/8$. Because $\delta < 1/4$, the minimum queried value is at least $\delta/2$, proving (A.1). $\square$

The minimizer in this example lies in the relative interior of the edge $x = 0$, where the inward normal derivative is one. Smoothness does not control that linear term. Boundary completeness keeps the active coordinate fixed and uses only directions in which first-order optimality holds. It supplies the face estimate needed here at unchanged fixed-dimensional asymptotic cost. The convention and smallest-face principle are prior art; only the elementary exact-minimum formulation and the diagnostic are recorded here for a complete proof.

B The periodic cubic-spline bridge

We prove [Proposition 2.7](#). Let

$$B_p = \mathbf{1}_{[0, 1]}^{*(p+1)}$$

be the cardinal B-spline of degree p , and periodize it at period n by

$$B_{p,n}(t) = \sum_{\ell \in \mathbb{Z}} B_p(t - \ell n). \quad (\text{B.1})$$

These functions are nonnegative, satisfy the partition of unity

$$\sum_{j \in \mathbb{Z}/n\mathbb{Z}} B_{p,n}(t - j) = 1, \quad (\text{B.2})$$

and obey

$$B'_{p,n}(t) = B_{p-1,n}(t) - B_{p-1,n}(t - 1), \quad p = 2, 3. \quad (\text{B.3})$$

Define

$$G_a(x) = \sum_{v \in \Lambda_n} a_v \prod_{j=1}^3 B_{3,n}(x_j - v_j). \quad (\text{B.4})$$

The function is C^2 and $n\mathbb{Z}^3$ -periodic. By (B.2), it is a convex combination of the coefficients, so $|G_a(x)| \leq An^2$.

At any point, at most four indices per coordinate carry positive cubic-spline weight. Thus the value and all three gradient coordinates in one exact query use at most $4^3 = 64$ coefficients; period aliases can only reduce this number.

Differentiating (B.4), using (B.3), and shifting the periodic summation index give

$$\partial_r G_a(x) = \sum_v \Delta_r a_v B_{2,n}(x_r - v_r) \prod_{t \neq r} B_{3,n}(x_t - v_t). \quad (\text{B.5})$$

For $r \neq s$, with $t \notin \{r, s\}$,

$$\partial_{rs} G_a(x) = \sum_v \Delta_r \Delta_s a_v B_{2,n}(x_r - v_r) B_{2,n}(x_s - v_s) B_{3,n}(x_t - v_t), \quad (\text{B.6})$$

while

$$\partial_{rr} G_a(x) = \sum_v \Delta_r^2 a_v B_{1,n}(x_r - v_r) \prod_{t \neq r} B_{3,n}(x_t - v_t). \quad (\text{B.7})$$

Every weight family in (B.5)–(B.7) is nonnegative and sums to one. The promise therefore bounds every Hessian entry by K in absolute value. The maximum absolute row sum is at most $3K$, and hence

$$\|\nabla^2 G_a(x)\|_{\text{op}} \leq 3K, \quad \text{Lip}(\nabla G_a) \leq 3K. \quad (\text{B.8})$$

It remains to prove the decoder. Let $c = \lfloor x \rfloor \bmod n$. The positive weights in (B.5) use only indices in $S_r(c)$: the quadratic spline contributes the three indices $c_r - 2, c_r - 1, c_r$, while each cubic spline contributes the four indices in $I_t(c)$. Thus $\partial_r G_a(x)$ is a convex combination of $\{\Delta_r a_v : v \in S_r(c)\}$. If $\|\nabla G_a(x)\| \leq 1/2$, then

$$\min_{v \in S_r(c)} \Delta_r a_v \leq \frac{1}{2}, \quad \max_{v \in S_r(c)} \Delta_r a_v \geq -\frac{1}{2}.$$

The differences are integers, so their minimum is at most zero and their maximum is at least zero. This is (2.11).

Finally, G_a attains a minimum on its compact torus, where its gradient vanishes. The decoder therefore produces a balanced cell for every promised array, proving totality and completing Proposition 2.7.

C Discrete increment transport and bookkeeping

For completeness, we record the signed identities behind Lemma 3.6. Define the two oriented increments

$$D_{r,+}(v) = a_{v+e_r} - a_v = \Delta_r a_{v+e_r}, \quad D_{r,-}(v) = a_{v-e_r} - a_v = -\Delta_r a_v. \quad (\text{C.1})$$

At a relative-interior face minimum u , both are nonnegative and

$$D_{r,+}(u) + D_{r,-}(u) = \Delta_r^2 a_{u+e_r}. \quad (\text{C.2})$$

The right side has absolute value at most K ; since the left side is nonnegative, each initial oriented increment belongs to $[0, K]$.

Shifting the base of $D_{r,+}$ by $+e_s$ changes it by

$$D_{r,+}(v + e_s) - D_{r,+}(v) = \Delta_s \Delta_r a_{v+e_r+e_s}, \quad (\text{C.3})$$

and shifting by $-e_s$ gives the negative of a translated mixed difference. The same calculation for $D_{r,-}$ differs only by an overall sign and an index shift. This includes $r = s$, when the difference is pure. Consequently, moving the base of any fixed oriented increment through one face edge changes its value by at most K in absolute value. Transporting the orientation of the t th path edge back to u through at most $t - 1$ base shifts proves the bound Kt used in [Lemma 3.6](#).

A cyclic one-dimensional mesh with gaps at most h puts every lattice point within cyclic distance $\lfloor h/2 \rfloor$ of a mesh point; the same statement holds for an endpoint-complete interval mesh. Taking one shortest move in each free coordinate gives the known path-length budget $L_h = 2\lfloor h/2 \rfloor$. This budget depends only on the chosen mesh, not on the unknown face minimizer, so the algorithmic stopping rule is executable.

For odd lattice side lengths, choose either median layer and use inclusive halves. A side of R edges becomes sides of at most $\lceil R/2 \rceil$ edges. Within any three consecutive longest-side cuts, every coordinate that could remain longer than $\lceil R/2 \rceil$ is cut. Hence $R_{t+3} \leq \lceil R_t/2 \rceil$ until the constant terminal scale, which implies (3.22). Query caching can only reduce the count: each strict move requires at most the six previously unseen neighbor coefficients.

D A tight quadratic-regularization witness

The wall dichotomy in [Theorem 3.7](#) is qualitative. Ordinary quadratic regularization realizes its radius–bias tradeoff with matching scales.

Proposition D.1 (Quadratic radius versus bias). *For $0 < \lambda \leq 1$, define*

$$P_\lambda(x) = \frac{\lambda}{2} \|x\|^2, \quad f_\lambda(x) = \frac{1}{2} - \frac{1}{4} \tanh(\sqrt{\lambda} x_1). \quad (\text{D.1})$$

Then $f_\lambda : \mathbb{R}^d \rightarrow [1/4, 3/4]$ is 1-smooth, and $f_\lambda + P_\lambda$ has a unique stationary point

$$x_\lambda = (z_*/\sqrt{\lambda}, 0, \dots, 0), \quad (\text{D.2})$$

where $z_* \in (0, 1/4)$ is the unique solution of

$$z_* = \frac{1}{4} \operatorname{sech}^2 z_*. \quad (\text{D.3})$$

At this point,

$$\|x_\lambda\| = \frac{z_*}{\sqrt{\lambda}}, \quad \|\nabla f_\lambda(x_\lambda)\| = z_* \sqrt{\lambda}. \quad (\text{D.4})$$

Proof. The range is immediate. Writing $s = \sqrt{\lambda} x_1$, the only nonzero Hessian entry of f_λ has magnitude

$$\frac{\lambda}{2} |\operatorname{sech}^2(s) \tanh(s)| \leq \frac{\lambda}{2} \leq 1.$$

Thus f_λ is 1-smooth.

Stationarity in coordinates $2, \dots, d$ forces $x_j = 0$. In the first coordinate, after setting $z = \sqrt{\lambda} x_1$, the equation is

$$z = \frac{1}{4} \operatorname{sech}^2 z.$$

There is no negative solution. On $z > 0$, the left side is strictly increasing and the right side is positive and strictly decreasing; their values cross once between 0 and 1/4. This proves uniqueness and (D.2). The stationarity equation also gives

$$\|\nabla f_\lambda(x_\lambda)\| = \frac{\sqrt{\lambda}}{4} \operatorname{sech}^2 z_* = z_* \sqrt{\lambda},$$

and the radius formula is immediate. □

For a guarantee uniform over the function class, forcing the original gradient in (D.4) below ε requires $\lambda \leq \varepsilon^2/z_*^2$. At that choice, the stationary point may lie at distance at least z_*^2/ε . The adversarial objective is allowed to depend on the fixed regularization parameter; this is exactly the quantifier order in a worst-case uniform guarantee.