POLYLOG-TIGHT BOUNDS FOR STEP-SIZE ACCELERATION OF GD UNDER STRONG CONVEXITY

Anonymous authors
Paper under double-blind review

ABSTRACT

We study how much a predefined nonnegative step-size schedule can accelerate gradient descent on L-smooth, μ-strongly convex functions. For condition number $\kappa = L/\mu \geq 2$ and $\varepsilon \in (0, 1/2]$, we prove lower complexity bounds of rate $\Omega(\kappa^{\log_\rho 2}(\log\kappa)^{-\gamma_*}\log(1/\varepsilon)) \approx \Omega(\kappa^{0.786}(\log\kappa)^{-0.322}\log(1/\varepsilon))$ for relative squared distance to the minimizer in both *non-anytime* and *anytime* settings. Here, $\rho = 1+\sqrt{2}$ is the silver ratio. We also find that the periodic Silver step-size schedule (Altschuler & Parrilo, 2025a) achieves the *anytime* upper complexity bound $O(\kappa^{\log_\rho 2}\log(1/\varepsilon)) \approx O(\kappa^{0.786}\log(1/\varepsilon))$, extending its known bound at completed-block endpoints to every iterate beyond the accuracy threshold. Hence, our lower bounds tightly match the upper bounds up to $\mathrm{polylog}(\kappa)$. We also derive bounds for the function value gap and squared gradient norm.

1 INTRODUCTION

What is the *limit* of accelerating gradient descent by designing a step-size schedule in advance? We study this question for unconstrained minimization of an L-Lipschitz smooth, μ-strongly convex function $f : \mathbb{R}^d \to \mathbb{R}$, where $0 < \mu \leq L$. *Gradient descent* (GD) (Cauchy, 1847) is a simple first-order method that iteratively updates the current iterate $\boldsymbol{x}_t$ by moving in the direction of the negative gradient $-\nabla f(\boldsymbol{x}_t)$, scaled by a step size $\eta_t \geq 0$. In convex optimization, it is well known to converge to the unique minimizer $\boldsymbol{x}^\star$ of f at a *linear* rate. For $\varepsilon > 0$, GD with a constant step size can find an ε-approximate minimizer within $O(\kappa\log(1/\varepsilon))$ steps (Polyak, 1963; Bubeck, 2015; Nesterov, 2018), where $\kappa = L/\mu \geq 1$ is the *condition number* of f. This bound is known to be tight for *constant* step sizes (Arjevani et al., 2016), and it seemed impossible to improve.

Surprisingly, several recent works have shown that a *time-varying, non-adaptive* step-size schedule can accelerate GD beyond the $O(\kappa\log(1/\varepsilon))$ bound. For example, Altschuler & Parrilo (2025a;b) show that their *Silver step-size schedule* only takes $O(\kappa^\nu\log(1/\varepsilon)) \approx O(\kappa^{0.786}\log(1/\varepsilon))$ iterations[1] to achieve ε-accuracy. Their guarantees hold at prescribed endpoints (which are *non-anytime* guarantees), including the endpoints of periodically repeated Silver blocks. In addition, Zhang et al. (2025) propose a single infinite step-size schedule for GD that takes $O(\kappa^{(1+\nu)/2}\log(1/\varepsilon)) \approx O(\kappa^{0.893}\log(1/\varepsilon))$ iterations to achieve ε-accuracy no matter when it terminates (which is an *anytime* guarantee).

However, since GD with a predefined step-size schedule falls under a broader class of first-order methods, it is still subject to the lower bound of $\Omega(\sqrt{\kappa}\log(1/\varepsilon))$ for smooth strongly convex functions (Nemirovsky & Yudin, 1983; Arjevani & Shamir, 2016; Nesterov, 2018; Drori & Taylor, 2022). Thus, the two gaps below leave open whether a better predefined time-dependent schedule can further accelerate GD:

$$\tilde{\Omega}(\kappa^{0.5}) \xleftrightarrow{\text{non-anytime}} \tilde{O}(\kappa^{0.786}), \qquad \tilde{\Omega}(\kappa^{0.5}) \xleftrightarrow{\text{anytime}} \tilde{O}(\kappa^{0.893}). \tag{1.1}$$

1.1 PROBLEM SETUP

We denote the Euclidean ℓ_2 norm by $\|\cdot\|$ and the Euclidean inner product by $\langle\cdot,\cdot\rangle$. Let $\mathbb{N}$ be the set of all positive integers, $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$, and $[n] := \{1, \cdots, n\}$ for $n \in \mathbb{N}$. For a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, we say f is μ-strongly convex if $f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \langle\nabla f(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2$ $(\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d)$ for $\mu > 0$; it recovers (non-strong) convexity when $\mu = 0$. Also, for $L > 0$, we say f

[1] Here, we write $\nu = \log_\rho 2 = 0.786439\ldots$, where $\rho = 1+\sqrt{2}$ is the *silver ratio*.

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.

is L-Lipschitz smooth (or, L-smooth in short) if $\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \le L\|\boldsymbol{x} - \boldsymbol{y}\|$ $(\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d)$; in particular, it satisfies $f(\boldsymbol{x}) \le f(\boldsymbol{y}) + \langle \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2$ $(\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d)$. This is why the condition number of a μ-strongly convex L-smooth function satisfies $\kappa = L/\mu \ge 1$.

Let $\mathcal{F}_{\mu,L}(\mathbb{R}^d)$ be the class of μ-strongly convex, L-smooth functions on $\mathbb{R}^d$ (we may omit the domain and write $\mathcal{F}_{\mu,L}$ when it is clear from context). Let $\boldsymbol{x}^\star$ be the unique minimizer of f and let $f_\star = f(\boldsymbol{x}^\star)$. For a step-size schedule $\boldsymbol{\eta} = (\eta_t)_{t\ge 1} \in [0,\infty)^{\mathbb{N}}$, GD is written as the iteration

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} - \eta_t \nabla f(\boldsymbol{x}_{t-1}), \qquad t = 1, 2, \dots.$$

We focus on a deterministic (i.e., non-stochastic) schedule, fixed before running GD. Also, we often use the symbol $\boldsymbol{\eta}_{1:N} = (\eta_1, \cdots, \eta_N) \in [0,\infty)^N$ to denote the length-N prefix of $\boldsymbol{\eta}$; it denotes an empty schedule if $N = 0$. In the main text, our main convergence criterion is the relative squared distance to the minimizer, and we take its worst case over the function class:

$$R_D(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) = \|\boldsymbol{x}_N - \boldsymbol{x}^\star\|^2 / \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2, \quad R^{\sup}_{D,\kappa}(\boldsymbol{\eta}_{1:N}) = \sup_{d, f, \boldsymbol{x}_0 \ne \boldsymbol{x}^\star} R_D(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}), \tag{1.2}$$

where the supremum ranges over all finite $d \ge 1$, $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$, and $\boldsymbol{x}_0 \in \mathbb{R}^d \setminus \{\boldsymbol{x}^\star\}$. For an empty schedule (i.e., $N = 0$), we may write $R_D = R^{\sup}_{D,\kappa} = 1$. For $\varepsilon > 0$ and an infinite schedule $\boldsymbol{\eta}$, we study *iteration complexities* (Nemirovsky & Yudin, 1983) for non-anytime and anytime settings, respectively defined as

$$T^D_{\mathsf{na}}(\kappa, \varepsilon) = \inf\{N \in \mathbb{N}_0 : \exists \boldsymbol{\eta}_{1:N} \in [0,\infty)^N, \text{ s.t. } R^{\sup}_{D,\kappa}(\boldsymbol{\eta}_{1:N}) \le \varepsilon\}, \tag{1.3}$$

$$T^D_{\boldsymbol{\eta}}(\kappa, \varepsilon) = \inf\{N \in \mathbb{N}_0 : \forall n \in \mathbb{N}_0,\ n \ge N \implies R^{\sup}_{D,\kappa}(\boldsymbol{\eta}_{1:n}) \le \varepsilon\}. \tag{1.4}$$

In particular, $T^D_{\mathsf{na}}(\kappa, \varepsilon) \le T^D_{\boldsymbol{\eta}}(\kappa, \varepsilon)$ for every such $\boldsymbol{\eta}$. We also define iteration complexities for the function value gap $f(\boldsymbol{x}_N) - f_\star$ and squared gradient norm $\|\nabla f(\boldsymbol{x}_N)\|^2$ in Appx. F, where we state the bounds for each normalization separately.

1.2 Our Contributions

In this paper, we narrow the gaps (1.1) by proving a *nearly tight lower complexity bound* for arbitrary predetermined nonnegative step-size schedules, mostly in terms of the reduction rate of squared distance to the minimizer. Our lower bound identifies $\nu \approx 0.786$ as the optimal polynomial exponent of κ in both non-anytime and anytime settings. We also find that the periodic Silver construction of Altschuler & Parrilo (2025a) achieves an *anytime* upper bound at the same Silver complexity.

In the first theorem, we show the existence of a smooth, strongly convex hard instance function (whose construction can depend on the given step-size schedule) that satisfies a certain convergence-rate lower bound in terms of the relative squared distance to the minimizer.

Theorem 1.1 (Convergence rate lower bound)**.** *There is an absolute constant $C > 0$ satisfying the following: for every $\kappa = L/\mu \ge 2$, $N \in \mathbb{N}$, and step-size schedule $\boldsymbol{\eta}_{1:N} \in [0,\infty)^N$, there exist an L-smooth, μ-strongly convex function $f : \mathbb{R}^d \to \mathbb{R}$ with a (unique) minimizer $\boldsymbol{x}^\star \in \mathbb{R}^d$ for some $d \in \mathbb{N}$ and an initial point $\boldsymbol{x}_0 \ne \boldsymbol{x}^\star$ such that*

$$\frac{\|\boldsymbol{x}_N - \boldsymbol{x}^\star\|^2}{\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2} \ge \exp\left[-CN\kappa^{-\nu}(\log\kappa)^{\gamma_*}\right], \quad \textit{where } \gamma_* = (\pi^2\nu)/(8\sqrt{2}\,\rho\log\rho) \approx 0.322. \tag{1.5}$$

This convergence-rate lower bound gives the non-anytime iteration complexity lower bound below, which also applies in the anytime setting by definition (Eqs. (1.3) & (1.4)).

Corollary 1.2 (Iteration complexity lower bound)**.** *For every $\kappa \ge 2$ and $\varepsilon \in (0, \frac{1}{2}]$, if a nonnegative schedule of length N guarantees $\|\boldsymbol{x}_N - \boldsymbol{x}^\star\|^2 \le \varepsilon\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2$ uniformly over the L-smooth, μ-strongly convex function class with $L/\mu = \kappa$, then it is required that*

$$N \ge C^{-1}\kappa^{\nu}(\log\kappa)^{-\gamma_*}\log(1/\varepsilon).$$

We can use the same absolute constant $C > 0$ in Thm. 1.1.

Lastly, we provide an *anytime* complexity upper bound for the *periodic Silver schedule*. Although its non-anytime guarantee was previously established by Altschuler & Parrilo (2025a, Corollary 2), we extend this by proving prefix stability for the same schedule.

Theorem 1.3 (Anytime upper bound of periodic Silver schedule)**.** *For any $0<\mu\le L$ with $\kappa=\frac{L}{\mu}\ge 1$, there exists a (periodic) schedule $\boldsymbol{\eta}^{\mathsf{sil}}\in(0,\infty)^{\mathbb{N}}$ depending only on L and μ that generates the GD iterates $(\boldsymbol{x}_t)_{t\ge 1}$ satisfying the following: For any $\varepsilon\in(0,\frac{1}{2}]$, $d\in\mathbb{N}$, L-smooth μ-strongly convex function on $\mathbb{R}^d$ with minimizer $\boldsymbol{x}^\star$, and every initial point $\boldsymbol{x}_0$, the sufficient number of iterations T_0 to guarantee $\|\boldsymbol{x}_T-\boldsymbol{x}^\star\|^2\le\varepsilon\|\boldsymbol{x}_0-\boldsymbol{x}^\star\|^2$ for all $T\ge T_0$ is*

$$T_0=O(\kappa^\nu\log(1/\varepsilon)).$$

Therefore, we tighten the complexity bounds roughly to

$$T^D_{\mathsf{na}}(\kappa,\varepsilon)=\Omega(\kappa^{0.786}(\log\kappa)^{-0.322}\log(1/\varepsilon))\ \lesssim\ T^D_{\boldsymbol{\eta}^{\mathsf{sil}}}(\kappa,\varepsilon)=O(\kappa^{0.786}\cdot\log(1/\varepsilon))$$

for both non-anytime and anytime settings, differing by only a polylogarithmic factor in κ. In addition, it strictly separates the deterministic GD for strongly convex smooth functions from accelerated first-order methods (Nesterov, 2018) and Chebyshev-scheduled GD on quadratics (Young, 1953), both of which attain $O(\kappa^{0.5}\log(1/\varepsilon))$ complexity.

1.3 Related Work

The acceleration of GD with a predefined step-size schedule has been studied over the larger class of L-smooth convex functions. In this setting, Altschuler & Parrilo (2025a) show that their Silver schedule yields a function value gap of order $O(N^{-1/\nu})\approx O(N^{-1.271})$ at horizons $N=2^k-1$. Grimmer et al. (2025) improve the constant factor while keeping the same rate in N, and obtain accelerated squared-gradient-norm guarantees by reversing the step-size order. Zhang & Jiang (2026) use concatenation to give a family of schedules with this polynomial rate for each prescribed finite horizon and improve the constant factor. These works allow choosing schedules separately for each N, so they fall under *non-anytime* upper-bound guarantees. However, they do not automatically provide a single infinite schedule that achieves an *anytime* upper-bound guarantee. Kornowski & Shamir (2024) point this out and ask whether a single schedule can improve the $O(N^{-1})$ last-iterate guarantee of constant-step-size GD. They also demonstrate that the infinite Silver schedule has no uniform last-iterate upper convergence bound tending to zero for smooth non-strongly convex functions, which concerns a larger function class than the periodic Silver guarantee in Thm. 1.3, where we use strong convexity to obtain strict shrinkages. Later, Zhang et al. (2025) answered the open question by providing an infinite schedule achieving the rate $O(N^{-2/(1+\nu)})\approx O(N^{-1.119})$.

In fact, Zhang et al. (2025) extend their result to the smooth strongly convex setup, which gives an anytime complexity bound $\tilde{O}(\kappa^{(1+\nu)/2})\approx\tilde{O}(\kappa^{0.893})$ to achieve $f(\boldsymbol{x}_N)-f(\boldsymbol{x}^\star)\le\frac{L}{2}\|\boldsymbol{x}_0-\boldsymbol{x}^\star\|^2\cdot\varepsilon$. Note that the L-smoothness gives $f(\boldsymbol{x}_N)-f(\boldsymbol{x}^\star)\le\frac{L}{2}\|\boldsymbol{x}_N-\boldsymbol{x}^\star\|^2$. Hence, our anytime upper bound of rate $\tilde{O}(\kappa^{0.786})$ directly applies to their convergence metric, improving their result.

Until recently, it was open whether these guarantees could be improved beyond constant factors. For the smooth convex class, several preprints released over the last few months have rapidly narrowed the gap by providing convergence lower-bound guarantees (Tsai et al., 2026; Ma & Chen, 2026; Jung et al., 2026; Ye & Liu, 2026a;b). Most notably, Ye & Liu (2026b) provide near-tight lower bounds of the non-anytime rate $N^{-1/\nu-o(1)}$ and the anytime rate $N^{-2/(1+\nu)-o(1)}$ in terms of the function value gap normalized by $\|\boldsymbol{x}_0-\boldsymbol{x}^\star\|^2$, where the $o(1)$ terms in the exponents are of order $O(\sqrt{\log\log N/\log N})$. To the best of our knowledge, however, the gap between the upper and lower complexity bounds for *strongly* convex functions has not yet been narrowed.

Together with the smooth convex upper bounds (Altschuler & Parrilo, 2025a; Grimmer et al., 2025; Zhang et al., 2025; Zhang & Jiang, 2026), the lower-bound results by Ye & Liu (2026b) identify different optimal polynomial decay exponents of N in the non-anytime and anytime settings. Perhaps surprisingly, our results (Cor. 1.2 & Thm. 1.3) instead identify a *common* optimal exponent of the condition number for (non-)anytime settings as $\nu\approx 0.786$ under a smooth *strongly*-convex setup, consistent with the known non-anytime upper bound by Altschuler & Parrilo (2025b).

Very recently, a preprint by Liu & Ye (2026) was released right before the submission deadline, showing a lower complexity bound of rate $\Omega(\kappa^\nu\,[e^{\Theta(\sqrt{\log\kappa})}\log\kappa]^{-1}\log(1/\varepsilon))$ under the same setup. Although it shares several proof ideas with ours (e.g., checkpoints, curved local trajectories and the repetition trick), our hard-instance constructions are very different. Indeed, our lower bound is tighter by a super-polylogarithmic factor $e^{\Theta(\sqrt{\log\kappa})}(\log\kappa)^{1-\gamma_*}$. We also provide an *anytime* upper bound, implying a stronger match between upper/lower bounds in (non-)anytime settings.

2 Proof Sketches of Main Results

Here, we prove Thms. 1.1 & 1.3, deferring the proofs of key lemmas to later sections/appendices.

2.1 Proof of Theorem 1.1: Non-Anytime Lower Bound

To prove Thm. 1.1, we combine three key facts: Lem. 2.1, Prop. 2.2, and Lem. D.2. In particular, the proof of Lem. 2.1 constructs a *hard instance* $f \in \mathcal{F}_{\mu,L}$ and $\boldsymbol{x}_0$, which we overview in Secs. 3 & 4.

For $L > 0$, replacing a function f by f/L and each step η_t by $L\eta_t$ preserves the GD iterates. So, without loss of generality, fix $L = 1$ and $\mu = 1/\kappa$. Besides, for simplicity, we define:

$$\mathcal{Q}(\kappa) := \kappa^{\nu}(\log \kappa)^{-\gamma_*}, \qquad \alpha := \log_2 \rho = 1/\nu \approx 1.271. \tag{2.1}$$

The desired lower bound in Eq. (1.5) can then be written as $R_D(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) \geq \exp\left[-CN/\mathcal{Q}(\kappa)\right]$. To show this, we construct a *hard instance* function $f \in \mathcal{F}_{1/\kappa,1}$ based on a given schedule $\boldsymbol{\eta}_{1:N}$, as well as an initial point $\boldsymbol{x}_0 \neq \boldsymbol{x}^\star$. By our construction of f and an explicitly chosen $\boldsymbol{x}_0$, we have an inequality similar to Eq. (1.5) but with a tiny κ-dependent factor χ_κ in front of the exponential term:

Lemma 2.1. *There are absolute constants $\kappa_* \geq 2$ and $C_0, C_1 > 0$ such that, for any $\kappa \geq \kappa_*$, $N \in \mathbb{N}$, and step-size schedule $\boldsymbol{\eta}_{1:N} \in [0,\infty)^N$, there exists $f \in \mathcal{F}_{1/\kappa,1}$ and $\boldsymbol{x}_0 \neq \boldsymbol{x}^\star$ satisfying*

$$R_D(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) \geq \chi_\kappa \exp\left[-C_0 N/\mathcal{Q}(\kappa)\right], \qquad \chi_\kappa = \kappa^{-C_1}.$$

Directly applying this lemma does not yield Eq. (1.5) because it only gives $N \geq C_0^{-1}\mathcal{Q}(\kappa)\log(\chi_\kappa/\varepsilon)$, which is non-vacuous only when $\varepsilon < \chi_\kappa$, i.e., the target accuracy is nonconstant and shrinks with κ. To alleviate this, we apply the *repetition trick* described below:

Proposition 2.2. *Fix any $\kappa \geq 2$, $0 < \chi \leq 1$, $C', C_{\mathsf{rep}} > 0$, and $N \in \mathbb{N}$. Assume that each $m \in \mathbb{N}$ and $\tilde{\boldsymbol{\eta}}_{1:m} \in [0,\infty)^m$ gives some $\tilde{f} \in \mathcal{F}_{1/\kappa,1}$ and $\tilde{\boldsymbol{x}}_0$ so that $R_D(\tilde{f}, \tilde{\boldsymbol{x}}_0; \tilde{\boldsymbol{\eta}}_{1:m}) \geq \chi \exp(-C'm)$. Then, for any $\boldsymbol{\eta}_{1:N} \in [0,\infty)^N$, there exists $f \in \mathcal{F}_{1/\kappa,1}$ and $\boldsymbol{x}_0$ such that*

$$R_D(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) \geq \exp\left[-(C' + C_{\mathsf{rep}})N\right].$$

Proof. Fix any schedule $\boldsymbol{\eta}_{1:N}$. Consider a periodic schedule $(\boldsymbol{\eta}_{1:N}, \ldots, \boldsymbol{\eta}_{1:N})$ consisting of n_{rep} copies of $\boldsymbol{\eta}_{1:N}$, where $n_{\mathsf{rep}} := \max\{1, \lceil \log(1/\chi)/(C_{\mathsf{rep}}N)\rceil\}$. Applying the assumption with $m = n_{\mathsf{rep}}N$, take $\tilde{f}$ and $\boldsymbol{y}_0 = \tilde{\boldsymbol{x}}_0 \neq \boldsymbol{x}^\star$ such that, if $\boldsymbol{y}_j$ is the iterate after j copies of $\boldsymbol{\eta}_{1:N}$ starting from $\boldsymbol{y}_0$, then $\|\boldsymbol{y}_{n_{\mathsf{rep}}} - \boldsymbol{x}^\star\|^2 \geq \chi \exp\left(-C' n_{\mathsf{rep}} N\right) \|\boldsymbol{y}_0 - \boldsymbol{x}^\star\|^2 > 0$. This implies that

$$\chi \exp\left(-C' n_{\mathsf{rep}} N\right) \leq \prod_{j \in [n_{\mathsf{rep}}]} \frac{\|\boldsymbol{y}_j - \boldsymbol{x}^\star\|^2}{\|\boldsymbol{y}_{j-1} - \boldsymbol{x}^\star\|^2} \leq \left(\max_{j \in [n_{\mathsf{rep}}]} \frac{\|\boldsymbol{y}_j - \boldsymbol{x}^\star\|^2}{\|\boldsymbol{y}_{j-1} - \boldsymbol{x}^\star\|^2} \right)^{n_{\mathsf{rep}}}.$$

Taking the n_{rep}-th root of both sides, there must exist $j \in [n_{\mathsf{rep}}]$ such that

$$\exp\left[-(C' + C_{\mathsf{rep}})N\right] \leq \chi^{1/n_{\mathsf{rep}}} \exp\left(-C'N\right) \leq \|\boldsymbol{y}_j - \boldsymbol{x}^\star\|^2 / \|\boldsymbol{y}_{j-1} - \boldsymbol{x}^\star\|^2,$$

where the first inequality holds by definition of n_{rep}. Hence, we have $R_D(\tilde{f}, \boldsymbol{y}_{j-1}; \boldsymbol{\eta}_{1:N}) \geq \exp\left[-(C' + C_{\mathsf{rep}})N\right]$; choosing $f = \tilde{f}$ and $\boldsymbol{x}_0 = \boldsymbol{y}_{j-1}$ concludes the proof. □

Indeed, this proposition allows us to eliminate the κ-dependent prefactor in Lem. 2.1 using a repetitive schedule extension. The remaining hurdle is the assumption "$\kappa \geq \kappa_*$" since κ_* can be extremely huge. We circumvent this by applying the classical quadratic lower bound (e.g., Nemirovsky & Yudin 1983; Arjevani & Shamir 2016): for every $\kappa \geq 2$, $N \in \mathbb{N}$, and any predetermined N-step schedule, a class of strongly convex quadratics is enough to show $R_D \geq [(\sqrt{\kappa}-1)/(\sqrt{\kappa}+1)]^{2N}$; see Lem. D.2 for its simple derivation. Combining these, we finally prove our lower bound:

Proof of Thm. 1.1. For $\kappa \geq \kappa_*$, we apply Prop. 2.2 with $C' = C_0/\mathcal{Q}(\kappa)$ and $C_{\mathsf{rep}} = 1/\mathcal{Q}(\kappa)$. Then, Lem. 2.1 gives a finite-dimensional hard instance with $R_D \geq \exp\left[-(C_0+1)N/\mathcal{Q}(\kappa)\right]$. In contrast, for $2 \leq \kappa \leq \kappa_*$, use the quadratic lower bound (e.g., Lem. D.2) and the finite constant $C_{\mathsf{small}} := \max_{\kappa \in [2,\kappa_*]} 2\mathcal{Q}(\kappa)\log(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1})$. By continuity, the finite positive maximum on the compact interval is attained. Thus, we conclude the proof of the theorem for any $\kappa \geq 2$ by setting $C = \max\{C_0 + 1, C_{\mathsf{small}}\}$. □

As a result, we obtain a non-anytime lower bound $T^D_{\mathsf{na}}(\kappa, \varepsilon) \geq C^{-1}\mathcal{Q}(\kappa)\log(1/\varepsilon)$ (Cor. 1.2); the anytime lower bound follows immediately. We present the complete proofs in Appcs. A–D.

2.2 Proof of Theorem 1.3: Anytime Upper Bound

We now prove the *anytime* upper bound in Thm. 1.3 by combining the key facts: Lems. 2.3 & 2.4. Let us recall the Silver schedule (Altschuler & Parrilo, 2025a); define *Silver blocks* S_k at level k by

$$S_0 = (), \quad S_{k+1} = (S_k, \eta_k^{\mathsf{join}}, S_k), \quad \eta_k^{\mathsf{join}} = 1 + \rho^{k-1}. \qquad (\rho = 1+\sqrt{2};\ \ k = 0, 1, 2, \dots)$$

Repeating a fixed block gives a periodic Silver schedule (Altschuler & Parrilo, 2025b, Footnote 2):

$$\boldsymbol{\eta}^{\mathsf{sil}} = (S_{k_*}, S_{k_*}, S_{k_*}, \dots), \qquad \text{repeated forever, for some } k_* \geq 1. \tag{2.2}$$

Our goal is to choose k_* (depending on $\kappa \geq 1$) so that $\boldsymbol{\eta}^{\mathsf{sil}}$ gives anytime convergence over $\mathcal{F}_{1/\kappa,1}$. The following endpoint inequality specializes the primitive certificate and concatenation argument of Zhang & Jiang (2026); Appx. E.2 derives it.

Lemma 2.3 (Silver endpoint contraction)**.** *For a convex 1-smooth f minimized at $\boldsymbol{x}^\star$, write $f_\star = f(\boldsymbol{x}^\star)$. For every integer $k \geq 0$ and starting point $\boldsymbol{y}$, the endpoint $\boldsymbol{z}$ of S_k satisfies*

$$\|\boldsymbol{z} - \boldsymbol{x}^\star\|^2 + 2(\rho^k - 1)(f(\boldsymbol{z}) - f_\star) + \rho^k(\rho^k - 1)\|\nabla f(\boldsymbol{z})\|^2 \leq \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2. \tag{2.3}$$

Next, the following lemma controls the iterates between endpoints; we prove it in Appx. E.3.

Lemma 2.4 (Silver prefix stability)**.** *Fix a convex 1-smooth f minimized at $\boldsymbol{x}^\star$. For every integer $k \geq 0$ and starting point $\boldsymbol{y}$, let $\boldsymbol{y}_t$ be the iterate after t steps of S_k, with $\boldsymbol{y}_0 = \boldsymbol{y}$. Then*

$$\|\boldsymbol{y}_t - \boldsymbol{x}^\star\|^2 \leq \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2 \qquad (0 \leq t \leq 2^k - 1). \tag{2.4}$$

In short, we manage the distance from $\boldsymbol{x}^\star$ at intermediate iterates using the endpoint distance. Combined with the strong convexity, we can finally prove the theorem:

Proof of Thm. 1.3. Set $L = 1$. Choose the block level $k_* = \lceil \log_\rho(\kappa + 1) \rceil$ in Eq. (2.2), and observe that each block has length $m = 2^{k_*} - 1 < 2(\kappa+1)^\nu$. For every block input $\boldsymbol{y}$ and its endpoint $\boldsymbol{z}$, the strong convexity of f gives $2(f(\boldsymbol{z}) - f_\star) \geq \kappa^{-1}\|\boldsymbol{z} - \boldsymbol{x}^\star\|^2$. Substituting into (2.3) yields

$$\|\boldsymbol{z} - \boldsymbol{x}^\star\|^2 \leq \tfrac{1}{1+(\rho^{k_*}-1)/\kappa}\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2 \leq \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2,$$

since $\rho^{k_*} - 1 \geq \kappa$. Thus, every completed block contracts the squared distance by a factor of at most $1/2$, uniformly over starting points. Now, consider any $N \in \mathbb{N}_0$ and write $N = jm + r$ with $0 \leq r < m$. Repeated block contraction and Lem. 2.4 for the last r steps give $\|\boldsymbol{x}_N - \boldsymbol{x}^\star\|^2 \leq \|\boldsymbol{x}_{jm} - \boldsymbol{x}^\star\|^2 \leq 2^{-\lfloor N/m \rfloor}\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2$. Hence, by taking suprema, we obtain

$$R_{D,\kappa}^{\sup}(\boldsymbol{\eta}_{1:N}^{\mathsf{sil}}) \leq 2^{-\lfloor N/m \rfloor} \qquad N \in \mathbb{N}_0. \tag{2.5}$$

For $\varepsilon \in (0, 1/2]$, it gives $T_{\boldsymbol{\eta}^{\mathsf{sil}}}^D(\kappa, \varepsilon) \leq T_0 := m\,\lceil \log_2(1/\varepsilon) \rceil = O(\kappa^\nu \log(1/\varepsilon))$.

For general $L > 0$, repeating S_{k_*}/L on f reproduces the normalized iterates on f/L and yields the same bound, with a schedule depending only on L and μ. □

3 Checkpoint Selection and Construction of Core Functions

For sufficiently large κ and a prescribed schedule, we construct a function minimized at the origin such that GD retains a large enough coordinate. We select updates with large step sizes, called **checkpoints** (Ma & Chen, 2026; Jung et al., 2026; Ye & Liu, 2026a;b), and use them to obtain an amplitude in the next coordinate. Repeating this construction gives a lower bound on the final R_D.

3.1 From a Single Coordinate Pair to the Full Coordinate Chain

Consider n consecutive steps, indexed locally by $1, \dots, n$. Suppose that we have checkpoints $1 \leq t_1 < \cdots < t_k = n$ and let $t_0 = 0$. The unselected updates before the checkpoint t_i form a **gap**; write $b_i = \eta_{t_i}$ and $s_i = \sum_{t=t_{i-1}+1}^{t_i-1} \eta_t$. On coordinates $(i, i+1)$, we arrange a **transfer** from $(D_i, 0)$ at t_{i-1} to $(*, D_{i+1})$ at t_i, with $D_i, D_{i+1} > 0$ and $D_1 = 1$. Thus, $\|\boldsymbol{z}_{t_i}\| \geq D_{i+1}$, and the next transfer starts from this new amplitude (Jung et al., 2026). We realize these transfers within nonnegative convex 1-smooth $g_i : \mathbb{R}^2 \to \mathbb{R}$ vanishing at the origin, and define

$$h(\boldsymbol{z}) = \frac{2\|\boldsymbol{z}\|^2}{4\kappa - 1} + \frac{2\kappa - 4}{4\kappa - 1}\sum_{i=1}^{k} D_i^2 g_i\left(\frac{z^{(i)}}{D_i}, \frac{z^{(i+1)}}{D_i}\right). \tag{3.1}$$

We call this function, corresponding to a coordinate chain, a **phase core**. Each coordinate belongs to at most two pairs, giving $h \in \mathcal{F}_{4/(4\kappa-1),\,4(\kappa-1)/(4\kappa-1)}$; we keep the smoothness upper bound below 1 so that we can later add functions **coupling** this core to other cores. The quadratic fixes the minimizer at zero, while the smoothness bound leaves room for later couplings. To make the overlapping pairs *compatible*, the preceding function's derivative in the shared coordinate vanishes while the next function's gradient remains zero. The active pair therefore follows its prescribed transfer, regardless of how previously used coordinates move (Fig. 1; Lem. B.6).

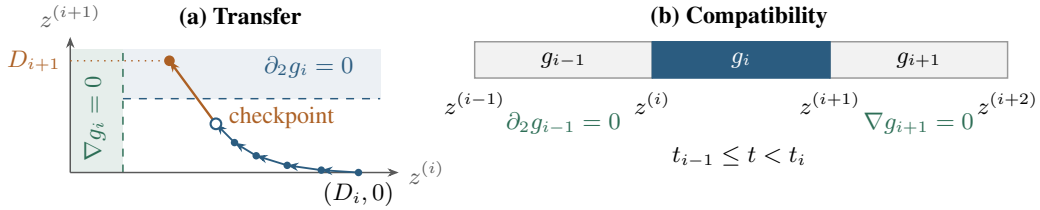


Figure 1: Schematic transfer and compatibility (Lem. B.6). At the scaled arguments in (3.1), the neighboring derivatives vanish before each update, so the active pair realizes the local transfer.

The gradient of each g_i rotates so that the first coordinate decreases while the next increases (Ye & Liu, 2026b); we use the *logarithmic spiral* suggested in their Section 5, Footnote 7. An *ordinary transfer* adjusts this rotation to reduce the amplitude loss. Its *startup parameter* a is a checkpoint-size threshold that ensures compatibility with the chosen rotation, so $a \le b$; it does not represent an extra update. The bound on D_i/D_{i+1} separates the numerical factor $c(a)(a+s)/b$ from damping and compatibility factors (Lem. B.3, Eqs. (B.10)–(B.11)). Here, a also bounds the cost of positioning the rotation and $c(a) > 1$ is the associated *geometric cost correction*. Increasing a reduces $c(a)$ but increases $a + s$, creating the trade-off used in checkpoint selection. Additional checkpoints may require exceptional transfers, discussed in Sec. 3.3.

We also observe a further difficulty, (possibly) created by the quadratic term in h: long gaps can substantially reduce the amplitude. Then, selecting more checkpoints does not resolve a long run containing only small steps. We therefore assign different consecutive portions of the schedule to different cores, calling each assigned portion a **phase**. This multi-level construction, using several phases rather than one coordinate chain, is our key technical contribution.

3.2 Checkpoint Selection and the Silver Time Scale

Within a phase, we exclude cases where small step sizes persist for too long. To make this precise, call a maximal consecutive run of steps ≤ 64 a *small-step run*. We assign phases so that each ends in a step size > 64 and every nonleading small-step run has sum at most $(4\kappa-1)/512$; we do not impose such a restriction on the leading one. Sec. 4.2 describes how to assign phases satisfying these conditions to the full schedule.

The goal of checkpoint selection is to minimize the product of transfer costs, meaning we want to keep the iterate amplitude as large as possible. So, we design a cost $\Psi_0(\mathcal{T})$ of a trial checkpoint set $\mathcal{T} = \{t_1 < \cdots < t_k\}$ as

$$\Psi_0(\mathcal{T}) = \mathcal{N}(s_{k+1}) \prod_{i=1}^{k} \frac{\mathcal{N}_{b_i}(s_i)}{b_i}, \qquad \Psi_0(\varnothing) = \mathcal{N}(s_{\mathrm{tot}}). \tag{3.2}$$

Here, $\mathcal{N}(s)$ is the minimum of the ordinary transfer cost's numerator $c(a)(a+s)$ over the ordinary startup range $a_{\min} \le a \le (4\kappa-1)/8$, where $a_{\min} > 64$, and $\mathcal{N}_b(s)$ imposes an additional constraint $a \le b$. Also, s_{k+1} is the terminal gap sum and $s_{\mathrm{tot}} = \sum_{t\in[n]} \eta_t$ is the total step sum.

We choose the checkpoint set by minimizing Eq. (3.2). To see why this choice is useful, consider a schedule for which the empty checkpoint selection is already optimal. If an eligible checkpoint could improve the cost, we would select it and split the problem into two shorter gaps. Hence, the difficult case is precisely when *no improvement is possible*. We bound the total step sum of such a schedule by using $W(\cdot)$, where m is the number of timesteps (Lem. A.12).

Why $\nu = \log_\rho 2$? The Silver exponent is already clear from an idealized balanced split (which is arguably similar to what the Silver schedule does). Suppose a new checkpoint splits a gap, originally

with step sum J, into two with equal step sum X. If selecting this checkpoint has the same cost as selecting none, then Eq. (3.2), together with $b = J - 2X$, gives $(J - 2X)\mathcal{N}(J) = \mathcal{N}_b(X)\mathcal{N}(X)$. Then, temporarily ignoring the startup and geometric cost corrections ($\mathcal{N}(s) = \mathcal{N}_b(s) = s$), we obtain $(J - 2X)J = X^2$ and $J = (1 + \sqrt{2})X = \rho X$. Thus, roughly doubling the number of positions multiplies the admissible step sum by ρ, suggesting the Silver exponent $\alpha = \log_2 \rho$.

Why $\gamma_* = (\pi^2\nu)/(8\sqrt{2}\,\rho\,\log\rho)$? The geometric cost correction explains the additional logarithmic factor. For large gap step sums s, even after optimizing the startup parameter $a \geq a_{\min}$ without an upper cap, the cost's numerator exceeds the ideal value s by a relative correction asymptotic to $\pi^2/(4\log s)$ (Lem. A.10). To absorb this slowly vanishing relative correction in the cost, we let the elasticity $d\log W(m)/d\log m$, the effective exponent of $W(m)$, slightly exceed α. We balance the correction against this increase and recognize the first-order term $\alpha\gamma_*/\log m$ (Appx. A.4). Integrating it with respect to $\log m$ contributes $\alpha\gamma_*\log\log m$ to $\log W(m)$, explaining the logarithmic factor in $W(m) = \Theta(m^\alpha[\log(em)]^{\alpha\gamma_*})$. The full argument controls the higher-order terms and also handles unbalanced splits and checkpoints for which the constraint $a \leq b$ is active (Prop. A.11 & Lem. A.12).

We use this step-sum bound locally by dividing each phase into consecutive **windows** of at most q updates. For sufficiently large κ, choose q as the largest positive integer satisfying $W(4q) \leq \frac{4\kappa-1}{256}$. With this choice, inverting the growth rate of W gives $q = \Theta(\kappa^\nu(\log\kappa)^{-\gamma_*}) = \Theta(\mathcal{Q}(\kappa))$ (Eqs. (A.21) & (A.22)). We use this window length in checkpoint selection to control the damping caused by the core's quadratic term.

3.3 A Phase Core Certificate

We now combine the window-wise checkpoint selections into a checkpoint set for a phase of length n. The same set must have controlled cost at every selected prefix; otherwise different stopping points could require different hard functions. We achieve this by carrying unfinished gaps across window boundaries and inserting additional checkpoints into long internal gaps. When an inserted checkpoint is too small for an ordinary transfer's startup condition, we use an *exceptional transfer*, which satisfies the required compatibility conditions at a larger polynomial cost in κ. These modifications preserve the window scale q and give one fixed checkpoint set with uniformly controlled prefix costs (Lems. B.4, A.13, & A.15).

Combining the transfer costs and gap damping, we choose absolute constants $C_{\mathsf{amp}} > 0$ and $C_{\mathsf{win}} \geq 2$ and define

$$\mathsf{amp}(n) := (\kappa - 1)^{-C_{\mathsf{amp}}} \exp[-C_{\mathsf{win}}\lceil n/q\rceil]\,. \tag{3.3}$$

Intuitively, it is the amplitude that we can guarantee survives throughout a phase of length n. The polynomial factor absorbs the fixed geometric losses, while the exponential term pays a constant factor per window, so the total decay is governed by the number n/q of windows. For this phase, we construct cores $h_{\boldsymbol{p}}$ parameterized by $\boldsymbol{p}$ in a nonempty compact convex set Trials. Every trial has the curvature bounds of Sec. 3.1, agrees with $2\|\boldsymbol{z}\|^2/(4\kappa - 1)$ on $\|\boldsymbol{z}\| \leq \mathsf{amp}(n)$, and has the same nonzero initial gradient $\boldsymbol{g}_* := \nabla h_{\boldsymbol{p}}(\boldsymbol{e}^{(1)})$, supported on the first two coordinates; here $\boldsymbol{e}^{(1)} = (1, 0, \ldots, 0)$. These are static properties: they hold before any trial is matched to an actual GD trajectory (Lem. B.8).

When coupled to earlier phases, the core may receive an *incoming correction* $-\theta_t\boldsymbol{g}_*$, where $0 \leq \theta_t \leq 1$ is the scalar correction coefficient at local time t. Under the readout and projection conditions established in Sec. 4.1, choosing a trial consistently with its actual correction coefficients matches its intended transfers to its GD trajectory and gives

$$\|\boldsymbol{z}_t\| \geq \mathsf{amp}(n), \qquad 0 \leq t \leq n, \qquad \boldsymbol{z}_0 = \boldsymbol{e}^{(1)}. \tag{3.4}$$

The first phase has zero incoming correction; Sec. 4 explains how we match later phases (Lem. B.9).

4 Coupling Phase Cores into our Hard Function

Simply summing cores defined on different coordinate blocks would allow every block to move from the first update. Each core, however, must start from its initial point when its assigned phase begins. We add a coupling function that cancels the new core's initial gradient, holding its coordinate block at its initial point until that time. Since the coupling function can also change the gradients on the earlier coordinates, we construct it so that this effect vanishes exactly while the new block is held fixed.

Suppose that we use $P+1$ phases in total. Denote the coordinate block and length of each phase $j \in \{0, \dots, P\}$ by $\boldsymbol{x}^{[j]}$ and $n^{[j]}$, respectively. For the core $h^{[j]}$ constructed with the step sizes in phase j (defined in Eq. (3.1)), we define $\boldsymbol{x}^{[j]}_{\mathsf{init}} = A^{[j]}\boldsymbol{e}^{(1)}$ with a unit vector $\boldsymbol{e}^{(1)} = (1, 0, \dots, 0)$ and

$$F^{[j]}(\boldsymbol{x}^{[j]}) = (A^{[j]})^2 h^{[j]}\left(\boldsymbol{x}^{[j]}/A^{[j]}\right), \qquad A^{[j]} > 0, \quad A^{[0]} = 1. \tag{4.1}$$

This scaling preserves the core's curvature bounds, and the normalized coordinates $\boldsymbol{x}^{[j]}/A^{[j]}$ use the original step sizes without modification. Finally, our hard instance function has the form

$$f(\boldsymbol{x}) = \sum_{j=0}^{P} F^{[j]}(\boldsymbol{x}^{[j]}) + \sum_{j=1}^{P} H^{[j]}(\boldsymbol{x}^{[0:j]}), \qquad \text{where } \boldsymbol{x}^{[i:j]} = (\boldsymbol{x}^{[i]}, \dots, \boldsymbol{x}^{[j]}).$$

4.1 Holding, Releasing, and Matching One New Core

We use a smooth nonnegative scalar $r^{[j]}(\boldsymbol{x}^{[0:j]})$, called the *readout*, to measure the weighted size of the existing blocks. We choose its weights so that a sufficiently small readout places every old block inside its quadratic region and keeps the total coupling curvature bounded.

To add block $j+1$, write $\boldsymbol{g}^{[j+1]} := \nabla F^{[j+1]}(\boldsymbol{x}^{[j+1]}_{\mathsf{init}})$ for its initial gradient, fixed across all trials by Sec. 3.3. We construct a coupling $H^{[j+1]}$ whose gradients satisfy

$$\nabla_{\boldsymbol{x}^{[j+1]}} H^{[j+1]} = -\theta \boldsymbol{g}^{[j+1]}, \qquad \nabla_{\boldsymbol{x}^{[0:j]}} H^{[j+1]} = -(1-\theta)\|\boldsymbol{g}^{[j+1]}\| \nabla r^{[j]}, \qquad 0 \le \theta \le 1.$$

At the new block's initial point, $\theta = 1$ above a prescribed *release band*, an interval between two readout levels: the new core gradient is canceled, and the old gradients are unchanged. Below the band, $\theta = 0$, so the incoming correction vanishes. We choose the band so that the readout stays above it at every earlier iterate and crosses below it at one state, which we call the *release*. Thus, we preserve the old trajectory and the new initial point through the release state; the new block can first move at the following update. A drop by a factor of two after the old phase supplies such a band, without requiring monotonicity (Lems. C.1, C.4, & C.6).

After release, the incoming correction may return. In normalized coordinates, the new core evolves as $\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \eta_{t+1}\left(\nabla h_{\boldsymbol{p}}(\boldsymbol{z}_t) - \theta_t \boldsymbol{g}_*\right)$. Because $\boldsymbol{g}_*$ is supported on the first two coordinates, only the first two transfers can be directly affected. The correction is zero when the readout is below the lower release level and the projection satisfies $|\langle \boldsymbol{g}_*/\|\boldsymbol{g}_*\|, \boldsymbol{z}_t\rangle| \le 8(\kappa - 1)$. Steps at most $(4\kappa-1)/2$ preserve the readout bound. These conditions verify the first transfer and, if $b_1 \le (4\kappa-1)/2$, the second (Lems. C.3 & B.9). For larger b_1, we use a *controlled second transfer*, if needed, whose local construction allows the bounded incoming correction. After that transfer reaches coordinate three, the correction acts only on retired coordinates and does not affect later transfers.

We must design the controlled transfer for the correction generated by its own full trajectory. For a prescribed vector $\boldsymbol{u}$ of normalized scalar corrections during the second gap, the local construction provides a continuous map $\mathsf{Shoot}(\boldsymbol{p}, \boldsymbol{u})$ into Trials; a fixed point realizes those corrections (Lem. B.5). For each trial $\boldsymbol{p}$, we collect its actual normalized corrections into $\boldsymbol{u}(\boldsymbol{p})$, which lies in the allowed control range. The finite GD recursion and the trial gradients are continuous, so $\boldsymbol{p} \mapsto \mathsf{Shoot}(\boldsymbol{p}, \boldsymbol{u}(\boldsymbol{p}))$ is a continuous self-map of the compact convex trial set. Brouwer's theorem gives a fixed point (Bonsall, 1962, Lemma 2.4). This matches the core to its own incoming correction, so Eq. (3.4) applies, while all previously certified iterates remain unchanged (Lems. C.5 & C.6).

4.2 Serial Assembly and Total Loss

We now repeat this extension along the prescribed schedule. For positive steps, we append $\eta_{N+1} = 128$ so that the last phase ends at a checkpoint; the target remains $\boldsymbol{x}_N$. We mark each small-step run whose step sum exceeds $(4\kappa-1)/512$ and end each phase before the next marked run, or at $N+1$. The run containing the phase's first update is allowed as its unrestricted leading run. Thus, every phase satisfies the conditions of Sec. 3.2.

We match the first core with zero incoming correction. After a phase, we stop if it reaches $N+1$ or its terminal readout already gives the required lower bound on $\|\boldsymbol{x}_N\|$; otherwise, the readout drops enough to release another core. Each extension passes a distinct marked run, and each such run contains $\Omega(\kappa)$ updates. Thus, the number P of extensions satisfies $P = O(N/\kappa)$. Let $n^{[j]}$ be phase

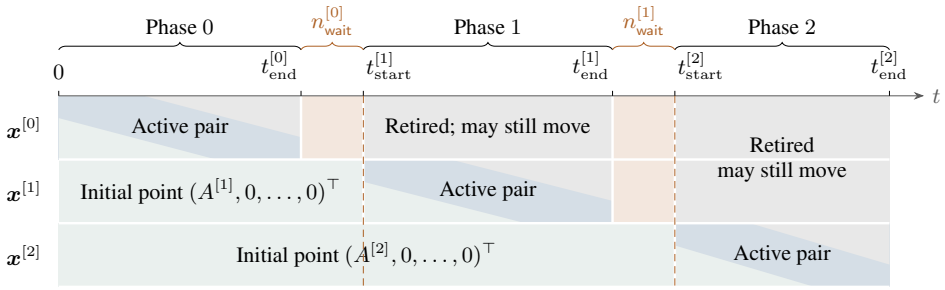


Figure 2: Schematic coordinate activity across phases. Orange bands are waiting intervals; dashed lines mark release states.

j's length and $n^{[j]}_{\mathsf{wait}}$ the number of updates from its endpoint through the next release state. The phase and waiting intervals are disjoint (Fig. 2), so

$$\sum_{j=0}^{P} n^{[j]} + \sum_{j=0}^{P-1} n^{[j]}_{\mathsf{wait}} \le N + 1.$$

The weighted readouts keep the total coupling curvature within the margin left by the phase cores. Hence the final fixed objective $f = \sum_{j=0}^{P} F^{[j]} + \sum_{j=1}^{P} H^{[j]}$ belongs to $\mathcal{F}_{1/\kappa,1}$, is minimized at zero, and has $1 \le \|\boldsymbol{x}_0\|^2 < 2$ (Lems. C.2 & C.7).

The release construction also controls the loss between consecutive block scales:

$$\frac{A^{[j+1]}}{A^{[j]}} \gtrsim \frac{\mathsf{amp}(n^{[j]})}{\kappa^{3/2}(1 + n^{[j]}_{\mathsf{wait}})^2}.$$

Multiplying these bounds, using Eq. (3.3) and the length bounds, and including the polynomial terminal comparison in Lem. C.7 gives

$$-\log R_D \lesssim \log\kappa + \tfrac{N}{q} + \tfrac{N}{\kappa}(\log\kappa + \log q).$$

All implicit constants are absolute. The second term is the accumulated window cost; the last is the total release cost. Since $q = \Theta(\mathcal{Q}(\kappa))$, we have $\log q = O(\log\kappa)$ and $q\log\kappa/\kappa \to 0$, so the last term is absorbed into $O(N/q)$. For absolute constants $C_0, C_1 > 0$, we obtain

$$R_D(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) \ge \kappa^{-C_1} \exp[-C_0\, N/\mathcal{Q}(\kappa)]\,.$$

This proves Lem. 2.1; we account for the hidden constants completely in the appendix (Appx. D).

5 Conclusion

We provided a near-tight pair of upper and lower bounds (up to polylogarithmic factors) for the iteration complexity of GD equipped with an oblivious, deterministic, nonnegative step-size schedule. Our (non-)anytime lower bound (Thm. 1.1) and the periodic-Silver anytime upper bound (Thm. 1.3) jointly identify $\nu = \log_{1+\sqrt{2}} 2 \approx 0.786$ as the optimal polynomial condition-number exponent in both *non-anytime* and *anytime* stopping models. Both bounds share the $\log(1/\varepsilon)$ dependence in accuracy level $\varepsilon > 0$ and differ by at most a factor of order $(\log\kappa)^{\gamma_*} \approx (\log\kappa)^{0.322}$.

It remains open whether the polylogarithmic factor can be removed or is necessary, which we leave as a challenging future work. Extending our technical tools, such as hierarchical segmentation of a given step-size schedule, to other first-order optimization algorithms, such as Heavy-ball momentum (Polyak, 1964; Ma & Zhang, 2026), Nesterov momentum (Nesterov, 1983), or methods with a randomized schedule (e.g., Altschuler & Parrilo 2024), is also an interesting future direction.

AI use statement

In this work, we used generative AI tools for helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical

claims, assisting in the writing of proofs, proposing or refining hypotheses, and assisting with translation. We have not used generative AI tools for designing or providing feedback on research methodology or experiments, and interpreting results. Tasks involving generating synthetic data sets, implementing methods, cleaning and reformatting datasets, and supporting qualitative and thematic data analysis are not applicable to this work. Additionally, we used generative AI tools for creating or modifying scientific figures or images, creating artifacts, drafting parts of a research paper, brainstorming, and editing a research paper to improve readability. We reviewed all AI-assisted work: the first author polished the text, reinterpreted the results, carefully compared them with previous work, verified the correctness of the refined proofs generated by LLMs, and reorganized them. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Jason M. Altschuler and Pablo A. Parrilo. Acceleration by random stepsizes: Hedging, equalization, and the Arcsine stepsize schedule. *arXiv preprint arXiv:2412.05790*, 2024. doi: 10.48550/arXiv.2412.05790. URL https://arxiv.org/abs/2412.05790v2.

Jason M. Altschuler and Pablo A. Parrilo. Acceleration by stepsize hedging: Silver Stepsize Schedule for smooth convex optimization. *Mathematical Programming*, 213:1105–1118, 2025a. doi: 10.1007/s10107-024-02164-2.

Jason M. Altschuler and Pablo A. Parrilo. Acceleration by stepsize hedging: Multi-step descent and the Silver stepsize schedule. *Journal of the ACM*, 72(2):12:1–12:38, 2025b. doi: 10.1145/3708502.

Yossi Arjevani and Ohad Shamir. On the iteration complexity of oblivious first-order optimization algorithms. In *International Conference on Machine Learning*, pp. 908–916. PMLR, 2016.

Yossi Arjevani, Shai Shalev-Shwartz, and Ohad Shamir. On lower and upper bounds in smooth and strongly convex optimization. *Journal of Machine Learning Research*, 17(126):1–51, 2016.

F. F. Bonsall. *Lectures on Some Fixed Point Theorems of Functional Analysis*. Number 26 in Lectures on Mathematics and Physics, Mathematics. Tata Institute of Fundamental Research, Bombay, 1962. URL https://mathweb.tifr.res.in/Documents/Publications/Lectures/tifr26.pdf. Notes by K. B. Vedak.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3–4):231–358, 2015. doi: 10.1561/2200000050.

Augustin Cauchy. Méthode générale pour la résolution des systèmes d'équations simultanées. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 25(16):536–538, 1847.

Yoel Drori and Adrien Taylor. On the oracle complexity of smooth strongly convex minimization. *Journal of Complexity*, 68:101590, 2022.

Benjamin Grimmer, Kevin Shu, and Alex L. Wang. Accelerated objective gap and gradient norm convergence for gradient descent via long steps. *INFORMS Journal on Optimization*, 7(2):156–169, 2025. doi: 10.1287/ijoo.2024.0057.

Minchan Jung, Hanseul Cho, and Chulhee Yun. Stronger lower bounds for (non-)anytime acceleration of gradient descent. *arXiv preprint arXiv:2609.04032*, 2026. URL https://arxiv.org/abs/2609.04032v2.

Guy Kornowski and Ohad Shamir. Open problem: Anytime convergence rate of gradient descent. In *The Thirty Seventh Annual Conference on Learning Theory*, pp. 5335–5339. PMLR, 2024.

Kaizhao Liu and Yuhan Ye. Silver rate is (almost) optimal for gradient descent: The strongly convex case. *arXiv preprint arXiv:2609.26873*, 2026.

Jianhao Ma and Yuxin Chen. A lower bound for stepsize-based acceleration of gradient descent. *arXiv preprint arXiv:2608.10418*, 2026.

Jianhao Ma and Jingzhao Zhang. A lower bound for the heavy-ball method on smooth convex functions. *arXiv preprint arXiv:2609.08656*, 2026.

Arkadii S. Nemirovsky and David B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons, Chichester, 1983. ISBN 978-0-471-10345-5.

Yurii Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, 2 edition, 2018. doi: 10.1007/978-3-319-91578-4.

Yurii E. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics*, 4(5):1–17, 1964.

Boris Teodorovich Polyak. Gradient methods for minimizing functionals (in russian). *Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki*, 3(4):643–653, 1963.

Chung-En Tsai, Ilyas Fatkhullin, Liang Zhang, and Niao He. Lower bounds for anytime acceleration of gradient descent. *arXiv preprint arXiv:2607.02053*, 2026.

Yuhan Ye and Kaizhao Liu. Improved gradient descent lower bounds beyond nesterov. *arXiv preprint arXiv:2609.02855*, 2026a.

Yuhan Ye and Kaizhao Liu. Silver rate is (almost) optimal for gradient descent. *arXiv preprint arXiv:2609.09152*, 2026b. URL https://arxiv.org/abs/2609.09152v2.

David Young. On Richardson's method for solving linear systems with positive definite matrices. *Journal of Mathematics and Physics*, 32(1–4):243–255, 1953. doi: 10.1002/sapm1953321243.

Zehao Zhang and Rujun Jiang. Accelerated gradient descent by concatenation of stepsize schedules. *SIAM Journal on Optimization*, 36(2):1182–1210, 2026. doi: 10.1137/25M173898X.

Zihan Zhang, Jason D. Lee, Simon S. Du, and Yuxin Chen. Anytime acceleration of gradient descent. In *Proceedings of the Thirty-Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 5991–6013. PMLR, 2025. URL https://proceedings.mlr.press/v291/zhang25a.html.

Table of Contents

1 Introduction **1**
1.1 Problem Setup . . . 1
1.2 Our Contributions . . . 2
1.3 Related Work . . . 3

2 Proof Sketches of Main Results **4**
2.1 Proof of Theorem 1.1: Non-Anytime Lower Bound . . . 4
2.2 Proof of Theorem 1.3: Anytime Upper Bound . . . 5

3 Checkpoint Selection and Construction of Core Functions **5**
3.1 From a Single Coordinate Pair to the Full Coordinate Chain . . . 5
3.2 Checkpoint Selection and the Silver Time Scale . . . 6
3.3 A Phase Core Certificate . . . 7

4 Coupling Phase Cores into our Hard Function **7**
4.1 Holding, Releasing, and Matching One New Core . . . 8
4.2 Serial Assembly and Total Loss . . . 8

5 Conclusion **9**

Guide to the Proof **14**

A Checkpoint Selection and Cost Bounds **16**
A.1 Scalar costs and eligible startups . . . 16
A.2 Buffered optimization . . . 18
A.3 The possible extremal splits . . . 21
A.4 Checkpoint-numerator expansion and the step-sum bound . . . 23
A.5 Step-sum bounds for empty-optimal schedules . . . 29
A.6 Buffered windows and all-prefix composition . . . 30
A.7 Repairing internal gaps without changing the scale . . . 31

B Building and Realizing Phase Cores **36**
B.1 Projection geometry and backward shooting . . . 36
B.2 Ordinary and exceptional transfers . . . 41
B.3 A controlled second transfer . . . 44
B.4 Compatible transfer chains . . . 47
B.5 Uniform families and matched phases . . . 48

C Assembling the Hard Instance **54**
C.1 Core and coupling definitions . . . 54
C.2 Cancellation and release . . . 56
C.3 Readout weights and initial point . . . 57
C.4 Global curvature . . . 57

C.5 A safe region for the actual earlier coordinates 59
C.6 Fresh readout crossings 60
C.7 Matching and extending one phase 60
C.8 The schedule partition 62
C.9 The finite serial certificate 63

D Completing the Lower Bound **66**
D.1 Accumulated scalar loss 66
D.2 The finite lower bound 67
D.3 Quadratic baseline 67

E Silver-Block Contraction and Prefix Stability **68**
E.1 Primitive certificates and balanced concatenation 68
E.2 The Silver endpoint inequality 70
E.3 Joining-step control and prefix stability 71

F Other Convergence Metrics **72**
F.1 Relative function value gap and relative squared gradient norm 72
F.2 Normalization by initial squared distance 74

G Summary of Notation **76**

GUIDE TO THE PROOF

Sec. 2 assembles the main lower and upper bounds from results proved in the appendices. The lower bound requires one hard instance for each prescribed finite schedule; the upper bound requires one infinite schedule that works for every admissible function at every iterate after the accuracy-dependent threshold. We explain the two arguments below, with their principal dependencies collected in Fig. 3. Both arguments first normalize to $L = 1$ and $\mu = 1/\kappa$.

Lower bound: choosing checkpoints. The construction keeps amplitude in a succession of fresh coordinates. A *checkpoint* is a selected update that transfers amplitude to the next coordinate; consecutive unselected updates form a *gap*. Long gaps can lose too much amplitude through the strongly convex quadratic term. Appx. A therefore groups positions into *windows* and controls both numerical checkpoint costs and unselected step sums. The step-sum majorant in Prop. A.11 determines the window length $q = \Theta(\mathcal{Q}(\kappa))$, which sets the lower-bound time scale. Lems. A.12, A.13, & A.15 then select one checkpoint set, control all its selected prefixes, and insert checkpoints to shorten nonleading gaps. These conclusions apply to the phase segments specified in Appx. B.5; the leading gap remains uncut. Window boundaries organize this accounting without restarting GD.

Lower bound: constructing and realizing a phase core. Appx. B constructs ordinary, exceptional, and controlled transfers, and explains how compatibility lets consecutive coordinate pairs follow their local trajectories. The numerical checkpoint costs alone do not include all amplitude losses: Lem. B.7 also accounts for damping and threshold factors. Its formal amplitude bounds hold for every trial and, together with the local zero radii, yield the static family in Lem. B.8, with a common quadratic ball and a common initial gradient. These two properties allow us to choose a coupling before selecting the trial. The actual trajectory bound is a separate conclusion: under the readout and projection conditions, and after matching any controlled second transfer to its incoming correction, Lem. B.9 bounds every iterate in the phase away from zero.

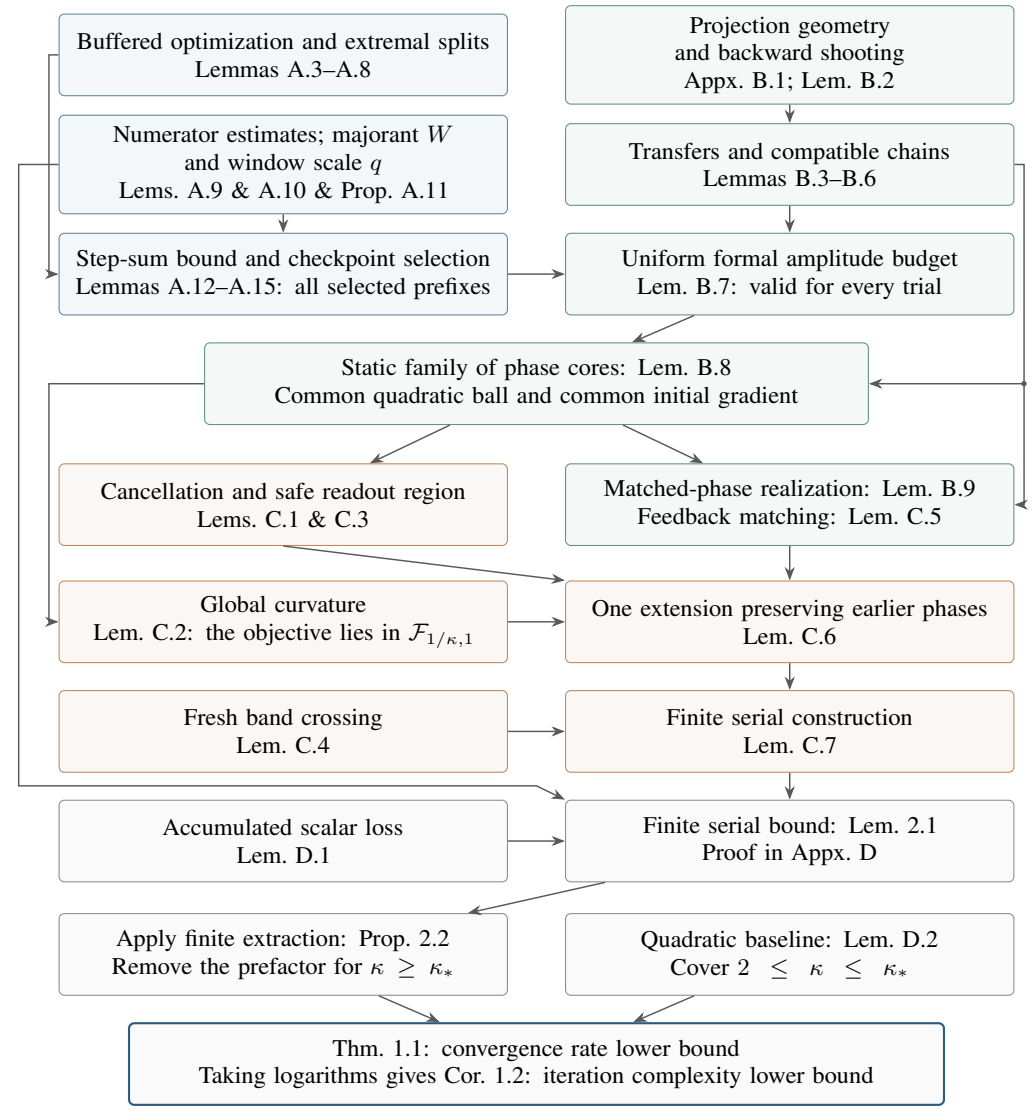
Lower bound: assembling one fixed objective. Appx. C couples the phase cores so that each future coordinate block remains at its initial point until release. Cancellation, the safe readout region, and a fresh band crossing control this release. The feedback-matching lemma selects a trial consistent with the correction generated by its own GD run. Together, these ingredients give Lem. C.6: append a matched phase while preserving the earlier certified trajectory exactly. The global curvature bound keeps the objective in $\mathcal{F}_{1/\kappa,1}$. The schedule partition and induction in Lem. C.7 then give a finite serial construction certificate for every positive schedule. This construction chooses the function from the schedule; the final function is fixed throughout GD.

Lower bound: summing losses and removing the prefactor. In Appx. D, Lem. D.1 combines the phase, waiting, and scale losses in that certificate. The bounds on q from Prop. A.11 convert the result into Lem. 2.1, namely $R_D \geq \kappa^{-C_1} \exp[-C_0 N/\mathcal{Q}(\kappa)]$. Zero steps can be deleted without changing the iterates. The construction applies above an absolute lower threshold on κ, with loss constants fixed before the schedule and its length. Prop. 2.2 removes the prefactor by applying this bound to a repeated schedule and selecting one copy with sufficiently little contraction. The quadratic baseline, Lem. D.2, covers the remaining bounded range of κ. Sec. 2 combines these results to prove Thm. 1.1; taking logarithms gives Cor. 1.2.

Upper bound: endpoint contraction and prefix stability. Appx. E proves the two block properties used in Sec. 2.2. Lem. E.1 gives the balanced primitive-certificate concatenation argument of Zhang & Jiang (2026); applying it along the Silver recursion yields the endpoint inequality in Lem. 2.3. That inequality also controls the joining step between two smaller blocks. Induction then proves Lem. 2.4: every prefix stays within the starting distance of its block, although individual steps can increase distance. Both lemmas hold for convex, 1-smooth functions with a minimizer. For every $\kappa \geq 1$, Sec. 2.2 uses strong convexity to contract the squared distance by a fixed factor over a block of length $O(\kappa^{\nu})$. Repeating that block and controlling the final incomplete block by prefix stability proves Thm. 1.3. Dividing the normalized step sizes by L gives a schedule that depends only on L, μ, independent of the horizon and target accuracy.

Appx. F derives the guarantees for the other convergence metrics; Appx. G collects the shared notation.

(a) Lower bound: construction and final reduction



(b) Upper bound: from Silver blocks to every iterate

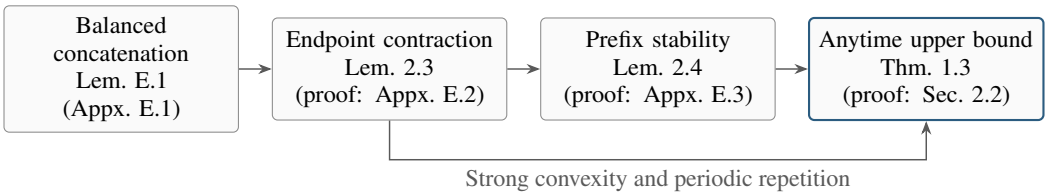


Figure 3: Principal proof dependencies. Arrows point from inputs to results that use them; grouped boxes omit internal dependencies and subsidiary estimates. In (a), the long left arrow records the second use of the majorant: its window-scale bounds also enter the finite serial bound. The static family holds for every trial, whereas the trajectory conclusion requires matching. In (b), both endpoint contraction and prefix stability enter the anytime bound.

A Checkpoint Selection and Cost Bounds

To construct one phase core, we must choose checkpoints that keep transfer costs small and leave gaps that the local geometry can realize. The numerical cost $c(a)(a+s)/b$ couples these choices: selecting a position changes both the checkpoint size b and the surrounding gap step sums s, while the startup must remain eligible, $a \le b$. We first bound the worst optimized cost by reducing it to the step sum of an empty-optimal schedule. Splitting such schedules into shorter gaps gives the majorant W and the window length q used in Sec. 3.2.

The remaining task is to make this local optimization serve one fixed core at every iterate. We choose one set in each buffered window, prove bounds for every prefix of that same set, and compose the windows while retaining any unfinished gap. We then insert checkpoints to shorten internal gaps that exceed the local transfer bounds. The final result, Lem. A.15, supplies fixed checkpoint positions and startup parameters, prefix cost bounds, and gap bounds. These are the inputs used in Appx. B to construct transfer functions and control their amplitudes; no GD trajectory is asserted by the scalar selection alone. Throughout this section, $(\eta_1, \dots, \eta_n)$ is a finite segment of the prescribed nonnegative schedule. Its index labels positions in this segment. Every gap sum and virtual buffer uses these same step units. Throughout the construction, κ is sufficiently large; every implicit constant and the required lower threshold are absolute. Write $\rho = 1+\sqrt{2}$, $\nu = \log_\rho 2$, and $\alpha = 1/\nu$; γ_* is the constant in Eq. (1.5).

This optimization follows the checkpoint viewpoint of Jung et al. (2026), with startup-dependent costs and the eligibility constraint retained throughout.

A.1 Scalar costs and eligible startups

Geometric parameterization. To bound the gap-end second coordinate and the derivative threshold relative to the output amplitude, choose $\tau_\zeta = 3\zeta^2/2 \le \min\{\zeta/2, 3/32\}$ for $0 < \zeta \le 1/4$. The scalar quantities used in the transfer costs are

$$\Phi_\zeta = \arctan(1/(2\zeta)), \qquad \mathsf{rad}_{\zeta,0} = \sqrt{1+4\zeta^2}\, e^{\zeta\Phi_\zeta}, \qquad \mathsf{rad}_{\mathsf{turn}}(\zeta) = \sqrt{1+4\zeta^2}\, e^{(\zeta+\zeta^{-1})\Phi_\zeta}.$$

The startup threshold below makes the ordinary derivative-threshold ratio at most $\tau_\zeta(1-\tau_\zeta)$, and $c(a_\zeta)$ is its cost correction:

$$a_\zeta = \frac{4\kappa-1}{4(\kappa-1)} \frac{4\zeta \mathsf{rad}_{\mathsf{turn}}(\zeta)}{\tau_\zeta(1-\tau_\zeta)^2}, \quad c(a_\zeta) = \frac{\mathsf{rad}_{\zeta,0}}{(1-\tau_\zeta)^2}, \quad a_{\min} = a_{1/4}. \tag{A.1}$$

Ordinary transfers allow $a \le (4\kappa-1)/4$; the buffered checkpoint analysis uses $4\kappa - 1 \ge 16 a_{\min}$ and $a \le (4\kappa-1)/8$. Smaller pitches will reduce the correction but increase the startup threshold. To express the cost derivative, define

$$\mathsf{slope}(\zeta) = \Phi_\zeta + \frac{2\zeta}{1+4\zeta^2} + \frac{6\zeta}{1-3\zeta^2/2} > 0, \qquad \mathsf{bal}(\zeta) := -(\log a_\zeta)'/\mathsf{slope}(\zeta) - 1.$$

Here primes denote derivatives with respect to ζ.

Lemma A.1 (Calibration). *The map $\zeta \mapsto a_\zeta$ is continuously differentiable and strictly decreases from infinity to $a_{\min}$ on $(0, 1/4]$, and its derivative never vanishes. Its inverse $\zeta(a)$ makes the correction $c(a)$ well-defined and decreasing for $a \ge a_{\min}$, with*

$$1 < c(a) \le c(a_{\min}) < 9/5, \qquad a_{\min} > 64. \tag{A.2}$$

The function $\mathsf{bal}(\zeta)$ is positive and strictly decreasing, and $a_\zeta \mathsf{bal}(\zeta)$ is strictly decreasing. For $0 < \zeta \le 1/4$ and every gap step sum $s \ge 0$,

$$\begin{gathered} \mathsf{bal}(\zeta) > \frac{1}{3\zeta^2}, \qquad \mathsf{bal}(1/4) > 6, \\ \frac{d}{d\zeta}[c(a_\zeta)(a_\zeta + s)] = c(a_\zeta)\mathsf{slope}(\zeta)[s - a_\zeta \mathsf{bal}(\zeta)]. \end{gathered} \tag{A.3}$$

Moreover, $\mathsf{slope}(\zeta) = (\log c(a_\zeta))' > 0$ and $(\frac{4(\kappa-1)}{4\kappa-1})a_\zeta > e^{1/\zeta}$. Writing $a_\zeta = (\frac{4\kappa-1}{4(\kappa-1)})a^\circ_\zeta$ removes all κ dependence from a°_ζ, $c(a_\zeta)$, and $\mathsf{bal}(\zeta)$. Consequently $a_\zeta/a_{\min}$ is independent of κ.

Proof. Differentiating the factors $\mathsf{rad}_{\zeta,0}$ and $(1-\tau_\zeta)^{-2}$ in $c(a_\zeta)$ gives

$$(\log \mathsf{rad}_{\zeta,0})' = \Phi_\zeta + \frac{2\zeta}{1+4\zeta^2}, \qquad (\log \mathsf{rad}_{\zeta,0})'' = -\frac{16\zeta^2}{(1+4\zeta^2)^2}.$$

These identities and the derivative of $(1-\tau_\zeta)^{-2}$ give the asserted formula for $(\log c(a_\zeta))'$. The definition of $\mathsf{bal}(\zeta)$ gives

$$\begin{aligned} \mathsf{bal}(\zeta) &= \zeta^{-2} - 2 - \frac{5+3\zeta^2/2}{\zeta \mathsf{slope}(\zeta)(1-3\zeta^2/2)}, \\ (\log a_\zeta)' &= -(\mathsf{bal}(\zeta)+1)\mathsf{slope}(\zeta). \end{aligned} \tag{A.4}$$

Since $(\log \mathsf{rad}_{\zeta,0})'$ decreases, its endpoint value gives $(\log \mathsf{rad}_{\zeta,0})' > 3/2$. Also $6 \le 6/(1-3\zeta^2/2) \le 64/7$. Consequently,

$$\begin{aligned} (2/3 - 2\zeta^2)\mathsf{slope}(\zeta) - \zeta\left(\frac{6}{1-3\zeta^2/2} - 1\right) &\ge 13/16 + \zeta - \frac{11\zeta}{4(1-3\zeta^2/2)} \\ &\ge 13/16 - (67/21)\zeta \ge 5/336 > 0. \end{aligned}$$

The left side equals $\zeta^2 \mathsf{slope}(\zeta)(\mathsf{bal}(\zeta) - 1/(3\zeta^2))$, proving the lower bound and $a'_\zeta < 0$. For monotonicity, rewrite the same identity as

$$\zeta^2 \mathsf{bal}(\zeta) = 1 - 2\zeta^2 - \frac{5+3\zeta^2/2}{6 + (\zeta^{-1} - 3\zeta/2)\,(\Phi_\zeta + 2\zeta/(1+4\zeta^2))}.$$

The fraction has increasing positive numerator and decreasing positive denominator, so $(\zeta^2 \mathsf{bal}(\zeta))' < 0$. Since $\mathsf{bal}(\zeta) > 0$, this implies $\mathsf{bal}'(\zeta) < 0$; hence $a_\zeta \mathsf{bal}(\zeta)$ also strictly decreases. The product rule gives Eq. (A.3), and $a'_\zeta < 0$ makes the inverse differentiable. The exact startup formula is

$$a_\zeta = \frac{4\kappa - 1}{4(\kappa-1)} \frac{8}{3\zeta} c(a_\zeta) e^{\Phi_\zeta/\zeta} \longrightarrow \infty \quad (\zeta \downarrow 0). \tag{A.5}$$

Thus the range is $[a_{\min}, \infty)$, and $c(a)$ decreases because $\mathsf{slope}(\zeta) > 0$. Its formula gives $c(a) > 1$. For the upper bound, use $\Phi_{1/4} = \pi/4 + \arctan(1/3) < 10/9$, $\sqrt{5} < 161/72$, and $e^{5/18} \le 1 + 5/18 + (5/18)^2/[2(1-5/54)] < 119/90$. Consequently,

$$c(a_{\min}) = \frac{512\sqrt{5}}{841} e^{\Phi_{1/4}/4} < \frac{512 \cdot 161 \cdot 119}{841 \cdot 72 \cdot 90} < \frac{9}{5}.$$

At $\zeta = 1/4$, Eq. (A.5) gives $a_{\min} > (32/3)e^4 > 64$, since $\Phi_{1/4} > 1$, $c(a_{\min}) > 1$, and $\frac{4\kappa-1}{4(\kappa-1)} > 1$. For cost minimization, (A.3) says that an interior optimal pitch satisfies $s = a_\zeta \mathsf{bal}(\zeta)$. The checkpoint analysis in Appx. A needs only the elementary bound

$$\mathsf{slope}(1/4) = \arctan 2 + \frac{2}{5} + \frac{48}{29} > 3, \qquad \mathsf{bal}(1/4) = 14 - \frac{652}{29\mathsf{slope}(1/4)} > \frac{566}{87} > 6.$$

Here $\arctan 2 = \pi/4 + \arctan(1/3) > 3/4 + 1/3 - 1/81 > 1$. Also $\Phi_\zeta \ge \arctan 2 > 1$, so $(\frac{4(\kappa-1)}{4\kappa-1})a_\zeta > e^{1/\zeta}$ and the inverse-logarithmic pitch lower-bound argument applies. The normalization claims follow directly from Eq. (A.1). □

For an ordinary transfer, $c(a)(a+s)/b$ is the numerical transfer cost: a is the startup parameter, s the gap step sum, and b the checkpoint size. We optimize over a while retaining the eligibility constraint $a \le b$. Virtual buffers add step sum at the two ends of a local problem; they are bookkeeping quantities, not actual steps. The terminal numerator records the cost still pending after the last selected position. For $s \ge 0$ and $4\kappa - 1 \ge 16 a_{\min}$, define the *terminal* and *eligible checkpoint numerators* by

$$\begin{aligned} \mathcal{N}(s) &= \min_{a_{\min} \le a \le (4\kappa-1)/8} c(a)(a+s), \\ \mathcal{N}_b(s) &= \min_{a_{\min} \le a \le \min\{b,(4\kappa-1)/8\}} c(a)(a+s), \quad b \ge a_{\min}. \end{aligned} \tag{A.6}$$

The minima are attained. Both functions are positive, continuous, strictly increasing, and concave, with increments between $s_2 - s_1$ and $c(a_{\min})(s_2 - s_1)$ for $s_2 > s_1$. Indeed, they are pointwise minima of affine functions with slopes in $[1, c(a_{\min})]$. In particular,

$$a_{\min} + s \le \mathcal{N}(s) \le \mathcal{N}_b(s) \le 2(a_{\min} + s). \tag{A.7}$$

Only positions with size strictly greater than $a_{\min}$ are numerically selectable. The local parameter at that position may equal its size.

We also define the uncapped checkpoint numerator and the numerical exceptional cost by

$$\begin{aligned} \mathcal{N}_\infty(s) &= \min_{0<\zeta\le 1/4} c(a_\zeta)(a_\zeta + s), \\ \text{numerical exceptional cost} &= \frac{(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3}{b} \quad (b > 64). \end{aligned} \tag{A.8}$$

The uncapped minimum is attained because $a_\zeta \to \infty$ as $\zeta \downarrow 0$. The exceptional factor is a numerical budget; Lem. B.4 will realize it with a transfer function.

Lemma A.2 (Eligible startup and elasticity)**.** *For $s \ge 0$ and $a_{\max} \ge a_{\min}$, the minimizing startup in $[a_{\min}, a_{\max}]$ is unique and nondecreasing in s. It equals $a_{\min}$ when $s \le a_{\min}\mathsf{bal}(1/4)$; otherwise it is the smaller of $a_{\max}$ and the unique root $a > a_{\min}$ of $s = a\mathsf{bal}(\zeta(a))$. For $a_{\min} < b \le (4\kappa-1)/8$, if the eligible endpoint $a = b$ minimizes, then $s \ge b\mathsf{bal}(\zeta(b)) > 6b$. Every one-sided negative logarithmic elasticity of $\mathcal{N}_b(s)/b$ with respect to $b > a_{\min}$ lies in $[1, 8/7]$.*

Proof. Lem. A.1 gives

$$\partial_a \log[c(a)(a+s)] = \frac{1}{a+s} - \frac{1}{a(\mathsf{bal}(\zeta(a)) + 1)}.$$

Its sign is that of $a\mathsf{bal}(\zeta(a)) - s$. Both $\mathsf{bal}(\zeta(a))$ and $a\mathsf{bal}(\zeta(a))$ strictly increase. The minimizing startup is consequently unique and nondecreasing in s, with its unique turning point clipped to the allowed interval. For $a_{\min} < b < (4\kappa-1)/8$, if the eligible endpoint $a = b$ minimizes, then

$$s \ge b\mathsf{bal}(\zeta(b)) > 6b, \qquad -b\partial_b \log \frac{\mathcal{N}_b(s)}{b} = 1 + \frac{1}{\mathsf{bal}(\zeta(b)) + 1} - \frac{b}{b+s} \in [1, 8/7]. \tag{A.9}$$

Indeed, $\mathsf{bal}(\zeta(b)) > 6$ and $1/[\mathsf{bal}(\zeta(b)) + 1] < 1/7$. If eligibility does not bind, the elasticity is one. At $b = (4\kappa-1)/8$, increasing b leaves the startup cap fixed, so the right elasticity is one. If the cap minimizes, $s \ge b\mathsf{bal}(\zeta(b))$ and the left elasticity is given by the active formula; otherwise, the left elasticity is one. At a transition between active and inactive eligibility below the cap, $s = b\mathsf{bal}(\zeta(b))$, and the active formula also equals one. Thus, both one-sided elasticities lie in $[1, 8/7]$ at the endpoints as well, which is the bound needed for the perturbations below. □

A.2 Buffered optimization

Virtual buffers allow a window to be analyzed before its actual neighboring gaps are known. They add step sum to the first and last gaps without adding schedule positions. Keeping the terminal numerator in the objective is essential: a later checkpoint may still use the gap pending at the end of the window. For a nonempty checkpoint set $\mathcal{T} = \{t_1 < \cdots < t_k\}$, set $t_0 = 0$ and define

$$b_i = \eta_{t_i} > a_{\min}, \qquad s_i = \sum_{t=t_{i-1}+1}^{t_i-1} \eta_t \quad (1 \le i \le k), \qquad s_{k+1} = \sum_{t=t_k+1}^{n} \eta_t.$$

Thus, s_i is the step sum before checkpoint i, and s_{k+1} is the unselected terminal step sum. These sums exclude the selected steps; consecutive checkpoints have a gap sum of zero. For virtual left and right buffers $s_{\mathsf{L}}, s_{\mathsf{R}} \ge 0$, define

$$\begin{aligned} \Psi(\mathcal{T}) &= \mathcal{N}(s_{\mathsf{R}} + s_{k+1}) \frac{\mathcal{N}_{b_1}(s_{\mathsf{L}} + s_1)}{b_1} \prod_{i=2}^{k} \frac{\mathcal{N}_{b_i}(s_i)}{b_i}, \qquad \mathcal{T} \ne \varnothing, \\ \Psi(\varnothing) &= \mathcal{N}(s_{\mathsf{L}} + s_{\mathsf{R}} + s_{\mathsf{tot}}), \qquad s_{\mathsf{tot}} = \sum_{j=1}^{n} \eta_j. \end{aligned} \tag{A.10}$$

A schedule is *empty-optimal* if $\Psi(\varnothing) \le \Psi(\mathcal{T})$ for every eligible checkpoint set $\mathcal{T}$. A checkpoint set is *tight* if equality holds. The product of the checkpoint factors in Eq. (A.10), excluding the terminal numerator, is the selected product.

We also allow a general positive, continuous, strictly increasing, concave terminal numerator $\mathcal{N}_{\mathsf{term}}$. Then

$$\Psi_{\mathsf{term}}(\mathcal{T}) = \mathcal{N}_{\mathsf{term}}(s_{k+1}) \frac{\mathcal{N}_{b_1}(s_{\mathsf{L}} + s_1)}{b_1} \prod_{i=2}^{k} \frac{\mathcal{N}_{b_i}(s_i)}{b_i}, \qquad \Psi_{\mathsf{term}}(\varnothing) = \mathcal{N}_{\mathsf{term}}(s_{\mathsf{L}} + s_{\mathsf{tot}}).$$

The original terminal numerator is $\mathcal{N}_{\mathsf{term}}(s) = \mathcal{N}(s_{\mathsf{R}} + s)$. Later, we will need $\mathcal{N}_{\mathsf{term}}(s) = \mathcal{N}_{b_{\mathsf{rec}}}(s_{\mathsf{rem}} + s)$ with an already fixed next checkpoint of size b_{rec}.

The next two proofs adapt the attainment and intersection arguments of Jung et al. (2026, Lemmas C.1–C.2 and Appendix C.6) and Ye & Liu (2026b, Lemmas B.3–B.4) to these eligible numerators and general terminal functions.

Lemma A.3 (Intersections of minimizing checkpoint sets)**.** *For a fixed finite nonnegative schedule, a left buffer $s_{\mathsf{L}} \ge 0$, and any positive, continuous, strictly increasing, concave terminal numerator $\mathcal{N}_{\mathsf{term}}$, minimizing checkpoint sets are closed under union and intersection. The common intersection of all nonempty minimizing checkpoint sets, when such checkpoint sets exist, is nonempty and minimizing.*

Proof. For $s_{\mathsf{inc}} > 0$, eligibility and positive concavity give

$$\frac{\mathcal{N}_b(s + s_{\mathsf{inc}})}{\mathcal{N}_b(s)} \ge \frac{s + b + s_{\mathsf{inc}}}{s + b}, \qquad \frac{\mathcal{N}_{\mathsf{term}}(s)}{\mathcal{N}_{\mathsf{term}}(s + s_{\mathsf{inc}})} > \frac{s}{s + s_{\mathsf{inc}}}. \tag{A.11}$$

To prove the first inequality, choose a startup $a \le b$ that attains the numerator at $s + s_{\mathsf{inc}}$, and bound the denominator using that same line. The resulting ratio is $(a+s+s_{\mathsf{inc}})/(a+s) \ge (s+b+s_{\mathsf{inc}})/(s+b)$. For the second, concavity gives $\mathcal{N}_{\mathsf{term}}(s) \ge [s/(s + s_{\mathsf{inc}})]\mathcal{N}_{\mathsf{term}}(s + s_{\mathsf{inc}}) + [s_{\mathsf{inc}}/(s + s_{\mathsf{inc}})]\mathcal{N}_{\mathsf{term}}(0)$, and $\mathcal{N}_{\mathsf{term}}(0) > 0$.

Inserting a checkpoint of size b after step sum s_1 and before step sum s_2, with next numerator $\mathcal{N}_{\mathsf{term}}$, changes the cost by

$$\mathsf{ins}(s_1, s_2; b, \mathcal{N}_{\mathsf{term}}) = \frac{\mathcal{N}_b(s_1)\mathcal{N}_{\mathsf{term}}(s_2)}{b\mathcal{N}_{\mathsf{term}}(s_1 + b + s_2)}.$$

This is the cost after insertion divided by the cost before insertion: factors outside the affected gap cancel. Thus, a ratio below one means that inserting the checkpoint lowers the cost. For $\varepsilon_{\mathsf{pert}} > 0$, (A.11) yields

$$\frac{\mathsf{ins}(s_1 + \varepsilon_{\mathsf{pert}}, s_2; b, \mathcal{N}_{\mathsf{term}})}{\mathsf{ins}(s_1, s_2; b, \mathcal{N}_{\mathsf{term}})} > \frac{s_1 + b + \varepsilon_{\mathsf{pert}}}{s_1 + b} \frac{s_1 + b + s_2}{s_1 + b + s_2 + \varepsilon_{\mathsf{pert}}} \ge 1.$$

Thus, ins strictly increases with s_1. It weakly increases with s_2: positivity and concavity of $\mathcal{N}_{\mathsf{term}}$ make $\log \mathcal{N}_{\mathsf{term}}$ concave, so its increment from s_2 to $s_2 + s_1 + b$ is nonincreasing in s_2. The logarithm of the insertion ratio contains the negative of this increment. Adding a checkpoint before b strictly decreases s_1. If instead a new checkpoint of size b' becomes its next checkpoint, with preceding step sum s_{next}, put $s_{\mathsf{inc}} = s_1 + b$; the old s_2 satisfies $s_2 \ge s_{\mathsf{next}} + b'$, and

$$\frac{\mathcal{N}_{b'}(s_{\mathsf{next}})}{\mathcal{N}_{b'}(s_{\mathsf{next}} + s_{\mathsf{inc}})} \le \frac{s_{\mathsf{next}} + b'}{s_{\mathsf{next}} + b' + s_{\mathsf{inc}}} \le \frac{s_2}{s_2 + s_{\mathsf{inc}}} < \frac{\mathcal{N}_{\mathsf{term}}(s_2)}{\mathcal{N}_{\mathsf{term}}(s_2 + s_{\mathsf{inc}})}.$$

Hence, adding checkpoints can only decrease an insertion ratio, strictly when its preceding or next checkpoint changes. The same proof includes the left buffer in the first step sum s_1.

Taking logarithms turns each insertion ratio into the change in $\log \Psi_{\mathsf{term}}$ when one checkpoint is added. The preceding comparison says that this change can only decrease if more checkpoints are already selected. For checkpoint sets $\mathcal{T}_1, \mathcal{T}_2$, adding the elements of $\mathcal{T}_1 \setminus \mathcal{T}_2$ successively to $\mathcal{T}_1 \cap \mathcal{T}_2$ and to $\mathcal{T}_2$ gives

$$\log \Psi_{\mathsf{term}}(\mathcal{T}_1) + \log \Psi_{\mathsf{term}}(\mathcal{T}_2) \ge \log \Psi_{\mathsf{term}}(\mathcal{T}_1 \cup \mathcal{T}_2) + \log \Psi_{\mathsf{term}}(\mathcal{T}_1 \cap \mathcal{T}_2).$$

The inequality is strict when $\mathcal{T}_1, \mathcal{T}_2$ are disjoint and nonempty, since at least one comparison changes a preceding or next checkpoint. Consequently, minimizing sets are closed under union and intersection. Strictness for disjoint nonempty sets implies that any two nonempty minimizers intersect. Successive intersections therefore give a nonempty common minimizing set. □

Lemma A.4 (Attainment of the optimized value). *For fixed length $n \geq 0$, left buffer $s_{\mathsf{L}} \geq 0$, and any positive, continuous, strictly increasing, concave terminal numerator $\mathcal{N}_{\mathsf{term}}$, the supremum over $\boldsymbol{\eta} \in [0,\infty)^n$ of $\min_{\mathcal{T}} \Psi_{\mathsf{term}}(\mathcal{T})$ is positive, finite, and attained. Every maximizing schedule is empty-optimal. In particular, the statements hold for Ψ with fixed nonnegative buffers.*

Proof. For $b > a_{\min}$, $\mathcal{N}_b(s)$ depends continuously on (b, s), because its compact feasible interval varies continuously. Assign a checkpoint-set cost the value $+\infty$ when a selected size is at most $a_{\min}$. At an eligible schedule the cost is continuous; at an ineligible schedule its assigned value is $+\infty$, which makes the upper-semicontinuity inequality automatic. Thus, each extended cost is upper semicontinuous, and so is their finite minimum.

Finiteness follows by induction on n, for every fixed $s_{\mathsf{L}}, \mathcal{N}_{\mathsf{term}}$. At $n = 0$, the value is $\mathcal{N}_{\mathsf{term}}(s_{\mathsf{L}})$. If all steps are at most $a_{\min}$, the empty checkpoint set bounds it by $\mathcal{N}_{\mathsf{term}}(s_{\mathsf{L}} + na_{\min})$. Otherwise, select the largest step $b_{\max} > a_{\min}$ as the first checkpoint. Its preceding step sum is at most $nb_{\max}$, so its factor is at most $2(a_{\min} + s_{\mathsf{L}} + nb_{\max})/b_{\max} \leq 2(1 + s_{\mathsf{L}}/a_{\min} + n)$. After this selected step, optimize the shorter suffix with initial buffer zero and the same terminal numerator. Induction bounds its value uniformly, regardless of the remaining step sizes. Only finitely many suffix lengths exist.

If a maximizing sequence were unbounded, take a subsequence along which every step size either converges to a finite limit or tends to infinity, and at least one tends to infinity. This extraction is possible because there are finitely many positions. Select the earliest divergent position as the first checkpoint, of size b. Its preceding step sum s stays bounded, and its selected factor satisfies

$$\frac{\mathcal{N}_b(s_{\mathsf{L}} + s)}{b} \leq \frac{2(a_{\min} + s_{\mathsf{L}} + s)}{b} \longrightarrow 0.$$

Its optimized suffix remains uniformly bounded by the preceding induction. The total optimized value would then tend to zero, contradicting the positive value of the all-zero schedule. A maximizing sequence is therefore bounded, and upper semicontinuity gives attainment.

At a value maximizer, suppose the empty checkpoint set does not minimize. Lem. A.3 supplies a checkpoint common to every minimizing checkpoint set. Decrease its size slightly, keeping it above $a_{\min}$. Its gap step sums in these checkpoint sets do not change, its numerator weakly increases as eligibility shrinks, and its denominator strictly decreases. Thus, every minimizing cost increases. The finitely many other eligible checkpoint sets have a positive cost gap above the minimum; continuity preserves this gap for a sufficiently small change. Decreasing a size creates no newly eligible set. This contradicts value maximality. □

Lemma A.5 (Closing an eligible terminal numerator). *For $b \geq a_{\min}$ and $0 \leq s_1 \leq s_2$,*

$$\frac{\mathcal{N}_b(s_1)}{\mathcal{N}_b(s_2)} \leq \frac{\mathcal{N}(s_1)}{\mathcal{N}(s_2)}. \tag{A.12}$$

Replacing a terminal numerator $\mathcal{N}_b(s_{\mathsf{R}} + s)$ by $\mathcal{N}(s_{\mathsf{R}} + s)$ therefore preserves empty optimality with the same selected factors.

Proof. It suffices to show that $\mathcal{N}_b/\mathcal{N}$ is nondecreasing. Choose a_{s_1} attaining $\mathcal{N}(s_1)$ and a_{s_2} attaining $\mathcal{N}_b(s_2)$ in Eq. (A.6). If $a_{s_1} \leq b$, it is also eligible for $\mathcal{N}_b(s_1)$, so $\mathcal{N}_b(s_1)/\mathcal{N}(s_1) = 1 \leq \mathcal{N}_b(s_2)/\mathcal{N}(s_2)$. Otherwise, $a_{s_2} \leq b < a_{s_1}$, so $c(a_{s_2})(a_{s_2} + s)/[c(a_{s_1})(a_{s_1} + s)]$ increases with s. Using the minimizing properties at s_1 and s_2, respectively, gives

$$\begin{aligned}\frac{\mathcal{N}_b(s_1)}{\mathcal{N}(s_1)} &\leq \frac{c(a_{s_2})(a_{s_2} + s_1)}{c(a_{s_1})(a_{s_1} + s_1)} \\ &\leq \frac{c(a_{s_2})(a_{s_2} + s_2)}{c(a_{s_1})(a_{s_1} + s_2)} \leq \frac{\mathcal{N}_b(s_2)}{\mathcal{N}(s_2)}.\end{aligned}$$

All denominators are positive, so rearrangement proves Eq. (A.12).

For a fixed nonempty checkpoint set, write s_1 for the terminal argument after the last selected position and s_2 for the argument of the empty checkpoint set. Then $s_1 \leq s_2$, and the ratio of the two costs is the selected product times $\mathcal{N}_b(s_1)/\mathcal{N}_b(s_2)$. The replacement by $\mathcal{N}$ increases this ratio by Eq. (A.12). An empty-optimal schedule therefore remains empty-optimal. □

Lemma A.6 (Compactness of the empty-optimal schedules)**.** *For every fixed length, left buffer, and positive, continuous, strictly increasing, concave terminal numerator, the set of empty-optimal nonnegative schedules is nonempty and compact.*

Proof. The all-zero schedule belongs to it. Closedness follows by passing each selectable checkpoint-set inequality to a limit: a checkpoint set eligible at the limit is selectable nearby. For boundedness, an eligible singleton at position i and monotonicity of the terminal numerator give

$$1 \le \frac{\mathcal{N}_{\eta_i}(s_{\mathsf{L}} + \sum_{j<i} \eta_j)}{\eta_i} \frac{\mathcal{N}_{\mathsf{term}}(\sum_{j>i} \eta_j)}{\mathcal{N}_{\mathsf{term}}(s_{\mathsf{L}} + \sum_j \eta_j)} \le \frac{\mathcal{N}_{\eta_i}(s_{\mathsf{L}} + \sum_{j<i} \eta_j)}{\eta_i}.$$

Thus, $\eta_i \le \mathcal{N}_{\eta_i}(s_{\mathsf{L}} + \sum_{j<i} \eta_j) \le 2(a_{\min} + s_{\mathsf{L}} + \sum_{j<i} \eta_j)$. The same final bound holds if $\eta_i \le a_{\min}$. Induction in i bounds every coordinate, completing compactness. □

A.3 The possible extremal splits

A schedule is empty-optimal when selecting no checkpoint minimizes the buffered objective. Bounding its maximum step sum will bound the optimized value for arbitrary schedules by Lem. A.4. A nonempty tight checkpoint set ties with the empty checkpoint set; its common checkpoints split an extremal schedule into shorter empty-optimal problems. Eligibility is active at a checkpoint when its minimizing startup equals the checkpoint size. The proof distinguishes these active splits from the single inactive checkpoint that yields the binary merge. We next maximize the actual step sum over empty-optimal schedules of fixed length and buffers. Lem. A.6 ensures that this continuous step sum attains a maximum. For nonnegative X, Y, define $J_{\mathcal{N}}(X, Y) > X + Y$ by the unique root

$$(J_{\mathcal{N}} - X - Y)\mathcal{N}(J_{\mathcal{N}}) = \mathcal{N}(X)\mathcal{N}(Y).$$

The left side, as a function of the unknown greater than $X + Y$, strictly increases from zero to infinity. The root is unique. Increasing an input decreases that left side at any old candidate and increases its right side, proving monotonicity.

We add the same reserve $s_{\mathsf{res}} = C_{\mathsf{res}} a_{\min}$, with $C_{\mathsf{res}} \ge 1$, to each child gap. This reserve is a bookkeeping addition to the step sum, not a change to the startup threshold or to the transfer family. Splitting at k checkpoints supplies k additional copies of the reserve, which help absorb the selected step sizes. Lem. A.8 chooses an absolute C_{res} for that purpose.

Lemma A.7 (The possible extremal splits)**.** *At an empty-optimal schedule with maximum step sum and at least one eligible checkpoint, the common intersection $\mathcal{T}_0$ of all nonempty tight checkpoint sets is nonempty and has size $k \le 3$. If b is the largest common size, every common size lies in $[7b/8, b]$. At most one common checkpoint is inactive, and no consecutive common checkpoints are both active, where activity is measured in $\mathcal{T}_0$. For any reserve $s_{\mathsf{res}} = C_{\mathsf{res}} a_{\min} \ge a_{\min}$, put*

$$s_{\mathsf{buf}} = s_{\mathsf{L}} + s_{\mathsf{R}} + s_{\mathsf{tot}}, \quad X_{\mathsf{tot}} = s_{\mathsf{res}} + s_{\mathsf{buf}},$$
$$X_0 = s_{\mathsf{res}} + s_{\mathsf{L}} + s_1, \quad X_i = s_{\mathsf{res}} + s_{i+1}\ (0 < i < k), \quad X_k = s_{\mathsf{res}} + s_{\mathsf{R}} + s_{k+1}.$$

Each child gap is empty-optimal after closing its terminal onto $\mathcal{N}$. Only the first and last children inherit the corresponding outer buffers. If $\mathcal{T}_0$ is a singleton with inactive eligibility, then $X_{\mathsf{tot}} \le J_{\mathcal{N}}(X_0, X_1)$.

Proof. The proof first restricts the possible common checkpoints of the tight sets. Perturbations of their sizes leave every tight set's gaps unchanged, which lets us compare the cost change with the gain in total step sum. Once the common checkpoint set has at most three checkpoints, its gaps give the child problems. **Step 1: restrict the common checkpoints.** At least one nonempty checkpoint set must be tight: otherwise, slightly increasing one already selectable size preserves all finitely many strict constraints and increases step sum. Lem. A.3 supplies the nonempty common intersection $\mathcal{T}_0$. Call a common checkpoint active when eligibility binds at its gap in $\mathcal{T}_0$. Every other common checkpoint is locally inactive in every nonempty tight checkpoint set, because tight checkpoint supersets shorten its gap and therefore cannot increase its minimizing startup.

If two common sizes $b < b'$ satisfy $b' > (8/7)b$, perturb them to $b - \varepsilon_{\mathsf{pert}}, b' + \mathsf{tilt}\varepsilon_{\mathsf{pert}}$, where $\mathsf{tilt} > 1$ is close enough to one. Both checkpoints belong to every nonempty tight checkpoint set, so neither perturbed size contributes to a gap in any of these sets. Their gaps therefore stay unchanged. By

Eq. (A.9), the one-sided derivative of every difference between a tight checkpoint-set cost and the empty checkpoint-set cost is at least

$$\mathcal{N}(s_{\mathsf{buf}})(1/b - 8\mathsf{tilt}/(7b')) - \mathcal{N}'_+(s_{\mathsf{buf}})(\mathsf{tilt} - 1) > 0$$

for a suitable $\mathsf{tilt} > 1$. All other constraints retain slack. The step sum increases, a contradiction. Thus, if b is the largest common checkpoint size,

$$7b/8 \le \eta_t \le b \qquad (t \in \mathcal{T}_0). \tag{A.13}$$

There are no two unequal inactive common sizes under the same perturbation with both elasticities equal to one. For two equal inactive sizes b, use $b - \varepsilon_{\mathsf{pert}}, b + \varepsilon_{\mathsf{pert}} + \varepsilon^2_{\mathsf{pert}}/(2b)$. Inactivity leaves a neighborhood in which neither size restricts its minimizing startup. Their numerators therefore remain unchanged, and their denominator product is $b^2 - \varepsilon^2_{\mathsf{pert}}/2 + O(\varepsilon^3_{\mathsf{pert}})$. Every nonempty tight checkpoint-set cost increases by $\mathcal{N}(s_{\mathsf{buf}})\varepsilon^2_{\mathsf{pert}}/(2b^2) + o(\varepsilon^2_{\mathsf{pert}})$, while the empty checkpoint-set cost increases by $\mathcal{N}'_+(s_{\mathsf{buf}})\varepsilon^2_{\mathsf{pert}}/(2b) + o(\varepsilon^2_{\mathsf{pert}})$. Their difference is positive because positive concavity gives $\mathcal{N}(s_{\mathsf{buf}}) \ge s_{\mathsf{buf}}\mathcal{N}'_+(s_{\mathsf{buf}}) + \mathcal{N}(0) > b\mathcal{N}'_+(s_{\mathsf{buf}})$. Again, the other constraints retain slack, and the step sum increases. Hence, at most one common checkpoint is inactive.

Two consecutive common checkpoints cannot both be active. Let their sizes be b, b' and their preceding step sums $s_1 \ge 6b$, $s_2 \ge 6b'$. Their insertion ratio for the first checkpoint is bounded below by

$$\frac{\mathcal{N}_b(s_1)\mathcal{N}_{b'}(s_2)}{b\mathcal{N}_{b'}(s_1 + b + s_2)} \ge \frac{s_1 s_2}{2b(s_1 + s_2 + b + b')} \ge \frac{6^2}{2(6+1)(1+8/7)} = \frac{6}{5} > 1.$$

For the first bound, the two numerator factors are at least s_1 and s_2, while the denominator numerator is at most $2(a_{\min} + s_1 + b + s_2) \le 2(s_1 + b + s_2 + b')$. Next, $b'/b \in [7/8, 8/7]$ by Eq. (A.13); the middle fraction increases in s_1, s_2, and substituting their lower endpoints gives the last bound. Deleting the first checkpoint would strictly decrease a tight checkpoint-set cost, contrary to empty optimality. With at most one inactive checkpoint and no adjacent active checkpoints, $k \le 3$, with only the pattern active-inactive-active possible when $k = 3$.

Step 2: identify the children and handle an inactive singleton. Keep the checkpoints of $\mathcal{T}_0$ fixed and insert checkpoints only inside one gap. All factors outside the gap and its next checkpoint cancel. The ratio of full costs is therefore exactly the ratio to the empty checkpoint-set cost of that child problem. It cannot be below one. The child is thus empty-optimal for its next eligible terminal numerator, or for the final terminal numerator. Lem. A.5 closes eligible terminal numerators onto $\mathcal{N}$ and preserves the inherited buffers.

For an inactive singleton, tightness and inactivity give $b\mathcal{N}(s_{\mathsf{buf}}) = \mathcal{N}(s_{\mathsf{L}} + s_1)\mathcal{N}(s_{\mathsf{R}} + s_2)$. If $b \le s_{\mathsf{res}}$, then $X_{\mathsf{tot}} \le X_0 + X_1 < J_{\mathcal{N}}(X_0, X_1)$. Otherwise, positive concavity gives $\mathcal{N}(s_{\mathsf{buf}} + s_{\mathsf{res}}) \le (1 + s_{\mathsf{res}}/s_{\mathsf{buf}})\mathcal{N}(s_{\mathsf{buf}})$; also $(b - s_{\mathsf{res}})(1 + s_{\mathsf{res}}/s_{\mathsf{buf}}) \le b$, since $s_{\mathsf{buf}} \ge b$. Therefore

$$(X_{\mathsf{tot}} - X_0 - X_1)\mathcal{N}(X_{\mathsf{tot}}) = (b - s_{\mathsf{res}})\mathcal{N}(s_{\mathsf{buf}} + s_{\mathsf{res}}) \le b\mathcal{N}(s_{\mathsf{buf}}) \le \mathcal{N}(X_0)\mathcal{N}(X_1).$$

If $X_{\mathsf{tot}} \le X_0 + X_1$, the bound is immediate; otherwise, it follows from strict increase on the domain defining the root. □

A.3.1 The power bound for active splits

We choose the reserve large enough to handle all active splits by one estimate.

Lemma A.8 (The power bound for active splits)**.** *There is an absolute $C_{\mathsf{res}} \ge 2$ such that, for every sufficiently large κ, each configuration in Lem. A.7 containing an active checkpoint satisfies, with $s_{\mathsf{res}} = C_{\mathsf{res}} a_{\min}$,*

$$X^\nu_{\mathsf{tot}} \le \sum_{i=0}^{k} X^\nu_i. \tag{A.14}$$

The choice of C_{res} is independent of the schedule and its length.

Proof. Choose an absolute $C_{\mathsf{act}} \ge 1$ so large that

$$\nu(49/16)C^{\nu-1}_{\mathsf{act}} \le 16^{-\nu}.$$

This is possible because $0 < \nu < 1$. Next choose $C_{\mathsf{res}} \geq 2$ so that $\mathsf{bal}(\zeta(C_{\mathsf{res}} a_{\min}/2)) \geq 2C_{\mathsf{act}}$. The choice is independent of κ: by Lem. A.1, $a_\zeta/a_{\min}$ is independent of κ, as is $\mathsf{bal}(\zeta)$, and $\mathsf{bal}(\zeta(a)) \to \infty$ as $a/a_{\min} \to \infty$.

Let b be the largest common checkpoint size. If $b \leq s_{\mathsf{res}}$, then $X_{\mathsf{tot}} = \sum_i X_i + \sum_j b_j - k s_{\mathsf{res}} \leq \sum_i X_i$, and subadditivity proves (A.14). Otherwise, choose any active checkpoint of size b_j and let s be its preceding gap sum, including the left buffer when $j = 1$. By (A.13), $b_j \geq b/2 > C_{\mathsf{res}} a_{\min}/2$. Activity and (A.9) give

$$s \geq b_j \mathsf{bal}(\zeta(b_j)) \geq C_{\mathsf{act}} b,$$

so its preceding child $X_* = s_{\mathsf{res}} + s$ is at least $C_{\mathsf{act}} b$.

Tightness forces another child to have a value $X_i \geq b/16$. Indeed, in the identity

$$\mathcal{N}(s_{\mathsf{buf}}) \prod_{j=1}^{k} b_j = \mathcal{N}(X_k - s_{\mathsf{res}}) \prod_{j=1}^{k} \mathcal{N}_{b_j}(X_{j-1} - s_{\mathsf{res}}),$$

the active factor is $c(b_j)(b_j + s) \leq 4s$, while $\mathcal{N}(s_{\mathsf{buf}}) \geq s_{\mathsf{buf}} \geq s > 0$. After dividing by that factor, the product of the remaining k numerators is at least $\frac{1}{4} \prod_j b_j$. Each remaining numerator is at most twice its corresponding X_i by (A.7) and $s_{\mathsf{res}} \geq a_{\min}$. Since $b_j \geq b/2$, it follows that

$$\prod_{i \neq *} X_i \geq \frac{b^k}{4 \cdot 4^k}, \qquad \max_{i \neq *} X_i \geq \frac{b}{4 \cdot 4^{1/k}} \geq b/16.$$

Finally, $k \leq 3$ gives $X_{\mathsf{tot}} \leq \sum_i X_i + 3b$. For fixed $b > 0$, the deficit $\sum_i X_i^\nu - (\sum_i X_i + 3b)^\nu$ increases in each nonnegative coordinate, by differentiation and continuity. Lower one child to $C_{\mathsf{act}} b$, another to $b/16$, and all others to zero. Concavity then gives

$$\sum_i X_i^\nu - X_{\mathsf{tot}}^\nu \geq b^\nu \left[C_{\mathsf{act}}^\nu + 16^{-\nu} - (C_{\mathsf{act}} + 49/16)^\nu \right] \geq b^\nu \left[16^{-\nu} - \nu C_{\mathsf{act}}^{\nu-1} (49/16) \right] \geq 0.$$

This proves (A.14) for every active configuration with the same absolute reserve. □

A.4 Checkpoint-numerator expansion and the step-sum bound

The split classification turns the optimization into an induction on the number of positions. To close that induction, one majorant must dominate every permitted combination of child gaps, including unequal binary splits and splits with active eligibility. The idealized balanced identity in Sec. 3.2 identifies the polynomial exponent, but the positive geometric correction prevents the pure power n^α from closing the actual induction. We first expand the optimized numerator to quantify this correction, then absorb it by a slowly varying increase in the majorant's logarithmic growth rate.

We construct a continuous increasing function W and write $W_n := W(n)$ at integer indices. These values bound step sums for empty-optimal schedules, with additional indices accounting for the virtual buffers. The resulting window bound controls the unselected step sum of an arbitrary schedule. Thus W measures step sum, whereas q counts positions; (A.21) chooses q so that $W(4q)$ fits the damping budget. The exact integral defining W gives finite bounds as well as the logarithmic exponent in its asymptotic growth. The next proof extends the concavity argument of Jung et al. (2026, Lemma C.4) and Ye & Liu (2026b, Lemma B.8) to every $1 < \alpha' < 2$.

Lemma A.9 (Generalized Silver inequality). *For $1 < \alpha' < 2$ and a split fraction $\mathsf{frac} \in [0, 1]$,*

$$\mathsf{frac}^{\alpha'} + (1 - \mathsf{frac})^{\alpha'} + (2^{2\alpha'} - 2^{\alpha'+1})[\mathsf{frac}(1 - \mathsf{frac})]^{\alpha'} \leq 1 \qquad (0 \leq \mathsf{frac} \leq 1). \tag{A.15}$$

Moreover, $(2^{2\alpha} - 2^{\alpha+1}) = 1$ and

$$\frac{d}{d\alpha'}(2^{2\alpha'} - 2^{\alpha'+1})\Big|_{\alpha'=\alpha} = 2\sqrt{2}\rho \log 2, \qquad 2\alpha\sqrt{2}\rho \log 2 > 11/2. \tag{A.16}$$

Proof. The expression is symmetric about $1/2$ and equals one at 0 and $1/2$. We will show that on this half interval it first decreases and then increases, so no interior point can exceed those endpoint

values. On $(0, 1/2)$, the derivative of the left side of Eq. (A.15), divided by $\alpha'[\mathsf{frac}(1-\mathsf{frac})]^{\alpha'-1}$, is

$$\mathsf{der}(\mathsf{frac}) = (1-\mathsf{frac})^{1-\alpha'} - \mathsf{frac}^{1-\alpha'} + (2^{2\alpha'} - 2^{\alpha'+1})(1-2\mathsf{frac}),$$
$$\mathsf{der}''(\mathsf{frac}) = \alpha'(\alpha'-1)[(1-\mathsf{frac})^{-\alpha'-1} - \mathsf{frac}^{-\alpha'-1}] < 0.$$

We have $\mathsf{der}(0+) = -\infty$, $\mathsf{der}'(0+) = +\infty$, $\mathsf{der}(1/2) = 0$, and $\mathsf{der}'(1/2) = 2^{\alpha'+1}(\alpha'+1-2^{\alpha'}) < 0$. The last sign follows because $2^{\alpha'} - \alpha' - 1$ vanishes at one and has a positive increasing derivative thereafter. Strict concavity implies that der changes sign once, from negative to positive, before vanishing at $1/2$. The original expression therefore decreases and then increases on $[0, 1/2]$; its endpoint values are one. Symmetry proves Eq. (A.15). Writing $\rho = 2^{\alpha}$ gives Eq. (A.16). The inequality $\log x > 2(x-1)/(x+1)$ for $x > 1$, proved by differentiation, implies $2\alpha\sqrt{2}\rho \log 2 > 4\sqrt{2} > 11/2$. □

Lemma A.10 (Two-sided checkpoint-numerator expansion)**.** *For the uncapped checkpoint numerator in Eq. (A.8), as $s \to \infty$,*

$$\frac{\mathcal{N}_\infty(s)}{s} = 1 + \frac{\pi^2}{4\log s} + O\left(\frac{\log\log s}{(\log s)^2}\right). \tag{A.17}$$

For $4\kappa - 1 \geq 16 a_{\min}$,

$$\mathcal{N}(s) = \mathcal{N}_\infty(s) \quad (0 \leq s \leq (4\kappa-1)/2), \qquad \mathcal{N}(s) \geq \mathcal{N}_\infty(s) \quad (s \geq 0).$$

All constants in the expansion are absolute, uniformly for sufficiently large κ, since $1 < (4\kappa - 1)/[4(\kappa-1)] < 2$.

Proof. By Lem. A.2, the minimizing pitch is $\zeta = 1/4$ for $s \leq a_{\min}\mathsf{bal}(1/4)$ and otherwise uniquely solves $s = a_\zeta \mathsf{bal}(\zeta)$. This follows because $a_\zeta \mathsf{bal}(\zeta)$ strictly decreases from infinity to $a_{\min}\mathsf{bal}(1/4)$, and the cost derivative has the sign of $s - a_\zeta \mathsf{bal}(\zeta)$. In particular, the minimizing pitch tends to zero as $s \to \infty$.

Taylor expansion of the explicit functions at $\zeta = 0$, using $\Phi_\zeta = \pi/2 - \arctan(2\zeta)$ and (A.5), gives

$$\log a_\zeta = \pi/(2\zeta) - \log\zeta + \log[(8/3)(\frac{4\kappa-1}{4(\kappa-1)})] - 2 + \pi\zeta/2 + O(\zeta^2), \qquad c(a_\zeta) = 1 + \pi\zeta/2 + O(\zeta^2). \tag{A.18}$$

The remainders are bounded uniformly on a fixed sufficiently small interval: $\arctan(2\zeta) = 2\zeta + O(\zeta^3)$ and $\log c(a_\zeta) = \pi\zeta/2 + O(\zeta^2)$ give the displayed expressions directly.

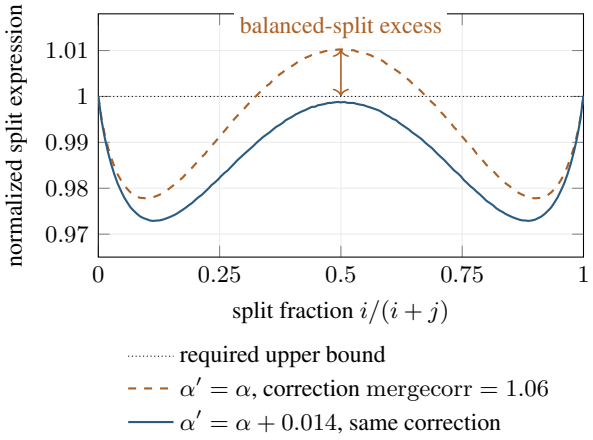
At the interior minimizer for large s, Eq. (A.4) gives $\mathsf{bal}(\zeta) = \zeta^{-2} + O(\zeta^{-1})$, hence $a_\zeta/s = 1/\mathsf{bal}(\zeta) = \zeta^2(1 + O(\zeta))$. Taking logarithms in this relation and using Eq. (A.18) yields $\pi/(2\zeta) = \log s + 3\log\zeta - \log[(8/3)(\frac{4\kappa-1}{4(\kappa-1)})] + 2 + O(\zeta)$. Equivalently, $\log s = \pi/(2\zeta) + 3\log(1/\zeta) + O(1)$. First, $1/\zeta \sim 2\log s/\pi$; then $\pi/(2\zeta) = \log s - 3\log\log s + O(1)$, and therefore

$$\zeta = \frac{\pi}{2\log s} + O\left(\frac{\log\log s}{(\log s)^2}\right), \qquad \frac{\mathcal{N}_\infty(s)}{s} = c(a_\zeta)(1 + a_\zeta/s) = 1 + \frac{\pi^2}{4\log s} + O\left(\frac{\log\log s}{(\log s)^2}\right).$$

This proves Eq. (A.17).

To check the cap uniformly, an interior minimizer satisfies $s = a_\zeta \mathsf{bal}(\zeta)$ and $\mathsf{bal}(\zeta) > 1/(3\zeta^2)$. Thus, $a_\zeta < 3\zeta^2 s \leq 3(4\kappa-1)/32 < (4\kappa-1)/8$ when $s \leq (4\kappa-1)/2$. At the other possible minimizer $\zeta = 1/4$, its startup is $a_{\min} \leq (4\kappa-1)/16$. Each uncapped minimizer is therefore admissible under $a \leq (4\kappa-1)/8$ on that range. The global inequality follows by minimizing over fewer parameters. □

Fig. 4 isolates the obstruction at a balanced split: a positive numerator correction pushes the split expression above one at exponent α. The proof absorbs the actual, scale-dependent correction by a varying elasticity; a fixed increase in the exponent is only a local illustration.



In the proof
The correction decreases with the scale.
$\alpha' - \alpha \sim \dfrac{\alpha\gamma_*}{\log n}$
Integrating this elasticity gives the factor $[\log(en)]^{\alpha\gamma_*}$ in $W(n)$.

Figure 4: Illustration of the merge test in Prop. A.11 & Eq. (A.25): $\mathrm{frac}^{\alpha'} + (1-\mathrm{frac})^{\alpha'} + \mathrm{mergecorr}[\mathrm{frac}(1-\mathrm{frac})]^{\alpha'}$, with $\mathrm{frac} = i/(i+j)$ and illustrative $\mathrm{mergecorr} = 1.06$. Increasing α' above α offsets the excess at balanced splits. The proof uses scale-dependent elasticity to obtain the logarithmic factor in W.

Why the correction is a power of a logarithm. The *elasticity* of a positive differentiable function W is $nW'(n)/W(n) = d\log W(n)/d\log n$. For $W(n) = n^{\alpha'}$ it equals the constant exponent α'; allowing it to vary lets the majorant absorb a correction that changes with scale. At the step-sum bound $W_n \approx n^{\alpha}$, the checkpoint-numerator correction in a balanced merge exceeds one by a leading term $\pi^2/(4\alpha\log n)$. The Silver coefficient responds as $(2^{2\alpha'} - 2^{\alpha'+1}) - 1 \sim 2\sqrt{2}\rho(\log 2)(\alpha' - \alpha)$. Balancing the two requires local elasticity $\alpha' - \alpha \sim \alpha\gamma_*/\log n$, where $\alpha\gamma_* = \pi^2/(8\sqrt{2}\rho\log\rho)$. Integrating this elasticity gives $W_n \approx n^{\alpha}(\log n)^{\alpha\gamma_*}$. The proposition below uses the exact checkpoint numerator to obtain a computable bound at exactly $\alpha\gamma_*$.

The following majorant controls the startup, buffers, and actual step sum of empty-optimal schedules, as proved in Lem. A.12.

Proposition A.11 (An explicit majorant at the sharp logarithmic scale)**.** *For every fixed absolute reserve multiplier $C_{\mathsf{res}} \geq 1$ and all sufficiently large κ, there is an increasing convex function $W : [0, \infty) \to [0, \infty)$ with $W(0) = 0$, $W(1) > C_{\mathsf{res}} a_{\min}$, and elasticity in $[\alpha, \alpha_{\max}]$ for $n > 0$ and an absolute $\alpha < \alpha_{\max} < 2$. It obeys*

$$W(n)^\nu \leq \frac{n}{n_*} W(n_*)^\nu \qquad (0 \leq n \leq n_*,\ n_* > 0) \tag{A.19}$$

and the finite merge bound

$$J_{\mathcal{N}}(W_i, W_j) \leq W_{i+j} \qquad (i, j \in \mathbb{N},\ W_{i+j} \leq (4\kappa - 1)/4). \tag{A.20}$$

Moreover,

$$W(n) = \Theta(n^\alpha [\log(en)]^{\alpha\gamma_*}) \qquad (n \geq 1).$$

The initialization, comparison constants, and lower threshold on κ may depend on the fixed C_{res}, but not on a schedule or its length. After fixing C_{res}, the comparison constants are uniform for sufficiently large κ. Use the integer window length

$$q = \max\big(\{0\} \cup \{j \in \mathbb{N} : W(4j) \leq (4\kappa - 1)/256\}\big). \tag{A.21}$$

For sufficiently large κ, this integer satisfies

$$\begin{gathered} (4\kappa - 1)/1024 < W_{4q} \leq (4\kappa - 1)/256, \\ q = \Theta((\kappa - 1)^\nu / (\log(\kappa - 1))^{\gamma_*}), \qquad q \log(\kappa - 1)/(\kappa - 1) \longrightarrow 0. \end{gathered} \tag{A.22}$$

In particular, $q = \Theta(\mathcal{Q}(\kappa))$ and, after enlarging the absolute lower threshold on κ, $q \geq 2(\log(\kappa - 1) + 2)^{1/(\alpha-1)}$.

Proof. Let $\mathsf{corr}(s) = \mathcal{N}_\infty(s)/s$ for $s > 0$. The selected geometric family has $\mathsf{bal}(1/4) > 1$.

Monotonicity of $\mathsf{corr}(s) = \mathcal{N}_\infty(s)/s$. The unique minimizing startup a_s equals $a_{\min}$ up to $s = a_{\min}\mathsf{bal}(1/4)$, and satisfies $s = a_s \mathsf{bal}(\zeta(a_s))$ above it. The positive logarithmic decay rate of corr is

$$\mathsf{decay}(s) := -\frac{d \log \mathsf{corr}(s)}{d \log s} = \frac{a_s}{a_s + s} > 0.$$

Indeed, differentiating at the minimizing startup gives $\mathcal{N}'_\infty(s) = c(a_s)$ on both branches. Here decay decreases: it equals $a_{\min}/(a_{\min} + s)$ on the endpoint branch and $1/[1 + \mathsf{bal}(\zeta(a_s))]$ on the interior branch. The formulas agree at the transition.

Choosing absolute initialization constants. Since $(2^{2\alpha} - 2^{\alpha+1}) = 1 < 2c(a_{\min})$, choose $\alpha < \alpha_{\max} < 2$ close enough to α that $(2^{2\alpha_{\max}} - 2^{\alpha_{\max}+1}) < 2c(a_{\min})$. By Lem. A.10, $\mathsf{corr}(s) \to 1$, so choose $C_{\mathsf{init}} > C_{\mathsf{res}}$ large enough that $\mathsf{corr}(C_{\mathsf{init}} a_{\min})^2 / \mathsf{corr}(2^{\alpha_{\max}} C_{\mathsf{init}} a_{\min}) < (2^{2\alpha_{\max}} - 2^{\alpha_{\max}+1})$. Finally choose $C_{\mathsf{cut}} > 0$ sufficiently small: since $\alpha > 1$,

$$[1 + \mathsf{corr}(C_{\mathsf{init}} a_{\min})] C_{\mathsf{cut}}^\alpha + (1 - C_{\mathsf{cut}})^\alpha = 1 - \alpha C_{\mathsf{cut}} + o(C_{\mathsf{cut}}) < 1.$$

These choices are independent of κ. Indeed, by Lem. A.1, writing $a_\zeta = \frac{4\kappa-1}{4(\kappa-1)} a^\circ_\zeta$ removes all κ dependence from a°_ζ, $c(a_\zeta)$, and $\mathsf{bal}(\zeta)$. Consequently $\mathsf{corr}(a_{\min} t)$ is the same function of t for every κ.

These choices have $C_{\mathsf{init}} \geq 1$, $\alpha < \alpha_{\max} < 2$, and $0 < C_{\mathsf{cut}} < 1/2$ and satisfy

$$\begin{gathered} \frac{\mathsf{corr}(C_{\mathsf{init}} a_{\min})^2}{\mathsf{corr}(2^{\alpha_{\max}} C_{\mathsf{init}} a_{\min})} \leq (2^{2\alpha_{\max}} - 2^{\alpha_{\max}+1}), \qquad [1 + \mathsf{corr}(C_{\mathsf{init}} a_{\min})] C_{\mathsf{cut}}^\alpha + (1 - C_{\mathsf{cut}})^\alpha \leq 1, \\ (2^{2\alpha_{\max}} - 2^{\alpha_{\max}+1}) < 2c(a_{\min}). \end{gathered} \tag{A.23}$$

Construction of the majorant. Define the inverse W^{-1} by

$$\begin{aligned}\mathsf{mergecorr}(s) &= \frac{\mathsf{corr}(C_{\mathsf{cut}}^{\alpha_{\max}} s)\mathsf{corr}((1-C_{\mathsf{cut}})^{\alpha_{\max}} s)}{\mathsf{corr}(s)},\\ \alpha_{\mathsf{el}}(s) &= \min\{\alpha_{\max}, \log_2(1+\sqrt{1+\mathsf{mergecorr}(s)})\},\\ \log W^{-1}(s) &= \int_{C_{\mathsf{init}}a_{\min}}^{s} \frac{d\log t}{\alpha_{\mathsf{el}}(t)}, \qquad s \ge C_{\mathsf{init}}a_{\min}.\end{aligned} \tag{A.24}$$

Thus, $W(1) = C_{\mathsf{init}}a_{\min}$ and $nW'(n)/W(n) = \alpha_{\mathsf{el}}(W(n))$ for $n \ge 1$. This gives $W_1 = C_{\mathsf{init}}a_{\min} > C_{\mathsf{res}}a_{\min}$. For $0 \le n \le 1$, set $W(n) = C_{\mathsf{init}}a_{\min}\, n^{\alpha_{\mathsf{el}}(C_{\mathsf{init}}a_{\min})}$. The elasticity α_{el} is chosen so that the generalized Silver inequality absorbs the numerator correction in every binary merge. Using the decrease of corr established above, we prove that the resulting W is convex and check the merge for balanced and unbalanced splits. The expansion of the defining integral will give the exact logarithmic exponent. On the varying branch, the formula for $\alpha_{\mathsf{el}}(s)$ is the positive-root solution of $2^{2\alpha_{\mathsf{el}}(s)} - 2^{\alpha_{\mathsf{el}}(s)+1} = \mathsf{mergecorr}(s)$. Thus, its generalized Silver coefficient is exactly the correction that the merge must absorb. The cap $\alpha_{\max}$ handles the initial range, where the first condition in Eq. (A.23) supplies the required bound instead.

Decreasing elasticity and convexity. The logarithmic derivative of $\mathsf{mergecorr}(s)$ is $-\mathsf{decay}(C_{\mathsf{cut}}^{\alpha_{\max}} s) - \mathsf{decay}((1-C_{\mathsf{cut}})^{\alpha_{\max}} s) + \mathsf{decay}(s) < 0$. Also $\mathsf{mergecorr} > 1$ and $\mathsf{mergecorr} \to 1$. Thus, α_{el} is continuous and nonincreasing, and $\alpha < \alpha_{\mathsf{el}} \le \alpha_{\max}$. It has at most one transition from a constant initial value $\alpha_{\max}$ to the varying branch. On that branch, $\mathsf{mergecorr}(s) \le (2^{2\alpha_{\max}} - 2^{\alpha_{\max}+1}) < 2c(a_{\min})$. Since $\mathsf{mergecorr}(s) \ge \mathsf{corr}(C_{\mathsf{cut}}^{\alpha_{\max}} s)$, the endpoint identity $\mathsf{corr}(s) = c(a_{\min})(1 + a_{\min}/s)$ for $s \le a_{\min}$ implies $C_{\mathsf{cut}}^{\alpha_{\max}} s > a_{\min}$; also $(1-C_{\mathsf{cut}})^{\alpha_{\max}} s > a_{\min}$. Consequently, $\mathsf{decay}(C_{\mathsf{cut}}^{\alpha_{\max}} s), \mathsf{decay}((1-C_{\mathsf{cut}})^{\alpha_{\max}} s) < \mathsf{decay}(a_{\min}) = 1/2$, and

$$\begin{aligned}\frac{d\alpha_{\mathsf{el}}}{d\log s} &= \frac{2^{\alpha_{\mathsf{el}}} - 2}{2(\log 2)(2^{\alpha_{\mathsf{el}}} - 1)}[-\mathsf{decay}(C_{\mathsf{cut}}^{\alpha_{\max}} s) - \mathsf{decay}((1-C_{\mathsf{cut}})^{\alpha_{\max}} s) + \mathsf{decay}(s)]\\ &> -\frac{2^{\alpha_{\mathsf{el}}} - 2}{2(\log 2)(2^{\alpha_{\mathsf{el}}} - 1)} > -(\alpha_{\mathsf{el}} - 1).\end{aligned}$$

For the last inequality, strict convexity of $\alpha_{\mathsf{el}} \mapsto 2^{2\alpha_{\mathsf{el}}} - 2^{\alpha_{\mathsf{el}}+1}$ places its tangent slope at α_{el} above its secant slope from 1. Thus, the ratio of its value to its derivative, namely $(2^{\alpha_{\mathsf{el}}} - 2)/[2(\log 2)(2^{\alpha_{\mathsf{el}}} - 1)]$, is smaller than $\alpha_{\mathsf{el}} - 1$. Therefore

$$W''(n) = \frac{W(n)\alpha_{\mathsf{el}}}{n^2}\left[\alpha_{\mathsf{el}} - 1 + \frac{d\alpha_{\mathsf{el}}}{d\log s}\right] > 0.$$

Both W and W' match at a branch transition and at $n = 1$. For $n \ge 1$, $d\log(W(n)/n^{\alpha})/d\log n = \alpha_{\mathsf{el}}(W(n)) - \alpha \ge 0$, so the ratio $W(n)/n^{\alpha}$ increases. For $0 < n \le 1$, the defining power $W(n) = C_{\mathsf{init}}a_{\min}\, n^{\alpha_{\mathsf{el}}(C_{\mathsf{init}}a_{\min})}$ gives the same monotonicity. Since $\alpha\nu = 1$ and $W(0) = 0$, this monotonicity proves Eq. (A.19). This scaling inequality will let us add the child bounds in an active split: their indices sum to the parent index, so the factors n/n_* sum to one.

Every binary split. Fix $i, j \ge 1$ with $W_{i+j} \le (4\kappa - 1)/4$. The defining root is at most W_{i+j} exactly when $(W_{i+j} - W_i - W_j)\mathcal{N}(W_{i+j}) \ge \mathcal{N}(W_i)\mathcal{N}(W_j)$. All three arguments lie below $(4\kappa - 1)/4$, so uncapped agreement turns this condition into

$$\frac{W_i + W_j}{W_{i+j}} + \frac{\mathcal{N}_\infty(W_i)\mathcal{N}_\infty(W_j)}{W_{i+j}\mathcal{N}_\infty(W_{i+j})} \le 1. \tag{A.25}$$

If $\alpha_{\mathsf{el}}(W_{i+j}) = \alpha_{\max}$, monotonicity of α_{el} gives $\alpha_{\mathsf{el}}(W(n)) = \alpha_{\max}$ for $1 \le n \le i + j$. Integrating the defining elasticity with $W(1) = C_{\mathsf{init}}a_{\min}$ yields $W(n) = C_{\mathsf{init}}a_{\min}\, n^{\alpha_{\max}}$ on this interval. For this pure power, the merge correction is $\mathsf{corr}(C_{\mathsf{init}}a_{\min}i^{\alpha_{\max}})\mathsf{corr}(C_{\mathsf{init}}a_{\min}j^{\alpha_{\max}})/\mathsf{corr}(C_{\mathsf{init}}a_{\min}(i+j)^{\alpha_{\max}})$. Its logarithmic derivative with respect to $\log i$ is

$$-\alpha_{\max}\mathsf{decay}(C_{\mathsf{init}}a_{\min}i^{\alpha_{\max}}) + \frac{\alpha_{\max}i}{i+j}\mathsf{decay}(C_{\mathsf{init}}a_{\min}(i+j)^{\alpha_{\max}}) < 0.$$

The same holds for j. Its maximum on $i, j \ge 1$ is therefore $\mathsf{corr}(C_{\mathsf{init}}a_{\min})^2/\mathsf{corr}(2^{\alpha_{\max}}C_{\mathsf{init}}a_{\min})$, attained at $i = j = 1$. The first condition in Eq. (A.23) bounds this correction by $2^{2\alpha_{\max}} - 2^{\alpha_{\max}+1}$. After substituting $W_i/W_{i+j} = (i/(i+j))^{\alpha_{\max}}$, Lem. A.9 proves Eq. (A.25).

On the varying branch, decreasing elasticity gives

$$\left(\frac{i}{i+j}\right)^{\alpha_{\max}} \le \frac{W_i}{W_{i+j}} \le \left(\frac{i}{i+j}\right)^{\alpha_{\mathsf{el}}(W_{i+j})},$$

with the analogous bounds for j. For $C_{\mathsf{cut}} \le i/(i+j) \le 1 - C_{\mathsf{cut}}$, decrease of corr yields

$$\frac{\mathsf{corr}(W_i)\mathsf{corr}(W_j)}{\mathsf{corr}(W_{i+j})} \le \mathsf{mergecorr}(W_{i+j}) = (2^{2\alpha_{\mathsf{el}}(W_{i+j})} - 2^{\alpha_{\mathsf{el}}(W_{i+j})+1}).$$

Indeed, for fixed s, the product $\mathsf{corr}(s\mathsf{frac}^{\alpha_{\max}})\mathsf{corr}(s(1-\mathsf{frac})^{\alpha_{\max}})$ is symmetric about $\mathsf{frac} = 1/2$, and its logarithmic derivative on $(0, 1/2)$ is

$$-\frac{\alpha_{\max}\mathsf{decay}(s\mathsf{frac}^{\alpha_{\max}})}{\mathsf{frac}} + \frac{\alpha_{\max}\mathsf{decay}(s(1-\mathsf{frac})^{\alpha_{\max}})}{1-\mathsf{frac}} < 0.$$

The generalized Silver inequality, applied with $\mathsf{frac} = i/(i+j)$ and $\alpha' = \alpha_{\mathsf{el}}(W_{i+j})$, proves the balanced merge. For $i/(i+j) \le C_{\mathsf{cut}}$, cancel the terminal numerator to obtain

$$\frac{\mathcal{N}_\infty(W_i)\mathcal{N}_\infty(W_j)}{W_{i+j}\mathcal{N}_\infty(W_{i+j})} \le \mathsf{corr}(C_{\mathsf{init}}a_{\min})\frac{W_i}{W_{i+j}}.$$

Here, $\mathcal{N}_\infty(W_j) \le \mathcal{N}_\infty(W_{i+j})$ and $\mathsf{corr}(W_i) \le \mathsf{corr}(W_1) = \mathsf{corr}(C_{\mathsf{init}}a_{\min})$. Thus, the left side of Eq. (A.25) is at most

$$[1 + \mathsf{corr}(C_{\mathsf{init}}a_{\min})]\left(\frac{i}{i+j}\right)^{\alpha} + \left(\frac{j}{i+j}\right)^{\alpha}.$$

As a function of the split fraction $i/(i+j)$, this is convex, equals one at zero, and is at most one at C_{cut} by the second initialization condition. Its chord bounds it by one throughout $[0, C_{\mathsf{cut}}]$. The other endpoint is symmetric.

Exact logarithmic exponent. For the fixed fractions $C_{\mathsf{cut}}^{\alpha_{\max}}$ and $(1-C_{\mathsf{cut}})^{\alpha_{\max}}$, Lem. A.10 gives

$$\mathsf{mergecorr}(s) = 1 + \frac{\pi^2}{4\log s} + O\left(\frac{\log\log s}{(\log s)^2}\right).$$

The denominator subtracts one of the two first-order numerator terms. Hence, $\alpha_{\mathsf{el}}(s) = \alpha + \pi^2/[8\sqrt{2}\rho\log 2\,\log s] + O(\log\log s/(\log s)^2)$. Expanding the reciprocal divides the first-order coefficient by α^2 and changes its sign. The remaining error is integrable with respect to $d\log s$, since $\int^\infty (\log t)t^{-2}\,dt < \infty$ after setting $t = \log s$. Thus, $\log W^{-1}(s) - \nu\log s + \gamma_*\log\log s$ converges to a finite real number; in particular,

$$\log W^{-1}(s) = \nu\log s - \frac{\pi^2}{8\alpha^2\sqrt{2}\rho\log 2}\log\log s + O(1).$$

The logarithmic coefficient is exactly γ_*. Inverting $W^{-1}(s) = \Theta(s^\nu/(\log s)^{\gamma_*})$, using $\log W(n) \sim \alpha\log n$, gives $W(n) = \Theta(n^\alpha(\log n)^{\alpha\gamma_*})$. Finite values remain specified by Eq. (A.24). The remaining scale factor $\frac{4\kappa-1}{4(\kappa-1)}$ lies between one and two, so the constants in the majorant's asymptotic bounds are uniform.

The maximal window. Since W is increasing and unbounded, q is finite, and it is positive for large κ. Maximality and upper elasticity give

$$(4\kappa-1)/256 < W_{4(q+1)} \le ((q+1)/q)^2 W_{4q} \le 4W_{4q}.$$

The integral asymptotics just proved give $W_n = \Theta(n^\alpha[\log(en)]^{\alpha\gamma_*})$. Since $\log((\kappa-1)^\nu/(\log(\kappa-1))^{\gamma_*}) = \nu\log(\kappa-1) + O(\log\log(\kappa-1))$, this scale satisfies $W(4(\kappa-1)^\nu/(\log(\kappa-1))^{\gamma_*}) = \Theta((\kappa-1))$. The elasticity bounds bracket the maximal integer q between fixed positive multiples of $(\kappa-1)^\nu/(\log(\kappa-1))^{\gamma_*}$, proving its stated scale. Since $\nu < 1$, that scale also gives $q\log(\kappa-1)/(\kappa-1) \to 0$, completing Eq. (A.22). Since $\mathcal{Q}(\kappa)$ dominates every fixed power of $\log\kappa$, the stated repair condition on q also holds for sufficiently large κ. Every finite value of W or its inverse is specified by the exact one-dimensional integral, independently of the constants hidden in these asymptotic comparisons. □

From the majorant to a window length. The every-split argument in Prop. A.11 establishes the scale

$$W(n) = \Theta\big(n^{\alpha}[\log(en)]^{\alpha\gamma_*}\big), \qquad \alpha\gamma_* = \frac{\pi^2}{8\sqrt{2}\rho\log\rho}.$$

The next lemma makes W a bound on the step sum of an empty-optimal schedule. The selected window length q satisfies $W(4q) \le (4\kappa-1)/256$. For sufficiently large κ, the largest admissible integer has scale

$$q = \Theta\left(\frac{(\kappa-1)^{\nu}}{(\log(\kappa-1))^{\gamma_*}}\right).$$

The positive logarithmic correction in W therefore shortens the window available at a fixed damping budget. This conclusion uses the every-split majorant, not only the balanced-split heuristic preceding it.

A.5 Step-sum bounds for empty-optimal schedules

The extremal split and the majorant now meet in a single induction. Each child receives the number of positions and inherited buffers belonging to its gap; their target indices add to the parent's index. This proves the step-sum bound needed to choose checkpoints, without yet committing to a window partition.

Lemma A.12 (Step-sum bound for empty-optimal schedules)**.** *Use the reserve $s_{\mathsf{res}} = C_{\mathsf{res}}a_{\min}$ of Lem. A.8 and the corresponding majorant of Prop. A.11. Let integers $m_L, m_R \ge 1$ satisfy $s_{\mathsf{L}} \le W_{m_L} - W_1$ and $s_{\mathsf{R}} \le W_{m_R} - W_1$. Every empty-optimal schedule of length $n \ge 0$ and actual step sum s_{tot} satisfies*

$$s_{\mathsf{res}} + s_{\mathsf{L}} + s_{\mathsf{R}} + s_{\mathsf{tot}} \le W_{n+m_L+m_R-1}, \quad \textit{provided } W_{n+m_L+m_R-1} \le (4\kappa-1)/4. \tag{A.26}$$

Proof. For fixed buffers, Lem. A.6 lets us bound a schedule maximizing step sum over the empty-optimal schedules. We induct on actual length n, simultaneously for every admissible buffer pair; this simultaneous formulation is needed because the children inherit different buffers. With the selected initialization $W_1 = C_{\mathsf{init}}a_{\min} > s_{\mathsf{res}} = C_{\mathsf{res}}a_{\min}$, the induction invariant is $s_{\mathsf{res}} + s_{\mathsf{L}} + s_{\mathsf{R}} + s_{\mathsf{tot}} \le W_{n+m_L+m_R-1}$. This uses exactly the quantities X_i for the child gaps in Lem. A.7. Convexity of W and $W(0) = 0$ imply that integer increments are nondecreasing and at least $W_1 \ge s_{\mathsf{res}}$. Also $W_i + W_j - W_1 \le W_{i+j-1}$ for $i, j \ge 1$, by comparing the $j-1$ increments starting at i with those starting at one. Hence, $s_{\mathsf{res}} + s_{\mathsf{L}} + s_{\mathsf{R}} \le W_{m_L} + W_{m_R} - 2W_1 + s_{\mathsf{res}} \le W_{m_L} + W_{m_R} - W_1 \le W_{m_L+m_R-1}$, which proves the base case. If every actual step is at most $a_{\min}$, adding n integer increments proves the result as well.

Otherwise, apply Lems. A.7 & A.8. If its child lengths are $n_0, \dots, n_k$, their target indices in the induction are

$$n_0 + m_L, \qquad n_i + 1 \ (0 < i < k), \qquad n_k + m_R.$$

They sum to $n + m_L + m_R - 1$, since $\sum_i n_i = n - k$. Each child has smaller actual length, valid inherited buffers, and a closed terminal numerator. Its target index is no larger than the parent's, so its target step-sum bound is also at most $(4\kappa-1)/4$. For an inactive singleton, apply induction to its two children and then use monotonicity of the merge root:

$$X_{\mathsf{tot}} \le J_{\mathcal{N}}(X_0, X_1) \le J_{\mathcal{N}}(W_{n_0+m_L}, W_{n_1+m_R}) \le W_{n+m_L+m_R-1}.$$

The last inequality is (A.20), whose parent-index condition holds by assumption.

For an active split, apply Eq. (A.19) with upper argument $n_* = n + m_L + m_R - 1$ and lower argument equal to each child index displayed above. Induction bounds each X_i by W at that child index. Since the child indices sum to n_*, their ratios to n_* sum to one. Combining these bounds with Eq. (A.14) gives

$$X_{\mathsf{tot}}^{\nu} \le \sum_i X_i^{\nu} \le W_{n+m_L+m_R-1}^{\nu}.$$

Taking the positive νth root proves Eq. (A.26). The argument uses the assumed cap only for the target majorant values, without assuming a bound on the unknown extremal step sum. □

A.6 BUFFERED WINDOWS AND ALL-PREFIX COMPOSITION

We use the step-sum bound to choose one minimizing set per window. The key extra requirement is that every prefix of this same set must be controlled: choosing a new optimal set at each cut would not define one hard objective. The next checkpoint after a cut has a numerator that depends on the pending gap step sum. Keeping that checkpoint fixed makes prefix comparisons compatible with the already chosen suffix. After fixing the minimizing startup parameters, the selected product and pending gap step sum evolve through positive linear maps. This lets one buffered input bound every actual input and allows the same maps to pass an unfinished gap across a window boundary. The bound therefore applies to the costs of the actual global checkpoint set, even when its gaps span several windows.

Lemma A.13 (One buffered window and all its prefixes)**.** *Let $q_0 \geq 1$ be an integer with $W_{4q_0} \leq (4\kappa - 1)/4$. For a window of at most q_0 positions, set $s_{\mathsf{L}} = s_{\mathsf{R}} = W_{q_0}$ and choose a checkpoint set minimizing the buffered checkpoint-set cost Ψ in Eq. (A.10) over all eligible checkpoint sets, including the empty set. If it has k checkpoints and gap sums $s_1, \ldots, s_{k+1}$, then*

$$2s_{\mathsf{R}} + \sum_{i=1}^{k+1} s_i + (k+1)a_{\min} \leq W_{4q_0}. \tag{A.27}$$

Its full selected product is at most 16, its total unselected step sum is at most W_{4q_0}, and every selected prefix of this same checkpoint set has product at most

$$\Pi_* = \max\{16, 1 + 2W_{4q_0}/a_{\min}\}. \tag{A.28}$$

Fix affine parameters attaining the numerators at the gaps of its buffered window. For the state (Π, mass), with Π the selected product, s the pending gap step sum, and $\mathsf{mass} = \Pi s$, the prescribed maps are

$$\begin{aligned} \textit{skip } \eta: &\quad (\Pi, \mathsf{mass}) \mapsto (\Pi, \mathsf{mass} + \eta\Pi), \\ \textit{select } b: &\quad (\Pi, \mathsf{mass}) \mapsto \left(\frac{c(a)(a\Pi + \mathsf{mass})}{b}, 0 \right). \end{aligned} \tag{A.29}$$

For the buffered window input $(1, s_{\mathsf{R}})$, the potential $\Pi + \mathsf{mass}/s_{\mathsf{R}}$ is at most 32 at the output and at most $16\Pi_*$ at every prefix. The same maps multiply this potential for any nonnegative input by at most these respective factors.

Proof. We keep one minimizing checkpoint set throughout the window argument. First, we use the step-sum bound to control its full cost and gaps. Next, fixing its suffix turns each cut into a problem with a fixed next checkpoint, which controls every prefix of that same set. Finally, the positive state maps transfer these buffered window bounds to actual gaps that may cross window boundaries.

A value bound with a fixed next checkpoint. In a length-n problem with initial buffer $s_{\mathsf{L}} \geq 0$ and terminal numerator $\mathcal{N}_{b_{\mathsf{rec}}}(s_{\mathsf{rem}} + s)$, $b_{\mathsf{rec}} \geq a_{\min}$, let $s_{\mathsf{rem}} \geq 0$ and integers $m_L, m_R \geq 1$ satisfy $s_{\mathsf{L}} \leq W_{m_L} - W_1$, $s_{\mathsf{rem}} \leq W_{m_R} - W_1$, and $W_{n+m_L+m_R-1} \leq (4\kappa - 1)/4$. Its optimized value is at most $\mathcal{N}_{b_{\mathsf{rec}}}(W_{n+m_L+m_R-1})$. The analogous bound holds for terminal numerator $\mathcal{N}(s_{\mathsf{rem}} + s)$. We maximize the optimized value over schedules, rather than maximizing their raw step sums. This gives an empty-optimal schedule to which the step-sum bound applies, while retaining the numerator of the next checkpoint in the value comparison. Lem. A.4 gives a global value maximizer which is empty-optimal for this terminal numerator. Lem. A.5 makes the same schedule empty-optimal for $\mathcal{N}(s_{\mathsf{rem}} + s)$. If $s_{\mathsf{tot},*}$ is its actual step sum, Lem. A.12 gives $a_{\min} + s_{\mathsf{L}} + s_{\mathsf{rem}} + s_{\mathsf{tot},*} \leq W_{n+m_L+m_R-1}$. Its value for the original terminal is $\mathcal{N}_{b_{\mathsf{rec}}}(s_{\mathsf{L}} + s_{\mathsf{rem}} + s_{\mathsf{tot},*}) \leq \mathcal{N}_{b_{\mathsf{rec}}}(W_{n+m_L+m_R-1})$, proving the bound for every schedule. For terminal numerator $\mathcal{N}(s_{\mathsf{rem}} + s)$, the same argument omits the closing step.
A fixed buffered window. Let its actual length be $m \leq q_0$. Since $W_{q_0+1} - W_{q_0} \geq W_1$, the buffers $s_{\mathsf{R}} = W_{q_0}$ have indices $m_L = m_R = q_0 + 1$. The target index is $m + 2q_0 + 1 \leq 4q_0$. By Lems. A.4 & A.12, at a maximizing empty-optimal schedule with step sum $s_{\mathsf{tot},*}$, $a_{\min} + 2s_{\mathsf{R}} + s_{\mathsf{tot},*} \leq W_{4q_0}$. Thus, the optimized value of any window is at most $\mathcal{N}(2s_{\mathsf{R}} + s_{\mathsf{tot},*}) \leq 2(a_{\min} + 2s_{\mathsf{R}} + s_{\mathsf{tot},*}) \leq 2W_{4q_0}$, and also at most $\mathcal{N}(W_{4q_0})$. The selected product is its optimized value divided by $\mathcal{N}(s_{\mathsf{R}} + s_{k+1})$, unless the checkpoint set is empty, in which case it is one. Positive concavity gives $\mathcal{N}(s_2)/\mathcal{N}(s_1) \leq s_2/s_1$ for $s_2 \geq s_1 > 0$; hence

$$\Pi \leq \mathcal{N}(W_{4q_0})/\mathcal{N}(s_{\mathsf{R}}) \leq W_{4q_0}/W_{q_0} \leq 16.$$

The last inequality integrates the upper elasticity bound from q_0 to $4q_0$.

Every complete gap of the minimizing checkpoint set is empty-optimal for its next terminal numerator; otherwise, inserting in that gap would improve the full checkpoint set. Close eligible terminal numerators using Lem. A.5. For the corresponding gap lengths $n_0, \dots, n_k$ (so n_i counts the positions in the gap with step sum s_{i+1}), the child indices are n_0+q_0+1, n_i+1 for interior gaps, and n_k+q_0+1. Their sum is $m + 2q_0 + 1 \le 4q_0$. Convexity and $W(0) = 0$ imply $W(n_1) + W(n_2) \le W(n_1 + n_2)$ for $n_1, n_2 \ge 0$. Apply Lem. A.12 to each gap. The sum of the left sides contains both buffers once and one startup $a_{\min}$ per gap; superadditivity bounds the sum of the right sides by W at the sum of their indices. This proves Eq. (A.27). For an empty checkpoint set, this follows directly from Lem. A.12.

Every prefix of that same checkpoint set. Fix a cut and keep all checkpoints strictly after it fixed. If the selected suffix is nonempty, let b_{rec} be the size of its first checkpoint and let s_{rem} be the unselected step sum between the cut and the checkpoint of size b_{rec}. Let s be the pending gap step sum at the cut, including the initial buffer if no checkpoint has yet been selected. All factors after the checkpoint of size b_{rec}, and its denominator b_{rec}, cancel when comparing prefix checkpoint sets. Thus, our fixed minimizing checkpoint set minimizes the prefix problem with initial buffer s_{R} and terminal numerator $\mathcal{N}_{b_{\mathsf{rec}}}(s_{\mathsf{rem}} + s)$. Indeed, a better prefix choice could be joined to the unchanged suffix and improve the full-window objective, contradicting its minimality.

If this checkpoint of size b_{rec} is the window's first checkpoint, its complete preceding gap has at most $q_0 - 1$ positions; the step-sum bound gives $s_{\mathsf{rem}} \le W_{2q_0} - s_{\mathsf{R}} - a_{\min} \le W_{2q_0} - W_1$. Otherwise, its complete preceding gap has no virtual buffer, so $s_{\mathsf{rem}} \le W_{q_0} - a_{\min} \le W_{2q_0} - W_1$, where $W_{2q_0} - W_{q_0} \ge W_1$ is used in the last comparison. These complete-gap estimates also bound its suffix s_{rem}. Consequently, indices $m_L = q_0 + 1, m_R = 2q_0$ are valid for the prefix; its length is at most q_0 and its target index at most $4q_0$. The fixed-next-checkpoint value bound above therefore gives

$$\Pi_{\mathsf{prefix}} \mathcal{N}_{b_{\mathsf{rec}}}(s_{\mathsf{rem}} + s) \le \mathcal{N}_{b_{\mathsf{rec}}}(W_{4q_0}), \qquad \Pi_{\mathsf{prefix}} \le \frac{\mathcal{N}_{b_{\mathsf{rec}}}(W_{4q_0})}{\mathcal{N}_{b_{\mathsf{rec}}}(s_{\mathsf{rem}})} \le 1 + 2W_{4q_0}/a_{\min}.$$

For the last inequality, use increments at most 2 and $\mathcal{N}_{b_{\mathsf{rec}}}(s_{\mathsf{rem}}) \ge a_{\min}$ when $W_{4q_0} \ge s_{\mathsf{rem}}$; otherwise, the ratio is at most one. If no selected suffix remains, the selected prefix product is the full product, already bounded by 16. This proves Eq. (A.28) without choosing a new checkpoint set.

State and potential bounds. The first state coordinate Π is the selected product accumulated so far, and $\mathsf{mass} = \Pi s$ retains the pending gap step sum s. Passing (Π, mass) across a window boundary preserves the unfinished gap. Skipping a step adds its size to the pending gap, whereas selecting a checkpoint uses that gap and resets it to zero. The positive linear maps therefore compose actual global costs even when transfer blocks and windows have different boundaries. At the buffered window output, $\Pi > 0$ and the buffered cost is $\Pi\mathcal{N}(s_{\mathsf{R}} + \mathsf{mass}/\Pi) \ge s_{\mathsf{R}}\Pi + \mathsf{mass}$. Thus, its output potential is at most $2W_{4q_0}/s_{\mathsf{R}} \le 32$. At every prefix of the buffered window, (A.27) implies pending gap step sum $s \le s_{\mathsf{R}} + \sum_i s_i \le W_{4q_0} - s_{\mathsf{R}}$. Therefore $\Pi + \mathsf{mass}/s_{\mathsf{R}} = \Pi(1 + s/s_{\mathsf{R}}) \le 16\Pi \le 16\Pi_*$.

Choose attained affine parameters at the buffered window gaps. Every reached selected factor is then exactly the one just bounded. The maps (A.29) are linear with nonnegative coefficients. Every nonnegative input obeys $(\Pi, \mathsf{mass}) \le (\Pi + \mathsf{mass}/s_{\mathsf{R}})(1, s_{\mathsf{R}})$ componentwise. Applying any prefix of the maps and then the positive linear functional $(\Pi, \mathsf{mass}) \mapsto \Pi + \mathsf{mass}/s_{\mathsf{R}}$ proves the general input bound. The parameters stay fixed during this comparison: the positivity of a common sequence of linear maps allows the buffered window input to dominate every other input. This reasoning also includes $\Pi = 0$ without forming mass/Π. □

A.7 Repairing internal gaps without changing the scale

Window selection bounds the total unselected step sum but may leave an individual internal gap too large for the transfer geometry. We insert checkpoints in such gaps. An insertion changes both its own factors and the next retained checkpoint's numerator; accounting for that checkpoint is essential, especially for a prefix that stops before it. We use the already bounded total unselected step sum to pay for these insertions. Each repaired portion consumes a definite amount of that sum, so the number and total cost of repairs remain controlled by the number of windows. The leading gap is treated separately by its damping product and is left uncut. For a slice of $m \ge 1$ positions, let $j = \lceil m/q \rceil$.

When $j \le \log(\kappa-1)$, set

$$q_{\mathsf{chunk}} = \left\lfloor q/(j+1)^{1/(\alpha-1)} \right\rfloor .$$

Assume $q \ge 2(\log(\kappa-1)+2)^{1/(\alpha-1)}$. This holds for all sufficiently large κ by (A.22). Fig. 5 shows the cost comparison for a retained checkpoint. The inserted checkpoints contribute their own product, and the next retained checkpoint's numerator must also be recomputed using the remaining gap.

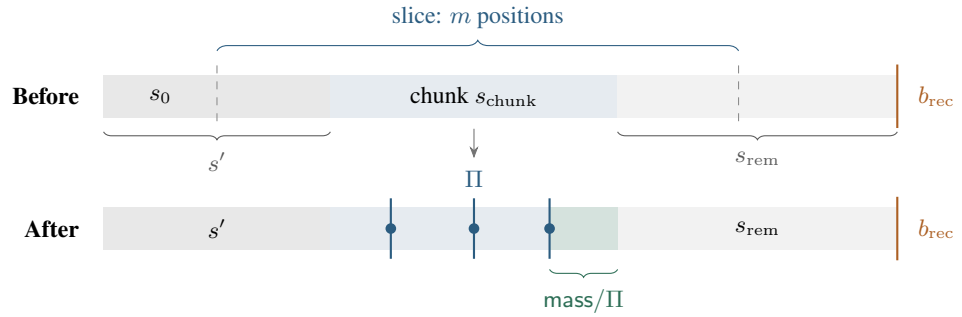


Figure 5: Checkpoint insertion in Lem. A.14, used to repair gaps in Lem. A.15. The diagram shows the short-slice case, $j = \lceil m/q \rceil \le \log(\kappa-1)$: a slice of m positions contains a chunk of at most q_{chunk} positions with step sum $s_{\mathrm{chunk}} \ge W_{4q_{\mathsf{chunk}}}$, which forces a nonempty group of ordinary checkpoints. The original head has step sum s_0; the actual head s' also includes the part of the slice before the chosen chunk. Blue marks are inserted checkpoints, with cost product Π; green shading is the chunk's pending gap, of step sum mass/Π. The suffix has step sum s_{rem} and ends at the next retained checkpoint of size b_{rec}. The cost changes from $\mathcal{N}_{\mathrm{term}}(s' + s_{\mathrm{chunk}} + s_{\mathrm{rem}})/b_{\mathrm{rec}}$ to $\Pi\mathcal{N}_{\mathrm{term}}(\mathsf{mass}/\Pi + s_{\mathrm{rem}})/b_{\mathrm{rec}}$, including the changed numerator of that checkpoint (Eq. (A.31)). For a long slice, $j > \log(\kappa-1)$, the insertion instead uses one exceptional checkpoint of size $b > 64$, provided such a step exists, at insertion ratio at most $(\kappa-1)^3 \le e^{3j}$; the small-step-run bound in Lem. A.15 guarantees its existence.

The elasticity bounds and (A.22) imply

$$\begin{aligned} &\frac{q}{2(j+1)^{1/(\alpha-1)}} \le q_{\mathsf{chunk}} \le \frac{q}{(j+1)^{1/(\alpha-1)}}, \\ &W_{4q_{\mathsf{chunk}}} \le W_{4q}(q_{\mathsf{chunk}}/q)^{\alpha}, \quad W_{q_{\mathsf{chunk}}} \ge W_q(q_{\mathsf{chunk}}/q)^2 > \frac{4\kappa-1}{2^{16}(j+1)^{2/(\alpha-1)}}. \end{aligned} \tag{A.30}$$

Indeed, $W_q \ge W_{4q}/16 > (4\kappa-1)/2^{14}$. These estimates use $5/4 < \alpha < 3/2$, so all exponents and constants are absolute.

Lemma A.14 (One insertion group)**.** *Suppose a slice of $m \ge 1$ positions has step sum s_{slice} and is preceded by an unselected head of step sum s_0, where*

$$(4\kappa-1)/256 \le s_{\mathsf{slice}} < (4\kappa-1)/128, \qquad (4\kappa-1)/256 \le s_0 < (4\kappa-1)/128.$$

The numerator of its next retained checkpoint is $\mathcal{N}_{\mathsf{term}} = \mathcal{N}_{b_{\mathsf{rec}}}$ for an eligible next checkpoint, or $\mathcal{N}_{\mathsf{term}} = \mathcal{N}$ for a provisional terminal checkpoint. If $j = \lceil m/q \rceil \le \log(\kappa-1)$, a nonempty group of ordinary eligible checkpoints in the slice has insertion ratio at most $65600(j+1)^{2/(\alpha-1)}$, and every selected prefix of this same group, starting from its actual head, has product at most $41.25(4\kappa-1)/(4a_{\min})$. The insertion ratio includes changes to the numerator of the next retained checkpoint and leaves the denominator fixed. If $j > \log(\kappa-1)$ and the slice contains a step above 64, one exceptional checkpoint has insertion ratio at most $(\kappa-1)^3 \le e^{3j}$.

Proof. In the short case, we will find a chunk whose step sum is too large for the empty checkpoint set to minimize. The window lemma then supplies a nonempty group, and its state bound controls both the actual head and the changed numerator of the next retained checkpoint. Partition the slice into consecutive chunks of at most q_{chunk} positions. There are at most $m/q_{\mathsf{chunk}} + 1 \le (j+1)q/q_{\mathsf{chunk}}$ chunks. Some chunk has step sum

$$s_{\mathsf{chunk}} \ge \frac{s_{\mathsf{slice}} q_{\mathsf{chunk}}}{(j+1)q} \ge W_{4q_{\mathsf{chunk}}}.$$

For the last inequality, $W_{4q} \le (4\kappa-1)/256 \le s_{\mathsf{slice}}$, and $W_{4q_{\mathsf{chunk}}} \le s_{\mathsf{slice}}(q_{\mathsf{chunk}}/q)^{\alpha} \le s_{\mathsf{slice}} q_{\mathsf{chunk}}/[(j+1)q]$, because $(q_{\mathsf{chunk}}/q)^{\alpha-1} \le 1/(j+1)$. Apply Lem. A.13 with buffers $s_{\mathsf{R}} = W_{q_{\mathsf{chunk}}}$ to that chunk. Its minimizing checkpoint set is nonempty; otherwise, the step-sum bound in Lem. A.13 would imply $a_{\min} + 2s_{\mathsf{R}} + s_{\mathsf{chunk}} \le W_{4q_{\mathsf{chunk}}}$. Keep its attained affine parameters fixed.

Let s' be the actual step sum from the previous checkpoint to the start of the chosen chunk. Then $s' \le s_0 + s_{\mathsf{slice}} < (4\kappa-1)/64$. The positive-map comparison, now with actual input $(1, s')$, gives at chunk output (Π, mass)

$$\Pi + \mathsf{mass}/s_{\mathsf{R}} \le 32(1 + s'/s_{\mathsf{R}}),$$
$$\Pi_{\mathsf{prefix}} \le 16\Pi_*(1 + s'/s_{\mathsf{R}}), \quad \Pi_* = \max\{16, 1 + 2W_{4q_{\mathsf{chunk}}}/a_{\min}\}.$$

For any remaining step sum s_{rem} between the chunk and its next retained checkpoint, the exact ratio is

$$\mathsf{ins} = \frac{\Pi \mathcal{N}_{\mathsf{term}}(s_{\mathsf{rem}} + \mathsf{mass}/\Pi)}{\mathcal{N}_{\mathsf{term}}(s_{\mathsf{rem}} + s' + s_{\mathsf{chunk}})}. \tag{A.31}$$

All reached state first coordinates are positive. Before insertion, the next retained checkpoint uses the head s', the entire chunk step sum s_{chunk}, and the unchanged suffix s_{rem}. After insertion, it sees only the chunk's pending gap step sum mass/Π and s_{rem}, while the inserted checkpoints contribute Π. This proves Eq. (A.31), including the next retained checkpoint whose denominator cancels. Using Eq. (A.7) and $s_{\mathsf{chunk}} \ge s_{\mathsf{R}}$,

$$\mathsf{ins} \le \frac{2[(a_{\min} + s_{\mathsf{rem}})\Pi + \mathsf{mass}]}{a_{\min} + s_{\mathsf{rem}} + s' + s_{\mathsf{chunk}}} \le 2(\Pi + \mathsf{mass}/s_{\mathsf{R}}) \le 64(1 + s'/s_{\mathsf{R}}) \le 65600(j+1)^{2/(\alpha-1)},$$

where (A.30) supplies the last inequality. For the middle comparison, the denominator is at least both $a_{\min} + s_{\mathsf{rem}}$ and $s_{\mathsf{chunk}} \ge s_{\mathsf{R}}$; hence its two numerator terms are bounded by Π and $\mathsf{mass}/s_{\mathsf{R}}$, respectively. Also $W_{4q_{\mathsf{chunk}}} \le 16 W_{q_{\mathsf{chunk}}}$, $a_{\min} \le s_{\mathsf{R}} \le (4\kappa-1)/256$, and $s' < (4\kappa-1)/64$, so $\Pi_* \le 33 s_{\mathsf{R}}/a_{\min}$ and $\Pi_{\mathsf{prefix}} \le 528(s_{\mathsf{R}} + s')/a_{\min} \le 41.25(4\kappa-1)/(4a_{\min})$. Both estimates concern that same fixed group.

In the long case, the numerical exceptional cost in Eq. (A.8) satisfies $(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b \le (\kappa-1)^3$ at any $b > 64$, since $\frac{4\kappa-1}{4(\kappa-1)} < 2$ for sufficiently large κ. Its actual preceding gap lies inside the head and slice, so its step sum is less than $(4\kappa-1)/64$. Inserting it shortens the gap before the next retained checkpoint, so its increasing numerator can only decrease. Its logarithmic insertion cost is at most $3\log(\kappa-1) < 3j$. □

Lemma A.15 (One final checkpoint set and all its prefixes)**.** *There are absolute constants $C_{\mathsf{sel}}, C_{\mathsf{sel,amp}}, C_{\mathsf{sel,win}} > 0$ such that the following holds for every sufficiently large κ. Consider a nonempty nonnegative schedule of length $n \ge 1$ ending in a step above 64. Assume every strictly internal maximal run of steps at most 64 has step sum at most $(4\kappa-1)/512$; the leading run is unrestricted. There are an initial window checkpoint set and one final checkpoint set extending it, with every final checkpoint above 64 and including the last step. After optionally forcing the terminal checkpoint, the first gap is uncut; every later final gap has step sum at most $(4\kappa-1)/64$. Its ordinary factors are $c(a)(a+s)/b$ with one fixed choice of eligible parameters $a_{\min} \le a \le \min\{b, (4\kappa-1)/8\}$, and its exceptional factors are $(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b$. The full product and every selected prefix of this same set are at most*

$$C_{\mathsf{sel}}(\kappa-1)^{C_{\mathsf{sel,amp}}} e^{C_{\mathsf{sel,win}}\lceil n/q \rceil}.$$

The step sum s_{gap} unselected by the initial window checkpoint set satisfies $s_{\mathsf{gap}} \le \lceil n/q \rceil W_{4q}$, and each such step is at most $W_{4q} \le (4\kappa-1)/256$. The final checkpoint set adds checkpoints only to the initial checkpoint set.

If the terminal checkpoint is already in the initial window set, it stays ordinary. Its first selected checkpoint is ordinary unless the forced terminal checkpoint is the sole checkpoint. The number of inserted exceptional checkpoints is at most $\lceil n/q \rceil$, apart from a possible forced terminal checkpoint.

Proof. The initial window selection already controls every selected prefix product, but only bounds the total unselected step sum. We repair each nonleading gap by inserting groups that consume a definite amount of that step sum. To preserve the prefix bounds, we charge a group together with the change in its next original checkpoint, then account for the one original checkpoint that may still be pending at a cut.

Global window composition. Partitioning a length-$n \geq 1$ schedule into $\lceil n/q \rceil$ windows gives one checkpoint set with eligible fixed parameters and

$$\begin{aligned} \Pi \leq 32^{\lceil n/q \rceil}, \qquad \Pi_{\mathsf{prefix}} \leq 16\Pi_* 32^{\lceil n/q \rceil}, \\ s_{\mathsf{gap}} \leq \lceil n/q \rceil W_{4q}, \qquad \eta_j \leq W_{4q} \text{ at every unselected position,} \end{aligned} \tag{A.32}$$

Here, s_{gap} is the step sum left unselected by this global window checkpoint set, Π_* uses $q_0 = q$, and every factor uses its actual global gap. For the global list, use buffers $s_{\mathsf{R}} = W_q$ in every buffered window. Start the actual state at $(1, 0)$, whose potential is one, and compose the fixed window maps. Each full window multiplies its potential by at most 32, and a prefix by at most $16\Pi_*$. The first state coordinate is no greater than the potential, so the product bounds in Eq. (A.32) follow. Summing the combinatorial unselected step sums gives its step-sum bounds. Actual gaps can cross window boundaries: the state retains the pending gap step sum. Eligibility depends only on the selected size, not on this changing step sum. Finally, replacing original affine numerators by the smaller $\mathcal{N}_b$ at their actual global gaps only decreases every product and prefix. These checkpoint numerators can be retained during the following repair, then assigned attaining parameters at their final gaps. **Initial checkpoint set and terminal checkpoint.** Use the global window composition just proved. It gives the step-sum bounds in the statement, the full product $e^{O(\lceil n/q \rceil)}$, and every prefix $(\kappa - 1)^{O(1)} e^{O(\lceil n/q \rceil)}$. Replace the initial fixed affine numerators by eligible checkpoint numerators at their actual gaps, which only decreases these products. Call these checkpoints original. The step sum s_{gap} still means the step sum unselected at this initial stage, even after some of its positions are subsequently selected.

If necessary, force the last step to be selected, provisionally with numerator $\mathcal{N}$. Its factor is at most $2(a_{\min} + s_{\mathsf{gap}})/64$, hence costs a polynomial in κ and $\lceil n/q \rceil$. At the end, use its actual exceptional geometry; the ratio $[(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b]/[\mathcal{N}(s)/b]$ is at most $(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/a_{\min}$, another polynomial. If the terminal checkpoint was already selected, keep it ordinary. Treat a provisional terminal checkpoint as an original checkpoint for scanning. All positions in the gaps to be modified were originally unselected and therefore have size at most $W_{4q} \leq (4\kappa - 1)/256$.

Scanning and admissibility. Leave the first gap uncut. In each other gap, while its remaining step sum exceeds $(4\kappa - 1)/64$, scan a head until its step sum first reaches $(4\kappa - 1)/256$, then a slice until its step sum first reaches $(4\kappa - 1)/256$. Individual sizes are at most $(4\kappa - 1)/256$, so both completed step sums lie in $[(4\kappa - 1)/256, (4\kappa - 1)/128)$: the step sum before the last scanned step is strictly less than $(4\kappa - 1)/256$. Use Lem. A.14 and resume immediately after the last inserted checkpoint. The preceding gap of the first inserted checkpoint lies within the head and slice, and all other newly completed gaps lie within the slice. Their combined available step sum is strictly less than $(4\kappa - 1)/64$, so each newly completed gap meets the cap. Both scans can be completed because their combined step sum is smaller than the remaining step sum that triggered the scan. The final remainder also meets the cap. Each insertion is nonempty and advances beyond the head; the finite procedure terminates, even when some steps are zero.

Whenever the long-slice case needs an exceptional checkpoint, that slice contains a step above 64. Otherwise, it lies in a strictly internal maximal run of steps at most 64: its surrounding original or inserted checkpoints exceed 64, and the first gap was not modified. Its step sum is at least $(4\kappa - 1)/256$, which contradicts the phase bound $(4\kappa - 1)/512$.

Group count and slice multiplicity. Every group selects at least one position in its slice, after the completed head of step sum at least $(4\kappa-1)/256$. Thus, each group advances by at least $(4\kappa-1)/256$ of the originally unselected step sum, and these advances are disjoint even if successive slices overlap. The number of groups groups therefore satisfies

$$\mathsf{groups} \leq 256 s_{\mathsf{gap}}/(4\kappa - 1) \leq 256 \lceil n/q \rceil W_{4q}/(4\kappa - 1) \leq \lceil n/q \rceil.$$

A slice of length m_i has $j_i = \lceil m_i/q \rceil$, and either insertion case costs at most $O(1 + j_i)$ in logarithm. The slice lengths, including zero-size positions, sum to at most $3n$. To see this, use cumulative step sum along an original gap. If s_2 is the cumulative step sum at the end of a position covered by a slice, its scan start s_1 satisfies $s_2 - (4\kappa - 1)/64 < s_1 \leq s_2 - (4\kappa - 1)/256$. Successive starts differ by at least $(4\kappa - 1)/256$, so at most three starts can cover the same position, including a zero-size position. The argument counts positions, not positive contributions to the step sum: a zero-size position has the

same cumulative coordinate as its neighbor but still belongs to at most the same three slices. Original gaps are disjoint, so the same multiplicity bound holds globally. Consequently,

$$\sum_i j_i \le \mathsf{groups} + q^{-1}\sum_i m_i \le \mathsf{groups} + 3n/q \le 4\lceil n/q\rceil.$$

The total logarithmic insertion cost is $O(\lceil n/q\rceil)$.

Exact cost accounting at the next checkpoint, including a cut. Fix an original checkpoint with numerator $\mathcal{N}_{\mathsf{term}}$. Let s_0 be its original gap step sum and s_j its remaining gap after j inserted groups. If Π_j is the selected product of group j with its actual head, its exact ratio is

$$\mathsf{ins}_j = \Pi_j \frac{\mathcal{N}_{\mathsf{term}}(s_j)}{\mathcal{N}_{\mathsf{term}}(s_{j-1})}, \qquad \prod_{j=1}^{j_*} \Pi_j = \left(\prod_{j=1}^{j_*} \mathsf{ins}_j\right) \frac{\mathcal{N}_{\mathsf{term}}(s_0)}{\mathcal{N}_{\mathsf{term}}(s_{j_*})}.$$

This is Eq. (A.31) with the common outside step sum retained. It holds for exceptional groups too; their ratios were bounded by their selected factors. Once the next original checkpoint is included, its changed numerator $\mathcal{N}_{\mathsf{term}}(s_{j_*})$ cancels the final denominator. Completed groups therefore cost $e^{O(\lceil n/q\rceil)}$ relative to the original product.

At a cut before this original checkpoint, its cancellation has not happened. Let j_* now count the completed groups before the cut in this original gap. Thus, s_{j_*} is the remaining gap after these groups, before any partially completed group is inserted. All earlier original checkpoints have been reached and their numerator ratios have canceled. The completed groups in the current gap leave exactly the ratio in the telescoping identity above, bounded by

$$\frac{\mathcal{N}_{\mathsf{term}}(s_0)}{\mathcal{N}_{\mathsf{term}}(s_{j_*})} \le \frac{2(a_{\min} + s_{\mathsf{gap}})}{a_{\min}} \le O((\kappa-1)(\lceil n/q\rceil + 1)).$$

Here $s_0 \le s_{\mathsf{gap}}$, while every checkpoint numerator is at least $a_{\min}$; the bound also covers $j_* = 0$, when the ratio is one. There is also at most one partially completed group. Its selected prefix multiplies the product of the completed groups by at most $O((4\kappa-1)/a_{\min})$ in the ordinary case and $(\kappa-1)^3$ in the exceptional case. For this partial group, use its own all-prefix bound instead of its full insertion ratio, so cancellation at its next original checkpoint is not assumed. The original all-prefix bound, these factors, and the polynomial cost of the forced terminal checkpoint give $(\kappa-1)^{O(1)}e^{O(\lceil n/q\rceil)}$ at every final selected prefix, since $\lceil n/q\rceil + 1 \le 2^{\lceil n/q\rceil}$. The full product has the same bound.

Assign attaining parameters to original checkpoints at their final gaps. Inserted groups retain the buffered window parameters used by their positive-map bounds, which remain eligible for every actual head. This gives one final checkpoint set and one parameter choice. The first gap has remained uncut, so its first checkpoint is original and ordinary unless only the forced terminal checkpoint exists. Every exceptional insertion is one of the groups groups, proving the final count. □

B Building and Realizing Phase Cores

A rotating-gradient transfer is the geometric idea of Ye & Liu (2026b). We use a logarithmic spiral, also suggested in their Section 5, and choose a pitch-dependent bound on the derivative threshold relative to the output amplitude. The purpose is to realize the checkpoint data from Appx. A as one phase core with a lower bound at every phase iterate. The projection geometry first supplies local transfer functions whose derivatives vanish in the regions needed to place them on overlapping coordinate pairs. Ordinary transfers realize the optimized numerical costs, while exceptional transfers handle the additional checkpoints used in gap repair. One controlled second transfer also allows the core to accommodate its later incoming coupling.

We then separate two uses of these local constructions. Formal transfer factors bound the scales of every trial and give a common quadratic ball and common initial gradient before a coupling is chosen. Only after the trial is matched to the actual incoming correction do we identify its local trajectories with the GD trajectory of the core. This order lets Appx. C fix the coupling from the common static data and then select a consistent trial.

A gap consists of $n_{\mathsf{gap}} \geq 0$ unselected steps immediately before a checkpoint of size b. A transfer starts at $(1,0)$ and places amplitude $B > 0$ in the second coordinate after the checkpoint, at cost B^{-1}. In a coordinate chain, the transfer function has zero gradient while its predecessor is active and zero derivative in the shared coordinate while its successor is active. The *derivative threshold* is the second-coordinate value at or above which this derivative vanishes. All transfers use the prescribed nonnegative steps η_t and a normalized pair $\boldsymbol{y} = (y^{(1)}, y^{(2)}) \in \mathbb{R}^2$, starting at $(1,0)$, with local objective

$$\frac{2\|\boldsymbol{y}\|^2}{4\kappa - 1} + \frac{2\kappa - 4}{4\kappa - 1} g(\boldsymbol{y}). \tag{B.1}$$

This objective generates the active-pair dynamics in (3.1) after dividing the pair by its incoming amplitude: transfer i uses $\boldsymbol{y} = (z^{(i)}, z^{(i+1)})/D_i$. Here $\kappa > 2$; the final parameter bounds require only that κ exceed an absolute constant. Overlapping pairs have smoothness $4(\kappa-1)/(4\kappa-1)$, leaving the coupling budget of Appx. C.1. We index each gap locally: $\boldsymbol{y}_0$ is its input, and step η_t produces $\boldsymbol{y}_t$ from $\boldsymbol{y}_{t-1}$; $y_t^{(i)}$ is its coordinate i. For the calculations in the proofs, we abbreviate the fixed coefficients by

$$\mu_{\mathsf{c}} := \frac{4}{4\kappa - 1}, \qquad \beta_{\mathsf{c}} := \frac{2(\kappa - 2)}{4\kappa - 1}.$$

Definition B.1 (Transfer data and compatibility)**.** A transfer has a convex 1-smooth nonnegative function g with $g(0) = 0$, a positive transfer factor B, an input threshold $\tau_{\mathsf{in}} \geq 0$, a derivative threshold $\delta \geq 0$, and a zero radius $\mathsf{rad}_{\mathsf{loc}} > 0$. Its static properties are

$$g = \nabla g = 0 \quad \text{if } y^{(1)} \leq \tau_{\mathsf{in}},\ y^{(2)} \geq 0 \quad \text{or if } \|\boldsymbol{y}\| \leq \mathsf{rad}_{\mathsf{loc}}, \qquad \partial_2 g(\boldsymbol{y}) = 0 \quad \text{if } y^{(2)} \geq \delta,$$

where the last assertion holds for every first coordinate. A matched local trajectory starts at $(1,0)$, follows the gap steps in (B.1) with any specified horizontal control, and has second coordinate B after its checkpoint. At every state before the checkpoint update, it has $y^{(1)} > \tau_{\mathsf{in}}$ and $0 \leq y^{(2)} < \epsilon_{\mathsf{gap}} B$, and its derivative threshold satisfies $\delta < \tau_{\mathsf{out}} B$. Two successive transfers are compatible when $\epsilon_{\mathsf{gap},i} \leq \tau_{\mathsf{in},i+1}$ and $\tau_{\mathsf{out},i} \leq \tau_{\mathsf{in},i+1}$. The leading transfer may instead use its stated positive first-coordinate bound. For a controlled family, B and δ are formal static data for every trial; the trajectory and the normalized threshold comparisons are asserted only at matching.

B.1 Projection geometry and backward shooting

Projection onto a convex set supplies the gradient; normal cones specify the points realizing each gradient value. For a nonempty compact convex set $\mathcal{C} \subset \mathbb{R}^2$ containing zero, define

$$\mathcal{E}_{\mathcal{C}}(\widehat{\boldsymbol{y}}) = \max_{\boldsymbol{v} \in \mathcal{C}} \left\{ \langle \boldsymbol{v}, \widehat{\boldsymbol{y}} \rangle - \frac{1}{2}\|\boldsymbol{v}\|^2 \right\} = \frac{1}{2}\|\widehat{\boldsymbol{y}}\|^2 - \frac{1}{2}\operatorname{dist}(\widehat{\boldsymbol{y}}, \mathcal{C})^2. \tag{B.2}$$

The projection envelope is nonnegative, convex, and 1-smooth, with $\nabla \mathcal{E}_{\mathcal{C}} = \operatorname{proj}_{\mathcal{C}}$ and $\mathcal{E}_{\mathcal{C}}(0) = 0$. We use the first-order projection characterization (Bubeck, 2015, Proposition 1.3 and Lemma 3.1)

$$\boldsymbol{v}_{\mathsf{proj}} = \operatorname{proj}_{\mathcal{C}}(\widehat{\boldsymbol{y}}) \quad \Longleftrightarrow \quad \boldsymbol{v}_{\mathsf{proj}} \in \mathcal{C}, \quad \langle \widehat{\boldsymbol{y}} - \boldsymbol{v}_{\mathsf{proj}}, \boldsymbol{v} - \boldsymbol{v}_{\mathsf{proj}} \rangle \leq 0 \quad (\boldsymbol{v} \in \mathcal{C}).$$

Applying this inequality at two projected points gives $\|\mathrm{proj}_{\mathcal{C}}(\widehat{\boldsymbol{y}}) - \mathrm{proj}_{\mathcal{C}}(\widehat{\boldsymbol{y}}')\|^2 \le \langle \mathrm{proj}_{\mathcal{C}}(\widehat{\boldsymbol{y}}) - \mathrm{proj}_{\mathcal{C}}(\widehat{\boldsymbol{y}}'), \widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}}'\rangle$, so projection is nonexpansive. Inserting the maximizing point at each of two nearby arguments into (B.2) gives the gradient identity above; nonexpansiveness then gives 1-smoothness. Convexity follows because the envelope is a supremum of affine functions, and the candidate zero gives nonnegativity.

The gradient arc. A positive first gradient component reduces the first coordinate; a negative second component builds the next coordinate. For pitch $0 < \zeta \le 1/4$, use $\Phi_\zeta, \mathsf{rad}_{\zeta,0}$ from Appx. A.1 and choose a logarithmic spiral ending at $(2\zeta, -1)$:

$$\begin{gathered} \boldsymbol{e}_\varphi = (\cos\varphi, -\sin\varphi), \\ \mathsf{rad}_{\zeta,\varphi} = \sqrt{1+4\zeta^2}\, e^{\zeta(\Phi_\zeta - \varphi)}, \qquad \boldsymbol{v}_{\zeta,\varphi} = \mathsf{rad}_{\zeta,\varphi}\boldsymbol{e}_\varphi \quad (0 \le \varphi \le \Phi_\zeta), \\ \mathcal{C}_\zeta = \mathrm{conv}\big(\{0\} \cup \{\boldsymbol{v}_{\zeta,\varphi} : 0 \le \varphi \le \Phi_\zeta\}\big). \end{gathered} \tag{B.3}$$

Along this arc, $v^{(1)}_{\zeta,\varphi}$ decreases and $-v^{(2)}_{\zeta,\varphi}$ increases. The largest radius is $\mathsf{rad}_{\zeta,0}$. Including zero creates the horizontal edge that makes the second coordinate of the projection vanish for points with nonnegative second coordinate. The tangent and outward-normal directions are

$$\mathbf{tangent}_{\zeta,\varphi} = \boldsymbol{e}'_\varphi - \zeta \boldsymbol{e}_\varphi, \qquad \mathbf{normal}_{\zeta,\varphi} = \boldsymbol{e}_\varphi + \zeta \boldsymbol{e}'_\varphi.$$

They are perpendicular, and $\mathsf{normal}^{(1)}_{\zeta,\varphi} = \cos\varphi - \zeta\sin\varphi > 0$. For $\Delta = \psi - \varphi$, the support calculation

$$\begin{gathered} \langle \mathbf{normal}_{\zeta,\varphi}, \boldsymbol{v}_{\zeta,\psi}\rangle = \mathsf{rad}_{\zeta,\varphi} e^{-\zeta\Delta}(\cos\Delta + \zeta\sin\Delta), \\ \frac{d}{d\Delta}\langle \mathbf{normal}_{\zeta,\varphi}, \boldsymbol{v}_{\zeta,\psi}\rangle = -(1+\zeta^2)\mathsf{rad}_{\zeta,\varphi} e^{-\zeta\Delta}\sin\Delta \end{gathered}$$

shows that the functional is maximal at $\psi = \varphi$, since $|\Delta| < \pi/2$; zero is on its inner side as well. The *normal cone* $N_{\mathcal{C}}(\boldsymbol{v}) = \{\mathbf{normal}_{\mathsf{aux}} : \langle \mathbf{normal}_{\mathsf{aux}}, \boldsymbol{v}' - \boldsymbol{v}\rangle \le 0 \ (\boldsymbol{v}' \in \mathcal{C})\}$ therefore satisfies

$$\begin{gathered} \mathrm{proj}_{\mathcal{C}_\zeta}(\boldsymbol{v}_{\zeta,\varphi} + \xi \mathbf{normal}_{\zeta,\varphi}) = \boldsymbol{v}_{\zeta,\varphi} \quad (\xi \ge 0), \\ N_{\mathcal{C}_\zeta}(\boldsymbol{v}_{\zeta,0}) = \{\mathbf{normal}_{\mathsf{aux}} : \mathsf{normal}^{(1)}_{\mathsf{aux}} \ge 0,\ \mathsf{normal}^{(2)}_{\mathsf{aux}} \ge -\zeta \mathsf{normal}^{(1)}_{\mathsf{aux}}\}. \end{gathered} \tag{B.4}$$

For the second identity, test the origin and the arc tangent at $\varphi = 0$ to obtain the two necessary inequalities. Conversely, the support inequality at angle zero gives $v^{(1)} - \mathsf{rad}_{\zeta,0} \le \zeta v^{(2)}$, so $\langle \mathbf{normal}_{\mathsf{aux}}, \boldsymbol{v} - \boldsymbol{v}_{\zeta,0}\rangle \le (\zeta \mathsf{normal}^{(1)}_{\mathsf{aux}} + \mathsf{normal}^{(2)}_{\mathsf{aux}}) v^{(2)} \le 0$ under those inequalities.

Translation and scale. We use

$$g(\boldsymbol{y}) = \upsilon^2 \mathcal{E}_{\mathcal{C}_\zeta}(\boldsymbol{y}/\upsilon - \boldsymbol{w}), \qquad \nabla g(\boldsymbol{y}) = \upsilon\, \mathrm{proj}_{\mathcal{C}_\zeta}(\boldsymbol{y}/\upsilon - \boldsymbol{w}), \qquad \boldsymbol{w} = (w^{(1)}, w^{(2)}), \quad \upsilon > 0.$$

The terminal point in relative coordinates $(2\zeta, -1)$ corresponds to $\upsilon(w^{(1)} + 2\zeta, w^{(2)} - 1)$. We construct its preceding iterates backward, then choose $w^{(2)}$ and υ to make the input $(1, 0)$. For $w^{(1)} > 0$, $w^{(2)} \ge 0$, and $w^{(1)} \ge (1 + 2\zeta^2) w^{(2)}/\zeta$, direct substitution gives

$$\langle \boldsymbol{w}, \mathbf{normal}_{\zeta,\varphi}\rangle \ge 0, \qquad \langle \boldsymbol{w}, \mathbf{tangent}_{\zeta,\varphi}\rangle < 0, \qquad \langle \boldsymbol{w}, \boldsymbol{e}_\varphi\rangle \ge 0, \qquad \langle \boldsymbol{w}, \boldsymbol{e}'_\varphi\rangle \le 0. \tag{B.5}$$

Global zero regions. For every translation satisfying the preceding inequalities and every positive scale, the following bounds hold. Globally,

$$\partial_1 g \ge 0, \quad \partial_2 g \le 0, \quad \|\nabla g\| \le \upsilon \mathsf{rad}_{\zeta,0}, \qquad \partial_2 g(\boldsymbol{y}) = 0 \quad \text{if } y^{(2)} \ge \upsilon w^{(2)}, \tag{B.6}$$

and $g = \nabla g = 0$ whenever $y^{(1)} \le \upsilon(w^{(1)} - w^{(2)}/(2\zeta))$ and $y^{(2)} \ge 0$. Both vanish on the ball centered at zero of radius

$$\mathsf{rad}_{\mathsf{loc}} = \upsilon \frac{2\zeta w^{(1)} - w^{(2)}}{\sqrt{1+4\zeta^2}} \ge \frac{\zeta}{2} \upsilon w^{(1)} > 0. \tag{B.7}$$

The hull lies in $[0, \mathsf{rad}_{\zeta,0}] \times [-1, 0]$ and contains the segment $[0, \mathsf{rad}_{\zeta,0}] \times \{0\}$; every point in the hull has norm at most $\mathsf{rad}_{\zeta,0}$. For a point $(\widehat{y}^{(1)}, \widehat{y}^{(2)})$ with $\widehat{y}^{(2)} \ge 0$, replacing any hull point $(v^{(1)}, v^{(2)})$ by $(v^{(1)}, 0)$ stays in this segment and cannot increase its distance to that point. Its projection therefore

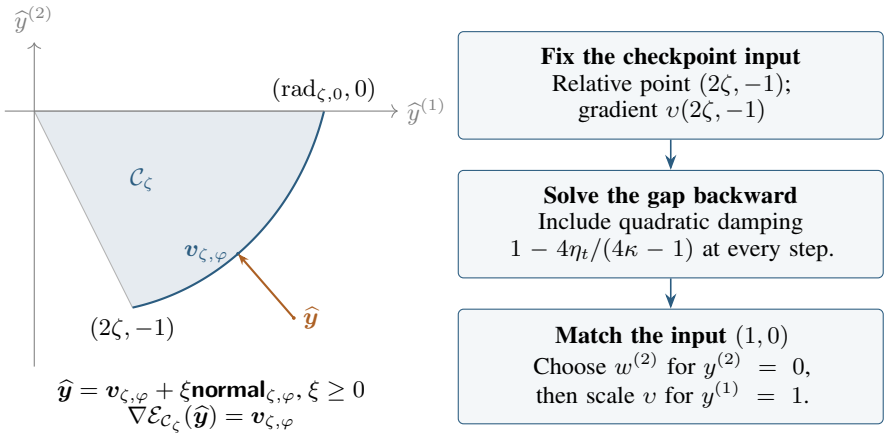


Figure 6: Left: the normal-ray projection in Eq. (B.4), which specifies the transfer gradient through $\nabla g(\boldsymbol{y}) = \upsilon \operatorname{proj}_{\mathcal{C}_\zeta}(\boldsymbol{y}/\upsilon - \boldsymbol{w})$, shown at $\zeta = 1/4$. Right: the backward recursion and choice of $w^{(2)}, \upsilon$ that match input $(1, 0)$ in Lem. B.2, Steps 1–3.

has second coordinate zero, regardless of $\widehat{y}^{(1)}$. This proves Eq. (B.6). Every $\boldsymbol{v} \in \mathcal{C}_\zeta$ satisfies $0 \leq -v^{(2)} \leq v^{(1)}/(2\zeta)$. If $y^{(1)}/\upsilon - w^{(1)} \leq -w^{(2)}/(2\zeta)$ and $y^{(2)} \geq 0$, then

$$\langle \boldsymbol{v}, \boldsymbol{y}/\upsilon - (w^{(1)}, w^{(2)}) \rangle \leq -\frac{w^{(2)}}{2\zeta} v^{(1)} - w^{(2)} v^{(2)} \leq 0.$$

The projection, envelope, and gradient are zero there. More generally, the zero set is $\upsilon((w^{(1)}, w^{(2)}) + \mathcal{C}_\zeta^\circ)$, where $\mathcal{C}_\zeta^\circ = \{\widehat{\boldsymbol{y}} : \langle \widehat{\boldsymbol{y}}, \boldsymbol{v} \rangle \leq 0 \text{ for every } \boldsymbol{v} \in \mathcal{C}_\zeta\}$ is the *polar cone*. Its largest origin-centered ball has radius $\upsilon \min_{0 \leq \varphi \leq \Phi_\zeta} (w^{(1)} \cos\varphi - w^{(2)} \sin\varphi)$. The minimum occurs at Φ_ζ, giving Eq. (B.7). Since $w^{(1)} \geq (1 + 2\zeta^2) w^{(2)}/\zeta$ and $\sqrt{1 + 4\zeta^2} < 2$, it is at least $(\zeta/2)\upsilon w^{(1)}$.

For the matching interval, define

$$w_{\min}(\zeta) = 2(1 + 2\zeta^2)\mathsf{rad}_{\mathsf{turn}}(\zeta).$$

Since $\Phi_\zeta \geq \arctan 2 > 1$ and $e^{\Phi_\zeta/\zeta} \geq \Phi_\zeta/\zeta$, we have $\zeta \mathsf{rad}_{\mathsf{turn}}(\zeta) \geq 1$. Thus $[1, 2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)]$ is nonempty, and $w^{(1)} \geq w_{\min}(\zeta)$ gives Eq. (B.5) throughout this height interval. The next proof shows that $\mathsf{rad}_{\mathsf{turn}}(\zeta)$ and $\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)$ bound the damped horizontal and vertical sizes during turning, making the height-matching map preserve this interval.

B.1.1 Solving a gap backward

Lemma B.2 (Backward shooting for a gap). *Fix $\kappa > 2$ and $0 < \zeta \leq 1/4$, with $\mathcal{C}_\zeta$ as in (B.3). Let the prescribed gap steps $\eta_1, \ldots, \eta_{n_{\mathsf{gap}}} \geq 0$ satisfy $\eta_t < (4\kappa - 1)/4$. Put $Q = \prod_{t=1}^{n_{\mathsf{gap}}} (1 - 4\eta_t/(4\kappa - 1)) > 0$. For every $w^{(1)} \geq w_{\min}(\zeta)$, there are $1 \leq w^{(2)} \leq 2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)$ and $\upsilon > 0$ such that*

$$\begin{gathered} g(\boldsymbol{y}) = \upsilon^2 \mathcal{E}_{\mathcal{C}_\zeta}(\boldsymbol{y}/\upsilon - (w^{(1)}, w^{(2)})), \\ \upsilon = \frac{Q}{w^{(1)} + 2\zeta + K}, \quad 0 \leq K \leq \mathsf{rad}_{\zeta,0} \frac{(\kappa - 2)(1 - Q)}{2}, \\ w^{(2)} - 1 \leq \frac{(\kappa - 2)\mathsf{rad}_{\mathsf{turn}}(\zeta)}{2w^{(1)}}. \end{gathered} \tag{B.8}$$

Gradient descent on (B.1) from $(1, 0)$ ends the gap at $\upsilon(w^{(1)} + 2\zeta, w^{(2)} - 1)$. During the gap, its first coordinate is positive and nonincreasing, and its second coordinate is nonnegative and nondecreasing. The global zero sets and derivative bounds in Eqs. (B.6) & (B.7) hold. The local objective is in $\mathcal{F}_{4/(4\kappa-1),\, 2\kappa/(4\kappa-1)}$ and is minimized at zero. The relative backward recursion is continuous in the trial height $w^{(2)} \in [1, 2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)]$ before matching; no continuous selection of matching heights is asserted.

Proof. **Step 1: certify one backward update.** Fix a trial $w^{(2)}$, write $\boldsymbol{w} = (w^{(1)}, w^{(2)})$, and solve backward from $\widehat{\boldsymbol{y}}_{n_{\mathsf{gap}}} = \boldsymbol{v}_{\zeta,\Phi_\zeta}$. Relative and physical coordinates are related by $\boldsymbol{y} = \upsilon(\widehat{\boldsymbol{y}} + \boldsymbol{w})$. Until the backward path reaches the horizontal branch, write $\widehat{\boldsymbol{y}}_t = \boldsymbol{v}_{\zeta,\psi} + \xi\mathbf{normal}_{\zeta,\psi}$, $\xi \geq 0$. For $\eta = \eta_t$, set $\widehat{\boldsymbol{y}}_{\mathsf{aux}} = \widehat{\boldsymbol{y}}_t + \mu_{\mathsf{c}}\eta\boldsymbol{w}$. If the previous projection is $\boldsymbol{v}_{\zeta,\varphi}$, inversion forces $\widehat{\boldsymbol{y}}_{t-1} = (\widehat{\boldsymbol{y}}_{\mathsf{aux}} + \beta_{\mathsf{c}}\eta\boldsymbol{v}_{\zeta,\varphi})/(1-\mu_{\mathsf{c}}\eta)$. Its normal-ray condition is the scalar equation $\langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \mathbf{tangent}_{\zeta,\varphi}\rangle/\mathsf{rad}_{\zeta,\varphi} = \zeta((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta - 1)$. For $0 \leq \varphi \leq \psi$, the $\boldsymbol{e}_\varphi$ component of $\widehat{\boldsymbol{y}}_{\mathsf{aux}}$ is positive: writing $\Delta = \psi - \varphi$, the contribution from $\widehat{\boldsymbol{y}}_t$ is

$$\langle\widehat{\boldsymbol{y}}_t, \boldsymbol{e}_\varphi\rangle = (\mathsf{rad}_{\zeta,\psi} + \xi)\cos\Delta - \zeta\xi\sin\Delta > 0.$$

Indeed, $\Delta \leq \Phi_\zeta$ implies $\cos\Delta - \zeta\sin\Delta > 0$, and the translation adds a nonnegative contribution by Eq. (B.5). Differentiation and (B.5) give

$$\frac{d}{d\varphi}\frac{\langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \mathbf{tangent}_{\zeta,\varphi}\rangle}{\mathsf{rad}_{\zeta,\varphi}} = -(1+\zeta^2)\frac{\langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \boldsymbol{e}_\varphi\rangle}{\mathsf{rad}_{\zeta,\varphi}} < 0,$$

$$\frac{\langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \mathbf{tangent}_{\zeta,\psi}\rangle}{\mathsf{rad}_{\zeta,\psi}} - \zeta((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta - 1) = \eta\left(\mu_{\mathsf{c}}\frac{\langle\boldsymbol{w}, \mathbf{tangent}_{\zeta,\psi}\rangle}{\mathsf{rad}_{\zeta,\psi}} - (\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\zeta\right) \leq 0.$$

If $\langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \mathbf{tangent}_{\zeta,0}\rangle/\mathsf{rad}_{\zeta,0} > \zeta((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta - 1)$, choose the unique root $\varphi \in (0, \psi]$ of $\langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \mathbf{tangent}_{\zeta,\varphi}\rangle/\mathsf{rad}_{\zeta,\varphi} = \zeta((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta - 1)$; otherwise, choose $\varphi = 0$. In either case, put $\widehat{\boldsymbol{y}}_{t-1} = (\widehat{\boldsymbol{y}}_{\mathsf{aux}} + \beta_{\mathsf{c}}\eta\boldsymbol{v}_{\zeta,\varphi})/(1-\mu_{\mathsf{c}}\eta)$.

To certify either choice, put

$$\mathbf{normal}_{\mathsf{aux}} = \widehat{\boldsymbol{y}}_{\mathsf{aux}} + ((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta - 1)\boldsymbol{v}_{\zeta,\varphi}, \qquad \widehat{\boldsymbol{y}}_{t-1} - \boldsymbol{v}_{\zeta,\varphi} = \frac{\mathbf{normal}_{\mathsf{aux}}}{1-\mu_{\mathsf{c}}\eta}.$$

We must show that $\mathbf{normal}_{\mathsf{aux}}$ belongs to the normal cone at $\boldsymbol{v}_{\zeta,\varphi}$. For an interior root, the scalar equation gives $\mathbf{normal}_{\mathsf{aux}} = (1-\mu_{\mathsf{c}}\eta)\xi_{\mathsf{new}}\mathbf{normal}_{\zeta,\varphi}$ for some scalar ξ_{new}, so only $\xi_{\mathsf{new}} \geq 0$ remains to be checked. For the horizontal choice, the branch-selection inequality already gives $\mathsf{normal}^{(2)}_{\mathsf{aux}} + \zeta\mathsf{normal}^{(1)}_{\mathsf{aux}} \geq 0$, so only $\mathsf{normal}^{(1)}_{\mathsf{aux}} \geq 0$ remains.

If $(\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta \geq 1$, then $\langle\mathbf{normal}_{\mathsf{aux}}, \boldsymbol{e}_\varphi\rangle = \langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \boldsymbol{e}_\varphi\rangle + ((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta - 1)\mathsf{rad}_{\zeta,\varphi} > 0$. This proves the required sign in both cases, because $\langle\mathbf{normal}_{\zeta,\varphi}, \boldsymbol{e}_\varphi\rangle = 1$ and $\boldsymbol{e}_0 = (1, 0)$. Otherwise, the old nonnegative normal coefficient and the translation sign give $\langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \mathbf{normal}_{\zeta,\psi}\rangle \geq \mathsf{rad}_{\zeta,\psi}$. Together with the support inequality $\langle\boldsymbol{v}_{\zeta,\varphi}, \mathbf{normal}_{\zeta,\psi}\rangle \leq \mathsf{rad}_{\zeta,\psi}$, this yields the common estimate

$$\begin{aligned}\langle\mathbf{normal}_{\mathsf{aux}}, \mathbf{normal}_{\zeta,\psi}\rangle &= \langle\widehat{\boldsymbol{y}}_{\mathsf{aux}}, \mathbf{normal}_{\zeta,\psi}\rangle - (1 - (\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta)\langle\boldsymbol{v}_{\zeta,\varphi}, \mathbf{normal}_{\zeta,\psi}\rangle \\ &\geq (\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta\mathsf{rad}_{\zeta,\psi} \geq 0.\end{aligned}$$

For an interior root, $\langle\mathbf{normal}_{\zeta,\varphi}, \mathbf{normal}_{\zeta,\psi}\rangle = (1+\zeta^2)\cos(\psi-\varphi) > 0$, so the multiple is nonnegative. For the horizontal choice, $\mathsf{normal}^{(1)}_{\mathsf{aux}} < 0$ would imply $\mathsf{normal}^{(2)}_{\mathsf{aux}} \geq -\zeta\mathsf{normal}^{(1)}_{\mathsf{aux}} > 0$. The signs $\mathsf{normal}^{(1)}_{\zeta,\psi} > 0$ and $\mathsf{normal}^{(2)}_{\zeta,\psi} < 0$ would then give $\langle\mathbf{normal}_{\mathsf{aux}}, \mathbf{normal}_{\zeta,\psi}\rangle < 0$, a contradiction. Thus, both branches satisfy the normal-cone conditions in Eq. (B.4), certifying their projections. Once $\widehat{\boldsymbol{y}}_t - \boldsymbol{v}_{\zeta,0} \in N_{\mathcal{C}_\zeta}(\boldsymbol{v}_{\zeta,0})$, all further backward steps stay on this branch because

$$\widehat{\boldsymbol{y}}_{t-1} - \boldsymbol{v}_{\zeta,0} = \frac{\widehat{\boldsymbol{y}}_t - \boldsymbol{v}_{\zeta,0} + \eta((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\boldsymbol{v}_{\zeta,0} + \mu_{\mathsf{c}}\boldsymbol{w})}{1-\mu_{\mathsf{c}}\eta} \in N_{\mathcal{C}_\zeta}(\boldsymbol{v}_{\zeta,0}).$$

This includes zero steps. On the curved branch, strict monotonicity of the scalar equation makes its unique root continuous in the previous point and $w^{(2)}$. As the root reaches zero, the curved and horizontal formulas agree. Induction over the finitely many backward steps therefore proves continuity in $w^{(2)}$, including branch changes and zero steps.

Step 2: bound the turning part. For a curved backward update, let $\Delta = \psi - \varphi \geq 0$. Its $\boldsymbol{e}'_\varphi$ component gives

$$\begin{aligned}(1-\mu_{\mathsf{c}}\eta)\zeta\xi_{\mathsf{new}} &= (\mathsf{rad}_{\zeta,\psi} + \xi_{\mathsf{old}})\sin\Delta + \zeta\xi_{\mathsf{old}}\cos\Delta + \mu_{\mathsf{c}}\eta\langle\boldsymbol{w}, \boldsymbol{e}'_\varphi\rangle \\ &\leq (\mathsf{rad}_{\zeta,\psi} + \xi_{\mathsf{old}})\sin\Delta + \zeta\xi_{\mathsf{old}}\cos\Delta.\end{aligned}$$

Since $1-\mu_{\mathsf{c}}\eta \leq 1$, $\mathsf{rad}_{\zeta,\varphi} = e^{\zeta\Delta}\mathsf{rad}_{\zeta,\psi}$, and $\mathsf{rad}_{\zeta,\psi} \leq (\mathsf{rad}_{\zeta,\psi} + \xi_{\mathsf{old}})$,

$$\begin{aligned}(1-\mu_{\mathsf{c}}\eta)(\mathsf{rad}_{\zeta,\varphi} + \xi_{\mathsf{new}}) &\leq (\mathsf{rad}_{\zeta,\psi} + \xi_{\mathsf{old}})\left(\cos\Delta + \frac{\sin\Delta}{\zeta}\right) + \mathsf{rad}_{\zeta,\psi}(e^{\zeta\Delta} - \cos\Delta) \\ &\leq (\mathsf{rad}_{\zeta,\psi} + \xi_{\mathsf{old}})e^{(\zeta+\zeta^{-1})\Delta}.\end{aligned}$$

The last step uses $\sin\Delta \le \Delta$ and $e^{\zeta\Delta} + \Delta/\zeta \le e^{\zeta\Delta}(1+\Delta/\zeta) \le e^{(\zeta+\zeta^{-1})\Delta}$. Backward angles never increase, and after reaching the horizontal branch they remain there. Thus, in forward time the steps whose projections, evaluated at the iterates before those updates, have positive angle form a contiguous terminal suffix. If this suffix is nonempty, let t_0 be its first step index and write

$$\widehat{\boldsymbol{y}}_{\mathsf{turn}} = \widehat{\boldsymbol{y}}_{t_0-1} = \boldsymbol{v}_{\zeta,\varphi} + \xi\,\mathbf{normal}_{\zeta,\varphi}, \qquad Q_{\mathsf{turn}} = \prod_{t=t_0}^{n_{\mathsf{gap}}} (1-\mu_{\mathsf{c}}\eta_t).$$

Each update in this suffix is a curved backward update; the sum of its angle increments is $\Phi_\zeta - \varphi$. Multiplying their radial estimates, starting from the terminal state $\boldsymbol{v}_{\zeta,\Phi_\zeta}$, gives

$$Q_{\mathsf{turn}}(\mathsf{rad}_{\zeta,\varphi} + \xi) \le \sqrt{1+4\zeta^2}\, e^{(\zeta+\zeta^{-1})(\Phi_\zeta-\varphi)} = \mathsf{rad}_{\mathsf{turn}}(\zeta) e^{-(\zeta+\zeta^{-1})\varphi}.$$

This estimate yields

$$Q_{\mathsf{turn}}\widehat{y}^{(1)}_{\mathsf{turn}} \le \mathsf{rad}_{\mathsf{turn}}(\zeta), \qquad -Q_{\mathsf{turn}}\widehat{y}^{(2)}_{\mathsf{turn}} \le \zeta\mathsf{rad}_{\mathsf{turn}}(\zeta). \tag{B.9}$$

Indeed, $\widehat{y}^{(1)}_{\mathsf{turn}} \le \mathsf{rad}_{\zeta,\varphi} + \xi$, $-\widehat{y}^{(2)}_{\mathsf{turn}} \le (\mathsf{rad}_{\zeta,\varphi} + \xi)(\sin\varphi + \zeta\cos\varphi)$, and

$$\frac{d}{d\varphi}\left[e^{-(\zeta+\zeta^{-1})\varphi}(\sin\varphi + \zeta\cos\varphi)\right] = -e^{-(\zeta+\zeta^{-1})\varphi}\left[\zeta^2\cos\varphi + (\zeta^{-1}+2\zeta)\sin\varphi\right] \le 0.$$

Step 3: match the input. Write the backward projections as $\boldsymbol{v}_{t-1}$, with $0 \le v^{(1)}_{t-1} \le \mathsf{rad}_{\zeta,0}$ and $-1 \le v^{(2)}_{t-1} \le 0$, and define

$$w_{\mathsf{out}}(w^{(2)}) = 1 + \sum_t \beta_{\mathsf{c}}\eta_t[-v^{(2)}_{t-1}] \prod_{l>t}(1-\mu_{\mathsf{c}}\eta_l).$$

Indeed, unrolling the physical recursion and substituting the prescribed endpoint gives

$$\upsilon(w^{(2)} - 1) = y^{(2)}_{n_{\mathsf{gap}}} = Qy^{(2)}_0 + \upsilon(w_{\mathsf{out}}(w^{(2)}) - 1), \qquad Qy^{(2)}_0 = \upsilon(w^{(2)} - w_{\mathsf{out}}(w^{(2)})).$$

Since $Q, \upsilon > 0$, the initial condition $y^{(2)}_0 = 0$ holds exactly when $w_{\mathsf{out}}(w^{(2)}) = w^{(2)}$. Only the curved suffix contributes. For a nonempty suffix, use the index t_0 from Step 2. Let $K_{\mathsf{turn}} = \sum_{t=t_0}^{n_{\mathsf{gap}}} \beta_{\mathsf{c}}\eta_t v^{(1)}_{t-1} \prod_{l>t}(1-\mu_{\mathsf{c}}\eta_l) \ge 0$ be its damped first-coordinate gradient sum. Unrolling the first coordinate from the suffix's physical input $\upsilon(\widehat{y}^{(1)}_{\mathsf{turn}} + w^{(1)})$ to its output $\upsilon(2\zeta + w^{(1)})$ gives

$$2\zeta + w^{(1)} = Q_{\mathsf{turn}}(\widehat{y}^{(1)}_{\mathsf{turn}} + w^{(1)}) - K_{\mathsf{turn}}, \qquad Q_{\mathsf{turn}}\widehat{y}^{(1)}_{\mathsf{turn}} = 2\zeta + K_{\mathsf{turn}} + w^{(1)}(1-Q_{\mathsf{turn}}).$$

Thus, $1 - Q_{\mathsf{turn}} \le (\mathsf{rad}_{\mathsf{turn}}(\zeta) - 2\zeta)/w^{(1)}$. Unrolling the second coordinate and using Eq. (B.9),

$$\begin{aligned} 1 \le w_{\mathsf{out}}(w^{(2)}) &= -Q_{\mathsf{turn}}\widehat{y}^{(2)}_{\mathsf{turn}} + w^{(2)}(1-Q_{\mathsf{turn}}) \\ &\le \zeta\mathsf{rad}_{\mathsf{turn}}(\zeta) + 2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)\frac{\mathsf{rad}_{\mathsf{turn}}(\zeta) - 2\zeta}{w^{(1)}} \le 2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta), \end{aligned}$$

because $w^{(1)} \ge w_{\min}(\zeta) = 2(1+2\zeta^2)\mathsf{rad}_{\mathsf{turn}}(\zeta) \ge 2\mathsf{rad}_{\mathsf{turn}}(\zeta)$. If the suffix is empty, $w_{\mathsf{out}}(w^{(2)}) = 1$. Thus w_{out} is a continuous self-map of $[1, 2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)]$, and the intermediate value theorem gives a fixed point.

Set $K = \sum_t \beta_{\mathsf{c}}\eta_t v^{(1)}_{t-1} \prod_{l>t}(1-\mu_{\mathsf{c}}\eta_l) \le \mathsf{rad}_{\zeta,0}\frac{\beta_{\mathsf{c}}(1-Q)}{\mu_{\mathsf{c}}}$ and choose υ as in Eq. (B.8). The unrolled identities $Qy^{(1)}_0 = \upsilon(w^{(1)} + 2\zeta + K)$ and $Qy^{(2)}_0 = \upsilon(w^{(2)} - w_{\mathsf{out}}(w^{(2)}))$ then give input $(1, 0)$. Every backward iterate has a certified projection, so reversing this finite recursion gives the actual GD trajectory from $(1, 0)$. Moreover,

$$w^{(2)} - 1 \le \sum_{t \text{ in suffix}} \beta_{\mathsf{c}}\eta_t \prod_{l>t}(1-\mu_{\mathsf{c}}\eta_l) = \frac{\beta_{\mathsf{c}}}{\mu_{\mathsf{c}}}(1-Q_{\mathsf{turn}}) \le \frac{\beta_{\mathsf{c}}\mathsf{rad}_{\mathsf{turn}}(\zeta)}{\mu_{\mathsf{c}} w^{(1)}}.$$

Step 4: verify the forward trajectory and gradient properties. The positive relative first coordinates and nonnegative first gradient components make the physical first coordinate positive and nonincreasing. The recursion $y^{(2)}_t = (1-\mu_{\mathsf{c}}\eta_t)y^{(2)}_{t-1} + \upsilon\beta_{\mathsf{c}}\eta_t[-v^{(2)}_{t-1}]$ is a convex combination with nondecreasing target levels $\upsilon\beta_{\mathsf{c}}[-v^{(2)}_{t-1}]/\mu_{\mathsf{c}}$. Starting from zero, it is therefore nonnegative and nondecreasing.

Finally, spatial scaling preserves convexity and smoothness, and adding the quadratic proves the class and minimizer claims. □

B.2 Ordinary and exceptional transfers

The startup thresholds a_ζ, geometric cost corrections $c(a_\zeta)$, and ordinary thresholds $\tau_\zeta = 3\zeta^2/2$ are defined in Eq. (A.1). We now realize their numerical transfer costs by local transfer functions.

B.2.1 Ordinary transfers and damping

Write $s = \sum_t \eta_t$ and $Q = \prod_t (1 - 4\eta_t/(4\kappa - 1))$ for the gap step sum and damping. The second coordinate at the end of the gap is $\upsilon(w^{(2)} - 1)$; the derivative threshold is $\upsilon w^{(2)}$, at or above which $\partial_2 g = 0$. An *input threshold* τ_{in} means that $g = \nabla g = 0$ when $y^{(1)} \le \tau_{\mathsf{in}}, y^{(2)} \ge 0$, while the gap trajectory has $y^{(1)} > \tau_{\mathsf{in}}$. We will compare these thresholds for successive transfers so that the predecessor has zero derivative in the shared coordinate and the successor has zero gradient during the active transfer.

Lemma B.3 (Ordinary transfer)**.** *Let κ be sufficiently large. Suppose a gap has steps $0 \le \eta_t < (4\kappa - 1)/4$, step sum $s = \sum_t \eta_t$, damping $Q = \prod_t (1 - 4\eta_t/(4\kappa - 1))$, checkpoint $b \ge a$, and $a \in [a_{\min}, (4\kappa - 1)/4]$. For every $0 \le \tau_{\mathsf{in}} < Q$, there is a transfer function with input threshold at least τ_{in}, gap-end second coordinate and derivative threshold below $\tau_{\zeta(a)} B$, and positive transfer factor satisfying*

$$B^{-1} \le \frac{c(a)(a+s)}{b} \frac{1 - \tau_\zeta}{1 - \tau_{\mathsf{in}}} \frac{1 - \tau_{\mathsf{in}}}{Q - \tau_{\mathsf{in}}}, \qquad \frac{1 - \tau_{\mathsf{in}}}{Q - \tau_{\mathsf{in}}} > 0. \tag{B.10}$$

For $s \le (4\kappa - 1)/64$, $\tau_{\mathsf{in}} \le 1/2$, the last factor is at most $e^{12s/(4\kappa-1)}$. The zero radius satisfies $\mathsf{rad}_{\mathsf{loc}} \ge (\zeta/2)\upsilon w^{(1)}$, and $\upsilon w^{(1)} \ge \tau_{\mathsf{in}}$. If a nonleading input threshold is a preceding ordinary τ or $1/2$, both its zero radius and gap first-coordinate bound exceed $(\kappa - 1)^{-2}$ for sufficiently large κ. For a first transfer with no requested input threshold, the gap may have arbitrary step sum if each gap step is below $(4\kappa - 1)/4$. The leading transfer satisfies

$$\begin{aligned} B^{-1} &\le \frac{c(a)(a+s)}{b} \frac{1 - \tau_\zeta}{Q}, \qquad \min_{\mathsf{gap}} y^{(1)} \ge Q/(2(\kappa - 1)), \\ \upsilon w^{(1)} &\ge Q/(2(\kappa - 1)), \qquad \mathsf{rad}_{\mathsf{loc}} \ge Q(\kappa - 1)^{-2}. \end{aligned} \tag{B.11}$$

Proof. **Step 1: choose the translation to enforce the input threshold.** With $\zeta = \zeta(a)$, choose

$$w^{(1)} = \max\left\{ w_{\min}(\zeta), \frac{\tau_{\mathsf{in}}(2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)}{2}) + Q\mathsf{rad}_{\mathsf{turn}}(\zeta)}{Q - \tau_{\mathsf{in}}} \right\}. \tag{B.12}$$

Lem. B.2 gives a matching $w^{(2)} \le 2\zeta \mathsf{rad}_{\mathsf{turn}}(\zeta)$. Its strip boundary satisfies

$$\upsilon(w^{(1)} - w^{(2)}/(2\zeta)) \ge \frac{Q(w^{(1)} - \mathsf{rad}_{\mathsf{turn}}(\zeta))}{w^{(1)} + 2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)}{2}} \ge \tau_{\mathsf{in}}.$$

The endpoint first coordinate, and hence every first coordinate in the gap, is strictly above that boundary.

Step 2: bound the derivative threshold relative to the checkpoint output. By Eq. (A.1) and $a \le (4\kappa - 1)/4$,

$$\frac{2w^{(2)}}{\kappa} \le \frac{4\zeta \mathsf{rad}_{\mathsf{turn}}(\zeta)}{\kappa} = \frac{(\frac{4(\kappa-1)}{4\kappa-1}) a \tau_\zeta (1 - \tau_\zeta)^2}{\kappa} \le \tau_\zeta.$$

The terminal transfer gradient is $\upsilon(2\zeta, -1)$, so

$$B = \upsilon[\frac{2(\kappa - 2w^{(2)})b}{4\kappa - 1} + w^{(2)} - 1] \ge \upsilon(1 - \tau_\zeta)\frac{2\kappa}{4\kappa - 1} b > 0. \tag{B.13}$$

Therefore

$$\frac{\upsilon w^{(2)}}{B} \le \frac{4(\frac{4\kappa-1}{4(\kappa-1)})\zeta \mathsf{rad}_{\mathsf{turn}}(\zeta)}{(1 - \tau_\zeta)b} \le \tau_\zeta(1 - \tau_\zeta) < \tau_\zeta.$$

Throughout the gap, the second coordinate is at most $\upsilon(w^{(2)} - 1) < \upsilon w^{(2)}$ by monotonicity.

Step 3: bound the cost of the translation and damping. Since $\mathsf{rad}_{\mathsf{turn}}(\zeta) \le w_{\min}(\zeta)$, (B.12) gives

$$w^{(1)} + 2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)}{2} \le \frac{Q}{Q-\tau_{\mathsf{in}}}(w_{\min}(\zeta) + 2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)}{2}).$$

The startup threshold a_ζ bounds the cost of the fixed backward shooting translation:

$$2(\frac{4\kappa-1}{4(\kappa-1)})(w_{\min}(\zeta) + 2\zeta) \le \mathsf{rad}_{\zeta,0} a_\zeta.$$

Substituting the formulas for $w_{\min}(\zeta)$ and a_ζ reduces this bound to

$$\tau_\zeta(1-\tau_\zeta)^2\left(\frac{1+2\zeta^2}{\zeta} + \frac{1}{\mathsf{rad}_{\mathsf{turn}}(\zeta)}\right) \le \mathsf{rad}_{\zeta,0}.$$

Indeed, $\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta) \ge 1$ and $\tau_\zeta = 3\zeta^2/2$ bound its left side by $(3/2)\zeta(1+3\zeta^2) \le 57/128 < 1 \le \mathsf{rad}_{\zeta,0}$. The gap-dependent part is bounded by $(\kappa-2)(1-Q) \le (\frac{4(\kappa-1)}{4\kappa-1})s$. Combining these bounds with $\frac{2\kappa}{4\kappa-1} \ge \frac{1}{2}\frac{4(\kappa-1)}{4\kappa-1}$ in Eq. (B.13) gives

$$\begin{aligned} B^{-1} &\le \frac{2(\frac{4\kappa-1}{4(\kappa-1)})(w^{(1)} + 2\zeta + K)}{Q(1-\tau_\zeta)b} \\ &\le \frac{(\frac{4\kappa-1}{4(\kappa-1)})[2(w_{\min}(\zeta)+2\zeta) + \mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)]}{(Q-\tau_{\mathsf{in}})(1-\tau_\zeta)b} \le \frac{\mathsf{rad}_{\zeta,0}(a_\zeta + s)}{(Q-\tau_{\mathsf{in}})(1-\tau_\zeta)b}. \end{aligned}$$

Using $c(a_\zeta) = \mathsf{rad}_{\zeta,0}/(1-\tau_\zeta)^2$ yields Eq. (B.10). For $s \le (4\kappa-1)/64$ and $\tau_{\mathsf{in}} \le 1/2$,

$$\log\frac{1-\tau_{\mathsf{in}}}{Q-\tau_{\mathsf{in}}} \le \frac{1-Q}{Q-\tau_{\mathsf{in}}} \le \frac{\mu_{\mathsf{c}} s}{1-\tau_{\mathsf{in}}-\mu_{\mathsf{c}} s} \le 3\mu_{\mathsf{c}} s.$$

Here $Q \ge 1-\mu_{\mathsf{c}} s$ follows by multiplying factors $1-\mu_{\mathsf{c}}\eta_t \in (0,1]$, and $1-\tau_{\mathsf{in}}-\mu_{\mathsf{c}} s \ge 7/16$ makes every denominator positive. For a first transfer, take $w^{(1)} = w_{\min}(\zeta)$ and omit the input threshold requirement. The same calculation yields Eq. (B.11) for arbitrary step sum. For the remaining first-coordinate bound, $\mathsf{rad}_{\zeta,0} < 2$ gives $K < (\kappa-1)$, while $2\zeta \le 1/2$ and $w_{\min}(\zeta) \ge 1$ give

$$\frac{w_{\min}(\zeta) + 2\zeta + K}{w_{\min}(\zeta)} \le (\kappa-1) + 3/2 \le 2(\kappa-1).$$

Thus, $\upsilon w_{\min}(\zeta) \ge Q/(2(\kappa-1))$. The certified relative first coordinate is positive throughout the gap, so every physical first coordinate is at least this large.

Uniform ordinary radii. For every allowed ordinary startup threshold, Lem. A.1 gives $(\frac{4(\kappa-1)}{4\kappa-1})a_\zeta > e^{1/\zeta}$. Together with $a_\zeta \le (4\kappa-1)/4$, this yields

$$\zeta(a) > 1/\log(\kappa-1), \qquad \tau_{\zeta(a)} > \frac{1}{100(\log(\kappa-1))^2}. \tag{B.14}$$

An ordinary transfer with ordinary predecessor has input threshold $\tau_{\zeta_{\mathsf{prev}}} > 3/[2(\log(\kappa-1))^2]$; hence Eq. (B.7) gives

$$\mathsf{rad}_{\mathsf{loc}} \ge (\zeta/2)\upsilon w^{(1)} \ge (\zeta/2)\tau_{\zeta_{\mathsf{prev}}} > \frac{3}{4(\log(\kappa-1))^3} > (\kappa-1)^{-2}$$

for sufficiently large κ. With input threshold $1/2$, the same bound gives $\mathsf{rad}_{\mathsf{loc}} \ge \zeta/4$. The gap first coordinate exceeds its assigned input threshold, so it too exceeds $(\kappa-1)^{-2}$ in these nonleading cases. For a leading transfer, (B.11) gives $\mathsf{rad}_{\mathsf{loc}} \ge \zeta Q/(4(\kappa-1)) \ge Q(\kappa-1)^{-2}$ eventually. All thresholds here are absolute, since $(\kappa-1)/(\log(\kappa-1))^2 \to \infty$ and $(\kappa-1)^2/(\log(\kappa-1))^3 \to \infty$.

□

B.2.2 EXCEPTIONAL CHECKPOINTS

For checkpoints satisfying only $b > 16(\frac{4\kappa-1}{4(\kappa-1)})$, fix $\zeta = 1/4$ and increase $w^{(1)}$ to reduce the gap-end second coordinate. Use tolerance

$$\sigma_\kappa = \frac{1}{100(\log(\kappa-1))^2}.$$

All exceptional transfers use the fixed pitch $\zeta = 1/4$. Assume κ is sufficiently large that $\mathsf{rad}_{\mathsf{turn}}(1/4)/\kappa \le \sigma_\kappa$. The successor of an exceptional transfer, when present, will use input threshold $1/2$.

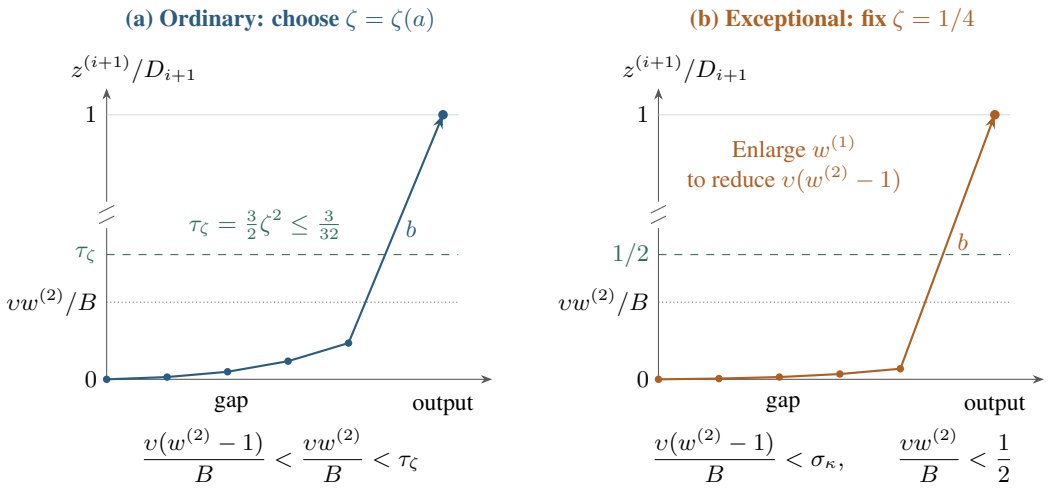
Fixed-pitch parameter bounds. At fixed pitch $\zeta = 1/4$, the quantities $\mathsf{rad}_{\zeta,0}$, $\mathsf{rad}_{\mathsf{turn}}(\zeta)$, and $w_{\min}(\zeta)$ are absolute constants. Moreover, $\mathsf{rad}_{\zeta,0} = (29/32)^2 c(a_{\min}) < 3/2$ by (A.2), and $(\frac{4(\kappa-1)}{4\kappa-1})a_{\min}$ is constant. Thus $4\kappa - 1 \ge 16a_{\min}$ eventually, and

$$\frac{4\zeta \mathsf{rad}_{\mathsf{turn}}(\zeta)}{\kappa\sigma_\kappa} \longrightarrow 0, \qquad \sigma_\kappa \longrightarrow 0.$$

In particular, $2w^{(2)}/\kappa \le \sigma_\kappa < 1/2$ throughout the trial height interval for sufficiently large κ.

Lemma B.4 (Exceptional transfer)**.** *For a gap of nonnegative steps with step sum $s \le (4\kappa-1)/64$ and a checkpoint $b > 16(\frac{4\kappa-1}{4(\kappa-1)})$, there is a transfer function with input threshold $1/2$, gap-end second coordinate below $\sigma_\kappa B$, derivative threshold below $B/2$, and transfer cost at most $(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b$. For a sole first transfer with arbitrary gap step sum and each gap step below $(4\kappa-1)/4$, the transfer cost is at most $(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/(Qb)$ and no input threshold is required. In the latter case, $vw^{(1)}$ and every gap first coordinate are at least $Q/(2(\kappa-1))$. The zero radius and gap first-coordinate bound are at least $(\kappa-1)^{-2}$ in the short-gap case and $Q(\kappa-1)^{-2}$ in the sole-first case.*

Fig. 7 compares this guarantee with Lem. B.3. An ordinary transfer bounds both the gap-end second coordinate and the derivative threshold, relative to the output, by the same small threshold τ_ζ. An exceptional transfer instead has two separate requirements: a small gap-end second coordinate for compatibility and a derivative threshold below half the output.



Broken ordinates show threshold order, not numerical spacing; paths are schematic.

Figure 7: Bounds on the gap-end second coordinate and the derivative threshold for (a) ordinary and (b) exceptional transfers (Lems. B.3 & B.4). In the phase construction of Lem. B.7, ordinary-to-ordinary transitions use input threshold τ_ζ, while a transfer following an exceptional one uses $1/2$.

Proof. At $\zeta = 1/4$, choose the translation to satisfy the conditions for backward shooting, the input threshold $1/2$, and the tolerance on the gap-end second coordinate:

$$w^{(1)} = \max\left\{ w_{\min}(\zeta), \frac{(2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)}{2})/2 + Q\mathsf{rad}_{\mathsf{turn}}(\zeta)}{Q - 1/2}, \frac{\frac{(\kappa-2)\mathsf{rad}_{\mathsf{turn}}(\zeta)}{2}}{\sigma_\kappa} \right\}.$$

Since $Q \geq 15/16$, the middle term makes

$$\upsilon(w^{(1)} - w^{(2)}/(2\zeta)) \geq \frac{Q(w^{(1)} - \mathsf{rad}_{\mathsf{turn}}(\zeta))}{w^{(1)} + 2\zeta + \mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)/2} \geq \frac{1}{2},$$

which gives the input threshold; the last term gives $w^{(2)} - 1 \leq \sigma_\kappa$ by Lem. B.2. The middle term in the maximum defining the exceptional $w^{(1)}$ is $O(\kappa)$, since $Q \geq 15/16$ and $1 - Q \leq 1/16$. The last term is of order $\kappa(\log(\kappa-1))^2$ and eventually dominates that maximum, uniformly in the gap. Hence $(\kappa - 1) \leq w^{(1)} \leq (\kappa-1)^2/16$ for sufficiently large κ. The same bounds hold when the middle term is omitted for a leading transfer. These facts justify the scale and cost estimates in Lem. B.4 without imposing a bound on the leading gap sum.

For sufficiently large κ, therefore, $2w^{(2)}/\kappa \leq \sigma_\kappa < 1/2$ and $(\kappa - 1) \leq w^{(1)} \leq (\kappa-1)^2/16$. Therefore

$$B/\upsilon \geq (1 - \sigma_\kappa)\frac{2\kappa}{4\kappa - 1} b > 8(1 - \sigma_\kappa).$$

Dividing $w^{(2)} - 1 \leq \sigma_\kappa$ and $w^{(2)} \leq 1 + \sigma_\kappa$ by this bound gives $\upsilon(w^{(2)} - 1) < \sigma_\kappa B$ and $\upsilon w^{(2)} < B/2$. Monotonicity extends the second-coordinate bound to all earlier gap iterates. Since $K \leq \mathsf{rad}_{\zeta,0}(\kappa - 1)/32$, $\mathsf{rad}_{\zeta,0} < 3/2$, and $Q \geq 15/16$, the scale formula gives $\upsilon^{-1} \leq 2w^{(1)}$ and

$$B^{-1} \leq \frac{2(\frac{4\kappa-1}{4(\kappa-1)})\upsilon^{-1}}{(1-\sigma_\kappa)b} \leq \frac{8(\frac{4\kappa-1}{4(\kappa-1)})w^{(1)}}{b} \leq \frac{(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3}{b}.$$

For a sole first transfer, omit the middle term in the maximum. Now $K < \mathsf{rad}_{\zeta,0}(\kappa-1)/2 < 3(\kappa-1)/4$, so $w^{(1)} \geq (\kappa - 1)$ gives $\upsilon^{-1} = (w^{(1)} + 2\zeta + K)/Q \leq 2w^{(1)}/Q$. This adds only Q^{-1} to the cost, for arbitrary gap sum. The first-coordinate estimate follows from $\upsilon w^{(1)} = Qw^{(1)}/(w^{(1)} + 2\zeta + K) \geq Q/2 \geq Q/(2(\kappa-1))$ and positivity of the relative first coordinates. For a short gap the input threshold gives $\upsilon w^{(1)} \geq 1/2$ and $\mathsf{rad}_{\mathsf{loc}} \geq \zeta/4$ by (B.7); for a sole first transfer it gives $\mathsf{rad}_{\mathsf{loc}} \geq \zeta Q/4$. At fixed $\zeta = 1/4$, these imply the stated radius and first-coordinate floors for sufficiently large κ. □

B.3 A CONTROLLED SECOND TRANSFER

An incoming coupling corrects the core gradient in the direction of its initial gradient, which is supported on the first two coordinates. If the first checkpoint is too large to preserve the readout invariant, the second pair must tolerate this correction; after launch into coordinate three, the correction acts only on retired coordinates. We construct a family continuous in all bounded control vectors before choosing the matching trajectory.

Suppose the first transfer is ordinary and its checkpoint is $b_1 > (4\kappa - 1)/2$. The core's initial gradient $\boldsymbol{g}_*$ and first transfer factor B_1 are fixed. The second-pair horizontal correction has magnitude bounded by

$$u_{\max} = \frac{|g_*^{(2)}|}{B_1} \leq \frac{1}{(1-\tau_1)b_1}, \qquad u_{\max} s_2 \leq \frac{2s_2}{(4\kappa-1)(1-\tau_1)}, \tag{B.15}$$

where s_2 obeys the selected gap cap. Since $\tau_1 \leq 3/32$ and $s_2 \leq (4\kappa - 1)/64$, this is at most $1/29$. Indeed, $|g_*^{(2)}| \leq [2(\kappa-2)/(4\kappa-1)]\upsilon_1$, whereas $B_1 \geq \upsilon_1(1-\tau_1)\frac{2\kappa}{4\kappa-1}b_1$ and $2(\kappa-2)/(4\kappa-1) \leq 2\kappa/(4\kappa-1)$. An incoming core correction $-\theta_{t-1}\boldsymbol{g}_*$, $0 \leq \theta_{t-1} \leq 1$, therefore changes the normalized second-pair update to

$$\boldsymbol{y}_t = \boldsymbol{y}_{t-1} - \eta_t \left[\frac{4}{4\kappa-1}\boldsymbol{y}_{t-1} + \frac{2\kappa-4}{4\kappa-1}\nabla g(\boldsymbol{y}_{t-1}) + u_t \boldsymbol{e}^{(1)} \right], \qquad u_t = u_{\max}\theta_{t-1} \in [0, u_{\max}]. \tag{B.16}$$

The plus sign is due to $g_*^{(2)} \le 0$.

Set $\zeta = 1/4$ and use the fixed translation

$$w_{\mathsf{ctrl}} = 16\left(w_{\min}(\zeta) + 2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-1)}{32} + \frac{(\kappa-1)\mathsf{rad}_{\mathsf{turn}}(\zeta)}{\sigma_\kappa}\right). \tag{B.17}$$

Lemma B.5 (Compact controlled shooting family)**.** *Let κ be sufficiently large, $\eta_1, \dots, \eta_{n_{\mathsf{gap}}} \ge 0$, $\sum_t \eta_t = s \le s_{\max} := (4\kappa-1)/64$, $u_{\max} \ge 0$, $u_{\max} s \le 1/29$, and $b > 16(\frac{4\kappa-1}{4(\kappa-1)})$. Set $\zeta = 1/4$, $\overline{w}^{(2)} = 2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)$, and $\tau_{\mathsf{in},*} = 1/2$. The transfer functions $g_{\upsilon,w^{(2)}}(\boldsymbol{y}) = \upsilon^2\mathcal{E}_{\mathcal{C}_\zeta}(\boldsymbol{y}/\upsilon - (w^{(1)}, w^{(2)}))$, with $w^{(1)} = w_{\mathsf{ctrl}}$ from Eq. (B.17), have a nonempty compact convex trial rectangle Trials and a jointly continuous map $\mathsf{Shoot} : \mathsf{Trials} \times [0, u_{\max}]^{n_{\mathsf{gap}}} \to \mathsf{Trials}$. For each fixed control vector, this map is a self-map of the trial rectangle. Every trial vanishes on $\{y^{(1)} \le \tau_{\mathsf{in},*}, y^{(2)} \ge 0\}$ and on a common ball of radius at least $(\kappa-1)^{-2}$. Its derivative threshold is $\delta = \upsilon w^{(2)}$, independent of the first coordinate; its positive scales and formal transfer factors are continuous in the trial, and $(\boldsymbol{y}, \upsilon, w^{(2)}) \mapsto \nabla g_{\upsilon,w^{(2)}}(\boldsymbol{y})$ is jointly continuous. Its formal transfer factor satisfies*

$$B^{\mathsf{trial}} = \upsilon[\frac{2(\kappa-2w^{(2)})b}{4\kappa-1} + w^{(2)} - 1] \ge (\frac{4(\kappa-1)}{4\kappa-1})b/(8(\kappa-1)^3).$$

For every control vector $\boldsymbol{u} \in [0, u_{\max}]^{n_{\mathsf{gap}}}$ and trial $\boldsymbol{p} = (\upsilon, w^{(2)})$ satisfying $\boldsymbol{p} = \mathsf{Shoot}(\boldsymbol{p}, \boldsymbol{u})$, the controlled gap from $(1,0)$ ends at $\upsilon(w^{(1)} + 2\zeta, w^{(2)} - 1)$; its first coordinate is nonincreasing and above $\tau_{\mathsf{in},*}$, and its second is nonnegative and nondecreasing. The formal factor is the actual checkpoint output, with gap-end second coordinate below $\sigma_\kappa B^{\mathsf{trial}}$ and derivative threshold below $\tau_{\mathsf{in},*}B^{\mathsf{trial}}$. The checkpoint's horizontal control may be arbitrary.

Proof. **Step 1: fix a trial rectangle.** Put $Q = \prod_t(1-\mu_{\mathsf{c}}\eta_t)$. Set $w^{(1)} = w_{\mathsf{ctrl}}$ from (B.17) and let

$$\mathsf{Trials} = \left[\frac{(Q - u_{\max}s)}{w^{(1)} + 2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)}{2}}, \frac{Q}{w^{(1)} + 2\zeta}\right] \times [1, \overline{w}^{(2)}]. \tag{B.18}$$

We first verify, uniformly on this rectangle,

$$\begin{aligned} \upsilon(w^{(1)} - w^{(2)}/(2\zeta)) > \tau_{\mathsf{in},*}, \qquad 2w^{(2)}/\kappa \le \sigma_\kappa, \\ w^{(1)} \ge w_{\min}(\zeta), \qquad \zeta\mathsf{rad}_{\mathsf{turn}}(\zeta) + \overline{w}^{(2)}\mathsf{rad}_{\mathsf{turn}}(\zeta)/w^{(1)} \le \overline{w}^{(2)}. \end{aligned} \tag{B.19}$$

The control bound gives $Q - u_{\max}s \ge 15/16 - 1/29 > 7/8$. We have

$$2\zeta + \mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)/2 \le 2\zeta + \frac{\mathsf{rad}_{\zeta,0}(\kappa-1)}{32} \le \frac{w_{\mathsf{ctrl}}}{16}, \qquad \frac{w^{(2)}}{2\zeta} \le \mathsf{rad}_{\mathsf{turn}}(\zeta) \le \frac{w_{\mathsf{ctrl}}}{16}.$$

The lower scale endpoint in Eq. (B.18) therefore gives, for every trial,

$$\upsilon(w_{\mathsf{ctrl}} - w^{(2)}/(2\zeta)) > \frac{7}{8}\frac{1-1/16}{1+1/16} > \frac{1}{2}.$$

Also $w_{\mathsf{ctrl}} \ge w_{\min}(\zeta)$ and $\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta) + \overline{w}^{(2)}\mathsf{rad}_{\mathsf{turn}}(\zeta)/w_{\mathsf{ctrl}} \le (9/8)\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta) < \overline{w}^{(2)}$. This proves every condition in Eq. (B.19). Since $w_{\mathsf{ctrl}} = O(\kappa(\log(\kappa-1))^2) = o((\kappa-1)^2)$, it is at most $(\kappa-1)^2/16$ for sufficiently large κ. The same scale bounds give $\upsilon^{-1} < 2w_{\mathsf{ctrl}}$ uniformly, so

$$(B^{\mathsf{trial}})^{-1} \le \frac{2(\frac{4\kappa-1}{4(\kappa-1)})\upsilon^{-1}}{(1-\sigma_\kappa)b} \le \frac{8(\frac{4\kappa-1}{4(\kappa-1)})w_{\mathsf{ctrl}}}{b} \le \frac{(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^2}{2b} \le \frac{8(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3}{b}.$$

The height cap equals $2\zeta\mathsf{rad}_{\mathsf{turn}}(\zeta)$, so every trial lies in the translation range of Lem. B.2. The signs in Eq. (B.5) therefore hold for every trial, before its actual controls are identified.

Step 2: construct the backward trajectory for every trial and control. For $(\upsilon, w^{(2)}) \in \mathsf{Trials}$ and $\boldsymbol{u} \in [0, u_{\max}]^{n_{\mathsf{gap}}}$, start at $\widehat{\boldsymbol{y}}_{n_{\mathsf{gap}}} = (2\zeta, -1)$ and use

$$\widehat{\boldsymbol{y}}_{\mathsf{aux}} = \widehat{\boldsymbol{y}}_t + \mu_{\mathsf{c}}\eta_t(w^{(1)}, w^{(2)}) + (\eta_t u_t/\upsilon)\boldsymbol{e}^{(1)}, \qquad \widehat{\boldsymbol{y}}_{t-1} = (\widehat{\boldsymbol{y}}_{\mathsf{aux}} + \beta_{\mathsf{c}}\eta_t\boldsymbol{v}_{t-1})/(1-\mu_{\mathsf{c}}\eta_t).$$

Use the scalar projection rule in Step 1 of Lem. B.2, with this $\widehat{\boldsymbol{y}}_{\mathsf{aux}}$ and the same threshold $\zeta((\mu_{\mathsf{c}}+\beta_{\mathsf{c}})\eta_t-1)$. The added term has nonnegative coefficient $\eta_t u_t/\upsilon$ and direction satisfying

$$\langle \boldsymbol{e}^{(1)}, \mathbf{normal}_{\zeta,\varphi}\rangle > 0, \quad \langle \boldsymbol{e}^{(1)}, \mathbf{tangent}_{\zeta,\varphi}\rangle < 0, \quad \langle \boldsymbol{e}^{(1)}, \boldsymbol{e}_{\varphi}\rangle > 0, \quad \langle \boldsymbol{e}^{(1)}, \boldsymbol{e}'_{\varphi}\rangle \le 0.$$

These signs preserve the root monotonicity, the old-angle threshold, and the normal-cone inequalities used in Step 1 of Lem. B.2. The horizontal branch is preserved because $\boldsymbol{e}^{(1)} \in N_{\mathcal{C}_\zeta}(\boldsymbol{v}_{\zeta,0})$. The nonpositive $\boldsymbol{e}'_{\varphi}$ contribution also preserves the radial bound (B.9). Unique roots, agreement at branch boundaries, and the positive lower scale bound give joint continuity in $(\upsilon, w^{(2)}, \boldsymbol{u})$, including zero steps.

Step 3: define the self-map. Keep controls as independent inputs, so the serial construction can later substitute each trial objective's actual controls. Write its projections as $\boldsymbol{v}_{t-1}$ and set

$$K = \sum_t \beta_{\mathsf{c}}\eta_t v_{t-1}^{(1)} \prod_{l>t}(1-\mu_{\mathsf{c}}\eta_l) \in [0, \frac{\mathsf{rad}_{\zeta,0}(\kappa-2)(1-Q)}{2}],$$
$$u_{\mathsf{damp}} = \sum_t \eta_t u_t \prod_{l>t}(1-\mu_{\mathsf{c}}\eta_l) \in [0, u_{\max} s],$$
$$w_{\mathsf{out}} = 1 + \sum_t \beta_{\mathsf{c}}\eta_t[-v_{t-1}^{(2)}] \prod_{l>t}(1-\mu_{\mathsf{c}}\eta_l).$$

For the curved suffix, let K_{turn} and $u_{\mathsf{damp,turn}}$ denote the corresponding partial sums. First-coordinate unrolling gives

$$Q_{\mathsf{turn}}\widehat{y}_{\mathsf{turn}}^{(1)} = 2\zeta + K_{\mathsf{turn}} + w^{(1)}(1-Q_{\mathsf{turn}}) + u_{\mathsf{damp,turn}}/\upsilon.$$

Both additional sums are nonnegative, so the radial bound still implies $1-Q_{\mathsf{turn}} \le \mathsf{rad}_{\mathsf{turn}}(\zeta)/w^{(1)}$. Horizontal controls do not enter the second-coordinate identity, which remains

$$w_{\mathsf{out}} = -Q_{\mathsf{turn}}\widehat{y}_{\mathsf{turn}}^{(2)} + w^{(2)}(1-Q_{\mathsf{turn}}).$$

Thus, $1 \le w_{\mathsf{out}} \le \zeta\mathsf{rad}_{\mathsf{turn}}(\zeta) + \overline{w}^{(2)}\mathsf{rad}_{\mathsf{turn}}(\zeta)/w^{(1)} \le \overline{w}^{(2)}$ by Eq. (B.19); if the suffix is empty, $w_{\mathsf{out}} = 1$. Consequently,

$$\mathsf{Shoot}((\upsilon, w^{(2)}), \boldsymbol{u}) = \left(\frac{Q-u_{\mathsf{damp}}}{w^{(1)}+2\zeta+K},\ w_{\mathsf{out}}\right) \tag{B.20}$$

is a continuous self-map of Eq. (B.18) for every allowed control vector. The bounds on $K, u_{\mathsf{damp}}, w_{\mathsf{out}}$ place both components inside the rectangle. For an empty gap, $Q=1$, $K=u_{\mathsf{damp}}=0$, and $w_{\mathsf{out}}=1$, so the same formula gives a constant self-map.

Step 4: identify a matching trajectory. Fix an allowed control vector and a point satisfying $\mathsf{Shoot}((\upsilon, w^{(2)}), \boldsymbol{u}) = (\upsilon, w^{(2)})$. First-coordinate unrolling gives the initial physical coordinate $[\upsilon(w^{(1)}+2\zeta+K)+u_{\mathsf{damp}}]/Q = 1$; second-coordinate unrolling gives $Qy_0^{(2)} = \upsilon(w^{(2)}-w_{\mathsf{out}}) = 0$. Reversing the certified recursion gives the controlled trajectory from $(1,0)$. The nonnegative horizontal control preserves first-coordinate monotonicity above the endpoint strip boundary; second-coordinate monotonicity follows exactly as in Lem. B.2. On the curved suffix, $w^{(2)} - 1 \le \frac{(\kappa-2)\mathsf{rad}_{\mathsf{turn}}(\zeta)}{2w^{(1)}}$, while

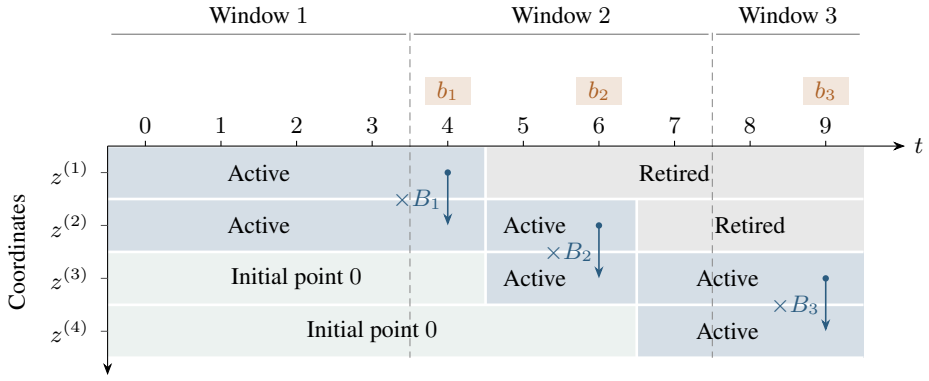
$$B^{\mathsf{trial}} \ge \upsilon(1-\sigma_\kappa)\frac{2\kappa}{4\kappa-1}b,$$

which follows from $2w^{(2)}/\kappa \le \sigma_\kappa$ in Eq. (B.19). At matching, $w^{(2)}-1 \le \sigma_\kappa/32$. Because $b > 16(\frac{4\kappa-1}{4(\kappa-1)})$ and $2\kappa/(4\kappa-1) \ge (\frac{4(\kappa-1)}{4\kappa-1})/2$, the checkpoint lower bound divided by υ exceeds $8(1-\sigma_\kappa)$. Thus the ratio of the gap-end second coordinate to the output is below σ_κ, and the derivative-threshold ratio is below $1/2$, for sufficiently large κ. For every trial, (B.7) and its zero strip give $\mathsf{rad}_{\mathsf{loc}} \ge (\zeta/2)\upsilon w^{(1)} > \zeta/4 = 1/16 \ge (\kappa-1)^{-2}$ eventually. The positive lower scale bound and the projection formula give joint gradient continuity on the full rectangle. Finally, a horizontal checkpoint control does not change its second-coordinate output. □

B.4 Compatible transfer chains

The preceding results supply local transfer functions and their selected positions. Adding these functions on overlapping pairs must preserve their prescribed local updates. The required condition concerns only derivatives entering the active pair: the predecessor may still act on a retired coordinate, provided its derivative in the shared coordinate is zero. This is the compatibility used below, and it is why no bound on retired coordinates is needed. Fix a phase j and suppress its superscript $[j]$ on the core and its local data. Write its selected checkpoints as $t_1 < \cdots < t_{d_j}$, with $b_i = \eta_{t_i}$ and $s_i = \sum_{t=t_{i-1}+1}^{t_i-1} \eta_t$. Here we index the phase locally, with $t_0 = 0$ and initial state at time zero; d_j is the size k of the final checkpoint set from Appx. A. The unscaled core has coordinates $\boldsymbol{z} = (z^{(1)}, \ldots, z^{(d_j+1)}) \in \mathbb{R}^{d_j+1}$. We place successive functions on overlapping coordinate pairs. The predecessor's derivative in the shared coordinate and the successor's gradient vanish along the active trajectory, preventing interference. Fig. 8 illustrates these vanishing derivatives on successive active pairs. It uses the same checkpoint positions as the first phase of Fig. 11; the window boundaries do not alter the active pair.

(a) Successive active pairs in one core



Columns show input iterates at indices $0, \ldots, 9$; the final output is at 10.
Window boundaries do not launch a new pair.

(b) Why only the active pair reaches a fresh coordinate

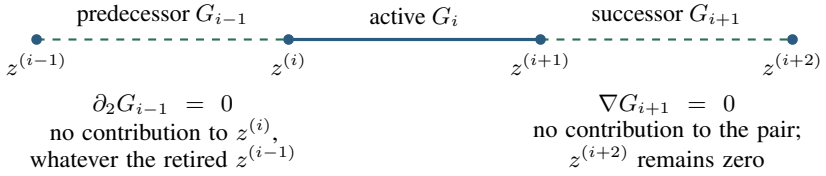


Figure 8: The transfer-chain induction in Lem. B.6. (a) Checkpoint updates advance the active coordinate pair; $B_i = D_{i+1}/D_i$ compares checkpoint output to transfer-entry amplitude. (b) At the iterates before gap and checkpoint updates, compatibility makes the predecessor's derivative in $z^{(i)}$ and the successor's gradient vanish, so the active pair follows its prescribed local GD update, including the quadratic term.

For ordinary transfer i, write its pitch as ζ_i and abbreviate $a_i := a_{\zeta_i}$ and $\tau_i := \tau_{\zeta_i}$ in calculations; thus $\zeta_i = \zeta(a_i)$. Denote its translation coordinates by $w_i^{(1)}, w_i^{(2)}$ and its spatial scale by υ_i. We use the transfer data and compatibility conditions in Def. B.1. These comparisons use the scaled coordinates in the chain below. For selected transfers $i = 1, \ldots, d_j$, install the coordinate chain

$$h(\boldsymbol{z}) = \frac{2\|\boldsymbol{z}\|^2}{4\kappa - 1} + \frac{2\kappa - 4}{4\kappa - 1} \sum_{i=1}^{d_j} D_i^2 g_i(z^{(i)}/D_i, z^{(i+1)}/D_i), \qquad D_1 = 1, \quad D_{i+1} = D_i B_i. \tag{B.21}$$

For a two-coordinate argument, use the scaled transfer function

$$G_i(\boldsymbol{y}) := D_i^2 g_i(\boldsymbol{y}/D_i), \qquad G_i(z^{(i)}, z^{(i+1)}) = D_i^2 g_i(z^{(i)}/D_i, z^{(i+1)}/D_i).$$

For $z^{(i+1)} \geq 0$, the guaranteed input strip $z^{(i)} \leq \tau_{\mathsf{in},i} D_i$ has $G_i = \nabla G_i = 0$. The second gradient component $\partial_2 G_i$ vanishes when $z^{(i+1)} \geq \upsilon_i w_i^{(2)} D_i$. These regions are illustrated in Fig. 1(a), where the derivative labels refer to g_i evaluated at $(z^{(i)}, z^{(i+1)})/D_i$.

Lemma B.6 (Realization of a compatible transfer chain)**.** *Fix $\kappa > 2$, checkpoints $0 = t_0 < t_1 < \cdots < t_{d_j} = n$, and the prescribed nonnegative steps. For each pair, suppose a matched local transfer has the data in Def. B.1, and every two successive transfers are compatible. Install (B.21) with positive amplitudes $D_1 = 1$, $D_{i+1} = D_i B_i$ and start at $\boldsymbol{e}^{(1)}$. If gradient corrections are present during transfer i, their components on coordinates $i, i+1$ must equal D_i times the specified local correction, and their components on coordinates $i+2, \ldots, d_j+1$ must be zero; their components on retired coordinates may be arbitrary. Then every active pair follows its prescribed local trajectory, all future coordinates remain zero before launch, and the checkpoint outputs satisfy $z_{t_i}^{(i+1)} = D_{i+1}$. No bound on the retired coordinates is required.*

Proof. At the beginning of transfer i, suppose $z^{(i)} = D_i$ and all later coordinates are zero. This holds initially. To propagate this assertion, consider the prescribed local trajectory on coordinates $i, i+1$, scaled by D_i. The proposed active trajectory keeps $z^{(i)} > \tau_{\mathsf{in},i} D_i$. If $i > 1$, the predecessor's scaled derivative threshold is below $\tau_{\mathsf{out},i-1} D_i \leq \tau_{\mathsf{in},i} D_i$, so its derivative in coordinate i is zero, independently of $z^{(i-1)}$. If a successor exists, the second coordinate of the active pair satisfies $z^{(i+1)} < \epsilon_{\mathsf{gap},i} D_{i+1} \leq \tau_{\mathsf{in},i+1} D_{i+1}$, which lies in that successor's zero strip, and $z^{(i+2)} = 0$. Every still later transfer function sees $(0, 0)$. The assumed correction also vanishes on future coordinates. Hence, no future coordinate moves, and the core gradient on the active pair is

$$\mu_{\mathsf{c}}(z^{(i)}, z^{(i+1)}) + \beta_{\mathsf{c}} D_i \nabla g_i(z^{(i)}/D_i, z^{(i+1)}/D_i).$$

Dividing the active coordinates and their update by D_i gives exactly the local update with the same prescribed step from Eq. (B.1), including the specified active-pair correction when present. For gap i, the local state and step in Appx. B correspond to

$$\boldsymbol{y}_r = \frac{(z_{t_{i-1}+r}^{(i)}, z_{t_{i-1}+r}^{(i+1)})}{D_i}, \quad 0 \leq r < t_i - t_{i-1}; \qquad \text{gap step } r: \ \eta_{t_{i-1}+r} \quad (1 \leq r < t_i - t_{i-1}).$$

The checkpoint uses $b_i = \eta_{t_i}$ at state index $t_i - 1$. The proposed trajectory and the full GD trajectory start at the same active state and have the same update at every iterate, so induction over the gap steps identifies them. This induction includes the iterate before the checkpoint update. The checkpoint produces $D_i B_i = D_{i+1}$ in coordinate $i+1$ and leaves the later coordinates zero, proving the next induction hypothesis. A retired coordinate may move arbitrarily far at the checkpoint; the predecessor's derivative in the shared coordinate vanishes regardless. The exceptional transfers of Appx. B obey the same induction with their specified input thresholds and bounds on the gap-end second coordinate. The same argument applies to a controlled transfer after matching, because its transfer data include that same active-pair correction.

□

B.5 Uniform families and matched phases

The checkpoint result controls products; the assembly with fixed couplings also needs a common quadratic ball, a fixed initial gradient, and a phase lower bound. A parameter trial is one member of the family before its controlled shooting parameters are matched. The static properties hold for every trial; the trajectory bound will require matching with its actual incoming correction.

A *phase* is a nonempty finite nonnegative schedule of length n ending in a step greater than 64, with every strictly internal maximal run of steps at most this threshold having step sum at most $(4\kappa - 1)/512$. Its leading run is unrestricted. Use the checkpoint selection in Lem. A.15, with window q from (A.21). The cores and couplings use the same prescribed schedule $\boldsymbol{\eta}$. We write $n = n^{[j]}$ for the phase length and index its iterates $\boldsymbol{z}_t$ by $0 \leq t \leq n$.

Formal amplitudes before matching. The numerical prefix costs must bound the scales of every candidate core before a coupling is chosen. We use formal transfer factors for this purpose; their identification with actual checkpoint outputs is a separate step.

Lemma B.7 (Uniform formal amplitude budget)**.** *There are absolute constants $C_D, C_{D,\mathsf{pow}}, C_{D,\mathsf{exp}} > 0$ and an absolute lower threshold on κ, chosen before the schedule, phase, or trial, with the following property. For every κ above that threshold and every phase above, take the one fixed checkpoint set and numerical parameters supplied by Lem. A.15. There is a family of local transfer functions with a fixed first transfer and fixed later local functions except possibly the second. The trial domain* Trials *is a singleton unless at least two checkpoints exist and $b_1 > (4\kappa-1)/2$; in that case the second transfer uses Lem. B.5 and* Trials *is its rectangle. The fixed first transfer has factor $B_1 > 0$; define $\boldsymbol{g}_*$ to have first two components $\mu_{\mathsf{c}}(1,0) + \beta_{\mathsf{c}}\nabla g_1(1,0)$ and all remaining components zero. In the controlled case, $u_{\max} = |g_*^{(2)}|/B_1$ is independent of the trial and satisfies $u_{\max} s_2 \le 1/29$. Every formal transfer amplitude $D_1^{\mathsf{trial}} = 1$, $D_{i+1}^{\mathsf{trial}} = D_i^{\mathsf{trial}} B_i^{\mathsf{trial}}$ is positive and continuous on* Trials*.*

$$(D_i^{\mathsf{trial}})^{-1} \le C_D(\kappa-1)^{C_{D,\mathsf{pow}}} e^{C_{D,\mathsf{exp}}\lceil n/q\rceil} \quad (1 \le i \le d_j+1,\ \boldsymbol{p} \in \mathsf{Trials}), \qquad Q_1^{-1} \le e^{\lceil n/q\rceil/32}. \tag{B.22}$$

Here Q_1 is the damping product in the first gap; these assertions do not assume a matched phase trajectory.

Proof. **Step 1: convert checkpoint costs into amplitudes.** The numerical checkpoint product controls reciprocal transfer amplitudes. The additional geometric costs are the threshold factors at the starts of ordinary runs and the damping within gaps; we account for these separately so that the number of transfers does not create an extra exponential loss. The rules and repair in Appx. A supply a fixed checkpoint set, with every checkpoint greater than 64 and every nonleading gap within its selected cap. Since $\frac{4\kappa-1}{4(\kappa-1)} < 2$, these checkpoints also satisfy $b > 16(\frac{4\kappa-1}{4(\kappa-1)})$. Ordinary factors are $c(a_i)(a_i+s_i)/b_i$; exceptional factors are bounded by $(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b_i$. The leading gap is uncut. Let s_{gap} be the originally unselected step sum. Repair only selects additional positions, so the sum of the final gap step sums is at most s_{gap}. By Lem. A.15, $s_{\mathsf{gap}} \le (4\kappa-1)\lceil n/q\rceil/256$, and every originally unselected step is at most $(4\kappa-1)/256$.

At ordinary checkpoints, the retained startup parameters satisfy $a_{\min} \le a_i \le b_i$ and $a_i \le (4\kappa-1)/8$. For a nonleading gap, its damping is at least $1-\mu_{\mathsf{c}}s_i$, which exceeds every assigned input threshold: $15/16 > 1/2$. For the leading gap, the requested input threshold is zero and $Q_1 > 0$, since each of its steps is below $(4\kappa-1)/4$. Thus, the ordinary transfers satisfy the hypotheses of Lem. B.3; use Lem. B.4 for the remaining checkpoints. Mark every ordinary successor of an exceptional transfer with input threshold $1/2$. Ordinary-to-ordinary input thresholds are τ_{i-1}. These fixed local transfers obey both compatibility comparisons in Def. B.1. More generally, consider an ordinary run $i_0, \dots, i_1$ with $1 \le i_0 \le i_1 \le d_j$, including $i_0 = 1$ when the leading gap is short. Assume

$$a_{\min} \le a_{\zeta_\ell} \le \min\{b_\ell, (4\kappa-1)/4\}, \qquad s_\ell \le (4\kappa-1)/64 \quad (i_0 \le \ell \le i_1),$$

with first input threshold $0 \le \tau_{\mathsf{in},i_0} \le 1/2$ and subsequent input thresholds $\tau_{\mathsf{in},\ell} = \tau_{\zeta_{\ell-1}}$ for $i_0 < \ell \le i_1$. These assumptions give $Q_\ell \ge 15/16 > \tau_{\mathsf{in},\ell}$ throughout the run. Multiplying (B.10) gives

$$\begin{aligned}\prod_{\ell=i_0}^{i_1} B_\ell^{-1} &\le \frac{1-\tau_{\zeta_{i_1}}}{1-\tau_{\mathsf{in},i_0}} \exp\left(\frac{12}{4\kappa-1}\sum_{\ell=i_0}^{i_1} s_\ell\right) \prod_{\ell=i_0}^{i_1} \frac{c(a_{\zeta_\ell})(a_{\zeta_\ell}+s_\ell)}{b_\ell} \\ &\le 2\exp\left(\frac{12}{4\kappa-1}\sum_{\ell=i_0}^{i_1} s_\ell\right) \prod_{\ell=i_0}^{i_1} \frac{c(a_{\zeta_\ell})(a_{\zeta_\ell}+s_\ell)}{b_\ell}.\end{aligned} \tag{B.23}$$

The factors $(1-\tau_i)/(1-\tau_{\mathsf{in},i})$ cancel at every interior boundary, leaving $(1-\tau_{i_1})/(1-\tau_{\mathsf{in},i_0}) \le 2$. The hypotheses $s_i \le (4\kappa-1)/64$ and $\tau_{\mathsf{in},i} \le 1/2$ give the damping bound in Lem. B.3, proving Eq. (B.23). Stopping the multiplication earlier gives the same bound at every prefix. Since $D_{i+1}/D_{i_0} = \prod_{\ell=i_0}^{i} B_\ell$ with $B_\ell > 0$, taking reciprocals gives, for every $i_0 \le i \le i_1$,

$$\frac{D_{i+1}}{D_{i_0}} \ge \frac{1}{2}\exp\left(-\frac{12}{4\kappa-1}\sum_{\ell=i_0}^{i} s_\ell\right) \prod_{\ell=i_0}^{i} \frac{b_\ell}{c(a_{\zeta_\ell})(a_{\zeta_\ell}+s_\ell)}.$$

We apply this bound to nonleading ordinary runs here; a leading transfer with unrestricted gap sum is treated separately using (B.11). Input thresholds are never reset at numerical-window boundaries.

Fig. 9 displays the threshold assignments and the cancellation inside each ordinary run. An exceptional checkpoint changes these assignments but remains a transfer in the same phase and coordinate block.

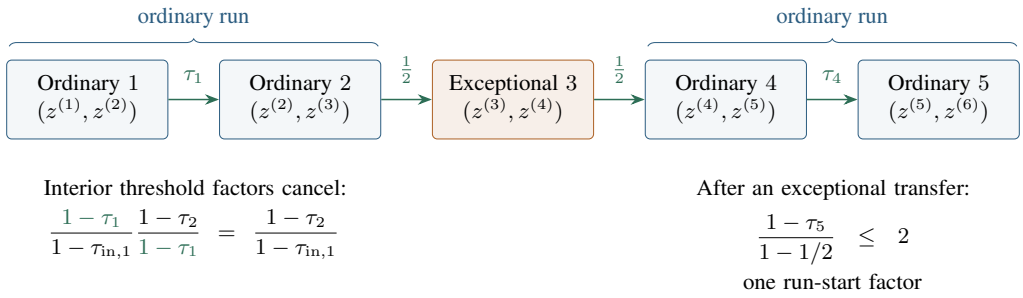


Figure 9: Input-threshold assignments used in Lem. B.7. Threshold factors telescope within each ordinary run (Eq. (B.23)), leaving a factor at most two per run; exceptional transfers separate these runs.

The leading gap has

$$Q_1^{-1} = \prod_{\text{first gap}} (1-\mu_{\mathsf{c}}\eta)^{-1} \le e^{2\mu_{\mathsf{c}} s_{\mathsf{gap}}} \le e^{\lceil n/q\rceil/32}, \tag{B.24}$$

since $\mu_{\mathsf{c}}\eta \le 1/64$ and $-\log(1-x) \le 2x$ for $0 \le x \le 1/2$. This estimate depends only on the prescribed gap steps, before any controlled trial is matched.

Step 2: isolate the trial-dependent transfer. The first transfer is chosen once and for all by its unforced shooting construction. Define $\boldsymbol{g}_*$ by its first two components $\mu_{\mathsf{c}}(1,0) + \beta_{\mathsf{c}}\nabla g_1(1,0)$ and zeros thereafter; thus $g_*^{(2)} \le 0$. This definition uses only the fixed first transfer, not a trajectory of the whole core. If at least two checkpoints exist and $b_1 > (4\kappa-1)/2$, the first transfer is ordinary by the selected checkpoint rule. Replace the second transfer by Lem. B.5, using Eq. (B.15), and mark its successor to use $1/2$. When that successor is ordinary, choose its transfer once with this input threshold and its unchanged startup parameter; keep it fixed over the controlled rectangle. The replacement has a reciprocal formal transfer factor of at most $8(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b_2$ uniformly over its rectangle. The replaced numerical factor is at least $a_{\min}/b_2$ when ordinary and $(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b_2$ when exceptional; in either case it is at least $(\frac{4\kappa-1}{4(\kappa-1)})/b_2$. Replacing that factor by $8(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b_2$ and the at most one adjacent run boundary will cost only one polynomial factor, as quantified below. The fixed first transfer factor is B_1.

Define the next formal transfer amplitude as $D_3^{\mathsf{trial}} = B_1 B_2^{\mathsf{trial}}$, and define every later amplitude by multiplying the preceding amplitude by the next fixed transfer factor. Here the local transfer functions after the second transfer are fixed; only their placement scales change with D_3^{trial}. The controlled transfer factor is positive and continuous on the rectangle; hence so is every formal transfer amplitude. These amplitudes will be identified with the actual checkpoint amplitudes only after matching. If no replacement is needed, use a singleton family. In either case, the chain uses $d_j + 1 \le n+1$ coordinates.

By Lem. A.15, each numerical prefix product is at most $C_{\mathsf{sel}}(\kappa-1)^{C_{\mathsf{sel,amp}}} e^{C_{\mathsf{sel,win}}\lceil n/q\rceil}$ for the absolute constants in that lemma. The $O(\lceil n/q\rceil)$ exceptional checkpoints create $O(\lceil n/q\rceil)$ ordinary runs, whose start factors cost $2^{O(\lceil n/q\rceil)}$ by (B.23). All gap damping, including the leading gap by (B.24), costs at most $\exp(5\mu_{\mathsf{c}} s_{\mathsf{gap}}) \le e^{5\lceil n/q\rceil/64}$. A controlled replacement costs at most $8(\kappa-1)^3$ and creates at most one extra run boundary: its reciprocal formal factor is at most $8(\frac{4\kappa-1}{4(\kappa-1)})(\kappa-1)^3/b$, whereas the replaced numerical factor is at least $(\frac{4\kappa-1}{4(\kappa-1)})/b$. Consequently (B.22) holds for absolute $C_{D,\mathsf{pow}}, C_{D,\mathsf{exp}}, C_D > 0$, uniformly over all formal transfer amplitudes and trials.

□

One common lower bound. We choose one scale small enough for two different conclusions. For every trial it must fit inside all scaled zero regions of the transfer functions, so the core is purely quadratic near zero. For a matched trial it must also lie below the active-coordinate bounds during gaps and the exact outputs at checkpoints. The polynomial margin in the following choice pays for the local thresholds in both comparisons. Fix absolute $C_{\mathsf{amp}} > C_{D,\mathsf{pow}} + 2$ and $C_{\mathsf{win}} \geq \max\{C_{D,\mathsf{exp}}, 2\}$, and enlarge the absolute lower threshold on κ to absorb C_D. Define

$$\mathsf{amp}(n) = (\kappa - 1)^{-C_{\mathsf{amp}}} e^{-C_{\mathsf{win}}\lceil n/q \rceil}. \tag{B.25}$$

These choices make $D_i^{\mathsf{trial}}(\kappa-1)^{-2} \geq \mathsf{amp}(n)$ and $Q_1(\kappa-1)^{-2} \geq \mathsf{amp}(n)$ for every trial. We use the same constants for the common quadratic ball and the matched trajectory bounds.

Lemma B.8 (Static family of phase cores)**.** *For every sufficiently large κ and every phase of length n as defined in Appx. B.5, the local transfers of Lem. B.7 give cores $h_{\boldsymbol{p}}$ of the form (B.21) on one fixed coordinate space of dimension $d_j + 1 \leq n + 1$. Their parameter domain Trials is nonempty, compact, and convex, and their checkpoint set includes the last step and is independent of $\boldsymbol{p}$. For every $\boldsymbol{p} \in \mathsf{Trials}$,*

$$h_{\boldsymbol{p}} \in \mathcal{F}_{4/(4\kappa-1),\, 4(\kappa-1)/(4\kappa-1)}, \qquad h_{\boldsymbol{p}}(\boldsymbol{z}) = \frac{2\|\boldsymbol{z}\|^2}{4\kappa - 1} \quad (\|\boldsymbol{z}\| \leq \mathsf{amp}(n)),$$

and the minimizer is zero. The nonzero initial gradient $\nabla h_{\boldsymbol{p}}(\boldsymbol{e}^{(1)}) = \boldsymbol{g}_*$ is independent of $\boldsymbol{p}$, is supported on coordinates one and two, and satisfies $g_*^{(2)} \leq 0$. All transfer scales are positive and continuous, and $(\boldsymbol{z}, \boldsymbol{p}) \mapsto \nabla h_{\boldsymbol{p}}(\boldsymbol{z})$ is jointly continuous. In the controlled case, write $n_{\mathsf{gap}} = t_2 - t_1 - 1$. The family supplies the jointly continuous map $\mathsf{Shoot} : \mathsf{Trials} \times [0, u_{\max}]^{n_{\mathsf{gap}}} \to \mathsf{Trials}$ of Lem. B.5, the fixed normalization $u_{\max} = |g_*^{(2)}|/B_1$, and the fixed second-gap state indices $t_1, \ldots, t_2 - 2$. For each fixed control vector, this map is a self-map of the trial domain. No claim about an unmatched trial's actual phase trajectory is needed for these static properties.

Proof. **Curvature and minimizer.** The scaled transfer function $D_i^2 g_i(z^{(i)}/D_i, z^{(i+1)}/D_i)$ is convex and 1-smooth on its pair. For a displacement $\mathbf{dir}$, $\sum_i((\mathsf{dir}^{(i)})^2 + (\mathsf{dir}^{(i+1)})^2) \leq 2\|\mathbf{dir}\|^2$. The quadratic gives lower curvature $\mu_{\mathsf{c}} = 4/(4\kappa-1)$, while the upper curvature is $\mu_{\mathsf{c}} + 2\beta_{\mathsf{c}} = 4(\kappa - 1)/(4\kappa - 1)$. Each transfer function is nonnegative and zero at zero. Thus, the quadratic term also fixes the minimizer at zero.

A common quadratic ball for all trials. We need a radius valid before matching, because the serial coupling will be chosen from this radius. We therefore use the formal transfer amplitudes from Lem. B.7 and the strip bounds valid for every controlled trial, rather than properties of a matched trajectory. The ordinary startup cap gives (B.14), including $\tau_{\zeta(a)} > \sigma_\kappa$. For an ordinary transfer with an ordinary predecessor, its chosen input threshold and (B.7) give

$$\mathsf{rad}_{\mathsf{loc},i} \geq (\zeta_i/2)\upsilon_i w_i^{(1)} \geq \zeta_i \tau_{i-1}/2 > 1/[200(\log(\kappa-1))^3].$$

An assigned input threshold of $1/2$ is stronger. Exceptional and controlled trials have fixed pitch and a common zero strip with input threshold $1/2$. Lems. B.4 & B.5 give local radii at least $(\kappa-1)^{-2}$ for sufficiently large κ. The first transfer has $\upsilon_1 w_1^{(1)} \geq Q_1/(2(\kappa-1))$; for an ordinary pitch use Eq. (B.14), and for a sole exceptional transfer use its fixed pitch. In either case, its zero radius is at least $Q_1(\kappa-1)^{-2}$ for sufficiently large κ.

Scaling a pair by D_i gives a zero ball of radius $D_i\mathsf{rad}_{\mathsf{loc},i}$ in the full space: restricting a vector to a coordinate pair never increases its norm. Thus, there is no dimension factor. The uniform all-prefix transfer-amplitude estimate and (B.24) prove the common radius (B.25) with the absolute $C_{\mathsf{amp}}, C_{\mathsf{win}}$ chosen above. Every transfer function vanishes there, leaving exactly $2\|\boldsymbol{z}\|^2/(4\kappa-1)$.

Common initial gradient and continuity. On input $\boldsymbol{e}^{(1)}$, every transfer after the first sees $(0,0)$ and has zero gradient there, since it is differentiable, nonnegative, and vanishes there. Thus, $\nabla h_{\boldsymbol{p}}(\boldsymbol{e}^{(1)})$ is exactly the fixed vector $\boldsymbol{g}_*$ defined in Lem. B.7, is supported on the first two coordinates, and has $g_*^{(2)} \leq 0$; strong convexity makes it nonzero. Every later formal transfer amplitude is positive and continuous on the compact trial domain, and the scaled projection formulas are jointly continuous in position and parameters. Positivity on this compact domain keeps the coordinate rescalings away from division by zero. This proves joint continuity of the entire gradient and uniform positive transfer

amplitudes for this fixed finite phase. The control normalization in Eq. (B.15) depends only on the fixed first transfer, and (B.20) supplies the required jointly continuous map, which is a self-map for each fixed control vector.

□

The static family now permits the coupling to be fixed, but its actual trajectory still needs to be identified. The incoming correction acts on the first two coordinates, so we first keep it zero while the unforced transfer bounds apply and then match the controlled second transfer if necessary. To state this argument independently of a coupling formula, let $r_t \geq 0$ be a readout of the earlier coordinate blocks, let $\ell_{\mathsf{in}} > 0$ be a release threshold, and let the core receive the correction $-\theta_t \boldsymbol{g}_*$, $0 \leq \theta_t \leq 1$, where $\boldsymbol{g}_*$ is its initial gradient. Assume the following readout and projection conditions whenever needed:

$$\begin{aligned} r_t \leq \ell_{\mathsf{in}}, \quad |\langle \boldsymbol{g}_*/\|\boldsymbol{g}_*\|, \boldsymbol{z}_t\rangle| \leq 8(\kappa-1) \quad &\Longrightarrow \quad \theta_t = 0, \\ r_t \leq \ell_{\mathsf{in}}, \quad \eta_{t+1} \leq (4\kappa-1)/2 \quad &\Longrightarrow \quad r_{t+1} \leq \ell_{\mathsf{in}}. \end{aligned} \tag{B.26}$$

Here, θ_t is the *correction coefficient*, distinct from the vector correction $-\theta_t \boldsymbol{g}_*$. The index t denotes the current state, so the second line concerns step η_{t+1}, which produces the next readout. These are assumptions for Lem. B.9; the actual serial coupling will satisfy them by Lems. C.1 & C.3. The first implication switches off the correction when the earlier blocks and the new block's projection are both small enough. The second propagates the readout condition across the steps for which this invariance is valid. If a larger first checkpoint breaks that invariance, matching the controlled second transfer replaces the need to propagate it further. The first core uses no incoming correction and needs no readout assumptions.

Lemma B.9 (Realization of a matched phase)**.** *Fix the phase family in Lem. B.8 and a trial $\boldsymbol{p} \in \mathsf{Trials}$. Suppose $\boldsymbol{z}_0 = \boldsymbol{e}^{(1)}$ and the actual updates are*

$$\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \eta_{t+1}\big(\nabla h_{\boldsymbol{p}}(\boldsymbol{z}_t) - \theta_t \boldsymbol{g}_*\big), \qquad 0 \leq \theta_t \leq 1 \quad (0 \leq t < n).$$

Thus the only gradient correction is $-\theta_t \boldsymbol{g}_*$. Assume $r_0 \leq \ell_{\mathsf{in}}$ and the readout and projection conditions (B.26). If the second transfer is controlled, also assume

$$\boldsymbol{p} = \mathsf{Shoot}(\boldsymbol{p}, \boldsymbol{u}), \qquad u_r = u_{\max}\theta_{t_1+r-1} \quad (1 \leq r \leq n_{\mathsf{gap}} := t_2 - t_1 - 1).$$

The empty second gap uses the empty control vector; the checkpoint coefficient θ_{t_2-1} is not part of that vector. Then each checkpoint output is exactly $z_{t_i}^{(i+1)} = D_{i+1}^{\mathsf{trial}}$; every iterate with $0 \leq t \leq n$ has norm at least $\mathsf{amp}(n)$, and the same lower bound holds for the magnitude of every transfer amplitude. The pre-update states in the first gap, including the state before its checkpoint, satisfy the $8(\kappa-1)$ projection condition and have zero incoming correction. If a second transfer exists and $b_1 \leq (4\kappa-1)/2$, the same holds in its gap and before its checkpoint update. For larger b_1, that second transfer is controlled; after its launch into coordinate three, the incoming correction touches only retired coordinates and no readout invariant is needed. When $\theta_t \equiv 0$, the readout hypotheses can be omitted; the matching condition is still required when a controlled transfer is installed.

Proof. **Step 1: identify the first two pairs under actual incoming correction.** The argument using these conditions is an induction on iterates: agreement up to an iterate gives its coordinate bound, that bound makes the correction zero, and the resulting update gives agreement at the next iterate. The readout invariant is propagated only across steps at most $(4\kappa-1)/2$.

During the formal first gap and before its checkpoint update, $0 < z^{(1)} \leq 1$ and $0 \leq z^{(2)} \leq \upsilon_1 w_1^{(2)} \leq \zeta/(1+2\zeta^2) < 1$, since $\upsilon_1 w_1^{(1)} \leq 1$ and $w_1^{(2)}/w_1^{(1)} \leq \zeta/(1+2\zeta^2)$. Thus, the first two coordinates satisfy the $8(\kappa-1)$ projection condition. All steps producing these iterates are gap steps less than $(4\kappa-1)/4 < (4\kappa-1)/2$. Starting with $r_0 \leq \ell_{\mathsf{in}}$, induction using Eq. (B.26) keeps the readout below threshold and makes $\theta_t = 0$ at every such iterate. This identifies the actual first-gap updates with the formal unforced ones. The first checkpoint may be arbitrarily large: the iterate at which its gradient is evaluated has already been identified before that update is taken. If it is the sole checkpoint, its exact output ends the phase.

Suppose a second transfer exists. If $b_1 \leq (4\kappa-1)/2$, that checkpoint preserves the readout threshold. Since $\upsilon_1 w_1^{(1)} \leq 1$ and $w_1^{(1)} \geq w_{\min}(\zeta) \geq \mathsf{rad}_{\zeta,0}$, we have $\beta_{\mathsf{c}} \upsilon_1 \mathsf{rad}_{\zeta,0} \leq \mu_{\mathsf{c}} + 2\beta_{\mathsf{c}} =$

$4(\kappa-1)/(4\kappa-1)$. The checkpoint update therefore gives

$$|z^{(1)}| \le 1+\mu_c b_1+(\frac{4(\kappa-1)}{4\kappa-1})b_1 \le 3+2(\kappa-1), \qquad 0<B_1 \le 1+(\frac{4(\kappa-1)}{4\kappa-1})b_1 \le 2\kappa-1.$$

During the second gap, the first coordinate is retired; write $s_{\max}=(4\kappa-1)/64$. Its incoming correction is zero as long as the projection condition holds, and its transfer gradient is bounded by $4(\kappa-1)/(4\kappa-1)$. The damping factors are nonnegative, so its absolute value can increase by at most $[4(\kappa-1)/(4\kappa-1)]s_{\max} \le (\kappa-1)/16$. The active second coordinate stays between zero and B_1. Their Euclidean norm is at most $|z^{(1)}|+|z^{(2)}| \le 4(\kappa-1)+4+(\kappa-1)/16 < 8(\kappa-1)$. Only these two coordinates enter the projection, since $\boldsymbol{g}_*$ is supported on them; the third coordinate need not satisfy this bound. Every second-gap step is at most $s_{\max}=(4\kappa-1)/64<(4\kappa-1)/2$. The same induction on iterates therefore keeps the incoming correction zero through the iterate before the second checkpoint update. The ordinary or exceptional second pair is therefore identified exactly.

If $b_1>(4\kappa-1)/2$, no readout invariance is claimed after it. The second-pair correction is exactly (B.16) with trial-independent $u_{\max}$ from Eq. (B.15). The first coordinate of this pair is core coordinate two, so the correction is horizontal in the local transfer coordinates. At matching, Lem. B.5 identifies its actual gap and its exact launch into coordinate three. The control at its checkpoint can be arbitrary.

At matching, Lem. B.5 bounds the ratio of the gap-end second coordinate to the output by $\sigma_\kappa<1/2$ and the ratio of the derivative threshold to the output by $1/2$. The successor's input threshold is $1/2$, so both threshold comparisons required for compatibility hold.

In either case, after the second launch, the incoming correction touches only retired coordinates one and two. The preceding transfer function's derivative in the shared active coordinate vanishes regardless of the retired coordinates, so the induction in Lem. B.6 identifies every later pair without assumptions on the earlier blocks. We do not need to conclude that all earlier transfer functions become identically zero.

Step 2: obtain the phase lower bound. It remains to cover the states between checkpoints as well as the checkpoint outputs. Every ordinary gap after the first has an active coordinate above $\tau_{\mathsf{in},i}D_i$, where $\tau_{\mathsf{in},i} \ge \sigma_\kappa$ by Eq. (B.14), unless it has the stronger input threshold $1/2$. Exceptional and controlled gaps have input threshold $1/2$. The first gap has first coordinate at least $Q_1/(2(\kappa-1))$, and every checkpoint produces the exact next amplitude. By Lem. B.7 and the choice of $C_{\mathsf{amp}}, C_{\mathsf{win}}$, the transfer-amplitude and damping estimates therefore bound every iterate norm, endpoint norm, and transfer amplitude below by (B.25). This last step concerns matched trajectories; we proved the common quadratic ball and continuity conclusions for all trials independently. □

An uncoupled phase. Taking $\theta_t=0$ in Lem. B.9 gives the unforced phase bound for any matching trial. If the family has a controlled second transfer, its zero-control matching equation is $\boldsymbol{p}=\mathsf{Shoot}(\boldsymbol{p},0)$; the compact-domain fixed-point step is separate from the static family and the realization implication.

C Assembling the Hard Instance

A phase core handles the schedule segment assigned to it, but a long small-step run may exhaust its amplitude. We attach another core in a fresh coordinate block and hold it at its initial point until release. The construction must accomplish three tasks together: preserve every earlier certified trajectory, realize the new phase under its actual incoming correction, and keep the total curvature within the target bounds. We first define the scaled cores, weighted readouts, and couplings using only properties shared by every trial. Cancellation then holds the new block without changing the earlier gradients, while a fresh readout crossing determines its release state. We match the new core to the correction produced by its full trial run and apply the phase realization result. Repeating this extension gives the finite serial certificate, whose scalar losses are summed in Appx. D.

C.1 Core and coupling definitions

Fix sufficiently large κ and the prescribed schedule $\boldsymbol{\eta}_{1:N}$. Each core already reserves curvature for the couplings: its lower and upper curvature bounds are $4/(4\kappa-1)$ and $4(\kappa-1)/(4\kappa-1)$. The coupling coefficients below ensure that their total negative curvature has magnitude at most $1/[4(\kappa-1)(4\kappa-1)]$ and their total positive curvature is at most $3/(4\kappa-1)$. Consequently, the complete objective satisfies

$$\left[\frac{16\kappa-17}{4(\kappa-1)(4\kappa-1)},\,1\right]\subseteq[1/\kappa,1], \tag{C.1}$$

and hence

$$f\in\mathcal{F}_{1/\kappa,1}. \tag{C.2}$$

Lem. C.2 proves these bounds uniformly over every core trial. The schedule remains $\boldsymbol{\eta}$ throughout the construction, including all coupling corrections.

We use global iteration indices in this section. The iterate at index t is $\boldsymbol{x}_t$; its coordinate block j is $\boldsymbol{x}_t^{[j]}$. Fix the absolute constants in Lem. B.8 before any schedule-dependent choices, enlarging them so $0<\mathsf{amp}(n)\le 1$.

Exact cancellation holds each future block at its initial point until release. Installing and matching it preserves the earlier trajectory; weighted readouts keep the total coupling curvature bounded independently of the number of blocks.

C.1.1 The objective and its parameters

A readout is a nonnegative scalar measuring the size of earlier coordinate blocks. Its release band has an upper level above which the next coordinate block remains at its initial point and a lower level at which the incoming correction vanishes at that point. The readout weights must also keep the sum of all coupling curvatures bounded independently of the number of phases. Let $P\ge 0$ be an integer indexing the last of $P+1$ phases, and let $\boldsymbol{x}=(\boldsymbol{x}^{[0]},\dots,\boldsymbol{x}^{[P]})$ be the orthogonal concatenation of their coordinate blocks. Coordinate block j receives a phase of length $n^{[j]}$ and an unscaled core $h^{[j]}$ from Lem. B.8. Its phase-start index $t_{\mathsf{start}}^{[j]}$ is the state index immediately before its first phase update. Thus, local state index t_{loc} corresponds to global index $t_{\mathsf{start}}^{[j]}+t_{\mathsf{loc}}$; for $0\le t_{\mathsf{loc}}<n^{[j]}$, its next phase update uses the prescribed step $\eta_{t_{\mathsf{start}}^{[j]}+t_{\mathsf{loc}}+1}$. Choose each block's initial point $\boldsymbol{x}_{\mathsf{init}}^{[j]}\in\{A\boldsymbol{e}^{(1)}:A>0\}$, with $\boldsymbol{x}_{\mathsf{init}}^{[0]}=\boldsymbol{e}^{(1)}$. Recall from (4.1) that $A^{[j]}=\|\boldsymbol{x}_{\mathsf{init}}^{[j]}\|>0$ and $A^{[0]}=1$; thus $\boldsymbol{x}_{\mathsf{init}}^{[j]}=A^{[j]}\boldsymbol{e}^{(1)}$. Put

$$F^{[j]}(\boldsymbol{x}^{[j]})=(A^{[j]})^2h^{[j]}(\boldsymbol{x}^{[j]}/A^{[j]}),\quad \mathsf{rad}^{[j]}=\frac{A^{[j]}\mathsf{amp}(n^{[j]})}{4}. \tag{C.3}$$

We apply the unscaled core lemma to the normalized block

$$\widehat{\boldsymbol{z}}_{t_{\mathsf{loc}}}:=\frac{\boldsymbol{x}^{[j]}_{t_{\mathsf{start}}^{[j]}+t_{\mathsf{loc}}}}{A^{[j]}},\qquad 0\le t_{\mathsf{loc}}\le n^{[j]}.$$

The variable $\boldsymbol{z}$ in Appx. B.5 is the argument of this unscaled core; in the full objective that argument is $\widehat{\boldsymbol{z}}_{t_{\mathsf{loc}}}$. Spatial normalization and local indexing leave the prescribed step sizes unchanged. The

ball of radius $\mathsf{rad}^{[j]}$ lies inside the common quadratic ball of the trials of $F^{[j]}$. Thus, $F^{[j]} \in \mathcal{F}_{4/(4\kappa-1),\,4(\kappa-1)/(4\kappa-1)}$ and $F^{[j]}(\boldsymbol{x}^{[j]}) = 2\|\boldsymbol{x}^{[j]}\|^2/(4\kappa-1)$ when $\|\boldsymbol{x}^{[j]}\| \le \mathsf{rad}^{[j]}$. Define the weighted linear map $\mathsf{Weight}^{[j]}$ and scalar readout $r^{[j]}$ on the concatenation of blocks $0,\dots,j$ by

$$\mathsf{Weight}^{[0]}\boldsymbol{x}^{[0:0]} = \boldsymbol{x}^{[0]}, \qquad \mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]} = \left(\boldsymbol{x}^{[j]}, \frac{4\mathsf{rad}^{[j]}}{\mathsf{rad}^{[j-1]}}\mathsf{Weight}^{[j-1]}\boldsymbol{x}^{[0:j-1]}\right),$$
$$\ell^{[j]} = \frac{\mathsf{rad}^{[j]}}{\sqrt{8(\kappa-1)+1}}, \tag{C.4}$$
$$r^{[j]}(\boldsymbol{x}^{[0:j]}) = \sqrt{\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\|^2 + (4(\kappa-1)\ell^{[j]})^2} - 4(\kappa-1)\ell^{[j]}.$$

In particular, $r^{[j]}(\boldsymbol{x}^{[0:j]}) = \ell^{[j]}$ exactly when $\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\| = \mathsf{rad}^{[j]}$.

Fig. 10 shows how the earlier coordinate blocks enter one weighted readout and thereby control the next coupling. The weights act on coordinates, not on the values of the core functions.

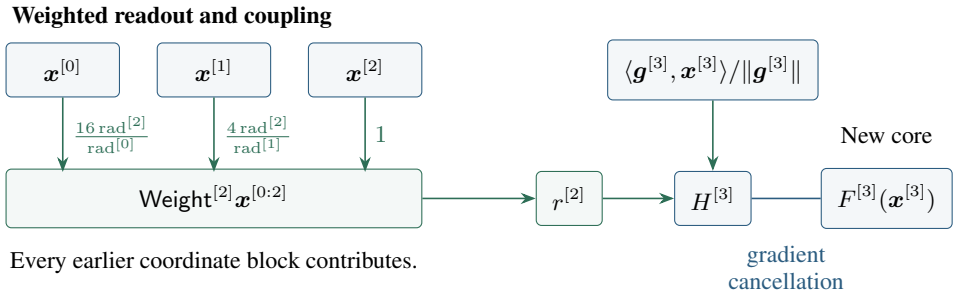


Figure 10: The readout $r^{[2]}(\boldsymbol{x}^{[0:2]})$ measures the weighted size of blocks $0, 1, 2$ (Eq. (C.4)). The coupling $H^{[3]}$ also reads block 3's coordinate along its common initial gradient $\boldsymbol{g}^{[3]}$ (Eq. (C.7)). When $\boldsymbol{x}^{[3]} = \boldsymbol{x}^{[3]}_{\mathsf{init}}$ and $r^{[2]} \ge \ell_k^+$, its gradient cancels the new core's gradient, and its correction to earlier blocks is zero (Lem. C.1).

A transition $j \to j+1$ chooses an integer band index $k \ge 1$. At the level $\ell^{[j]}$, set

$$\ell_k^+ = \frac{\ell^{[j]}}{2} + \frac{\ell^{[j]}}{2k}, \quad \ell_k^- = \frac{\ell^{[j]}}{2} + \frac{\ell^{[j]}}{2(k+1)}. \tag{C.5}$$

Choose

$$\boldsymbol{x}^{[j+1]}_{\mathsf{init}} = \frac{\ell_k^+ - \ell_k^-}{2(16(\kappa-1)+1)}\boldsymbol{e}^{(1)}, \qquad \boldsymbol{g}^{[j+1]} = \nabla F^{[j+1]}(\boldsymbol{x}^{[j+1]}_{\mathsf{init}}). \tag{C.6}$$

The offset $\ell_k^- + 8(\kappa-1)A^{[j+1]}$ fixes the release location. The gradient is trial-independent by Lem. B.8; the common zero minimizer, strong convexity, and smoothness give $0 < \|\boldsymbol{g}^{[j+1]}\| \le [4(\kappa-1)/(4\kappa-1)]A^{[j+1]}$. Let

$$\mathsf{ramp}(x) = \begin{cases} 0 & x \le 0, \\ x^2/4 & 0 < x < 2, \\ x-1 & x \ge 2. \end{cases}$$

At the initial point, the saturated ramp will cancel the new core gradient. To leave earlier gradients unchanged in the same regime, we subtract its linear dependence on the readout. This gives a coupling function $H^{[j+1]}$ on the concatenation of blocks $0,\dots,j+1$:

$$H^{[j+1]}(\boldsymbol{x}^{[0:j+1]}) = (4\kappa-1)\|\boldsymbol{g}^{[j+1]}\|^2 \times \mathsf{ramp}\left(\frac{r^{[j]}(\boldsymbol{x}^{[0:j]}) - \langle \boldsymbol{g}^{[j+1]}, \boldsymbol{x}^{[j+1]}\rangle/\|\boldsymbol{g}^{[j+1]}\| - \left(\ell_k^- + 8(\kappa-1)A^{[j+1]}\right)}{(4\kappa-1)\|\boldsymbol{g}^{[j+1]}\|}\right) - \|\boldsymbol{g}^{[j+1]}\| r^{[j]}(\boldsymbol{x}^{[0:j]}). \tag{C.7}$$

All coupling functions and their parameters are fixed before the run and remain unchanged during gradient descent. Our objective and initial point are

$$f(\boldsymbol{x}) = \sum_{j=0}^{P} F^{[j]}(\boldsymbol{x}^{[j]}) + \sum_{j=1}^{P} H^{[j]}(\boldsymbol{x}^{[0:j]}),$$
$$\boldsymbol{x}_0 = \boldsymbol{x}_{\mathsf{init}} = \boldsymbol{x}_{\mathsf{init}}^{[0:P]} = (\boldsymbol{x}_{\mathsf{init}}^{[0]}, \dots, \boldsymbol{x}_{\mathsf{init}}^{[P]}). \tag{C.8}$$

Each $H^{[j]}(\boldsymbol{x}^{[0:j]})$ depends on $\boldsymbol{x}^{[0:j-1]}$ solely through $r^{[j-1]}(\boldsymbol{x}^{[0:j-1]})$ and on $\boldsymbol{x}^{[j]}$ solely through $\langle \boldsymbol{g}^{[j]}, \boldsymbol{x}^{[j]} \rangle$. Thus, (C.8) is a serial sum with explicitly weighted overlaps with earlier coordinate blocks.

Writing the correction coefficient $\theta^{[j+1]}(\boldsymbol{x}^{[0:j+1]})$ for ramp' at the argument in Eq. (C.7), differentiation gives

$$\nabla_{\boldsymbol{x}^{[j+1]}} H^{[j+1]}(\boldsymbol{x}^{[0:j+1]}) = -\theta^{[j+1]} \boldsymbol{g}^{[j+1]},$$
$$\nabla_{\boldsymbol{x}^{[0:j]}} H^{[j+1]}(\boldsymbol{x}^{[0:j+1]}) = -\|\boldsymbol{g}^{[j+1]}\|(1-\theta^{[j+1]})\nabla r^{[j]}(\boldsymbol{x}^{[0:j]}). \tag{C.9}$$

Here and below, we omit the argument of $\theta^{[j+1]}$ when it is clear from context. When $\theta^{[j+1]} = 1$ and $\boldsymbol{x}^{[j+1]} = \boldsymbol{x}_{\mathsf{init}}^{[j+1]}$, the first term cancels the new core gradient exactly and the second vanishes. The chosen parameters guarantee saturation at the new block's initial point whenever $r^{[j]} \geq \ell_k^+$, and zero incoming correction whenever

$$r^{[j]} \leq \ell_k^-, \qquad |\langle \boldsymbol{g}^{[j+1]}/\|\boldsymbol{g}^{[j+1]}\|, \boldsymbol{x}^{[j+1]} \rangle| \leq 8(\kappa-1)A^{[j+1]}. \tag{C.10}$$

C.2 Cancellation and release

The coupling has to do more than cancel the new core's initial gradient. It must simultaneously leave all earlier gradients unchanged, so appending a future block preserves the already constructed trajectory. The two identities in (C.9) achieve both requirements in the saturated regime. At a low readout, the incoming correction instead vanishes under the projection condition (C.10); this is the condition used to start the phase realization argument.

Lemma C.1 (Cancellation and release)**.** *For a transition $j \to j+1$, fix a trial-independent vector $\boldsymbol{g}^{[j+1]}$ with $0 < \|\boldsymbol{g}^{[j+1]}\| \leq [4(\kappa-1)/(4\kappa-1)]A^{[j+1]}$, and use the band and scale in (C.5)–(C.6). Then the coupling (C.7) obeys (C.9). Its coefficient $\theta^{[j+1]}$ is continuous and belongs to $[0,1]$ at every state. Its parameters are independent of the new core's shooting trial. At $\boldsymbol{x}^{[j+1]} = \boldsymbol{x}_{\mathsf{init}}^{[j+1]}$, it is saturated when $r^{[j]}(\boldsymbol{x}^{[0:j]}) \geq \ell_k^+$. Its correction to the new coordinate block is zero whenever (C.10) holds. Moreover,*

$$A^{[j+1]} \leq \frac{1}{8(\kappa-1)}\ell^{[j]}, \qquad \|\boldsymbol{g}^{[j+1]}\|/\ell^{[j]} \leq 1/[2(4\kappa-1)]. \tag{C.11}$$

Proof. The two required regimes correspond to the two constant parts of ramp': saturation requires a ramp argument at least two, and release requires an argument at most zero. We check these conditions using the same trial-independent initial gradient. The derivative $\mathsf{ramp}'(x) = \min\{1, \max\{0, x/2\}\}$ is continuous, nondecreasing, and $1/2$-Lipschitz on the whole real line, including its two joins. Differentiation gives the new-coordinate correction $-\theta^{[j+1]}\boldsymbol{g}^{[j+1]}$ and the old-coordinate correction $-\|\boldsymbol{g}^{[j+1]}\|(1-\theta^{[j+1]})\nabla r^{[j]}$ in Eq. (C.9). The assumed gradient bound gives $0 < \|\boldsymbol{g}^{[j+1]}\| \leq [4(\kappa-1)/(4\kappa-1)]A^{[j+1]}$. Consequently,

$$\begin{aligned}\left(\ell_k^- + 8(\kappa-1)A^{[j+1]}\right) + \frac{\langle \boldsymbol{g}^{[j+1]}, \boldsymbol{x}_{\mathsf{init}}^{[j+1]} \rangle}{\|\boldsymbol{g}^{[j+1]}\|} + 2(4\kappa-1)\|\boldsymbol{g}^{[j+1]}\| &\leq \ell_k^- + A^{[j+1]}(16(\kappa-1)+1) \\ &= (\ell_k^- + \ell_k^+)/2 < \ell_k^+.\end{aligned} \tag{C.12}$$

Thus, $r^{[j]} \geq \ell_k^+$ makes the ramp argument greater than two at the new block's initial point. Under Eq. (C.10), the argument's numerator is at most $\ell_k^- + 8(\kappa-1)A^{[j+1]} - (\ell_k^- + 8(\kappa-1)A^{[j+1]}) = 0$. At the new block's initial point, the absolute projection is at most $A^{[j+1]} \leq 8(\kappa-1)A^{[j+1]}$, so the condition holds when $r^{[j]} \leq \ell_k^-$. Finally, $A^{[j+1]} = (\ell_k^+ - \ell_k^-)/[2(16(\kappa-1)+1)] \leq \frac{1}{8(\kappa-1)}(\ell_k^+ - \ell_k^-)/2 \leq \frac{1}{8(\kappa-1)}\ell^{[j]}$. The first budget and the assumed gradient bound give the second budget. □

C.3 Readout weights and initial point

For $0 \le i \le j \le P$, put

$$\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\|^2 = \sum_{i=0}^{j} 16^{j-i}\frac{(\mathsf{rad}^{[j]})^2}{(\mathsf{rad}^{[i]})^2}\|\boldsymbol{x}^{[i]}\|^2. \tag{C.13}$$

The weighted linear maps satisfy

$$\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\| \le \|\boldsymbol{x}\|, \qquad \sum_{j=0}^{P}\|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\|^2 \le 2\|\mathbf{dir}\|^2. \tag{C.14}$$

The main issue is that each coupling reads several earlier blocks. We first show that the squared readout weights form summable geometric sequences. This will bound the total coupling curvature independently of the number of blocks. From $\mathsf{rad}^{[j+1]} \le A^{[j+1]}$, (C.11), and $\ell^{[j]} \le \mathsf{rad}^{[j]}$,

$$(4\mathsf{rad}^{[j+1]}/\mathsf{rad}^{[j]})^2 \le \frac{1}{4(\kappa-1)^2}(\ell^{[j]})^2/(\mathsf{rad}^{[j]})^2 \le \frac{1}{4(\kappa-1)^2} \le 1/4.$$

The coefficient $16^{j-i}(\mathsf{rad}^{[j]})^2/(\mathsf{rad}^{[i]})^2$ is a product of these squared edge weights, or one when $i = j$. Thus, $16^{j-i}(\mathsf{rad}^{[j]})^2/(\mathsf{rad}^{[i]})^2 \le 4^{-(j-i)}$. Formula (C.13) gives the contraction of each $\mathsf{Weight}^{[j]}$. For a fixed i, $\sum_{j=i}^{P} 16^{j-i}(\mathsf{rad}^{[j]})^2/(\mathsf{rad}^{[i]})^2 \le \sum_{l\ge 0} 4^{-l} = 4/3 \le 2$. Summing first over j and then i gives Eq. (C.14).

The readout definition also gives $r^{[j]}(0) = 0$, $\nabla r^{[j]}(0) = 0$, and $r^{[j]}(\boldsymbol{x}^{[0:j]}) \le \|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\|$. It is increasing in the weighted norm and equals $\ell^{[j]}$ exactly at weighted norm $\mathsf{rad}^{[j]}$.

At the initial point, $\|\mathsf{Weight}^{[j]}\boldsymbol{x}_{\mathsf{init}}^{[0:j]}\| \ge A^{[j]} \ge \mathsf{rad}^{[j]}$, so $r^{[j]}(\boldsymbol{x}_{\mathsf{init}}^{[0:j]}) \ge \ell^{[j]} \ge \ell_k^+$. Every coupling is saturated by Eq. (C.12); its correction to earlier coordinates is zero, and its correction to the new coordinate block cancels that block's core gradient. Only the gradient of the first core remains. Finally, $A^{[j+1]} \le \frac{1}{8(\kappa-1)}\ell^{[j]} \le \frac{1}{8(\kappa-1)}A^{[j]}$, and therefore

$$1 \le \|\boldsymbol{x}_{\mathsf{init}}\|^2 = \sum_{j=0}^{P}(A^{[j]})^2 \le \sum_{j\ge 0}(8(\kappa-1))^{-2j} = (1-(8(\kappa-1))^{-2})^{-1} < 2. \tag{C.15}$$

C.4 Global curvature

To estimate the couplings, we compare function values with their first-order approximations. For any differentiable scalar-valued function h, write

$$\mathsf{Rem}_h(\boldsymbol{x}, \mathbf{dir}) := h(\boldsymbol{x}+\mathbf{dir}) - h(\boldsymbol{x}) - \langle\nabla h(\boldsymbol{x}), \mathbf{dir}\rangle.$$

The same notation applies to a real function of one variable.

Lemma C.2 (Global curvature)**.** *Every trial objective in Eq. (C.8) belongs to $\mathcal{F}_{1/\kappa,1}$ and is minimized at zero. For every state $\boldsymbol{x}$ and displacement* **dir***, its couplings satisfy*

$$-\frac{\|\mathbf{dir}\|^2}{4(\kappa-1)(4\kappa-1)} \le 2\sum_{j=0}^{P-1}\mathsf{Rem}_{H^{[j+1]}}(\boldsymbol{x}^{[0:j+1]}, \mathbf{dir}^{[0:j+1]}) \le \frac{3\|\mathbf{dir}\|^2}{4\kappa-1}.$$

Proof. We will bound the remainder above and below by multiples of $\|\mathbf{dir}\|^2$, then add these bounds over the cores and couplings.

The readout is smooth everywhere, with

$$\nabla r^{[j]}(\boldsymbol{x}^{[0:j]}) = \frac{(\mathsf{Weight}^{[j]})^*\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}}{\sqrt{\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\|^2 + (4(\kappa-1)\ell^{[j]})^2}}.$$

Differentiating once more gives the following bounds; Cauchy–Schwarz gives the lower bound, and the denominator is at least $4(\kappa-1)\ell^{[j]}$:

$$0 \le \langle \mathbf{dir}^{[0:j]}, \nabla^2 r^{[j]} \mathbf{dir}^{[0:j]} \rangle \le \frac{\|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\|^2}{4(\kappa-1)\ell^{[j]}},$$
$$|\langle \nabla r^{[j]}, \mathbf{dir}^{[0:j]} \rangle| \le \|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\|.$$

Integrating along $\boldsymbol{x}^{[0:j]} + t\mathbf{dir}^{[0:j]}$, $0 \le t \le 1$, yields

$$0 \le \mathsf{Rem}_{r^{[j]}}(\boldsymbol{x}^{[0:j]}, \mathbf{dir}^{[0:j]}) \le \frac{\|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\|^2}{8(\kappa-1)\ell^{[j]}},$$
$$|r^{[j]}(\boldsymbol{x}^{[0:j]} + \mathbf{dir}^{[0:j]}) - r^{[j]}(\boldsymbol{x}^{[0:j]})| \le \|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\|.$$

Likewise, ramp' is continuous, nondecreasing, and $1/2$-Lipschitz on the whole real line, so integration gives $0 \le \mathsf{Rem}_{\mathsf{ramp}}(x_{\mathsf{ramp}}, x'_{\mathsf{ramp}} - x_{\mathsf{ramp}}) \le (x'_{\mathsf{ramp}} - x_{\mathsf{ramp}})^2/4$ for every $x_{\mathsf{ramp}}, x'_{\mathsf{ramp}} \in \mathbb{R}$.

Fix coupling $j+1$ and a displacement $\mathbf{dir}$. Let x_{ramp} and x'_{ramp} be its ramp arguments at $\boldsymbol{x}^{[0:j+1]}$ and $\boldsymbol{x}^{[0:j+1]} + \mathbf{dir}^{[0:j+1]}$, respectively. Expanding Eq. (C.7) and cancelling the first-order terms gives the exact identity

$$\begin{aligned}\mathsf{Rem}_{H^{[j+1]}}(\boldsymbol{x}^{[0:j+1]}, \mathbf{dir}^{[0:j+1]}) &= (4\kappa-1)\|\boldsymbol{g}^{[j+1]}\|^2 \mathsf{Rem}_{\mathsf{ramp}}(x_{\mathsf{ramp}}, x'_{\mathsf{ramp}} - x_{\mathsf{ramp}}) \\ &\quad - \|\boldsymbol{g}^{[j+1]}\|(1-\theta^{[j+1]})\mathsf{Rem}_{r^{[j]}}(\boldsymbol{x}^{[0:j]}, \mathbf{dir}^{[0:j]}).\end{aligned}$$

The first term is nonnegative and the second nonpositive. Moreover, the readout bound gives

$$|x'_{\mathsf{ramp}} - x_{\mathsf{ramp}}| \le \frac{\|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\| + |\langle \boldsymbol{g}^{[j+1]}/\|\boldsymbol{g}^{[j+1]}\|, \mathbf{dir}^{[j+1]} \rangle|}{(4\kappa-1)\|\boldsymbol{g}^{[j+1]}\|}.$$

Using $(x_1+x_2)^2 \le 2x_1^2 + 2x_2^2$ in the upper bound, we obtain

$$\begin{aligned}&-\frac{\|\boldsymbol{g}^{[j+1]}\|}{4(\kappa-1)\ell^{[j]}}\|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\|^2 \\ &\le 2\mathsf{Rem}_{H^{[j+1]}}(\boldsymbol{x}^{[0:j+1]}, \mathbf{dir}^{[0:j+1]}) \\ &\le \frac{1}{4\kappa-1}\big(\|\mathsf{Weight}^{[j]}\mathbf{dir}^{[0:j]}\|^2 + |\langle \boldsymbol{g}^{[j+1]}/\|\boldsymbol{g}^{[j+1]}\|, \mathbf{dir}^{[j+1]} \rangle|^2\big).\end{aligned} \tag{C.16}$$

These are global bounds, including ramp joins. The remainder of a sum is the sum of its remainders. By Eq. (C.14), the upper bounds in Eq. (C.16) sum to at most $3\|\mathbf{dir}\|^2/(4\kappa-1)$: the readout terms contribute at most $2\|\mathbf{dir}\|^2$ and the projections onto the separate new blocks contribute at most $\|\mathbf{dir}\|^2$. For the negative part, $\|\boldsymbol{g}^{[j+1]}\|/\ell^{[j]} \le 1/[2(4\kappa-1)]$ and the same readout sum give a magnitude at most $\|\mathbf{dir}\|^2/[4(\kappa-1)(4\kappa-1)]$. Adding the core quadratic bounds gives the global curvature constants

$$\text{lower curvature} \ge \frac{4}{4\kappa-1} - \frac{1}{4(\kappa-1)(4\kappa-1)} = \frac{1}{\kappa} + \frac{3\kappa-4}{4\kappa(\kappa-1)(4\kappa-1)} \ge \frac{1}{\kappa},$$
$$\text{upper curvature} \le \frac{4(\kappa-1)}{4\kappa-1} + \frac{3}{4\kappa-1} = 1.$$

In particular, for every $\boldsymbol{x}, \boldsymbol{y}$,

$$\frac{1}{2\kappa}\|\boldsymbol{y}-\boldsymbol{x}\|^2 \le f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x} \rangle \le \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2.$$

These are the standard quadratic-remainder characterizations of strong convexity and, for a convex differentiable function, Lipschitz continuity of the gradient; see (Bubeck, 2015, Lemmas 3.4–3.5 and Section 3.4). They prove Eqs. (C.1) & (C.2) globally.

At zero, each core gradient vanishes, $r^{[j]}(0) = 0$, $\nabla r^{[j]}(0) = 0$, and every ramp numerator is $-\big(\ell_k^- + 8(\kappa-1)A^{[j+1]}\big) < 0$. Hence, $\nabla f(0) = 0$, and strong convexity gives the unique minimizer. □

C.5 A SAFE REGION FOR THE ACTUAL EARLIER COORDINATES

After release, the earlier blocks still participate in the same GD run and still receive coupling gradients. We therefore need an invariant region for their actual coupled updates. The readout weights place all these cores inside their common quadratic balls, where their gradients are linear; the remaining coupling corrections preserve the required weighted-norm bound for the stated step range.

Lemma C.3 (Safe readout region)**.** *Fix any trial objective (C.8), $0 \le j \le P$, and a full state $\boldsymbol{x}$. If $\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\| \le \mathsf{rad}^{[j]}$, every update with $0 \le \eta_t \le (4\kappa-1)/2$ obeys*

$$\|\mathsf{Weight}^{[j]}\big(\boldsymbol{x}^{[0:j]} - \eta_t \nabla_{\boldsymbol{x}^{[0:j]}} f(\boldsymbol{x})\big)\| \le \|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\|. \tag{C.17}$$

Every incoming correction coefficient on coordinate blocks $1, \dots, j$ is zero there. The conclusion is uniform over all core trials and all later coordinate block states. In particular, for every $0 \le \ell_{\mathsf{safe}} \le \ell^{[j]}$, the condition $r^{[j]} \le \ell_{\mathsf{safe}}$ remains true after these updates.

Proof. The weighted norm hypothesis serves two purposes: it puts every earlier core inside its quadratic ball and switches off its incoming correction. The remaining coupling corrections are radial, so each earlier block is multiplied by a scalar of absolute value at most one. Indeed, the coefficient of $\|\mathsf{Weight}^{[i]}\boldsymbol{x}^{[0:i]}\|^2$ in the recursive expansion of $\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\|^2$ is $16^{j-i}(\mathsf{rad}^{[j]})^2/(\mathsf{rad}^{[i]})^2$. Keeping this nonnegative term and using $\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\| \le \mathsf{rad}^{[j]}$ gives, for $0 \le i \le j$,

$$\|\mathsf{Weight}^{[i]}\boldsymbol{x}^{[0:i]}\| \le \mathsf{rad}^{[i]}/4^{j-i}, \qquad \|\boldsymbol{x}^{[i]}\| \le \mathsf{rad}^{[i]}.$$

Every earlier coordinate block is therefore inside its core's common quadratic ball. For an incoming coupling $i \le j$, the weighted norm of the earlier blocks is at most $\mathsf{rad}^{[i-1]}/4$. The scalar function $r_{\mathsf{norm}} \mapsto \sqrt{r_{\mathsf{norm}}^2 + (4(\kappa-1)\ell^{[i-1]})^2} - 4(\kappa-1)\ell^{[i-1]}$ is convex, equals zero at $r_{\mathsf{norm}} = 0$, and equals $\ell^{[i-1]}$ at $r_{\mathsf{norm}} = \mathsf{rad}^{[i-1]}$. When the weighted norm of the earlier blocks is at most $\mathsf{rad}^{[i-1]}/4$, convexity therefore gives $r^{[i-1]}(\boldsymbol{x}^{[0:i-1]}) \le \ell^{[i-1]}/4$. The new block's norm is at most $\mathsf{rad}^{[i]} \le A^{[i]} \le \frac{1}{8(\kappa-1)}\ell^{[i-1]}$, while its offset is at least $\ell^{[i-1]}/2$. The absolute value of the projection of the new block in the ramp numerator is bounded by this block norm. Its ramp numerator is at most $\ell^{[i-1]}(1/4 + 1/8 - 1/2) < 0$, so $\theta^{[i]} = 0$.

On a coordinate block $i \le j$, the core gradient is exactly $4\boldsymbol{x}^{[i]}/(4\kappa-1)$, and its incoming correction has just vanished. All other corrections to it are radial readout subtractions from later couplings. By Eq. (C.13), its complete gradient is

$$\nabla_{\boldsymbol{x}^{[i]}} f(\boldsymbol{x}) = (4/(4\kappa-1) - \omega^{[i]}(\boldsymbol{x}))\boldsymbol{x}^{[i]},$$
$$\omega^{[i]}(\boldsymbol{x}) = \sum_{l=i}^{P-1} 16^{l-i}\frac{(\mathsf{rad}^{[l]})^2}{(\mathsf{rad}^{[i]})^2}\frac{\|\boldsymbol{g}^{[l+1]}\|(1-\theta^{[l+1]}(\boldsymbol{x}^{[0:l+1]}))}{\sqrt{\|\mathsf{Weight}^{[l]}\boldsymbol{x}^{[0:l]}\|^2 + (4(\kappa-1)\ell^{[l]})^2}}.$$

There is no restriction on the later correction coefficients in this expression. Each denominator is at least $4(\kappa-1)\ell^{[l]}$, so (C.11) bounds the corresponding scalar factor by $1/[8(\kappa-1)(4\kappa-1)]$. Summing the readout weights as in (C.14) gives

$$0 \le \omega^{[i]}(\boldsymbol{x}) \le \frac{1}{8(\kappa-1)(4\kappa-1)}\sum_{l=i}^{P-1} 16^{l-i}\frac{(\mathsf{rad}^{[l]})^2}{(\mathsf{rad}^{[i]})^2} \le \frac{1}{4(\kappa-1)(4\kappa-1)} < \frac{2}{4\kappa-1}.$$

The estimate uses only $0 \le \theta^{[l+1]} \le 1$ and the uniform denominator bound, so it remains valid for arbitrary later block states and every core trial. Thus, $2/(4\kappa-1) \le 4/(4\kappa-1) - \omega^{[i]}(\boldsymbol{x}) \le 4/(4\kappa-1)$, and an update multiplies $\boldsymbol{x}^{[i]}$ by $1 - \eta_t[4/(4\kappa-1) - \omega^{[i]}(\boldsymbol{x})] \in [-1, 1]$ for $0 \le \eta_t \le (4\kappa-1)/2$. The norm of each earlier coordinate block cannot increase. Their fixed positive weighted sum (C.13) proves Eq. (C.17). Since $r^{[j]}$ is an increasing function of $\|\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\|$, any upper bound on this readout remains valid after the update. □

The step restriction in Lem. C.3 is not a restriction on the schedule in Lem. 2.1. It is used only during the first one or two gaps where the readout and projection conditions must hold. The controlled second transfer handles a larger first checkpoint; during later transfers, the preceding transfer function has zero derivative in the shared active coordinate, regardless of the retired coordinates.

C.6 Fresh readout crossings

A readout need not decrease monotonically after large steps. We instead locate a band above which all earlier readouts lie and below which the current readout lies. Saturation at all earlier iterates then preserves the old gradients exactly, while the crossing releases the new block.

The fresh crossing has two gradient consequences. Before the crossing, cancellation preserves the old gradients and holds the new block at its initial point. At the crossing iterate the incoming correction vanishes, while the state is still at its initial point; only the next update can move the new block.

Lemma C.4 (A fresh band crossing)**.** *Let $r_0, \dots, r_T$ be real numbers with $r_0 \geq \ell$ and $r_T \leq \ell/2$, where $T \geq 1$ and $\ell > 0$. Define the bands as in Eq. (C.5) with level ℓ. There exist integers $1 \leq k \leq n_{\mathsf{wait}} \leq T$ such that $r_t \geq \ell_k^+$ for all $t < n_{\mathsf{wait}}$ and $r_{n_{\mathsf{wait}}} \leq \ell_k^-$. In fact, one may take $k = n_{\mathsf{wait}}$.*

Proof. Use the band index itself as a candidate crossing time. The identity $\ell_t^- = \ell_{t+1}^+$ then makes the failure to cross at an earlier time imply the required upper-band bound for a later one. Choose $n_{\mathsf{wait}} = \min\{t \in \{1, \dots, T\} : r_t \leq \ell_t^-\}$, which exists because $r_T \leq \ell/2 < \ell_T^-$, and set $k = n_{\mathsf{wait}}$. For $1 \leq t < n_{\mathsf{wait}}$, minimality of n_{wait} and the decreasing band levels give

$$r_t > \ell_t^- = \ell_{t+1}^+ \geq \ell_{n_{\mathsf{wait}}}^+.$$

Also $r_0 \geq \ell = \ell_1^+ \geq \ell_{n_{\mathsf{wait}}}^+$, while $r_{n_{\mathsf{wait}}} \leq \ell_{n_{\mathsf{wait}}}^-$ by definition. No monotonicity of the readouts is required. □

C.7 Matching and extending one phase

The fixed-point step uses the correction produced by each complete trial objective. The common initial gradient is crucial here: it fixes the coupling while the new core's trial parameter varies. Each trial therefore defines its own unique finite GD run and correction coefficients before any matching condition is imposed. The following elementary consequence of Brouwer's theorem states the needed consistency condition independently of the construction.

Lemma C.5 (Matching a continuous feedback)**.** *Let $\mathsf{Trials} \subseteq \mathbb{R}^d$ be nonempty, compact, and convex, where d is finite, and let $\mathcal{U}$ be a topological space. Suppose $\mathsf{Shoot} : \mathsf{Trials} \times \mathcal{U} \to \mathsf{Trials}$ and $\Theta : \mathsf{Trials} \to \mathcal{U}$ are continuous. Then some $\boldsymbol{p}^\star \in \mathsf{Trials}$ satisfies $\boldsymbol{p}^\star = \mathsf{Shoot}(\boldsymbol{p}^\star, \Theta(\boldsymbol{p}^\star))$.*

Proof. The composition $\boldsymbol{p} \mapsto \mathsf{Shoot}(\boldsymbol{p}, \Theta(\boldsymbol{p}))$ is a continuous self-map of the nonempty compact convex set Trials. Brouwer's fixed-point theorem gives the conclusion (Bonsall, 1962, Lemma 2.4). □

The following lemma takes the prescribed $N+1$ steps as input; Appx. C.8 specifies $\eta_{N+1} = 128$ for the finite serial construction.

Lemma C.6 (One extension preserving earlier phases)**.** *Fix a partial objective f^{old} of the form (C.8), with the core and coupling data of Appx. C.1, on coordinate blocks $0, \dots, j$. Fix its trajectory under the prescribed steps $\boldsymbol{\eta}_{1:N+1}$, with every earlier phase identified through its certified endpoint. Suppose the integer $t_{\mathrm{start}}^{[j+1]} \in \{0, \dots, N\}$ is strictly greater than every earlier certified endpoint, and, for a band $k^{[j]} \geq 1$,*

$$r^{[j]}(\boldsymbol{x}_t^{[0:j]}) \geq \ell_{k^{[j]}}^+ \quad (0 \leq t < t_{\mathrm{start}}^{[j+1]}), \qquad r^{[j]}(\boldsymbol{x}_{t_{\mathrm{start}}^{[j+1]}}^{[0:j]}) \leq \ell_{k^{[j]}}^-.$$

Suppose also that the assigned steps $t_{\mathrm{start}}^{[j+1]}+1, \dots, t_{\mathrm{end}}^{[j+1]} \leq N+1$ form a phase of length $n^{[j+1]} \geq 1$. One can append a core $F^{[j+1]}$ and coupling $H^{[j+1]}$ with parameters (C.6) so that the extended objective belongs to $\mathcal{F}_{1/\kappa,1}$ and is minimized at zero. Its old-coordinate trajectory and the new block's initial point are preserved through $t_{\mathrm{start}}^{[j+1]}$, with zero incoming coefficient there. On its assigned phase, the new block realizes the matched core and satisfies

$$\|\boldsymbol{x}_t^{[j+1]}\| \geq A^{[j+1]}\mathsf{amp}(n^{[j+1]}) = 4\mathsf{rad}^{[j+1]} \quad (t_{\mathrm{start}}^{[j+1]} \leq t \leq t_{\mathrm{end}}^{[j+1]}).$$

Every older certified trajectory and its actual correction coefficients remain unchanged, and the new block is held at its initial point before its phase starts.

Proof. Use the family of Lem. B.8 for the assigned phase. Construct its initial amplitude and coupling by Eq. (C.6), using the selected band $k^{[j]}$ and the already fixed $\ell^{[j]}$. The assigned length fixes the radius of its common quadratic ball. Its fixed initial gradient then fixes its coupling width and direction before choosing a trial.

Claim: exact prefix preservation for every trial. For every trial of the new core, the old-coordinate iterates remain exactly $\boldsymbol{x}_t$ through iteration $t_{\mathrm{start}}^{[j+1]}$, the new coordinate block remains at its initial point through that iteration, and its incoming correction coefficient is zero at $t_{\mathrm{start}}^{[j+1]}$.

Proof of the claim. At iteration zero, the old coordinates are unchanged and the new coordinate block is at its initial point. Suppose this holds at some $t < t_{\mathrm{start}}^{[j+1]}$. Its readout is still the old $r^{[j]}(\boldsymbol{x}_t^{[0:j]}) \geq \ell_{k^{[j]}}^{+}$, because $\mathsf{Weight}^{[j]}$ reads only old coordinates. By Lem. C.1, $\theta^{[j+1]} = 1$. Equation (C.9) makes the old-coordinate gradient exactly ∇f^{old}, and the sum of the new core gradient and its incoming correction exactly zero. The next update therefore preserves both assertions, for any prescribed step magnitude. Induction proves them through iteration $t_{\mathrm{start}}^{[j+1]}$. The update producing that iterate uses the still-saturated gradient at the preceding iterate; release changes only the gradient used by the next update. At that iterate, the readout is at most $\ell_{k^{[j]}}^{-}$, and $|\langle \boldsymbol{g}^{[j+1]}/\|\boldsymbol{g}^{[j+1]}\|, \boldsymbol{x}_{\mathsf{init}}^{[j+1]}\rangle| \leq A^{[j+1]} \leq 8(\kappa-1)A^{[j+1]}$, so (C.10) makes the incoming correction zero. All coupling data and the new core's initial gradient are fixed across trials. The proof consequently holds for every trial, not just a future matching one. This claim concerns equality of old-coordinate iterates. The extended full state has an additional fixed coordinate block, and its objective value before release need not equal the old value. Gradient equality, not value equality, is what the GD construction needs.

Match only the new core. There is no need to rematch an earlier core. Prefix preservation holds for every new trial, and the trial-independent initial gradient fixes the new coupling before shooting. We may therefore run each full trial objective to read its controls and insert those controls into the local self-map. If the new phase has no controlled transfer, call its singleton trial $\boldsymbol{p}^\star$. Otherwise, let $\boldsymbol{p}$ range over its compact rectangle. For each $\boldsymbol{p}$, build the complete extended objective $f_{\boldsymbol{p}}^{\mathrm{trial}}$, initialize at the extended initial point, and run the prescribed $N+1$ steps. Read the incoming correction coefficients at the designated iterates of the new phase's second gap. A step with phase-local index $t_{\mathsf{loc}} \geq 1$ uses the iterate at index $t_{\mathrm{start}}^{[j+1]} + t_{\mathsf{loc}} - 1$. If $t_1 < t_2$ are the first two local checkpoints, this vector queries exactly the global states $t_{\mathrm{start}}^{[j+1]} + t_1, \ldots, t_{\mathrm{start}}^{[j+1]} + t_2 - 2$. Call the resulting vector $\boldsymbol{\theta}(\boldsymbol{p})$, with $n_{\mathsf{gap}} = t_2 - t_1 - 1$; when $n_{\mathsf{gap}} = 0$ it is the empty vector. The second-checkpoint coefficient at state $t_{\mathrm{start}}^{[j+1]} + t_2 - 1$ is not part of this vector. By Lem. C.1, the vector of correction coefficients belongs to $[0,1]^{n_{\mathsf{gap}}}$ for every trial, whether or not that trial's later trajectory is yet identified. There is no implicit trajectory choice here: each trial defines a complete differentiable objective, and the prescribed finite schedule uniquely determines its GD iterates.

The incoming correction after spatial scaling uses the trial-independent initial gradient. Indeed, using $A^{[j+1]} = \|\boldsymbol{x}_{\mathsf{init}}^{[j+1]}\|$ and (C.3), with $\widehat{\boldsymbol{z}} = \boldsymbol{x}^{[j+1]}/A^{[j+1]}$,

$$\frac{\eta_t}{A^{[j+1]}}\left[\nabla F^{[j+1]}(\boldsymbol{x}^{[j+1]}) - \theta \nabla F^{[j+1]}(\boldsymbol{x}_{\mathsf{init}}^{[j+1]})\right] = \eta_t[\nabla h_{\boldsymbol{p}}(\widehat{\boldsymbol{z}}) - \theta \nabla h_{\boldsymbol{p}}(\boldsymbol{e}^{(1)})].$$

Thus, the second-pair controls are $u_{\max}\boldsymbol{\theta}(\boldsymbol{p})$, with $u_{\max}$ fixed as in Eq. (B.15); the sign $g_*^{(2)} \leq 0$ gives the plus sign in (B.16). Consider the map

$$\boldsymbol{p} \longmapsto \mathsf{Shoot}(\boldsymbol{p}, u_{\max}\boldsymbol{\theta}(\boldsymbol{p})). \tag{C.18}$$

Lem. B.8 supplies the map $\mathsf{Shoot} : \mathsf{Trials} \times [0, u_{\max}]^{n_{\mathsf{gap}}} \to \mathsf{Trials}$, so this is a self-map before identification. It is continuous: every core gradient is jointly continuous by Lem. B.8, and the coupling parameters are fixed. Starting from the same initial point, induction in the finite recursion $\boldsymbol{x}_{t+1}(\boldsymbol{p}) = \boldsymbol{x}_t(\boldsymbol{p}) - \eta_{t+1}\nabla f_{\boldsymbol{p}}^{\mathrm{trial}}(\boldsymbol{x}_t(\boldsymbol{p}))$ proves continuity of every trial iterate. Reading finitely many correction coefficients and applying the jointly continuous local map preserves continuity. Lem. C.5 gives a fixed point $\boldsymbol{p}^\star$. At this point, the control vector used by the local shooting construction is exactly the one produced by the full trial objective. The next step verifies that the actual core enters that controlled transfer at its required input.

Identify the new phase and preserve earlier ones. Fix this extended objective permanently at $\boldsymbol{p}^\star$. At $t_{\mathrm{start}}^{[j+1]}$, its new coordinate block is exactly $\boldsymbol{x}_{\mathsf{init}}^{[j+1]}$ and the readout of the earlier blocks is

at most $\ell^-_{k^{[j]}} < \ell^{[j]}$. Apply Lem. B.9 with the normalized new block $\widehat{\boldsymbol{z}} = \boldsymbol{x}^{[j+1]}/A^{[j+1]}$, readout $r_t = r^{[j]}(\boldsymbol{x}^{[0:j]}_{t^{[j+1]}_{\text{start}}+t})$, and release threshold $\ell_{\text{in}} = \ell^-_{k^{[j]}}$. The normalized new block starts at $\boldsymbol{e}^{(1)}$.

Whenever the readout of the earlier blocks is at most $\ell^-_{k^{[j]}}$, the weighted norm of the earlier blocks is below $\mathsf{rad}^{[j]}$, since $\ell^-_{k^{[j]}} < \ell^{[j]}$. Lem. C.3 then preserves this same scalar threshold under every step at most $(4\kappa - 1)/2$, which is the second condition in Eq. (B.26). For the first condition, $\boldsymbol{g}^{[j+1]} = A^{[j+1]}\boldsymbol{g}_*$, so spatial scaling preserves the initial-gradient direction. Dividing the projection bound in (C.10) by $A^{[j+1]}$ gives exactly the projected $8(\kappa - 1)$ bound in Eq. (B.26). Lem. C.1 therefore makes the incoming correction zero whenever that condition holds. The new coordinate block has no outgoing coupling at this construction stage, so its only correction is the incoming correction, with the spatial scaling given above. The guarded first-gap induction, and then the unforced or matched second-pair induction in Lem. B.9, identify the phase through its endpoint. Its coordinate block norm is at least $4\mathsf{rad}^{[j+1]}$ at every assigned iterate and endpoint.

The new coupling preserved the whole earlier trajectory through $t^{[j+1]}_{\text{start}} > t^{[j]}_{\text{end}}$. In particular, it preserved all older certified phase trajectories and their actual correction coefficients, not only their initial points. None of the matching parameters change. The extended objective belongs to the target class and has minimizer zero by Lem. C.2. This supplies the same static data and certified trajectories for a later extension. □

C.8 The schedule partition

At each stage, the current cores, couplings, and matched phase trajectories are fixed. Only the new core's trial parameters will vary. An extension must preserve every earlier certified endpoint; termination is certified either by a large final readout or by a phase reaching the appended last step. The construction selects the final function in finitely many stages; at each stage, every trial determines a GD run of finite length. The resulting function is fixed throughout its own GD run. For now, take $N \geq 1$ and a positive prescribed schedule $\boldsymbol{\eta}_{1:N}$. Append $\eta_{N+1} = 128$. The iterate $\boldsymbol{x}_t$ is the state after t updates; update $t+1$ uses step size η_{t+1} and produces $\boldsymbol{x}_{t+1}$. Mark every maximal consecutive run of steps of size at most 64 whose step sum exceeds $(4\kappa - 1)/512$. The appended step creates no new such run. Each marked run of n_{run} positions satisfies $64n_{\text{run}} \geq \sum \eta_t > (4\kappa - 1)/512$, hence $n_{\text{run}} > (4\kappa - 1)/32768 > (\kappa - 1)/8192$. Thus,

$$\#\{\text{marked runs}\} < 8192N/(\kappa - 1). \tag{C.19}$$

For a phase-start index $t^{\text{start}} \in \{0, \ldots, N\}$, ignore the marked run containing step $t^{\text{start}} + 1$, when one exists. If the next marked run starts at step $t_{\text{first}} > t^{\text{start}} + 1$, set $t^{\text{end}}(t^{\text{start}}) = t_{\text{first}} - 1$; if none exists, set $t^{\text{end}}(t^{\text{start}}) = N + 1$. The phase assigned to t^{start} uses steps $t^{\text{start}} + 1, \ldots, t^{\text{end}}(t^{\text{start}})$ and has length $n(t^{\text{start}}) = t^{\text{end}}(t^{\text{start}}) - t^{\text{start}} \geq 1$. Its last step exceeds 64, either because the following run is maximal or because it is the appended step. A marked run containing step $t^{\text{start}} + 1$ is part of its unrestricted leading run. All strictly internal small-step runs are unmarked, with step sum at most $(4\kappa - 1)/512$; hence the phase meets Lem. B.8.

Fig. 11 separates the roles of this partition and checkpoint selection. A phase assigns a coordinate block, a waiting interval precedes the next release, and a window only groups positions for the scalar cost bounds. In particular, a gap can cross a window boundary while the same coordinate pair remains active.

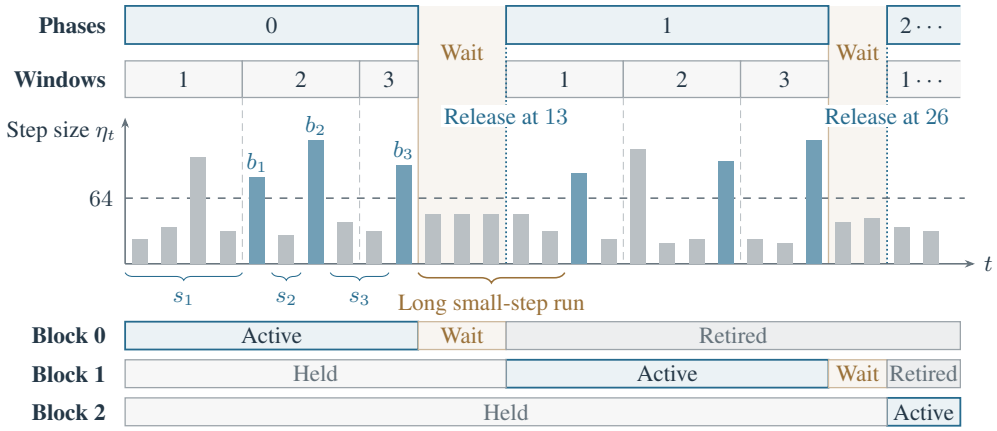


Figure 11: Schematic schedule and block activity, with one core per phase. Blue bars mark checkpoints of sizes b_i; s_i are gap step sums. Windows (defined in Sec. 3.2) account for checkpoint costs without resetting GD. Held blocks remain fixed, and active blocks execute their phase. During a waiting interval, earlier blocks continue GD while the next block remains held; retired blocks may also move. A readout crossing determines each release, illustrated at states 13 and 26, before the first updates 14 and 27 of phases 1 and 2.

C.9 The finite serial certificate

The following result gathers the consequences needed to finish the lower bound. Its trajectory assertion ensures that all phases occur in one fixed objective; its length and scale bounds quantify what each extension costs. The terminal comparison covers both stopping alternatives, so no later matching or construction step remains when we apply the scalar loss bound.

Lemma C.7 (Finite serial construction)**.** *For every sufficiently large κ and every positive schedule of length $N \geq 1$, there are a fixed finite-dimensional $f \in \mathcal{F}_{1/\kappa,1}$ with minimizer zero and an initial point satisfying $1 \leq \|\boldsymbol{x}_0\|^2 < 2$. The construction gives an integer $P \geq 0$, positive integer phase lengths $n^{[0]}, \ldots, n^{[P]}$, positive integer waits $n^{[0]}_{\mathsf{wait}}, \ldots, n^{[P-1]}_{\mathsf{wait}}$, and positive scales $A^{[0]}, \ldots, A^{[P]}$, with $A^{[0]} = 1$, such that*

$$\sum_{j=0}^{P} n^{[j]} + \sum_{j=0}^{P-1} n^{[j]}_{\mathsf{wait}} \leq N + 1, \qquad P < 8192N/(\kappa - 1),$$

$$\frac{(A^{[j+1]})^2}{(A^{[j]})^2} \geq \frac{\mathsf{amp}(n^{[j]})^2}{665856(\kappa-1)^3(1 + n^{[j]}_{\mathsf{wait}})^4} \quad (0 \leq j < P),$$

$$R_D(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) \geq \frac{(\mathsf{rad}^{[P]})^2}{8(8(\kappa-1)+1)}, \qquad \mathsf{rad}^{[P]} = A^{[P]}\mathsf{amp}(n^{[P]})/4.$$

One can choose the coordinate blocks, couplings, and core trials successively so that, for every constructed coordinate block j, its assigned phase begins at that block's initial point at its phase-start index $t^{[j]}_{\mathrm{start}}$ and is identified through $t^{[j]}_{\mathrm{end}} = t^{\mathrm{end}}(t^{[j]}_{\mathrm{start}})$. Its first update uses $\eta_{t^{[j]}_{\mathrm{start}}+1}$. Every later extension preserves the entire trajectory through that certified endpoint. The procedure terminates after $P < 8192N/(\kappa-1)$ extensions with either $t^{[P]}_{\mathrm{end}} = N + 1$ or

$$r^{[P]}(\boldsymbol{x}_N^{[0:P]}) > \ell^{[P]}/2 \tag{C.20}$$

on the final objective.

Proof. We maintain the actual trajectories through the certified phase endpoints, not just the separate cores' local formulas. An extension has three steps: find a fresh crossing on the current objective, choose a new core consistent with the correction generated by its own trial run, and verify its phase without changing any earlier certified trajectory.

First core. Assign the phase starting at $t_{\mathrm{start}}^{[0]} = 0$ and set $\boldsymbol{x}_{\mathsf{init}}^{[0]} = \boldsymbol{e}^{(1)}$. Use the family supplied by Lem. B.8 for this phase. Choose a matching trial for that phase with zero incoming correction. If it has a controlled transfer, Lems. C.5 & B.8 give a fixed point of $\boldsymbol{p} \mapsto \mathsf{Shoot}(\boldsymbol{p}, 0)$; otherwise, no trial is needed. Run this objective on one coordinate block through all $N+1$ appended steps. Lem. B.9 identifies its phase and gives $\|\boldsymbol{x}_t^{[0]}\| \geq A^{[0]}\mathsf{amp}(n^{[0]}) = 4\mathsf{rad}^{[0]}$ for $0 \leq t \leq t_{\mathrm{end}}^{[0]}$.

Inductive state and the decision to extend. Suppose coordinate blocks $0, \ldots, j$ and all their parameters have been fixed. Identify their assigned phases, and note that all earlier certified trajectories remain valid on the current objective. If $t_{\mathrm{end}}^{[j]} = N+1$, stop. Otherwise, inspect the terminal readout of this current objective. If (C.20) holds for coordinate block j, stop. In the remaining case, $r^{[j]}(\boldsymbol{x}_{t_{\mathrm{end}}^{[j]}}^{[0:j]}) > \ell^{[j]}$ by the phase lower bound and $r^{[j]}(\boldsymbol{x}_N^{[0:j]}) \leq \ell^{[j]}/2$. These two inequalities imply $t_{\mathrm{end}}^{[j]} < N$. Apply Lem. C.4 to the finite sequence of readouts at iterations $t_{\mathrm{end}}^{[j]}, \ldots, N$. It supplies $1 \leq k^{[j]} = n_{\mathsf{wait}}^{[j]} \leq N - t_{\mathrm{end}}^{[j]}$ with

$$r^{[j]}(\boldsymbol{x}_t^{[0:j]}) \geq \ell_{k^{[j]}}^{+} \quad (t_{\mathrm{end}}^{[j]} \leq t < t_{\mathrm{end}}^{[j]} + n_{\mathsf{wait}}^{[j]}), \qquad r^{[j]}(\boldsymbol{x}_{t_{\mathrm{end}}^{[j]}+n_{\mathsf{wait}}^{[j]}}^{[0:j]}) \leq \ell_{k^{[j]}}^{-}.$$

The upper inequality also holds at every earlier iterate. Before the phase starts for coordinate block j, it remains at its initial point, so $\|\mathsf{Weight}^{[j]}\boldsymbol{x}_t^{[0:j]}\| \geq A^{[j]} \geq \mathsf{rad}^{[j]}$. During its certified phase, $\|\mathsf{Weight}^{[j]}\boldsymbol{x}_t^{[0:j]}\| \geq \|\boldsymbol{x}_t^{[j]}\| \geq 4\mathsf{rad}^{[j]}$. These facts cover all $0 \leq t < t_{\mathrm{end}}^{[j]}$.

Set $t_{\mathrm{start}}^{[j+1]} = t_{\mathrm{end}}^{[j]} + n_{\mathsf{wait}}^{[j]}$ and assign the phase ending at $t^{\mathrm{end}}(t_{\mathrm{start}}^{[j+1]})$. The phase is admissible by the partition above, and the crossing inequalities hold at all earlier states. Apply Lem. C.6: it matches the new core, identifies its assigned phase, and preserves the whole earlier trajectory through $t_{\mathrm{start}}^{[j+1]} > t_{\mathrm{end}}^{[j]}$. None of the earlier matching parameters change. This proves the induction hypothesis for the new current objective. A later extension has the same property, so eventually the final objective realizes every certified trajectory.

Termination. When an extension is needed, a marked run starts at step $t_{\mathrm{end}}^{[j]} + 1$. Since $n_{\mathsf{wait}}^{[j]} \geq 1$, the next phase-start index is at least $t_{\mathrm{end}}^{[j]} + 1$, after the first step of that run. That run is then either over or belongs to the new leading run. Indeed, if its last step is t_{last}, a new start at least t_{last} is after the run, whereas a new start below t_{last} has its next update inside that run. It cannot cause a strictly internal marked run again for a later coordinate block. Hence, each extension accounts for a distinct marked run, and the number of marked runs is finite. The bound (C.19) therefore gives $P < 8192N/(\kappa - 1)$. The procedure must therefore reach one of its two stopping alternatives. All functions used to select parameters are finite-dimensional. Once the final parameters are chosen, the entire sum (C.8) is fixed from initialization; the argument adds no term during any run on that final objective. A terminal readout inspected at an intermediate construction stage may change after a subsequent extension. This causes no difficulty: that inspection is used only to choose the next fresh band for the current objective. The preserved objects in the induction are the earlier trajectories through the phase-start indices, not the earlier terminal tests.

Scale losses and terminal distance. For band index $k^{[j]} \leq n_{\mathsf{wait}}^{[j]}$, (C.5) and (C.6) give

$$A^{[j+1]} = \frac{\ell^{[j]}}{4k^{[j]}(k^{[j]}+1)(16(\kappa-1)+1)} \geq \frac{\mathsf{rad}^{[j]}}{204(\kappa-1)^{3/2}(n_{\mathsf{wait}}^{[j]}+1)^2}.$$

Here we used $\sqrt{8(\kappa-1)+1} \leq 3\sqrt{(\kappa-1)}$, $16(\kappa-1)+1 \leq 17(\kappa-1)$, and $k^{[j]}(k^{[j]}+1) \leq (n_{\mathsf{wait}}^{[j]}+1)^2$. Since $\mathsf{rad}^{[j]} = A^{[j]}\mathsf{amp}(n^{[j]})/4$,

$$\frac{(A^{[j+1]})^2}{(A^{[j]})^2} \geq \frac{\mathsf{amp}(n^{[j]})^2}{665856(\kappa-1)^3(n_{\mathsf{wait}}^{[j]}+1)^4}.$$

The phase intervals of updates $t_{\mathrm{start}}^{[j]}+1, \ldots, t_{\mathrm{end}}^{[j]}$ and the waiting intervals $t_{\mathrm{end}}^{[j]}+1, \ldots, t_{\mathrm{start}}^{[j+1]}$ are disjoint and partition all updates through $t_{\mathrm{end}}^{[P]}$:

$$\sum_{j=0}^{P} n^{[j]} + \sum_{j=0}^{P-1} n_{\mathsf{wait}}^{[j]} = t_{\mathsf{end}}^{[P]} \leq N+1, \qquad P < 8192N/(\kappa-1).$$

The dimension is finite: block j uses at most $n^{[j]}+1$ coordinates, so the total is at most $N+P+2$. By Lem. C.2, the final objective belongs to $\mathcal{F}_{1/\kappa,1}$ and is minimized at zero; (C.15) gives $1 \le \|\boldsymbol{x}_0\|^2 < 2$. If construction stops with $r^{[P]}(\boldsymbol{x}_N^{[0:P]}) > \ell^{[P]}/2$, then

$$\|\boldsymbol{x}_N\| \ge \|\mathsf{Weight}^{[P]}\boldsymbol{x}_N^{[0:P]}\| \ge r^{[P]}(\boldsymbol{x}_N^{[0:P]}) > \ell^{[P]}/2 = \mathsf{rad}^{[P]}/(2\sqrt{8(\kappa-1)+1}).$$

Otherwise $t_{\mathsf{end}}^{[P]} = N+1$, so iteration N belongs to the final matched phase and $\|\boldsymbol{x}_N\| \ge \|\boldsymbol{x}_N^{[P]}\| \ge A^{[P]}\mathsf{amp}(n^{[P]}) = 4\mathsf{rad}^{[P]}$. This includes a final phase starting at iteration N, whose local state zero is the target state. Squaring either bound and dividing by the initial squared norm gives $R_D \ge (\mathsf{rad}^{[P]})^2/[8(8(\kappa-1)+1)]$. When $P=0$, the waiting sums and scale products are empty, and the same construction and comparison apply. □

D Completing the Lower Bound

The serial construction has produced one fixed objective and reduced its terminal distance to positive scalar quantities: the phase lengths, waiting lengths, and ratios of consecutive block scales. We now multiply these scale ratios and sum their logarithmic losses to prove Lem. 2.1. The window scale makes the loss proportional to $N/\mathcal{Q}(\kappa)$, apart from a polynomial prefactor in κ independent of the schedule and horizon. Sec. 2 removes that prefactor by finite extraction and uses the quadratic baseline given below for bounded condition numbers.

D.1 Accumulated scalar loss

Each phase costs an exponential in its number of windows and a polynomial in κ. Each release adds a polynomial loss in κ and in its waiting length. There are only $O(N/\kappa)$ releases, and the sum of phase and waiting lengths is at most $N+1$. These two bounds prevent either many short waits or one long wait from producing an additional horizon-dependent prefactor. The next lemma states this accounting independently of the already completed objective construction.

Lemma D.1 (Accumulated scalar loss)**.** *Fix constants $C_{\mathsf{amp}} > 0$, $C_{\mathsf{win}} \geq 2$, $C_{\log q} \geq 0$, and $C_{q\log\kappa} > 0$. There are constants $C_1, C > 0$, depending only on these four constants, with the following property. Let $\kappa \geq 2$ and $q \geq 1$ satisfy*

$$\log q \leq C_{\log q}\log\kappa, \qquad \frac{q\log\kappa}{(\kappa-1)} \leq C_{q\log\kappa}, \qquad \mathsf{amp}(n) = (\kappa-1)^{-C_{\mathsf{amp}}} e^{-C_{\mathsf{win}}\lceil n/q\rceil}.$$

Let $N \geq 1$ and $P \geq 0$ be integers, let $n_0, \dots, n_P \geq 1$ and $n_{\mathsf{wait},0}, \dots, n_{\mathsf{wait},\mathsf{P}-1} \geq 1$ be integers, and let $A_0, \dots, A_P > 0$, with $A_0 = 1$. Suppose

$$\begin{aligned} \sum_{j=0}^{P} n_j + \sum_{j=0}^{P-1} n_{\mathsf{wait},\mathsf{j}} \leq N+1, \qquad P < \frac{8192N}{(\kappa-1)}, \\ \frac{A_{j+1}^2}{A_j^2} \geq \frac{\mathsf{amp}(n_j)^2}{665856(\kappa-1)^3(1+n_{\mathsf{wait},\mathsf{j}})^4} \quad (0 \leq j < P). \end{aligned} \tag{D.1}$$

If $\mathsf{rad} = A_P\mathsf{amp}(n_P)/4$ and $\mathsf{err} \geq \mathsf{rad}^2/[8(8(\kappa-1)+1)]$, then

$$\mathsf{err} \geq \kappa^{-C_1} e^{-CN/q}. \tag{D.2}$$

Proof. The lower bound on err is positive, so its logarithm is defined. We first bound the number of charged windows by the total phase length plus one rounding term per phase. For waiting intervals, a linear charge of n_{wait}/q and a fixed charge of $\log q$ per release dominate the logarithmic waiting loss. Moreover, $qe^{n_{\mathsf{wait}}/q} \geq q + n_{\mathsf{wait}} \geq 1 + n_{\mathsf{wait}}$ yields

$$\sum_{j=0}^{P} \lceil n_j/q\rceil \leq q^{-1}\sum_{j=0}^{P} n_j + P + 1, \qquad \sum_{j=0}^{P-1} \log(1+n_{\mathsf{wait},\mathsf{j}}) \leq P\log q + q^{-1}\sum_{j=0}^{P-1} n_{\mathsf{wait},\mathsf{j}}.$$

Multiply the scale inequalities in (D.1), using $A_0 = 1$ and $\mathsf{rad}^2 = A_P^2\mathsf{amp}(n_P)^2/16$. Taking negative logarithms gives

$$\begin{aligned} -\log\mathsf{err} &\leq 2C_{\mathsf{win}}\sum_{j=0}^{P}\lceil n_j/q\rceil + [2C_{\mathsf{amp}}(P+1)+3P]\log(\kappa-1) \\ &\quad + \log 128 + \log(8(\kappa-1)+1) + P\log 665856 + 4\sum_{j=0}^{P-1}\log(1+n_{\mathsf{wait},\mathsf{j}}) \\ &\leq \frac{2C_{\mathsf{win}}}{q}(N+1) + 2C_{\mathsf{amp}}\log(\kappa-1) + 2C_{\mathsf{win}} + \log 128 + \log(8(\kappa-1)+1) \\ &\quad + P[(2C_{\mathsf{amp}}+3)\log(\kappa-1) + 4\log q + 2C_{\mathsf{win}} + \log 665856]. \end{aligned}$$

Here $2C_{\mathsf{win}} \geq 4$ lets us combine phase and waiting lengths. Since $q \geq 1$, $8(\kappa-1)+1 \leq 9\kappa$, and $\log\kappa \geq \log 2$, the terms outside the bracket are at most

$$\frac{2C_{\mathsf{win}}N}{q} + \left(2C_{\mathsf{amp}} + 1 + \frac{4C_{\mathsf{win}} + \log 1152}{\log 2}\right)\log\kappa.$$

The bracket is at most $[2C_{\mathsf{amp}} + 3 + 4C_{\log q} + (2C_{\mathsf{win}} + \log 665856)/\log 2]\log\kappa$. Using the bounds on P and $q\log\kappa/(\kappa-1)$ proves (D.2), for example with

$$C_1 = 2C_{\mathsf{amp}} + 1 + \frac{4C_{\mathsf{win}} + \log 1152}{\log 2},$$
$$C = 2C_{\mathsf{win}} + 8192C_{q\log\kappa}\left(2C_{\mathsf{amp}} + 3 + 4C_{\log q} + \frac{2C_{\mathsf{win}} + \log 665856}{\log 2}\right).$$

Empty sums and products give the same argument when $P = 0$. □

D.2 The finite lower bound

The geometric work now enters only through the serial certificate and the admissible window scale. All constants and lower thresholds on κ are chosen from these two results before the prescribed schedule is considered.

Proof of Lem. 2.1. Fix the absolute constants $C_{\mathsf{amp}}, C_{\mathsf{win}}$ supplied by Lem. C.7. By Prop. A.11, after increasing one absolute threshold on κ, there are absolute $C_{\log q}, C_{q\log\kappa}, C_q > 0$ such that

$$\log q \leq C_{\log q}\log\kappa, \qquad q\log\kappa/(\kappa-1) \leq C_{q\log\kappa}, \qquad q \geq C_q\mathcal{Q}(\kappa).$$

For a positive schedule, Lem. C.7 supplies one fixed objective and every scalar input in (D.1), with $\mathsf{err} = R_D$ and $\mathsf{rad} = \mathsf{rad}^{[P]}$. Apply Lem. D.1 and use $q \geq C_q\mathcal{Q}(\kappa)$ to obtain $R_D \geq \kappa^{-C_1}\exp[-(C/C_q)N/\mathcal{Q}(\kappa)]$. Thus $\chi_\kappa = \kappa^{-C_1}$ and $C_0 = C/C_q$ are independent of the schedule and its length. All thresholds and constants are fixed before the schedule and its length are chosen. Zero steps leave iterates unchanged and can be deleted; the resulting length is at most N, so its bound implies the same bound with N. An all-zero schedule has $R_D = 1$, for example on $f(x) = x^2/(2\kappa)$ from $x_0 = 1$. □

D.3 Quadratic baseline

Lemma D.2 (Quadratic baseline)**.** *For every $\kappa \geq 2$ and every finite schedule of length $N \geq 0$, a one-dimensional quadratic in $\mathcal{F}_{1/\kappa,1}$ satisfies*

$$R_D \geq \operatorname{sech}^2\left(N\log\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right) \geq \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2N}. \tag{D.3}$$

Proof. This is the classical Chebyshev lower bound; see also the proof of Lemma 2 in Appendix B of the supplementary material of Arjevani & Shamir (2016). For $f_\lambda(x) = \lambda x^2/2$, $\lambda \in [1/\kappa, 1]$, GD from 1 has relative squared distance $\mathrm{poly}(\lambda)^2$, where $\mathrm{poly}(\lambda) = \prod_{j=1}^N(1-\eta_j\lambda)$. For $N = 0$ the conclusion is immediate. Otherwise $\deg\mathrm{poly} \leq N$ and $\mathrm{poly}(0) = 1$. The affine variable $x = (1+\kappa^{-1}-2\lambda)/(1-\kappa^{-1})$ maps the curvature interval onto $[-1,1]$ and zero to $x_{\mathsf{out}} = (\kappa+1)/(\kappa-1) > 1$. The Chebyshev polynomial Cheb_N, defined by $\mathrm{Cheb}_N(\cos\varphi) = \cos(N\varphi)$, has degree N and satisfies $\mathrm{Cheb}_N(\cosh x) = \cosh(Nx)$. The classical Chebyshev minimax inequality gives

$$\max_{\lambda\in[1/\kappa,1]}|\mathrm{poly}(\lambda)| \geq \frac{1}{\mathrm{Cheb}_N(x_{\mathsf{out}})} = \operatorname{sech}\left(N\log\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right).$$

The maximum is attained by continuity, and $\cosh x \leq e^x$ gives the second inequality in (D.3). □

E SILVER-BLOCK CONTRACTION AND PREFIX STABILITY

We prove the two block properties used in Sec. 2.2. First, we show that balanced concatenation preserves a primitive certificate, then use the Silver recursion to derive the endpoint inequality in Lem. 2.3. Next, we apply that inequality at each joining step and prove Lem. 2.4 by induction. Throughout, $f : \mathbb{R}^d \to \mathbb{R}$ is convex and 1-smooth; a minimizer is assumed only where stated.

E.1 PRIMITIVE CERTIFICATES AND BALANCED CONCATENATION

We use the standard smooth convex inequality (Nesterov, 2018, Theorem 2.1.5, Eq. (2.1.10)):

$$f(\boldsymbol{u}) - f(\boldsymbol{v}) + \langle \nabla f(\boldsymbol{u}), \boldsymbol{v} - \boldsymbol{u} \rangle + \tfrac{1}{2}\|\nabla f(\boldsymbol{u}) - \nabla f(\boldsymbol{v})\|^2 \le 0 \qquad (\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d). \tag{E.1}$$

This inequality requires neither a minimizer nor strong convexity.

We use the step sum to specify the joining step and the coefficients of the primitive certificate. For a finite nonnegative schedule $\boldsymbol{\eta}_{1:n}$, write

$$s(\boldsymbol{\eta}_{1:n}) := \sum_{t=1}^{n} \eta_t, \qquad s(()) = 0.$$

To track the inequality preserved by concatenation, run this schedule from $\boldsymbol{x}_0 = \boldsymbol{y}$ and put $f_t = f(\boldsymbol{x}_t)$ and $\boldsymbol{g}_t = \nabla f(\boldsymbol{x}_t)$. Define its certificate by

$$\begin{aligned} \mathcal{P}(\boldsymbol{\eta}_{1:n}; \boldsymbol{y}) = \sum_{t=1}^{n} \eta_t \left(2(f_{t-1} - f_n) + (1 - \eta_t)\|\boldsymbol{g}_{t-1}\|^2\right) \\ - s(\boldsymbol{\eta}_{1:n})\big(s(\boldsymbol{\eta}_{1:n}) + 1\big)\|\boldsymbol{g}_n\|^2. \end{aligned} \tag{E.2}$$

The dependence on f is suppressed; every certificate in a statement uses the same function. Empty schedules are allowed, with $\mathcal{P}((); \boldsymbol{y}) = 0$. Nonnegativity of this certificate will imply an endpoint distance bound in Appx. E.2.

The next lemma follows the equal-step-sum concatenation argument of Zhang & Jiang (2026), specifically Theorem 3.1 and the calculation in Lemma 3.5. Their joining coefficient becomes $-\sigma + \sqrt{2}(\sigma + 1) = 1 + (\sigma + 1)/\rho$ when both step sums equal σ. Their primitive-schedule condition holds uniformly over functions and starting points. Here we need only two certificate inequalities on a fixed trajectory, so we give the calculation establishing that local implication.

Lemma E.1 (Balanced concatenation of primitive certificates)**.** *Fix a convex 1-smooth function $f : \mathbb{R}^d \to \mathbb{R}$. Let $\boldsymbol{\eta}^{\mathsf{L}}$ and $\boldsymbol{\eta}^{\mathsf{R}}$ be finite nonnegative schedules, possibly empty, such that*

$$s(\boldsymbol{\eta}^{\mathsf{L}}) = s(\boldsymbol{\eta}^{\mathsf{R}}).$$

Set

$$\eta^{\mathsf{join}} = 1 + \frac{s(\boldsymbol{\eta}^{\mathsf{L}}) + 1}{\rho}, \qquad \boldsymbol{\eta} = (\boldsymbol{\eta}^{\mathsf{L}}, \eta^{\mathsf{join}}, \boldsymbol{\eta}^{\mathsf{R}}).$$

Run $\boldsymbol{\eta}$ from any $\boldsymbol{y}$, and let $\boldsymbol{w}$ be the point immediately after the joining step. If

$$\mathcal{P}(\boldsymbol{\eta}^{\mathsf{L}}; \boldsymbol{y}) \ge 0, \qquad \mathcal{P}(\boldsymbol{\eta}^{\mathsf{R}}; \boldsymbol{w}) \ge 0,$$

then

$$\mathcal{P}(\boldsymbol{\eta}; \boldsymbol{y}) \ge 0.$$

The two schedules need not have the same length or entries. No minimizer is required.

Proof. Write

$$\ell = |\boldsymbol{\eta}^{\mathsf{L}}|, \quad n = \ell + 1 + |\boldsymbol{\eta}^{\mathsf{R}}|, \quad \sigma = s(\boldsymbol{\eta}^{\mathsf{L}}) = s(\boldsymbol{\eta}^{\mathsf{R}}), \quad \eta = \eta^{\mathsf{join}}.$$

Thus $\boldsymbol{x}_0 = \boldsymbol{y}$, $\boldsymbol{x}_{\ell+1} = \boldsymbol{w} = \boldsymbol{x}_\ell - \eta \boldsymbol{g}_\ell$, and $s(\boldsymbol{\eta}) = 2\sigma + \eta$. In this proof abbreviate

$$\mathcal{P}_{\mathsf{L}} = \mathcal{P}(\boldsymbol{\eta}^{\mathsf{L}}; \boldsymbol{x}_0), \qquad \mathcal{P}_{\mathsf{R}} = \mathcal{P}(\boldsymbol{\eta}^{\mathsf{R}}; \boldsymbol{x}_{\ell+1}).$$

The interpolation residuals from Eq. (E.1) are

$$\mathcal{I}_{i,j} = f_i - f_j + \langle \boldsymbol{g}_i, \boldsymbol{x}_j - \boldsymbol{x}_i \rangle + \tfrac{1}{2}\|\boldsymbol{g}_i - \boldsymbol{g}_j\|^2 \le 0.$$

We combine residuals connecting the right block to its two surrounding points:

$$\mathcal{I}_{\mathsf{R}} = \sum_{i=\ell+1}^{n-1} \eta_{i+1}(\mathcal{I}_{i,\ell} + \mathcal{I}_{i,n}) + (\eta - 1)\mathcal{I}_{\ell,n} + \mathcal{I}_{n,\ell}. \tag{E.3}$$

All weights are nonnegative: the schedule is nonnegative and $\eta - 1 = (\sigma + 1)/\rho > 0$. Hence $\mathcal{I}_{\mathsf{R}} \le 0$.

We now expand (E.3) without using a minimizer. GD on the right block gives

$$\sum_{i=\ell+1}^{n-1} \eta_{i+1}\boldsymbol{g}_i = \boldsymbol{x}_{\ell+1} - \boldsymbol{x}_n,$$

$$\sum_{i=\ell+1}^{n-1} \left(2\eta_{i+1}\langle \boldsymbol{g}_i, \boldsymbol{x}_i \rangle - \eta_{i+1}^2\|\boldsymbol{g}_i\|^2\right) = \|\boldsymbol{x}_{\ell+1}\|^2 - \|\boldsymbol{x}_n\|^2.$$

Expanding the paired residuals with these two identities yields

$$\begin{aligned} &\sum_{i=\ell+1}^{n-1} \eta_{i+1}(\mathcal{I}_{i,\ell} + \mathcal{I}_{i,n}) \\ &= \mathcal{P}_{\mathsf{R}} - \sigma(f_\ell - f_n) + \tfrac{\sigma}{2}\|\boldsymbol{g}_\ell\|^2 + \left(\sigma(\sigma+1) + \tfrac{\sigma}{2}\right)\|\boldsymbol{g}_n\|^2 \\ &\quad + \langle \boldsymbol{x}_{\ell+1} - \boldsymbol{x}_n, (\eta - 1)\boldsymbol{g}_\ell - \boldsymbol{g}_n \rangle. \end{aligned} \tag{E.4}$$

For the two remaining residuals, use $\boldsymbol{x}_\ell - \boldsymbol{x}_n = \eta\boldsymbol{g}_\ell + \boldsymbol{x}_{\ell+1} - \boldsymbol{x}_n$ to obtain

$$\begin{aligned} (\eta - 1)\mathcal{I}_{\ell,n} + \mathcal{I}_{n,\ell} = (\eta - 2)(f_\ell - f_n) + \left(\tfrac{3\eta}{2} - \eta^2\right)\|\boldsymbol{g}_\ell\|^2 + \tfrac{\eta}{2}\|\boldsymbol{g}_n\|^2 \\ + \langle \boldsymbol{x}_{\ell+1} - \boldsymbol{x}_n, \boldsymbol{g}_n - (\eta - 1)\boldsymbol{g}_\ell \rangle. \end{aligned} \tag{E.5}$$

The inner products cancel when (E.4) and (E.5) are added. Therefore

$$\begin{aligned} \mathcal{I}_{\mathsf{R}} = \mathcal{P}_{\mathsf{R}} - (\sigma + 2 - \eta)(f_\ell - f_n) \\ + \tfrac{1}{2}(\sigma + 3\eta - 2\eta^2)\|\boldsymbol{g}_\ell\|^2 + \tfrac{1}{2}(2\sigma^2 + 3\sigma + \eta)\|\boldsymbol{g}_n\|^2. \end{aligned} \tag{E.6}$$

Independently, splitting the definition of the full certificate into the left block, joining step, and right block gives

$$\begin{aligned} \mathcal{P}(\boldsymbol{\eta}; \boldsymbol{x}_0) = \mathcal{P}_{\mathsf{L}} + \mathcal{P}_{\mathsf{R}} + 2(\sigma + \eta)(f_\ell - f_n) \\ + \left[\sigma(\sigma+1) + \eta - \eta^2\right]\|\boldsymbol{g}_\ell\|^2 \\ + \left[\sigma(\sigma+1) - (2\sigma + \eta)(2\sigma + \eta + 1)\right]\|\boldsymbol{g}_n\|^2. \end{aligned} \tag{E.7}$$

Here the left block's terminal function value is f_ℓ, the right block's is f_n, and the full step sum is $2\sigma + \eta$.

Substitute $\eta = 1 + (\sigma + 1)/\rho$ and use $\rho^2 = 2\rho + 1$. The coefficients satisfy

$$\begin{aligned} \rho(\sigma + 2 - \eta) &= \sigma + \eta, \\ \rho(\sigma + 3\eta - 2\eta^2) &= \eta^2 - \eta - \sigma(\sigma + 1), \\ \rho(2\sigma^2 + 3\sigma + \eta) &= (2\sigma + \eta)(2\sigma + \eta + 1) - \sigma(\sigma + 1). \end{aligned} \tag{E.8}$$

For example, $\eta = \sqrt{2}(\sigma + 1) - \sigma$, so $\sigma + 2 - \eta = (2 - \sqrt{2})(\sigma + 1) > 0$; the other two identities follow by expanding the same affine expression for η.

Combining (E.6)–(E.8) proves the exact identity

$$\mathcal{P}(\boldsymbol{\eta}; \boldsymbol{x}_0) = \mathcal{P}_{\mathsf{L}} + (1 + 2\rho)\mathcal{P}_{\mathsf{R}} - 2\rho\,\mathcal{I}_{\mathsf{R}}. \tag{E.9}$$

Each term on the right is nonnegative under the hypotheses, so the conclusion follows. All formulas allow empty sums. If one block is empty, its step sum is zero, and equality of step sums forces every step in the other block to be zero; the same proof still applies. □

When a minimizer exists, let

$$\mathcal{I}_{t-1,\star} = f_{t-1} - f_\star - \langle \boldsymbol{g}_{t-1}, \boldsymbol{x}_{t-1} - \boldsymbol{x}^\star \rangle + \tfrac{1}{2}\|\boldsymbol{g}_{t-1}\|^2.$$

The GD distance identity gives

$$\begin{aligned}\mathcal{P}(\boldsymbol{\eta}_{1:n}; \boldsymbol{x}_0) = \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2 - \|\boldsymbol{x}_n - \boldsymbol{x}^\star\|^2 + 2\sum_{t=1}^{n} \eta_t \mathcal{I}_{t-1,\star} \\ - 2s(\boldsymbol{\eta}_{1:n})(f_n - f_\star) - s(\boldsymbol{\eta}_{1:n})\big(s(\boldsymbol{\eta}_{1:n}) + 1\big)\|\boldsymbol{g}_n\|^2.\end{aligned}$$

Thus $\mathcal{P}/2$ is the slack in the primitive inequality of Zhang & Jiang (2026, Definition 2.7), with smoothness normalized to 1. The uniform quantifier in their definition and the two particular certificate evaluations in Lem. E.1 remain distinct.

E.2 The Silver endpoint inequality

The recursion defining S_k gives $|S_{k+1}| = 2|S_k| + 1$ and

$$s(S_{k+1}) = 2s(S_k) + 1 + \rho^{k-1}.$$

Starting from $S_0 = ()$, these recurrences give $|S_k| = 2^k - 1$ and $s(S_k) = \rho^k - 1$, since $\rho^2 = 2\rho + 1$. We now specialize the primitive certificate to these blocks, as in the endpoint argument of Zhang & Jiang (2026).

Proof of Lem. 2.3. First prove $\mathcal{P}(S_k; \boldsymbol{y}) \geq 0$ for every starting point $\boldsymbol{y}$ by induction on k. The empty block gives the case $k = 0$. If the assertion holds for S_k, both copies in S_{k+1} satisfy it at their respective starting points. Their step sums equal $s(S_k) = \rho^k - 1$, and the joining step in Lem. E.1 is

$$1 + \frac{s(S_k) + 1}{\rho} = 1 + \rho^{k-1} = \eta_k^{\mathsf{join}}.$$

Lem. E.1 therefore proves the assertion for S_{k+1}. Uniformity over starting points is part of this induction; it permits applying the assertion to the second copy at its actual starting point.

Next, for any finite schedule whose certificate is nonnegative, apply Eq. (E.1) with $(\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{x}_{t-1}, \boldsymbol{x}^\star)$. Since $\nabla f(\boldsymbol{x}^\star) = \boldsymbol{0}$,

$$\langle \boldsymbol{g}_{t-1}, \boldsymbol{x}_{t-1} - \boldsymbol{x}^\star \rangle \geq f_{t-1} - f_\star + \tfrac{1}{2}\|\boldsymbol{g}_{t-1}\|^2.$$

Multiplying by $2\eta_t \geq 0$ and using the GD distance identity gives

$$\begin{aligned}\|\boldsymbol{x}_{t-1} - \boldsymbol{x}^\star\|^2 - \|\boldsymbol{x}_t - \boldsymbol{x}^\star\|^2 &= 2\eta_t \langle \boldsymbol{g}_{t-1}, \boldsymbol{x}_{t-1} - \boldsymbol{x}^\star \rangle - \eta_t^2 \|\boldsymbol{g}_{t-1}\|^2 \\ &\geq 2\eta_t (f_{t-1} - f_\star) + \eta_t(1 - \eta_t)\|\boldsymbol{g}_{t-1}\|^2.\end{aligned}$$

Summing and using the definition of $\mathcal{P}$ yields

$$\begin{aligned}\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2 - \|\boldsymbol{x}_n - \boldsymbol{x}^\star\|^2 \geq 2s(\boldsymbol{\eta}_{1:n})(f_n - f_\star) \\ + s(\boldsymbol{\eta}_{1:n})\big(s(\boldsymbol{\eta}_{1:n}) + 1\big)\|\boldsymbol{g}_n\|^2 + \mathcal{P}(\boldsymbol{\eta}_{1:n}; \boldsymbol{x}_0).\end{aligned}$$

Take $\boldsymbol{\eta}_{1:n} = S_k$, $\boldsymbol{x}_0 = \boldsymbol{y}$, and $\boldsymbol{x}_n = \boldsymbol{z}$. The certificate is nonnegative by the induction, and $s(S_k) = \rho^k - 1$, proving the claim. For $k = 0$, both additional terms vanish and $\boldsymbol{z} = \boldsymbol{y}$. Equivalently, the bound is

$$\|\boldsymbol{z} - \boldsymbol{x}^\star\|^2 + 2(\rho^k - 1)(f(\boldsymbol{z}) - f_\star) + \rho^k(\rho^k - 1)\|\nabla f(\boldsymbol{z})\|^2 \leq \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2. \tag{E.10}$$

□

For $f \in \mathcal{F}_{1/\kappa,1}$, strong convexity gives $2(f(\boldsymbol{z}) - f_\star) \geq \kappa^{-1}\|\boldsymbol{z} - \boldsymbol{x}^\star\|^2$. Substituting into Eq. (E.10) and dropping the nonnegative gradient term yields the block contraction used in Sec. 2.2:

$$\|\boldsymbol{z} - \boldsymbol{x}^\star\|^2 \leq \frac{\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2}{1 + (\rho^k - 1)/\kappa}. \tag{E.11}$$

E.3 JOINING-STEP CONTROL AND PREFIX STABILITY

The joining step can exceed two, so we use the preceding block's endpoint inequality to control its distance from the minimizer. This calculation supplies the middle step in the induction over the two copies of the Silver block.

Proof of Lem. 2.4. Induct on k, with the assertion quantified over every starting point. The empty block S_0 gives the base case. Assume the assertion for S_k, and run

$$S_{k+1} = (S_k, \eta_k^{\mathsf{join}}, S_k)$$

from $\boldsymbol{y}$. Let $\boldsymbol{z}$ be the endpoint of the left copy and let $\boldsymbol{w} = \boldsymbol{z} - \eta_k^{\mathsf{join}} \nabla f(\boldsymbol{z})$. For $0 \le t \le 2^k - 1$, the induction hypothesis on the left copy gives $\|\boldsymbol{y}_t - \boldsymbol{x}^\star\|^2 \le \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2$.

To control the joining step, write

$$\sigma = s(S_k) = \rho^k - 1, \qquad \eta = \eta_k^{\mathsf{join}} = 1 + \rho^{k-1}, \qquad \boldsymbol{g} = \nabla f(\boldsymbol{z}).$$

Then $\sigma \ge 0$, $\boldsymbol{w} = \boldsymbol{z} - \eta \boldsymbol{g}$, and $0 < \eta \le 1 + \rho^k = \sigma + 2$. Apply (E.1) to $(\boldsymbol{z}, \boldsymbol{x}^\star)$ and $(\boldsymbol{x}^\star, \boldsymbol{z})$, respectively, to obtain

$$\langle \boldsymbol{g}, \boldsymbol{z} - \boldsymbol{x}^\star \rangle \ge f(\boldsymbol{z}) - f_\star + \tfrac{1}{2}\|\boldsymbol{g}\|^2, \qquad f(\boldsymbol{z}) - f_\star \ge \tfrac{1}{2}\|\boldsymbol{g}\|^2.$$

Expanding the squared distance and applying the first inequality, followed by Lem. 2.3 at $\boldsymbol{z}$, gives

$$\begin{aligned}
\|\boldsymbol{w} - \boldsymbol{x}^\star\|^2 &= \|\boldsymbol{z} - \boldsymbol{x}^\star\|^2 - 2\eta\langle \boldsymbol{g}, \boldsymbol{z} - \boldsymbol{x}^\star \rangle + \eta^2 \|\boldsymbol{g}\|^2 \\
&\le \|\boldsymbol{z} - \boldsymbol{x}^\star\|^2 - 2\eta(f(\boldsymbol{z}) - f_\star) + (\eta^2 - \eta)\|\boldsymbol{g}\|^2 \\
&\le \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2 - 2(\sigma + \eta)(f(\boldsymbol{z}) - f_\star) + \left[\eta^2 - \eta - \sigma(\sigma + 1)\right]\|\boldsymbol{g}\|^2.
\end{aligned}$$

Since $\sigma + \eta \ge 0$, the second interpolation inequality bounds the last two terms by

$$\begin{aligned}
&\left[\eta^2 - \eta - \sigma(\sigma + 1) - (\sigma + \eta)\right]\|\boldsymbol{g}\|^2 \\
&\qquad = (\sigma + \eta)(\eta - \sigma - 2)\|\boldsymbol{g}\|^2 \le 0.
\end{aligned}$$

Consequently,

$$\|\boldsymbol{w} - \boldsymbol{x}^\star\|^2 \le \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2.$$

This is the desired bound at time 2^k, just after the joining step. Finally, the induction hypothesis on the right copy, starting from $\boldsymbol{w}$, gives

$$\|\boldsymbol{y}_{2^k + r} - \boldsymbol{x}^\star\|^2 \le \|\boldsymbol{w} - \boldsymbol{x}^\star\|^2 \le \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2 \quad (0 \le r \le 2^k - 1).$$

These ranges cover every prefix of S_{k+1}. When $k = 0$, the two copies are empty and $\sigma = 0$; the same joining-step calculation applies without division by σ. □

F Other Convergence Metrics

We retain the function class, nonnegative step-size schedules fixed in advance, and the terminal iterate of Sec. 1.1. Function value gap and squared gradient norm admit two different initial normalizations. We state their bounds separately, since the normalization changes the accuracy dependence. The upper bounds use the periodic Silver schedule of Thm. 1.3 and the endpoint inequality (E.10); we give the metric conversions explicitly.

F.1 Relative function value gap and relative squared gradient norm

In addition to R_D from (1.2), define

$$R_E(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) = \frac{f(\boldsymbol{x}_N) - f_\star}{f(\boldsymbol{x}_0) - f_\star},$$
$$R_G(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) = \frac{\|\nabla f(\boldsymbol{x}_N)\|^2}{\|\nabla f(\boldsymbol{x}_0)\|^2},$$
$$R^{\sup}_{M,\kappa}(\boldsymbol{\eta}_{1:N}) = \sup_{d, f, \boldsymbol{x}_0 \neq \boldsymbol{x}^\star} R_M(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}), \quad M \in \{E, G\}.$$

The suprema range over $d \in \mathbb{N}$, $f \in \mathcal{F}_{1/\kappa,1}(\mathbb{R}^d)$, and $\boldsymbol{x}_0 \neq \boldsymbol{x}^\star$, as for $R^{\sup}_{D,\kappa}$ under the normalization $L = 1$. We omit the arguments of R_D, R_E, R_G when the instance and schedule are fixed. Define T^M_{na} and $T^M_{\boldsymbol{\eta}}$ by replacing $R^{\sup}_{D,\kappa}$ with $R^{\sup}_{M,\kappa}$ in Eq. (1.3) and (1.4). Equivalently, T^M_{na} is the shortest horizon for which some finite schedule guarantees error at most ε, whereas $T^M_{\boldsymbol{\eta}}$ is the earliest threshold after which every prefix of the fixed infinite schedule $\boldsymbol{\eta}$ guarantees that error. An empty feasible set gives $+\infty$. Strong convexity makes each initial denominator positive. For all $f \in \mathcal{F}_{1/\kappa,1}$,

$$\frac{\|\boldsymbol{x} - \boldsymbol{x}^\star\|^2}{2\kappa} \leq f(\boldsymbol{x}) - f(\boldsymbol{x}^\star) \leq \frac{\|\boldsymbol{x} - \boldsymbol{x}^\star\|^2}{2}, \qquad \frac{\|\boldsymbol{x} - \boldsymbol{x}^\star\|}{\kappa} \leq \|\nabla f(\boldsymbol{x})\| \leq \|\boldsymbol{x} - \boldsymbol{x}^\star\|. \tag{F.1}$$

The basic inequalities (F.1) imply, on every fixed trajectory,

$$\kappa^{-1} R_D \leq R_E \leq \kappa R_D, \qquad \kappa^{-2} R_D \leq R_G \leq \kappa^2 R_D. \tag{F.2}$$

The comparison factors depend on κ but not on the schedule or its length. To remove these factors from the lower bounds, we extend the finite extraction argument of Prop. 2.2 to relative errors.

Proposition F.1. *Fix $\kappa \geq 2$, $0 < \chi \leq 1$, and $C' > 0$. For every $d \in \mathbb{N}$ and $f \in \mathcal{F}_{1/\kappa,1}(\mathbb{R}^d)$, fix an error function $\mathsf{err}_f : \mathbb{R}^d \to [0, \infty)$ such that $\mathsf{err}_f(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x}$ is the minimizer of f. Assume that, for every $m \in \mathbb{N}$ and every schedule $\tilde{\boldsymbol{\eta}}_{1:m} \in [0, \infty)^m$, there exist a finite-dimensional $\tilde{f} \in \mathcal{F}_{1/\kappa,1}$ and an initial point $\tilde{\boldsymbol{x}}_0$ distinct from its minimizer whose GD iterates satisfy*

$$\mathsf{err}_{\tilde{f}}(\tilde{\boldsymbol{x}}_m)/\mathsf{err}_{\tilde{f}}(\tilde{\boldsymbol{x}}_0) \geq \chi e^{-C'm}.$$

Then, for every $C_{\mathsf{rep}} > 0$, every $N \in \mathbb{N}$, and every schedule $\boldsymbol{\eta}_{1:N} \in [0, \infty)^N$, there exist a finite-dimensional $f \in \mathcal{F}_{1/\kappa,1}$ and an initial point $\boldsymbol{x}_0 \neq \boldsymbol{x}^\star$ whose GD iterates satisfy

$$\mathsf{err}_f(\boldsymbol{x}_N)/\mathsf{err}_f(\boldsymbol{x}_0) \geq e^{-(C' + C_{\mathsf{rep}})N}.$$

Proof. Fix $C_{\mathsf{rep}} > 0$, $N \in \mathbb{N}$, and a schedule $\boldsymbol{\eta}_{1:N} \in [0, \infty)^N$. Repeat this schedule $n_{\mathsf{rep}} = \max\{1, \lceil \log(1/\chi)/(C_{\mathsf{rep}} N) \rceil\}$ times and apply the assumption to these $n_{\mathsf{rep}} N$ steps. Keep the resulting objective f fixed, and let $\boldsymbol{y}_j$ be the state after j copies, for $j = 0, \ldots, n_{\mathsf{rep}}$. The initial error is positive, so the assumed lower bound makes the terminal error positive as well. If any block input had zero error, it would be the minimizer; GD would remain there and the terminal error would be zero. Thus every block input has positive error, and the ratios telescope:

$$\prod_{j=0}^{n_{\mathsf{rep}}-1} \frac{\mathsf{err}_f(\boldsymbol{y}_{j+1})}{\mathsf{err}_f(\boldsymbol{y}_j)} = \frac{\mathsf{err}_f(\boldsymbol{y}_{n_{\mathsf{rep}}})}{\mathsf{err}_f(\boldsymbol{y}_0)} \geq \chi e^{-C' n_{\mathsf{rep}} N}.$$

At least one factor is at least $\chi^{1/n_{\mathsf{rep}}} e^{-C'N} \geq e^{-(C' + C_{\mathsf{rep}})N}$, since $\log(1/\chi)/n_{\mathsf{rep}} \leq C_{\mathsf{rep}} N$. Starting from that block's input on the same objective reproduces the block under the original schedule and proves the conclusion. □

The choices $\mathrm{err}_f(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{x}^\star\|^2$, $\mathrm{err}_f(\boldsymbol{x}) = f(\boldsymbol{x}) - f_\star$, and $\mathrm{err}_f(\boldsymbol{x}) = \|\nabla f(\boldsymbol{x})\|^2$ give R_D, R_E, R_G, respectively. Strong convexity ensures that each error vanishes only at the minimizer, so Prop. F.1 applies to all three relative metrics.

For later use, collect the endpoint conversions for one Silver block S_k, $k \geq 1$, from $\boldsymbol{y} \neq \boldsymbol{x}^\star$ to $\boldsymbol{z}$. Write $s_k^{\mathsf{Sil}} = \rho^k - 1$ for the step sum of the block. Dropping nonnegative terms in (E.10) gives the last two rows of Tab. 1; $\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2 \leq 2\kappa(f(\boldsymbol{y}) - f_\star)$ and $\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2 \leq \kappa^2\|\nabla f(\boldsymbol{y})\|^2$ give the relative rows. The distance row is (E.11). If a block starts at the minimizer, it stays there and all unnormalized bounds hold with zero on both sides.

Table 1: Endpoint bounds from the same Silver endpoint inequality. Here $s_k^{\mathsf{Sil}} = \rho^k - 1 > 0$.

Output normalized by block input	Upper bound
$\|\boldsymbol{z} - \boldsymbol{x}^\star\|^2/\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2$	$(1 + s_k^{\mathsf{Sil}}/\kappa)^{-1}$
$(f(\boldsymbol{z}) - f_\star)/(f(\boldsymbol{y}) - f_\star)$	$\kappa/s_k^{\mathsf{Sil}}$
$\|\nabla f(\boldsymbol{z})\|^2/\|\nabla f(\boldsymbol{y})\|^2$	$\kappa^2/[s_k^{\mathsf{Sil}}(s_k^{\mathsf{Sil}} + 1)]$
$2(f(\boldsymbol{z}) - f_\star)/\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2$	$1/s_k^{\mathsf{Sil}}$
$\|\nabla f(\boldsymbol{z})\|^2/\|\boldsymbol{y} - \boldsymbol{x}^\star\|^2$	$1/[s_k^{\mathsf{Sil}}(s_k^{\mathsf{Sil}} + 1)]$

Below, we take a block level k_* differently from the proof of Thm. 1.3:

$$k_* = \lceil \log_\rho(2\kappa + 1) \rceil, \qquad m = 2^{k_*} - 1 < 2(2\kappa + 1)^\nu < 6\kappa^\nu. \tag{F.3}$$

Throughout this appendix, $\boldsymbol{\eta}^{\mathsf{sil}}$ denotes the infinite repetition of this block S_{k_*}, and $\boldsymbol{\eta}^{\mathrm{fin},N}$ consists of $\lfloor N/m \rfloor$ copies of this block followed by zero steps to reach length N.

Corollary F.2 (Bounds for relative metrics)**.** *With the same absolute constant C chosen in the proof of Thm. 1.1, the following hold. For every $\kappa \geq 2$, every finite schedule of length $N \geq 0$, and $M \in \{E, G\}$,*

$$R_{M,\kappa}^{\sup}(\boldsymbol{\eta}_{1:N}) \geq e^{-CN/\mathcal{Q}(\kappa)}.$$

For $N \geq 1$, a fixed finite-dimensional hard instance attains at least this bound, and may depend on the metric. With the period m from Eq. (F.3), the finite schedule $\boldsymbol{\eta}^{\mathrm{fin},N}$ defined above satisfies

$$R_{M,\kappa}^{\sup}(\boldsymbol{\eta}_{1:N}^{\mathrm{fin},N}) \leq 2^{-\lfloor N/m \rfloor}, \tag{F.4}$$

and the single periodic schedule satisfies

$$R_{E,\kappa}^{\sup}(\boldsymbol{\eta}_{1:N}^{\mathsf{sil}}) \leq \kappa\, 2^{-\lfloor N/m \rfloor}, \qquad R_{G,\kappa}^{\sup}(\boldsymbol{\eta}_{1:N}^{\mathsf{sil}}) \leq \kappa^2\, 2^{-\lfloor N/m \rfloor} \quad (N \geq 0). \tag{F.5}$$

Consequently, for $0 < \varepsilon \leq 1/2$ and every infinite schedule $\boldsymbol{\eta}$,

$$\begin{aligned} T_{\mathsf{na}}^M,\, T_{\boldsymbol{\eta}}^M &\geq \left\lceil \frac{\mathcal{Q}(\kappa)}{C} \log(1/\varepsilon) \right\rceil, \\ T_{\mathsf{na}}^M &\leq m \lceil \log_2(1/\varepsilon) \rceil, \\ T_{\boldsymbol{\eta}^{\mathsf{sil}}}^E &\leq m \lceil \log_2(\kappa/\varepsilon) \rceil, \qquad T_{\boldsymbol{\eta}^{\mathsf{sil}}}^G \leq m \lceil \log_2(\kappa^2/\varepsilon) \rceil. \end{aligned}$$

Proof. For $\kappa \geq \kappa_*$, combine Lem. 2.1 with (F.2) and apply Prop. F.1 to either error, with $\chi = \kappa^{-2}\chi_\kappa$ and the same $C' = C_0/\mathcal{Q}(\kappa)$, $C_{\mathsf{rep}} = 1/\mathcal{Q}(\kappa)$ as in Sec. 2. This gives a finite-dimensional hard instance with rate constant $C_0 + 1$. For bounded κ, the quadratic from Lem. D.2 has $R_D = R_E = R_G$, so the same C_{small} applies. Thus $C = \max\{C_0 + 1, C_{\mathsf{small}}\}$ works for all three relative metrics. For $N = 0$, every relative ratio is one. To prove the upper bounds, first check contraction in each metric at a complete block endpoint; then distinguish a prescribed terminal horizon from an arbitrary prefix of the periodic schedule. Since $\rho^{k_*} - 1 \geq 2\kappa$, Tab. 1 contracts each of D, E, G by at least a factor of two. At a prescribed horizon N, use $\lfloor N/m \rfloor$ complete blocks followed by $N - m\lfloor N/m \rfloor$ zero

steps to obtain Eq. (F.4). This finite schedule is common to D, E, G. For the periodic schedule, an unfinished block can have a larger function-value gap or gradient norm than its input. Its distance is nevertheless controlled by (2.5); combining that estimate with Eq. (F.2) gives Eq. (F.5). Inverting the bounds proves the iteration statements; an anytime threshold includes its own prefix. □

The additional $\log \kappa$ in the relative anytime upper bounds is a feature of this guarantee; no lower bound here proves it necessary. The quadratic baseline also holds for E, G: on $f_\lambda(x) = \lambda x^2/2$, all three relative ratios equal $\prod_{t=1}^{N}(1-\eta_t\lambda)^2$. Thus, the same quadratic chosen in Lem. D.2 works for each relative metric.

Combining the two lower bounds gives, for $M \in \{E, G\}$,

$$R^{\sup}_{M,\kappa}(\boldsymbol{\eta}_{1:N}) \geq \max\left\{e^{-CN/\mathcal{Q}(\kappa)}, \operatorname{sech}^2\left(N \log \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)\right\}.$$

The hard instances for the two terms need not coincide. In particular, each relative iteration threshold is also at least $\left\lceil \operatorname{arcosh}(\varepsilon^{-1/2}) / \log[(\sqrt{\kappa}+1)/(\sqrt{\kappa}-1)] \right\rceil$.

F.2 Normalization by initial squared distance

We use the labels E/D and G/D for the function-value gap and the squared gradient norm, normalized by the initial squared distance, respectively. For $M \in \{E/D, G/D\}$, define the pointwise and worst-case quantities

$$R_{E/D}(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) = \frac{2(f(\boldsymbol{x}_N) - f_\star)}{\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2}, \qquad R_{G/D}(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}) = \frac{\|\nabla f(\boldsymbol{x}_N)\|^2}{\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2},$$
$$R^{\sup}_{M,\kappa}(\boldsymbol{\eta}_{1:N}) = \sup_{d, f, \boldsymbol{x}_0 \neq \boldsymbol{x}^\star} R_M(f, \boldsymbol{x}_0; \boldsymbol{\eta}_{1:N}),$$
$$R^{*}_{M,\kappa}(N) = \inf_{\boldsymbol{\eta}_{1:N} \in [0,\infty)^N} R^{\sup}_{M,\kappa}(\boldsymbol{\eta}_{1:N}).$$

The supremum uses the same $d, f, \boldsymbol{x}_0$ as above, and the infimum ranges over all nonnegative length-N schedules, including the empty schedule when $N = 0$. The factor 2 in $R_{E/D}$ makes both worst-case initial values equal to one. To write the bounds for both metrics together, set $\Gamma_{E/D} = 1$ and $\Gamma_{G/D} = 2$. On every trajectory,

$$\kappa^{-\Gamma_M} R_D \leq R_M \leq R_D. \tag{F.6}$$

For general smoothness L, the denominators in $R_{E/D}, R_{G/D}$ are $L\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2$ and $L^2\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2$, respectively. Replacing f by f/L and each step by $L\eta_t$ preserves these dimensionless quantities. Thus, bounds for the unscaled gap and squared gradient norm, divided by initial squared distance, acquire factors $L/2$ and L^2.

These normalizations by initial squared distance do not telescope under repetition. For example, on $f_\lambda(x) = \lambda x^2/2$, take $\lambda \in [1/\kappa, 1)$ and a block multiplier $\operatorname{poly}(\lambda) = \prod_{t=1}^{N}(1-\eta_t\lambda) \neq 0$. Over $n_{\mathsf{rep}} \geq 2$ copies, the total error is $\lambda^{\Gamma_M} \operatorname{poly}(\lambda)^{2n_{\mathsf{rep}}}$, whereas the product of the n_{rep} block errors is $\lambda^{\Gamma_M n_{\mathsf{rep}}} \operatorname{poly}(\lambda)^{2n_{\mathsf{rep}}}$. Hence Prop. F.1 cannot remove the factor $\kappa^{-\Gamma_M}$ in (F.6).

Define T^M_{na} and $T^M_{\boldsymbol{\eta}}$ by replacing $R^{\sup}_{D,\kappa}$ with $R^{\sup}_{M,\kappa}$ in Eqs. (1.3) & (1.4).

Proposition F.3 (Two accuracy regimes)**.** *For $\kappa \geq 4$, $0 < \varepsilon \leq 1/16$, and $M \in \{E/D, G/D\}$, set $\kappa_{M,\varepsilon} = \min\{\kappa, \varepsilon^{-1/\Gamma_M}\}$. With absolute implicit constants,*

$$T^M_{\mathsf{na}}(\kappa, \varepsilon) \gtrsim \mathcal{Q}(\kappa_{M,\varepsilon}) \left(1 + \log \frac{1}{\kappa_{M,\varepsilon}^{\Gamma_M}\varepsilon}\right),$$
$$T^M_{\mathsf{na}}(\kappa, \varepsilon) \lesssim \kappa_{M,\varepsilon}^{\nu} \left(1 + \log \frac{1}{\kappa_{M,\varepsilon}^{\Gamma_M}\varepsilon}\right).$$

The lower bound also holds for every anytime threshold $T^M_{\boldsymbol{\eta}}$. The accuracy transition is $\varepsilon = \kappa^{-1}$ for function value gap and $\varepsilon = \kappa^{-2}$ for squared gradient norm.

Proof. For $2 \le \kappa' \le \kappa$, the inclusion $\mathcal{F}_{1/\kappa',1} \subseteq \mathcal{F}_{1/\kappa,1}$ lets us apply Thm. 1.1 and (F.6) on each smaller class. Every fixed schedule has worst-case error at least $(\kappa')^{-\Gamma_M} e^{-CN/\mathcal{Q}(\kappa')}$ for every such κ'. The bound is independent of the schedule, so taking the supremum over κ' and then the infimum over schedules gives

$$R^*_{M,\kappa}(N) \ge \sup_{2\le\kappa'\le\kappa} (\kappa')^{-\Gamma_M} e^{-CN/\mathcal{Q}(\kappa')} \ge \kappa^{-\Gamma_M} e^{-CN/\mathcal{Q}(\kappa)}. \tag{F.7}$$

Consequently, both stopping models satisfy

$$T^M_{\mathsf{na}}(\kappa,\varepsilon),\ T^M_{\boldsymbol{\eta}}(\kappa,\varepsilon) \ge \frac{1}{C} \sup_{2\le\kappa'\le\kappa} \mathcal{Q}(\kappa') \max\left\{\log \frac{1}{(\kappa')^{\Gamma_M}\varepsilon}, 0\right\}.$$

Since $\kappa_{M,\varepsilon} \ge 4$, choose $\kappa' = \kappa_{M,\varepsilon}/2 \in [2,\kappa]$; using $\kappa_{M,\varepsilon}$ itself could make the logarithm vanish at the transition. Then

$$\begin{aligned} \mathcal{Q}(\kappa_{M,\varepsilon}/2) &\ge 2^{-\nu}\mathcal{Q}(\kappa_{M,\varepsilon}), \\ \log \frac{1}{(\kappa_{M,\varepsilon}/2)^{\Gamma_M}\varepsilon} &= \Gamma_M \log 2 + \log \frac{1}{\kappa_{M,\varepsilon}^{\Gamma_M}\varepsilon} \\ &\ge (\log 2)\left(1 + \log \frac{1}{\kappa_{M,\varepsilon}^{\Gamma_M}\varepsilon}\right), \end{aligned}$$

since the last logarithm is nonnegative. This proves the lower bound.

For the upper bound, if $\kappa_{M,\varepsilon} = \varepsilon^{-1/\Gamma_M}$, use the shortest Silver block with $s^{\mathsf{Sil}}_k = \rho^k - 1 \ge \kappa_{M,\varepsilon}$. The last two rows of Tab. 1 give $R_M \le (s^{\mathsf{Sil}}_k)^{-\Gamma_M} \le \varepsilon$. The block has length $2^k - 1 < 2(1+\kappa_{M,\varepsilon})^\nu = O(\kappa^\nu_{M,\varepsilon})$, and $\log(1/(\kappa^{\Gamma_M}_{M,\varepsilon}\varepsilon)) = 0$. If $\kappa_{M,\varepsilon} = \kappa$, repeat the block S_{k_*} with $s^{\mathsf{Sil}}_{k_*} = \rho^{k_*} - 1 \ge 2\kappa$. The first $j-1$ copies contract squared distance by $2^{-(j-1)}$; the final copy converts that distance to the output metric, giving

$$R^{\sup}_{M,\kappa}(\boldsymbol{\eta}^{\mathrm{fin},N}_{1:N}) \le (s^{\mathsf{Sil}}_{k_*})^{-\Gamma_M} 2^{-(j-1)} \le \kappa^{-\Gamma_M} 2^{-j}, \qquad j = \lfloor N/m \rfloor \ge 1. \tag{F.8}$$

Zero padding preserves the endpoint. Choose $j = \max\{1, \lceil \log_2(1/(\kappa^{\Gamma_M}\varepsilon))\rceil\}$ and $N = jm$ to get error at most ε in $O(\kappa^\nu[1 + \log(1/(\kappa^{\Gamma_M}\varepsilon))])$ steps. □

Fixed-horizon bounds. The proof of (F.7) holds for every $\kappa \ge 2$ and $N \ge 0$, and (F.8) holds for every $N \ge m$. Since $\lfloor N/m \rfloor \ge N/m - 1$ and $m < 6\kappa^\nu$, they give

$$\kappa^{-\Gamma_M} e^{-CN/\mathcal{Q}(\kappa)} \le R^*_{M,\kappa}(N) \le 2\kappa^{-\Gamma_M} e^{-(\log 2)N/(6\kappa^\nu)} \qquad (N \ge m).$$

For general L, μ, the lower prefactors for the unscaled gap and squared gradient norm are $\mu/2$ and μ^2, respectively. For every $N \ge 1$, one can instead use S_k with $k = \lfloor \log_2(N+1) \rfloor \ge 1$ and zero padding. Since $(N+1)/2 < 2^k \le N+1$ and $\rho^k - 1 \asymp (2^k)^\alpha$, the last two rows of Tab. 1 give

$$R^*_{M,\kappa}(N) = O((N+1)^{-\Gamma_M \alpha}),$$

with an absolute constant independent of κ. Thus the non-anytime complexity stays bounded as κ increases at fixed accuracy; a uniform lower bound of order $\mathcal{Q}(\kappa)\log(1/\varepsilon)$ cannot hold for this normalization.

Anytime guarantee. For the single periodic schedule, use (2.5) and $R_M \le R_D$ to obtain

$$R^{\sup}_{M,\kappa}(\boldsymbol{\eta}^{\mathsf{sil}}_{1:N}) \le 2^{-\lfloor N/m \rfloor} \qquad (N \ge 0).$$

For $\kappa \ge 2$ and $0 < \varepsilon \le 1/2$, this gives

$$T^M_{\boldsymbol{\eta}^{\mathsf{sil}}}(\kappa,\varepsilon) \le m\,\lceil \log_2(1/\varepsilon) \rceil = O(\kappa^\nu \log(1/\varepsilon)).$$

Unlike the relative E, G guarantee, this upper bound has no added $\log\kappa$. The sharper upper bound in Prop. F.3 uses endpoints of complete Silver blocks and does not control every intermediate iterate.

G SUMMARY OF NOTATION

The tables collect symbols shared across the main arguments. Local variables are defined where they are used. Unless stated otherwise, $L = 1$ and $\mu = 1/\kappa$.

PROBLEM, RATES, AND CONVERGENCE METRICS

See Secs. 1.1 & 2.2 & Appx. F.

Symbol	Meaning
$\mu, L, \kappa, \mathcal{F}_{\mu,L}$	Strong convexity and smoothness constants, condition number L/μ, and the corresponding function class.
$\boldsymbol{x}_t^{[j]}, \boldsymbol{z}_t$	Block j of the full iterate, and the local iterate of one unscaled phase core. Its scalar coordinates are $z_t^{(i)}$.
$\boldsymbol{y}, \widehat{\boldsymbol{z}}_{t_{\mathsf{loc}}}$	Normalized two-coordinate transfer input $(z^{(i)}, z^{(i+1)})/D_i$, and normalized block $\boldsymbol{x}^{[j]}_{t^{[j]}_{\mathsf{start}}+t_{\mathsf{loc}}}/A^{[j]}$.
$\boldsymbol{x}_t, \boldsymbol{x}^\star, f_\star$	GD iterate, minimizer, and minimum value.
$\boldsymbol{\eta}, \eta_t, \boldsymbol{\eta}_{1:N}$	Prescribed schedule, its step at update t, and its length-N prefix; $N = 0$ gives the empty schedule.
ρ, ν, α	$\rho = 1 + \sqrt{2}$, $\nu = \log_\rho 2$, and $\alpha = \log_2 \rho = 1/\nu$.
$\gamma_*, \mathcal{Q}(\kappa)$	Logarithmic exponent $\pi^2\nu/(8\sqrt{2}\rho \log \rho)$ and lower-bound time scale $\kappa^\nu/(\log \kappa)^{\gamma_*}$.
C	Absolute rate constant in Thm. 1.1.
$\mathsf{err}_f, \mathsf{err}$	Error function in Prop. F.1 and scalar error in Lem. D.1; e denotes the natural exponential constant.
R_D, R_E, R_G	Relative squared distance to the minimizer, relative function value gap, and relative squared gradient norm; each output error is divided by the same error at the initial point.
$R^{\sup}_{D,\kappa}, R^{\sup}_{E,\kappa}, R^{\sup}_{G,\kappa}$	Worst-case relative errors for a fixed schedule, over all finite dimensions, admissible functions, and nonoptimal initial points.
$T^M_{\mathsf{na}}, T^M_{\boldsymbol{\eta}}$	For $M \in \{D, E, G, E/D, G/D\}$, the shortest successful finite horizon and the threshold after which every prefix of $\boldsymbol{\eta}$ succeeds.
$R_{E/D}, R_{G/D}$	$2(f(\boldsymbol{x}_N) - f_\star)$ and $\|\nabla f(\boldsymbol{x}_N)\|^2$, divided by initial squared distance; for general L, also divide by L and L^2, respectively.
$R^{\sup}_{M,\kappa}, R^*_{M,\kappa}(N)$	Worst-case R_M for a fixed schedule and its infimum over length-N schedules, $M \in \{E/D, G/D\}$.
Γ_M	$\Gamma_{E/D} = 1$ and $\Gamma_{G/D} = 2$, the exponents in the bounds for normalization by initial squared distance.
$\kappa_{M,\varepsilon}$	Effective condition number $\min\{\kappa, \varepsilon^{-1/\Gamma_M}\}$ in Prop. F.3.
Silver schedules (Sec. 2.2 & Appx. E)	
$S_k, \eta_k^{\mathsf{join}}$	Silver block and joining step: $S_0 = ()$, $S_{k+1} = (S_k, \eta_k^{\mathsf{join}}, S_k)$, $\eta_k^{\mathsf{join}} = 1 + \rho^{k-1}$; k is the block level.
$2^k - 1, s(S_k) = \rho^k - 1$	Block length and step sum, where $s(\boldsymbol{\eta}_{1:n}) = \sum_{t=1}^n \eta_t$ and $s(()) = 0$.
k_*, m	Block index $\lceil \log_\rho(\kappa + 1) \rceil$ in Sec. 2.2 and $\lceil \log_\rho(2\kappa + 1) \rceil$ in Appx. F; period $m = 2^{k_*} - 1$.
$\boldsymbol{\eta}^{\mathsf{sil}}$	Infinite repetition of S_{k_*} with the block index chosen in the respective section; divide the steps by L for general smoothness.
$\boldsymbol{\eta}^{\mathrm{fin},N}$	In Appx. F, $\lfloor N/m \rfloor$ complete blocks followed by zero steps to reach the prescribed horizon N.

TRANSFERS, CHECKPOINTS, AND COUPLED CORES

See Secs. 3 & 4 and their proof appendices.

Symbol	Meaning
$\mu_{\mathsf{c}}, \beta_{\mathsf{c}}$	Proof abbreviations for $4/(4\kappa-1)$ and $2(\kappa-2)/(4\kappa-1)$; core smoothness is $4(\kappa-1)/(4\kappa-1)$.
$\mathcal{C}_\zeta, \mathcal{E}_{\mathcal{C}_\zeta}, g_i$	Spiral gradient hull, its projection envelope, and a transfer function with input $(1,0)$; $\nabla\mathcal{E}_{\mathcal{C}_\zeta} = \mathrm{proj}_{\mathcal{C}_\zeta}$ (Appx. B.1).
G_i	Transfer function at amplitude D_i: $G_i(\boldsymbol{y}) = D_i^2 g_i(\boldsymbol{y}/D_i)$, applied to $(z^{(i)}, z^{(i+1)})$ in the core.
$\zeta_i, a_{\zeta_i}, a_{\min}, c(a)$	Pitch and startup threshold of transfer i, $a_{\min} = a_{1/4}$, and the geometric cost correction; proofs may write $a_i := a_{\zeta_i}$ (Eq. (A.1)).
D_i, B_i, τ_i	Formal transfer amplitude, transfer factor D_{i+1}/D_i, and ordinary threshold $\tau_i = \tau_{\zeta_i}$. For an ordinary transfer, this threshold bounds the gap-end second coordinate and derivative threshold relative to the output amplitude. At matching, the formal amplitudes are actual checkpoint amplitudes; D_i is distinct from the block scale $A^{[j]}$.
$\mathsf{Trials}, \boldsymbol{p}, \mathsf{Shoot}$	Controlled trial rectangle, trial parameter, and shooting self-map (Lem. B.5).
$\mathcal{T}, t_i, b_i, s_i$	Checkpoint set, its ordered positions, sizes $b_i = \eta_{t_i}$, and preceding unselected gap sums.
$s_{\mathsf{tot}}, s_{\mathsf{gap}}$	Total schedule step sum and step sum left unselected by the initial window checkpoint set, respectively; s_{gap} remains fixed during repair.
d_j	Number of checkpoints in phase j; its core uses $d_j + 1$ coordinates. This is $k = \lvert\mathcal{T}\rvert$ in the general selection problem.
$\mathcal{N}, \mathcal{N}_b, \mathcal{N}_\infty$	Terminal numerator, eligible checkpoint numerator, and uncapped checkpoint numerator (Eqs. (A.6) & (A.17)).
$s_{\mathsf{res}}, s_{\mathsf{L}}, s_{\mathsf{R}}$	Reserve $C_{\mathsf{res}} a_{\min}$ and virtual left and right step-sum buffers in checkpoint selection.
$\Psi_0(\mathcal{T}), \Psi(\mathcal{T})$	Selection cost with zero boundary buffers (Eq. (3.2)) and its buffered version (Eq. (A.10)).
W, W_n, q	Step-sum majorant, $W_n = W(n)$, and window length with $q = \Theta(\mathcal{Q}(\kappa))$ (Prop. A.11 & Eq. (A.21)).
$\mathsf{amp}(n)$	Common quadratic radius for every trial and norm lower bound at every matched phase iterate: $(\kappa-1)^{-C_{\mathsf{amp}}} e^{-C_{\mathsf{win}}\lceil n/q\rceil}$, for absolute $C_{\mathsf{amp}} > 0, C_{\mathsf{win}} \geq 2$.
$h_{\boldsymbol{p}}(\boldsymbol{z}), \boldsymbol{g}_*$	Unscaled trial core and trial-independent initial gradient $\nabla h_{\boldsymbol{p}}(\boldsymbol{e}^{(1)})$ (Lem. B.8); matched dynamics are in Lem. B.9.
Serial construction (Appx. C)	
$P, \boldsymbol{x} = (\boldsymbol{x}^{[0]}, \ldots, \boldsymbol{x}^{[P]})$	Number of extensions and full state partitioned into $P+1$ coordinate blocks.
$\boldsymbol{x}^{[0:j]}$	Concatenation $(\boldsymbol{x}^{[0]}, \ldots, \boldsymbol{x}^{[j]})$ of blocks $0, \ldots, j$; the same notation applies to iterates, initial points, and displacements.
$h^{[j]}$	Unscaled core for phase j (Eq. (3.1)).
$A^{[j]}, \boldsymbol{x}^{[j]}_{\mathsf{init}}, F^{[j]}$	Block scale, initial point $\boldsymbol{x}^{[j]}_{\mathsf{init}} = A^{[j]}\boldsymbol{e}^{(1)}$, and scaled core $F^{[j]}(\boldsymbol{x}^{[j]}) = (A^{[j]})^2 h^{[j]}(\boldsymbol{x}^{[j]}/A^{[j]})$; $A^{[0]} = 1$.
$n^{[j]}, \mathsf{rad}^{[j]}, \ell^{[j]}$	Assigned phase length, radius $\mathsf{rad}^{[j]} = A^{[j]}\mathsf{amp}(n^{[j]})/4$, and readout level $\ell^{[j]} = \mathsf{rad}^{[j]}/\sqrt{8(\kappa-1)+1}$.
$\mathsf{Weight}^{[j]}, r^{[j]}$	Weighted map and scalar readout on $\boldsymbol{x}^{[0:j]}$; $r^{[j]}(\boldsymbol{x}^{[0:j]}) = \ell^{[j]}$ exactly when $\lVert\mathsf{Weight}^{[j]}\boldsymbol{x}^{[0:j]}\rVert = \mathsf{rad}^{[j]}$ (Appx. C.1).
$H^{[j]}, \theta^{[j]}, \boldsymbol{g}^{[j]}$	Incoming coupling $H^{[j]}(\boldsymbol{x}^{[0:j]})$, coefficient $\theta^{[j]}(\boldsymbol{x}^{[0:j]})$ of its correction $-\theta^{[j]}\boldsymbol{g}^{[j]}$, and initial scaled-core gradient $\boldsymbol{g}^{[j]} = \nabla F^{[j]}(\boldsymbol{x}^{[j]}_{\mathsf{init}})$.
$\ell_k^-, \ell_k^+, n^{[j]}_{\mathsf{wait}}$	Release-band levels and waiting length between phases; the band index is $k^{[j]} = n^{[j]}_{\mathsf{wait}}$.
f	Fixed objective $f(\boldsymbol{x}) = \sum_{j=0}^{P} F^{[j]}(\boldsymbol{x}^{[j]}) + \sum_{j=1}^{P} H^{[j]}(\boldsymbol{x}^{[0:j]})$, in $\mathcal{F}_{1/\kappa,1}$.
$\kappa_*, C_0, \chi_\kappa$	Absolute construction threshold, rate constant, and positive prefactor independent of horizon and schedule (Lem. 2.1).