NO UNIVERSAL REMAINDER RATE FOR CHAMBOLLE–DOSSAL ACCELERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Chambolle–Dossal acceleration guarantees $F(x_n) - F^* = o(n^{-2})$ for every fixed smooth convex loss with a minimizer. We show that this qualitative improvement admits no universal quantitative rate. For every damping parameter $\alpha > 3$ and positive nondecreasing gain $G(n) \to \infty$, we construct a fixed one-dimensional smooth convex loss whose exact CD orbit satisfies

$$\sup_{n \geq 1} n^2 G(n)\big(F(x_n) - F^*\big) = \infty.$$

Thus no divergent gain improves the n^{-2} scale for all fixed losses, even with instance-dependent constants. The construction prescribes queried gradients and realizes infinitely many slow blocks within one smooth convex objective. Under local p-power growth with $p > 2$ and sufficiently strong damping, we also construct a fixed loss whose exact CD orbit satisfies $F(x_n) - F^* \sim Dn^{-2p/(p-2)}$, $D > 0$, establishing the sharpness of the known convergence rate. Both main results are formally verified in Lean 4.

1 INTRODUCTION

The accelerated $O(n^{-2})$ function-value guarantee has a strict fixed-instance refinement: Chambolle–Dossal (CD) acceleration achieves $o(n^{-2})$ on every fixed smooth convex objective with a minimizer (Chambolle & Dossal, 2015; Attouch & Peypouquet, 2016). This means that the normalized residual $n^2(F(x_n) - F^*)$ tends to zero. Does this vanishing factor admit any common quantitative decay rate, with constants allowed to depend on the objective and initialization? We show that it does not, even for one-dimensional coercive losses with a prescribed smoothness budget and initial distance.

Consider CD acceleration with damping $\alpha > 3$. For a convex loss F with globally L-Lipschitz gradient and a fixed step $0 < s < L^{-1}$, its update is

$$\begin{aligned} &y_n = x_n + \beta_n(x_n - x_{n-1}), \quad x_{n+1} = y_n - s\nabla F(y_n), \\ &\beta_n = \frac{n-1}{n+\alpha-1}, \qquad\qquad n \geq 1. \end{aligned} \tag{1}$$

The normalized residual $q_n = n^2(F(x_n) - F^*)$ tends to zero on every fixed loss. A universal remainder rate would give $q_n = O(1/G(n))$ for some common positive nondecreasing $G(n) \to \infty$, with instance-dependent constants. For each such G, we construct one fixed loss with $\sup_n G(n)q_n = \infty$. Thus no common gain exists, however slowly it grows.

We prescribe nonincreasing queried gradients and invert the CD recurrence to recover an exact orbit, using separated constant-gradient blocks to defeat the proposed gain along a subsequence. The inverse construction and choice of scales have precedents in scalar gradient flow (Siegel & Wojtowytsch, 2023, Lemma 2.7 and Example 2.9). We impose weighted summability to ensure convergence and construct the loss by interpolating its derivative, with bounded slopes enforcing global smoothness (see also Taylor et al., 2017a, for the general finite-data interpolation framework). The resulting loss is one-dimensional, even, coercive, and uniquely minimized at zero; spatial rescaling sets any prescribed positive initial distance without changing the gradient Lipschitz constant.

Using the same inverse-orbit construction with gradients that decay as a power, we also obtain fixed losses whose exact CD orbits satisfy $F(x_n) - F^* \sim Dn^{-2p/(p-2)}$, $D > 0$. These instances establish

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the sharpness of the known $O(n^{-2p/(p-2)})$ rate under local p-power growth with suitable flatness or sufficiently strong damping.

We establish the following two sharpness results for CD acceleration. Both conclusions are formally verified in Lean 4 (de Moura & Ullrich, 2021); see the supplementary materials in Appendix G.

- **No universal remainder rate on fixed losses.** For every positive nondecreasing $G(n) \to \infty$, we construct one fixed one-dimensional smooth convex loss whose exact CD orbit satisfies
$$F(x_n) - F^* = o(n^{-2}), \qquad \sup_{n \geq 1} n^2 G(n)\big(F(x_n) - F^*\big) = \infty.$$
The smoothness constant and initial distance can be prescribed, and the construction extends to every fixed finite dimension. To our knowledge, this is the first no-universal-gain result for exact discrete CD iterates on a single fixed scalar loss (Theorem 1).
- **Sharp rates under local geometry.** For every admissible parameter choice in Theorem 2, we construct fixed smooth convex losses with local p-power growth whose exact CD orbits satisfy
$$F(x_n) - F^* \sim D n^{-2p/(p-2)}, \qquad D > 0.$$
These instances attain the established upper rates (Apidopoulos et al., 2021; Aujol et al., 2024). To our knowledge, they are the first exact discrete constructions proving sharpness throughout both stated geometry regimes, including endpoint flatness.

2 Related work

Our results connect four strands of work: accelerated convergence guarantees, finite-horizon lower bounds and interpolation, slow convergence on fixed objectives, and rates under local geometry. We distinguish horizon-dependent worst cases from one fixed loss, and continuous trajectories from exact discrete orbits.

2.1 Accelerated convergence guarantees

Nesterov's accelerated methods achieve the classical $O(n^{-2})$ bound (Nesterov, 2013), which FISTA extends to composite convex minimization (Beck & Teboulle, 2009). Chambolle & Dossal (2015) modify the momentum schedule to establish convergence of the iterates; Attouch & Peypouquet (2016) prove the strict $o(n^{-2})$ function-value rate for $\alpha > 3$. Weighted residual summability, $\sum_n n(F(x_n) - F^*) < \infty$, is also established for this discrete schedule (Chambolle & Dossal, 2015; Su et al., 2016).

Neither little-o convergence nor this summability bound supplies a common eventual pointwise remainder rate. Theorem 1 shows that for every positive nondecreasing $G(n) \to \infty$, there is one scalar loss whose error is not $O(1/(n^2 G(n)))$.

2.2 Finite-horizon lower bounds and interpolation

Classical oracle lower bounds describe how large the optimization error can remain after a prescribed number of iterations (Nesterov, 2013). Arjevani & Shamir (2016) obtain lower bounds valid in every fixed dimension for a class of first-order methods whose linear update coefficients depend only on the iteration number and supplied smoothness and strong convexity parameters. In these results, the difficult objective may vary with the iteration budget. Our question requires one fixed objective for which the proposed rate fails along an infinite sequence of CD iterates.

Performance estimation formulates the largest possible error after a fixed number of iterations as an optimization problem over objective functions and the values and gradients observed by the algorithm (Drori & Teboulle, 2014). Smooth convex interpolation determines whether prescribed function values and gradients at finitely many points can come from a convex function with a specified gradient Lipschitz bound. It also constructs such functions, enabling exact worst-case analyses (Taylor et al., 2017a;b). We use a related construction, prescribing gradients at infinitely many query points and building one smooth convex function whose CD iterates produce these queries. The construction must satisfy the CD update equations and produce errors that violate the proposed rate at arbitrarily large iteration indices.

2.3 SLOW CONVERGENCE ON FIXED OBJECTIVES

The closest antecedent is Siegel & Wojtowytsch (2023). Their scalar inverse construction and sparse-scale example (Siegel & Wojtowytsch, 2023, Lemma 2.7 and Example 2.9) show that gradient flow admits no common divergent gain over its $1/t$ scale, even in one dimension. Their slow inertial construction with an attained minimum uses an infinite-dimensional quadratic objective (Siegel & Wojtowytsch, 2023, Lemma 4.2). These results already establish the principles of prescribing a trajectory and placing slow portions at widely separated scales.

These constructions do not supply a scalar objective for the exact CD recurrence. For damping $3/t$, the ODE approximation of Su et al. (2016, Theorem 2) takes the step size to zero on bounded time intervals; it does not transfer infinite-time lower asymptotics to a fixed step size. For any proposed improved rate, we construct a one-dimensional smooth convex objective by prescribing the gradients observed by CD and recovering its iterates from the update equations. On this fixed objective, the iterates converge to the unique minimizer, but the objective errors fail to satisfy that rate. See Sections 4.1 and 4.2 and Appendix B for the construction and proofs.

2.4 LOCAL GEOMETRY AND MATCHING POWER RATES

Flatness and growth conditions quantify how the objective behaves near its minimizers. For continuous inertial dynamics, power potentials with $p > 2$ already give matching trajectories at the $t^{-2p/(p-2)}$ scale (Attouch et al., 2018), and Aujol et al. (2019) develop sharp geometry-dependent ODE rates. In discrete time, Apidopoulos et al. (2021) establish $O(n^{-2p/(p-2)})$ under local p-power growth, suitable flatness, uniqueness of the minimizer, and a damping threshold. They also present numerical evidence for sharpness using power potentials, while leaving discrete optimality unproved. Aujol et al. (2024, Theorem 1) obtain the same exponent for coercive objectives under local growth alone when $\alpha > 5 + 8/(p-2)$, without requiring flatness or a unique minimizer.

We complement these upper bounds by constructing fixed smooth convex objectives whose CD iterates satisfy $F(x_n) - F^* \sim Dn^{-2p/(p-2)}$, $D > 0$. Thus the rates are sharp under both sets of assumptions in Theorem 2, including the case where the flatness and growth exponents coincide. See Section 4.3 and Appendices C–D for the construction and proofs.

3 SETTING AND MAIN RESULTS

3.1 OBJECTIVES AND THE CD ITERATION

We consider convex differentiable losses $F : \mathbb{R}^d \to \mathbb{R}$, in a fixed finite dimension $d \geq 1$, that attain a minimum F^* and have a globally L-Lipschitz gradient:

$$\|\nabla F(x) - \nabla F(z)\| \leq L\|x - z\| \qquad \text{for all } x, z \in \mathbb{R}^d.$$

Here $L > 0$, and norms and inner products are Euclidean. We use $C^{1,1}$ for differentiability with a globally Lipschitz gradient, and F' for the derivative in dimension one. A loss is coercive if $F(x) \to \infty$ as $\|x\| \to \infty$.

For damping $\alpha > 3$, a fixed step $0 < s < L^{-1}$, and initialization $x_0, x_1 \in \mathbb{R}^d$, the Chambolle–Dossal (CD) iteration is

$$\begin{aligned} &y_n = x_n + \beta_n(x_n - x_{n-1}), \quad x_{n+1} = y_n - s\nabla F(y_n), \\ &\beta_n = \frac{n-1}{n+\alpha-1}, \qquad\qquad n \geq 1. \end{aligned} \tag{2}$$

Here x_n is the iterate, β_n the momentum coefficient, and y_n the extrapolated query point. Write $g_n := \nabla F(y_n)$ for the queried gradient. The objective error is $r_n := F(x_n) - F^*$ and the normalized residual is $q_n := n^2 r_n$. Since $\beta_1 = 0$, x_0 does not affect later iterates; our constructions use $x_0 = x_1$. All convergence rates are as $n \to \infty$.

3.2 NO UNIVERSAL REMAINDER RATE

A gain is a positive nondecreasing sequence $G(n) \to \infty$. A universal remainder rate would give $q_n = O(1/G(n))$ on every fixed loss satisfying the assumptions above, for every CD initialization.

Both the constant and the starting index may depend on the loss and initialization. The following theorem excludes such a rate.

Theorem 1 (No universal gain). *Fix $\alpha > 3$, $L > 0$, $0 < s < L^{-1}$, an initial distance $R > 0$, and a gain G. There is a fixed even, coercive convex loss $F : \mathbb{R} \to \mathbb{R}$ with an L-Lipschitz derivative, uniquely minimized at zero with $F(0) = 0$, whose exact CD orbit from $x_0 = x_1 = R$ satisfies*

$$\sup_{n \geq 1} n^2 G(n)\big(F(x_n) - F^*\big) = \infty. \tag{3}$$

The construction extends to every fixed finite dimension, preserving smoothness, coercivity, the unique minimizer, initial distance, and objective errors.

For each proposed G, one loss remains fixed along the entire orbit. Its normalized residual satisfies $q_n \to 0$ by the established little-o guarantee, while $G(n)q_n$ is unbounded along a subsequence. Thus even an eventual bound with an instance-dependent constant fails. The loss may depend on G: taking $G(n) = n^{\varepsilon}$, $\varepsilon > 0$, excludes any common improved polynomial exponent, and $G(n) = \log(n+1)$ excludes a common logarithmic gain. The theorem also covers arbitrarily slower divergent gains.

3.3 Local geometry and sharp power rates

Suppose first that F has a unique minimizer x^*. For $\beta > 2$, the local flatness condition $H(\beta)$ means that, for some $\eta > 0$,

$$H(\beta): \qquad F(x) - F^* \leq \frac{1}{\beta}\langle \nabla F(x), x - x^* \rangle \qquad \text{if } \|x - x^*\| \leq \eta. \tag{4}$$

For $p > 2$, the local growth condition $L(p)$ means that, for some $K, \varepsilon > 0$,

$$L(p): \qquad F(x) - F^* \geq K\|x - x^*\|^p \qquad \text{if } \|x - x^*\| \leq \varepsilon. \tag{5}$$

Both inequalities apply to every x in the indicated neighborhood. Here β is a flatness exponent, distinct from the momentum coefficient β_n.

To allow multiple minimizers, let X^* denote the minimizers of F and $\operatorname{dist}(x, X^*)$ the Euclidean distance from x to the nearest minimizer. For $p > 2$, the local growth condition $L_{\mathrm{set}}(p)$ means that, for some $K, \varepsilon > 0$,

$$L_{\mathrm{set}}(p): \qquad F(x) - F^* \geq K \operatorname{dist}(x, X^*)^p \qquad \text{if } \operatorname{dist}(x, X^*) \leq \varepsilon. \tag{6}$$

The constants apply throughout this neighborhood of the minimizers. When the minimizer is unique, $L_{\mathrm{set}}(p)$ reduces to $L(p)$.

Theorem 2 (Sharp power rates under local geometry). *Fix $\alpha > 3$, $L > 0$, $0 < s < L^{-1}$, and $p > 2$. Let $F : \mathbb{R}^d \to \mathbb{R}$ be convex, attain its minimum, and have an L-Lipschitz gradient. For every CD initialization,*

$$F(x_n) - F^* = O\big(n^{-2p/(p-2)}\big) \tag{7}$$

holds under either of the following conditions:

- *(i) F has a unique minimizer and satisfies $H(\beta)$ and $L(p)$, with $2 < \beta \leq p$ and $\alpha \geq \frac{\beta+2}{\beta-2}$.*
- *(ii) F satisfies $L_{\mathrm{set}}(p)$ and $\alpha > 5 + 8/(p-2)$.*

In each regime, for every admissible parameter choice and every fixed finite dimension, there is a coercive loss satisfying these assumptions and having a unique minimizer, whose exact CD orbit for some initialization satisfies

$$F(x_n) - F^* \sim D n^{-2p/(p-2)}, \qquad D > 0. \tag{8}$$

Thus the upper rate is attained with a positive asymptotic constant D, and no common divergent gain improves it. The constants and onset of the upper bound may depend on the loss and initialization.

The upper bounds in Theorem 2 build on established results. Apidopoulos et al. (2021) prove the rate in part (i) under flatness and growth conditions. Aujol et al. (2024) obtain the growth-only

rate used in part (ii) with an additional coercivity assumption. We extend their upper bound to the finite-dimensional setting of part (ii) without requiring coercivity.

Our contribution is a matching construction for every parameter choice allowed by the theorem: one fixed smooth convex loss has an exact CD orbit with $F(x_n) - F^* \sim Dn^{-2p/(p-2)}$, $D > 0$. These instances prove that the upper rate is sharp in both regimes.

As $p \to \infty$, the exponent $2p/(p-2)$ approaches 2. The case $p = 2$ lies outside this power law.

4 Constructing the fixed losses

This section constructs the fixed losses in Theorem 1 and the sharpness instances in Theorem 2. The construction proceeds in three steps: prescribe a sequence of queried gradients, recover the corresponding CD orbit, and interpolate and integrate the prescribed derivative values to obtain a smooth convex loss. Section 4.1 develops this common procedure and explains how it gives the required properties. Section 4.2 uses separated constant-gradient blocks to obtain the no-universal-gain instance, while Section 4.3 uses a shifted power sequence to obtain a loss with the exact power-rate asymptotic.

4.1 Recovering the orbit and the loss

We reconstruct a fixed loss from a prescribed gradient sequence so that the CD algorithm follows the designed orbit exactly. Choose $(g_n)_{n\geq 1}$ with

$$g_n > 0, \qquad g_{n+1} \leq g_n, \qquad \sum_{n=1}^{\infty} n g_n < \infty.$$

Write $v_n = x_{n-1} - x_n$ for the displacement towards zero. Recover the displacements, iterates, and query points by

$$\begin{aligned} v_{n+1} &= \beta_n v_n + s g_n, \quad v_1 = 0, \\ x_n &= \sum_{j=n+1}^{\infty} v_j, \qquad x_0 = x_1, \qquad y_n = x_n - \beta_n v_n. \end{aligned} \tag{9}$$

These sequences satisfy

$$v_{n+1} > 0, \qquad \sum_{n=1}^{\infty} v_{n+1} < \infty, \qquad x_n \downarrow 0, \qquad y_n \downarrow 0,$$

with $x_{n+1} = x_n - v_{n+1}$ and $y_n - s g_n = x_{n+1}$. Thus the total displacement is finite, and the query points form an ordered mesh accumulating at zero.

Join consecutive points (y_n, g_n) by line segments to define the derivative on the positive half-axis, including its extension at zero and beyond the first query:

$$\psi(x) = \begin{cases} 0, & x = 0, \\ g_{n+1} + \dfrac{g_n - g_{n+1}}{y_n - y_{n+1}}(x - y_{n+1}), & y_{n+1} \leq x \leq y_n, \quad n \geq 1, \\ g_1, & x \geq y_1. \end{cases} \tag{10}$$

Finally, define the loss on the whole real line by

$$F(x) = \int_0^{|x|} \psi(u)\, du. \tag{11}$$

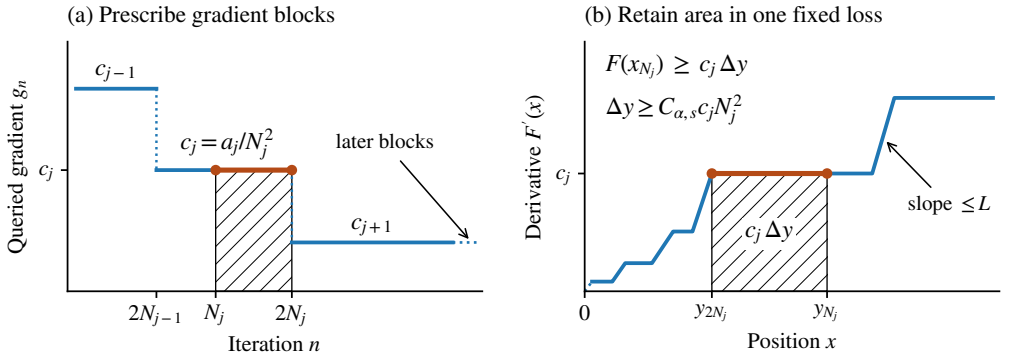
The resulting loss is even, coercive, convex, and $C^{1,1}$, with unique minimizer zero. Its derivative is L-Lipschitz whenever

$$0 \leq \frac{g_n - g_{n+1}}{y_n - y_{n+1}} \leq L \qquad (n \geq 1).$$

Each gradient schedule below enforces this bound. Moreover, $F'(y_n) = g_n$ and $x_{n+1} = y_n - sF'(y_n)$, so the prescribed points form an exact CD orbit of this one fixed loss. All these properties are proved in Lemma 3 in Appendix B.

4.2 THEOREM 1: MAKING SLOW EPISODES COMPATIBLE WITH CONVERGENCE

For a prescribed gain G, we construct increasingly long gradient plateaus with decreasing heights. Lowering the heights keeps the total displacement finite, while lengthening the plateaus retains enough error to violate the proposed rate $1/(n^2 G(n))$ at arbitrarily late iterates. Figure 1 summarizes how these two requirements are met within one gradient sequence.



(c) Amplitudes ensure convergence

$a_j = 2^{-j}, \quad \sum_j a_j < \infty$

$\sum_{k=2N_{j-1}+1}^{2N_j} k g_k \leq 3a_j$

$\sum_k k g_k < \infty \implies x_n, y_n \downarrow 0$

(d) Late blocks defeat the gain

$N_j^2 F(x_{N_j}) \geq C_{\alpha,s} a_j^2$

Choose $N_j > 2N_{j-1}$ with $G(N_j) \geq j 4^j$

$N_j^2 G(N_j) F(x_{N_j}) \geq C_{\alpha,s} j \to \infty$

Figure 1: Block construction; axes are not to scale. Orange marks the same constant-gradient interval in time (a) and space (b), retaining area $c_j \Delta y \leq F(x_{N_j})$. Summable amplitudes ensure convergence (c); late blocks amplify the residual (d). $C_{\alpha,s} > 0$ is the block constant.

First partition the positive iteration indices into consecutive blocks. Set $a_j = 2^{-j}$ and $N_0 = 0$, and choose integers $N_j \geq 8$ with $N_j > 2N_{j-1}$; their dependence on G will be specified below. On each block, prescribe the constant queried gradient

$$g_k = c_j = \frac{a_j}{N_j^2} \qquad \text{for } 2N_{j-1} < k \leq 2N_j. \tag{12}$$

These blocks cover all positive iteration indices, and the heights c_j decrease from block to block. Thus the gradient sequence is positive and nonincreasing. Each plateau consists of equal values of $g_k = F'(y_k)$; the loss values continue to decrease along the orbit.

To preserve convergence as the blocks grow longer, their heights decrease in proportion to N_j^{-2}. A block contains $O(N_j)$ indices with weights $k = O(N_j)$. More precisely, its weighted gradient sum satisfies

$$\sum_{k=2N_{j-1}+1}^{2N_j} k g_k \leq 3 c_j N_j^2 = 3a_j.$$

Since $\sum_j a_j < \infty$, this gives $\sum_k k g_k < \infty$. The inverse construction therefore has finite total displacement and $x_n, y_n \downarrow 0$, as established in Lemma 3. Crucially, each block's contribution is controlled by a_j independently of N_j: we can place a block arbitrarily late without increasing this bound.

At the same time, a long plateau retains a lower bound on the error. The separation $N_j > 2N_{j-1}$ places $[N_j, 2N_j]$ entirely within block j. Over this interval, momentum accumulates the repeated gradient contributions and produces a spatial width of at least $C_{\alpha,s} c_j N_j^2$, where $C_{\alpha,s} > 0$ depends only on the damping and step. The interpolated derivative equals c_j on $[y_{2N_j}, y_{N_j}]$, so the area under

this segment gives

$$F(x_{N_j}) \geq c_j\big(y_{N_j} - y_{2N_j}\big) \geq C_{\alpha,s}c_j^2 N_j^2 = \frac{C_{\alpha,s}a_j^2}{N_j^2}.$$

Lemma 4 proves this estimate independently of the gradients before the block. Hence the normalized residual satisfies $N_j^2 F(x_{N_j}) \geq C_{\alpha,s}a_j^2$ at the selected indices. This is a subsequence lower bound; its vanishing coefficient a_j^2 is consistent with the strict $o(n^{-2})$ guarantee.

Finally, choose the blocks late enough to defeat the proposed gain. Since $G(n) \to \infty$, each N_j can be chosen large enough that $G(N_j) \geq j4^j$, in addition to the separation requirements. Multiplying the retained lower bound by this gain gives

$$N_j^2 G(N_j)F(x_{N_j}) \geq C_{\alpha,s}G(N_j)a_j^2 \geq C_{\alpha,s}j \longrightarrow \infty. \tag{13}$$

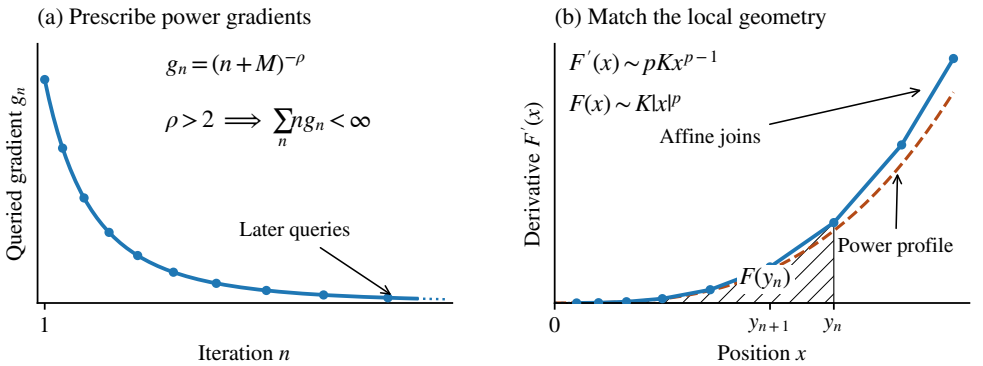
Thus the same amplitudes that ensure convergence leave a residual lower bound that the growing gain amplifies without bound along one subsequence.

It remains to meet the prescribed smoothness bound. Within each block, the interpolated derivative is constant; only the joins between gradient levels require control. These joins occur across the space between consecutive query points. The velocity built up during the preceding block provides enough separation to keep their slopes below L once the first scale is sufficiently large. The explicit choice is given in Algorithm 1, and the estimates are proved in Appendix B.3. A concrete instance with these plateaus and continuous affine joins is shown in Appendix C.6.

All blocks are chosen for the given G and incorporated into a single derivative, with their query points accumulating at zero. The same fixed loss therefore realizes every lower bound in (13).

4.3 Theorem 2: Matching the orbit to power geometry

We now use the inverse construction of Section 4.1 to obtain a loss with local p-power geometry and an exact power-rate tail. Following the gradient-design approach of Section 4.2, we prescribe a power-law sequence whose exponent links the shape of the loss to the decay of its orbit. Figure 2 summarizes the construction.



(c) Match the exponents

$x_n \sim y_n \sim An^{2-\rho}$

$\rho = (p-1)(\rho-2) \implies \rho = 2 + \frac{2}{p-2}$

$F(x_n) \sim Dn^{-2p/(p-2)}, \quad D > 0$

(d) Control smoothness and flatness

$\text{slope}_n \leq \frac{\rho}{s(n+M)}$

$M \geq \alpha + \frac{\rho}{sL}$

$\implies F'$ is L–Lipschitz, $H(p)$

Figure 2: Power construction; axes are not to scale. Blue shows the gradients (a) and affine derivative (b); orange marks the power profile, and hatching represents $F(y_n)$. Exponent matching yields the rate (c); the shift ensures smoothness and flatness (d). Here $p > 2$ and $A, K, D > 0$.

For the prescribed $p > 2$, choose

$$g_n = (n + M)^{-\rho}, \qquad \rho = 2 + \frac{2}{p-2}, \qquad M > 0. \tag{14}$$

The inverse construction yields

$$x_n \sim y_n \sim An^{-2/(p-2)}, \qquad A > 0.$$

With this choice of ρ, the gradients $g_n = F'(y_n)$ are asymptotically proportional to y_n^{p-1}, the derivative behavior of a p-power loss. This is the link between the temporal gradient sequence and the spatial geometry we want to construct.

Interpolating the derivative and integrating it as in Section 4.1 gives

$$F(x) \sim K|x|^p \quad (x \to 0), \qquad F(x_n) \sim Dn^{-2p/(p-2)} \quad (n \to \infty), \qquad K, D > 0.$$

Thus the same fixed loss has local p-power growth and attains the upper rate in Theorem 2 with a positive asymptotic constant. This establishes sharpness at the stated exponent.

Choosing the fixed shift M sufficiently large ensures global smoothness and the required flatness without changing the asymptotic power. Appendix C gives the parameter checks, interpolation estimates, and extension to any fixed finite dimension, together with a concrete $p = 4$ illustration. Algorithm 1 records the complete choices for both constructions.

Algorithm 1 Complete fixed-loss construction for Theorems 1 and 2

Require: Parameters of Theorem 1 or 2, with $\alpha > 3$, $L > 0$, $0 < s < L^{-1}$, dimension $d \geq 1$, and initial distance $R > 0$ for Theorem 1.

1: For Theorem 1, set $N_0 = 0$ and choose integers $N_j > 2N_{j-1}$ with
$N_j \geq \max\{8, 3(\alpha+1)e^\alpha/(sL)\}$ and $G(N_j) \geq j4^j$.
Set $g_k = 2^{-j}/N_j^2$ for $2N_{j-1} < k \leq 2N_j$.
2: For Theorem 2, set $\rho = 2 + 2/(p-2)$, choose $M \geq \alpha + \rho/(sL)$, and set $g_n = (n+M)^{-\rho}$.
3: Set $v_1 = 0$ and $v_{n+1} = \beta_n v_n + sg_n$ for every $n \geq 1$.
4: Set $x_n = \sum_{j=n+1}^{\infty} v_j$, $x_0 = x_1$, and $y_n = x_n - \beta_n v_n$.
5: Set $\psi(0) = 0$, $\psi(y_n) = g_n$; interpolate affinely between consecutive nodes.
6: Set $\psi(x) = g_1$ for $x \geq y_1$ and $F(x) = \int_0^{|x|} \psi(u)\, du$.
7: For Theorem 1, set $h(t) = Lt^2/2$ and $\delta = R/x_1$; replace (F, x_n, y_n) by
$(\delta^2 F(\,\cdot\,/\delta), \delta x_n, \delta y_n)$.
8: For Theorem 2, set $h(0) = 0$ and $h'(t) = \frac{L}{p-1}\operatorname{sgn}(t)\min\{|t|^{p-1}, 1\}$.
9: Define $\widehat{F}(z) = F(z_1) + \sum_{i=2}^{d} h(z_i)$.
10: **return** $\widehat{F}$ and its exact CD orbit $\widehat{x}_n = (x_n, 0, \ldots, 0)$, $\widehat{y}_n = (y_n, 0, \ldots, 0)$.

5 Numerical experiments on the fixed losses

We run CD on the fixed losses with $L = 1$, $s = 1/2$, and $x_0 = x_1 = 1$. Figure 3 summarizes the results; Appendix F gives the protocols and numerical checks.

5.1 Repeated exceedances for a double-logarithmic gain

For Theorem 1, take $\alpha = 4$ and $G(n) = \log\log(e^e + n)$. With $a_j = (1-r)r^{j-1}$, $r = 0.999$, the scales in Appendix F give $N_j^2 G(N_j) F(x_{N_j})/K \geq 2j$, with a fixed $K > 0$. The first five computed ratios are approximately 200, 394, 590, 786, and 983. Local time n/N_j makes each crossing visible. Because $N_5 = 2^{34437}$, constant-gradient intervals are traversed by the closed-form discrete recurrence, with individual oracle steps near joins. Direct-step and higher-precision checks validate this batching. The five displayed episodes illustrate the construction; infinitely many exceedances follow from the bound $2j$.

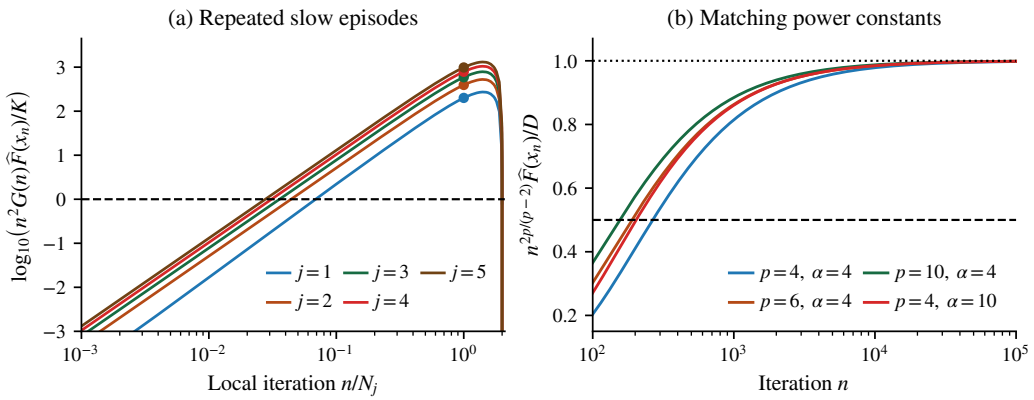


Figure 3: CD on the fixed witnesses. Left: five windows of one orbit; crossing zero means exceeding $K/[n^2G(n)]$, and dots mark N_j. Right: power witnesses approach 1 (dotted), exceeding $1/2$ (dashed). $\widehat{F}$ omits a bounded integration tail; K and D are fixed by construction.

5.2 Attaining the power-rate constants

For Theorem 2, take $g_n = (n+32)^{-\rho}$, $\rho = 2+2/(p-2)$, with $(p,\alpha) = (4,4), (6,4), (10,4), (4,10)$. Each run executes 10^5 iterations. Using the analytic constant D without fitting, the final ratios $n^{2p/(p-2)}\widehat{F}(x_n)/D$ lie between 0.9978 and 0.9987. All four observed tails stay above $1/2$, consistent with the predicted positive asymptotic constant.

6 Conclusion and Limitation

We prove that the strict $o(n^{-2})$ guarantee for CD has no universal remainder rate: every divergent gain is defeated along a subsequence by one fixed smooth convex loss. We also construct fixed losses with $F(x_n) - F^* \sim Dn^{-2p/(p-2)}$, $D > 0$, attaining the established upper rates under the stated geometry and damping assumptions. Both results follow from prescribing queried gradients and recovering a loss whose exact discrete orbit realizes the design.

Our analysis concerns deterministic full-gradient CD on convex objectives. The no-gain loss depends on G and has infinitely many curvature transitions near its minimizer; it does not imply that every loss is slow. Stochastic gradients, adaptive restarts, and nonconvex objectives require separate analysis.

Two questions remain unresolved. First, what is the weakest damping that guarantees the exponent $2p/(p-2)$ under local p-power growth alone? Our examples establish sharpness of the exponent in the stated regimes, but do not establish necessity of the damping threshold. Second, how can these local asymptotic rates be converted into sharp finite-time bounds with explicit constants and onset times? This requires quantifying how local geometry controls the rate and what global information controls entry into the neighborhood where it applies.

AI Use Statement

The initial research was conducted by a fully autonomous in-house research agent system powered by GPT-5.6 Sol and GPT-6 Astra. The human authors' initial contribution was to pose the task of improving the convergence rate of Chambolle–Dossal acceleration or proving that it cannot be further refined. Starting from this task, the agent carried out the entire initial literature search, formulated the mathematical claims, developed and wrote the proofs, produced the manuscript draft, and implemented and checked the Lean 4 formalization. The human authors subsequently used GPT-6 Astra to assist in verifying and correcting the results, refining the arguments, and preparing the final manuscript. The scope and limitations of the formal verification are described in Appendix G.2. The human authors take responsibility for the final content of this work, including all AI-generated text, claims, proofs, and code.

Reproducibility Statement

To support independent verification of the theoretical results, we state the assumptions explicitly and provide complete mathematical proofs in Appendices B–E. The provided supplementary materials include Lean 4 sources for the formalized results, pinned toolchain and dependency versions, and an axiom audit. They also include the numerical experiment code, recorded trajectories, and scripts for rerunning the experiments and regenerating the figures, with experimental settings documented in Appendix F. Appendix G introduces these materials and provides setup and reproduction commands, with further instructions in the accompanying README files. Appendix G.2 identifies the formalized results and their correspondence to the manuscript.

References

Vassilis Apidopoulos, Jean-François Aujol, Charles Dossal, and Aude Rondepierre. Convergence rates of an inertial gradient descent algorithm under growth and flatness conditions. *Mathematical Programming*, 187(1):151–193, May 2021. ISSN 1436-4646. doi: 10.1007/s10107-020-01476-3. URL https://doi.org/10.1007/s10107-020-01476-3.

Yossi Arjevani and Ohad Shamir. On the iteration complexity of oblivious first-order optimization algorithms. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 908–916, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/arjevani16.html.

Hedy Attouch and Juan Peypouquet. The rate of convergence of nesterov's accelerated forward-backward method is actually faster than $1/k^2$. *SIAM Journal on Optimization*, 26(3):1824–1834, 2016. doi: 10.1137/15M1046095. URL https://doi.org/10.1137/15M1046095.

Hedy Attouch, Zaki Chbani, Juan Peypouquet, and Patrick Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168 (1):123–175, Mar 2018. ISSN 1436-4646. doi: 10.1007/s10107-016-0992-8. URL https://doi.org/10.1007/s10107-016-0992-8.

Jean-Francois Aujol, Charles Dossal, and Aude Rondepierre. Optimal convergence rates for nesterov acceleration. *SIAM Journal on Optimization*, 29(4):3131–3153, 2019. doi: 10.1137/18M1186757. URL https://doi.org/10.1137/18M1186757.

Jean-François Aujol, Charles Dossal, Hippolyte Labarrière, and Aude Rondepierre. Strong convergence of fista iterates under hölderian and quadratic growth conditions, 2024. URL https://arxiv.org/abs/2407.17063.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542. URL https://doi.org/10.1137/080716542.

A. Chambolle and Ch. Dossal. On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". *Journal of Optimization Theory and Applications*, 166(3):968–982, Sep 2015. ISSN 1573-2878. doi: 10.1007/s10957-015-0746-4. URL https://doi.org/10.1007/s10957-015-0746-4.

Leonardo de Moura and Sebastian Ullrich. The lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28: 28th International Conference on Automated Deduction, Virtual Event, July 12–15, 2021, Proceedings*, pp. 625–635, Berlin, Heidelberg, 2021. Springer-Verlag. ISBN 978-3-030-79875-8. doi: 10.1007/978-3-030-79876-5_37. URL https://doi.org/10.1007/978-3-030-79876-5_37.

DLMF. *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/, Release 1.2.8 of 2026-09-15. URL https://dlmf.nist.gov/. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1):451–482, Jun 2014. ISSN 1436-4646. doi: 10.1007/s10107-013-0653-0. URL https://doi.org/10.1007/s10107-013-0653-0.

Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 1999.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2 edition, 2012.

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*. Springer Science & Business Media, 2013.

Jonathan W. Siegel and Stephan Wojtowytsch. A qualitative difference between gradient flows of convex functions in finite- and infinite-dimensional hilbert spaces, 2023. URL https://arxiv.org/abs/2310.17610.

Weijie Su, Stephen Boyd, and Emmanuel J. Candès. A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17 (153):1–43, 2016. URL http://jmlr.org/papers/v17/15-084.html.

Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161 (1):307–345, Jan 2017a. ISSN 1436-4646. doi: 10.1007/s10107-016-1009-3. URL https://doi.org/10.1007/s10107-016-1009-3.

Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Exact worst-case performance of first-order methods for composite convex optimization. *SIAM Journal on Optimization*, 27(3): 1283–1313, 2017b.

A PROOF ROADMAP AND STANDING NOTATION

The proofs have two construction branches and one upper-bound branch. Both constructions first realize prescribed gradients as an exact CD orbit of a single smooth convex loss. Constant-gradient blocks then give the arbitrary-gain obstruction, while a shifted power sequence gives the sharp geometry-dependent examples. The upper bounds use established energy estimates, with a localization argument that removes coercivity in finite dimension. Figure 4 records these dependencies. The notation and proof targets are restated below so that the appendix can be read independently of the main exposition.

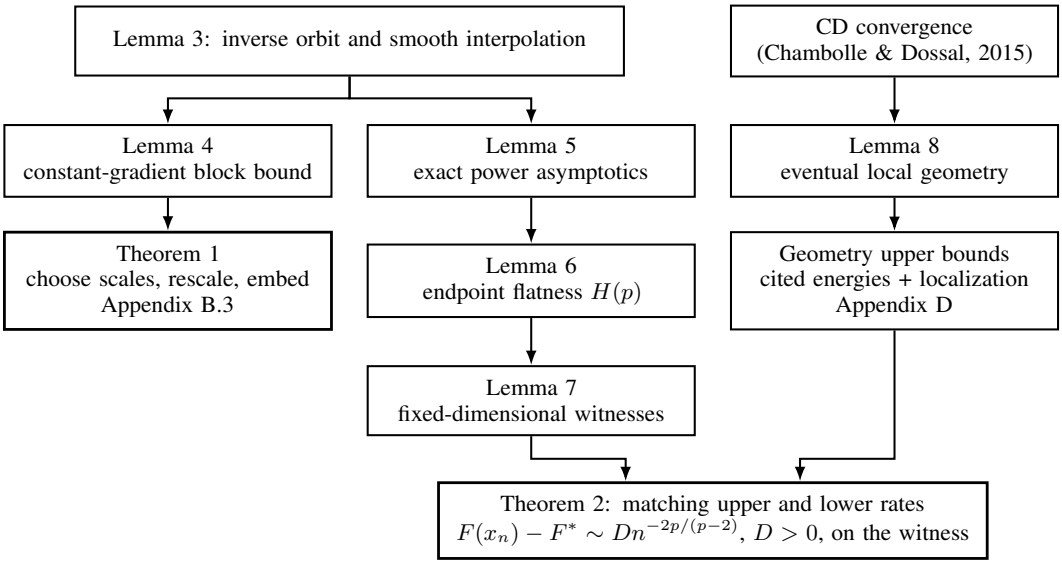


Figure 4: Proof dependencies. An arrow indicates an input to the next argument. Numbered lemmas are proved in this appendix. CD convergence is taken from Chambolle & Dossal (2015); the upper-rate arguments use Apidopoulos et al. (2021) and Aujol et al. (2024). The block and power constructions are independent of those upper-rate arguments.

A.1 WHAT EACH PART OF THE PROOF MUST ACCOMPLISH

Appendix B first recovers positions from the full velocity tail. The main difficulty is to make all infinitely many queries compatible with a globally Lipschitz derivative at their accumulation point. The general interpolation bound is $1/s$, which is larger than the prescribed L when $s < L^{-1}$. The block construction therefore needs a separate curvature estimate. Its scale choices balance weighted summability against the lower bound at infinitely many selected indices.

Appendix C replaces blocks by a shifted power sequence. Leading asymptotics determine both the residual exponent and local growth, but do not settle the endpoint inequality $pF(x) \le xF'(x)$. The decisive step is to retain the first correction, choose its sign through the shift, and show that affine interpolation introduces a smaller error. A separable extension then preserves this geometry in every fixed finite dimension.

Appendix D proves the upper statements. It identifies the published iteration convention and explains exactly where local geometry becomes valid. For growth alone, the technical difficulty is that the Lyapunov energy has a negative distance term; the growth inequality controls that term and closes the residual estimate. Appendix E treats the quadratic boundary separately, and Appendix G introduces the supplementary materials and records the scope of the Lean 4 formalization.

A.2 SETTING, RECURRENCE, AND LOCAL GEOMETRY

Let $F : \mathbb{R}^d \to \mathbb{R}$ be convex, attain its minimum F^*, and have a globally L-Lipschitz gradient, where $d \geq 1$ is fixed and $L > 0$. Norms and inner products are Euclidean. We write $C^{1,1}$ for a differentiable loss with globally Lipschitz gradient, and call F coercive when $F(x) \to \infty$ as $\|x\| \to \infty$. For $\alpha > 3$ and $0 < s < L^{-1}$, the Chambolle–Dossal (CD) recurrence is

$$\beta_n = \frac{n-1}{n+\alpha-1}, \qquad y_n = x_n + \beta_n(x_n - x_{n-1}), \qquad x_{n+1} = y_n - s\nabla F(y_n), \quad n \geq 1. \tag{15}$$

The queried gradients are $g_n = \nabla F(y_n)$; the residual and its normalization are $r_n = F(x_n) - F^*$ and $q_n = n^2 r_n$. Since $\beta_1 = 0$, x_0 does not affect later iterates. We use $x_0 = x_1$ and normalize our constructed losses so that $F^* = F(0) = 0$.

For a unique minimizer x^*, the local conditions used here are

$$H(\beta): \quad F(x) - F^* \leq \frac{1}{\beta}\langle \nabla F(x), x - x^* \rangle \qquad (\|x - x^*\| \leq \eta), \tag{16}$$

$$L(p): \quad F(x) - F^* \geq K\|x - x^*\|^p \qquad (\|x - x^*\| \leq \varepsilon), \tag{17}$$

for some $\eta, \varepsilon, K > 0$. The flatness exponent $\beta > 2$ is distinct from the momentum coefficient β_n. If the minimizer set X^* need not be a singleton, we use

$$L_{\mathrm{set}}(p): \quad F(x) - F^* \geq K\,\mathrm{dist}(x, X^*)^p \quad \text{whenever } \mathrm{dist}(x, X^*) \leq \varepsilon. \tag{18}$$

The same constants apply throughout the indicated neighborhood. Here $p > 2$, and $\mathrm{dist}(x, X^*)$ is the Euclidean distance to the minimizer set. For a unique minimizer, $L_{\mathrm{set}}(p)$ is $L(p)$. The symbols O, o, and $\sim$ refer to the stated limit; in particular, $a_n \sim b_n$ means $a_n/b_n \to 1$.

A.3 THE CONCLUSIONS TO BE PROVED

For Theorem 1, fix any positive nondecreasing gain $G(n) \to \infty$ and initial distance $R > 0$, in addition to α, L, s. We construct one even, coercive scalar convex loss with L-Lipschitz derivative, unique minimizer zero, and an exact CD orbit from $x_0 = x_1 = R$ such that

$$\sup_{n \geq 1} n^2 G(n) F(x_n) = \infty.$$

The same residual sequence must be realizable in every fixed finite dimension. This unboundedness conclusion coexists with the established $r_n = o(n^{-2})$ guarantee (Attouch & Peypouquet, 2016); the gain-weighted residual is unbounded on a subsequence, not necessarily divergent at every index.

For Theorem 2, put $q = 2p/(p-2)$. We establish the upper bound $r_n = O(n^{-q})$ for every initialization under either of these regimes:

(i) unique minimizer, $H(\beta)$ and $L(p)$, with $2 < \beta \leq p$ and $\alpha \geq (\beta+2)/(\beta-2)$;

(ii) $L_{\mathrm{set}}(p)$ and $\alpha > 5 + 8/(p-2)$.

For every admissible parameter choice and fixed finite dimension, we also construct a coercive loss with unique minimizer and some initialization for which $r_n \sim Dn^{-q}$, $D > 0$. These examples satisfy $H(p)$ as well as both growth formulations. Their positive asymptotic constant rules out any common divergent multiplicative improvement of the upper rate.

B INVERSE CONSTRUCTION AND THE NO-GAIN THEOREM

The task is to build one loss that defeats a prescribed gain at infinitely many indices. We first give a general procedure for realizing a positive gradient sequence as an exact orbit. A block estimate then quantifies how much objective error a constant-gradient interval retains. Finally, we choose all the blocks at once, verifying that their accumulation at zero preserves the required smoothness.

Prescribing an orbit and recovering a scalar objective has a continuous precedent in Siegel & Wojtowytsch (2023, Lemma 2.7). Interpolation of prescribed first-order data is also a standard tool in smooth convex optimization (Taylor et al., 2017a). Here the query locations are determined by the discrete momentum recurrence, and the interpolation must work at countably many nodes accumulating at the minimizer. The next lemma supplies this compatibility directly.

B.1 From a gradient sequence to one smooth convex loss

The velocity $v_n = x_{n-1} - x_n$ measures motion toward zero. Its recurrence can be solved forward from the gradients, but the positions are defined by summing the entire remaining velocity tail. This is what fixes the limiting point at zero without knowing the loss in advance.

Lemma 3 (Inverse orbit and smooth interpolation)**.** *Let $\alpha > 3$ and $s > 0$. Let $(g_n)_{n\geq 1}$ be positive, nonincreasing, and satisfy $\sum_{n\geq 1} n g_n < \infty$. With $\beta_n = (n-1)/(n+\alpha-1)$, define*

$$\begin{aligned} &v_1 = 0, \qquad v_{n+1} = \beta_n v_n + s g_n, \\ &x_n = \sum_{j=n+1}^{\infty} v_j, \qquad y_n = x_n - \beta_n v_n, \qquad n \geq 1, \qquad x_0 = x_1. \end{aligned} \tag{19}$$

These sums are finite. In particular, $\sum_n v_n < \infty$, $x_n \downarrow 0$, $y_n \downarrow 0$, and $y_n - s g_n = x_{n+1}$. There is a convex $C^{1,1}$ function $F : \mathbb{R} \to \mathbb{R}$ with $F'(y_n) = g_n$ for all n, so these points are an exact CD orbit. The function may be chosen even, coercive, and uniquely minimized at zero. Its derivative is $1/s$-Lipschitz; if all interpolation slopes are at most L, it is L-Lipschitz.

Proof of Lemma 3. 1. Summability of the recovered velocities For integers $1 \leq k \leq n$, with the empty-product convention $P_{k,k} = 1$, the products of momentum coefficients satisfy

$$P_{k,n} := \prod_{i=k+1}^{n} \beta_i = \frac{\Gamma(n)\Gamma(k+\alpha)}{\Gamma(k)\Gamma(n+\alpha)}, \qquad \sum_{n=k}^{\infty} P_{k,n} = \frac{k+\alpha-1}{\alpha-1}. \tag{20}$$

Here Γ is the gamma function. Its recurrence $\Gamma(z+1) = z\Gamma(z)$ for $z > 0$ gives

$$\prod_{i=k+1}^{n} \beta_i = \frac{k(k+1)\cdots(n-1)}{(k+\alpha)(k+\alpha+1)\cdots(n+\alpha-1)} = \frac{\Gamma(n)/\Gamma(k)}{\Gamma(n+\alpha)/\Gamma(k+\alpha)}.$$

To compute the infinite sum, use the classical beta integral (DLMF, Eq. 5.12.1):

$$\frac{\Gamma(n)}{\Gamma(n+\alpha)} = \frac{1}{\Gamma(\alpha)} \int_0^1 t^{n-1}(1-t)^{\alpha-1}\, dt.$$

All integrands are nonnegative, so Tonelli's theorem permits exchanging the sum and the integral (Folland, 1999, Theorem 2.37). The geometric series $\sum_{n=k}^{\infty} t^{n-1} = t^{k-1}/(1-t)$ for $0 < t < 1$ then yields

$$\begin{aligned} \sum_{n=k}^{\infty} P_{k,n} &= \frac{\Gamma(k+\alpha)}{\Gamma(k)\Gamma(\alpha)} \int_0^1 \left(\sum_{n=k}^{\infty} t^{n-1} \right) (1-t)^{\alpha-1}\, dt \\ &= \frac{\Gamma(k+\alpha)}{\Gamma(k)\Gamma(\alpha)} \int_0^1 t^{k-1}(1-t)^{\alpha-2}\, dt \\ &= \frac{\Gamma(k+\alpha)}{\Gamma(k)\Gamma(\alpha)} \frac{\Gamma(k)\Gamma(\alpha-1)}{\Gamma(k+\alpha-1)} = \frac{k+\alpha-1}{\alpha-1}. \end{aligned}$$

The second beta integral is finite because $k \geq 1$ and $\alpha > 1$, as ensured by our assumption $\alpha > 3$. The last equality uses $\Gamma(k+\alpha) = (k+\alpha-1)\Gamma(k+\alpha-1)$ and $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. We next expand the velocity recurrence. The base case is $v_2 = s g_1 = s g_1 P_{1,1}$. If the expansion holds at index n, then

$$v_{n+2} = \beta_{n+1} s \sum_{k=1}^{n} g_k P_{k,n} + s g_{n+1} = s \sum_{k=1}^{n+1} g_k P_{k,n+1},$$

because $\beta_{n+1} P_{k,n} = P_{k,n+1}$ and $P_{n+1,n+1} = 1$. Thus induction and the product sum give

$$v_{n+1} = s \sum_{k=1}^{n} g_k P_{k,n}, \qquad \sum_{n\geq 1} v_{n+1} = \frac{s}{\alpha-1} \sum_{k\geq 1} (k+\alpha-1) g_k < \infty. \tag{21}$$

For completeness, the sum exchange and the finiteness assertion in (21) are

$$\sum_{n=1}^{\infty} v_{n+1} = s\sum_{n=1}^{\infty}\sum_{k=1}^{n} g_k P_{k,n} = s\sum_{k=1}^{\infty} g_k \sum_{n=k}^{\infty} P_{k,n}$$
$$= \frac{s}{\alpha-1}\sum_{k=1}^{\infty}(k+\alpha-1)g_k \le \frac{s\alpha}{\alpha-1}\sum_{k=1}^{\infty} k g_k < \infty.$$

Tonelli applies because every summand is nonnegative; the inequality uses $k+\alpha-1 \le \alpha k$ for $k \ge 1$ and $\alpha > 1$. Also $0 \le \beta_n < 1$ and $g_n > 0$ imply $v_{n+1} \ge s g_n > 0$. Consequently, the tails defining x_n are finite, $x_n - x_{n+1} = v_{n+1} > 0$, and $x_n \downarrow 0$.

To express these tails directly in terms of the gradients, the same beta-integral calculation, now starting at $m = n$, gives for $1 \le k \le n$

$$\sum_{m=n}^{\infty} P_{k,m} = \frac{\Gamma(k+\alpha)}{\Gamma(k)\Gamma(\alpha)}\int_0^1 t^{n-1}(1-t)^{\alpha-2}\,dt$$
$$= \frac{\Gamma(k+\alpha)\Gamma(n)}{\Gamma(k)(\alpha-1)\Gamma(n+\alpha-1)}.$$

Substituting the velocity expansion into x_n and again exchanging nonnegative sums yields

$$x_n = \sum_{m=n}^{\infty} v_{m+1} = s\sum_{k=1}^{\infty} g_k \sum_{m=\max\{n,k\}}^{\infty} P_{k,m}$$
$$= s\sum_{k=1}^{n} g_k \sum_{m=n}^{\infty} P_{k,m} + s\sum_{k=n+1}^{\infty} g_k \sum_{m=k}^{\infty} P_{k,m}.$$

Use the tail identity in the first term and (20) in the second to obtain

$$x_n = \frac{s}{\alpha-1}\left[\frac{\Gamma(n)}{\Gamma(n+\alpha-1)}\sum_{k=1}^{n}\frac{\Gamma(k+\alpha)}{\Gamma(k)} g_k + \sum_{k=n+1}^{\infty}(k+\alpha-1)g_k\right]. \tag{22}$$

2. Ordering the query points The tail differences give $x_{n-1} - x_n = v_n$ for $n \ge 2$; the same identity holds at $n = 1$ because $x_0 = x_1$ and $v_1 = 0$. Using $\beta_n v_n = v_{n+1} - s g_n$, we therefore have

$$y_n = x_n - \beta_n v_n = x_n - v_{n+1} + s g_n = x_{n+1} + s g_n > 0.$$

Subtracting this identity at consecutive indices gives

$$y_n - y_{n+1} = v_{n+2} + s(g_n - g_{n+1}) > 0. \tag{23}$$

The right side is strictly positive because $v_{n+2} > 0$ and $g_n \ge g_{n+1}$. Moreover, $\sum_n n g_n < \infty$ implies $n g_n \to 0$, hence $g_n \to 0$. Together with $x_{n+1} \to 0$, the identity $y_n = x_{n+1} + s g_n$ proves $y_n \downarrow 0$.

3. Interpolating through the accumulation point Set $\psi(0) = 0$ and $\psi(y_n) = g_n$, interpolate affinely on each $[y_{n+1}, y_n]$, and set $\psi(x) = g_1$ for $x \ge y_1$. Explicitly, on an interpolation interval,

$$\psi(x) = g_{n+1} + \frac{g_n - g_{n+1}}{y_n - y_{n+1}}(x - y_{n+1}).$$

The derivative so defined is continuous, positive on $(0, \infty)$, and nondecreasing. Its affine slopes satisfy

$$0 \le \ell_n = \frac{g_n - g_{n+1}}{v_{n+2} + s(g_n - g_{n+1})} \le \frac{1}{s}. \tag{24}$$

Here the denominator is positive. If $g_n = g_{n+1}$ the slope is zero; otherwise the denominator is at least $s(g_n - g_{n+1})$, giving the upper bound. For $0 \le x \le y_n$, monotonicity gives $0 \le \psi(x) \le g_n \to 0$, proving continuity at zero.

Let $\Lambda = 1/s$, or $\Lambda = L$ when all $\ell_n \le L$. Given $0 < u < v$, only finitely many nodes lie in $[u, v]$, because $y_n \to 0$. If $u = t_0 < \cdots < t_r = v$ partitions this interval at those nodes, the slope bounds and the constant outer extension give

$$0 \le \psi(v) - \psi(u) = \sum_{i=1}^{r} \big(\psi(t_i) - \psi(t_{i-1})\big) \le \Lambda \sum_{i=1}^{r} (t_i - t_{i-1}) = \Lambda(v - u).$$

Letting $u \downarrow 0$ also gives $\psi(v) \le \Lambda v$. Extend ψ oddly to $\mathbb{R}$. The same-sign Lipschitz bounds follow by reflection; for $u < 0 < v$,

$$|\psi(v) - \psi(u)| = \psi(v) + \psi(-u) \le \Lambda(v - u).$$

The extension is continuous and nondecreasing, so its integral

$$F(x) = \int_0^x \psi(u)\, du = \int_0^{|x|} \psi(u)\, du$$

is even and convex, with $F' = \psi$ globally Λ-Lipschitz. Thus $F \in C^{1,1}$, including at zero. Since $\psi(u) > 0$ for $u > 0$, $F(x) > F(0) = 0$ whenever $x \ne 0$. Finally, for $|x| \ge y_1$,

$$F(x) = F(y_1) + g_1(|x| - y_1) \longrightarrow \infty \quad \text{as } |x| \to \infty,$$

which proves coercivity. The identities $F'(y_n) = g_n$, $y_n = x_n + \beta_n(x_n - x_{n-1})$, and $y_n - s g_n = x_{n+1}$ verify both updates of the exact CD recurrence. □

The automatic bound $1/s$ does not yet meet the standing requirement $L < 1/s$. Each construction below must therefore control its own interpolation slopes. This is the only extra smoothness check needed once the gradient sequence has been chosen.

B.2 Why a constant-gradient block retains error

Consider a block $n \le k \le 2n$ on which $g_k = c > 0$. Each update adds sc to the velocity. The momentum products within this block have a positive lower bound depending only on α, so these contributions accumulate: during the latter part of the block, the velocity is at least a constant times scn. Summing over a number of steps proportional to n gives a spatial segment of length at least a constant times scn^2. The interpolated derivative equals c on this segment, so its integral contributes at least $C_{\alpha,s} c^2 n^2$ to $F(x_n)$, for some $C_{\alpha,s} > 0$. Earlier gradients contribute only nonnegative terms to the velocity and can be discarded in this estimate. The lower bound therefore depends only on the current block.

Lemma 4 (Constant-gradient block)**.** *If $g_k = c$ for $n \le k \le 2n$ and $n \ge 8$, then the objective from Lemma 3 satisfies*

$$F(x_n) \ge \frac{s e^{-\alpha}}{6} c^2 n^2.$$

Proof of Lemma 4. Since $0 < y_n \le x_n$ and F is nondecreasing on the positive axis, $F(x_n) \ge F(y_n)$. On $[y_{k+1}, y_k]$, the derivative is at least g_{k+1}. Partitioning $[0, y_n]$ into these intervals therefore gives

$$F(x_n) \ge F(y_n) \ge \sum_{k=n}^{\infty} g_{k+1}(y_k - y_{k+1}) \ge \sum_{k=n}^{\infty} g_{k+1} v_{k+2}.$$

The first sum follows by integrating interval by interval and taking the limit at zero; the second uses (23) and $g_k - g_{k+1} \ge 0$. For $n \le j \le k + 1 \le 2n$, apply $\log(1 - u) \ge -u/(1 - u)$ for $0 \le u < 1$ to the product in (20):

$$\log P_{j,k+1} = \sum_{i=j+1}^{k+1} \log\left(1 - \frac{\alpha}{i + \alpha - 1}\right) \ge -\alpha \sum_{i=j+1}^{k+1} \frac{1}{i - 1} \ge -\alpha.$$

The logarithmic inequality follows from $-\log(1 - u) = \int_0^u (1 - t)^{-1}\, dt \le u/(1 - u)$. Here $i \ge n + 1 \ge 9$ makes $u = \alpha/(i + \alpha - 1)$ lie in $(0, 1)$, and

$$\sum_{i=j+1}^{k+1} \frac{1}{i - 1} \le \frac{k + 1 - j}{j} \le \frac{2n - n}{n} = 1.$$

Thus $P_{j,k+1} \geq e^{-\alpha}$ throughout this range. For each $\lceil 3n/2 \rceil \leq k \leq 2n-1$, retain only the block terms $n \leq j \leq k+1$ in the velocity expansion:

$$v_{k+2} = s\sum_{j=1}^{k+1} g_j P_{j,k+1} \geq sc\sum_{j=n}^{k+1} P_{j,k+1} \geq sce^{-\alpha}(k-n+2) \geq \frac{1}{2}sce^{-\alpha}n.$$

There are $2n - \lceil 3n/2 \rceil = \lfloor n/2 \rfloor \geq n/3$ such indices when $n \geq 8$, and $g_{k+1} = c$ at all of them. Hence

$$F(x_n) \geq \sum_{k=\lceil 3n/2 \rceil}^{2n-1} cv_{k+2} \geq \frac{n}{3}\, c\, \frac{1}{2}sce^{-\alpha}n = \frac{se^{-\alpha}}{6}c^2n^2.$$

□

B.3 Choosing all scales and completing the no-gain proof

The choice of scales must satisfy two competing demands. The weighted gradient sum must remain finite so that the orbit converges to zero, while the gain-weighted normalized residual must become arbitrarily large along a subsequence. We first choose a summable sequence (a_j) to control each block's contribution to $\sum_k kg_k$, then place each block late enough that G is large. This choice of separated scales is analogous to the scalar flow construction in Siegel & Wojtowytsch (2023, Example 2.9); the interpolation-slope estimate below is what makes it compatible with the fixed discrete step.

Proof of Theorem 1. 1. Choose one infinite gradient sequence Work first in dimension one. Let $a_j = 2^{-j}$ and $N_0 = 0$. Recursively choose integers N_j so that

$$N_j > 2N_{j-1}, \qquad N_j \geq 8, \qquad G(N_j) \geq j4^j, \qquad N_1 \geq \frac{3(\alpha+1)e^{\alpha}}{sL}. \tag{25}$$

Such choices exist because $G(n) \to \infty$. Set $c_j = a_j/N_j^2$ and $g_k = c_j$ for $2N_{j-1} < k \leq 2N_j$. Indeed, after N_{j-1} is fixed, the gain inequality holds at every sufficiently large integer, so the separation and size conditions can be imposed simultaneously. Since $N_j \to \infty$, the blocks cover all positive integers without gaps. Their heights satisfy

$$\frac{c_{j+1}}{c_j} = \frac{1}{2}\left(\frac{N_j}{N_{j+1}}\right)^2 < \frac{1}{8},$$

so the full gradient sequence is positive and nonincreasing. Moreover,

$$\sum_{k\geq 1} kg_k \leq \sum_{j\geq 1} c_j \sum_{k=1}^{2N_j} k \leq 3\sum_{j\geq 1} c_j N_j^2 = 3\sum_{j\geq 1} a_j < \infty.$$

Here $\sum_{k=1}^{2N_j} k = N_j(2N_j+1) \leq 3N_j^2$ since $N_j \geq 1$, and $\sum_{j\geq 1} a_j = 1$. Thus Lemma 3 produces a single even, coercive loss with unique minimizer zero and an exact CD orbit.

2. Enforce the prescribed smoothness at every block transition Inside a block, consecutive gradients agree, so the interpolation slope vanishes. At a transition $n = 2N_j$, the last N_j terms in (21), indexed by $N_j + 1 \leq k \leq 2N_j$, lie inside block j. The product estimate in Lemma 4 gives

$$v_{n+1} \geq s\sum_{k=N_j+1}^{2N_j} c_j P_{k,2N_j} \geq sc_jN_je^{-\alpha}.$$

At this transition $g_n = c_j$ and $g_{n+1} = c_{j+1}$. Since $y_n = x_{n+1} + sg_n$, the spacing between the interpolation nodes is

$$\begin{aligned} y_n - y_{n+1} &= x_{n+1} - x_{n+2} + s(g_n - g_{n+1}) \\ &= v_{n+2} + s(g_n - g_{n+1}) \\ &= \beta_{n+1}v_{n+1} + sc_{j+1} + s(c_j - c_{j+1}) \end{aligned}$$

$$= \beta_{n+1} v_{n+1} + s c_j \geq s c_j \left(1 + \frac{N_j e^{-\alpha}}{\alpha + 1}\right).$$

The last inequality uses $\beta_{n+1} = n/(n+\alpha) \geq 1/(\alpha+1)$, equivalent to $\alpha(n-1) \geq 0$. The interpolation slope on $[y_{n+1}, y_n]$ is the gradient difference divided by this spacing. Since $0 < c_j - c_{j+1} \leq c_j$,

$$\begin{aligned} \ell_{2N_j} &= \frac{g_n - g_{n+1}}{y_n - y_{n+1}} = \frac{c_j - c_{j+1}}{y_n - y_{n+1}} \\ &\leq \frac{1}{s(1 + N_j e^{-\alpha}/(\alpha+1))} \leq \frac{L}{3}. \end{aligned}$$

For the final inequality, put $t_j = N_j e^{-\alpha}/(\alpha+1) \geq 3/(sL)$ and use $1/[s(1+t_j)] \leq 1/(st_j) \leq L/3$. The lower bound on N_1 therefore controls every transition, including those arbitrarily close to zero. The loss has an L-Lipschitz derivative.

3. Violate the gain along infinitely many iterates The interval $[N_j, 2N_j]$ lies inside block j. Applying Lemma 4 at $n = N_j$ and substituting $c_j = a_j/N_j^2$ gives

$$N_j^2 F(x_{N_j}) \geq \frac{s e^{-\alpha}}{6} a_j^2, \qquad N_j^2 G(N_j) F(x_{N_j}) \geq \frac{s e^{-\alpha}}{6} j \longrightarrow \infty.$$

Indeed, $N_j^2 (c_j^2 N_j^2) = a_j^2 = 4^{-j}$ and $G(N_j) 4^{-j} \geq j$. The objective remains fixed as j grows. Since every finite prefix has finite residuals, this also rules out any eventual bound with an instance-dependent constant and starting index.

4. Prescribe the initial distance and ambient dimension The recovered initial point x_1 is positive and finite. For any prescribed $R > 0$, put $\delta = R/x_1$ and $F_\delta(t) = \delta^2 F(t/\delta)$. Its derivative is $F'_\delta(t) = \delta F'(t/\delta)$, with the same Lipschitz constant. More explicitly,

$$|F'_\delta(u) - F'_\delta(v)| \leq \delta L |u/\delta - v/\delta| = L|u - v|, \qquad \delta y_n - s F'_\delta(\delta y_n) = \delta(y_n - s F'(y_n)) = \delta x_{n+1}.$$

The extrapolation identity also scales by δ, so the scaled orbit starts at $\delta x_0 = \delta x_1 = R$ and follows the same CD recurrence. Since $F_\delta(\delta x_n) = \delta^2 F(x_n)$, its gain-weighted residual remains unbounded.

For any prescribed $d > 1$, set

$$\widetilde{F}(z) = F_\delta(z_1) + \frac{L}{2} \sum_{i=2}^{d} z_i^2 \quad \text{and initialize at } (R, 0, \ldots, 0).$$

This loss is coercive, has unique minimizer zero, and has an L-Lipschitz gradient in the Euclidean norm: for $z, w \in \mathbb{R}^d$,

$$\|\nabla \widetilde{F}(z) - \nabla \widetilde{F}(w)\|^2 = |F'_\delta(z_1) - F'_\delta(w_1)|^2 + L^2 \sum_{i=2}^{d} |z_i - w_i|^2 \leq L^2 \|z - w\|^2.$$

To check coercivity, a sublevel $\widetilde{F}(z) \leq H$ with $H \geq 0$ forces $F_\delta(z_1) \leq H$ and $\sum_{i=2}^d z_i^2 \leq 2H/L$. The scalar coercivity bounds z_1, so every such sublevel is bounded. Continuity then gives coercivity in finite dimension. Each summand vanishes only at zero, proving uniqueness of the minimizer. If two successive transverse coordinates equal zero, their extrapolated coordinate and gradient also equal zero; induction therefore keeps them zero. The first coordinate follows the scaled scalar orbit, and $\widetilde{F}(\delta x_n, 0, \ldots, 0) = \delta^2 F(x_n)$. Both the initial distance and the scalar residual sequence are preserved. This completes the theorem. □

C Exact power witnesses and endpoint flatness

We next build the sharp instances for the geometry-dependent rates. Unlike the block construction, a power gradient sequence has a regular tail, which makes its velocity, position, and loss asymptotics explicit. The main additional work is at the endpoint $H(p)$: a limit of $xF'(x)/F(x)$ equal to p does not say on which side of p the ratio lies. We will choose a positive first correction and show that it survives interpolation between the query points.

C.1 Parameter choice and the required exponent

For the prescribed $p > 2$, set

$$\rho = 2 + \frac{2}{p-2}, \qquad \delta = \rho - 2, \qquad g_n = (n+M)^{-\rho}.$$

Choose a fixed positive shift M, subject to the bounds below. In part (i), the map $t \mapsto 1 + 4/(t-2)$ decreases for $t > 2$, so

$$\alpha \geq \frac{\beta+2}{\beta-2} = 1 + \frac{4}{\beta-2} \geq 1 + \frac{4}{p-2} = 2\rho - 3.$$

Thus $\rho \leq (\alpha+3)/2 < \alpha$, where the strict inequality uses the standing assumption $\alpha > 3$. In part (ii),

$$\alpha > 5 + \frac{8}{p-2} = 4\rho - 3 > \rho,$$

because $\rho > 2$. Consequently, both regimes give

$$2 < \rho < \alpha, \qquad a := \alpha - \rho > 0. \tag{26}$$

The sequence (g_n) is positive and nonincreasing, and $\sum_n n g_n \leq \sum_n n^{1-\rho} < \infty$, since $M > 0$ and $\rho > 2$. We use the exact inverse orbit from Lemma 3, with $x_0 = x_1$.

C.2 Leading asymptotics, local growth, and global smoothness

The first lemma identifies the exponent and the positive asymptotic constant. It also checks local growth throughout a neighborhood, rather than only at the query points, and enforces the prescribed smoothness.

Lemma 5 (Exact power asymptotics)**.** *Fix $\alpha > 3$, $L > 0$, $0 < s < L^{-1}$, $2 < \rho < \alpha$, and $M \geq \rho/(sL)$. For $g_n = (n+M)^{-\rho}$, construct F using the inverse orbit and the piecewise-affine derivative in the proof of Lemma 3, with $F(0) = 0$. Set*

$$\delta = \rho - 2, \quad p = \frac{2(\rho-1)}{\rho-2}, \quad B = \frac{s}{\alpha+1-\rho}, \quad A = \frac{B}{\delta}, \quad D = \frac{B}{2(\rho-1)}.$$

The resulting loss is even, coercive, convex, and has an L-Lipschitz derivative and unique minimizer zero. Its exact orbit satisfies

$$v_{n+1} \sim Bn^{1-\rho}, \qquad x_n \sim y_n \sim An^{-\delta}, \qquad F(x_n) \sim Dn^{-2p/(p-2)}.$$

Moreover, $F(x) \sim (D/A^p)|x|^p$ as $x \to 0$, so $L(p)$ holds.

Proof of Lemma 5. 1. Recover the velocity and position asymptotics The gamma-product formula (21) gives

$$v_{n+1} = s\frac{\Gamma(n)}{\Gamma(n+\alpha)} \sum_{k=1}^{n} \frac{\Gamma(k+\alpha)}{\Gamma(k)} (k+M)^{-\rho}.$$

The standard gamma-ratio asymptotic (DLMF, Eq. 5.11.12) gives separately

$$\frac{\Gamma(k+\alpha)}{\Gamma(k)} \sim k^{\alpha}, \qquad (k+M)^{-\rho} \sim k^{-\rho}, \qquad \frac{\Gamma(n)}{\Gamma(n+\alpha)} \sim n^{-\alpha}.$$

The summand is therefore asymptotic to k^a, with $a = \alpha - \rho > 0$. Integral comparison gives

$$\frac{n^{a+1}}{a+1} = \int_0^n t^a \, dt \leq \sum_{k=1}^{n} k^a \leq \int_1^{n+1} t^a \, dt \sim \frac{n^{a+1}}{a+1}.$$

For any fixed relative error, the original summand lies between the corresponding multiples of k^a once k is large. Its finite initial sum is $o(n^{a+1})$, so the same asymptotic holds for the full sum. Multiplying by $sn^{-\alpha}$ yields

$$v_{n+1} \sim Bn^{1-\rho}, \qquad B = \frac{s}{\alpha+1-\rho}.$$

To sum the velocity tail, use $\rho > 2$ and

$$\int_n^\infty t^{1-\rho}\,dt \le \sum_{m=n}^\infty m^{1-\rho} \le n^{1-\rho} + \int_n^\infty t^{1-\rho}\,dt, \qquad \int_n^\infty t^{1-\rho}\,dt = \frac{n^{2-\rho}}{\rho-2}.$$

The extra term is $o(n^{2-\rho})$. Bounding v_{m+1} between $(1 \pm \varepsilon)Bm^{1-\rho}$ for every sufficiently large m therefore gives $x_n = \sum_{m=n}^\infty v_{m+1} \sim Bn^{2-\rho}/(\rho-2)$. Also $x_{n+1}/x_n \to 1$ and $sg_n/x_n = O(n^{-2}) \to 0$. Thus $y_n = x_{n+1} + sg_n$ yields

$$x_n \sim y_n \sim An^{-\delta}, \qquad A = \frac{B}{\delta}.$$

2. Integrate the derivative and pass from nodes to a neighborhood Write $\psi = F'$ for the derivative from this construction. It is affine on each interpolation interval $[y_{k+1}, y_k]$, so

$$F(y_n) = \sum_{k=n}^\infty \frac{g_k + g_{k+1}}{2}(y_k - y_{k+1}).$$

Indeed, each term is the exact trapezoidal integral over $[y_{k+1}, y_k]$, and these intervals partition $(0, y_n]$. The finite sum through $k = N$ equals $F(y_n) - F(y_{N+1})$; letting $N \to \infty$ gives the formula because $y_{N+1} \to 0$ and F is continuous with $F(0) = 0$. The mean value theorem gives $g_k - g_{k+1} = O(k^{-\rho-1})$, so $s(g_k - g_{k+1})/v_{k+2} = O(k^{-2}) \to 0$. Hence (23) implies

$$y_k - y_{k+1} = v_{k+2} + s(g_k - g_{k+1}) \sim Bk^{1-\rho}, \qquad \frac{g_k + g_{k+1}}{2} \sim k^{-\rho}.$$

The area summand is asymptotic to $Bk^{1-2\rho}$. Applying the same tail comparison, now with exponent $1 - 2\rho < -1$, gives $\sum_{k=n}^\infty k^{1-2\rho} \sim n^{2-2\rho}/(2\rho-2)$, and thus

$$F(y_n) \sim Dn^{2-2\rho}, \qquad D = \frac{B}{2(\rho-1)}. \tag{27}$$

For $n \ge 2$, the identities $y_n = x_n - \beta_n v_n \le x_n$ and $y_{n-1} = x_n + sg_{n-1} \ge x_n$ show that

$$F(y_n) \le F(x_n) \le F(y_{n-1}).$$

Both bounding quantities divided by $Dn^{2-2\rho}$ tend to one, because $((n-1)/n)^{2-2\rho} \to 1$. This proves the same asymptotic for $F(x_n)$. The exponent conversion is

$$p - 2 = \frac{2}{\delta}, \qquad \frac{2p}{p-2} = p\delta = 2(\rho-1).$$

In particular, $F(y_n)/y_n^p \to D/A^p$. For $x \in [y_{n+1}, y_n]$, monotonicity gives

$$\frac{F(y_{n+1})}{y_n^p} \le \frac{F(x)}{x^p} \le \frac{F(y_n)}{y_{n+1}^p}.$$

For example, the left bound equals $[F(y_{n+1})/y_{n+1}^p](y_{n+1}/y_n)^p \to D/A^p$; the right bound has the same limit. This proves the asserted asymptotic for every $x \downarrow 0$, not just at the nodes. Choose $\varepsilon > 0$ such that $F(x) \ge [D/(2A^p)]x^p$ for $0 \le x \le \varepsilon$. By evenness, $F(x) \ge K|x|^p$ for $|x| \le \varepsilon$, with $K = D/(2A^p) > 0$. This is $L(p)$.

Similarly, on the same interval,

$$\frac{g_{n+1}}{y_n^{p-1}} \le \frac{\psi(x)}{x^{p-1}} \le \frac{g_n}{y_{n+1}^{p-1}}.$$

Since $\delta(p-1) = \rho$ and $y_{n+1}/y_n \to 1$, both bounds tend to $A^{-(p-1)}$. Consequently, as $x \downarrow 0$,

$$\frac{xF'(x)}{F(x)} \longrightarrow \frac{A^{-(p-1)}}{D/A^p} = \frac{A}{D} = p,$$

where $A/D = 2(\rho-1)/(\rho-2) = p$. Since F is even and F' is odd, the same limit holds as $x \uparrow 0$.

3. Bound all interpolation slopes Let ℓ_n denote the slope of ψ on $[y_{n+1}, y_n]$, as defined in (24). To enforce the prescribed smoothness, the recurrence gives $v_{n+2} = \beta_{n+1}v_{n+1} + sg_{n+1} \geq sg_{n+1}$. Thus the slope denominator obeys $v_{n+2} + s(g_n - g_{n+1}) \geq sg_n > 0$, and

$$0 \leq \ell_n \leq \frac{g_n - g_{n+1}}{sg_n} = \frac{1}{s}\left[1 - \left(\frac{n+M}{n+M+1}\right)^{\rho}\right] \leq \frac{\rho}{s(n+M)}.$$

The last inequality follows, with $t = n + M > 0$, from

$$1 - \left(\frac{t}{t+1}\right)^{\rho} = 1 - (1 + 1/t)^{-\rho} = \int_0^{1/t} \rho(1+u)^{-\rho-1}\, du \leq \frac{\rho}{t}.$$

Consequently, $M \geq \rho/(sL)$ gives $\ell_n \leq \rho/[s(n+M)] \leq \rho/(sM) \leq L$ on every interpolation interval. The construction sets $\psi(0) = 0$, continues ψ constantly as g_1 for $x \geq y_1$, and extends it oddly to the negative axis. Step 3 of the proof of Lemma 3 shows that these extensions preserve continuity, monotonicity, and the global L-Lipschitz bound, including at zero. It also proves that $F(x) = \int_0^x \psi(u)\, du$ is even and coercive, has unique minimizer zero, and realizes the prescribed CD orbit.

□

C.3 The first correction and endpoint flatness

The previous lemma implies $xF'(x)/F(x) \to p$, which suffices for $H(\beta)$ when $\beta < p$. The endpoint $H(p)$ needs more: the ratio must approach p from above. The shift M controls this sign without changing the leading exponent.

Lemma 6 (Endpoint flatness)**.** *Under the notation and assumptions of Lemma 5, choose $M \geq \alpha + \rho/(sL)$. Then the same loss satisfies $H(p)$, and hence $H(\beta)$ for every $2 < \beta \leq p$.*

Proof of Lemma 6. 1. Retain the first correction in the nodal data Write $K_\alpha = \alpha(\alpha-1)/2$. With M fixed, the gamma-ratio expansion (DLMF, Eq. 5.11.13) and Taylor expansion give

$$\frac{\Gamma(k+\alpha)}{\Gamma(k)} = k^{\alpha}\left(1 + \frac{K_\alpha}{k} + O(k^{-2})\right), \qquad (k+M)^{-\rho} = k^{-\rho}\left(1 - \frac{\rho M}{k} + O(k^{-2})\right).$$

Multiplying these expressions, and inverting the gamma ratio for the prefactor in the velocity formula, yields

$$\frac{\Gamma(k+\alpha)}{\Gamma(k)}(k+M)^{-\rho} = k^{a}\left(1 + \frac{K_\alpha - \rho M}{k} + O(k^{-2})\right),$$
$$\frac{\Gamma(n)}{\Gamma(n+\alpha)} = n^{-\alpha}\left(1 - \frac{K_\alpha}{n} + O(n^{-2})\right).$$

Because $a = \alpha - \rho > 0$, Euler–Maclaurin summation (DLMF, Eq. 2.10.1) gives

$$\sum_{k=1}^{n} k^a = \frac{n^{a+1}}{a+1} + \frac{1}{2}n^a + O(1 + n^{a-1}),$$
$$\sum_{k=1}^{n} k^{a-1} = \frac{n^a}{a} + o(n^a),$$
$$\sum_{k=1}^{n} O(k^{a-2}) = \begin{cases} O(1), & 0 < a < 1, \\ O(\log n), & a = 1, \\ O(n^{a-1}), & a > 1. \end{cases}$$

For the first line, the integral and upper-endpoint terms are displayed; the lower-endpoint and derivative-remainder terms are $O(1 + n^{a-1})$. The second line follows by integral comparison, and the third by summing the absolute remainder bound. Every remainder is $o(n^a)$ for the indicated range of a. Adding the leading and first-correction terms therefore yields

$$\sum_{k=1}^{n} \frac{\Gamma(k+\alpha)}{\Gamma(k)}(k+M)^{-\rho} = \frac{n^{a+1}}{a+1} + \left(\frac{1}{2} + \frac{K_\alpha - \rho M}{a}\right)n^a + o(n^a).$$

Multiplying this sum by $sn^{-\alpha}(1 - K_\alpha/n + O(n^{-2}))$ gives the leading coefficient $B = s/(a+1)$ and the relative n^{-1} coefficient $(a+1)[1/2 + (K_\alpha - \rho M)/a] - K_\alpha$. Thus

$$v_{n+1} = Bn^{-\delta-1}\left(1 + \frac{V_1}{n} + o(n^{-1})\right), \qquad V_1 = (a+1)\left(\frac{1}{2} + \frac{K_\alpha - \rho M}{a}\right) - K_\alpha. \tag{28}$$

For the position correction the starting index of the tail matters. Because $x_{n+1} = \sum_{m=n+1}^{\infty} v_{m+1}$, use

$$\sum_{m=n+1}^{\infty} m^{-\delta-1} = \frac{n^{-\delta}}{\delta} - \frac{1}{2}n^{-\delta-1} + O(n^{-\delta-2}),$$
$$\sum_{m=n+1}^{\infty} m^{-\delta-2} = \frac{n^{-\delta-1}}{\delta+1} + O(n^{-\delta-2}).$$

These are the tail forms of Euler–Maclaurin; the minus sign in the first endpoint correction reflects exclusion of $m = n$. The velocity remainder is $o(m^{-\delta-2})$, whose tail is $o(n^{-\delta-1})$: for any $\varepsilon > 0$, its absolute value is eventually bounded by $\varepsilon m^{-\delta-2}$, and the second tail estimate applies. Hence

$$x_{n+1} = \frac{B}{\delta}n^{-\delta} + B\left(-\frac{1}{2} + \frac{V_1}{\delta+1}\right)n^{-\delta-1} + o(n^{-\delta-1}).$$

Adding $sg_n = O(n^{-\delta-2}) = o(n^{-\delta-1})$ to get y_n and factoring out $An^{-\delta}$, we obtain

$$y_n = An^{-\delta}\left(1 + \frac{Y_1}{n} + o(n^{-1})\right), \qquad Y_1 = -\frac{\delta}{2} + \frac{\delta V_1}{\delta+1}. \tag{29}$$

Since $p - 1 = \rho/\delta$, Taylor expansion of the denominator gives

$$y_n^{p-1} = A^{p-1}n^{-\rho}\left(1 + \frac{(p-1)Y_1}{n} + o(n^{-1})\right).$$

Divide $g_n = n^{-\rho}(1 - \rho M/n + O(n^{-2}))$ by this expression, using $(1+t)^{-1} = 1 - t + O(t^2)$ as $t \to 0$. The resulting coefficient is $H_1 = -\rho M - (p-1)Y_1$, so

$$\frac{g_n}{y_n^{p-1}} = A^{-(p-1)}\left(1 + \frac{H_1}{n} + o(n^{-1})\right), \qquad H_1 = -\rho M + \frac{\rho}{2} - \frac{\rho V_1}{\rho - 1}. \tag{30}$$

2. Choose the sign through the shift First simplify the velocity coefficient to $V_1 = (a+1)/2 + K_\alpha/a - \rho M(a+1)/a$. Then

$$\begin{aligned}\frac{H_1}{\rho} &= \left[-1 + \frac{\rho(a+1)}{a(\rho-1)}\right]M + \frac{1}{2} - \frac{a+1}{2(\rho-1)} - \frac{K_\alpha}{a(\rho-1)} \\ &= \frac{\alpha M}{a(\rho-1)} + \frac{2\rho - \alpha - 2}{2(\rho-1)} - \frac{\alpha(\alpha-1)}{2a(\rho-1)}.\end{aligned}$$

Here the coefficient of M uses $a + \rho = \alpha$. Factoring out $\alpha/[a(\rho-1)]$ proves

$$H_1 = \frac{\alpha\rho}{a(\rho-1)}\left[M - \frac{\alpha-1}{2} - \frac{(\alpha-\rho)(\alpha+2-2\rho)}{2\alpha}\right].$$

The coefficient of M is positive. To bound the two subtracted terms, if $\alpha + 2 - 2\rho \le 0$ their sum is at most $(\alpha-1)/2 < \alpha$. Otherwise, $0 < a < \alpha$ and $\rho > 2$ imply

$$0 < a(\alpha + 2 - 2\rho) < \alpha(\alpha - 2), \qquad \frac{\alpha-1}{2} + \frac{a(\alpha+2-2\rho)}{2\alpha} < \frac{2\alpha-3}{2} < \alpha.$$

Thus in either case $M \ge \alpha + \rho/(sL)$ ensures both $H_1 > 0$ and the smoothness bound from Lemma 5.

3. Show that interpolation preserves the correction A nodal inequality alone would not establish $H(p)$ for every nearby point. We therefore compare the affine interpolant with the smooth function described by the first two terms of the nodal expansion. Set

$$\varepsilon = \delta^{-1}, \qquad C_0 = A^{-(p-1)}, \qquad c = H_1 A^{-\varepsilon} > 0.$$

Raising $y_n = An^{-\delta}(1 + O(n^{-1}))$ to the power $\varepsilon = 1/\delta$ gives $y_n^\varepsilon = A^\varepsilon n^{-1}(1 + O(n^{-1}))$. Thus $n^{-1} = A^{-\varepsilon} y_n^\varepsilon (1 + o(1))$, and (30) becomes

$$g_n = C_0 y_n^{p-1} \left(1 + c y_n^\varepsilon + o(y_n^\varepsilon)\right).$$

Let $h(x) = C_0 x^{p-1}(1 + cx^\varepsilon)$ for $x > 0$. On the interval $I_n = [y_{n+1}, y_n]$, the leading mesh estimate gives

$$\frac{|I_n|}{y_n} = \frac{y_n - y_{n+1}}{y_n} \sim \frac{B}{An} = \frac{\delta}{n}, \qquad \frac{y_{n+1}}{y_n} \to 1.$$

Write $e_n = g_n - h(y_n) = o(y_n^{p-1}/n)$. At $x = (1-t)y_{n+1} + ty_n$, $0 \le t \le 1$, let $\mathcal{I}_n h(x) = (1-t)h(y_{n+1}) + th(y_n)$. Since ψ uses the same affine weights,

$$|\psi(x) - \mathcal{I}_n h(x)| \le \max\{|e_n|, |e_{n+1}|\} = o(y_n^{p-1}/n)$$

uniformly on I_n. The last equality uses the adjacent-node ratio and $(n+1)/n \to 1$. Also

$$h''(x) = C_0(p-1)(p-2)x^{p-3} + C_0 c(p-1+\varepsilon)(p-2+\varepsilon)x^{p-3+\varepsilon}.$$

On each sufficiently late I_n, $y_n/2 \le x \le y_n$, so $\sup_{x \in I_n} |h''(x)| = O(y_n^{p-3})$, with an n-independent constant. The classical linear interpolation remainder (DLMF, Eq. 3.3.5) gives $h(x) - \mathcal{I}_n h(x) = h''(\zeta_x)(x - y_{n+1})(x - y_n)/2$ for an intermediate point $\zeta_x \in I_n$. Since $|(x - y_{n+1})(x - y_n)| \le |I_n|^2/4$, it bounds the error by

$$\frac{1}{8} \sup_{x \in I_n} |h''(x)|\, |I_n|^2 = O(y_n^{p-1} n^{-2}) = o(y_n^{p-1}/n).$$

Thus the interpolation error is one order smaller than the positive correction. The estimate remains valid for $2 < p < 3$, when h'' may be unbounded at zero, because it is used separately on each I_n, where x is comparable to y_n. Combining the two error bounds gives $\sup_{x \in I_n} |\psi(x) - h(x)| = o(y_n^{p-1}/n)$. Uniformly on I_n, $x/y_n \to 1$ and $nx^\varepsilon \to A^\varepsilon > 0$; dividing by $x^{p-1+\varepsilon}$ therefore makes this error tend to zero. It follows, uniformly as $x \downarrow 0$, that

$$\psi(x) = C_0 x^{p-1} \left(1 + cx^\varepsilon + o(x^\varepsilon)\right). \tag{31}$$

4. Integrate and verify the endpoint inequality To justify integration of the remainder, write $\psi(x) = C_0 x^{p-1} + C_0 c x^{p-1+\varepsilon} + x^{p-1+\varepsilon}\eta(x)$, where $\eta(x) \to 0$. Then

$$\left| \int_0^x u^{p-1+\varepsilon} \eta(u)\, du \right| \le \sup_{0 < u \le x} |\eta(u)| \, \frac{x^{p+\varepsilon}}{p+\varepsilon} = o(x^{p+\varepsilon}).$$

Integrating the two explicit powers therefore gives

$$F(x) = \frac{C_0}{p} x^p \left(1 + \frac{cp}{p+\varepsilon} x^\varepsilon + o(x^\varepsilon)\right).$$

Divide the expansion of $xF'(x)$ by this expression, again using $(1+t)^{-1} = 1 - t + O(t^2)$. The relative correction coefficient is $c - cp/(p+\varepsilon) = c\varepsilon/(p+\varepsilon) > 0$, so

$$\frac{xF'(x)}{F(x)} = p\left(1 + \frac{c\varepsilon}{p+\varepsilon} x^\varepsilon + o(x^\varepsilon)\right) > p$$

for all sufficiently small $x > 0$: the $o(x^\varepsilon)$ term can be bounded in absolute value by half of the displayed positive correction. Evenness gives $xF'(x) = |x|F'(|x|)$ for $x < 0$, proving the same inequality for negative x, while both sides of $pF(x) \le xF'(x)$ vanish at zero. This proves $H(p)$. Since $F(x) \ge 0$ and $\beta \le p$, $\beta F(x) \le pF(x) \le xF'(x)$ on the same neighborhood, proving every required $H(\beta)$.

□

C.4 Preserving the geometry in a prescribed dimension

The quadratic transverse terms used for Theorem 1 would violate $H(p)$ when $p > 2$. Here the extra coordinates must instead have local p-power geometry. Capping their derivatives away from zero enforces global smoothness while retaining coercivity.

Lemma 7 (Finite-dimensional power witness)**.** *Let F_1 be the scalar loss from Lemma 6. For every fixed $d \geq 1$, there is a coercive convex loss $\widetilde{F} : \mathbb{R}^d \to \mathbb{R}$ with L-Lipschitz gradient, unique minimizer zero, and local conditions $L(p)$ and $H(p)$, whose exact CD orbit from a suitable initialization satisfies $\widetilde{F}(z_n) \sim Dn^{-2p/(p-2)}$ with the same $D > 0$.*

Proof of Lemma 7. The case $d = 1$ is Lemmas 5 and 6. Choose $r_1, K_1 > 0$ so that both $F_1(t) \geq K_1|t|^p$ and $tF_1'(t) \geq pF_1(t)$ hold for $|t| \leq r_1$. For $d > 1$, choose $R, \kappa > 0$ with $\kappa(p-1)R^{p-2} \leq L$, and define $\varphi(0) = 0$ and

$$\varphi'(t) = \kappa \, \mathrm{sgn}(t) \min\{|t|^{p-1}, R^{p-1}\}, \qquad \widetilde{F}(z) = F_1(z_1) + \sum_{i=2}^{d} \varphi(z_i).$$

For example, $R = 1$ and $\kappa = L/(p-1)$ meet this condition. Integrating the derivative gives the explicit extension

$$\varphi(t) = \begin{cases} \kappa|t|^p/p, & |t| \leq R, \\ \kappa R^p/p + \kappa R^{p-1}(|t| - R), & |t| \geq R. \end{cases}$$

Its derivative is continuous and nondecreasing. Away from the two joins, its slope is $\kappa(p-1)|t|^{p-2} \leq L$ inside $[-R, R]$ and zero outside. Splitting an interval at the joins proves the global L-Lipschitz bound. Thus φ is convex, coercive, nonnegative, and vanishes only at zero. For any prescribed height $H > 0$, choose a common radius beyond which each of the finitely many summands exceeds H. If $\|z\|_2$ exceeds $\sqrt{d}$ times that radius, at least one coordinate does, so $\widetilde{F}(z) > H$. This proves coercivity of the sum; its unique minimizer is zero because every summand is nonnegative and vanishes only there. Since F_1' and φ' are both L-Lipschitz, for all $z, w \in \mathbb{R}^d$,

$$\|\nabla\widetilde{F}(z) - \nabla\widetilde{F}(w)\|^2 \leq L^2 \sum_{i=1}^{d} |z_i - w_i|^2 = L^2\|z - w\|^2.$$

For $\|z\|_2 \leq \min\{r_1, R\}$, all scalar local inequalities apply, and $z_i\varphi'(z_i) = p\varphi(z_i)$ for $i \geq 2$. The norm comparison follows from convexity of $t \mapsto t^{p/2}$:

$$\left(\frac{1}{d}\sum_{i=1}^{d} |z_i|^2\right)^{p/2} \leq \frac{1}{d}\sum_{i=1}^{d} |z_i|^p, \qquad \sum_{i=1}^{d} |z_i|^p \geq d^{1-p/2}\|z\|_2^p.$$

Combining the scalar growth bounds with this comparison, and summing the scalar flatness inequalities, gives

$$\widetilde{F}(z) \geq d^{1-p/2} \min\{K_1, \kappa/p\}\|z\|_2^p, \qquad \langle\nabla\widetilde{F}(z), z\rangle \geq p\widetilde{F}(z).$$

Consequently, $L(p)$ and $H(p)$ hold in every prescribed finite dimension. Initialize $z_0 = z_1 = (x_1, 0, \ldots, 0)$, where x_1 is the scalar initial point. If two consecutive values of coordinate $i \geq 2$ are zero, its extrapolated value is zero and $\varphi'(0) = 0$ keeps the next value zero. Induction therefore gives $z_n = (x_n, 0, \ldots, 0)$ and $\widetilde{F}(z_n) = F_1(x_n)$ for all n. Equation (27) and the interlacing argument therefore give the claimed exact asymptotic. □

C.5 Completion of the sharpness statement

For either parameter regime in Appendix A.3, the choice $\rho = 2 + 2/(p-2)$ satisfies $2 < \rho < \alpha$, as verified in (26). Choose $M \geq \alpha + \rho/(sL)$ and apply Lemmas 5 to 7. The resulting loss has all the required properties. In regime (i), $H(p)$ implies the prescribed $H(\beta)$; in regime (ii), the unique minimizer makes $L(p)$ identical to $L_{\mathrm{set}}(p)$. For every gain $G(n) \to \infty$, the exact asymptotic gives

$$n^{2p/(p-2)}G(n)\big(\widetilde{F}(z_n) - \widetilde{F}^*\big) \sim DG(n) \longrightarrow \infty.$$

Thus no divergent gain improves the rate, proving the sharpness assertion of Theorem 2. The upper bounds for the whole class are established in Appendix D; their proofs are independent of this construction.

C.6 Illustrating the two constructions

Figure 5 shows concrete instances of the block and power constructions developed in Appendices B and C. Each row follows the same sequence: prescribe the queried gradients, recover the query points and interpolate the derivative, then integrate to obtain the loss. The plots illustrate the functions whose properties were proved above. The figure generator records the recurrence checks and bounds on the omitted integration tails; Appendix G describes the accompanying code and outputs.

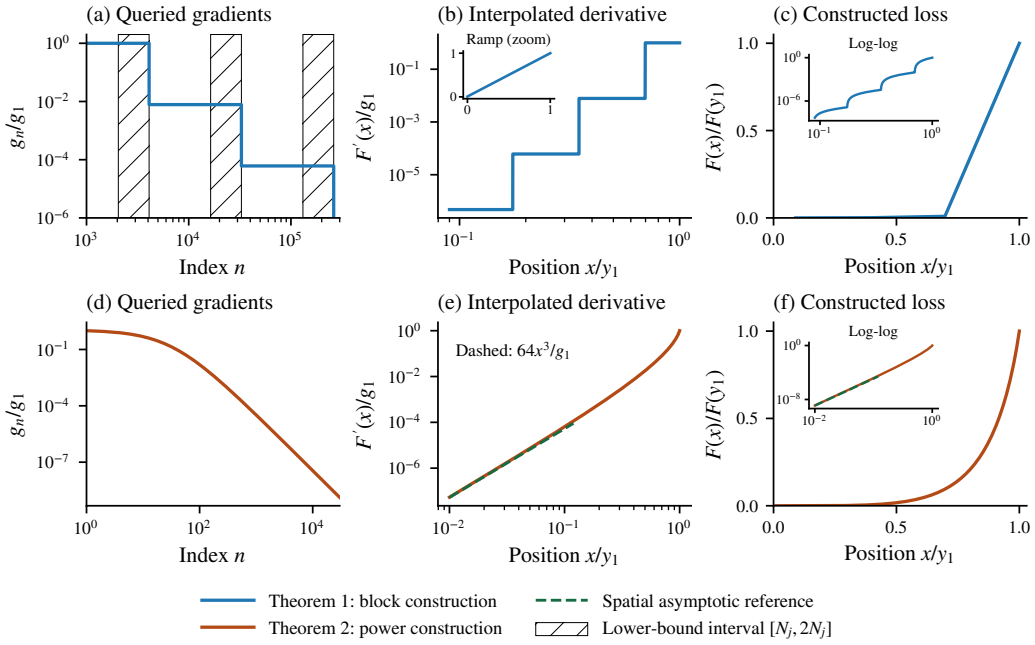


Figure 5: Two choices of gradients and the losses they produce, with $\alpha = 4$, $s = 1/2$, and $L = 1$. Top: a no-gain instance for $G(n) = n$, using $N_j = 2048 \cdot 8^{j-1}$ and $a_j = 2^{-j}$. Hatching marks the intervals $[N_j, 2N_j]$ used in the lower bound. The inset in (b) magnifies one continuous affine join in F'. Bottom: the power instance with $p = 4$ and $M = 32$, in the flatness regime of Theorem 2. Panels (c) and (f) use linear axes to display convexity; their insets retain log–log views of the same curves. The other main panels use log–log axes. Dashed curves in (e) and the inset of (f) show the spatial asymptotics $F'(x) \sim 64x^3$ and $F(x) \sim 16x^4$, respectively.

D Local-geometry upper bounds

This section proves that the power instances attain rates valid for every objective in the stated classes. The rate mechanisms come from Apidopoulos et al. (2021) and Aujol et al. (2024). Our technical extension is to remove the coercivity assumption in the growth-only result in finite dimension. We first match the iteration conventions and establish eventual local geometry, then give the two upper-bound arguments separately.

D.1 Matching the iteration conventions

The upper-rate analyses of Apidopoulos et al. (2021) and Aujol et al. (2024) use a different iteration index. We shift our index to match their recurrence and apply the published energy estimates to the same CD orbit.

Let the setting and recurrence be those of Appendix A.2. Let $T_s = I - s\nabla F$, set $z_{-1} = x_1$, and put $z_m = x_{m+1}$ for $m \geq 0$. Since $\beta_{m+1} = m/(m+\alpha)$, for $m \geq 1$ we have $y_{m+1} = z_m + [m/(m+\alpha)](z_m - z_{m-1})$ and $z_{m+1} = T_s(y_{m+1})$. For $m = 0$, both expressions reduce to

$y_1 = x_1 = z_0$ and $z_1 = T_s(z_0)$ because $\beta_1 = 0$. Thus at every index,

$$z_{m+1} = T_s\left(z_m + \frac{m}{m+\alpha}(z_m - z_{m-1})\right), \qquad m \geq 0. \tag{32}$$

The coefficient is exactly $m/(m+\alpha)$, as in the cited upper-rate arguments; in Apidopoulos et al. (2021), the damping parameter is denoted b, so $b = \alpha$. In particular, the gradient is evaluated at the extrapolated point inside T_s. Our recurrence starts at $m = 0$ with $z_{-1} = z_0$ and zero inertial coefficient. The subsequent energy inequalities apply to the resulting finite initial states z_0, z_1; their validity does not require these states to coincide. Because $\beta_1 = 0$, an arbitrary x_0 has no effect on the original iterates beginning with x_1.

D.2 Local geometry along the convergent orbit

The local assumptions need only hold eventually, but this must be justified for both the iterates and the extrapolated query points. Finite-dimensional convergence provides that localization even when the objective is not coercive.

Lemma 8 (Eventual local geometry)**.** *Let $F : \mathbb{R}^d \to \mathbb{R}$ be convex, attain its minimum, and have a globally L-Lipschitz gradient, where $d \geq 1$ is finite and $L > 0$. For $\alpha > 3$ and a fixed step $0 < s < L^{-1}$, let (x_n) and (y_n) be generated by the CD iteration* (15) *from any $x_0, x_1 \in \mathbb{R}^d$. Then there is a minimizer $\bar{x}$ such that $x_n \to \bar{x}$ and $y_n \to \bar{x}$ in Euclidean norm. Consequently, if $H(\beta)$ and $L(p)$ hold at the unique minimizer, they hold at all sufficiently late iterates and query points. If $L_{\mathrm{set}}(p)$ holds, its inequality likewise applies at all sufficiently late iterates and query points, without uniqueness or coercivity.*

Proof of Lemma 8. The CD convergence theorem (Chambolle & Dossal, 2015) gives weak convergence of (x_n) to a minimizer for $\alpha > 3$. In finite dimension this is norm convergence. Since $x_n - x_{n-1} \to 0$ and $0 \leq \beta_n < 1$, the query points satisfy

$$\begin{aligned} \|y_n - \bar{x}\| &\leq \|x_n - \bar{x}\| + \beta_n \|x_n - x_{n-1}\| \\ &\leq 2\|x_n - \bar{x}\| + \|x_{n-1} - \bar{x}\| \longrightarrow 0. \end{aligned}$$

For a unique minimizer, both sequences therefore enter and remain in each neighborhood in (16)–(17). In the set-valued case, $\operatorname{dist}(x_n, X^*) \leq \|x_n - \bar{x}\| \to 0$, and the same holds for y_n. This proves eventual validity of (18). □

We apply the published asymptotic energy inequalities from this eventual index onward. The original momentum coefficients $m/(m+\alpha)$ are retained throughout; localization does not restart the algorithm. A finite prefix affects only the constants in the estimates.

D.3 Upper rate under flatness and growth

Numbered references to Apidopoulos et al. (2021) in this subsection use the authors' HAL version 3. For a unique minimizer, their Theorem 3.2 uses the local conditions (16)–(17). In our notation it gives $F(z_m) - F^* = O(m^{-2p/(p-2)})$ whenever $p \geq \beta > 2$ and

$$\alpha \geq \frac{\beta + 2}{\beta - 2}.$$

That result uses the same exponents β, p, and its step γ is our s. The result allows $0 < s \leq 1/L$, so our strict step restriction is included. Its damping threshold includes equality, as also recorded in Aujol et al. (2024, Table 1). By Lemma 8, the local inequalities are valid along the required tail. The cited proof uses an eventual one-step energy estimate (their Lemma 4.3 and equation (4.28)), then sums it from an index at which the local conditions hold. That one-step estimate depends on the current states and the coefficient $m/(m+\alpha)$, not on equality of the two initial states. Its initial energy is finite for our z_0, z_1, so the same argument applies to the initialization in (32).

To spell out the final index conversion, put $q = 2p/(p-2)$ and choose C, N such that $F(z_m) - F^* \leq Cm^{-q}$ for $m \geq N$. For $n \geq \max\{N+1, 2\}$,

$$F(x_n) - F^* = F(z_{n-1}) - F^* \leq C(n-1)^{-q} \leq 2^q C n^{-q},$$

because $n - 1 \geq n/2$. This proves part (i) of Theorem 2.

D.4 Upper rate under growth alone, without coercivity

Here the minimizer set may have more than one point. The energy uses distance to that set and does not require flatness. There are two issues to check: how local growth becomes available without coercivity, and how to extract a residual bound from an energy with a negative distance term.

1. Replace coercivity by convergence in the local-growth step Set $\widehat{L} = 1/s$. Since the gradient is L-Lipschitz and $L < \widehat{L}$, it is also $\widehat{L}$-Lipschitz. The step $s = 1/\widehat{L}$ is the one used in Aujol et al. (2024, Theorem 1), with nonsmooth term zero. Their coercivity assumption is used in Lemma 4, through Lemma 1, to obtain eventual validity of the local growth inequality along the orbit. Lemma 8 already gives that conclusion for the present finite-dimensional CD orbit. The following calculation records the remainder of the rate argument without coercivity.

2. Relate distance to residual and use the published energy estimate Let z_m^* be the Euclidean projection of z_m onto the closed convex set X^*, and define

$$w_m = \frac{2}{\widehat{L}}\big(F(z_m) - F^*\big), \qquad h_m = \|z_m - z_m^*\|^2, \qquad b_m = \frac{m}{m+\alpha}, \qquad \theta = 1 + \frac{4}{p-2}.$$

By $L_{\mathrm{set}}(p)$ and $\mathrm{dist}(z_m, X^*) \to 0$, there is an index N such that $K h_m^{p/2} \le F(z_m) - F^* = (\widehat{L}/2) w_m$ for $m \ge N$. Both sides are nonnegative, so taking the power $2/p$ gives

$$h_m \le C_K w_m^{2/p} \quad (m \ge N), \qquad C_K = \left(\frac{\widehat{L}}{2K}\right)^{2/p}. \tag{33}$$

This is the estimate required in equations (49)–(50) of Aujol et al. (2024).

Put $\lambda = \alpha - 1 - \theta > 0$ and $\xi = \lambda(\lambda + 1 - \alpha) = -\lambda\theta < 0$, and consider

$$\begin{aligned} E_m = m^2 w_m + \|\lambda(z_m - z_m^*) + m b_m (z_m - z_{m-1})\|^2 + \xi h_m \\ + \lambda m b_m^2 \|z_m - z_{m-1}\|^2, \qquad J_m = m^\theta E_m. \end{aligned}$$

We use the one-step energy estimate of Aujol et al. (2024, Lemma 6, Eq. (52)) as a cited input. To verify its applicability and the signs needed below, first note the two elementary inequalities underlying that estimate. If $u^+ = y - s\nabla F(y)$, smoothness with constant $\widehat{L} = 1/s$ and convexity imply, for any x,

$$\begin{aligned} F(u^+) - F(x) &\le \langle \nabla F(y), y - x\rangle - \frac{s}{2}\|\nabla F(y)\|^2 \\ &= \frac{1}{2s}\big(\|y - x\|^2 - \|u^+ - x\|^2\big). \end{aligned}$$

Here the equality follows by expanding $\|y - s\nabla F(y) - x\|^2$. Using $y = z_m + b_m(z_m - z_{m-1})$, first with $x = z_m$ and then with $x = z_m^*$, gives respectively

$$\begin{aligned} w_{m+1} - w_m &\le b_m^2 \|z_m - z_{m-1}\|^2 - \|z_{m+1} - z_m\|^2, \\ w_{m+1} &\le \|y - z_m^*\|^2 - \|z_{m+1} - z_m^*\|^2. \end{aligned}$$

Also, projection onto the closed convex minimizer set gives $\langle z_m - z_m^*, v - z_m^*\rangle \le 0$ for every $v \in X^*$. This follows by minimizing squared distance along $z_m^* + t(v - z_m^*)$ and taking the right derivative at $t = 0$. In particular, with

$$\Delta_m^* = z_{m+1}^* - z_m^*, \qquad Q_m = \|\Delta_m^*\|^2 - 2\langle z_m - z_m^*, \Delta_m^*\rangle,$$

we have $Q_m \ge 0$. These facts require neither coercivity nor a unique minimizer.

To state the cited estimate without confusing its notation with ours, put $U_m = \|\lambda(z_m - z_m^*) + m b_m (z_m - z_{m-1})\|^2$. The source's growth exponent γ is our p, its energy weight p is our θ, its inertial coefficient α_m is our b_m, and its squared-norm term b_m is our U_m. Its estimate reads

$$\begin{aligned} J_{m+1} - J_m \le [C_1(m+1)^{\theta+1} + R_1(m)] w_{m+1} \\ + [C_2(m+1)^{\theta-1} + R_2(m)] U_{m+1} \end{aligned}$$

$$+ [C_3(m+1)^{\theta-1} + R_3(m)]h_{m+1} - m^\theta B_4(m) Q_m,$$

where $R_1(m) = O(m^\theta)$, $R_2(m), R_3(m) = O(m^{\theta-2})$, and substitution of $\lambda = \alpha - 1 - \theta$ gives

$$C_1 = 2 - \lambda + \theta = 3 + 2\theta - \alpha < 0,$$
$$C_2 = 2(\lambda + 1 - \alpha) + \theta = -\theta < 0,$$
$$C_3 = \lambda(\lambda + 1 - \alpha)(\theta - 2\lambda) = \lambda\theta(2\lambda - \theta) > 0,$$
$$B_4(m) = 2\lambda\theta - \frac{\alpha^2\lambda}{m + 1 + \alpha} \longrightarrow 2\lambda\theta > 0.$$

The damping assumption is exactly $\alpha > 3 + 2\theta = 5 + 8/(p-2)$; it gives $\lambda > \theta + 2 > 0$, so all the displayed signs follow. Since each remainder is of lower order than its leading coefficient, enlarge N until

$$C_1(m+1)^{\theta+1} + R_1(m) \le \tfrac{1}{2} C_1 (m+1)^{\theta+1},$$
$$C_2(m+1)^{\theta-1} + R_2(m) \le 0,$$
$$C_3(m+1)^{\theta-1} + R_3(m) \le 2C_3(m+1)^{\theta-1}, \qquad B_4(m) \ge 0.$$

The U_{m+1} and Q_m terms are then nonpositive and can be dropped. With $c_0 = -C_1/2 > 0$ and $c_1 = 2C_3 > 0$, this proves

$$J_{m+1} - J_m \le -c_0(m+1)^{\theta+1} w_{m+1} + c_1(m+1)^{\theta-1} h_{m+1}. \tag{34}$$

This is the source's equation (57), with its hypotheses and coefficient choices made explicit.

3. Bound the energy increments The exponent identity needed for the substitution is

$$\theta - 1 = \frac{4}{p-2}, \qquad \theta + 1 = 2 + \frac{4}{p-2} = \frac{2p}{p-2}, \qquad \frac{2(\theta+1)}{p} = \theta - 1.$$

Thus substituting (33) into (34) bounds its right side by $-c_0 u + c_1 C_K u^{2/p}$, where $u = (m+1)^{\theta+1} w_{m+1}$. In detail, $(m+1)^{\theta-1} w_{m+1}^{2/p} = u^{2/p}$ by the displayed identity. To bound the scalar expression explicitly, let $\nu = 2/p \in (0,1)$, $A_0 = c_1 C_K > 0$, and $T_0 = (2A_0/c_0)^{1/(1-\nu)}$. For $u \ge T_0$, $A_0 u^\nu \le (c_0/2)u$ and the expression is nonpositive; for $0 \le u \le T_0$, it is at most $M_0 := A_0 T_0^\nu$. Therefore

$$J_m = J_N + \sum_{j=N}^{m-1} (J_{j+1} - J_j) \le J_N + M_0(m - N) \le Cm \qquad (m \ge N),$$

where $N \ge 1$ and, for example, $C = M_0 + |J_N| + 1 > 0$ suffices.

4. Control the negative term and recover the residual rate An upper bound on J_m alone is not yet a residual bound because $\xi < 0$. All terms in E_m other than ξh_m are nonnegative, so dividing $J_m \le Cm$ by $m > 0$ and using $J_m = m^\theta E_m$ gives

$$m^{\theta+1} w_m - |\xi| m^{\theta-1} h_m \le C.$$

Using (33) once more, we obtain

$$X_m - |\xi| C_K X_m^{2/p} \le C, \qquad X_m = m^{\theta+1} w_m.$$

Indeed, $h_m \le C_K w_m^\nu$ implies $-|\xi| m^{\theta-1} h_m \ge -|\xi| C_K X_m^\nu$, so replacing the negative term preserves the stated upper inequality for the smaller expression. To make boundedness explicit, put $B_0 = |\xi| C_K$ and $V_0 = (2B_0)^{1/(1-\nu)}$. If $X_m \ge V_0$, then

$$B_0 X_m^\nu \le \tfrac{1}{2} X_m, \qquad \tfrac{1}{2} X_m \le X_m - B_0 X_m^\nu \le C.$$

Consequently $X_m \le \max\{V_0, 2C\}$ in both cases. Since $F(z_m) - F^* = (\widehat{L}/2) w_m$, we conclude

$$F(z_m) - F^* = O(m^{-\theta-1}) = O(m^{-2p/(p-2)}).$$

This reproduces the rate step in equations (58)–(62) of Aujol et al. (2024) with eventual growth supplied by finite-dimensional CD convergence. It proves part (ii), including noncoercive objectives and non-singleton minimizer sets. Finally, the fixed shift $z_m = x_{m+1}$ does not change either upper-rate exponent.

E THE QUADRATIC BOUNDARY

The power constructions require $p > 2$. To illustrate the different behavior at the quadratic boundary, we give an explicit quadratic loss and prove a geometric bound for its exact discrete CD orbit. The same argument then covers every fixed convex quadratic in finite dimension.

E.1 THE LOSS AND THE CLAIMED RATE

Fix $0 < \lambda \le L$ and consider the scalar loss

$$F_\lambda(x) = \frac{\lambda}{2}x^2, \qquad x \in \mathbb{R}.$$

It is smooth, convex, coercive, and uniquely minimized at zero, with $F_\lambda^* = 0$ and $F_\lambda'(x) = \lambda x$. Run CD with the same damping $\alpha > 3$ and step $0 < s < L^{-1}$ as in the main results, from $x_0 = x_1 = R > 0$. We will show that, for every fixed $\theta \in (1 - s\lambda, 1)$, there is a constant $C > 0$ such that

$$F_\lambda(x_n) \le C\theta^n \qquad (n \ge 1).$$

This is the geometric decay claimed here: the error is bounded by an exponentially decreasing envelope, even if it oscillates between steps. For example, $F(x) = x^2/2$, $\alpha = 4$, $s = 1/2$, and $x_0 = x_1 = 1$ give $F(x_n) = O((3/4)^n)$.

E.2 WHY THE EXACT CD RECURRENCE CONTRACTS

For this loss, the gradient step simply multiplies the query point by $a := 1 - s\lambda \in (0, 1)$. Substituting $F_\lambda'(y_n) = \lambda y_n$ into the CD update gives

$$x_{n+1} = ay_n = a\big((1 + \beta_n)x_n - \beta_n x_{n-1}\big), \qquad \beta_n = \frac{n-1}{n+\alpha-1}.$$

Because momentum involves two successive iterates, we track their joint state $u_n = (x_n, x_{n-1})^\mathsf{T}$. It satisfies

$$u_{n+1} = M_n u_n, \qquad M_n = \begin{pmatrix} a(1+\beta_n) & -a\beta_n \\ 1 & 0 \end{pmatrix} \longrightarrow M = \begin{pmatrix} 2a & -a \\ 1 & 0 \end{pmatrix}.$$

The key is to find one fixed norm in which all sufficiently late updates contract. This is a standard adapted-norm argument (Horn & Johnson, 2012, Chapter 5); here the norm can be written explicitly:

$$\|(z, w)\|_a^2 := (1 - a)z^2 + a(z - w)^2.$$

Since $0 < a < 1$, this positive definite quadratic form defines a norm. Indeed, it is the Euclidean norm after the invertible linear map $(z, w) \mapsto (\sqrt{1-a}\, z, \sqrt{a}\,(z - w))$. For $u = (z, w)^\mathsf{T}$, substitute $Mu = (2az - aw, z)^\mathsf{T}$ and expand:

$$\begin{aligned} \|Mu\|_a^2 &= (1 - a)(2az - aw)^2 + a((2a - 1)z - aw)^2 \\ &= az^2 - 2a^2zw + a^2w^2 = a\big((1 - a)z^2 + a(z - w)^2\big) = a\|u\|_a^2. \end{aligned}$$

Thus the limiting update contracts by exactly $\sqrt{a} < 1$ in this norm. To transfer this contraction to the varying matrices, observe that

$$(M_n - M)u = (a(\beta_n - 1)(z - w), 0)^\mathsf{T}.$$

The norm of $(t, 0)$ is $|t|$, and $a(z - w)^2 \le \|u\|_a^2$. Consequently,

$$\|(M_n - M)u\|_a \le \sqrt{a}\,(1 - \beta_n)\|u\|_a = \frac{\alpha\sqrt{a}}{n+\alpha-1}\|u\|_a,$$

$$\|M_n u\|_a \le \left(\sqrt{a} + \frac{\alpha\sqrt{a}}{n+\alpha-1}\right)\|u\|_a.$$

Fix $\theta \in (a, 1)$ and put $r = \sqrt{\theta} > \sqrt{a}$. Choose $N \ge 1$ with $\alpha\sqrt{a}/(N + \alpha - 1) \le r - \sqrt{a}$. Then $\|M_n u\|_a \le r\|u\|_a$ for every u and $n \ge N$. Applying this inequality successively gives

$$\|u_n\|_a \le r^{n-N}\|u_N\|_a \qquad (n \ge N).$$

Finally, $(1-a)x_n^2 \le \|u_n\|_a^2$, so the state contraction controls the loss itself:

$$F_\lambda(x_n) = \frac{\lambda}{2}x_n^2 \le \frac{\lambda\|u_N\|_a^2}{2(1-a)}\,\theta^{n-N} \qquad (n \ge N).$$

For example, take C to be the maximum of $\lambda\|u_N\|_a^2\theta^{-N}/[2(1-a)]$ and the finite set $\{F_\lambda(x_n)\theta^{-n} : 1 \le n \le N\}$. Then the claimed bound holds for every $n \ge 1$.

E.3 Finite-dimensional quadratics and the scope of the argument

More generally, take

$$F(x) = F^* + \frac{1}{2}(x-x^*)^{\mathsf{T}}Q(x-x^*),$$

where Q is a fixed symmetric positive semidefinite matrix whose eigenvalues λ_i lie in $[0, L]$. In an orthonormal eigenbasis, write $z_{i,n}$ for the coordinates of $x_n - x^*$. Orthogonal change of coordinates turns $\nabla F(x) = Q(x-x^*)$ into coordinatewise multiplication by λ_i. Thus

$$z_{i,n+1} = (1-s\lambda_i)\big((1+\beta_n)z_{i,n} - \beta_n z_{i,n-1}\big),$$

and the objective is

$$F(x_n) - F^* = \frac{1}{2}\sum_{i=1}^{d}\lambda_i z_{i,n}^2.$$

Zero eigenvalues contribute no error. If $Q \neq 0$, let λ_+ be its smallest positive eigenvalue. For any $\theta \in (1-s\lambda_+, 1)$, the scalar argument bounds each positive term by $C_i\theta^n$: indeed, for $\lambda_i > 0$, $a_i = 1 - s\lambda_i \le 1 - s\lambda_+ < \theta$. The scalar proof applies to any finite initial state u_1, so no sign restriction on these coordinates is needed. Summing the finitely many bounds gives

$$F(x_n) - F^* \le \left(\sum_{i:\lambda_i>0} C_i\right)\theta^n.$$

If $Q = 0$, the error is identically zero.

Consequently, fixed finite-dimensional quadratics cannot attain the nonzero polynomial asymptotic of Theorem 2. This calculation concerns quadratic losses themselves; it does not classify all smooth convex losses satisfying a local quadratic growth inequality.

F Forward CD experiments on the fixed witnesses

The experiments run CD on the scalar losses constructed in Appendices B and C. They display repeated exceedances of the subsequence lower bound and agreement with the predicted power constant. The infinitely-many-indices and arbitrary-gain statements are established by the proofs; the numerical runs cover the finite horizons stated below. Appendix G describes the accompanying code, sampled traces, numerical checks, and reproduction commands.

F.1 Fixed losses and independent gradient queries

Every run uses $L = 1$, $s = 1/2$, and $x_0 = x_1 = 1$. Starting from the inverse construction, set $\lambda = 1/x_1$ and replace the unscaled loss by $\lambda^2 F(\cdot/\lambda)$; below, F denotes this rescaled loss, whose minimum value is zero. The derivative interpolant is fixed before optimization. Gradient evaluations use the current forward query point; constant-gradient intervals also permit the batched updates described below. The prescribed gradient sequence and inverse-orbit positions are used only for building the loss and checking the resulting run.

The query mesh incorporates the infinite weighted gradient tail. For the block instance a geometric sum captures its leading part, with an explicit bound on the remaining correction. For the power instances Euler–Maclaurin summation supplies the tail and a remainder bound. Additional nodes below the queried region are used to integrate the objective. If y_{cut} is the last integration node, omitting the area below it gives the truncation estimate

$$0 \le F(x) - \widehat{F}(x) \le y_{\mathrm{cut}}F'(y_{\mathrm{cut}}), \qquad x \ge y_{\mathrm{cut}}.$$

In fact, exact integration over the retained affine pieces gives $\widehat{F}(x) = \int_{y_{\mathrm{cut}}}^{x} F'(u)\,du$, so

$$F(x) - \widehat{F}(x) = \int_{0}^{y_{\mathrm{cut}}} F'(u)\,du.$$

The bound follows from $0 \le F'(u) \le F'(y_{\mathrm{cut}})$ on $[0, y_{\mathrm{cut}}]$. It controls truncation of the analytic integral; floating-point roundoff is assessed separately by the precision comparisons below. All gradient evaluations remain inside the retained mesh or its prescribed constant outer extension. The implementation stops if a query leaves this region toward zero. No completion of the loss below the retained mesh is needed for the runs.

The power runs use extended precision with 64 mantissa bits, with a second run using float64 optimizer states on the same oracle. Their largest sampled relative position discrepancy from the inverse-orbit reference is 1.23×10^{-17}, and the largest relative objective difference between precisions is 1.02×10^{-13}. The block run uses arbitrary precision and batches of the discrete recurrence, as described below. Precision comparisons are measured consistency checks, separate from the analytic tail bounds and not interval-arithmetic certificates.

F.2 Theorem 1: repeated exceedances with a double-logarithmic gain

We take $G(n) = \log\log(e^e + n)$ and $\alpha = 4$. There is no slowest divergent gain; this choice grows more slowly than every positive power of $\log n$. To display several episodes, use the summable amplitudes $a_j = (1-r)r^{j-1}$ with $r = 0.999$, in the same block construction:

$$N_0 = 0, \qquad N_j = 2^{\left\lceil \exp\left(2jr^{-2(j-1)}\right)/\log 2 \right\rceil}, \qquad g_k = \frac{a_j}{N_j^2} \quad (2N_{j-1} < k \le 2N_j).$$

The weighted-summability and transition-slope estimates in Appendix B.3 apply unchanged to these amplitudes. The scales satisfy the separation and first-scale conditions, and $G(N_j)a_j^2 \ge 2j(1-r)^2$ for every j. Thus the single fixed loss has the infinite sequence of bounds

$$\frac{N_j^2 G(N_j) F(x_{N_j})}{K} \ge 2j, \qquad K = \lambda^2 \frac{se^{-\alpha}}{6}(1-r)^2 > 0.$$

Here is the verification of the scale conditions. Write $b_j = 2jr^{-2(j-1)}$ and $q_j = e^{b_j}/\log 2$, so that $N_j = 2^{\lceil q_j \rceil}$. Then

$$b_{j+1} - b_j = 2r^{-2j}\big((j+1) - jr^2\big) > 2, \qquad q_{j+1} > e^2 q_j.$$

Since $q_j \ge q_1 = e^2/\log 2 > 10$, this implies $\lceil q_{j+1} \rceil - \lceil q_j \rceil \ge 2$ and hence $N_{j+1} \ge 4N_j > 2N_j$. The first scale is $N_1 = 2^{11} = 2048 > 30e^4 = 3(\alpha+1)e^{\alpha}/(sL)$ for the stated parameters. Moreover, $\log N_j \ge e^{b_j}$ implies

$$G(N_j) = \log\log(e^e + N_j) \ge \log\log N_j \ge b_j, \qquad G(N_j)a_j^2 \ge 2j(1-r)^2.$$

The amplitudes satisfy $\sum_j a_j = 1$ and $c_{j+1}/c_j = r(N_j/N_{j+1})^2 < 1$, so the summability and monotonicity arguments still apply. Multiplying the block lower bound by the rescaling factor λ^2 and dividing by K yields the displayed bound $2j$. The plotted reference is $K/[n^2 G(n)]$; K is determined by the construction, without fitting the observed errors.

The first five selected indices are

$$(N_1, N_2, N_3, N_4, N_5) = (2^{11}, 2^{80}, 2^{597}, 2^{4513}, 2^{34437}).$$

Since $\log_{10} N_5 \approx 10366.57$, executing every step separately is impractical. We instead sum the discrete recurrence exactly on each constant-gradient interval. For $g_n = c$ on $m \le n < k$, with $2 \le m < k$, let $A = sc/5$ and $H = (v_m - A(m+3))(m-1)m(m+1)(m+2)$. Then

$$v_k = A(k+3) + \frac{H}{(k-1)k(k+1)(k+2)},$$
$$x_k = x_m - \frac{A}{2}(k-m)(k+m+7) - \frac{H}{3}\left(\frac{1}{m(m+1)(m+2)} - \frac{1}{k(k+1)(k+2)}\right).$$

To derive these updates, for $\alpha = 4$ the velocity recurrence is $v_{n+1} = [(n-1)/(n+3)]v_n + sc$. The particular solution $v_n = A(n+3)$ satisfies it because $sc = 5A$. Subtracting this solution gives a homogeneous recurrence, and therefore

$$\begin{aligned} v_k - A(k+3) &= (v_m - A(m+3)) \prod_{i=m}^{k-1} \frac{i-1}{i+3} \\ &= (v_m - A(m+3)) \frac{(m-1)m(m+1)(m+2)}{(k-1)k(k+1)(k+2)}. \end{aligned}$$

For the position, $x_k = x_m - \sum_{j=m+1}^{k} v_j$ and

$$\sum_{j=m+1}^{k} (j+3) = \frac{1}{2}(k-m)(k+m+7),$$

$$\frac{1}{(j-1)j(j+1)(j+2)} = \frac{1}{3}\left(\frac{1}{(j-1)j(j+1)} - \frac{1}{j(j+1)(j+2)}\right).$$

The latter identity telescopes when summed over $m+1 \le j \le k$, giving the displayed position formula. At the initial index $m = 1$, the zero momentum coefficient instead gives $v_2 = sc$ and $x_2 = x_1 - sc$ directly. Each batch starts from the current forward state and is accepted only when its endpoint queries lie in the same constant part of the fixed spatial derivative. Positivity makes the intervening queries monotone, so they remain in that interval. Steps near joins are evaluated individually. This computes the discrete orbit without replacing it by a continuous-time approximation or resetting it to inverse positions.

The calculation through $2N_5$ uses 20,894 decimal digits, 883 batches, and 57 individual steps. A repeat with 100 additional digits leaves the displayed values unchanged. A separate check through iteration 5000 compares batching with individual oracle steps; their state difference is below 10^{-179}. Six retained blocks suffice for the spatial oracle and integration; the relative omitted-area bound is below $10^{-128393}$. All tail and precision checks are recorded with the code.

Figure 6 uses n/N_j to magnify each episode and $G(N_j)$ to compress the global time scale. The five selected ratios $N_j^2 G(N_j)\widehat{F}(x_{N_j})/K$ are approximately 200, 394, 590, 786, and 983; every displayed window crosses the reference from below. These are five observed episodes of the same loss. The bound $2j$ above establishes their infinite continuation.

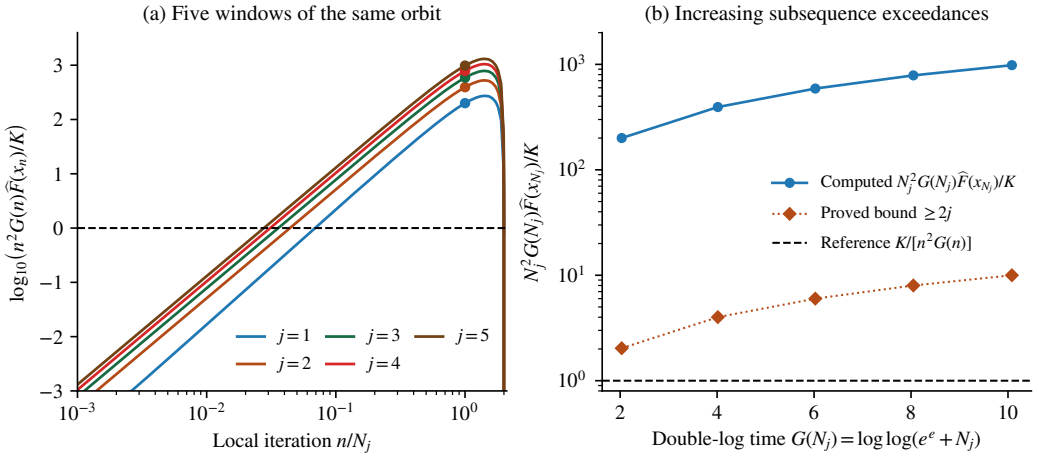


Figure 6: Batched discrete CD on one fixed loss with $G(n) = \log\log(e^e + n)$. Left: five windows of the same orbit, rescaled by N_j; positive values mean that the error exceeds $K/[n^2 G(n)]$. Dots mark $n = N_j$. Right: selected ratios on a double-logarithmic time axis, with their proved lower bounds. Connections between selected points are visual guides. The dashed reference appears at 0 on the left because of the logarithm, and at 1 on the right.

F.3 THEOREM 2: ATTAINING THE PREDICTED POWER CONSTANT

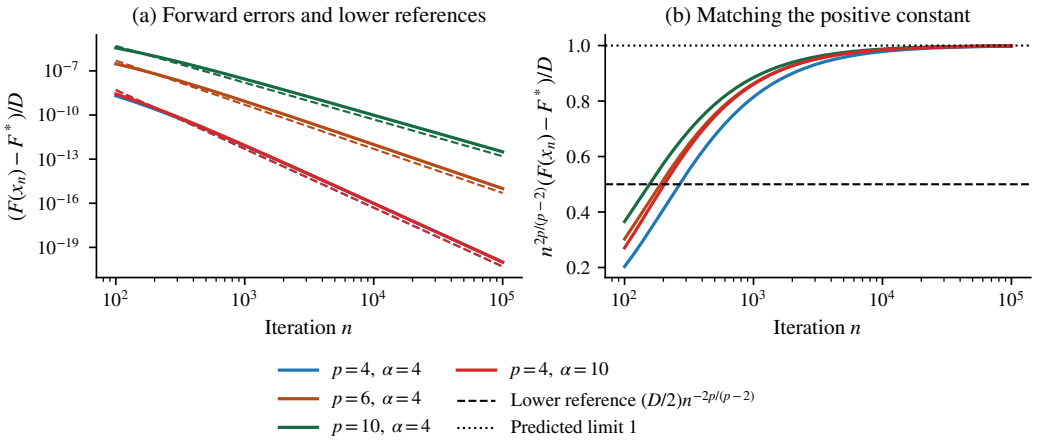


Figure 7: Forward CD on four fixed power witnesses. Left: computed errors divided by the predicted D, together with dashed references $\frac{1}{2}n^{-2p/(p-2)}$. Right: the normalized errors approach the predicted limit 1 and exceed $1/2$ throughout the reported observed tails. The same normalizations are used for the reference lines in both panels. These simulations evaluate the constructed interpolants directly.

We use $g_n = (n+32)^{-\rho}$ with $\rho = 2 + 2/(p-2)$. The cases $p = 4, 6, 10$ at $\alpha = 4$ lie in the flatness regime. An additional case $p = 4$, $\alpha = 10$ also satisfies the growth-only damping condition in part (ii); the constructed witness itself still has $H(p)$. Each forward run reaches $n = 10^5$. Objective integration uses five million nodes, giving a maximum relative omitted-area bound of 5.67×10^{-4} across the four runs.

The normalization uses the analytic constant from Lemma 5, including the spatial rescaling:

$$D = \lambda^2 \frac{s}{2(\rho-1)(\alpha+1-\rho)}.$$

No constant is fitted to the observed curve.

Figure 7 compares the computed errors with $(D/2)n^{-2p/(p-2)}$, an eventual lower reference implied by the positive asymptotic constant. The final normalized errors $n^{2p/(p-2)}\widehat{F}(x_n)/D$ are respectively 0.997791, 0.998458, 0.998677, and 0.998471. In the tested ranges, this ratio stays above $1/2$ from indices 267, 190, 156, and 204 onward, respectively. The proof establishes the eventual lower bound for the full infinite orbit; the runs show its finite-horizon manifestation.

G GUIDE TO THE SUPPLEMENTARY MATERIALS

The supplementary materials contain the Lean 4 development, construction figure generators, and forward CD experiments. This appendix describes their organization, gives reproduction commands, and specifies which claims are formally checked and which are illustrated numerically. All paths below are relative to the root of the accompanying materials.

G.1 CONTENTS AND ORGANIZATION

- `nag_cd_rate_nonimprovability/` contains the Lean 4 sources, pinned toolchain and dependencies, and the public axiom audit. Its theorem declarations and verification scope are described below.
- `figures/` contains the generators for the two construction schematics and the concrete instances in Appendix C.6, together with generated plots and numerical checks. The schematics explain the construction steps; the concrete instances record recurrence checks and integration-tail bounds. The proof dependency diagram is supplied as a separate TikZ source.

- `experiments/fixed_loss_lower_bounds/` contains the fixed-loss gradient oracles, forward CD implementations, sampled traces, plots, and numerical checks used in Section 5 and Appendix F. The Python runner `reproduce.py` invokes `run_cd.c` for individual CD steps; `slow_gain.py` computes the double-logarithmic-gain experiment using arbitrary-precision batched updates. The script `plot_main.py` assembles the main-text experiment figure from the saved data.

The accompanying README files describe the inputs, outputs, dependencies, and recorded checks for each component.

G.2 Scope of the Lean 4 formalization

The Lean 4 development formalizes the no-universal-gain conclusion, including every prescribed initial distance $R > 0$, and the sharp power-rate comparison.

The following declarations belong to the namespace `NAGPointwiseRates` and support the corresponding conclusions of Theorems 1 and 2.

Conclusion	Lean 4 declaration
Scalar no-gain with prescribed R	`nagCD_noPointwiseUniversalGain_dimOne_prescribedRadius`
Finite-dimensional no-gain with prescribed R	`nagCD_noPointwiseUniversalGain_prescribedRadius`
Flatness–growth upper rate	`sharpPowerBoundary_flatnessGrowth`
Growth-only upper rate	`sharpPowerBoundary_growthOnly`
Fixed power witness	`sharpPowerBoundary_noFurtherGain`

The manuscript uses an affine block interpolant for the no-gain construction and a shifted forcing sequence for the power example; the formal development uses its own interpolants and a spliced exact power tail. The correspondence is at the level of the specified theorem conclusions, not an identification of these particular witness functions. The intermediate lemmas and dependency diagram in the proof appendices organize the manuscript proofs; their presence does not assert separate formal verification of each displayed calculation.

G.3 Rebuilding and auditing the Lean 4 development

The package pins both Lean 4 and mathlib to version `v4.32.0`. With the Lean 4 toolchain manager installed, run the following commands from the root of the accompanying materials:

```
cd nag_cd_rate_nonimprovability
lake exe cache get
lake build
lake env lean AxiomAudit.lean
```

The public audit reports only `propext`, `Classical.choice`, and `Quot.sound` for its 16 declarations. Formal checking applies to the listed declarations. The remaining manuscript arguments are supplied as mathematical proofs; they are not claimed to have been checked by Lean 4.

G.4 Reproducing the figures and numerical runs

From the root of the accompanying materials, the following commands create a Python environment, install the pinned numerical dependencies, regenerate the construction figures, and rerun the forward experiments:

```
python3 -m venv .venv
.venv/bin/pip install -r \
  experiments/fixed_loss_lower_bounds/slow_gain_requirements.txt
.venv/bin/python figures/generate_constructions.py
make experiments PYTHON=.venv/bin/python
```

The dependency file includes the requirements for both experiment engines and the construction figures. The individual-step runner uses a C compiler and requires `long double` with at least 64 mantissa bits; the dependencies include a compiler fallback. The batched experiment repeats its calculation at higher precision and compares a shorter run against individual CD updates. The scripts record precision comparisons, source information, and analytic bounds on omitted integration tails.

To redraw Figure 3 from the supplied traces without rerunning the simulations, use

```
make experiment-overview PYTHON=.venv/bin/python
```

These numerical checks concern the reported finite runs. The analytic tail bounds do not enclose every floating-point rounding error; the infinite-time and arbitrary-gain conclusions are established by the mathematical proofs.