Dimension gaps for positive gradient-descent schedules from horizon three onward

Logan Bell

September 2026

Abstract

The worst-case terminal objective value of gradient descent under a prescribed step schedule can depend on the ambient dimension. For every $M \geq 3$, a full-dimensional cube of strictly positive M-step schedules satisfies $W_1^{(M)}(h) < W_2^{(M)}(h)$. The same convex 1-smooth function on $\mathbf{R}^2$ has a terminal value above the scalar worst case throughout all these cubes. We also give a single infinite positive schedule for which this function gives a dimension gap at every finite prefix of at least three steps. The proof combines an exact three-step gap with a trajectory stability estimate that is uniform in dimension.

1 Prescribed schedules

Let $\mathcal{F}_d$ be the class of differentiable convex functions $f : \mathbf{R}^d \to \mathbf{R}$ whose gradients are 1-Lipschitz and whose sets of minimizers are nonempty. Given a schedule $h = (h_1, \ldots, h_M) \in \mathbf{R}^M$, gradient descent takes the steps

$$x_k = x_{k-1} - h_k \nabla f(x_{k-1}), \qquad 1 \leq k \leq M. \tag{1}$$

We measure the normalized terminal objective gap by

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

Embedding an instance in one additional coordinate gives $W_d^{(M)}(h) \leq W_{d+1}^{(M)}(h)$. We construct schedules with $h_k > 0$ for every k and $W_1^{(M)}(h) < W_2^{(M)}(h)$.

Performance estimation represents sampled geometry by a positive semidefinite Gram matrix, whose rank controls the realization dimension [2]. Computational work has optimized variable-step schedules in the unrestricted formulation [3], and exact worst-case values are known for particular three-step schedules [4]. Luner and Grimmer report a numerical dimension effect for an extrapolated output after seven unit steps [5]. The results below concern the ordinary terminal iterate, positive prescribed schedules, and the extension of one certified three-step gap to every later horizon.

2 The three-step construction

The companion paper *Three-step dimension gaps for gradient descent* [1] proves the three-step gap used below by exact scalar enumeration and an explicit planar interpolating function. We

recall the function and estimates needed here. The rational center schedule is

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

The companion paper constructs an explicit convex function $f^\circ : \mathbf{R}^2 \to \mathbf{R}$ with a 1-Lipschitz gradient as the Fenchel conjugate of a maximum of five quadratic functions. The construction uses smooth-convex interpolation [2] of exact rational data at the schedule a. It also fixes an initial point $x_0 \in \mathbf{R}^2$ satisfying $\|x_0\| < 1$. The function is minimized at the origin with $f^\circ(0) = 0$. We use this function and initial point for every horizon and schedule below. For a schedule h of any length, write

$$z_0^h = x_0, \qquad z_k^h = z_{k-1}^h - h_k \nabla f^\circ(z_{k-1}^h), \qquad J_M^\circ(h) = f^\circ(z_M^h) - f^\circ(0), \tag{4}$$

and write $y_k = z_k^a$ for the center iterates.

We use the exact constants

$$U = \frac{25}{602}, \qquad L = \frac{828889}{12500000}, \qquad G = L - U = \frac{93245589}{3762500000}, \qquad q = \frac{149}{100}, \tag{5}$$

and the three-step gap in the next proposition. In the companion's three-step notation, $W_d(a) = W_d^{(3)}(a)$.

Proposition 1 (The three-step gap)**.** *At the schedule* (3)*,*

$$W_1^{(3)}(a) \le U < L = J_3^\circ(a) \le W_2^{(3)}(a). \tag{6}$$

Proof. The scalar identity $W_1^{(3)}(a) = U$ is proved by exact stationary-point enumeration in the companion paper. The equality $J_3^\circ(a) = L$ follows from the center interpolation identities in its explicit conjugate construction: the gradients of f° at y_0, y_1, y_2 drive the recurrence (1) through explicit rational points, and $f^\circ(y_3) = L$. The pair (f°, x_0) is an admissible planar instance, so $J_3^\circ(a) \le W_2^{(3)}(a)$. Finally, the middle inequality is $G = L - U > 0$ from (5). □

The constant q enters through the schedule alone: the first and third center steps lie in $[0, 2]$, while $|1 - a_2| = q$. Section 3 gives the corresponding Lipschitz bounds.

3 Uniform stability with respect to the schedule

For $t \in \mathbf{R}$, define

$$\ell(t) = \max\{1, |1 - t|\}. \tag{7}$$

At the center schedule, $\ell(a_1) = \ell(a_3) = 1$ and $\ell(a_2) = q$.

Lemma 2 (Update stability)**.** *If $f \in \mathcal{F}_d$, then the factor in* (7) *satisfies*

$$\mathrm{Lip}(I - t\nabla f) \le \ell(t) \qquad (t \in \mathbf{R}). \tag{8}$$

If $x_\star$ minimizes f, then every trajectory of (1) *satisfies*

$$\|x_k - x_\star\| \le \|x_0 - x_\star\| \prod_{i=1}^{k} \ell(h_i). \tag{9}$$

Proof. For $u = x - y$ and $v = \nabla f(x) - \nabla f(y)$, adding the two smooth-convex interpolation inequalities for x, y gives $\langle u, v\rangle \geq \|v\|^2$ [2, Corollary 1]. Lipschitz continuity gives $\|v\| \leq \|u\|$. For $t \geq 0$, expansion gives

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

For $0 \leq t \leq 2$, the second term is nonpositive. For $t > 2$, using $\|v\| \leq \|u\|$ bounds the right-hand side by $(t-1)^2\|u\|^2$. For $t < 0$, the triangle inequality gives $\|u - tv\| \leq (1-t)\|u\|$. Thus in all three cases $\|u - tv\| \leq \ell(t)\|u\|$, which is (8). Apply that inequality successively with $y = x_\star$, noting that $\nabla f(x_\star) = 0$, to obtain (9). □

Fix $M \geq 3$, and pad the seed schedule with zeros:

$$\bar{h}^{(M)} = (a_1, a_2, a_3, 0, \ldots, 0) \in \mathbf{R}^M. \tag{10}$$

For a fixed admissible instance, write x_k^h for the iterate under schedule h, and abbreviate $\bar{h}^{(M)}$ to $\bar{h}$ in superscripts.

Lemma 3 (Performance stability)**.** *Let $0 < \rho < 1$ and $\left\|h - \bar{h}^{(M)}\right\|_\infty \leq \rho$. For every normalized admissible instance in every dimension,*

$$\left|f(x_M^h) - f(x_M^{\bar{h}})\right| \leq \varepsilon_M(\rho), \qquad \varepsilon_M(\rho) = M\rho(q+\rho)^2(1+\rho)^{2M-3}. \tag{11}$$

Consequently,

$$\left|W_d^{(M)}(h) - W_d^{(M)}(\bar{h}^{(M)})\right| \leq \varepsilon_M(\rho) \qquad (d \geq 1). \tag{12}$$

Proof. Compare the trajectories of one admissible instance under h and $\bar{h}^{(M)}$, and set $e_k = \left\|x_k^h - x_k^{\bar{h}}\right\|$. Adding and subtracting the update with step h_k at the base trajectory gives

$$e_k \leq \ell(h_k)e_{k-1} + \rho\left\|\nabla f(x_{k-1}^{\bar{h}})\right\|. \tag{13}$$

Since $\|\nabla f(x)\| \leq \|x - x_\star\|$, Lemma 2 bounds the base gradient in the kth term by $\prod_{i<k}\ell(\bar{h}_i)$. Unrolling (13) therefore gives

$$e_M \leq \rho\sum_{k=1}^{M}\left(\prod_{i<k}\ell(\bar{h}_i)\right)\left(\prod_{j>k}\ell(h_j)\right). \tag{14}$$

At the base schedule, the update factors are $1, q, 1, \ldots, 1$. The function ℓ is 1-Lipschitz, so every summand in (14) contains at most one factor bounded by $q + \rho$, while each other factor is at most $1 + \rho$. Hence

$$e_M \leq M\rho(q+\rho)(1+\rho)^{M-2}. \tag{15}$$

Both terminal points lie within $R = (q+\rho)(1+\rho)^{M-1}$ of a minimizer. Their joining segment lies in the same ball, where $\|\nabla f\| \leq R$. Integrating the gradient along that segment and using (15) proves (11).

The estimate holds for every admissible instance before taking a supremum. Applying it in each direction and taking suprema proves (12), whether or not either supremum is attained. □

4 Positive schedule cubes

For each $M \geq 3$, set

$$\rho_M = \frac{G}{20M}, \qquad c_k^{(M)} = \begin{cases} a_k, & 1 \leq k \leq 3, \\ \rho_M/2, & 4 \leq k \leq M. \end{cases} \tag{16}$$

The second line in (16) is empty when $M = 3$.

Theorem 4 (Positive schedule cubes)**.** *For every $M \geq 3$ and every schedule in the closed cube*

$$\left\| h - c^{(M)} \right\|_\infty \leq \frac{\rho_M}{4}, \tag{17}$$

all step sizes are strictly positive, and the fixed function $f^\circ : \mathbf{R}^2 \to \mathbf{R}$ of Section 2 satisfies

$$W_1^{(M)}(h) < U + \frac{G}{8} = \frac{191892227}{4300000000} < \frac{1902719123}{30100000000} = L - \frac{G}{8} < J_M^\circ(h) \leq W_2^{(M)}(h). \tag{18}$$

In particular, the interior of (17) *is a nonempty open set of positive schedules. Every schedule in this set satisfies $W_1^{(M)}(h) < W_2^{(M)}(h)$; the difference exceeds $3G/4 = 279736767/15050000000$.*

Proof. We first work on the larger cube $\left\| h - \bar{h}^{(M)} \right\|_\infty \leq \rho_M$. Since $\rho_M < 1/100$, we have $q + \rho_M < 3/2$. Moreover, $(2M-3)\rho_M < G/10 < 1/10$. The binomial estimate $\binom{n}{k} \leq n^k$ therefore gives

$$(1 + \rho_M)^{2M-3} \leq \frac{1}{1 - (2M-3)\rho_M} < \frac{10}{9}. \tag{19}$$

Using $M\rho_M = G/20$ in (11), we obtain

$$\varepsilon_M(\rho_M) < \frac{G}{8}. \tag{20}$$

The zero tail in $\bar{h}^{(M)}$, defined in (10), leaves every instance fixed after its third iterate. Proposition 1 consequently gives

$$W_1^{(M)}(\bar{h}^{(M)}) = W_1^{(3)}(a) \leq U, \qquad J_M^\circ(\bar{h}^{(M)}) = L.$$

Apply Lemma 3 in dimension one and apply its individual-instance form to f°. The error cap (20) yields the two outside strict inequalities in (18). The middle inequality has margin

$$\left(L - \frac{G}{8}\right) - \left(U + \frac{G}{8}\right) = \frac{3G}{4} = \frac{279736767}{15050000000} > 0.$$

Finally, f° is an admissible two-dimensional instance, so $J_M^\circ(h) \leq W_2^{(M)}(h)$.

For a schedule satisfying (17), each tail coordinate lies in $[\rho_M/4, 3\rho_M/4]$, and each of the first three coordinates lies within $\rho_M/4$ of the corresponding coordinate of a. Thus the whole cube lies strictly inside the padded radius-ρ_M cube used above. All tail steps are positive, and the first three remain positive because $\rho_M < 1/100$. This proves the theorem. □

5 An infinite positive schedule

We now keep the first three steps fixed and append a summable positive tail to obtain one infinite schedule. First let nonnegative steps $t_4, \dots, t_M$ follow the three seed steps, and write

$$S_M = \sum_{k=4}^{M} t_k. \tag{21}$$

Lemma 5 (A small-tail gap)**.** *Suppose* $0 \le t_k < 2$ *and the sum in* (21) *obeys*

$$S_M \le S_0, \qquad S_0 = \frac{G}{4q^2} = \frac{93245589}{33412505000}. \tag{22}$$

Then the schedule $h = (a_1, a_2, a_3, t_4, \dots, t_M)$ *satisfies*

$$W_1^{(M)}(h) \le U < L - \frac{G}{4} \le J_M^\circ(h) \le W_2^{(M)}(h). \tag{23}$$

Proof. For $0 \le t < 2$, the smoothness inequality gives

$$f(x - t\nabla f(x)) \le f(x) - t\left(1 - \frac{t}{2}\right)\|\nabla f(x)\|^2 \le f(x).$$

Every scalar objective is therefore nonincreasing along the tail, so $W_1^{(M)}(h) \le W_1^{(3)}(a) \le U$.

For the fixed function f°, the first three update factors relative to its minimizer are $1, q, 1$. Lemma 2 and $\|x_0\| < 1$ give $\|y_3\| \le q$. Each tail update is nonexpansive, hence all later iterates also remain in the radius-q ball. The total displacement along the tail is bounded by

$$\left\|z_M^h - y_3\right\| \le \sum_{k=4}^{M} t_k \left\|\nabla f^\circ(z_{k-1}^h)\right\| \le qS_M. \tag{24}$$

The segment from y_3 to z_M^h stays in the same ball, where the gradient norm is at most q. Integrating along this segment and using (24) gives

$$J_M^\circ(h) \ge L - q^2 S_M \ge L - \frac{G}{4}. \tag{25}$$

Since $L - G/4 = U + 3G/4 > U$, this proves the chain (23). □

Define the infinite schedule h^∞ by

$$h_1^\infty = \frac{9}{5}, \qquad h_2^\infty = \frac{249}{100}, \qquad h_3^\infty = \frac{123}{100}, \qquad h_{3+n}^\infty = \frac{S_0}{2^n} \quad (n \ge 1). \tag{26}$$

Theorem 6 (A gap at every finite prefix)**.** *Every step of the schedule* (26) *is positive. For each finite prefix* $h_{1:M}^\infty$ *with* $M \ge 3$*, the same function* f° *and the same initial point* x_0 *satisfy*

$$W_1^{(M)}(h_{1:M}^\infty) \le U < L - \frac{G}{4} < J_M^\circ(h_{1:M}^\infty) \le W_2^{(M)}(h_{1:M}^\infty). \tag{27}$$

The difference $W_2^{(M)}(h_{1:M}^\infty) - W_1^{(M)}(h_{1:M}^\infty)$ *exceeds* $3G/4$ *at each prefix.*

Proof. The infinite tail has total mass S_0, while every finite tail has mass strictly less than S_0. Its steps are positive and smaller than two. Lemma 5 applies, and its lower bound in (25) is strict for every such finite prefix. This proves (27). □

Supplementary verification. The scalar stationary-point enumeration and planar interpolation calculations are checked in `supplement/scalar-schedules/verify.py` and `supplement/robust-interpolation/verify_robust_potential_gap.py`. The extension checker `supplement/all-horizon-dimension-gaps/verify_all_horizon_gap.py` verifies the rational identities and inequalities in the horizon and tail bounds. The interpolation construction and stability arguments are proved above and in the companion paper. All paths are relative to the public collection.

References

[1] L. Bell, *Three-step dimension gaps for gradient descent*, Companion paper (2026). Public collection.

[2] 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.

[3] 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.

[4] 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.

[5] A. Luner and B. Grimmer, *On averaging and extrapolation for gradient descent*, Mathematics of Operations Research, 2025. doi:10.1287/moor.2024.0449.