Anytime Last Iterates in Fixed Dimension

Pahan Dewasurendra
Johns Hopkins University

Abstract

The standard horizon-free subgradient schedules, $\eta_t \asymp t^{-1/2}$ for convex objectives and $\eta_t \asymp t^{-1}$ under strong convexity, have last-iterate guarantees worse than the optimal rates by a logarithmic factor. Known matching examples use dimension growing with the horizon. We prove that this loss disappears in every fixed dimension for deterministic projected subgradient descent. The key is a pathwise conversion theorem for arbitrary positive and adaptive steps. On any suffix, the terminal objective can exceed the best suffix value by at most twice the suffix trajectory rank times the Lipschitz constant times the largest realized step length. The proof combines last exits from consecutive objective levels with a rank bound for obtuse vectors in a common open halfspace. Combining this conversion with suffix regret yields $O(dDL/\sqrt{n})$ for $\eta_t = D/(L\sqrt{t})$ on a domain of diameter D, and $O(dL^2/(\mu n))$ for both standard strongly convex schedules. Thus their horizon dependence is minimax optimal for every fixed dimension. The result complements recent dimension-free anytime impossibility and sharp horizon-tuned constant-step theory. It also gives data-dependent bounds in terms of trajectory rank and realized, rather than worst-case, step lengths.

1 Introduction

Projected subgradient descent is among the simplest methods for nonsmooth convex optimization. Its averages attain the optimal $n^{-1/2}$ rate for Lipschitz convex objectives and the optimal n^{-1} rate under strong convexity. The last iterate is more delicate. For the standard horizon-free schedules

$$\eta_t \asymp t^{-1/2} \quad \text{and} \quad \eta_t \asymp t^{-1},$$

the classical guarantees lose a factor $\log n$ (Shamir and Zhang, 2013; Harvey et al., 2019). High-dimensional deterministic constructions show that these losses are real (Harvey et al., 2019). Horizon-dependent schedules can remove them (Jain et al., 2021; Zamani and Glineur, 2025), but they must know when the algorithm will stop.

Dimension changes this picture. Koren and Segal (2020) asked whether standard last iterates recover the optimal horizon rate when dimension is fixed. Liu and Lu (2021) proved lower bounds proportional to $\log d$, but left dimension-dependent upper bounds open. Very recent work resolves the deterministic question for a horizon-tuned constant step. For $\eta = \Theta(n^{-1/2})$, the sharp terminal excursion is $\Theta(\min\{d, \log n\})\eta L^2$ (Beretta et al., 2026b). That theorem does not cover a single decreasing schedule that is run without a stopping horizon.

This distinction matters. Kornowski and Shamir (2026) prove that no horizon-free step sequence has a dimension-free optimal last-iterate guarantee. Their obstruction is high-dimensional. It remains compatible with an optimal rate whose constant depends on fixed dimension. The decreasing-step analogue of the fixed-dimensional question has therefore remained open.

We answer it positively for deterministic projected subgradient descent. Our main result is not tied to a power law. It converts a good point on any suffix into a terminal guarantee for arbitrary positive steps, including steps selected adaptively from the past. The dimension is replaced by the rank of the observed suffix trajectory, and the scale is the largest realized step length rather than the worst-case bound $L \max_t \eta_t$.

Contributions. Let x_n be the last iterate and consider a suffix $x_s, \dots, x_n$. If $r_{s,n}$ is the dimension of the span of its terminal displacements, then

$$f(x_n) - \min_{s \le t \le n} f(x_t) \le 2 r_{s,n} L \max_{s \le t < n} \eta_t \|g_t\|. \quad (1)$$

This pathwise inequality has three consequences.

- For every nonincreasing step sequence on a bounded domain, it gives a last-iterate bound from one suffix regret calculation.

- For $\eta_t = \alpha D/(L\sqrt{t})$, it gives $(\alpha^{-1} + \alpha + 4\alpha d)DL/\sqrt{n}$. In particular, the rate is $O_d(n^{-1/2})$ for every fixed d.
- For a μ-strongly convex objective, both $\eta_t = 1/(\mu t)$ and $\eta_t = 2/\{\mu(t+1)\}$ give $O(dL^2/(\mu n))$. This removes the horizon logarithm left by existing last-iterate analyses of these standard schedules.

The result is deterministic. It does not extend mechanically to stochastic gradients because a realized noisy direction need not be a subgradient of the population objective. This boundary is substantive. Recent one-dimensional examples retain a logarithmic loss under general bounded-variance stochastic oracles (Beretta et al., 2026a).

2 Setting

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be nonempty, closed, and convex. Let f be convex on a neighborhood of $\mathcal{X}$, attain its constrained minimum f^*, and satisfy

$$\|g\| \le L \quad \text{for every } x \in \mathcal{X} \text{ and } g \in \partial f(x). \tag{2}$$

Starting from $x_1 \in \mathcal{X}$, projected subgradient descent selects arbitrary $g_t \in \partial f(x_t)$ and positive steps η_t, then updates

$$x_{t+1} = \Pi_{\mathcal{X}}(x_t - \eta_t g_t). \tag{3}$$

The steps may depend on the preceding trajectory. Every result below is pathwise, so no measurability condition is needed.

For $1 \le s < n$, define

$$\rho_{s,n} := \max_{s \le t < n} \eta_t \|g_t\|, \tag{4}$$

$$r_{s,n} := \dim \operatorname{span}\{x_n - x_t : s \le t \le n\}. \tag{5}$$

Always $r_{s,n} \le \min\{d, n-s\}$. It may be much smaller than d if the iterates remain in a data-dependent low-dimensional affine space.

The projection inequality gives, for every $z \in \mathcal{X}$,

$$\|x_{t+1} - z\|^2 \le \|x_t - z\|^2 + \eta_t^2 \|g_t\|^2 - 2\eta_t\{f(x_t) - f(z)\}. \tag{6}$$

Our proof extracts geometric information from this familiar estimate rather than summing it only against a minimizer.

3 A Rank-Adaptive Conversion

Theorem 1 (Suffix-to-last conversion)**.** *Every trajectory generated by* (3) *satisfies, for every* $1 \le s < n$*,*

$$f(x_n) - \min_{s \le t \le n} f(x_t) \le 2 r_{s,n} L \rho_{s,n}. \tag{7}$$

Consequently,

$$f(x_n) - \min_{s \le t \le n} f(x_t) \le 2dL^2 \max_{s \le t < n} \eta_t. \tag{8}$$

The theorem is meaningful even when the domain is unbounded and the steps are nonmonotone. Its scale $\rho_{s,n}$ is invariant under reciprocal rescaling of the step and selected subgradient.

Proof. The claim is immediate if $\rho_{s,n} = 0$. Assume it is positive. Translate by an arbitrary point, scale points by $\rho_{s,n}$, and scale objective values by $L\rho_{s,n}$. We may therefore work with iterates y_t, objective F, and numbers

$$a_t = \frac{\eta_t L}{\rho_{s,n}}, \qquad b_t = \frac{\eta_t \|g_t\|}{\rho_{s,n}} \le 1.$$

Writing $h_t = g_t/L \in \partial F(y_t)$, inequality (6) becomes

$$\|y_{t+1} - z\|^2 \le \|y_t - z\|^2 + b_t^2 - 2a_t\{F(y_t) - F(z)\}. \tag{9}$$

Also $|F(y_{t+1}) - F(y_t)| \le 1$ because the projected update has length at most b_t and F is one-Lipschitz.

Put $m = \min_{s \le t \le n} F(y_t)$, $H = F(y_n)$, and $q = H - m$. There is nothing to prove if $q = 0$. Otherwise set $J = \lceil q \rceil - 1$. For $0 \le j \le J$, let $I_j = \{t \in \{s, \dots, n\} : F(y_t) \le m + j\}$ and define

$$\tau_j = \max I_j, \qquad p_j = y_{\tau_j}, \qquad v_j = y_n - p_j.$$

The one-step objective bound implies

$$\tau_0 < \tau_1 < \cdots < \tau_J < n. \tag{10}$$

Use (9) with comparator p_j and sum from τ_j to $n-1$. Every later function gap is positive, while the first is zero, hence

$$\|v_j\|^2 \le \sum_{t=\tau_j}^{n-1} b_t^2. \tag{11}$$

Now take $i \le j-2$. By (10), every $t \ge \tau_j$ occurs after the last visit below $m+i+1$. Thus $F(y_t) - F(p_i) > 1$. Moreover,

$$\frac{b_t^2}{a_t} = \frac{\eta_t \|g_t\|^2}{L\rho_{s,n}} \le \frac{\eta_t \|g_t\|}{\rho_{s,n}} = b_t \le 1. \tag{12}$$

Summing (9) with comparator p_i from τ_j to $n-1$ and using (12) gives

$$\|y_n - p_i\|^2 < \|p_j - p_i\|^2 - \sum_{t=\tau_j}^{n-1} b_t^2 \le \|p_j - p_i\|^2 - \|v_j\|^2.$$

Since $p_j - p_i = v_i - v_j$, expansion yields

$$\langle v_i, v_j \rangle < 0 \quad \text{whenever } |i-j| \ge 2. \tag{13}$$

Choose $h_n \in \partial F(y_n)$. Convexity at y_n gives

$$\langle h_n, v_j \rangle \geq H - F(p_j) \geq q - j > 0. \tag{14}$$

Within each parity class, the v_j are pairwise obtuse and lie in the same open halfspace. Such vectors are linearly independent. A proof is recalled in the supplement. They all lie in the span defining $r_{s,n}$, so each parity class has at most $r_{s,n}$ members. Therefore $J + 1 \leq 2r_{s,n}$. Since $q \leq J + 1$, undoing the normalization proves (7). The bounds $r_{s,n} \leq d$ and $\rho_{s,n} \leq L \max_t \eta_t$ prove (8). □

The parity split is necessary because consecutive last-exit vectors need not be obtuse. The rank dependence is also unavoidable for a conversion that covers constant steps. The multiscale support-function trajectories of Beretta et al. (2026b) start at a minimizer and have terminal excursion linear in their dimension.

4 Convex Anytime Rates

We first combine Theorem 1 with a suffix regret estimate that retains the actual gradient norms.

Proposition 2 (Nonincreasing steps)**.** *Suppose $\mathcal{X}$ has diameter at most D and η_t is nonincreasing. For every $1 \leq s < n$,*

$$f(x_n) - f^* \leq 2r_{s,n} L \rho_{s,n} + \frac{D^2/\eta_{n-1} + \sum_{t=s}^{n-1} \eta_t \|g_t\|^2}{2(n-s)}. \tag{15}$$

Proof. Divide (6), with a minimizer as comparator, by $2\eta_t$ and sum over the suffix. Because η_t is nonincreasing and all squared distances are at most D^2, summation by parts bounds the distance contribution by $D^2/(2\eta_{n-1})$. At least one suffix iterate before n has gap no larger than the resulting average. Apply Theorem 1. □

Corollary 3 (The standard convex schedule)**.** *Let $n \geq 4$ and*

$$\eta_t = \frac{\alpha D}{L\sqrt{t}}, \qquad \alpha > 0. \tag{16}$$

Then

$$f(x_n) - f^* \leq \frac{DL}{\sqrt{n}} \{\alpha^{-1} + \alpha + 4\alpha r_{\lfloor n/2 \rfloor, n}\}. \tag{17}$$

In particular, $\alpha = 1$ gives $(4d + 2)DL/\sqrt{n}$. If d is known, taking $\alpha = (4d + 1)^{-1/2}$ gives $2\sqrt{4d+1}\, DL/\sqrt{n}$.

The proof is in the supplement. The classical dimension-free analysis is $O(DL \log n/\sqrt{n})$ (Shamir and Zhang, 2013). Combining it with (17) at $\alpha = 1$ gives

$$f(x_n) - f^* = O\left(\frac{DL}{\sqrt{n}} \min\{d+1, \log(n+1)\}\right). \tag{18}$$

The lower bound of Liu and Lu (2021) is $\Omega(DL \log(d+1)/\sqrt{n})$ for schedule (16). Thus a dimension gap remains, but the dependence on n is now exact for every fixed d. This is precisely what high-dimensional anytime impossibility cannot rule out.

The same argument identifies the horizon exponent for every polynomially decreasing schedule.

Corollary 4 (Polynomial schedules)**.** *Let $q \in (0, 1)$, $n \geq 4$, and*

$$\eta_t = \frac{\alpha D}{L t^q}.$$

With $r = r_{\lfloor n/2 \rfloor, n}$,

$$f(x_n) - f^* \leq DL \left\{ \frac{n^{q-1}}{\alpha} + (8r+2)\alpha n^{-q} \right\}. \tag{19}$$

Thus the best horizon exponent in this family is attained uniquely at $q = 1/2$. For fixed r, slower decay is limited by terminal excursion and faster decay is limited by reaching a good suffix point.

The linear rank factor in (17) cannot be removed uniformly over the usual class $\eta_t = \Theta(t^{-1/2})$. The following result uses a staircase schedule fixed once and for all, not a schedule chosen after seeing the stopping horizon.

Theorem 5 (An anytime linear-dimension lower bound)**.** *Define the nonincreasing horizon-free schedule*

$$\eta_t = 2^{-j/2} \quad \text{when } 2^j \leq t < 2^{j+1}. \tag{20}$$

It satisfies $t^{-1/2} \leq \eta_t < \sqrt{2}\, t^{-1/2}$. For every integer $k \geq 1$, there is a dimension $d = 2k + 1$ such that, at every sufficiently large dyadic horizon $n = 2^{J+1}$, a convex one-Lipschitz objective on the Euclidean unit ball, an initialization at a minimizer, and an admissible subgradient selection satisfy

$$f(x_n) - f^* \geq \frac{k}{2\sqrt{2n}}. \tag{21}$$

Consequently, the worst-case constant of this standard anytime schedule is $\Omega(d)$ in fixed dimension.

The proof, given in the supplement, places the finite multiscale trajectory of Beretta et al. (2026b) at the end of a sufficiently long dyadic plateau and selects the zero subgradient beforehand. The trajectory stays inside the unit ball because its squared distance from the starting minimizer grows by at most one per normalized update. This lower bound concerns the staircase schedule (20). It does not strengthen the $\Omega(\log d)$ lower bound known for the exact schedule $t^{-1/2}$.

5 Strongly Convex Anytime Rates

Assume now that f is μ-strongly convex on $\mathcal{X}$:

$$f(y) \geq f(x) + \langle g, y - x\rangle + \frac{\mu}{2}\|y - x\|^2$$

for $x, y \in \mathcal{X}$ and $g \in \partial f(x)$. The standard schedules admit a constant-order suffix best iterate, after scaling by $1/n$. Theorem 1 then prevents the terminal point from losing more than $O(d/n)$.

Theorem 6 (Standard strongly convex schedules)**.** *Let $n \geq 6$ and let $r = r_{\lfloor n/2\rfloor,n}$. Each of the following holds for every initialization and every admissible subgradient selection:*

1. *If $\eta_t = 1/(\mu t)$, then*

$$f(x_n) - f^* \leq \frac{(6r+3)L^2}{\mu n}. \tag{22}$$

2. *If $\eta_t = 2/\{\mu(t+1)\}$, then*

$$f(x_n) - f^* \leq \frac{(8r+4)L^2}{\mu n}. \tag{23}$$

The constants are not optimized.

Proof sketch. For the first schedule, the distance recursion gives

$$\|x_t - x^*\|^2 \leq \frac{L^2}{\mu^2(t-1)}, \qquad t \geq 2. \tag{24}$$

Strong convexity and (6) imply

$$f(x_t) - f^* \leq \frac{1}{2}(\eta_t^{-1} - \mu)\|x_t - x^*\|^2 - \frac{\|x_{t+1} - x^*\|^2}{2\eta_t} + \frac{\eta_t L^2}{2}.$$

Since $\eta_t^{-1} - \mu = \eta_{t-1}^{-1}$, these terms telescope over the final half of the run. This produces a suffix point with gap $O(L^2/(\mu n))$. Theorem 1 adds at most $2rL^2/(\mu\lfloor n/2\rfloor)$.

For the second schedule, $\|x_t - x^*\|^2 \leq 4L^2/(\mu^2 t)$ and $\eta_t^{-1} - \mu \leq \eta_{t-1}^{-1}$. The same suffix argument applies. The supplement gives the index calculations and stated constants. □

Existing dimension-free analyses give $O(L^2 \log n/(\mu n))$ for these horizon-free steps (Harvey et al., 2019; Liu and Zhou, 2024). Hence Theorem 6 yields

$$f(x_n) - f^* = O\left(\frac{L^2}{\mu n}\min\{d+1, \log(n+1)\}\right). \tag{25}$$

The deterministic lower bound is $\Omega(L^2 \log(d+1)/(\mu n))$ (Liu and Lu, 2021), with an $\Omega(L^2/(\mu n))$ bound already in one dimension. Thus the n^{-1} rate is optimal in every fixed dimension. This also answers, in the deterministic fixed-dimensional setting, the question noted by Liu and Zhou (2024) of whether the last iterate under $2/\{\mu(t+1)\}$ can match the n^{-1} rate of nonuniform averaging.

6 Discussion

Table 1 summarizes the deterministic nonsmooth picture. The new entries concern one fixed schedule used at every horizon. Rates suppress problem parameters and universal constants.

Table 1: Last-iterate rates for projected deterministic subgradient descent.

Setting	Prior bound	This paper
Convex, $t^{-1/2}$	$\log n/\sqrt{n}$	$d/\sqrt{n}$
Strong, t^{-1}	$\log n/n$	d/n
Convex, t^{-q}	generic logarithm	Eq. (19)
Arbitrary steps	constant-step result	rank conversion

Relation to constant-step geometry. Theorem 1 uses the last-exit and obtuse-vector mechanism introduced by Beretta et al. (2026b). Three changes are essential for the anytime result. The normalization uses the largest realized update, the comparator argument allows a different multiplier at every step, and the dimension count is localized to the observed suffix span. A late suffix is what joins this geometry to decreasing-step regret.

Why boundedness appears only in the convex corollary. The conversion theorem itself is independent of domain diameter. Boundedness enters Proposition 2 only to control the changing inverse step sizes in suffix regret. Strong convexity supplies its own distance decay, so Theorem 6 needs no diameter assumption.

Limits. Our result does not settle the sharp dependence on dimension for the exact power-law schedules. The known lower factor is $\log(d+1)$ and our upper factor at fixed scale is $d+1$. It also does not settle stochastic fixed-dimensional behavior. The pathwise ordering of objective levels fails when the update direction is only unbiased. Indeed, oracle structure determines even the one-dimensional answer (Beretta et al., 2026a).

Broader steps. Proposition 2 applies verbatim to positive adaptive steps whenever their realized se-

quence is nonincreasing. Theorem 1 itself needs no monotonicity. It can therefore be paired with sharper problem-specific suffix guarantees, including bounds that retain $\sum_t \eta_t \|g_t\|^2$. This separates the last-iterate difficulty from the task of finding one good suffix point.

7 Conclusion

Standard decreasing steps do not incur a horizon logarithm merely because the stopping time is unknown. In deterministic fixed dimension, their last iterates achieve the same $n^{-1/2}$ and n^{-1} horizon rates as averaging. A rank-adaptive suffix conversion explains why: only finitely many well-separated objective levels can be exited for the last time along a finite-dimensional trajectory. The remaining open question is quantitative rather than asymptotic in the horizon, namely whether the linear rank factor can be reduced toward the logarithmic dimension lower bound for the exact power-law schedules.

References

Beretta, G., Cesari, T., Colomboni, R., and Paudice, A. (2026a). New bounds for the last iterate of the stochastic subgradient method. *arXiv preprint arXiv:2606.24879*.

Beretta, G., Cesari, T., Colomboni, R., and Paudice, A. (2026b). Sharp dimension dependence for the last iterate of the subgradient method. *arXiv preprint arXiv:2607.15980*.

Harvey, N. J. A., Liaw, C., Plan, Y., and Randhawa, S. (2019). Tight analyses for non-smooth stochastic gradient descent. In *Proceedings of the 32nd Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1579–1613. PMLR.

Jain, P., Nagaraj, D. M., and Netrapalli, P. (2021). Making the last iterate of SGD information theoretically optimal. *SIAM Journal on Optimization*, 31(2):1108–1130.

Koren, T. and Segal, S. (2020). Open problem: Tight convergence of SGD in constant dimension. In *Proceedings of the 33rd Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 3847–3851. PMLR.

Kornowski, G. and Shamir, O. (2026). Gradient descent's last iterate is often (slightly) suboptimal. *arXiv preprint arXiv:2604.13870*.

Liu, D. and Lu, Z. (2021). The convergence rate of SGD's final iterate: Analysis on dimension dependence. *arXiv preprint arXiv:2106.14588*.

Liu, Z. and Zhou, Z. (2024). Revisiting the last-iterate convergence of stochastic gradient methods. In *The Twelfth International Conference on Learning Representations*.

Shamir, O. and Zhang, T. (2013). Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 71–79. PMLR.

Zamani, M. and Glineur, F. (2025). Exact convergence rate of the last iterate in subgradient methods. *SIAM Journal on Optimization*, 35(3):2182–2201.

A Linear Algebra Used in the Main Proof

Lemma 7 (Obtuse vectors in an open halfspace)**.** *Let $v_1, \dots, v_M$ lie in a vector space of dimension r. If there is a vector u such that $\langle u, v_i \rangle > 0$ for every i, and $\langle v_i, v_j \rangle < 0$ for every $i \neq j$, then $M \leq r$.*

Proof. Suppose $\sum_i \alpha_i v_i = 0$ is a nontrivial linear dependence. Taking its inner product with u shows that the nonzero coefficients cannot all have the same sign. Let $P = \{i : \alpha_i > 0\}$ and $N = \{j : \alpha_j < 0\}$. Both are nonempty and

$$w := \sum_{i \in P} \alpha_i v_i = \sum_{j \in N} (-\alpha_j) v_j.$$

Taking the inner product of these two representations gives

$$\|w\|^2 = \sum_{i \in P} \sum_{j \in N} \alpha_i (-\alpha_j) \langle v_i, v_j \rangle < 0,$$

a contradiction. The vectors are linearly independent, so there are at most r of them. □

B Convex Suffix Bound

We prove Proposition 2 and Corollary 3 from the main paper with all constants.

Let x^* be a constrained minimizer and put $A_t = \|x_t - x^*\|^2$. The projection inequality gives

$$2\{f(x_t) - f^*\} \leq \frac{A_t - A_{t+1}}{\eta_t} + \eta_t \|g_t\|^2. \tag{26}$$

For a nonincreasing positive sequence,

$$\begin{aligned} \sum_{t=s}^{n-1} \frac{A_t - A_{t+1}}{\eta_t} &= \frac{A_s}{\eta_s} + \sum_{t=s+1}^{n-1} A_t \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) - \frac{A_n}{\eta_{n-1}} \\ &\leq \frac{D^2}{\eta_{n-1}}. \end{aligned} \tag{27}$$

Summing (26), dividing by $2(n-s)$, and using the smallest gap among $x_s, \dots, x_{n-1}$ proves

$$\min_{s\le t<n}\{f(x_t)-f^*\} \le \frac{D^2/\eta_{n-1}+\sum_{t=s}^{n-1}\eta_t\|g_t\|^2}{2(n-s)}.$$

Theorem 1 of the main paper converts this point to the terminal iterate and proves Proposition 2.

For Corollary 3, let $s=\lfloor n/2\rfloor$ and $m=n-s$. When $n\ge 4$, $m\ge n/2$ and $s\ge n/4$. Under $\eta_t=\alpha D/(L\sqrt{t})$,

$$\begin{aligned} 2r_{s,n}L\rho_{s,n} &\le \frac{2\alpha r_{s,n}DL}{\sqrt{s}} \le \frac{4\alpha r_{s,n}DL}{\sqrt{n}},\\ \frac{D^2}{2m\eta_{n-1}} &\le \frac{DL}{\alpha\sqrt{n}},\\ \frac{L^2\sum_{t=s}^{n-1}\eta_t}{2m} &\le \frac{\alpha DL}{2\sqrt{s}} \le \frac{\alpha DL}{\sqrt{n}}. \end{aligned}$$

Adding these inequalities gives the stated bound. The choice $\alpha=(4d+1)^{-1/2}$ minimizes $\alpha^{-1}+(4d+1)\alpha$.

C Strongly Convex Suffix Bounds

Strong convexity gives, for every selected subgradient,

$$\langle g_t, x_t-x^*\rangle \ge f(x_t)-f^*+\frac{\mu}{2}A_t \ge \mu A_t. \tag{28}$$

The second inequality also uses $f(x_t)-f^*\ge \mu A_t/2$. The projection inequality first gives

$$A_{t+1}\le (1-\mu\eta_t)A_t+\eta_t^2L^2. \tag{29}$$

Using the second inequality in (28) gives the sharper version

$$A_{t+1}\le (1-2\mu\eta_t)A_t+\eta_t^2L^2. \tag{30}$$

Retaining the objective gap only once gives

$$f(x_t)-f^*\le \frac{1}{2}(\eta_t^{-1}-\mu)A_t-\frac{A_{t+1}}{2\eta_t}+\frac{\eta_tL^2}{2}. \tag{31}$$

C.1 The schedule $\eta_t=1/(\mu t)$

Put $C=L^2/\mu^2$. Equation (30) gives $A_2\le C$. If $A_t\le C/(t-1)$ for $t\ge 2$, then

$$A_{t+1}\le\left(1-\frac{2}{t}\right)\frac{C}{t-1}+\frac{C}{t^2}=\frac{C(t^2-t-1)}{t^2(t-1)}\le\frac{C}{t}.$$

Thus $A_t\le C/(t-1)$ for every $t\ge 2$.

Since $\eta_t^{-1}-\mu=\eta_{t-1}^{-1}$, summing (31) from s to $n-1$ gives

$$\begin{aligned}\sum_{t=s}^{n-1}\{f(x_t)-f^*\} &\le \frac{A_s}{2\eta_{s-1}}+\frac{L^2}{2}\sum_{t=s}^{n-1}\eta_t\\ &\le \frac{L^2}{2\mu}+\frac{L^2(n-s)}{2\mu s}.\end{aligned} \tag{32}$$

Let $s=\lfloor n/2\rfloor$ and $m=n-s$. For $n\ge 6$, $s\ge n/3$, $m\ge n/2$, and $m/s\le 2$. Therefore the best suffix gap is at most $3L^2/(\mu n)$. The conversion term is at most

$$2r_{s,n}L^2\eta_s=\frac{2r_{s,n}L^2}{\mu s}\le\frac{6r_{s,n}L^2}{\mu n}.$$

This proves the first claim of Theorem 6.

C.2 The schedule $\eta_t=2/\{\mu(t+1)\}$

Equation (29) gives $A_2\le C$. We claim $A_t\le 4C/t$. Indeed, if the claim holds at t, then

$$A_{t+1}\le\frac{t-1}{t+1}\frac{4C}{t}+\frac{4C}{(t+1)^2}\le\frac{4C}{t+1}.$$

The last inequality is equivalent to $(t-1)/t+1/(t+1)\le 1$.

Here $\eta_t^{-1}-\mu=\mu(t-1)/2\le\eta_{t-1}^{-1}$. Summing (31) and discarding the resulting nonpositive interior terms gives

$$\begin{aligned}\sum_{t=s}^{n-1}\{f(x_t)-f^*\} &\le \frac{A_s}{2\eta_{s-1}}+\frac{L^2}{2}\sum_{t=s}^{n-1}\eta_t\\ &\le \frac{L^2}{\mu}+\frac{L^2(n-s)}{\mu(s+1)}.\end{aligned} \tag{33}$$

For $s=\lfloor n/2\rfloor$, the best suffix gap is at most $4L^2/(\mu n)$. The conversion term is bounded by

$$2r_{s,n}L^2\eta_s=\frac{4r_{s,n}L^2}{\mu(s+1)}\le\frac{8r_{s,n}L^2}{\mu n}.$$

This proves the second claim of Theorem 6.

D Polynomial Schedules and Staircase Lower Bound

D.1 Proof of the polynomial-schedule corollary

Take $s=\lfloor n/2\rfloor$ and $m=n-s$. For $n\ge 4$, we have $m\ge n/2$ and $s\ge n/4$. Under $\eta_t=\alpha D/(Lt^q)$, Proposition 2 in the main paper gives

$$\begin{aligned} 2r_{s,n}L\rho_{s,n} &\le 2\alpha r_{s,n}DLs^{-q}\le 8\alpha r_{s,n}DLn^{-q},\\ \frac{D^2}{2m\eta_{n-1}} &\le \frac{DL}{\alpha}n^{q-1},\\ \frac{L^2\sum_{t=s}^{n-1}\eta_t}{2m} &\le \frac{\alpha DL}{2s^q}\le 2\alpha DLn^{-q}. \end{aligned}$$

Their sum is the polynomial-schedule display in the main paper. For fixed rank and scale, its two powers balance only when $q=1/2$.

D.2 Proof of Theorem 5

We use the following finite construction from Beretta et al. (2026b, Theorem 13). For every integer $k \geq 1$, there are a finite set $A_k \subset \mathbb{R}^{2k+1}$ containing zero, vectors $a_1, \ldots, a_{T_k+1} \in A_k \setminus \{0\}$, and an integer $T_k < 1000^{k+1}$ with these properties:

1. every vector in A_k has norm at most one
2. if $S_t = \sum_{i<t} a_i$, then $a_t \in \arg\min_{a \in A_k} \langle S_t, a \rangle$
3. $-\langle S_{T_k+1}, a_{T_k+1} \rangle = k/4$

For the support function

$$F_k(y) = \max_{a \in A_k} \langle a, y \rangle,$$

these properties imply that F_k is convex and one-Lipschitz, $F_k \geq 0$, and the unit-step subgradient method may follow

$$\begin{aligned} y_1 &= 0, \\ y_{t+1} &= y_t - a_t = -S_{t+1}, \\ F_k(y_{T_k+1}) &= \frac{k}{4}. \end{aligned}$$

Fix any J with $2^J \geq T_k$ and set $n = 2^{J+1}$. The final plateau of the staircase schedule in the main paper has step $\eta = 2^{-J/2}$. Run at the origin using the admissible subgradient zero until update $n - T_k$. During the final T_k updates, select $a_1, \ldots, a_{T_k}$. Positive scaling preserves the active pieces of a support function, so the resulting points are

$$x_{n-T_k+r} = \eta y_{r+1}, \qquad 0 \leq r \leq T_k.$$

These updates are unaffected by projection onto the unit ball. Indeed, the projection comparator inequality for the normalized trajectory, with comparator zero, gives

$$\|y_{r+1}\|^2 \leq \|y_r\|^2 + 1 - 2F_k(y_r) \leq \|y_r\|^2 + 1.$$

Thus $\|y_{r+1}\| \leq \sqrt{r}$, and

$$\|x_{n-T_k+r}\| \leq \eta\sqrt{T_k} \leq 2^{-J/2} 2^{J/2} = 1.$$

The origin is a minimizer on the ball. At the terminal point,

$$F_k(x_n) = \eta F_k(y_{T_k+1}) = \frac{k}{4\,2^{J/2}} = \frac{k}{2\sqrt{2n}}.$$

This proves Theorem 5. Since $k = (d-1)/2$, the fixed-dimensional constant grows linearly in d.

E Scope and Literature Audit

The following points delimit the contribution.

- Shamir and Zhang (2013) establish the classical $O(\log n/\sqrt{n})$ and $O(\log n/n)$ last-iterate upper bounds.
- Harvey et al. (2019) give matching deterministic high-dimensional constructions for the standard decreasing schedules.
- Jain et al. (2021) and Zamani and Glineur (2025) remove the logarithm using schedules that depend on the stopping horizon.
- Koren and Segal (2020) isolate fixed dimension and explicitly note the analogous question for decreasing steps.
- Liu and Lu (2021) prove $\Omega(\log d/\sqrt{n})$ and $\Omega(\log d/n)$ deterministic lower bounds, and leave dimension-dependent upper bounds open.
- Kornowski and Shamir (2026) rule out a dimension-free optimal anytime guarantee for every step sequence. Their construction allows dimension to grow with the horizon.
- Beretta et al. (2026b) prove the sharp $\min\{d, \log n\}$ terminal excursion for a horizon-tuned constant step. The main paper adapts their last-exit geometry to arbitrary realized step lengths and local trajectory rank.
- Liu and Zhou (2024) provide a broad stochastic and composite theory. Their March 2026 version retains $O(\log n/n)$ for the standard nonsmooth strongly convex anytime schedules and notes that an $O(1/n)$ last-iterate result for $2/\{\mu(t+1)\}$ is unknown.
- Beretta et al. (2026a) show that stochastic fixed-dimensional behavior depends on oracle structure. Their general bounded-variance lower bound can retain a logarithm already in one dimension.

We found no prior pathwise suffix-to-last inequality involving trajectory rank and realized step lengths, nor a deterministic fixed-dimensional $O(n^{-1/2})$ or $O(n^{-1})$ theorem for the standard decreasing last iterates. The new results do not claim a sharp dependence on d and do not apply to noisy stochastic directions.