

An $N + 1$ -dimensional bound for the terminal objective of gradient descent

Logan Bell

September 2026

Abstract

We consider the worst possible terminal function-value gap after N steps of gradient descent on a convex function with 1-Lipschitz gradient. We show that replacing the ambient dimension d by $\min\{d, N + 1\}$ leaves the worst-case value unchanged. Consequently the unrestricted worst case equals the worst case in dimension $N + 1$. We prove this by restricting each instance to the span of the initial displacement and the N gradients that drive the updates. This reduces by one the $N + 2$ dimension bound supplied by the natural performance-estimation Gram representation.

1 The terminal-objective problem

Let $\mathcal{F}_d$ be the class of differentiable convex functions $f : \mathbf{R}^d \rightarrow \mathbf{R}$ whose gradients are 1-Lipschitz and which have at least one minimizer. Fix an integer $N \geq 0$ and a real step schedule $h = (h_1, \dots, h_N) \in \mathbf{R}^N$. Starting from x_0 , gradient descent forms

$$x_{k+1} = x_k - h_{k+1} \nabla f(x_k), \quad 0 \leq k < N. \quad (1)$$

For each $d \geq 1$, define

$$W_d(h) = \sup \left\{ f(x_N) - f(x_\star) : \begin{array}{l} f \in \mathcal{F}_d, \quad x_\star \in \operatorname{argmin} f, \quad \|x_0 - x_\star\| \leq 1, \\ x_0, \dots, x_N \text{ obey (1)} \end{array} \right\},$$

and let

$$W(h) = \sup_{d \geq 1} W_d(h).$$

Adding coordinates on which a function is constant shows that $W_d(h)$ is nondecreasing in d . The theorem below shows that it is constant for $d \geq N + 1$.

2 Restriction to the trajectory span

Theorem 1. *For every $N \geq 0$, every $h \in \mathbf{R}^N$, and every $d \geq 1$,*

$$W_d(h) = W_{\min\{d, N+1\}}(h), \quad W(h) = W_{N+1}(h). \quad (2)$$

Proof. Take any admissible d -dimensional instance, and abbreviate $g_k = \nabla f(x_k)$. Consider the linear subspace

$$S = \text{span}\{x_0 - x_\star, g_0, \dots, g_{N-1}\}. \quad (3)$$

It has dimension at most $N + 1$. The recurrence (1) shows inductively that $x_k - x_\star \in S$ for every $0 \leq k \leq N$.

Restrict f to the affine subspace $x_\star + S$, and identify that subspace with S by writing

$$\hat{f}(s) = f(x_\star + s) - f(x_\star), \quad s \in S.$$

If P_S denotes orthogonal projection onto S , then the intrinsic gradient of the restriction is

$$\nabla \hat{f}(s) = P_S \nabla f(x_\star + s).$$

Restriction preserves convexity, and contractivity of projection shows that $\hat{f}$ has 1-Lipschitz gradient on S . Moreover, the origin minimizes $\hat{f}$. At every preterminal iterate $k < N$, the gradient g_k belongs to S by construction. Hence

$$\nabla \hat{f}(x_k - x_\star) = g_k, \quad 0 \leq k < N.$$

The restricted instance therefore follows exactly the same recurrence through $x_N - x_\star$, has the same initial distance, and has the same terminal gap $f(x_N) - f(x_\star)$.

Write $r = \dim S \leq \min\{d, N + 1\}$. Identify S with $\mathbf{R}^r$ and extend $\hat{f}$ constantly in any additional coordinates. This realizes the same terminal value in dimension $\min\{d, N + 1\}$; when $r = 0$, use the constant function. Thus $W_d(h) \leq W_{\min\{d, N+1\}}(h)$. The reverse inequality follows by embedding into $\mathbf{R}^d$. Taking the supremum over d proves the second equality in (2). $\square$

3 Relation to the performance-estimation Gram matrix

The usual performance-estimation formulation records the Gram matrix of

$$x_0 - x_\star, \quad g_0, \dots, g_N, \quad (4)$$

whose rank is the dimension needed to realize the sampled data [1, Theorem 5 and Proposition 2]. This gives a sufficient dimension bound of $N + 2$, also used in formulations that optimize the schedule [2].

The restriction in Theorem 1 keeps all $N + 2$ vectors in the Gram matrix but replaces g_N by $P_S g_N$. The resulting matrix has rank at most $N + 1$ and realizes the same terminal value. The terminal gradient drives no update, so this projection preserves the trajectory. It can decrease $\|g_N\|$; a bound for a terminal-gradient criterion would require a separate argument.

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] 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.