# The exact last-gradient bound for constant-step gradient descent in the strongly convex case

Dev D. Goyal

29 October 2026; revised 1 November 2026

**Abstract**

Taylor, Hendrickx and Glineur conjectured the exact worst-case final-gradient norm of constant-step gradient descent on smooth strongly convex functions, normalized by the initial distance to a minimizer [1, Conjecture 3]. We prove the conjectured formula for every normalized step in $[0, 2]$, every finite horizon and every condition ratio in $[0, 1]$, and show that it is attained. The proof combines an explicit finite interpolation certificate for the balanced step, which retains a joint terminal potential, with two reductions. On the progress branch, scaling the objective preserves a weighted interpolation surplus. On the oscillation branch, decreasing an auxiliary curvature parameter, sometimes below zero, brings the given step to balance; the scalar factors remain nonnegative throughout this auxiliary extension. The familiar one-dimensional strongly convex Huber spline and quadratic attain the two branches. The smooth-convex curve was established earlier [6]; the contribution is the complete positive strong-curvature case. The actual-function upper bound and normalized attainment, including every closed parameter boundary, are verified in Lean 4.19.0, and sharpness at general scale follows by an explicit rescaling. The literature review is bounded by the sources and versions cited here.

## Contents

**1 Introduction** **2**
- 1.1 Source of the problem and prior work . . . 3
- 1.2 Normalization . . . 4

**2 Interpolation from actual smooth functions** **4**

**3 The balanced scalar construction** **5**
- 3.1 Ratios and their signs . . . 6
- 3.2 Four explicit factors . . . 6
- 3.3 Reverse recurrence and product identities . . . 7

**4 The exact balanced history identity** **8**
- 4.1 Divergence and cuts . . . 8
- 4.2 Cancellation of every off-diagonal term . . . 9
- 4.3 Every diagonal coefficient . . . 9
- 4.4 The joint potential with surplus retained . . . 10

**5 Recovering both branches at every step** **11**
- 5.1 The progress branch . . . 11
- 5.2 The oscillation branch and negative auxiliary curvature . . . 12

**6 Endpoints and equal curvature** **13**

**7 Sharpness of both branches** **13**

**8 Formal verification scope** **14**

**9 Research process and AI disclosure** **16**

# 1 Introduction

This paper determines the exact worst case of the last gradient of constant-step gradient descent on smooth strongly convex functions, measured against the initial distance to a minimizer. The formula was conjectured by Taylor, Hendrickx and Glineur [1, Conjecture 3], who also supplied the sharp examples. Theorem 1.1 proves it for all parameters, and Section 1.1 explains which part of the conjecture this is and how the result relates to prior work. The case of zero strong convexity is due to Jones [6]; the contribution here is the positive strong-curvature case. The upper bound and the normalized attaining examples are formalized in Lean (Section 8); only the rescaling of the examples to general $L$ and $R$ is a conventional argument.

Let $E$ be a real Hilbert space, $L > 0$, and $0 \le \mu \le L$. Let $f : E \to \mathbb{R}$ be differentiable with an $L$-Lipschitz gradient and $\mu$-strongly convex:

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \tag{1.1}$$

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|^2. \tag{1.2}$$

For differentiable functions, (1.2) is equivalent to convexity of $x \mapsto f(x) - \mu\|x\|^2/2$. Suppose $x_*$ is a global minimizer and consider

$$x_{i+1} = x_i - \frac{h}{L}\nabla f(x_i), \qquad 0 \le i < N, \qquad 0 \le h \le 2. \tag{1.3}$$

Write $\kappa = \mu/L$ and $R = \|x_0 - x_*\|$. Define

$$p_N(\kappa, h) = \begin{cases} \dfrac{1}{1 + Nh}, & \kappa = 0, \\ \dfrac{\kappa(1 - \kappa h)^N}{1 - (1 - \kappa)(1 - \kappa h)^N}, & 0 < \kappa \le 1, \end{cases} \qquad \rho_N(\kappa, h) = \max\{p_N(\kappa, h),\, |1 - h|^N\}. \tag{1.4}$$

For $\kappa > 0$ the denominator is at least $\kappa$, since $|1 - \kappa h| \le 1$. Thus the formula includes $1 - \kappa h = 0$, $h = 0$, and $N = 0$. Every zeroth power is one. The progress expression may be negative when $1 - \kappa h < 0$ and $N$ is odd; the maximum is nevertheless nonnegative.

**Theorem 1.1** (Exact last-gradient bound)**.** *Under* (1.1)–(1.3)*, for every integer* $N \ge 0$*,*

$$\|\nabla f(x_N)\| \le LR\,\rho_N(\mu/L, h). \tag{1.5}$$

*For every admissible* $\mu, L, h, N$ *and* $R > 0$*, equality is attained by a one-dimensional function in the same class with initial distance* $R$*.*

When $\kappa > 0$ and $1 - \kappa h \ne 0$, the original conjectured progress expression is

$$\frac{\kappa}{(\kappa - 1) + (1 - \kappa h)^{-N}},$$

which equals (1.4). The rational form also specifies its continuous extension through a zero base. Its limit at $\kappa = 0$ is $1/(1 + Nh)$: with $a = 1 - \kappa h$, the finite geometric identity gives

$$\frac{\kappa a^N}{1 - (1 - \kappa)a^N} = \frac{a^N}{1 + (1 - \kappa)h \sum_{j=0}^{N-1} a^j}. \tag{1.6}$$

### 1.1 Source of the problem and prior work

Theorem 1.1 is the final-gradient, initial-distance problem in Conjecture 3, Section 4.1.3, of Taylor–Hendrickx–Glineur [1]. That paper supplies the two types of sharp examples used below. The normalization matters: a sharp result for $\|\nabla f(x_N)\|^2/(f(x_0) - f_*)$ does not itself establish the sharp value of $\|\nabla f(x_N)\|/\|x_0 - x_*\|$.

The relevant prior results have distinct scopes.

- Taylor–Hendrickx–Glineur [2] prove exact same-measure linear contraction results for proximal gradient. Their Section 4 and Table 1 explicitly leave the mixed final-residual/initial-distance criterion unknown. The mixed measures proved there run in different directions.
- Kim [3] proves the exact final-function-gap rate under initial-distance normalization and relates it to final-gradient rates under initial-function-gap normalization.
- Rotaru–Glineur–Patrinos [4] give exact gradient-norm rates for all constant steps with initial objective-gap normalization. This is not a sharp initial-distance result.
- Chikhi [5] proves the optimal constant step, value, and uniqueness in the smooth-convex final-gradient problem, using a factorized interpolation certificate. The full curve away from that optimum is expressly left open there.
- Jones [6] subsequently establishes the complete $\kappa = 0$ curve, equality trajectories, and related proximal results. That repository expressly excludes positive strong convexity from its established claim. The factorized-certificate approach and the retained interpolation surplus under scaling are prior ingredients and are credited accordingly.

The close prior exact results in [4, 3], combined with $f(x_0) - f_* \le LR^2/2$, give, when $0 < \kappa < 1$, $a = 1 - \kappa h > 0$, and $t = a^N$, the progress factor

$$\frac{\sqrt{\kappa}\, t}{\sqrt{1 - (1 - \kappa)t^2}}, \quad \text{whereas the target factor is} \quad \frac{\kappa t}{1 - (1 - \kappa)t}. \tag{1.7}$$

The second does not follow from that conversion: the difference of the relevant squared denominators is

$$[1 - (1 - \kappa)t]^2 - \kappa[1 - (1 - \kappa)t^2] = (1 - \kappa)(1 - t)^2.$$

For fixed $0 < \kappa < 1$ as $t \to 0$, the ratio of the target progress factor to the converted factor tends to $\sqrt{\kappa}$. The oscillation factor $|1 - h|^N$ is the same.

The intended contribution is the full positive-curvature curve. The established $\kappa = 0$ theorem is credited to prior work. The proof below does not use the convex full-curve theorem as a black box. Basic analytic and finite-sum infrastructure in the accompanying Lean development is adapted from the MIT-licensed Jones project; its license and provenance accompany the source distribution.

A 2023 Glineur lecture presentation [7] displays the target formula as an equality under *mixed performance criteria*. This is not treated as a prior analytic proof: the same presentation distinguishes numerical PEP output from rigorous proofs, and displays analogous exact-max formulas whose

full proof is expressly unavailable there. The cited 2018 paper's unknown entry and the later primary-source scope statements are the relevant distinctions. The review is bounded by sources consulted through 29 October 2026 and is not a worldwide priority claim.

### 1.2 Normalization

Replacing $f$ by $F(z) = (f(x_* + z) - f(x_*))/L$ reduces the proof to $L = 1$, $x_* = 0$, $f_* = 0$, and curvature $\kappa \in [0, 1]$. Retain notation $f, x_i, g_i$ for normalized data:

$$g_i = \nabla f(x_i), \qquad x_{i+1} = x_i - h g_i. \tag{1.8}$$

The balanced certificate uses a parameter $\theta$ that may be negative. The actual function has $\kappa \geq 0$; the negative parameter is an algebraic device.

## 2 Interpolation from actual smooth functions

For data $(x_i, g_i, f_i)$ and $\theta < 1$, put

$$\begin{aligned} \mathcal{Q}_\theta(i,j) = 2(1-\theta)(f_i - f_j) - 2\langle g_j - \theta g_i, x_i - x_j\rangle \\ - \|g_i - g_j\|^2 - \theta\|x_i - x_j\|^2. \end{aligned} \tag{2.1}$$

Starred data mean $x_* = g_* = f_* = 0$. The gradient at the actual minimizer is zero because every directional derivative there vanishes. The smooth descent inequality used below also follows directly from the Lipschitz hypothesis: for $d = z - x$,

$$f(z) - f(x) - \langle \nabla f(x), d\rangle = \int_0^1 \langle \nabla f(x + td) - \nabla f(x), d\rangle \, dt \leq \int_0^1 t\|d\|^2 \, dt = \tfrac{1}{2}\|d\|^2.$$

**Lemma 2.1** (Necessary interpolation inequality)**.** *If $f$ has a 1-Lipschitz gradient and satisfies the strong first-order inequality with parameter $\theta < 1$, then $\mathcal{Q}_\theta(i,j) \geq 0$ for its actual values and gradients.*

*Proof.* Let $x = x_i$, $y = x_j$, $v = x - y$, $w = \nabla f(x) - \nabla f(y)$, and set

$$u = \frac{w - \theta v}{1 - \theta}, \qquad z = x - u.$$

Strong first-order support at $y$ and smooth descent at $x$ give

$$\begin{aligned} f(y) + \langle \nabla f(y), z - y\rangle + \frac{\theta}{2}\|z - y\|^2 \leq f(z), \\ f(z) \leq f(x) + \langle \nabla f(x), z - x\rangle + \frac{1}{2}\|z - x\|^2. \end{aligned}$$

Their difference is nonnegative. Substituting $z - x = -u$ and $z - y = v - u$ shows that it equals

$$\begin{aligned} f(x) - f(y) - \langle \nabla f(y), v\rangle - \frac{\theta}{2}\|v\|^2 - \langle w - \theta v, u\rangle + \frac{1-\theta}{2}\|u\|^2 \\ = f(x) - f(y) - \langle \nabla f(y), v\rangle - \frac{\theta}{2}\|v\|^2 - \frac{\|w - \theta v\|^2}{2(1-\theta)}. \end{aligned}$$

Multiplication by $2(1 - \theta) > 0$ and expansion gives (2.1). □

Write $\Delta x = x_i - x_j$ and $\Delta g = g_i - g_j$.

**Lemma 2.2** (Changing curvature and scaling)**.** *For $\kappa < 1$ and arbitrary real $\nu, s$,*

$$\mathcal{Q}_\nu(i,j) = \frac{1-\nu}{1-\kappa}\mathcal{Q}_\kappa(i,j) + \frac{\kappa-\nu}{1-\kappa}\|\Delta x - \Delta g\|^2, \tag{2.2}$$

$$\mathcal{Q}^{(s)}_{s\kappa}(i,j) = \frac{s(1-s\kappa)}{1-\kappa}\mathcal{Q}_\kappa(i,j) + \frac{s(1-s)}{1-\kappa}\|\Delta g - \kappa\Delta x\|^2. \tag{2.3}$$

*Here $\mathcal{Q}^{(s)}$ uses $(x_i, sg_i, sf_i)$. Thus nonnegative $\mathcal{Q}_\kappa$ imply nonnegative $\mathcal{Q}_\nu$ for $\nu \le \kappa < 1$, and nonnegative $\mathcal{Q}^{(s)}_{s\kappa}$ for $0 \le s \le 1$, $0 \le \kappa < 1$.*

*Proof.* After multiplication by $1-\kappa$, the function-value terms cancel on subtracting the indicated multiple of $\mathcal{Q}_\kappa$. The remainder in the first identity is

$$(\kappa - \nu)(\|\Delta x\|^2 - 2\langle \Delta g, \Delta x\rangle + \|\Delta g\|^2),$$

and in the second it is

$$s(1-s)(\|\Delta g\|^2 - 2\kappa\langle \Delta g, \Delta x\rangle + \kappa^2\|\Delta x\|^2).$$

These expansions prove the identities and their sign conclusions. □

For actual normalized convex data, $\mathcal{Q}_0(i,*) \ge 0$ gives $\|g_i\|^2 \le 2f_i$, and

$$\|x_i\|^2 - \|g_i\|^2 = \mathcal{Q}_0(*,i) + \mathcal{Q}_0(i,*) + \|x_i - g_i\|^2 \ge 0. \tag{2.4}$$

These terminal estimates apply even with negative auxiliary curvature.

## 3 The balanced scalar construction

Assume

$$N \ge 1, \quad 1 < h < 2, \quad r = h - 1, \quad a > r, \quad \theta = \frac{1-a}{h}. \tag{3.1}$$

Thus $0 < r < 1$, $a > 0$, and $\theta < 1$. No assumption $a \le 1$ is made. Define

$$S_m = h\sum_{j=1}^{m} a^{-j}, \qquad Y_m = r^{-m}, \qquad P_m = \frac{Y_m^2 - 1}{2} \quad (m \ge 0), \tag{3.2}$$

and impose balance:

$$Y_N = 1 + S_N. \tag{3.3}$$

Empty sums are zero. Geometric sums give

$$aS_{m+1} = S_m + h, \qquad S_{m+1} - S_m = ha^{-m-1}, \tag{3.4}$$

$$rY_{m+1} = Y_m, \qquad r^2P_{m+1} - P_m = \frac{1-r^2}{2}, \tag{3.5}$$

$$a^{-m} = 1 + \theta S_m. \tag{3.6}$$

The last identity follows from $(a-1)S_m = h(1-a^{-m})$ and $a - 1 = -\theta h$, including $a = 1$.

### 3.1 Ratios and their signs

For $1 \le m \le N$, put

$$b_m = \frac{S_m}{Y_m + 1}, \qquad c_m = \frac{S_m}{Y_m - 1}. \tag{3.7}$$

Both are positive. Since $Y_m - 1 = (1 - r)\sum_{j=1}^{m} r^{-j}$,

$$c_m = \frac{h}{1 - r} \frac{\sum_{j=1}^{m} a^{-j}}{\sum_{j=1}^{m} r^{-j}}. \tag{3.8}$$

The fraction is a weighted average of $(r/a)^j$ with positive weights $r^{-j}$. Each added value is smaller than every previous one, so $c_m$ strictly decreases. Balance gives $c_N = 1$ and hence

$$S_m > Y_m - 1 \qquad (1 \le m < N). \tag{3.9}$$

For an explicit strict comparison, if $0 < p < q$ and $G_m(t) = \sum_{j=1}^{m} t^j$, then

$$G_m(p)q^{m+1} - p^{m+1}G_m(q) = \sum_{j=1}^{m} p^j q^j (q^{m+1-j} - p^{m+1-j}) > 0.$$

Using $G_{m+1}(t) = G_m(t) + t^{m+1}$ proves strict decrease of the ratio; apply $p = a^{-1}$ and $q = r^{-1}$.

### 3.2 Four explicit factors

Use natural boundary indexing:

$$U_0 = B_0 = 0, \quad U_1 = 1, \quad U_{m+1} = U_m \frac{1 + b_m}{ab_{m+1} - r} \quad (1 \le m < N). \tag{3.10}$$

Define

$$B_m = b_m U_m, \quad D_m = P_m / U_m, \quad V_m = c_m D_m \quad (1 \le m \le N), \qquad D_{N+1} = V_{N+1} = 0. \tag{3.11}$$

Set $D_0 = V_0 = 0$ for terms multiplied by $U_0$ or $B_0$. Every denominator used in (3.10) is positive:

$$ab_{m+1} - r = \frac{1 + S_m - Y_m}{Y_{m+1} + 1} > 0 \quad (1 \le m < N). \tag{3.12}$$

This follows from (3.4), $h = 1 + r$, and (3.9). The zero denominator $ab_1 - r$ is never used.

**Lemma 3.1** (Factor monotonicity)**.** *$U_m$ and $B_m$ strictly increase on indices $0, \dots, N$. The sequences $D_m, V_m$ are positive on indices $1, \dots, N$ and strictly decrease through their terminal value zero at index $N + 1$.*

*Proof.* Positivity of $U_m$ follows from its recursion. The exact difference

$$\begin{aligned} &(1 + b_m)P_m - (ab_{m+1} - r)P_{m+1} \\ &\quad = \frac{h(1 - r)Y_m}{2r} \left( \sum_{j=1}^{m} r^{-j} - \sum_{j=1}^{m} a^{-j} \right) > 0 \qquad (1 \le m < N) \end{aligned} \tag{3.13}$$

compares the growth of $U_m$ to that of $P_m$. To verify it, use $P_m b_m = S_m(Y_m - 1)/2$, $aS_{m+1} = S_m + h$, and $Y_{m+1} = Y_m/r$. Twice the left side becomes

$$(Y_m^2 - 1) + S_m(Y_m - 1) - (S_m + h)(Y_m/r - 1) + r(Y_m^2/r^2 - 1).$$

With $h = 1 + r$ this is $Y_m[h(Y_m - 1) - (1 - r)S_m]/r$. Insert the geometric formulas for $Y_m - 1$ and $S_m$. The parenthesis in (3.13) is positive since $r < a$. Since $P_{m+1} > P_m > 0$,

$$\frac{U_{m+1}}{U_m} = \frac{1 + b_m}{ab_{m+1} - r} > \frac{P_{m+1}}{P_m} > 1.$$

Thus $U$ increases and $D$ decreases, including the stated boundaries.

The recursion for $U$ gives

$$aB_{m+1} - B_m = rU_{m+1} + U_m \qquad (0 \le m < N). \tag{3.14}$$

At $m = 0$ this uses $ab_1 = r$. To show that $B$ increases even if $a > 1$, note

$$(a - 1)b_m < h : \tag{3.15}$$

indeed $(a - 1)S_m = h(1 - a^{-m}) < h < h(Y_m + 1)$. Multiplying by $U_m > 0$ gives

$$\begin{aligned} aB_m &= B_m + (a - 1)B_m < B_m + hU_m \\ &= B_m + U_m + rU_m < B_m + U_m + rU_{m+1} = aB_{m+1}. \end{aligned}$$

Since $a > 0$, this proves the claim; $B_1 > B_0$ follows from $b_1 > 0$. Finally both positive factors in $V_m = c_m D_m$ decrease, so $V$ does also, including its terminal zero. □

### 3.3 Reverse recurrence and product identities

The second recurrence is

$$V_m - aV_{m+1} = D_m + rD_{m+1} \qquad (1 \le m \le N). \tag{3.16}$$

First record

$$P_m b_m = \frac{S_m(Y_m - 1)}{2}, \qquad P_m c_m = \frac{S_m(Y_m + 1)}{2}, \tag{3.17}$$

$$P_m(c_m - b_m) = S_m, \qquad P_m b_m c_m = \frac{S_m^2}{2}. \tag{3.18}$$

For $m < N$ these give

$$P_m(c_m - 1)(1 + b_m) = \frac{S_m^2}{2} + S_m - P_m, \tag{3.19}$$

$$P_{m+1}(ac_{m+1} + r)(ab_{m+1} - r) = \frac{a^2 S_{m+1}^2}{2} - arS_{m+1} - r^2 P_{m+1}. \tag{3.20}$$

By (3.4) and (3.5), the second right side is

$$\frac{(S_m + h)^2}{2} - r(S_m + h) - P_m - \frac{1 - r^2}{2} = \frac{S_m^2}{2} + S_m - P_m,$$

because $h - r = 1$ and $h^2 - 2rh - (1 - r^2) = 0$. Thus

$$P_m(c_m - 1)(1 + b_m) = P_{m+1}(ac_{m+1} + r)(ab_{m+1} - r).$$

Divide by $U_m(1 + b_m)$ and use (3.10): $(c_m - 1)D_m = (ac_{m+1} + r)D_{m+1}$, the desired recurrence. At $m = N$, use $c_N = 1$ and the terminal zero tails. The needed product identities are

$$U_m D_m = P_m, \tag{3.21}$$

$$C_m := U_m V_m - B_m D_m = P_m(c_m - b_m) = S_m, \tag{3.22}$$

$$B_N D_N = P_N b_N = S_N^2/2. \tag{3.23}$$

The zero-index cases of the first two identities follow from the boundary conventions; the last uses $Y_N - 1 = S_N$.

# 4 The exact balanced history identity

Let data obey (1.8). No terminal estimate or sign assumption on function values is needed for the identity in this section. Define positive differences

$$\begin{aligned} u_i &= U_{i+1} - U_i, \quad s_i = B_{i+1} - B_i \quad (0 \le i < N), \\ v_j &= V_j - V_{j+1}, \quad t_j = D_j - D_{j+1} \quad (1 \le j \le N). \end{aligned} \tag{4.1}$$

For $0 \le i < j \le N$, set

$$\lambda_{ij} = u_i v_j, \qquad \lambda_{ji} = s_i t_j, \qquad \lambda_{ii} = 0. \tag{4.2}$$

Lemma 3.1 gives $\lambda_{ij} \ge 0$. Write

$$d_i = \sum_{j=0}^{N} (\lambda_{ij} - \lambda_{ji}), \qquad W = \sum_{i,j=0}^{N} \lambda_{ij} \mathcal{Q}_\theta(i, j), \qquad S = S_N. \tag{4.3}$$

## 4.1 Divergence and cuts

For each cut $0 \le p < N$, telescoping gives

$$\sum_{i \le p < j} \lambda_{ij} = \left( \sum_{i=0}^{p} u_i \right) \left( \sum_{j=p+1}^{N} v_j \right) = U_{p+1} V_{p+1},$$

$$\sum_{i \le p < j} \lambda_{ji} = \left( \sum_{i=0}^{p} s_i \right) \left( \sum_{j=p+1}^{N} t_j \right) = B_{p+1} D_{p+1}.$$

Internal edges cancel in the divergence sum, so $\sum_{i=0}^{p} d_i = C_{p+1} = S_{p+1}$. Together with $\sum_i d_i = 0$, this proves

$$d_i = h a^{-i-1} \quad (i < N), \qquad d_N = -S. \tag{4.4}$$

For coefficient bookkeeping let $T_{ij} = \mathbf{1}_{\{j<i\}}$, so $x_i = x_0 - h \sum_k T_{ik} g_k$. Bilinear expansion for arbitrary weights gives

$$W = 2(1 - \theta) \sum_i d_i f_i + 2 \sum_{p,q} M_{pq} \langle g_p, g_q \rangle - \sum_i I_i \|g_i\|^2, \tag{4.5}$$

where

$$I_i = \sum_j (\lambda_{ij} + \lambda_{ji}), \tag{4.6}$$

$$M_{pq} = \lambda_{pq} + h \sum_k (\lambda_{kp} + \theta \lambda_{pk})(T_{kq} - T_{pq}) - \frac{\theta h^2}{2} \sum_{k,l} \lambda_{kl} (T_{kp} - T_{lp})(T_{kq} - T_{lq}). \tag{4.7}$$

The squared gradient difference in (2.1) contributes $2\lambda_{pq}$ to the ordered inner-product coefficient and $-I_i$ to each diagonal. The linear gradient–position term contributes the middle sum in (4.7); the position square contributes its last sum. The common $x_0$ cancels from every $x_i - x_j$. This accounts for every term of (2.1).

### 4.2 Cancellation of every off-diagonal term

For $p < q$, the product $(T_{kp} - T_{lp})(T_{kq} - T_{lq})$ is one exactly when an edge joins an index at most $p$ to an index greater than $q$, and zero otherwise. Consequently

$$\sum_{k,l} \lambda_{kl} (T_{kp} - T_{lp})(T_{kq} - T_{lq}) = U_{p+1} V_{q+1} + B_{p+1} D_{q+1}. \tag{4.8}$$

The two linear sums in $M_{pq} + M_{qp}$ are

$$\sum_k (\lambda_{kp} + \theta \lambda_{pk})(T_{kq} - T_{pq}) = s_p D_{q+1} + \theta u_p V_{q+1},$$

$$\sum_k (\lambda_{kq} + \theta \lambda_{qk})(T_{kp} - T_{qp}) = -U_{p+1} v_q - \theta B_{p+1} t_q.$$

Indeed $T_{pq} = 0$, so the first sum only uses $k > q$; $T_{qp} = 1$, so the second only uses $k \le p$, with a minus sign. Insert the factorized weights and telescope.

The coefficient of $\langle g_p, g_q \rangle$ in $W$, divided by two, is therefore

$$\begin{aligned} M_{pq} + M_{qp} &= u_p v_q + s_p t_q + h(s_p D_{q+1} + \theta u_p V_{q+1} - U_{p+1} v_q - \theta B_{p+1} t_q) \\ &\quad - \theta h^2 (U_{p+1} V_{q+1} + B_{p+1} D_{q+1}) \\ &= (u_p - h U_{p+1})(v_q + \theta h V_{q+1}) + (s_p - \theta h B_{p+1})(t_q + h D_{q+1}). \end{aligned} \tag{4.9}$$

The recurrences, $1 - a = \theta h$, and $h = 1 + r$ give

$$s_p - \theta h B_{p+1} = a B_{p+1} - B_p = r U_{p+1} + U_p = h U_{p+1} - u_p,$$

$$v_q + \theta h V_{q+1} = V_q - a V_{q+1} = D_q + r D_{q+1} = t_q + h D_{q+1}.$$

Thus the two terms in (4.9) cancel.

### 4.3 Every diagonal coefficient

Let $e_i$ be the coefficient of $\|g_i\|^2$ in $W$. For $i < N$ abbreviate $U = U_{i+1}$, $B = B_{i+1}$, $D = D_{i+1}$, $V = V_{i+1}$, and retain $U_i, B_i$ for the preceding prefixes. The incident sum, divergence, and diagonal expansion are

$$I_i = u_i V + s_i D + B_i t_i + U_i v_i, \tag{4.10}$$

$$d_i = u_i V - s_i D + B_i t_i - U_i v_i, \tag{4.11}$$

$$e_i = -I_i + 2\theta h u_i V + 2h s_i D - \theta h^2 (UV + BD). \tag{4.12}$$

Terms $B_0t_0, U_0v_0$ at $i = 0$ mean zero. No reverse recurrence at index zero is assumed. Adding the last two identities yields

$$e_i + d_i = 2rs_iD - 2U_iv_i + \theta h[(2u_i - hU)V - hBD]. \tag{4.13}$$

The forward recurrence gives $s_i = \theta hB + rU + U_i$. For $i \geq 1$, the reverse recurrence gives $v_i = D_i + rD - \theta hV$; at $i = 0$ the substitution is valid after multiplication by $U_0 = 0$. Hence

$$2rs_iD - 2U_iv_i = 2r^2UD - 2U_iD_i + 2r\theta hBD + 2\theta hU_iV.$$

Use $u_i = U - U_i$, (3.21), and (3.22) in (4.13) to obtain

$$\begin{aligned} e_i + d_i &= 2(r^2P_{i+1} - P_i) + \theta h(1 - r)(UV - BD) \\ &= (1 - r)[h + \theta hS_{i+1}] = (1 - r)d_i. \end{aligned} \tag{4.14}$$

The last step uses $d_i - \theta hS_{i+1} = h$, which follows from $d_i = S_{i+1} - S_i$ and $aS_{i+1} = S_i + h$. Thus $e_i = -rd_i$ for every $i < N$. At $i = N$ all interval tails vanish:

$$e_N = -(U_NV_N + B_ND_N) = -S - 2B_ND_N = -S(S + 1). \tag{4.15}$$

This establishes the full history formula

$$\begin{aligned} W = 2(1 - \theta) &\left( \sum_{i=0}^{N-1} ha^{-i-1}f_i - Sf_N \right) \\ &- r \sum_{i=0}^{N-1} ha^{-i-1}\|g_i\|^2 - S(S + 1)\|g_N\|^2. \end{aligned} \tag{4.16}$$

## 4.4 The joint potential with surplus retained

The star quantity is

$$\mathcal{Q}_\theta(*, i) = -2(1 - \theta)f_i + 2\langle g_i, x_i\rangle - \|g_i\|^2 - \theta\|x_i\|^2. \tag{4.17}$$

Its weighted sum, added to (4.16), cancels all nonterminal function values. The remaining history term is

$$\sum_{i=0}^{N-1} ha^{-i-1} \left( 2\langle g_i, x_i\rangle - h\|g_i\|^2 - \theta\|x_i\|^2 \right).$$

The GD recurrence and $a = 1 - \theta h$ give exactly

$$ha^{-i-1} \left( 2\langle g_i, x_i\rangle - h\|g_i\|^2 - \theta\|x_i\|^2 \right) = a^{-i}\|x_i\|^2 - a^{-i-1}\|x_{i+1}\|^2. \tag{4.18}$$

Summing telescopically proves the central identity.

**Proposition 4.1** (Balanced joint identity)**.** *Under* (3.1)–(3.3) *and the GD recurrence, the explicitly constructed nonnegative weights satisfy*

$$\begin{aligned} \|x_0\|^2 = a^{-N}\|x_N\|^2 &+ 2(1 - \theta)Sf_N + S(S + 1)\|g_N\|^2 \\ &+ W + \sum_{i=0}^{N-1} ha^{-i-1}\mathcal{Q}_\theta(*, i). \end{aligned} \tag{4.19}$$

*Nonnegative history quantities imply $W \geq 0$. If the star quantities are also nonnegative, their displayed weighted sum is nonnegative.*

This proof makes no sign assertion about a terminal squared-norm coefficient in an alternative completion. Such a coefficient can be negative when $\theta < 0$. The joint identity (4.19) is the form needed here.

# 5 Recovering both branches at every step

Return to normalized actual functions with $0 \le \kappa < 1$, $N \ge 1$, and $0 < h < 2$. Put $a_0 = 1 - \kappa h$. All $\mathcal{Q}_\kappa$ are nonnegative by Lemma 2.1. Whenever $a_0 > 0$, write

$$S = h \sum_{j=1}^{N} a_0^{-j}, \qquad A = a_0^{-N} = 1 + \kappa S. \tag{5.1}$$

## 5.1 The progress branch

Assume $a_0 > 0$ and either $h \le 1$ or

$$(h-1)^N (1+S) \le 1. \tag{5.2}$$

Let $G = \sum_{j=1}^{N} a_0^{-j} > 0$ and

$$F(H) = (H-1)^N (1 + HG), \qquad H \ge 1.$$

This is continuous, strictly increasing on $(1, \infty)$, and $F(1) = 0$. Moreover

$$F(1 + a_0) = a_0^N (1 + (1 + a_0)G) \ge a_0^N + 1 + a_0 > 1, \tag{5.3}$$

since $G \ge a_0^{-N}$. There is a unique $H \in (1, 1 + a_0)$ with $F(H) = 1$. The branch condition implies $h \le H$: this is immediate for $h \le 1$, and otherwise follows from $F(h) \le 1$. Since $a_0 \le 1$, also $H < 2$.

Set $s = h/H \in (0, 1]$. The scaled data $(x_i, sg_i, sf_i)$ follow a step $H$ with parameter $s\kappa$, and

$$1 - (s\kappa)H = a_0, \qquad S' = H \sum_{j=1}^{N} a_0^{-j} = S/s, \qquad (H-1)^{-N} = 1 + S'.$$

Their history interpolation quantities are nonnegative by (2.3). Apply Proposition 4.1, retaining its star sum.

Define

$$E_i = \|g_i\|^2 - 2\kappa f_i. \tag{5.4}$$

The adjacent identity

$$\begin{aligned} h(1-\kappa)(E_i - E_{i+1}) = a_0 \mathcal{Q}_\kappa(i, i+1) + \mathcal{Q}_\kappa(i+1, i) \\ + (2-h)\|g_{i+1} - a_0 g_i\|^2 \end{aligned} \tag{5.5}$$

shows $E_i \ge E_{i+1}$. To verify it, write $g = g_i$, $k = g_{i+1}$ and substitute $x_i - x_{i+1} = hg$ in (2.1):

$$\begin{aligned} \mathcal{Q}_\kappa(i, i+1) &= 2(1-\kappa)(f_i - f_{i+1}) - 2h\langle k, g\rangle + 2\kappa h\|g\|^2 - \|g - k\|^2 - \kappa h^2 \|g\|^2, \\ \mathcal{Q}_\kappa(i+1, i) &= -2(1-\kappa)(f_i - f_{i+1}) + 2h\|g\|^2 - 2\kappa h\langle k, g\rangle - \|g - k\|^2 - \kappa h^2 \|g\|^2. \end{aligned}$$

In $a_0$ times the first plus the second plus $(2-h)\|k - a_0 g\|^2$, the coefficients of $f_i - f_{i+1}$, $\|g\|^2$, $\|k\|^2$, and $\langle g, k\rangle$ are respectively $-2\kappa h(1-\kappa)$, $h(1-\kappa)$, $-h(1-\kappa)$, and zero. This is (5.5). Every term on its right is nonnegative, so in particular $E_i \ge E_N$.

Substitution of star data gives the exact identity

$$\mathcal{Q}^{(s)}_{s\kappa}(*, i) = s\mathcal{Q}_\kappa(*, i) + s(1-s)E_i. \tag{5.6}$$

Since $Hs = h$, the retained star remainder satisfies

$$\sum_{i<N} Ha_0^{-i-1}\mathcal{Q}_{s\kappa}^{(s)}(*,i) = \sum_{i<N} ha_0^{-i-1}\mathcal{Q}_\kappa(*,i) + (1-s)\sum_{i<N} ha_0^{-i-1}E_i$$
$$\geq (1-s)SE_N. \tag{5.7}$$

The scaled certificate's terminal part is

$$A\|x_N\|^2 + 2(1-s\kappa)Sf_N + S(S+s)\|g_N\|^2.$$

Add (5.7) and expand $E_N$. The result is exactly

$$A\|x_N\|^2 + 2(1-\kappa)Sf_N + S(S+1)\|g_N\|^2 \leq \|x_0\|^2. \tag{5.8}$$

There is no limit argument, including at $\kappa = 0$.

Using (2.4), $A > 0$, $S > 0$, and $A = 1 + \kappa S$, the left side is at least

$$\big(A + (1-\kappa)S + S(S+1)\big)\|g_N\|^2 = (1+S)^2\|g_N\|^2.$$

Hence

$$\|g_N\| \leq \frac{\|x_0\|}{1+S}. \tag{5.9}$$

For $\kappa = 0$, $S = Nh$. For $\kappa > 0$, $A = 1 + \kappa S$ gives

$$\frac{1}{1+S} = \frac{\kappa a_0^N}{1-(1-\kappa)a_0^N} = p_N(\kappa, h).$$

Thus this is the claimed progress estimate.

### 5.2 The oscillation branch and negative auxiliary curvature

Assume $1 < h < 2$, let $r = h - 1$, and put $T = r^{-N} - 1 > 0$. The function

$$\mathcal{S}(b) = h\sum_{j=1}^{N} b^{-j}, \qquad b > 0,$$

is continuous, strictly decreasing, and tends to zero as $b \to \infty$. At $b = r$,

$$\mathcal{S}(r) = \frac{h}{1-r}(r^{-N} - 1) > T. \tag{5.10}$$

There is therefore a unique $b > r$ with $\mathcal{S}(b) = T$.

Suppose either $a_0 \leq 0$, or

$$a_0 > 0, \qquad \mathcal{S}(a_0) \geq r^{-N} - 1. \tag{5.11}$$

Then $b \geq a_0$: the first case is immediate and the second follows from strict decrease of $\mathcal{S}$. Set

$$\nu = \frac{1-b}{h} \leq \kappa.$$

Also $\nu < 1$ because $b > 0$ and $h > 1$. By (2.2), all history and star quantities $\mathcal{Q}_\nu$ are nonnegative. The balanced identity with $a = b$, $S = T$ therefore gives

$$b^{-N}\|x_N\|^2 + 2(1-\nu)Sf_N + S(S+1)\|g_N\|^2 \leq \|x_0\|^2. \tag{5.12}$$

Here $b^{-N} = 1 + \nu S > 0$ even if $\nu < 0$. The terminal bounds belong to the actual convex function, and all coefficients multiplying them are nonnegative. The left side of (5.12) is thus at least

$$(b^{-N} + (1-\nu)S + S(S+1))\|g_N\|^2 = (1+S)^2\|g_N\|^2 = r^{-2N}\|g_N\|^2.$$

It follows that

$$\|g_N\| \le (h-1)^N \|x_0\|. \tag{5.13}$$

The reductions cover every $0 < h < 2$. For $h \le 1$, $a_0 > 0$ and the progress condition applies. For $h > 1$, $a_0 > 0$, compare $(h-1)^N(1 + \mathcal{S}(a_0))$ with one: values at most one give the progress branch and values at least one give (5.11). If $a_0 \le 0$, the oscillation branch applies directly. At equality either proof is valid.

# 6 Endpoints and equal curvature

For $N = 0$ or $h = 0$, Lipschitz continuity and $\nabla f(x_*) = 0$ give $\|\nabla f(x_N)\| \le LR$. The factor (1.4) equals one.

For $h = 2$, normalized convex interpolation gives $\|g_i\|^2 \le \langle g_i, x_i\rangle$ by adding the two star inequalities. More generally, when $0 \le h \le 2$,

$$\begin{aligned}\|x_i - hg_i\|^2 &= \|x_i\|^2 - 2h\langle g_i, x_i\rangle + h^2\|g_i\|^2 \\ &\le \|x_i\|^2 - h(2-h)\|g_i\|^2 \le \|x_i\|^2.\end{aligned}$$

Thus $\|g_N\| \le \|x_N\| \le \|x_0\|$ at $h = 2$. Here $|1-h|^N = 1$ and $p_N(\kappa, h) \le 1$, so the exact maximum factor is one.

When $\mu = L$, adding the two strong first-order inequalities yields

$$L\|x - x_*\|^2 \le \langle \nabla f(x), x - x_*\rangle, \qquad \|\nabla f(x)\| \le L\|x - x_*\|.$$

Consequently

$$\|\nabla f(x) - L(x - x_*)\|^2 = \|\nabla f(x)\|^2 - 2L\langle \nabla f(x), x - x_*\rangle + L^2\|x - x_*\|^2 \le 0.$$

The centered gradient is exactly $L(x - x_*)$. Therefore

$$x_i - x_* = (1-h)^i(x_0 - x_*), \qquad \|\nabla f(x_N)\| = L|1-h|^N R.$$

Since $p_N(1, h) = (1-h)^N$, this is exactly (1.4). Together with the two reductions and normalization, this proves the bound in Theorem 1.1.

# 7 Sharpness of both branches

The two types of examples below already appear in the original conjecture [1]. Their trajectories are checked below to make the exactness of the bound explicit.

The normalized quadratic

$$f_{\text{quad}}(x) = x^2/2, \qquad x_0 = 1, \tag{7.1}$$

belongs to every class $0 \le \kappa \le 1$ and has $x_i = (1-h)^i$. Its final gradient norm is $|1-h|^N$, attaining the maximum whenever the oscillation branch dominates.

Suppose $N \geq 1$, $h > 0$, $0 \leq \kappa < 1$, and the progress branch dominates. Then $a = 1 - \kappa h > 0$. Indeed, if $a \leq 0$, then $h > 1$, $|a| = \kappa h - 1 < h - 1$, and $p_N(\kappa, h) \leq |a|^N < (h-1)^N$. The first inequality is immediate for $a^N \leq 0$; otherwise $\kappa/[1-(1-\kappa)a^N] \leq 1$.

Define $S_m = h\sum_{j=1}^{m} a^{-j}$ and $\tau = 1/(1+S_N)$. The strongly convex Huber spline is

$$f_\tau(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq \tau, \\ \frac{\kappa}{2}x^2 + (1-\kappa)\tau|x| - \frac{1-\kappa}{2}\tau^2, & |x| \geq \tau. \end{cases} \tag{7.2}$$

Its derivative is continuous at the joining points and satisfies

$$f'_\tau(x) = \kappa x + (1-\kappa)\,\mathrm{clip}_{[-\tau,\tau]}(x). \tag{7.3}$$

The clipped function is nondecreasing and 1-Lipschitz. Thus the derivative is 1-Lipschitz and $f_\tau(x) - \kappa x^2/2$ is convex. The function is nonnegative and zero is a minimizer.

For $0 \leq i \leq N$, set

$$x_i = \tau(1 + S_{N-i}). \tag{7.4}$$

Then $x_0 = 1$, $x_N = \tau$, and $x_i > \tau$ for $i < N$. Using $1 + \kappa S_m = a^{-m}$,

$$f'_\tau(x_i) = \kappa x_i + (1-\kappa)\tau = \tau a^{-(N-i)} \qquad (0 \leq i \leq N).$$

The increment formula for $S_m$ gives

$$x_i - x_{i+1} = \tau(S_{N-i} - S_{N-i-1}) = h\tau a^{-(N-i)} = h f'_\tau(x_i).$$

Thus (7.4) is the actual GD trajectory and $f'_\tau(x_N) = \tau = p_N(\kappa, h)$, including $\kappa = 0$.

The cases $N = 0$, $h = 0$, or $\kappa = 1$ are attained by the quadratic. For general $L$ and $R > 0$, rescale a normalized witness by

$$\widetilde{f}(x) = LR^2 f(x/R), \qquad \widetilde{x}_0 = R.$$

Its smoothness constant is $L$, its strong-convexity parameter is $\mu = L\kappa$, and its final gradient is $LR$ times the normalized one. This proves the sharpness assertion of Theorem 1.1.

# 8 Formal verification scope

**Complete upper bound and normalized sharpness checked in Lean.** The checked upper bound is an actual-function theorem. Its hypotheses are ordinary strong convexity, actual gradients, gradient Lipschitz continuity, a global minimizer, and the update (1.3). The explicit weights are constructed within the proof. There is no horizon or dimension cutoff, and no interpolation or certificate condition remains as a hypothesis of the public theorem.

The development uses Lean 4.19.0 and the following Mathlib revision:

`c44e0c8ee63ca166450922a373c7409c5d26b00b`.

The checked modular decomposition is:

| Modules | Mathematical role |
|---|---|
| `AnalyticBase`, `Analytic` | Actual smooth descent, strong first-order support, and interpolation. |
| `Core`, `Scaling` | Curvature change, scaling, and adjacent potential monotonicity. |
| `ScalarBase`, `Scalars` | Geometric sums, explicit factors, every factor sign, and both recurrences. |
| `Cuts`, `CutsFin`, `History`, `Coefficients` | Finite cut identities and all mixed and diagonal coefficients. |
| `Balanced` | The actual balanced joint certificate with constructed nonnegative weights. |
| `Balance`, `Reduction`, `AnalyticFinal` | Intermediate-value reductions, the full parameter range, and transfer to the original analytic theorem. |
| `Sharpness`, `Huber`, `Attainment` | Algebra of attaining trajectories, actual one-dimensional functions, and selection of the larger rate branch. |

The public upper-bound declarations are

```
THGStrong.taylor_hendrickx_glineur_conjecture3
THGStrong.taylor_hendrickx_glineur_conjecture3_with_radius.
```

They include arbitrary $L > 0$ and all $0 \le \mu \le L$, $0 \le h \le 2$, and $N \ge 0$. The second allows an initial-distance upper bound $R$. The declaration

```
THGStrong.Sharpness.attain_rate
```

constructs actual one-dimensional strongly convex functions, their gradients, minimizers, and GD trajectories attaining the rate for every normalized $0 \le \kappa \le 1$, $0 \le h \le 2$, and $N \ge 0$. Sharpness is formalized at $L = R = 1$; the rescaling for general $L, R$ is given explicitly in the preceding section. The upper bound is already formalized without that normalization.

The axiom reports of the three public declarations contain only

`propext`, `Classical.choice`, `Quot.sound`.

These are the standard foundational principles used by Lean and Mathlib. There is no `sorryAx` or added mathematical postulate. The distribution includes all proof sources, dependency pins and a portable verifier. The verifier rebuilds all 20 local modules from a source-only snapshot, excluding earlier project objects from its search path; every compiler process uses `-trust=0`, and compiler warnings cause failure. It then runs a separate theorem-type and axiom audit and records source hashes and compiler exits. The accompanying verification record identifies the checked source hashes, dependency revision, compiler results, and the three axiom reports.

**Contribution assessment.** The upper bound and attaining examples establish the exact finite-horizon formula over the full stated parameter range. The intended contribution is positive strong convexity; the zero-curvature curve and the sharp examples are credited to prior work. The reusable part is the balanced joint certificate with an unrestricted lower curvature parameter and its use in both reductions. The theorem gives a sharp worst-case guarantee for a standard first-order method. The bibliographic comparison is bounded by the cited sources and does not establish worldwide priority.

## 9 Research process and AI disclosure

This is fully autonomous agentic research. Automated systems produced the research, the proofs, the Lean formalization and the preparation of the manuscript. Dev D. Goyal is responsible for publication. The work has not been externally peer reviewed, and no expert validation is claimed.

Selected analytic, finite-sum and clipped-gradient infrastructure in the Lean development is adapted from the MIT-licensed project of Jones [6], whose notice accompanies the source distribution. On 1 November 2026, Opus 5.5 by Anthropic used with Claude Code revised the presentation of this manuscript (front matter, abstract, introduction and this section) without changing its mathematics or its formal statements.

The formal scope is stated in Section 8, and the limits of the literature comparison in Section 1.1.

## 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. Conjecture 3 and Section 4.1.3. Primary manuscript: `https://optimization-online.org/wp-content/uploads/2015/03/4814.pdf`.

[2] A. B. Taylor, J. M. Hendrickx, and F. Glineur. Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. *Journal of Optimization Theory and Applications* **178** (2018), 455–476. doi:10.1007/s10957-018-1298-1. Section 4 and Tables 1–2; latest arXiv revision consulted: `https://arxiv.org/abs/1705.04398`, v4, 29 February 2020.

[3] J. Kim. A proof of the exact convergence rate of gradient descent. `https://arxiv.org/abs/2412.04427`, v2, 26 March 2025. Theorems 1.3 and 1.5 concern different initial/final measures.

[4] T. Rotaru, F. Glineur, and P. Patrinos. Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over nonconvex and convex functions. *Mathematical Programming*, published online 26 May 2026. doi:10.1007/s10107-025-02313-1. Preprint: `https://arxiv.org/abs/2406.17506`.

[5] S. R. Chikhi. Last-iterate performance of gradient descent and relaxed proximal point via $s$-composability. `https://arxiv.org/abs/2609.13765`, v2, 20 September 2026; first version 12 September 2026.

[6] J. D. Jones. Sharp last-iterate bounds and extremal trajectories for constant-step gradient descent. *thg-gradient-curve*, 2026. `https://github.com/JD-Jones-ASES/thg-gradient-curve`. Convex full-curve result dated 18 October 2026; mathematical proof, verification scope, and MIT-licensed Lean sources consulted 28–29 October 2026. Sources pinned at revision `134788107eb28be0459555383a482a350bc953a4`. Positive strong convexity is explicitly outside the repository's established claim.

[7] F. Glineur. Performance estimation of first-order methods: extensions and recent results. LNMB Conference on the Mathematics of Operations Research, 17 January 2023, lecture slides. `https://www.lnmb.nl/conferences/2023/invitedspeakers/Glineur-LNMB2023-part2-final.pdf`.