

Three-step dimension gaps for gradient descent

Logan Bell

September 2026

Abstract

For a prescribed three-step schedule, the worst terminal objective value of gradient descent can be strictly larger in two dimensions than in one. We establish this separation throughout two explicit neighborhoods of a rational schedule. On the larger cube, of radius $1/59$, the same planar gradients and function values are interpolable along every schedule's trajectory and give a common terminal value above a certified scalar upper bound; the interpolating function may depend on the schedule. This radius is maximal for that geometry and terminal target. On the smaller cube, of radius $3/196$, a single explicit planar function chosen before the schedule witnesses the strict gap throughout. The scalar estimates combine an exact evaluation at the center with a perturbation bound, while the planar estimates use smooth-convex interpolation.

1 Introduction

Let $\mathcal{F}_d$ be the class of differentiable convex functions $f : \mathbf{R}^d \rightarrow \mathbf{R}$ whose gradients are 1-Lipschitz and whose sets of minimizers are nonempty. For a three-step schedule $h = (h_1, h_2, h_3)$, gradient descent forms

$$x_{k+1} = x_k - h_{k+1} \nabla f(x_k), \quad k = 0, 1, 2, \quad (1)$$

and its normalized worst case in dimension d is

$$W_d(h) = \sup \left\{ f(x_3) - f(x_\star) : \begin{array}{l} f \in \mathcal{F}_d, \quad x_0 \in \mathbf{R}^d, \quad x_\star \in \operatorname{argmin} f, \quad \|x_0 - x_\star\| \leq 1, \\ x_1, x_2, x_3 \text{ obey (1)} \end{array} \right\}. \quad (2)$$

Every d -dimensional instance embeds in $\mathbf{R}^{d+1}$, so $W_d(h) \leq W_{d+1}(h)$. Both constructions below are centered at the rational schedule

$$a = \left(\frac{9}{5}, \frac{249}{100}, \frac{123}{100} \right). \quad (3)$$

At this schedule, $W_1(a) = 25/602 = 0.041528\dots$, while the planar constructions attain terminal values above 0.064. We extend this separation to neighborhoods of the center schedule.

The two main theorems differ in the order of their quantifiers. Theorem 5 treats every schedule in the closed cube of radius $1/59$: it uses the same sampled values and gradients throughout the cube, and interpolation supplies a function f_h after h is specified. Theorem 7 instead chooses an explicit function f° before the schedule and proves the gap on the smaller cube of radius $3/196$; its off-center samples follow the trajectory of that fixed function.

Smooth-convex interpolation represents finitely many sampled iterates and gradients by a positive semidefinite Gram matrix, whose rank determines the required ambient dimension [1].

Daccache produced a numerical regime map for arbitrary fixed step sizes, and Eloi later reported rank-two solutions for three-step schedules [2, 3]. Luner and Grimmer reported a numerical dimension effect for an extrapolated output after seven unit steps [4]. Das Gupta, Van Parys, and Ryu developed computational global optimization for unrestricted three-step performance estimation [5], while Grimmer, Shu, and Wang derived an exact value for another three-step schedule that is attained in one dimension [6]. We prove an exact separation between scalar and planar worst cases for the ordinary terminal iterate on explicit schedule neighborhoods.

2 Planar interpolation data

A finite collection (x_i, g_i, F_i) is interpolable by a convex function with a 1-Lipschitz gradient if and only if

$$F_i - F_j - \langle g_j, x_i - x_j \rangle - \frac{1}{2} \|g_i - g_j\|^2 \geq 0 \quad \text{for every ordered pair } (i, j). \quad (4)$$

This is the smooth-convex interpolation theorem of Taylor, Hendrickx, and Glineur [1, Corollary 1], with smoothness constant one. It is convenient to rewrite (4) as the difference constraint

$$F_j - F_i \leq w_{ij} := \langle g_j, x_j - x_i \rangle - \frac{1}{2} \|g_i - g_j\|^2, \quad (5)$$

where the weights w_{ij} depend on the sampled locations and gradients.

Both constructions use the following points and gradients. In $\mathbf{R}^2$, fix the initial point and four gradients

$$\begin{aligned} x_0 &= \frac{1}{1000}(-995, 95), \\ g_0 &= \frac{1}{1000}(-993, 87), \quad g_1 = \frac{1}{1000}(442, 304), \\ g_2 &= \frac{1}{1000}(-240, -170), \quad g_3 = \frac{1}{1000}(-50, -122), \end{aligned} \quad (6)$$

adjoin a stationary sample $(x_*, g_*, F_*) = (0, 0, 0)$, and let the remaining locations follow the gradient-descent recurrence with the prescribed gradients:

$$x_i(h) = x_0 - \sum_{k < i} h_{k+1} g_k, \quad i = 1, 2, 3. \quad (7)$$

The gradients are fixed while the locations move affinely with the schedule. Whenever a convex 1-smooth function interpolates these data, its gradient-descent trajectory from x_0 under h passes exactly through the locations (7), because the prescribed gradients drive the recurrence (1). The stationary sample forces the interpolant to be minimized at the origin with value zero.

Figure 1 shows the resulting trajectory at the center schedule. Its iterates are not collinear.

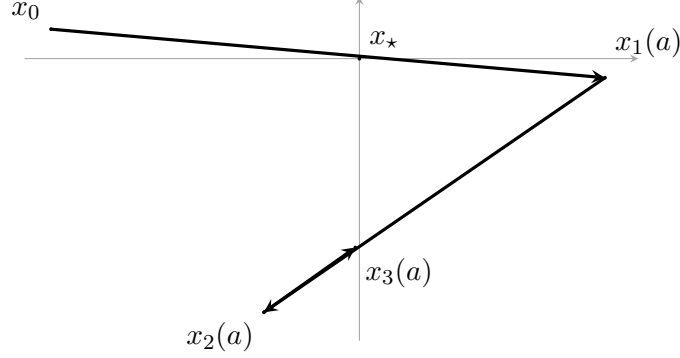


Figure 1: The planar trajectory at the center schedule a . The arrows show the three gradient-descent updates, and the origin is the stationary minimizer.

The initial norm and the determinant of the first two gradients are

$$\|x_0\|^2 = \frac{19981}{20000} < 1, \quad \det[g_0 \ g_1] = -\frac{170163}{500000} \neq 0. \quad (8)$$

The first inequality is the normalized initial-radius constraint in (2); the nonzero determinant shows that the sampled gradients span two dimensions.

The two constructions assign different values to these samples. Section 5 assigns values that interpolate throughout the larger cube. Section 6 uses another value vector to define a fixed function at the center.

3 Exact scalar values

We compute the scalar value at the center schedule exactly. The finite enumeration below also applies to an arbitrary rational schedule. For a horizon N , write $W_{N,1}(h)$ for its scalar worst case; thus $W_{3,1}(h) = W_1(h)$ in the notation of this paper.

3.1 Orders and interpolation polytopes

Translate a minimizer to zero and reflect the line if necessary. It suffices to start at 1. Indeed, an instance starting at $r \in (0, 1]$ rescales to $F(u) = (f(ru) - f(0))/r^2$ starting at 1; the same schedule applies and its gap is divided by r^2 . The zero starting point has zero gap.

Fix $h \in \mathbf{R}^N$ and write $g_i = f'(x_i)$, $0 \leq i \leq N$. Set $x_* = g_* = 0$ and

$$x_i = 1 - \sum_{j=0}^{i-1} h_{j+1} g_j.$$

Consider each ordering π of $*, 0, \dots, N$ in which $*$ precedes 0. Ties may be broken arbitrarily. For adjacent labels i, j in this increasing order impose

$$0 \leq g_j - g_i \leq x_j - x_i. \quad (9)$$

These $2(N+1)$ linear inequalities define a polytope P_π in $g_0, \dots, g_N$. They ensure the required spatial order as well as a nondecreasing 1-Lipschitz interpolating derivative. Conversely every scalar trajectory satisfies them for a suitable order.

Each nonempty P_π is compact. The interpolation inequalities with $(x_*, g_*) = (0, 0)$ imply $g_i = c_i x_i$ with $0 \leq c_i \leq 1$, so

$$|x_{i+1}| \leq \max\{1, |1 - h_{i+1}|\}|x_i|, \quad |g_i| \leq |x_i|.$$

Starting at 1 gives finite bounds on all variables. Closedness follows from (9).

3.2 The objective on an order

On an interval $[a, b]$ of length ℓ , prescribe derivative values $u \leq v$ with $\Delta = v - u \leq \ell$. Monotonicity and Lipschitz continuity give

$$u\ell + \frac{\Delta^2}{2} \leq \int_a^b f'(x) dx \leq v\ell - \frac{\Delta^2}{2}. \quad (10)$$

For the upper bound the derivative rises with slope one immediately and then stays constant; for the lower bound it stays constant and rises at the end. These choices attain the two bounds.

Follow the adjacent intervals from $*$ to the terminal label N , to the right if $x_N \geq 0$ and to the left otherwise. In each pair (i, j) , i is the starting endpoint and j the ending endpoint in this traversal; thus $x_j - x_i \leq 0$ on a leftward interval. The largest possible terminal gap for the prescribed gradient data is

$$Q_\pi(g) = \sum_{(i,j) \text{ on } * \rightarrow N} \left[g_j(x_j - x_i) - \frac{(g_j - g_i)^2}{2} \right]. \quad (11)$$

For a leftward interval this uses the negative of the lower integral in (10). Thus the same expression covers both orientations. The derivative values agree at shared endpoints, so these interval choices can be combined. Choose the maximizing derivative on the intervals between zero and x_N , either extremal derivative from (10) on the other intervals, and a constant derivative outside the extreme sample points. Integrating with $f(0) = 0$ gives a convex 1-smooth function minimized at zero. At coincident points, (9) forces the prescribed derivatives to coincide.

Theorem 1 (Finite order representation). *For any fixed real schedule,*

$$W_{N,1}(h) = \max_{\pi} \max_{g \in P_\pi} Q_\pi(g). \quad (12)$$

A maximizing function can be chosen piecewise quadratic, with derivative slopes only zero and one and at most $N + 1$ intervals of positive slope.

Proof. Every instance is bounded by its order objective through (10). Conversely the construction just given realizes each feasible order objective, with the prescribed gradient descent trajectory. Compactness and the finite number of orders give attainment. $\square$

3.3 Enumeration by stationarity equations

Write an order objective as $c + b^\top g + \frac{1}{2}g^\top Hg$. Among its maximizers choose one in the relative interior of a face of smallest dimension. On the tangent space of this face, stationarity holds and the Hessian is negative semidefinite. If the restricted Hessian has a nonzero null direction, both the linear and quadratic terms in the variation along that direction vanish. The objective is therefore constant on the corresponding line. Compactness lets us follow this line to the boundary

of the face and obtain a maximizer on a smaller face, a contradiction. Therefore the restricted Hessian is negative definite, unless the face is a vertex.

Choose independent active constraint rows A defining that face's affine hull, with $Ag = a$. The corresponding system

$$\begin{pmatrix} H & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} g \\ \mu \end{pmatrix} = \begin{pmatrix} -b \\ a \end{pmatrix} \quad (13)$$

is nonsingular: its homogeneous equations imply $Av = 0$ and $v^\top H v = 0$, whence $v = 0$ and then $\mu = 0$. Enumerate every subset of at most $N + 1$ active constraints, solve the nonsingular systems, discard infeasible solutions, and maximize their objective values. Multiplier signs need not be tested: every retained point is feasible, and at least one global maximizer is retained.

Corollary 2 (Rational fixed-schedule values). *If h is rational, $W_{N,1}(h)$ and a maximizing gradient vector are rational. A worst-case function has rational breakpoints and coefficients.*

There are $(N + 2)!/2$ orders and at most

$$\frac{(N + 2)!}{2} \sum_{j=0}^{N+1} \binom{2(N + 1)}{j}$$

systems in this enumeration. At $N = 3$ there are 9,780 systems.

3.4 Two exact three-step evaluations

Proposition 3 (Exact three-step values). *The following schedules have the stated scalar worst-case values:*

<i>Schedule</i>	<i>Exact scalar worst case</i>
(9/5, 249/100, 123/100)	25/602
(18353756, 25119018, 12271572)/10 ⁷	1250000/30372173

In particular,

$$W_1(a) = U := \frac{25}{602}, \quad \inf_{h \in [0, \infty)^3} W_1(h) \leq \frac{1250000}{30372173}. \quad (14)$$

Proof. For each schedule, the enumeration in `supplement/scalar-schedules/verify.py` solves the 9,780 rational systems described above and gives the displayed maximum. In both cases, a maximizing gradient vector is constant at all four iterates. Its value is 25/301 for the first schedule and 2500000/30372173 for the second.

These values are attained by the Huber function

$$f_\delta(x) = \begin{cases} x^2/2, & |x| \leq \delta, \\ \delta|x| - \delta^2/2, & |x| \geq \delta, \end{cases} \quad \delta = \frac{1}{1 + 2 \sum_{i=1}^3 h_i}.$$

Starting from $x_0 = 1$, all three updates remain on the positive affine branch. The terminal value is $\delta/2$, which gives the stated lower bounds and agrees with the enumeration. $\square$

4 Stability of the scalar bound

We transfer the scalar bound to nearby schedules by comparing terminal iterates under schedules that change one step at a time.

For a scalar convex function with a 1-Lipschitz derivative, every secant slope of f' belongs to $[0, 1]$. Hence the update

$$T_t(x) = x - tf'(x)$$

has universal Lipschitz factor

$$\text{Lip}(T_t) \leq \ell(t) := \max\{1, |1 - t|\}. \quad (15)$$

The same secant argument applied between x and $T_t(x)$ gives

$$|f'(T_t(x))| \leq \ell(t)|f'(x)|. \quad (16)$$

The function ℓ is 1-Lipschitz, and at the center schedule

$$\ell(a_1) = 1, \quad \ell(a_2) = \frac{149}{100}, \quad \ell(a_3) = 1.$$

Lemma 4 (Scalar schedule transfer). *Let h be a three-step schedule, write $d_i = |h_i - a_i|$, and suppose $d_3 \leq 77/100$. Define*

$$\Delta(h) = d_1 \left(\frac{149}{100} + d_2 \right) + d_2 + \frac{149}{100} d_3. \quad (17)$$

Then

$$W_1(h) \leq \frac{1}{2} \left(\sqrt{\frac{25}{301}} + \Delta(h) \right)^2. \quad (18)$$

Proof. Fix an admissible scalar function and starting point, and introduce the hybrid schedules

$$a^{(0)} = a, \quad a^{(1)} = (a_1, a_2, h_3), \quad a^{(2)} = (a_1, h_2, h_3), \quad a^{(3)} = h.$$

Compare the four terminal points consecutively. Using (15) and (16), the change from $a^{(0)}$ to $a^{(1)}$ is at most

$$d_3 \ell(a_1) \ell(a_2) |f'(x_0)| \leq \frac{149}{100} d_3.$$

The change from $a^{(1)}$ to $a^{(2)}$ is propagated through the common final update with factor $\ell(h_3) = 1$, because $h_3 \in [0, 2]$, and contributes at most d_2 . The change from $a^{(2)}$ to $a^{(3)}$ is propagated through the common last two updates and contributes at most

$$d_1 \ell(h_2) \ell(h_3) \leq d_1 \left(\frac{149}{100} + d_2 \right).$$

Here $|f'(x_0)| \leq 1$, since $f'(x_\star) = 0$ and f' is 1-Lipschitz. The triangle inequality therefore gives

$$|x_3^h - x_3^a| \leq \Delta(h).$$

For a convex 1-smooth function, the smoothness inequality shows that a unit gradient step decreases the function value by at least half the squared gradient norm. The new value is at least the minimum. Applying this at the center endpoint and using (14) gives

$$|f'(x_3^a)| \leq \sqrt{2(f(x_3^a) - f_\star)} \leq \sqrt{2U} = \sqrt{\frac{25}{301}}.$$

The smoothness inequality from x_3^a to x_3^h therefore gives

$$f(x_3^h) - f_\star \leq U + \sqrt{2U} \Delta(h) + \frac{1}{2} \Delta(h)^2 = \frac{1}{2} (\sqrt{2U} + \Delta(h))^2.$$

This holds for every admissible scalar instance. Taking the supremum gives (18). $\square$

If $\|h - a\|_\infty \leq r$, then

$$\Delta(h) \leq r \left(\frac{149}{100} + r \right) + r + \frac{149}{100} r = \frac{199}{50} r + r^2. \quad (19)$$

Both theorems apply Lemma 4 through this cube form, at the radii $1/59$ and $3/196$.

5 A dimension gap on the cube of radius $1/59$

For the geometry (6), fix the sampled values

$$\begin{aligned} F_0 &= \frac{499491}{1000000}, & F_1 &= \frac{13321727}{73750000}, \\ F_2 &= \frac{220981977}{1475000000}, & F_3 &= \frac{47267401}{737500000}, \end{aligned} \quad (20)$$

together with $F_\star = 0$. We use these values for every schedule. The proof verifies that all twenty directed interpolation slacks remain nonnegative throughout the cube and that the constraint for $(\star, 3)$ is tight at its boundary.

Theorem 5 (A dimension gap and its maximal radius). *Set*

$$C = F_3 = \frac{47267401}{737500000} \approx 0.0640914, \quad S = \frac{15507969961}{242347220000} \approx 0.0639907.$$

(i) *For every schedule h with $\|h - a\|_\infty \leq 1/59$, the geometry (6), the locations (7), and the fixed values (20) are interpolable by a convex 1-smooth function on $\mathbf{R}^2$, and*

$$W_1(h) < S < C \leq W_2(h), \quad C - S > 10^{-4}. \quad (21)$$

(ii) *The radius $1/59$ is the largest centered radius for this geometry and the terminal target C . More precisely, if $r > 1/59$, then some schedule with $\|h - a\|_\infty \leq r$ admits no interpolating values satisfying $F_\star = 0$ and $F_3 \geq C$, even when all nonterminal values may depend on h .*

Proof. For each ordered pair of distinct nodes, the slack of the difference constraint (5) under the fixed values (20) is affine in h , because the gradients are fixed and the locations (7) are affine in h . Consequently its minimum over the cube occurs at one of the eight vertices. Exact substitution at those vertices gives a nonnegative minimum for every ordered pair: all $20 \times 8 = 160$ directed slacks are nonnegative.¹ The four minima that vanish correspond to the pairs

$$\star \rightarrow 0, \quad \star \rightarrow 1, \quad \star \rightarrow 3, \quad 3 \rightarrow 2;$$

¹The function `verify` in `supplement/robust-interpolation/verify_robust_potential_gap.py` checks these interpolation slacks and the center data used below.

among the remaining minima, the smallest slack is $567/500000 > 0$. Thus all interpolation inequalities hold throughout the closed cube, and the interpolation theorem supplies a convex 1-smooth function f_h for each schedule. By the margins (8) and the stationary sample, each pair (f_h, x_0) is an admissible planar instance whose iterates under h are the locations (7), and

$$f_h(x_3(h)) - f_h(0) = C.$$

Hence $W_2(h) \geq C$ on the closed cube.

For the scalar bound, note $1/59 < 77/100$, so Lemma 4 applies throughout the cube, and the cube bound (19) gives

$$\Delta(h) \leq \frac{199}{50} \cdot \frac{1}{59} + \frac{1}{59^2} = \frac{11791}{174050}.$$

The radical has the exact rational cap

$$\sqrt{\frac{25}{301}} < \frac{29}{100}, \quad \left(\frac{29}{100}\right)^2 - \frac{25}{301} = \frac{3141}{3010000} > 0,$$

so (18) yields

$$W_1(h) < \frac{1}{2} \left(\frac{29}{100} + \frac{11791}{174050} \right)^2 = \frac{15507969961}{242347220000} = S.$$

Exact subtraction gives

$$C - S = \frac{7625162177}{75733506250000} > \frac{1}{10000}.$$

This proves the chain (21).

For the ordered pair $(\star, 3)$, the weight in (5) satisfies

$$\begin{aligned} w_{\star 3}(a) - C &= \frac{32741}{14750000}, \\ w_{\star 3}(h) - w_{\star 3}(a) &= \langle \beta_{\star 3}, h - a \rangle, \quad \beta_{\star 3} = \left(-\frac{9759}{250000}, \frac{14797}{250000}, -\frac{1637}{50000} \right), \\ \|\beta_{\star 3}\|_1 &= \frac{32741}{250000}. \end{aligned} \tag{22}$$

The ratio of the center margin to the sensitivity norm in (22) is $1/59$. For the sign choice $h - a = (r, -r, r)$, one has

$$w_{\star 3}(h) = w_{\star 3}(a) - r \|\beta_{\star 3}\|_1.$$

At $r = 1/59$ the weight equals C ; for every larger r it is strictly less than C . The interpolation constraint $F_3 - F_\star \leq w_{\star 3}(h)$ then rules out $F_3 \geq C$, regardless of the remaining function values. $\square$

6 A fixed planar function

We define a single function from interpolation data at the center schedule and bound its terminal value throughout the cube of radius $3/196$. Let $y_i = x_i(a)$ be the center locations from (7)—explicitly,

$$\begin{aligned} y_0 &= \left(-\frac{199}{200}, \frac{19}{200} \right), & y_1 &= \left(\frac{1981}{2500}, -\frac{77}{1250} \right), \\ y_2 &= \left(-\frac{15409}{50000}, -\frac{2558}{3125} \right), & y_3 &= \left(-\frac{649}{50000}, -\frac{30473}{50000} \right), \end{aligned} \tag{23}$$

with $y_\star = 0$, and assign the values

$$P_\star = 0, \quad \begin{aligned} P_0 &= \frac{499491}{1000000}, & P_1 &= \frac{469061}{2500000}, \\ P_2 &= \frac{3837603}{25000000}, & P_3 &= L = \frac{828889}{12500000}. \end{aligned} \quad (24)$$

We use the conjugate construction of Taylor, Hendrickx, and Glineur [1, Remark 2]. With $\mathcal{I} = \{\star, 0, 1, 2, 3\}$, define

$$\Phi(p) = \max_{i \in \mathcal{I}} \left\{ \langle y_i, p \rangle - P_i + \frac{1}{2} \|p - g_i\|^2 \right\}, \quad f^\circ = \Phi^*. \quad (25)$$

Here Φ^* is the Fenchel conjugate,

$$\Phi^*(x) = \sup_{p \in \mathbf{R}^2} \{ \langle x, p \rangle - \Phi(p) \}.$$

Proposition 6 (An explicit interpolating function). *The function f° in (25) is convex and has a 1-Lipschitz gradient. It satisfies*

$$\nabla f^\circ(y_i) = g_i, \quad f^\circ(y_i) = P_i \quad (i \in \mathcal{I}). \quad (26)$$

In particular, 0 minimizes f° and $f^\circ(0) = 0$.

Proof. Every branch of Φ has quadratic part $\|p\|^2/2$, so Φ is 1-strongly convex. Its conjugate is therefore differentiable, convex, and 1-smooth.

For distinct nodes $i, j \in \mathcal{I}$, put

$$\sigma_{ij} = w_{ij}(a) - (P_j - P_i), \quad (27)$$

with the weights w_{ij} of (5) taken at the center locations. Exact rational substitution gives $\sigma_{ij} \geq 0$ for all twenty ordered pairs. The tight pairs are

$$\star \rightarrow 0, \quad \star \rightarrow 1, \quad \star \rightarrow 3, \quad 3 \rightarrow 2, \quad (28)$$

and the smallest positive slack is $567/500000$.

At $p = g_i$, the value of branch i of Φ minus the value of branch j is σ_{ji} . Hence branch i is active at g_i , and its gradient there is y_i . It follows that $y_i \in \partial\Phi(g_i)$. Fenchel conjugacy gives $g_i \in \partial f^\circ(y_i)$, and the differentiability of f° gives the gradient identity in (26). Fenchel equality then yields

$$f^\circ(y_i) = \langle y_i, g_i \rangle - \Phi(g_i) = P_i.$$

The stationary node gives $\nabla f^\circ(0) = 0$ and $f^\circ(0) = 0$, completing the proof. $\square$

By (8), the pair (f°, x_0) is admissible and its sampled gradients span two dimensions. For any schedule h , let its iterates under the fixed function be

$$z_0^h = x_0, \quad z_{k+1}^h = z_k^h - h_{k+1} \nabla f^\circ(z_k^h), \quad k = 0, 1, 2. \quad (29)$$

Proposition 6 gives $z_i^a = y_i$ for $0 \leq i \leq 3$; in particular, the fixed function attains the terminal value L at the center. We compare this trajectory with the trajectories of the same function under nearby schedules.

Theorem 7 (A fixed function on a schedule cube). *Let f° be the single function defined in (25), and start it from the fixed point x_0 in (6). For every schedule satisfying*

$$\|h - a\|_\infty \leq \frac{3}{196}, \quad (30)$$

the trajectory (29) obeys

$$W_1(h) < \frac{1231939801}{20000000000} < \frac{770017}{12500000} < f^\circ(z_3^h) - f^\circ(0) \leq W_2(h). \quad (31)$$

The scalar upper bound and planar lower bound in this chain are approximately 0.0615970 and 0.0616014, respectively.

The proof needs a dimension-free version of the update factor (15).

Lemma 8 (Dimension-free update bound). *If f is convex with a 1-Lipschitz gradient, then*

$$\text{Lip}(I - t\nabla f) \leq \ell(t) = \max\{1, |1 - t|\} \quad (32)$$

in every Euclidean dimension.

Proof. Set $u = x - y$ and $v = \nabla f(x) - \nabla f(y)$. Adding the two interpolation inequalities (4) for x, y gives the cocoercivity bound $\langle u, v \rangle \geq \|v\|^2$. Lipschitz continuity gives $\|v\| \leq \|u\|$. For $t \geq 0$, expanding $\|u - tv\|^2$ and using cocoercivity bounds the change from $\|u\|^2$ by $(t^2 - 2t)\|v\|^2$, which is nonpositive when $0 \leq t \leq 2$. When $t > 2$, the same expansion is at most

$$\|u\|^2 + (t^2 - 2t)\|v\|^2 \leq (t - 1)^2 \|u\|^2.$$

For $t < 0$, the triangle inequality gives $\|u - tv\| \leq (1 - t)\|u\|$. This proves (32). $\square$

Lemma 9 (The fixed planar trajectory). *If $r = \|h - a\|_\infty \leq 3/196$, then*

$$\|z_3^h - y_3\| \leq r \left[\left(\frac{149}{100} + r \right) \|g_0\| + \|g_1\| + \|g_2\| \right] < \frac{223}{6250}. \quad (33)$$

Consequently,

$$f^\circ(z_3^h) - f^\circ(0) > \frac{770017}{12500000}. \quad (34)$$

Proof. Compare the terminal points under the four schedules

$$a, \quad (a_1, a_2, h_3), \quad (a_1, h_2, h_3), \quad h.$$

Changing the third step contributes at most $r\|g_2\|$. Changing the second step contributes at most $r\|g_1\|$, because the common third update has Lipschitz factor one. Changing the first step contributes at most $r(149/100 + r)\|g_0\|$, because it passes through the common second and third updates. Indeed, throughout the cube, $h_2 > 2$ and $0 < h_3 < 2$, so

$$h_2 \geq \frac{249}{100} - \frac{3}{196} = \frac{6063}{2450} > 2, \quad 0 < \frac{1488}{1225} \leq h_3 \leq \frac{3051}{2450} < 2,$$

and therefore

$$\ell(h_2) \leq \frac{149}{100} + r, \quad \ell(h_3) = 1.$$

Lemma 8 and the center interpolation identities (26) now give the first inequality in (33).

The rational norm bounds

$$\|g_0\| < \frac{99681}{100000}, \quad \|g_1\| < \frac{26823}{50000}, \quad \|g_2\| < \frac{29411}{100000}, \quad \|g_3\| < \frac{33}{250} \quad (35)$$

follow by squaring. At $r = 3/196$, they give the exact comparison

$$\frac{3}{196} \left[\left(\frac{149}{100} + \frac{3}{196} \right) \frac{99681}{100000} + \frac{26823}{50000} + \frac{29411}{100000} \right] = \frac{856669767}{2401000000} < \frac{223}{6250}. \quad (36)$$

The bound for smaller r follows because the corresponding expression as a function of r is increasing for $r \geq 0$.

Convexity at the center endpoint, together with $\nabla f^\circ(y_3) = g_3$, yields

$$f^\circ(z_3^h) \geq P_3 + \langle g_3, z_3^h - y_3 \rangle \geq P_3 - \|g_3\| \|z_3^h - y_3\|.$$

Using (35) and (33) gives

$$f^\circ(z_3^h) - f^\circ(0) > \frac{828889}{12500000} - \frac{33}{250} \cdot \frac{223}{6250} = \frac{770017}{12500000}, \quad (37)$$

which proves (34). $\square$

Proof of Theorem 7. For the scalar side, apply Lemma 4 through the cube bound (19). At $r = 3/196$, exact arithmetic gives

$$\frac{199}{50} \cdot \frac{3}{196} + \left(\frac{3}{196} \right)^2 = \frac{58731}{960400} < \frac{1529}{25000}, \quad (38)$$

and the radical has the rational upper bound

$$\sqrt{\frac{25}{301}} < \frac{28983}{100000}. \quad (39)$$

Since $3/196 < 77/100$, the transfer lemma applies and

$$W_1(h) < \frac{1}{2} \left(\frac{28983}{100000} + \frac{1529}{25000} \right)^2 = \frac{1231939801}{20000000000}. \quad (40)$$

The fixed planar lower bound is (34), and the two rational thresholds are separated by

$$\frac{770017}{12500000} - \frac{1231939801}{20000000000} = \frac{87399}{20000000000} > 0. \quad (41)$$

Finally, Proposition 6 and the margins (8) make (f°, x_0) admissible in the definition (2), so its terminal value is at most $W_2(h)$. This proves the complete chain (31). $\square$

Supplementary verification. The scalar supplement enumerates the stationary-point systems in Proposition 3 and reconstructs maximizing functions. The robust-interpolation supplement checks the interpolation inequalities and rational perturbation bounds in Theorems 5 and 7. Both accompany the public collection.

References

- [1] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Smooth strongly convex interpolation and exact worst-case performance of first-order methods*, Mathematical Programming 161 (2017), 307–345. doi:10.1007/s10107-016-1009-3.
- [2] A. Daccache, *Performance estimation of the gradient method with fixed arbitrary step sizes*, Master’s thesis, Université catholique de Louvain, 2019. institutional record.
- [3] D. Eloi, *Worst-case functions for the gradient method with fixed variable step sizes*, Master’s thesis, Université catholique de Louvain, 2022. institutional record.
- [4] A. Luner and B. Grimmer, *On averaging and extrapolation for gradient descent*, Mathematics of Operations Research, 2025. doi:10.1287/moor.2024.0449.
- [5] S. Das Gupta, B. P. G. Van Parys, and E. K. Ryu, *Branch-and-bound performance estimation programming: a unified methodology for constructing optimal optimization methods*, Mathematical Programming 204 (2024), 567–639. doi:10.1007/s10107-023-01973-1.
- [6] B. Grimmer, K. Shu, and A. L. Wang, *Composing optimized stepsize schedules for gradient descent*, Mathematics of Operations Research, 2025. doi:10.1287/moor.2024.0764.